COSMOLOGY

Prof. Dr. Alan H. Guth

Lecture Notes

Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge, MA, USA

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — Wednesday, September 5, 2018
Prof. Alan Guth

COURSE INFORMATION

INSTRUCTOR: Alan H. Guth, Room 6-322, Ext. 3-6265, guth@ctp.mit.edu.

TEACHING ASSISTANT: Honggeun Kim, hgkim@mit.edu.

LECTURE HOURS: Mondays and Wednesdays, 11:05 a.m. – 12:25 a.m., in Room 4-231.

REQUIRED TEXTBOOKS:

Introduction to Cosmology, Second Edition (Cambridge University Press, 2016), by Barbara Ryden.

The First Three Minutes, 2nd paper edition (Basic Books, 1993), by Steven Weinberg.

RECOMMENDED BOOKS:

An Introduction to Modern Cosmology, 2nd Edition (Wiley, 2003), by Andrew Liddle.

The Inflationary Universe (Perseus Books, 1997), by Alan H. Guth. This was written as a popular-level book, and therefore has no equations. It does not, however, shy away from trying to explain the relevant principles of physics and their logical connections. It attempts a kind of story-telling flavor, describing the history of twentieth century cosmology, and also the story of my own involvement in cosmology. The course will in no way follow this book, but you might like it.

LECTURE NOTES AND OTHER READING:

There is no textbook that I know of that is really appropriate for the intended content of this course, although Barbara Ryden's book, **Introduction to Cosmology**, comes much closer than any book I have seen previously. Steven Weinberg's **The First Three Minutes** is a superbly written book which gives an excellent description of cosmology in general, and the synthesis of the light chemical elements in particular. But it does not describe the mathematical details. It has a mathematical appendix, but the description there is very sketchy. We will try to fill in some of the mathematics behind Weinberg's descriptions in class.

The bulk of the course, nonetheless, will be based on lecture notes that will be posted periodically on the course website. The material in these lecture notes will be essential for doing the problem sets and quizzes, and will form the backbone of the course. (Incidentally, David Kaiser and I are currently working on an undergraduate textbook on cosmology, which will be mainly based on these lecture notes.)

For the first part of the course (classical cosmology), the lectures and the associated lecture notes will describe the subject at a level of detail that is much more mathematical than Weinberg's book, and a little beyond the level of Ryden's book. For the second part of the course (modern particle physics and its recent impact on cosmology), we will rely mostly on the lecture notes, although Ryden does have a good chapter on inflation. You will also be asked to read several articles from *Scientific American* or similar publications.

GRADING:

75% of the course grade will be based on quizzes, which will be given in class during the normal lecture period. There will be three of these quizzes, and there will be no final exam. The remaining 25% of the grade will be based on problem sets. Problem sets will normally be assigned every week, but there will be some breaks due to holidays and in-class quizzes. There will be 9 or 10 problem sets altogether.

TENTIVE DATES FOR IN-CLASS QUIZZES:

1) Wednesday, October 3, 2018 (8 preceding classes)
2) Monday, November 5, 2018 (7 classes since first quiz)
3) Wednesday, December 5, 2018 (7 classes since second quiz)

SPECIAL RELATIVITY:

I think that many of you have studied some special relativity, but special relativity is not a prerequisite for this course. For the benefit of those who have not studied special relativity, the basic results are summarized in Lecture Notes 1. I expect that you will be able to understand and occasionally use these statements, but we will not discuss how they are derived. For those who are interested, a few references are mentioned in Lecture Notes 1. I would be happy to talk to students outside of class about how the results of special relativity are derived, or anything else about special relativity. There will be a few more results from special relativity that will be needed as the course progresses ($E = mc^2$, for example), and I will try to point them out and summarize them carefully as we go along.

COURSE OUTLINE:

1. Doppler Effect (and a little Special Relativity)
2. Kinematics of Newtonian Cosmology
3. Dynamics of Newtonian Cosmology
4. Introduction to Non-Euclidean Spaces
5. Black-Body Radiation and the Early History of the Universe
6. The Accelerating Universe and the Cosmological Constant
7. Big-Bang Nucleosynthesis
8. Problems of the Conventional (Non-Inflationary) Hot Big Bang Model
9. Grand Unified Theories and the Magnetic Monopole Problem
10. The Inflationary Universe Model
11. Primordial Density Fluctuations and the Cosmic Microwave Background
12. Eternal Inflation and the Multiverse

HOMEWORK LOGISTICS

Problem sets will ordinarily be due at 5:00 pm on Fridays, to be turned in at the homework boxes at the intersection of buildings 8 and 16 (3rd floor bldg. 8, 4th floor bldg.16). You may also email your problem sets, sending them both to hgkim@mit.edu and guth@ctp.mit.edu. The first problem set will be due on Friday, September 14, 2018.

The problem sets will not all be assigned the same number of points. Your final problem set grade will be the total number of points you receive, divided by the number of points possible. Problem sets with more assigned points, therefore, will count more toward your grade.

All problem sets will count, none will be dropped. My reason for this policy is that I feel that the problem sets are an important component of the course, so I want to encourage you to do every one of them. However, I am fully aware that MIT students are active people who lead complicated lives, and that these complications can make it hard to turn in a problem set every week at 5 pm on Friday. So, to make up for the fact that no problem set grades will be dropped, I will be generous with extensions, while still expecting students to do all the problem sets during the term. If you find that you are having an unusually busy week and cannot fit in the 8.286 problem set, I'm okay with giving you an extension — just send an email describing the situation, and ask me for an extension.

HOMEWORK POLICY:

In this course I regard the problem sets primarily as an educational experience, rather than a mechanism of evaluation. I have allocated 25% of the grade to problem sets in order to encourage you to do them, and to make life easier for students who find it difficult to do well on quizzes. You should feel free to work on these problems in groups, and I would strongly encourage you to do so. With the right mix of students, the homework can be more fun and more illuminating. I will in fact soon be setting up a Class Contact webpage to help you make contact with each other.

However, it is important pedagogically that each student write up the solution independently. The simple copying of a friend's paper is not the kind of effort that the grading is intended to encourage. Using 8.286 solutions that have been circulated in previous years is strictly off limits. Using other sources, such as other textbooks or web documents, is considered perfectly okay, as long as you write up the solution in your own words.

A homework problem which appears to be copied from another student, from a solution circulated in a previous term, or copied more or less verbatim from some other source (without rewriting in your own words) will be given a reduced grade, possibly a zero. Except in blatant cases, however, students will be given a warning the first time this happens, and will be given an opportunity to redo the relevant solutions. Since the homework is intended primarily for learning, and not evaluation, there is nothing that you can do on the homework — in this course — that will lead to an interview with the Committee on Discipline. I say this because I want to strongly encourage you to work in groups on the homework, and I don't want you to feel that there are any hidden dangers. (Remember, however, that you should not assume that this policy holds in other classes; different professors have different points of view on these issues.)

MORE ADVANCED READING:

There are some excellent graduate-level textbooks on cosmology that some of you might want to look at. These books are well beyond the level of this course, but I mention them in case any of you become interested in pursuing some topic at a more advanced level. The first two are written from the astrophysical point of view, while the last five describe the early universe more from the particle physicists' slant:

Cosmological Physics (Cambridge University Press, 1999), by John A. Peacock.

Principles of Physical Cosmology (Princeton University Press, 1993), by P.J.E. Peebles.

Cosmology (Oxford University Press, 2008), by Steven Weinberg.

Modern Cosmology (Academic Press, 2003), by Scott Dodelson.

Physical Foundations of Cosmology (Cambridge University Press, 2005), by Viatcheslav Mukhanov.

The Early Universe (Addison-Wesley, 1990), by Edward W. Kolb and Michael S. Turner.

Particle Physics and Inflationary Cosmology (Harwood Academic publishers, 1990), by Andrei Linde.

THE COURSE WEBSITE:

$$\boxed{\text{http://web.mit.edu/8.286/www}}$$

We will use the Gradebook of the Stellar system, but all course information will be posted at the URL above.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — September 6, 2018
Prof. Alan Guth

Lecture Notes 1

THE DOPPLER EFFECT AND SPECIAL RELATIVITY

INTRODUCTION:

Probably the centerpiece of modern cosmology is what is usually called Hubble's law, attributed to a classic 1929 paper by Edwin Hubble.* The law states that all the distant galaxies are receding from us, with a recession velocity given by

$$v = Hr \ . \tag{1.1}$$

Here

$$v \equiv \text{recession velocity},$$

$$H \equiv \text{the Hubble constant},$$

$$\text{and } r \equiv \text{distance to galaxy}.$$

Starting about 2011 there has been some degree of dispute about the attribution of Hubble's law, because it turns out that the law was stated clearly in 1927 by the Belgian priest Georges Lemaître,† who deduced it theoretically from a model of an expanding universe, and estimated a value for the expansion rate based on published astronomical observations. It was certainly Hubble, however, who developed the observational case for what we now call Hubble's law. (We at MIT, however, have every reason to tout the contributions of Lemaître, who in the same year, 1927, received a Ph.D. in physics from MIT.) The controversy over the attribution of Hubble's law has led to a fascinating literature discussing paragraphs mysteriously missing from the English translation of Lemaître's 1927 paper, and ultimately the resolution of that mystery. The interested reader can pursue the links provided in the footnotes.¶ In any case, it seems clear that

---

* Edwin Hubble, "A relation between distance and radial velocity among extra-galactic nebulae," Proceedings of the National Academy of Science, vol. 15, pp. 168-173 (1929).

† Georges Lemaître, "Un Univers homogène de masse constante et de rayon croissant, rendant compte de la vitesse radiale des nébuleuses extra-galactiques," Annales de la Société Scientifique de Bruxelles, vol. A47, pp. 49-59 (1927). Translated into English as "A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae," Monthly Notices of the Royal Astronomical Society, vol. 91, pp. 483-490 (1931).

¶ See, for example, "Edwin Hubble in translation trouble," http://www.nature.com/news/2011/110627/full/news.2011.385.html#B5, and also "Hubble cleared," http://www.nature.com/nature/journal/v479/n7372/full/479150a.html.

Hubble's law will continue to be called Hubble's law, and that seems right to me. The question of whether the universe is expanding or not is really an observational one, and it was Hubble who made the first of these observations.

Later we will begin to talk about the implications of Hubble's law for cosmology, but for now I just want to discuss how the two ingredients — velocities and distances — are measured. Here we will consider the measurement of the velocities, which is done by means of the Doppler shift. The other ingredient in Hubble's law, the cosmic distance ladder, is described in Chapter 2 of Weinberg's **The First Three Minutes**, and will not be discussed in these notes. (You are expected, however, to learn about it from the reading assignment. It is also discussed in Sec. 7.4 of Ryden's **Introduction to Cosmology**, but we will not be reading that until later in the course, if at all.)

The Doppler shift formula for light requires special relativity, which is not a prerequisite for this course. For this course it will be sufficient for you to know the basic consequences of special relativity, which will be stated in these notes. If you would like to learn more about special relativity, however, you could look at **Special Relativity,** by Anthony P. French, **Introduction to Special Relativity,** by Robert Resnick, or Lecture Notes I and II of the 2009 Lecture Notes for this course.

THE NONRELATIVISTIC DOPPLER SHIFT:

It is a well-known fact that atoms emit and absorb radiation only at certain fixed wavelengths (or equivalently, at certain fixed frequencies). This fact was not understood until the development of quantum theory in the 1920's, but it was known considerably earlier. In 1814-15 the Munich optician Joseph Frauenhofer allowed sunlight to pass through a slit and then a glass prism, and noticed that the spectrum which was formed contained a pattern of hundreds of dark lines, which were always found at the same colors. Today we attribute these dark lines to the selective absorption by the cooler atoms in the atmosphere of the sun. In 1868 Sir William Huggins noticed that a very similar pattern of lines could be seen in the spectra of some bright stars, but that the lines were displaced from their usual positions by a small amount. He realized that this shift was presumably caused by the Doppler effect, and used it as a measurement of the velocity of these distant stars.

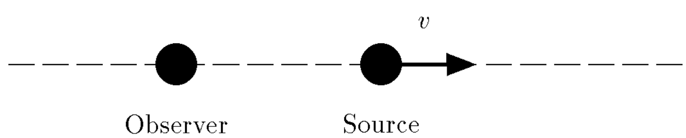
As long as the velocities of the stars in question are small compared to that of light, it is sufficient to use a nonrelativistic analysis. We will begin with the nonrelativistic case, and afterward we will discuss how the calculation is changed by the implications of special relativity. To keep the language manifestly nonrelativistic for now, let us consider first the Doppler shift of sound waves. Suppose for now that the source is moving and the observer is standing still (relative to the air), with all motion taking place along a line. We will let

$$u \equiv \text{velocity of sound waves,}$$

$$v \equiv \text{recession velocity of the source,}$$

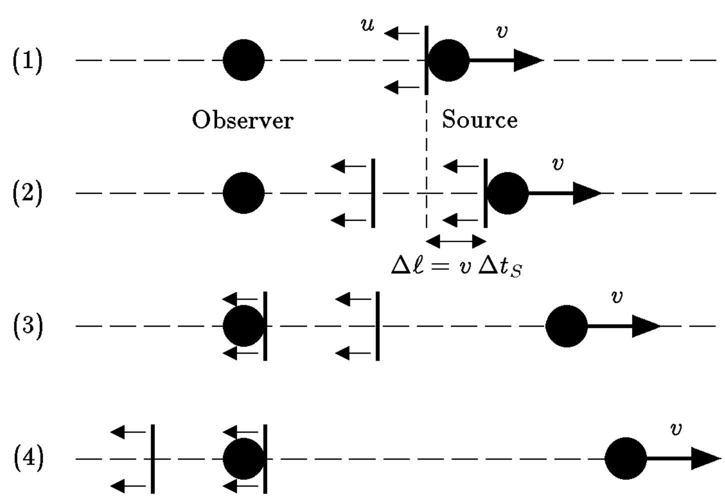
$$\Delta t_S \equiv \text{the period of the wave at the source,}$$

$$\Delta t_O \equiv \text{the period of the wave as observed.}$$



Now consider the following sequence, as illustrated below:

(1) The source emits a wave crest.
(2) At a time Δt_S later, the source emits a second wave crest. During this time interval the source has moved a distance $\Delta\ell = v\Delta t_S$ further away from the observer.
(3) The stationary observer receives the first wave crest.
(4) At some time Δt_O after (3), the observer receives the second wave crest. Our goal is to find Δt_O.



The time at which the first wave crest is received depends of course on the distance between the source and the observer, which was not specified in the description above.

We are interested, however, only in the time difference Δt_O between the reception of the first and second wave crests. This time difference does not depend on the distance between the source and the observer, since both wave crests have to travel this distance. The second crest, however, has to travel an extra distance

$$\Delta\ell = v\,\Delta t_S \ , \tag{1.2}$$

since the source moves this distance between the emission of the two crests. The extra time that it takes the second crest to travel this distance is $\Delta\ell/u$, so the time between the reception of the two crests is

$$\begin{aligned} \Delta t_O &= \Delta t_S + \frac{\Delta\ell}{u} \\ &= \Delta t_S + \frac{v\,\Delta t_S}{u} \\ &= \left(1 + \frac{v}{u}\right)\Delta t_S \ . \end{aligned} \tag{1.3}$$

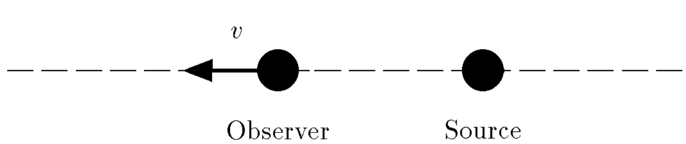
The result is usually described in terms of the "redshift" z, which is defined by the statement that the wavelength is increased by a factor of $(1 + z)$. Since the wavelength λ is related to the period Δt by $\lambda = u\Delta t$, we can write the definition of redshift as

$$\boxed{\frac{\lambda_O}{\lambda_S} = \frac{\Delta t_O}{\Delta t_S} \equiv 1 + z \ ,} \tag{1.4}$$

where λ_S and λ_O are the wavelength as measured at the source and at the observer, respectively. Combining this definition with Eq. (1.3), we find that the redshift for this case is given by

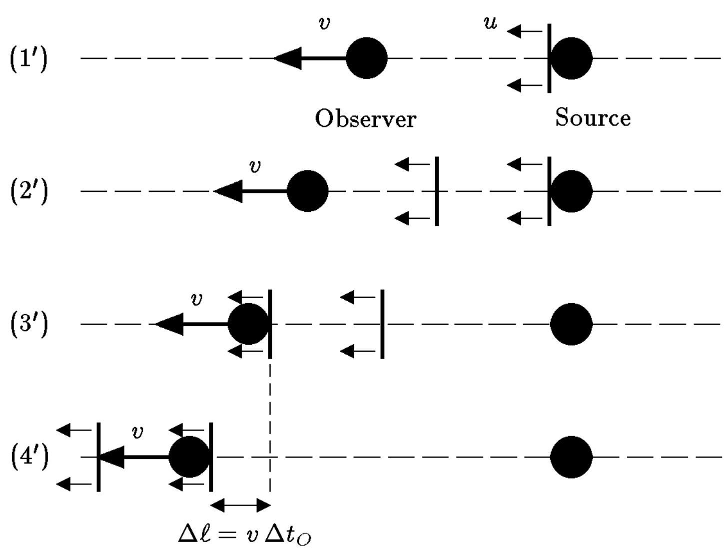
$$\boxed{z = v/u \qquad \text{(nonrelativistic, source moving).}} \tag{1.5}$$

Suppose now that the source stands still, but the observer is receding at a speed v:



In this case, the sequence becomes

(1′) The source emits a wave crest.
(2′) At a time Δt_S later, the source emits a second wave crest. The source is standing still.
(3′) The moving observer receives the first wave crest.
(4′) At a time Δt_O after (3′), the observer receives the second wave crest. During the time interval between (3′) and (4′), the observer has moved a distance $\Delta\ell = v\Delta t_O$ further from the source.



Using the same strategy as in the first case, we note that in this case, the second wave crest must travel an extra distance $\Delta\ell = v\Delta t_O$. Thus,

$$\Delta t_O = \Delta t_S + \frac{\Delta\ell}{u} = \Delta t_S + \frac{v\Delta t_O}{u} \ . \tag{1.6}$$

In this case Δt_O appears on both sides of the equation, but we can easily solve for Δt_O to find

$$\Delta t_O = \left(1 - \frac{v}{u}\right)^{-1} \Delta t_S \ . \tag{1.7}$$

Recalling the definition of z,

$$\begin{aligned} z &= \frac{\Delta t_O}{\Delta t_S} - 1 = \frac{1}{1-(v/u)} - 1 \\ &= \boxed{\frac{v/u}{1-(v/u)} \qquad \text{(nonrelativistic, observer moving).}} \end{aligned} \tag{1.8}$$

Notice that the difference between the two cases is given by

$$z_{\text{observer moving}} - z_{\text{source moving}} = \frac{(v/u)^2}{1-(v/u)} \;, \tag{1.9}$$

which is proportional to $(v/u)^2$. If the speed of recession is much smaller than the wave speed, $v/u \ll 1$, then the difference between the two expressions for z is very small, since it is proportional to the square of the small quantity v/u. Buf if the speed of recession is comparable to the wave speed, then the difference between the two expressions can be very significant.

THE DOPPLER SHIFT FOR LIGHT WAVES:

To derive the Doppler shift for light waves, one must decide which, if either, of the above calculations is applicable.

During the 19th century physicists thought that the situation for light waves was identical to that for sound waves. Sound waves propagate in air, and it was thought that light waves propagate in a medium called the aether which permeates all of space. The aether determines a privileged frame of reference, in which the laws of physics have their simplest form. In particular, Maxwell's equations were believed to have their usual form only in this frame, and it is in this frame that the speed of light was thought to have its standard value of $c = 3.0 \times 10^8$ m/sec in all directions. In a frame of reference which is moving with respect to the aether, the speed of light would be different. Light moving in the same direction as the frame of reference would appear to move more slowly, since the observer would be catching up to it. Light moving in the opposite direction would appear to move faster than normal. Thus, if the source is moving with respect to the aether and the observer is standing still, then the first calculation shown above would apply. If the observer is moving with respect to the aether and the source is standing still, then the second would apply. In either case one would of course replace the sound speed u by the speed of light, c.

In 1905 Albert Einstein published his landmark paper, "On the Electrodynamics of Moving Bodies", in which the theory of special relativity was proposed. The entire concept of the aether, after half a century of development, was removed from our picture of nature. In its place was the principle of relativity: **There exists no privileged frame of reference.** According to this principle, the speed of light will always be measured at the standard value of c, independent of the velocity of the source or the observer. The theory shook the very foundations of physics (which is in general a very risky thing to do), but it has become clear over time that the principle of relativity accurately describes the behavior of nature.

Since special relativity denies the existence of a privileged reference frame, it can make no difference whether it is the source or the observer that is moving. The Doppler shift, and for that matter any physically measurable effect, can depend only on the ***relative*** velocity of source and observer.

THE DEVELOPMENT OF SPECIAL RELATIVITY:

On the face of it, the principle of relativity appears to be self-contradictory. It does not seem possible that the speed of light could be independent of the velocity of the observer. Suppose, for example, that we observe a light pulse which passes us at speed c. Suppose then that a second observer takes off after the light pulse in "super-space-ship" that attains a speed of $0.5c$ relative to us. Surely, one would think, the space ship observer would tend to catch up to the light pulse, and would measure its speed at $0.5c$. How could it possibly be otherwise?

The genius of Albert Einstein is that he was able to figure out how it could be otherwise. The subtlety and the brilliance of the theory lie in the fact that it forces us to change our most fundamental beliefs about the nature of space and time. We have to accept the idea that at high velocities (i.e., velocities not negligible compared to that of light), some of our ingrained intuitions about space and time are no longer valid. In particular, we have to accept the notion that measurements of time intervals, measurements of lengths, and judgments about the simultaneity of events can all depend upon the velocity of the observer. We can, however, maintain our notion about what it means for two events to coincide: if two events appear to occur at the same place ***and*** time to one observer, then they will appear to occur at the same place and time to any observer. (It is standard practice in relativity jargon to use the word "event" to denote a point in spacetime— i.e., an ideal event occurs at a single point in space and at a single instant of time.) In addition, we have no need to change the definition of velocity, $\vec{v} = \mathrm{d}\vec{x}/\mathrm{d}t$, or the resulting equation $\Delta\vec{x} = \vec{v}\,\Delta t$, which holds when $\vec{v}$ is a constant. Furthermore, in contrast to the 19th century viewpoint, we now believe that the fundamental laws of physics have the same form in any inertial reference frame. While measurements of space and time depend on the observer, the fundamental laws of physics are universal.

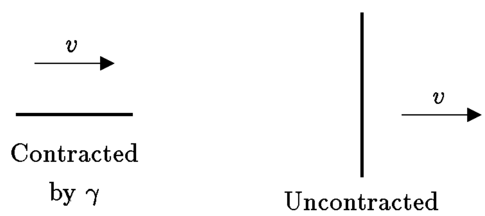
SUMMARY OF SPECIAL RELATIVITY:

We will not discuss the derivation of special relativity here, but the key consequences of special relativity for kinematics — i.e., for measurements of time and distance — can be summarized in three statements. Only the first of these — time dilation — will be needed for the Doppler shift calculation, but I include all three effects for completeness. All three statements use the word "appear," the precise meaning of which will be described later.

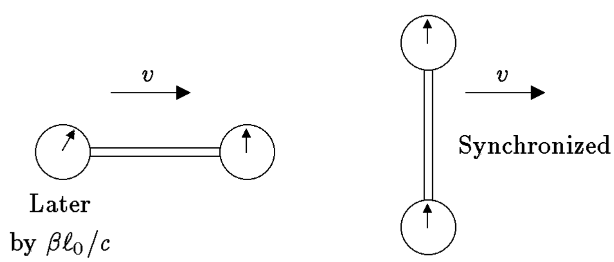
(1) TIME DILATION: Any clock which is moving at speed v relative to a given reference frame will "appear" (to an observer using that reference frame) to run slower than normal by a factor denoted by the Greek letter γ (gamma), and given by

$$\gamma \equiv \frac{1}{\sqrt{1-\beta^2}} \ , \qquad \beta \equiv v/c \ . \tag{1.10}$$

(2) LORENTZ-FITZGERALD CONTRACTION: Any rod which is moving at a speed v along its length relative to a given reference frame will "appear" (to an observer using that reference frame) to be shorter than its normal length by the same factor γ. A rod which is moving perpendicular to its length does not undergo a change in apparent length.



(3) RELATIVITY OF SIMULTANEITY: Suppose a rod which has rest length ℓ_0 is equipped with a clock at each end. The clocks can be synchronized in the rest frame of the system by using light pulses. (That is, a light pulse can be sent out from the center, and the clocks at both ends can be started when they receive the pulses.) If the system moves at speed v along its length, then the trailing clock will "appear" to read a time which is later than the leading clock by an amount $\beta\ell_0/c$. If, on the other hand, the system moves perpendicular to its length, then the synchronization of the clocks is not disturbed.



As mentioned above, the word "appear" in these statements has a special meaning. In plain English, the word "appear" normally refers to the perception of the human eyes. However, in these situations the perception of the human eyes would be very complicated. The complication is that one sees with light, and the speed of light is not infinite. Thus, when you look at an object, the light which you see coming from the parts of the object that are near you has left the object more recently than the light which you see coming from parts of the object that are further. Thus, you are seeing different parts of the object as they were at different times in the past. If the object is static, this makes no difference, but if it is moving, these effects can lead to complicated distortions. These distortions are not taken into account in the statements above. For purposes of interpreting these statements, one can imagine that each reference frame is covered by an infinite number of local observers, each of which observes only events so close that the time delay for light travel is negligible. Each local observer is at rest in the frame, and carries a clock that has been synchronized with the others by light pulses, taking into account the finite speed of light. The "appearance" is then the description that is assembled after the fact by combining the reports of these local observers.

The previous paragraph may sound more complicated then it is, so let's consider a simple example. Suppose that a straight rod is moving along the x-axis of a given reference frame. Suppose further that the positions of the two endpoints of the rod are measured by local observers, as a function of the reference frame time t, and found to be $x_1(t)$ and $x_2(t)$. We would then say that the length of the rod at time t "appears" in this reference frame to be

$$\ell(t) \equiv x_2(t) - x_1(t). \tag{1.11}$$

If $\ell(t)$ has some fixed value ℓ independent of t, then we would say that the rod "appears" to have a fixed length ℓ. We say that the rod "appears" to have this length even though most observers would not actually see this length. For most observers the two ends of the rod would not be equidistant, so the observer would see the location of the two ends at different times.

To complete the summary, we must state that these rules hold only for **inertial** reference frames — they do not hold for rotating or accelerating reference frames. Any reference frame which moves at a uniform velocity relative to an inertial reference frame is also an inertial reference frame.

THE RELATIVISTIC DOPPLER SHIFT:

We can now apply these ideas to the Doppler shift for light. We will first consider the case in which the source is moving relative to our reference frame, with the observer stationary. We will then consider the opposite possibility. The derivations will look very different in these two cases, but the principle of relativity guarantees us that the results must be the same — we are simply describing the same situation from the point of view of two different reference frames.

For the case of the moving source, we can refer back to the nonrelativistic derivation. We describe everything from the point of view of the reference frame shown in the diagrams, in which the observer is at rest. We will refer to this as "our" reference frame. The sequence of events is the same as in the nonrelativistic case, except for step (2). The source is a device that emits wave crests at fixed intervals in time, and hence it is a kind of clock. Since it is moving relative to our frame, it will appear to us to be running slowly, by a factor of γ. But Δt_S still refers to the time as measured on this clock, so the time interval between steps (1) and (2), as measured in our reference frame, is $\gamma\Delta t_S$. Thus, step (2) would read:

(2) At a time $\gamma\Delta t_S$ later, as measured on our clocks, the source emits a second wave crest. During this time interval the source has moved a distance $\Delta\ell = \gamma v\,\Delta t_S$ further away from the observer.

If the two crests traveled the same distance, the time between their reception would be the same as the time between their emission, which in our reference frame is $\gamma\,\Delta t_S$. Taking into account the extra distance $\Delta\ell = \gamma v\,\Delta t_S$ traveled by the second crest, and setting the wave speed u equal to the speed of light c, the time between the reception of the two crests is

$$\begin{aligned}\Delta t_O &= \gamma\Delta t_S + \frac{\Delta\ell}{c} = \gamma\Delta t_S + \frac{\gamma v\,\Delta t_S}{c}\\ &= \gamma\left(1+\frac{v}{c}\right)\Delta t_S = \sqrt{\frac{1+\beta}{1-\beta}}\,\Delta t_S\ .\end{aligned} \tag{1.12}$$

Now consider the case in which the observer is moving, with the source stationary. To describe this case we choose the reference frame of the diagrams ($1'$), etc., in which the source is at rest. We let $\Delta t'$ denote the time interval between the reception of the first and second crest, as measured in *our* frame. The distance that the observer travels between the receipt of the two crests is then given by $\Delta\ell = v\Delta t'$. Following the same strategy as in the nonrelativistic case, we can write $\Delta t'$ as the sum of the time between emissions plus the extra time needed for the second crest to travel the extra distances. Thus,

$$\Delta t' = \Delta t_S + \frac{v\,\Delta t'}{c}\ , \tag{1.13}$$

which can be solved to give

$$\Delta t' = \left(1-\frac{v}{c}\right)^{-1}\Delta t_S\ . \tag{1.14}$$

But now we must take into account the fact that the clock used by the observer is moving relative to our frame, so it will be running slowly compared to our clocks. Thus, the time Δt_O measured on the observer's clock is given by

$$\Delta t_O = \frac{\Delta t'}{\gamma}\ . \tag{1.15}$$

Combining Eqs. (1.14) and (1.15), we find

$$\Delta t_O = \frac{1}{\gamma}\left(1-\frac{v}{c}\right)^{-1}\Delta t_S = \sqrt{\frac{1+\beta}{1-\beta}}\,\Delta t_S\ . \tag{1.16}$$

As expected, the two answers agree. Eqs. (1.12) or (1.16) describe the relationship in special relativity between the Doppler shift and the velocity of recession. Here v denotes the relative speed between source and observer (assumed to lie on the line which joins the source and observer), and it is ***impossible*** to know which of the two is actually in motion. The quantity z is given by

$$\boxed{z = \sqrt{\frac{1+\beta}{1-\beta}} - 1 \qquad \text{(relativistic).}} \tag{1.17}$$

Now that we have the answer, I mention an important warning. While it is worthwhile for us to understand the special-relativistic Doppler shift, it is not the final picture for cosmology. The cosmological redshift involves also gravity, so it is properly described only in the context of general relativity. The good news, however, is that we will learn enough general relativity in this course to have a full understanding of the cosmological redshift.

ACCELERATING CLOCKS:

I'll close with a short discussion of accelerating clocks. Accelerating clocks are seldom relevant to cosmology, but they often show up in elementary problems in special relativity. There is a widespread rumor that special relativity describes clocks moving at a constant velocity relative to an inertial frame, while general relativity is needed to properly describe an accelerating clock. If you are a victim of this rumor, now is the time track down whoever told it to you and straighten him/her out.

We have learned that special relativity predicts that a moving clock runs slower by a factor of $\gamma = 1/\sqrt{1-\beta^2}$, but what should we say about an accelerating clock? After seeing the wondrous implications of special relativity for the behavior of moving clocks, it is tempting to think that general relativity might give us equally powerful insights about the effects of acceleration. A little common sense, however, is all that is needed to dispel this temptation. Consider, for example, a concrete experiment involving the effects of acceleration on a clock. To make the point, let us consider two clocks in particular. The first is a digital wristwatch — for definiteness, let's make it a data-bank-calculator-alarm-chronograph. For a second clock, let's think about an old-fashioned hourglass. To test the effects of acceleration on these two clocks, we can imagine holding each clock two feet above a concrete floor and then dropping it. (Is there anyone out there who still thinks that general relativity is important to understand the results of this experiment?) I'll admit I haven't actually tried this experiment, but I would guess that the hourglass would smash to smithereens, but that the data-bank-calculator-alarm-chronograph would probably survive the two foot drop.

In case you haven't gotten the drift, the conclusion is that the effects of acceleration on a clock are complicated, and strongly dependent on the details of the clock mechanism. In principle we can know the full equations of motion in our (inertial) reference frame, and these equations can be solved to describe the evolution of both the hourglass clock and the data-bank-calculator-alarm-chronograph as they hit the floor. While nature obeys a symmetry — Lorentz invariance — which determines the effect of uniform motion on a clock, there is no symmetry that determines the effect of acceleration.

It is possible to *define* an ideal clock, which runs at a rate that is unaffected by acceleration. That is, one can define an ideal clock as one that runs at the same rate as a nonaccelerating clock that is instantaneously moving at the same velocity. A truly ideal clock is impossible to construct, but there is nothing in principle that prevents one from coming arbitrarily close. Since acceleration (unlike uniform velocity) is detectable, it is always possible in principle to design a device to compensate for any effects that acceleration might otherwise produce. In any problem on a homework assignment or quiz in 8.286, you should assume that any accelerating clock is an ideal one.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — September 15, 2018
Prof. Alan Guth

Lecture Notes 2

THE KINEMATICS OF A HOMOGENEOUSLY EXPANDING UNIVERSE

INTRODUCTION:

Observational cosmology is of course a rich and complicated subject. It is described to some degree in Barbara Ryden's **Introduction to Cosmology** and in Steven Weinberg's **The First Three Minutes**, and I will not enlarge on that discussion here. I will instead concentrate on the basic results of observational cosmology, and on how we can build a simple mathematical model that incorporates these results. The key properties of the universe, which we will use to build a mathematical model, are the following:

(1) ISOTROPY

Isotropy means the same in all directions. The nearby region, however, is rather anisotropic (i.e., looks different in different directions), since it is dominated by the center of the Virgo supercluster of galaxies, of which our galaxy, the Milky Way, is a part. The center of this supercluster is in the Virgo cluster, approximately 55 million light-years from Earth. However, on scales of several hundred million light-years or more, galaxy counts which were begun by Edwin Hubble in the 1930's show that the density of galaxies is very nearly the same in all directions.

The most striking evidence for the isotropy of the universe comes from the observation of the cosmic microwave background (CMB) radiation, which is interpreted as the remnant heat from the big bang itself. Physicists have measured the temperature of the cosmic background radiation in different directions, and have found it to be extremely uniform. It is just slightly hotter in one direction than in the opposite direction, by about one part in 1000. Even this small discrepancy, however, can be accounted for by assuming that the solar system is moving through the cosmic background radiation, at a speed of about 400 km/s (kilometers/second). Once the effect of this motion is subtracted out, the resulting temperature pattern is uniform in all directions to an accuracy of a few parts in 100,000. * Thus, on the very large scales which are probed by the CMB, the universe is incredibly isotropic, as shown in Fig. 2.1:

---

* P. A. R. Ade et al. (Planck Collaboration), "Planck 2015 results, XIII: Cosmological parameters," Table 4, Column 6, arXiv:1502.01589. The Planck collaboration does not quote a value for $\Delta T/T$, the root-mean-square fractional variation of the CMB temperature, but it can be computed from their best-fit parameters, yielding $\Delta T/T = 4.14\times10^{-5}$.

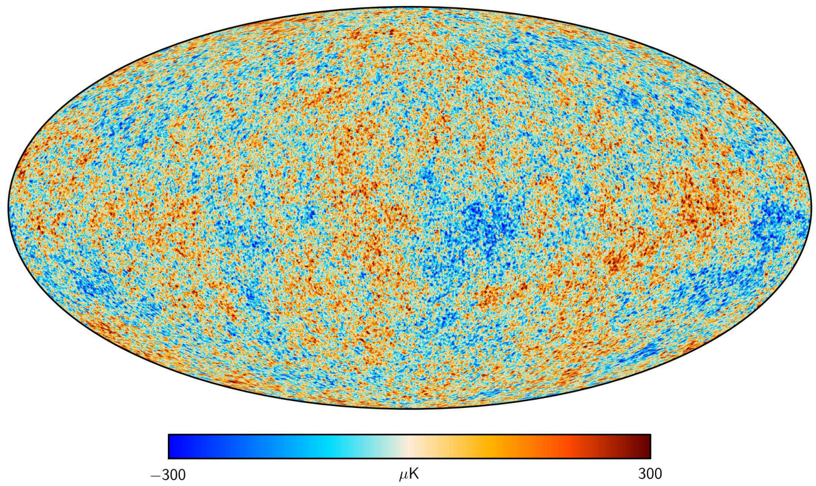


Figure 2.1: The cosmic microwave background radiation as detected by the *Planck* satellite, from the 2015 data release. After correcting for the motion of the Earth, the temperature of the radiation is nearly uniform across the entire sky, with average temperature $T_{\rm cmb} = 2.726$ K. Tiny deviations from the average temperature have been measured; they are so small that they must be depicted in a color scheme that greatly exaggerates the differences, to make them visible. As shown here, blue spots are slightly colder than $T_{\rm cmb}$ while red spots are slightly warmer than $T_{\rm cmb}$, across a range of $\Delta T/T_{\rm cmb} \sim 10^{-4}$ or 10^{-5}.

As an analogy, we can imagine a marble, say about 1 cm across, which is round to an accuracy of four parts in 100,000. That would make its radius constant to an accuracy of 2×10^{-7} m = 200 nm. For comparison, the wavelength of my green laser pointer is 532 nm, so the required accuracy is less than half the wavelength of visible light. Modern technology can certainly produce surfaces with that degree of accuracy, but it corresponds to a good quality photographic lens. In short, it is not easy to achieve spherical symmetry to an accuracy of a few parts in 100,000!

Note that the spherical symmetry stands as strong evidence against the popular misconception of the big bang as a localized explosion which occurred at some particular center. If that were the case, then we would expect the radiation to be hotter in the direction of the center. Thus, the big bang seems to have occurred everywhere. (A localized explosion could look isotropic if we happened to be living at the center, but since the time of Copernicus scientists have viewed with suspicion any assumption that we are at the center of the universe.)

(2) HOMOGENEITY

Homogeneity means the same at all locations. On scales of a few hundred million light-years and larger, the universe is believed to be homogeneous. The observational evidence for homogeneity, however, is not nearly as precise as the evidence for isotropy seen in the CMB. Our belief that the universe is homogeneous, in fact, is motivated significantly by our knowledge of its isotropy. It is conceivable that the universe appears isotropic because all the galaxies are arranged in concentric spheres about us, but such a picture would be at odds with the Copernican paradigm that has been central to our picture of the universe for centuries. So we assume instead that the universe is nearly homogeneous on large scales. That is, we assume that if one observes only large-scale structure, then the universe would look very much the same from any point.

The relationship between the two properties of homogeneity and isotropy is a little subtle. Note that a universe could conceivably be homogeneous without being isotropic — for example, the cosmic background radiation could be hotter in a certain direction, as seen from any point in space, or perhaps the angular momentum vectors of all the galaxies could have a prefered direction. Similarly, a universe could conceivably be isotropic (to one observer) without being homogeneous, if all the matter were arranged on spherical shells centered on the observer. However, if the universe is to be isotropic to all observers, then it must also be homogeneous.

The hypothesis of homogeneity can be tested to some degree of accuracy by galaxy counts. One can estimate the number of galaxies per volume as a function of radial distance from us, and one finds that it appears roughly independent of distance. This kind of analysis is hampered, however, by the difficulty in estimating distances. At large distances it is also hampered by evolution effects — as one looks out in space one is also looking back in time, and the brightness of a galaxy presumably varies with its age. Since we can only see galaxies down to some threshold brightness, the number that we see can depend on how their brightness evolves.

(3) HUBBLE'S LAW

Hubble's law, enunciated theoretically by Georges Lemaître in 1927 and first demonstrated observationally by Edwin Hubble in 1929, states that all the distant galaxies are receding from us, with a recession velocity given by

$$\boxed{v = Hr\ .} \tag{2.1}$$

Here

$$v \equiv \text{recession velocity}\ ,$$

$$H \equiv \text{Hubble expansion rate}\ ,$$

and

$$r \equiv \text{distance to galaxy}\ .$$

For the real universe Hubble's law is a good approximation, and Hubble's law will be an exact property of the mathematical model that we will construct.

The Hubble expansion rate H is often called "the Hubble constant" by astronomers, but it is constant only in the sense that its value changes very little over the lifetime of an astronomer. Over the lifetime of the universe, H varies considerably. The present value of the Hubble expansion rate is denoted by H_0, following a standard convention in cosmology: the present value of any time-dependent quantity is indicated by a subscript "0". Some authors, including Barbara Ryden, reserve the phrase "Hubble constant" for H_0, and refer to the time-dependent $H(t)$ as the "Hubble parameter." To me this is not much of an improvement, since in physics the word "parameter" is most often used to refer to a constant. I will call it the Hubble expansion rate, a terminology that is used by some other sources, including the Particle Data Group*.

For decades, the numerical value of H_0 proved difficult to determine, because of the difficulty in measuring distances. During the 1960s, 70s, and 80s, the Hubble expansion rate was merely known to lie somewhere in the range of

$$H_0 = \frac{0.5 - 1.0}{10^{10} \text{ years}} \ . \tag{2.2}$$

Note that H_0 has the units of 1/time, so that when it is multiplied by a distance it produces a velocity. However, since we rarely in practice talk about velocities in units of such and such a distance per year, H_0 is often quoted in a mixed set of units — for example, $1/(10^{10}$ yr) corresponds to about 30 km/s per million light-years. Astronomers usually quote distances in parsecs rather than light-years, where one parsec is the distance which corresponds to a parallax of 1 second of arc between the Earth and the Sun, when they are separated by their nominal average distance of 1 au (astronomical unit, $149.597870700 \times 10^9$ m), as illustrated at the right. One parsec (abbreviated pc) corresponds to 3.2616 light-years.† Astronomers usually quote the value of the Hubble expansion rate in units of km/s per

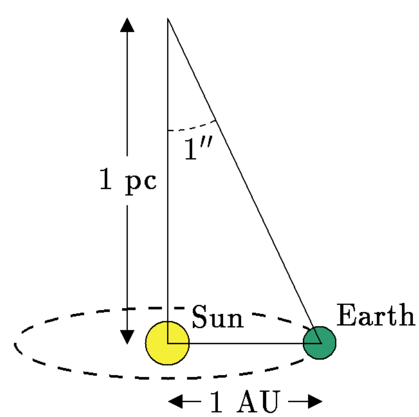


Figure 2.2

---

* *Astrophysical Constants and Parameters*, the Particle Data Group, http://pdg.lbl.gov/2015/reviews/rpp2015-rev-astrophysical-constants.pdf

† One drawback in using light-years is that the definition is tied to that of a year, and the International (SI) System of Units does not specify the definition of a year. This is a significant ambiguity, because the tropical year (vernal equinox to vernal equinox) and the sidereal year (full revolution about the Sun, relative to the fixed stars) differ by a

megaparsec, where 1 megaparsec (Mpc) is a million parsecs. The value of $1/(10^{10}$ yr) is equivalent to 97.8 km-s^{-1}-Mpc^{-1}, so the range of Eq. (2.2) corresponds roughly to a Hubble expansion rate between 50 and 100 km-s^{-1}-Mpc^{-1}. For convenience, astronomers also define the dimensionless quantity h_0 by

$$\boxed{H_0 \equiv h_0 \times (100 \text{ km-s}^{-1}\text{-Mpc}^{-1}) \ .} \tag{2.3}$$

The range of Eq. (2.2) translates into a value of h_0 between $\frac{1}{2}$ and 1.

While the actual value of the Hubble expansion rate certainly changes very little over the lifetime of an astronomer, the same cannot be said for its measured value. Recent precision measurements of the faint anisotropies in the cosmic microwave background radiation, using instruments on the *Planck* satellite, enabled cosmologists to determine*

$$H_0 = 67.66 \pm 0.42 \text{ km-s}^{-1}\text{-Mpc}^{-1} \ , \tag{2.4}$$

which corresponds to a time-scale $H_0^{-1} = 14.4 \pm 0.1$ billion years.† The uncertainty of ± 0.42 km-s^{-1}-Mpc^{-1} in Eq. (2.4), and all uncertainties in H_0 in the following discussion, are given as "1 σ" (one standard deviation) errors. Statistically one expects the correct value to lie inside the uncertainty range 68% of the time, and outside it 32% of the time.

When Hubble first measured the expansion rate, however, he found a value much larger than the value in Eq. (2.4). Due to a very bad estimate of the distance scale, he found $H_0 \sim 500$ km-s^{-1}-Mpc^{-1}, corresponding to $H_0^{-1} \sim 2$ billion years. Hubble's original published graph is reproduced here as Fig. 2.3‡:

---

fractional amount of about 4×10^{-5}. Both drift slowly with time due to changes in the Earth's orbit, and neither agrees with other conventions, such as the Julian or Gregorian years. The International Astronomical Union (IAU), however, does specify the meaning of a year, defining it as a Julian year, exactly 365.25 days (http://www.iau.org/science/publications/proceedings_rules/units/). The day is $24 \times 60 \times 60$ seconds, and the second is defined by atomic standards.

* N. Aghanim et al. (Planck Collaboration), "Planck 2018 results, VI: Cosmological parameters," Table 2, Column 6, arXiv:1807.06209.

† It may not be obvious why measurements of the anisotropies in the CMB should be related in any way to H_0, but cosmologists have developed a detailed theory of how these anisotropies were generated and how they have evolved, which we will pursue later in the course when we discuss inflation. By fitting the predictions of this theory with the observed anisotropies, it is possible to determine the values of a wide range of cosmological parameters, including H_0.

‡ Edwin Hubble, "A Relation Between Distance and Radial Velocity Among Extra-galactic Nebulae," *Proceedings of the National Academy of Science,* vol. 15, pp. 168-173 (1929), http://www.pnas.org/gca?gca=pnas;15/3/168.

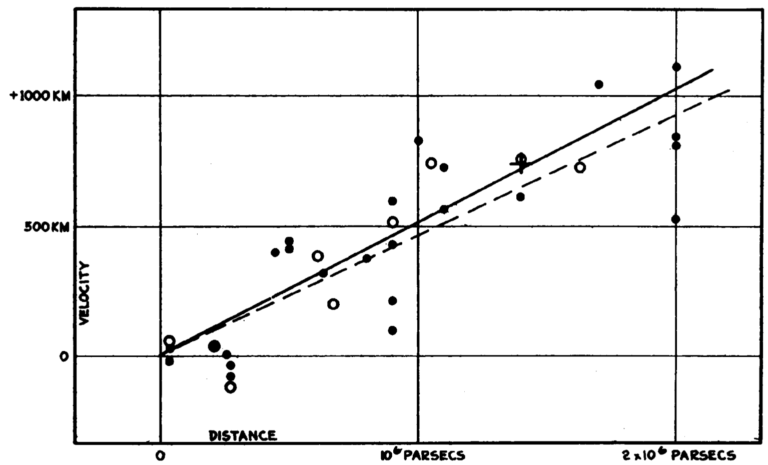


Figure 2.3: Edwin Hubble's original data, published in 1929, which introduced the first observational evidence for Hubble's law and the expansion of the universe.

The horizontal axis in Fig. 2.3 shows the estimated distance to the galaxies, and the vertical axis shows the recession velocity, corrected for the motion of the Sun, in kilometers per second (although it is labeled "km"). Each black dot represents a galaxy, and the solid line shows the best fit to these points. Each open circle represents a group of these galaxies, selected by their proximity in direction and distance; the broken line is the best fit to these points. The cross shows a statistical analysis of 22 galaxies for which individual distance measurements were not available. The evidence for a straight line is not completely convincing, but we must keep in mind that this was only the first paper on the subject. All the galaxies in Hubble's original sample were in fact quite close, so the local velocity perturbations were comparable to the Hubble velocities. Note that 1000 km/s, at the top of Hubble's graph, corresponds to $z \approx 0.03$, while modern tests of Hubble's law extend out to values of z of order 1. Hubble estimated the velocity of the Sun, relative to the mean motion of the galaxies in the sample, to be about 280 km/s, so the solar motion was a significant correction to the data.

After Hubble's original paper, the evidence for the linearity of Hubble's law improved very quickly. In 1931, Hubble and Humason published data that extended to much larger redshift:

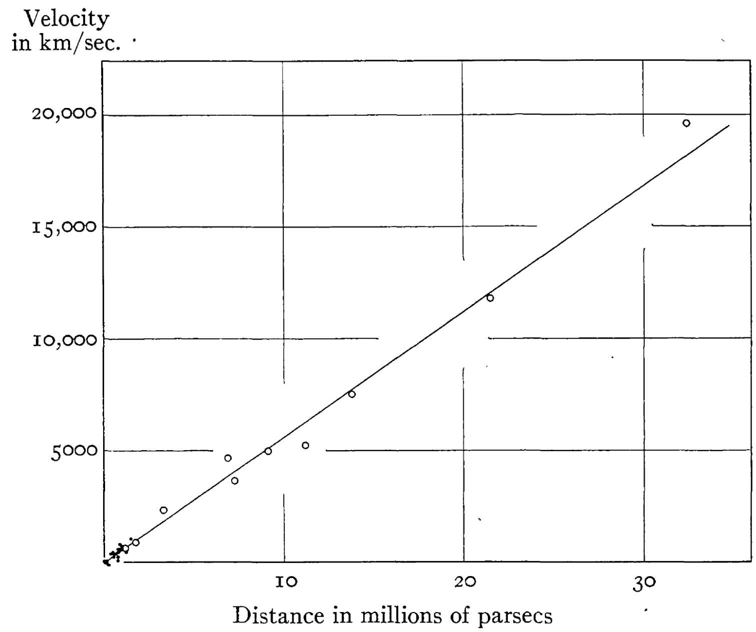


Figure 2.4: Data published by Edwin Hubble and Milton Humason in 1931*, extending Hubble's original measurements to significantly greater distances.

The data from the first paper are shown as dots in the lower left corner, all with velocities less than 1000 km/s. The new value for H_0 was 560 km-s^{-1}-Mpc^{-1}.

As we will see later, a value of the Hubble expansion rate as large as 500 or 560 km-s^{-1}-Mpc^{-1} would imply a very small age for the universe, and the inconsistency of this age with other estimates was a serious problem for big bang theorists for much of the 20th century. It was not until 1958 that the measured value came within the range of Eq. (2.2), primarily due to the work of Walter Baade and Allan Sandage. Summaries of these early measurements may be found in Kragh†, Tamman and Reindl‡, and Kirshner¶.

---

* Edwin Hubble and Milton L. Humason, "The velocity-distance relation among extra-galactic nebulae," Astrophysical Journal, vol. 74, pp. 43–80 (1931), http://adsabs.harvard.edu/abs/1931ApJ....74...43H.

† Helge Kragh, *Cosmology and Controversy: The Historical Development of Two Theories of the Universe* (Princeton: Princeton University Press, 1996).

‡ G. A. Tammann and B. Reindl, in the proceedings of the XXXVIIth Moriond Astrophysics Meeting, *The Cosmological Model,* Les Arcs, France, March 16-23, 2002. Available at http://arXiv.org/abs/astro-ph/0208176.

¶ R. P. Kirshner, "Hubble's diagram and cosmic expansion,“ *Proceedings of the National Academy of Sciences USA*, vol. 101, no. 1, pp. 8-13 (2004), http://www.pnas.org/content/101/1/8.

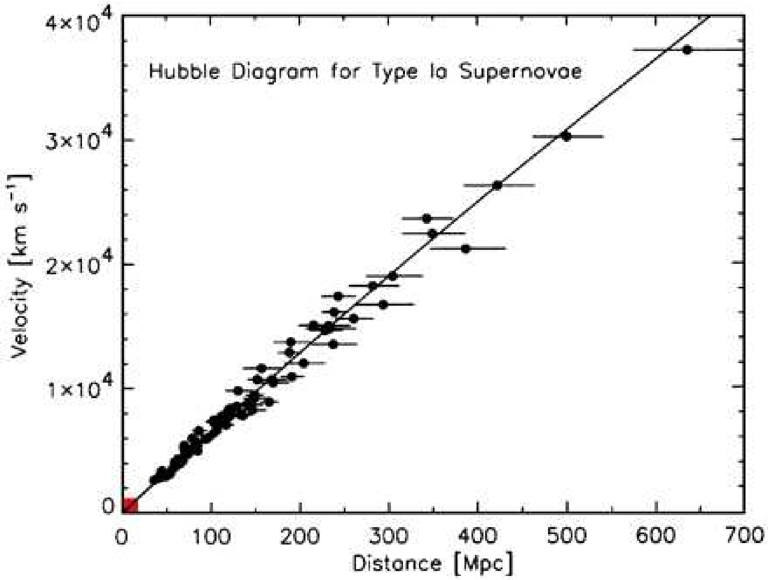


Figure 2.5: An extension of the Hubble diagram, showing observations up to 2002 of Type 1a supernovae. Error bars correspond to uncertainties in determining distances to each object. The small red box near the origin indicates the range covered in Hubble's original plot.

The situation improved dramatically during the 1990s, largely due to the ability of the Hubble Space Telescope to resolve Cepheid variable stars in a number of galaxies besides our own. Cepheids are variable stars, pulsing in a regular pattern, typically over a period of days. The period of the pulsations is a very good indicator of the star's intrinsic brightness — the brighter the star, the longer its period. By comparing the intrinsic brightness and the observed brightness of these stars, astronomers can estimate the distance, making Cepheids an invaluable tool for studying the relationship between distance and redshift. In addition to the Cepheids, supernovae of a type called 1a also began to play a major role in measurements of the Hubble constant. Type 1a supernovae explode once and then fade from view, unlike the periodic cycles of Cepheid stars. Nonetheless, the so-called "light-curves" from these supernovae — the way their brightness rises sharply to a peak and then falls over characteristic time-scales — can likewise be related quantitatively to their intrinsic brightness. Fig. 2.5 shows a more modern Hubble diagram, displaying measurements of Type 1a supernovae, all measured before 2002.

In 2001 the Hubble Key Project Team announced its final result,* $H_0 = 72 \pm 8$ km-s^{-1}-Mpc^{-1}, a considerable improvement over the large uncertainty expressed in Eq. (2.2).

---

* W. L. Freedman et al., "Final results from the Hubble Space Telescope Key Project to measure the Hubble Constant," *Astrophysical Journal,* vol. 553, pp. 47–72 (2001), http://arXiv.org/abs/astro-ph/0012376.

The Tammann and Sandage group* still advocated a slightly lower value, $H_0 = 60$ km-s^{-1}-Mpc^{-1}, "with a systematic error of probably less than 10%," but the difference between this number and the Hubble Key Project number is rather small.

Soon after that, astronomers reported new measurements of H_0 based on a complementary method. In February 2003 astronomers using the Wilkinson Microwave Anisotropy Probe (WMAP), a satellite dedicated to measuring the faint anisotropies in the cosmic background radiation, released an analysis of their first year of data.† By combining their data with several other experiments, they found the most precise value of H_0 that had yet been announced: 71 ± 4 km-s^{-1}-Mpc^{-1}. Since 2003 a number of new measurements have been announced, including WMAP measurements with 5 years,‡ 7 years,¶ and then 9 years§ of data, as well an estimate based on the higher resolution data from the Planck satellite, with data releases in 2013,♣ 2015,♢ and 2018.♡

Estimates based on CMB measurements, especially the most recent Planck results, have found values for H_0 a little lower than estimates based on more astronomical methods, such as the 2018 measurement by Riess et al.♠, who used Cepheid variables and supernovae of type Ia to recalibrate the cosmic distance scale, finding a value $H_0 = 73.52 \pm 1.62$ km-s^{-1}-Mpc^{-1}. The discrepancy between this value and the Planck

---

* G. A. Tammann, B. Reindl, F. Thim, A. Saha, and A. Sandage, in *A New Era in Cosmology* (Astronomical Society of the Pacific Conference Proceedings, Vol. 283), eds. T. Shanks and N. Metcalfe, http://arXiv.org/abs/astro-ph/0112489.

† D. N. Spergel et al., "First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters," *Astrophysical Journal Supplement,* vol. 148, pp. 175–194 (2003), http://arXiv.org/abs/astro-ph/0302209.

‡ E. Komatsu et al., "Five-year Wilkinson Microwave Anisotropy Probe observations: cosmological interpretation," *Astrophysical Journal Supplement*, vol. 180, pp. 330-376 (2009), Table 1, Column 6, http://arXiv.org/abs/arXiv:0803.0547.

¶ E. Komatsu et al., "Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Cosmological interpretation," *Astrophysical Journal Supplement,* vol. 192, article 18 (2011), Table 1, Column 6, http://arXiv.org/abs/1001.4538.

§ G. Hinshaw et al., "Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological parameter results," http://arXiv.org/abs/1212.5226, Table 3, Column 5.

♣ Planck Collaboration: P. A. R. Ade et al., "Planck 2013 results. XVI. Cosmological parameters," Table 2, Column 7, http://arXiv.org/abs/1303.5076.

♢ Planck 2015 results, XIII, op. cit.

♡ Planck 2018 results, VI, op. cit.

♠ A. G. Riess *et al.* (SH0ES Collaboration), "Milky Way Cepheid Standards for Measuring Cosmic Distances and Application to Gaia DR2: Implications for the Hubble Constant," Astrophys. J. **861**, 126 (2018), arXiv:1804.10655 [astro-ph.CO].

value of Eq. (2.4) is at the level of 3.5 σ, which means that if there are no systematic errors that are being overlooked, the probability that the two results should differ by this much is only about 1 in 2000. The discrepancy might nonetheless be a statistical fluke, or it could be due to some unknown systematic error. If neither of these is the case, it would seem to indicate that the contents of the universe include some new ingredient that is currently unknown.

These and a number of other measurements of the Hubble constant are listed in Table 2.1.¶

THE HOMOGENEOUSLY EXPANDING UNIVERSE:

Given the statements about isotropy, homogeneity, and Hubble's law described above, our task now is to build a mathematical model that incorporates these ideas.

In the real universe, of course, the properties of isotropy, homogeneity, and Hubble's law hold only approximately, and only if the complicated structure that exists on length scales less than a few hundred million light-years is ignored. For a first approximation, however, it is useful to construct a mathematical model describing an idealized universe in which these properties hold exactly.

¶ References that have not already been given are Georges Lemaître, "Un Univers homogène de masse constante et de rayon croissant, rendant compte de la vitesse radiale des nébuleuses extra-galactiques," *Annales de la Société Scientifique de Bruxelles,* vol. A47, pp. 49-59 (1927) [Translated into English as "A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae," *Monthly Notices of the Royal Astronomical Society,* vol. 91, pp. 483-490 (1931)]; W. Baade, *I.A.U. Trans. VIII* (Cambridge Univ. Press), p. 397 (quoted by Tammann and Reindl (2002), op. cit.); A. Sandage, "Current problems in the extragalactic distance scale," *Astrophysical Journal,* vol. 127, pp. 513–526 (1958), http://adsabs.harvard.edu/abs/1958ApJ...127..513S; G. de Vaucouleurs and G. Bollinger, "The extragalactic distance scale. VII - The velocity-distance relations in different directions and the Hubble ratio within and without the local supercluster," *Astrophysical Journal,* Part 1, vol. 233, pp. 433-452, http://adsabs.harvard.edu/abs/1979ApJ...233..433D; A. G. Riess, W. H. Press, and R. P. Kirshner, "A precise distance indicator: Type 1a supernova multicolor light-curve shapes," *Astrophysical Journal*, vol. 473, pp. 88-109 (1996), http://arxiv.org/abs/astro-ph/9604143; A. G Riess et al., "A 3% solution: Determination of the Hubble constant with the Hubble Space Telescope and Wide Field Camera 3," *Astrophysical Journal*, vol. 730, 119 (2011), http://arXiv.org/abs/1103.2976.; A. G. Riess *et al.* (SH0ES Collaboration), "A 2.4% Determination of the Local Value of the Hubble Constant," http://arxiv.org/abs/1604.01424 [astro-ph.CO]; J. N. Grieb *et al.* (BOSS Collaboration), "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: Cosmological implications of the Fourier space wedges of the final sample," Mon. Not. Roy. Astron. Soc. **467**, 2085-2112 (2017); S. Birrer *et al.* (H0LiCOW collaboration), "H0LiCOW-IX: Cosmographic analysis of the doubly imaged quasar SDSS 1206+4332 and a new measurement of the Hubble constant," arXiv:1809.01274 [astro-ph.CO].

Measurements of the Hubble Constant H_0		
Author	Date	Value (km-s^{-1}-Mpc^{-1})
Lemaître	1927	575 – 625
Hubble	1929	500
Hubble & Humason	1931	560
Baade	1952	250
Sandage	1958	75, with a possible uncertainty of a factor of 2
de Vaucouleurs & Bollinger	1979	100 ± 10
Riess et al. (SN 1a & cepheids)	1996	65 ± 6
Hubble Key Project	2001	72 ± 8
Tammann, Sandage, et al.	2001	$60\pm$ probably less than 10%
WMAP 1-year (with other data)	2003	71 ± 4
WMAP 5-year (with other data)	2008	70.5 ± 1.3
WMAP 7-year (with other data)	2011	70.2 ± 1.4
Riess et al. (SN 1a & cepheids)	2011	73.8 ± 2.4
WMAP 9-year (with other data)	2012	69.3 ± 0.8
Planck 2013 (with other data)	2013	67.3 ± 1.2
Planck 2015 (with other data)	2015	67.7 ± 0.5
Riess et al. (SH0ES collaboration, SN Ia & cepheids)	2016	73.2 ± 1.7
Grieb et al. (BOSS collaboration)	2016	67.6 ± 0.7
Riess et al. (SH0ES collaboration, SN Ia & cepheids)	2018	73.5 ± 1.6
Planck 2018 (with other data)	2018	67.7 ± 0.4
Birrer et al. (H0LiCOW collaboration, gravitationally lensed quasars)	2018	72.5 ± 2.2

Table 2.1

At first thought, one might think that the concept of homogeneity is inconsistent with Hubble's law — if the universe is expanding, there must be a unique point which is at rest. This argument would be valid ***if*** there were some physical way of telling if an object is at rest. However, the basic principle of the theory of relativity asserts that all inertial reference frames are equivalent, and that any reference frame traveling at a uniform velocity with respect to an inertial reference frame is also an inertial reference frame. For example, if a train moves at a constant speed in a fixed direction, then observers on the train would observe exactly the same laws of physics as observers on the ground. The viewpoint of observers on the train, for whom the ground is moving and the table in the dining car is at rest, is just as "real" as the viewpoint of observers on the ground. Thus, there is no meaning to being absolutely at rest. While special relativity dates from 1905, the basic principle that all inertial frames are equivalent was emphasized by Galileo as early as 1632 in his *Dialogue Concerning the Two Chief World Systems.* The concept was crucial to Galileo's view of the solar system, because it explained why we do not feel the huge velocities ($\sim$30,000 m/s $\approx$ 65,000 mph) associated with the rotation of the Earth and its motion around the Sun. (The principle that all inertial frames are equivalent was temporarily abandoned, however, in the 19th century, when the ether was introduced in the description of electromagnetism.)

To see how Hubble's law is consistent with homogeneity, it is easiest to begin with a one dimensional example. To this end, we will borrow a diagram from Steven Weinberg's book, **The First Three Minutes**, shown in Fig. 2.6

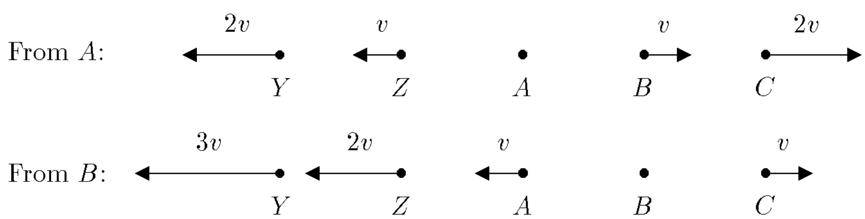


Figure 2.6: Hubble's Law is compatible with homogeneity in space. Each observer can consider herself at rest, and will observe other points moving away from her at speeds proportional to their distance from her.

This diagram shows a row of evenly spaced points. In the top part, the point A is shown in the center, with points B and C to the right, and Z and Y to the left. The picture is drawn from the point of view of an observer at A, so A is at rest in this reference frame. The observer at A sees a pattern of motion dictated by Hubble's law, which means that B and Z are each receding at some speed v, and C and Y are each receding at $2v$. (For now let us assume that $v \ll c$, so we need not worry yet about any of the peculiar effects associated with special relativity.) In this picture it looks as if A is special because it alone is at rest, and the picture is therefore not homogeneous. However, the lower portion the picture is shown from the point of view of an observer at B. The picture is shown in the rest frame of B, and so of course B is at rest. Each velocity in this picture is obtained from the velocity in the picture above by adding a velocity v to the left. One can see that an observer at B can also regard himself as the center of the motion, and he also sees a pattern of motion consistent with Hubble's law.

It is significantly harder to visualize this picture in three dimensions, so it is useful to introduce some mathematical machinery. The concept of a homogeneously expanding universe can be described most simply by using the analogy of a roadmap. A roadmap is of course much smaller than the area that it describes, but the distances are related by the scaling that is usually indicated in one of the corners of the map. It might read, for example, "1 inch = 7 miles." If some sorcerer somehow caused the entire region to uniformly double in size, we would be shocked, but we would not have to throw away the map. Instead we could just cross out the statement "1 inch = 7 miles" and replace it with "1 inch = 14 miles."

While it is not likely that we will meet such a sorcerer, the universe is to a good approximation expanding uniformly, and we can use the same map trick to describe it. Even though the universe is expanding, we can represent it by a map that does not change with time. The universe is three-dimensional, so the map takes the form of a three-dimensional coordinate system, with coordinates x, y, and z. The coordinate axes can be marked off in arbitrary units, which I will call "notches." We could measure the map in ordinary distance units, like centimeters, and in fact most cosmology textbooks do that. But by inventing a new unit, we can emphasize that distances on the map have no fixed relation to the physical distances between the actual objects that are pictured on the map. By using notches, we give ourselves an extra dimensional check on our calculations. If we keep track of our units and the answer is given in notches, then we will know that we calculated a map distance, and not the physical distance between real objects.

As time progresses, the expansion of the universe can be described by changing the relation between physical distances and the notch. At one time a notch might correspond to a million light-years, and at a later time it might correspond to one and a half million light-years. A coordinate system that expands with the universe in this way is called a *comoving coordinate system.* The expansion of a part of the universe, with the comoving coordinate system shown, could be depicted as in Fig. 2.7:

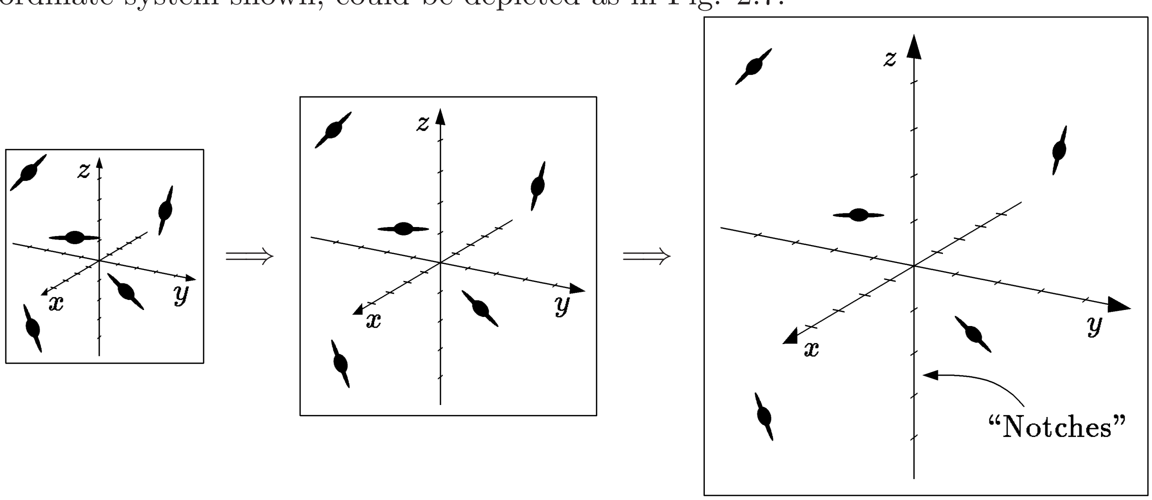


Figure 2.7: By employing "comoving coordinates," a single map can represent the locations of objects in an expanding universe. Distances between objects on the map are measured in "notches," while the relation between notches and physical units (such as centimeters or light-years) changes over time.

Objects that are moving with the Hubble expansion are at rest in these coordinates, and the motion is described entirely by the scale factor $a(t)$, which gives the physical distance that corresponds to one notch at any time t. The scale factor $a(t)$ might be measured, for example, in units such as m/notch. The physical distance between any two points at any given time is then given by

$$\boxed{\ell_p(t) = a(t)\ell_c \ .} \tag{2.5}$$

Here ℓ_c denotes the coordinate distance between the two objects (such as the galaxies depicted in Fig. 2.7). It is measured in notches and is independent of time. ℓ_p denotes the physical distance, which is measured in meters and increases with time as the universe expands.

(Note that the diagrams in Fig. 2.7 show that the distances between galaxies are growing uniformly, while the galaxies themselves are not expanding. Inside each galaxy the gravitational pull of the mass concentration has caused the expansion to halt. For now, however, we are interested only the properties of the universe that are seen when averaging over large regions with many galaxies, so the details of what happens inside these galaxies are not important.)

Since special relativity tells us that moving rulers contract in the direction of motion, the concept of "physical distance" needs to be carefully defined. Should the distance between us and a distant galaxy be measured with rulers at rest relative to us, or with rulers at rest relative to the distant galaxy? Neither of these choices is good, since either choice would require rulers on one end or the other that are moving at high speed relative to the matter around them. The relativistic contraction would distort the distances, so that the average separation between galaxies would appear to vary with the distance from the observer.

To avoid this problem, cosmologists use the concept of "comoving" rulers — rulers which move with the nearby matter. To define the physical distance between us and a far-away galaxy, one imagines marking off a line between us and the galaxy with closely spaced grid marks. The distance between each two grid marks is then measured with a ruler that is at rest with respect to the matter in the region between the two grid marks, and the distance between us and the galaxy is defined by adding the distances so measured. This is how the quantity $\ell_p(t)$ in Eq. (2.5) is defined. Distance defined in this way is called the *proper distance.* We will also refer to $\ell_p(t)$ as the *physical distance*, in contrast with the (comoving) coordinate distance ℓ_c.

We are now in a position to see how the homogeneous expansion implied by Eq. (2.5) leads directly to Hubble's law. To see this, one simply differentiates Eq. (2.5) in order to find the velocity. If ℓ_p denotes the distance between a particular distant galaxy and us, then the recession velocity of that galaxy is given by

$$v = \frac{\mathrm{d}\ell_p}{\mathrm{d}t} = \frac{\mathrm{d}a}{\mathrm{d}t}\ell_c = \left[\frac{1}{a(t)}\,\frac{\mathrm{d}a(t)}{\mathrm{d}t}\right] a(t)\ell_c \ . \tag{2.6}$$

Note that this can be rewritten as

$$v = \frac{\mathrm{d}\ell_p}{\mathrm{d}t} = H\ell_p \ , \tag{2.7}$$

where $H(t)$ is given by

$$\boxed{H(t) = \frac{1}{a(t)}\,\frac{\mathrm{d}a(t)}{\mathrm{d}t} \ .} \tag{2.8}$$

By comparing Eqs. (2.7) and (2.1), we see that the assumption of uniform expansion has led immediately to Hubble's law. Even better, in Eq. (2.8) we have derived an expression for the Hubble expansion rate, $H(t)$.

MOTION OF LIGHT RAYS:

To understand observations in a universe described by a comoving coordinate system, we will need to be able to trace the path of light rays through it. The rule is very simple: light travels in a straight line, with a speed that would be measured by each local observer, as the light ray passes, at the standard value $c = 299,792,458$ m/s. The key point is that the speed is fixed in the physical units, such as m/s, while the coordinate system is marked off in notches. Thus, at any given time one must use the conversion factor $a(t)$ to convert from meters to notches, in order to find the speed of a light pulse in comoving coordinates.

Consider, for simplicity, a light pulse moving along the x-axis. If the speed of light in m/s is c, and the number of meters per notch is given by $a(t)$, then the speed in notches per second is given by $c/a(t)$:

$$\boxed{\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{c}{a(t)} \ .} \tag{2.9}$$

To check our units, we can use square brackets $[A]$ to denote the units of some quantity A. Then

$$\left[\frac{c}{a(t)}\right] = \frac{\mathrm{m/s}}{\mathrm{m/notch}} = \frac{\mathrm{notch}}{\mathrm{s}} \ , \tag{2.10}$$

which gives the right units for $\mathrm{d}x/\mathrm{d}t$, since x is a coordinate measured in notches.

Since we have not studied general relativity, the reader might well be leery that the subtleties of spacetime might somehow lead to a flaw in this argument. Eq. (2.9), however, is in fact rigorously correct in general relativity. It can be derived in the context of hypothetical point particles that travel at the speed of light, as we argued here, or one can incorporate Maxwell's equations into general relativity, and then calculate the speed of electromagnetic waves.

THE SYNCHRONIZATION OF CLOCKS:

One of the key ideas discussed earlier in the context of special relativity was the notion that simultaneity is a frame-dependent concept — two clocks which appear synchronized to one observer will appear to be unsynchronized to an observer in relative motion. Thus, when we speak of $a(t)$ as a single function which characterizes the entire universe, we should ask ourselves how we will synchronize the clocks on which t will be measured.

The answer turns out to be simple, although a little subtle. Imagine that we are living in this idealized universe, so we can measure the expansion function a as a function of our own clock time, using our own choice of a notch. Similarly, we can imagine another civilization of creatures living in the galaxy M81, who measure a according to their own clocks, with their choice of a notch. We will assume that communication is possible, but time signals alone are not sufficient to synchronize clocks, since the signals travel with at most the speed of light, and the distance from the Earth to M81 is time-dependent and initially unknown. Thus, if we receive a signal from M81 saying that "this signal was sent at $t = 0$," we would have no way of knowing how much time had elapsed since the signal was sent. So, is it possible for the M81 creatures and us to agree on a definition of time and on the scale factor $a(t)$?

Common units for distance and time can in principle be established by using atomic standards, in the same way as we do on Earth — time can be defined in terms of a sharply defined atomic frequency, and distance can be defined in terms of how far light can travel in a unit of time. But one must still ask how the clocks are to be synchronized. One might think that one could synchronize the clocks by fixing the zero of time to be the instant when the scale factor a reaches a certain value, but this plan is complicated by the fact that it requires the creatures on M81 to understand not only what we mean by meters and seconds, but also what we mean by notches. Since the physical distance corresponding to a notch is time-dependent, we cannot communicate its definition until we have found a way to synchronize clocks.

The idea then is to find some physically measurable quantity and use its time dependence to synchronize clocks. One choice is the Hubble expansion rate $H(t)$. In principle, we and the M81 creatures could synchronize our clocks by setting them all to zero when $H(t)$ reaches some prescribed value. Alternatively, the temperature of the cosmic microwave background radiation could be used, resulting in the same synchronization.

(Note that the assumption of homogeneity implies that the relationship between $H(t)$ and the microwave background temperature $T(t)$ must be the same at all points in the universe.) Time defined in this way is called *cosmic time,* and it is this definition of time that will be used for the rest of this course, unless otherwise specified.

Once we agree with the M81 creatures on how to synchronize our clocks, we can also fix a definition of the notch by fixing its value in atomic units at the time of synchronization. They and we can then independently measure the scale factor $a(t)$ for all future times. Will we get the same value? By the assumption of homogeneity, of course we will — otherwise there would have to be some real distinction between the way the universe appears to them and the way it appears to us.

If one is looking for subtle problems, one might ask what would happen in a universe in which $H(t)$ just happens to be a constant (independent of time), and in which there is no microwave background radiation. A spacetime of this type was first studied in 1917 by the Dutch astronomer Willem de Sitter, and is called *de Sitter space.* The definition of cosmic time given above does not make sense in de Sitter space, and it turns out that there is no unique definition. Does this have any relevance to cosmology? Yes, as we will see later when we discuss inflation. Although the de Sitter model is no longer regarded as a viable description of the present universe, the model has become relevant in a different context. The inflationary universe scenario, which we will be discussing later in this course, is characterized by a phase in which the universe is accurately described by a de Sitter space. Furthermore, it is likely that the present acceleration of the cosmic expansion, discovered in 1998*, could indicate the beginning of a de Sitter space era in our future.

By using the time dependence of $H(t)$ or $T(t)$, we can define what it means to say that two events happened at the same time t, even if they occurred billions of light-years apart. In cosmology, in other words, we may single out a special class of observers: those who are moving with the Hubble expansion, and hence are at rest with respect to the matter in their own vicinity. Clocks carried by these special observers define the measurement of cosmic time. The special observers in different regions are moving with respect to each other, and thus the cosmic time system that they measure is not equal to the time that would be measured in any one inertial reference frame.

To summarize: the time variable t that we are using is called cosmic time, and any observer at rest relative to the galaxies in her vicinity can measure it on her own clock. The clocks throughout the universe can be synchronized by using the Hubble expansion rate $H(t)$ or the temperature $T(t)$ of the cosmic microwave background radiation.

---

* A. G. Riess et al., "Observational evidence from supernovae for an accelerating universe and a cosmological constant," *Astronomical Journal*, vol. 116, pp. 1009-10038 (1998), http://arxiv.org/abs/astro-ph/9805201; S. Perlmutter et al., "Measurements of Omega and Lambda from 42 high redshift supernovae," *Astrophysical Journal*, vol. 516, pp. 565-586 (1999), http://arxiv.org/abs/astro-ph/9812133.

THE COSMOLOGICAL REDSHIFT:

Suppose an atom on a distant galaxy is emitting light with wave crests separated by a fixed time interval Δt_S ("S" for "source"). We will receive these wave crests at a Doppler-shifted interval, which we will call Δt_O ("O" for "observer"). Our goal is to relate the Doppler shift to the behavior of the scale factor $a(t)$. We might think that we could just use the special relativity formula for the Doppler shift that we derived in Lecture Notes 1, but that would not properly take into account the motion of light rays in an expanding universe, as described by Eq. (2.9). To take this into account, we start the calculation from scratch.

Let us construct a coordinate system with ourselves at the origin, and let us align the x-axis so that the galaxy in question lies on it, as in Fig. 2.8:

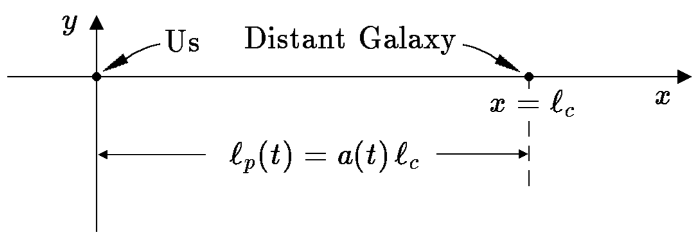


Figure 2.8: Diagram for discussing the transmission of a light signal from a distant galaxy to us. We are at the origin, and the galaxy is along the x-axis, at $x = \ell_c$. The light signal travels to us along the x-axis.

Let t_S be the cosmic time at which the first crest is emitted from the distant galaxy, with the second crest emitted at $t_S + \Delta t_S$. The atom is a kind of clock situated on the distant galaxy, so the time interval measured by the atom agrees with the interval of cosmic time. (Note that this is different from the relativistic Doppler shift calculation in Lecture Notes 1, in which we explicitly took into account the slowing down of a clock on a moving source. Here we are using a different kind of coordinate system, with a different definition of the time coordinate. Each clock is at rest in the non-inertial comoving coordinate system, and the cosmic time of the coordinate system is by definition the time as read on such clocks.)

The next step is to understand the relationship between the time interval of emission Δt_S and the time interval of observation Δt_O. Note that after the first crest is emitted, it travels a physical distance $\lambda_S \equiv c\Delta t_S$ before the second crest is emitted. If Δt_S is the time between the emission of wave crests, then

$$\lambda_S \equiv c\Delta t_S \tag{2.11}$$

is the wavelength of the emitted wave. The two crests are then separated by a coordinate distance

$$\Delta x = \lambda_S / a(t_S)\ . \tag{2.12}$$

We assume that the period of the wave Δt_S is very short compared to the time scale on which $a(t)$ varies, so it does not matter whether the denominator is written as $a(t_S)$ or $a(t_S + \Delta t_S)$. According to Eq. (2.9), the velocity of light in these coordinates depends on t, but is independent of spatial position. Thus, at any given time the two crests will travel at the same coordinate velocity $\mathrm{d}x/\mathrm{d}t$, and thus will stay the same coordinate distance apart. When they arrive at the observer they will still be separated by the same coordinate distance Δx with which they started. The physical separation at the observer will then be given by

$$\lambda_O = a(t_O)\Delta x = \frac{a(t_O)}{a(t_S)}\,\lambda_S\ , \tag{2.13}$$

and thus the wavelength is simply stretched with the expansion of the universe. The time separation between the arrival of the crests will be

$$\Delta t_O = \frac{\lambda_O}{c} = \frac{a(t_O)}{a(t_S)}\Delta t_S\ . \tag{2.14}$$

Finally, one has

$$\boxed{1 + z \equiv \frac{\Delta t_O}{\Delta t_S} = \frac{\lambda_O}{\lambda_S} = \frac{a(t_O)}{a(t_S)}\ .} \tag{2.15}$$

Thus, the Doppler shift factor $1 + z$ is just the ratio of the scale factors at the times of observation and emission. Equivalently, the wavelength of the light is stretched by the expansion of the universe.

It is natural to ask how this calculation is related to the calculation of the relativistic Doppler shift of Lecture Notes 1. Since this calculation did not involve any explicit reference to time dilation, one might think that this calculation is nonrelativistic. If you carefully go back over the calculation, however, you will find that there is no step that depends on these relativistic effects in any way. Eq. (2.15) is a rigorous consequence of Eq. (2.9) and the construction of the comoving coordinate system. In fact, Eq. (2.15) is an exact result of general relativity, which includes the effects of both special relativity and gravity. It is possible to apply Eq. (2.15) to the special case in which gravity is negligible, and the usual result of special relativity can, with some effort, be recovered. (You will be given the opportunity to carry out this exercise, with some hints, on a problem set later in the term.) However, the content of Eq. (2.15) differs from the special relativity result in two ways:

(1) The special relativity result holds exactly only in the absence of gravity, while Eq. (2.15) includes the effects of gravity — provided, of course, that one knows the effects of gravity on the scale factor $a(t)$.

(2) Eq. (2.15) expresses the Doppler shift in terms of the behavior of the scale factor $a(t)$ for objects at rest in a ***comoving*** coordinate system, while the

special relativity result expresses the Doppler shift in terms of the velocity as measured in an ***inertial*** coordinate system. Thus, the two results cannot be compared until one works out the relationship between these two coordinate systems. When Eq. (2.15) is applied to the special case in which gravity is negligible, one finds that the details of special relativity — time dilation, Lorentz-Fitzgerald contraction, etc. — must be used in order to relate these two coordinate systems.

While the cosmological Doppler shift is in general different from the special relativity Doppler shift, since it takes into account the effects of gravity, we will see in the next set of lecture notes that the effects of gravity grow with distance. So, if the source and observer are close, we would expect that the effects of gravity would be negligible and the two answers would agree.

To see this, we use the fact that if the source and observer are close, then the transmission time $\delta t \equiv t_O - t_S$ will be small. Over this small time interval, we can apporximate $a(t)$ by its first order Taylor expansion about t_S:

$$\begin{aligned} a(t) &= a(t_S) + \dot{a}(t_S)(t - t_S) + \ldots \\ &= a(t_S)\left[1 + H(t_S)(t - t_S) + \ldots\right] \ , \end{aligned} \tag{2.16}$$

where an overdot denotes a time derivative, and use was made of Eq. (2.8). Applying this eqation to $t = t_O$,

$$a(t_O) = a(t_S)\left[1 + H(t_S)\,\delta t + \ldots\right] \ . \tag{2.17}$$

The coordinate separation Δx between source and observer can be found by integrating the coordinate velocity given by Eq. (2.9):

$$\begin{aligned} \Delta x &= \int_{t_S}^{t_O} \frac{c\,dt}{a(t)} = \int_{t_S}^{t_S+\delta t} \frac{c\,dt}{a(t_S)\left[1 + H(t_S)(t - t_S) + \ldots\right]} \\ &= \frac{c}{a(t_S)} \int_{t_S}^{t_S+\delta t} dt\left[1 - H(t_S)(t - t_S) + \ldots\right] = \frac{c}{a(t_S)}\left[\delta t - \frac{1}{2}H(t_S)\delta t^2 + \ldots\right] \ . \end{aligned} \tag{2.18}$$

Since we are interested in very small δt, we use the lowest order result that $\Delta x = c\,\delta t/a(t_S)$. If we let δr denote the physical distance between source and observer at time t_S, then to lowest order in δt,

$$\delta r = a(t_S)\,\Delta x = c\,\delta t \ , \tag{2.19}$$

which we might well have foreseen. Eq. (2.19) is a consequence of the fact that if δt is small, then the effect of the expansion of the universe during the time δt is negligible. The cosmological redshift is then given by

$$1 + z = \frac{a(t_O)}{a(t_S)} = 1 + H(t_S)\,\delta t + \ldots \ , \tag{2.20}$$

where we used Eq. (2.16). Then by using Eqs. (2.20) and (2.19) we find

$$z = H(t_S)\,\delta t = \frac{H(t_S)c\,\delta t}{c} = \frac{H(t_S)\delta r}{c} = \frac{v}{c}\ , \tag{2.21}$$

where in the last step we used Hubble's law, Eq. (2.1). To lowest order in $\beta \equiv v/c$, this agrees with the special relativity Doppler formula,

$$z = \sqrt{\frac{1+\beta}{1-\beta}} - 1 \qquad \text{(relativistic)}, \tag{2.22}$$

where $\beta = v/c$.

Although the cosmological redshift is caused by both gravity and by motion, there is no natural way to divide it into these two parts. You might suggest, for example, that we define the part due to gravity by asking how much the Doppler shift would change if gravity were omitted from the calculation. The problem is that the trajectories of the source, the observer, and the light rays would all be different in the absence of gravity. Thus, we cannot ask what the redshift would be in a universe that is like ours, but without gravity. If gravity were not involved, there would not be any universe that is like ours.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe September 22, 2018
Prof. Alan Guth

Lecture Notes 3
THE DYNAMICS OF NEWTONIAN COSMOLOGY

INTRODUCTION:

The dynamics of the universe on the large scale seems to be controlled by gravity, and so any theoretical work in cosmology rests heavily on the foundations of a theory of gravity. Among professionals, the dynamics of cosmology is always treated in the context of the relativistic theory of gravity developed by Einstein in 1915 — the theory which is known by the misleading name of "General Relativity". We believe that, at the classical level, general relativity is almost certainly the correct theory of gravity. (At extraordinarily high energies, like those encountered during the first 10^{-45} second after the big bang, a quantum theory of gravity would be required. At present, however, gravity at the quantum level is not well-understood. Many physicists believe that string theory is likely to be the correct quantum theory of gravity, but there are many questions about string theory that have not yet been answered.) In cosmology, general relativity is necessary to make sure that the possibly non-Euclidean geometry of the universe is being treated correctly. General relativity is also required to give an accurate treatment of the gravitational effect of electromagnetic radiation (e.g., light), which is significant in the early universe and which is certainly a relativistic phenomenon. However, a good deal of cosmology can be understood strictly in terms of Newtonian gravity, and in these notes we will explore cosmology in that context. Even the gravitational effects of electromagnetic radiation can be inferred correctly by using Newtonian physics combined with some well-motivated guesses.

The universe is believed to be homogeneous, so the key problem is to understand the gravitational dynamics of a homogeneous distribution of mass. We will consider a distribution of mass with infinite extent, with a uniform mass density ρ.

This is a subtle problem, and in fact Isaac Newton himself got it wrong. Newton assumed that since the mass distribution is symmetric about any point, the gravitational field at any point must vanish, since there is no preferred direction in which it could point. He therefore believed that a static configuration of "fixed stars" could exist in equilibrium. Newton discussed this issue in a series of letters he wrote to the young theologian, Richard Bentley, during 1692-93*

---

* The original letters are still kept at Trinity College, Cambridge, and are published in H. W. Turnbull, ed., *The Correspondence of Isaac Newton, Volume III, 1688-1694* (Cambridge University Press, Cambridge, England, 1961, p. 233). They are also reprinted

> *"As to your first query, it seems to me that if the matter of our sun and planets and all the matter of the universe were evenly scattered throughout all the heavens, and every particle had an innate gravity toward all the rest, and the whole space throughout which this matter was scattered was but finite, the matter on the outside of this space would, by its gravity, tend toward all the matter on the inside and, by consequence, fall down into the middle of the whole space and there compose one great spherical mass. But if the matter was evenly disposed throughout an infinite space, it could never convene into one mass; but some of it would convene into one mass and some into another, so as to make an infinite number of great masses, scattered at great distances from one to another throughout all that infinite space. And thus might the sun and fixed stars be formed, supposing the matter were of a lucid nature.*[†]*" (December 10, 1692)*

The point of view that Newton described in his response to Bentley apparently represents a departure from his earlier reasoning. Previously Newton had believed that the fixed stars occupied a finite region in an infinite void, but now he realized that such a configuration would be driven by gravity to collapse. If the stars were distributed uniformly over the infinity of space, however, Newton concluded that static equilibrium could be maintained.*

---

in Milton K. Munitz, ed., *Theories of the Universe: From Babylonian Myth to Modern Science* (The Free Press, New York, 1957, p. 211). Best of all, thanks to Google Books and the Newton Project, the complete letters from Newton to Bentley are now available online: http://books.google.com/books?id=8DkCAAAAQAAJ&pg=PA201 and http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00254, http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00255, http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00256, and http://www.newtonproject.sussex.ac.uk/view/texts/normalized/THEM00258.

† By "lucid nature," Newton was apparently referring to the distinction that he supposed exists between the "lucid matter" of the sun and stars, and the "opaque" matter of the earth and other planets. The continuation of the text shows Newton's thoughts on this issue, and also on the role of divine intervention in the creation of the solar system: *"But how the matter should divide itself into two sorts, and that part of it which is to compose a shining body should fall down into one mass and make a sun and the rest which is fit to compose an opaque body should coalesce, not into one great body, like the shining matter, but into many little ones; or if the sun at first were an opaque body like the planets or the planets lucid bodies like the sun, how he alone should be changed into a shining body whilst all they continue opaque, or all they be changed into opaque ones whilst he remains unchanged, I do not think explicable by mere natural causes, but am forced to ascribe it to the counsel and contrivance of a voluntary Agent."*

* Newton's involvement in this problem was discussed in a fascinating article by Edward Harrison, "Newton and the Infinite Universe," *Physics Today*, February 1986, p. 24, which is available online with an MIT certificate at http://scitation.aip.org.

The fallacy of Newton's argument was not really understood until the beginning of the 20th century. When Einstein first developed his theory of general relativity, he very quickly tried to apply it to the universe as a whole, and at first he was rather shocked to learn that the theory did not allow a static solution. According to the mathematics of the theory, an initially static configuration would lead to a universal collapse, as each particle of matter in the universe attracted all of the others. Einstein chose to modify general relativity by adding a "cosmological term" — a kind of universal repulsion — so that a static solution would be possible. In hindsight, one can see that the same reasons which preclude a static solution in the theory of general relativity (without a cosmological constant) apply also to the Newtonian case.

The nonexistence of a static equilibrium for an infinite homogeneous distribution of mass can be seen very easily by using some mathematics that was unavailable to Newton. Newton formulated his law of universal gravitation in the language of an inverse square force law, but we now know how to reformulate such a law in terms of flux integrals. Just as Coulomb's law implies Gauss's law, Newton's inverse square law of gravity gives rise to a Gauss's law of gravity:

$$\vec{E} = \frac{q}{r^2}\hat{r} \quad \text{implies} \quad \oint \vec{E} \cdot d\vec{a} = 4\pi q_{\text{enclosed}} \tag{3.1}$$

$$\vec{g} = -\frac{GM}{r^2}\hat{r} \quad \text{implies} \quad \oint \vec{g} \cdot d\vec{a} = -4\pi G M_{\text{enclosed}} \ , \tag{3.2}$$

where $\vec{g}$ is the gravitational acceleration vector, and the integrals are over an arbitrary "Gaussian" surface. If Eq. (3.2) is applied to a uniform distribution of mass, then clearly $M_{\text{enclosed}} > 0$ for any Gaussian surface that encloses a nonzero volume. Thus the left hand side must also be nonzero, and so one cannot have $\vec{g} = 0$, as a static universe would demand.

Another formulation of Newtonian gravity takes the form of a gravitational Poisson's equation:

$$\nabla^2 \phi = 4\pi G \rho \ , \ \text{where } \vec{g} = -\vec{\nabla}\phi \ , \tag{3.3}$$

and ρ is the mass density. Here ∇^2 is the Laplacian,

$$\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \ ,$$

libproxy.mit.edu/content/aip/magazine/physicstoday/article/39/2/10.1063/1.881049 or for purchase at http://scitation.aip.org/content/aip/magazine/physicstoday/article/39/2/10.1063/1.881049.

and $\vec{\nabla}$ is the gradient,

$$\vec{\nabla} \equiv \hat{\imath}\,\frac{\partial}{\partial x} + \hat{\jmath}\frac{\partial}{\partial y} + \hat{k}\frac{\partial}{\partial z}\ .$$

In this formalism one can also see that $\vec{g} = 0$ implies $\phi = constant$, which in turn implies $\rho = 0$, so a static universe is possible only if it is empty.

Historically, I believe that the inconsistency of the static universe was overlooked in the context of Newtonian mechanics because Newtonian gravity is usually described in terms of an action at a distance. In this formulation, the relevant issues are subtle. General relativity, on the other hand, is always formulated in terms of local differential equations analogous to Eq. (3.3), and in this formulation the result is unmistakable.

I have now discussed the reasons why a homogeneous mass distribution must produce a gravitational field, but I have not yet discussed what goes wrong if one tries to calculate the force on a given particle by summing the Newtonian gravitational forces caused by all the other particles. Since these other particles extend with uniform density to infinity in all directions, it seems obvious that the integration over the mass distribution cannot pick out any preferred direction, and therefore must give no gravitational force. The problem with using this method, however, is that the integration is ambiguous. We will show that, due to the poor convergence properties of the integral, the integration has no unique answer, but instead can give any answer that one wants, depending on the order with which the different regions of the integration volume are included.

To see how this can happen, let us first consider some general properties of integrals. Suppose that $f(x)$ is a function such that

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x \tag{3.4}$$

converges, in the sense that

$$\lim_{L\to\infty} \int_{-L}^{L} f(x)\,\mathrm{d}x \tag{3.5}$$

exists. Suppose, however, that

$$\int_{-\infty}^{\infty} |f(x)|\;\mathrm{d}x \tag{3.6}$$

diverges (i.e., is infinite). Such integrals are called *conditionally convergent*, and in general their value is ambiguous. The answer depends on the order in which the different regions of the x-integration are added up. Conversely, if the integral (3.6) converges, then the integral (3.4) is called absolutely convergent, and its value is independent of the order in which the different regions of integration are added.

As a simple example, consider the function

$$f(x) = \begin{cases} +1 & \text{if } x > 0 \\ \\ -1 & \text{if } x < 0 \ , \end{cases} \tag{3.7}$$

the integral of which clearly satisfies the properties of conditional convergence as described above. To illustrate the ambiguity of the integral, note first that

$$\lim_{L\to\infty} \int_{-L}^{L} f(x)dx = 0. \tag{3.8}$$

This limit is trivial, since the integral is zero for any value of L. Now let's add the contributions in a different order, starting at some arbitrary point $x = a$. We take the region of integration larger and larger, but always centered on $x = a$. That is, we can define the integral

$$\lim_{L\to\infty} \int_{a-L}^{a+L} f(x)dx \ . \tag{3.9}$$

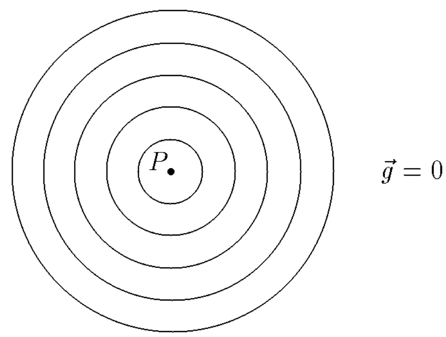
You should be able to convince yourself that the integral is equal to $2a$ for any $L \geq a$, and therefore the limit is $2a$. Since we can choose a to be anything we like, we can get any answer that we like. Note that the integrals shown as (3.5) and (3.9) are both ways of giving precise meaning to the integral (3.4), so one concludes that the integral (3.4) is ambiguous. Mathematically one can (and usually does) ***define*** the integral (3.4) to be the expression (3.5), but one must keep in mind that this is an arbitrary choice that is unlikely to have physical meaning. When x represents a spatial coordinate, as it does here, then the expressions (3.5) and (3.9) differ only by the choice of where the origin of the coordinate system is placed, while physically this choice is completely arbitrary.

For an infinite distribution of mass with uniform density ρ, the gravitational acceleration at a point P is given formally by the integral

$$\vec{g}(P) = \int G\rho \, d^3\vec{r}\,' \, \frac{\vec{r}\,' - \vec{r}_P}{|\vec{r}\,' - \vec{r}_P|^3} \ , \tag{3.10}$$

where $\vec{r}_P$ is a vector from the origin to the point P. We will see that this integral is conditionally convergent, and therefore $\vec{g}(P)$ can have any value, depending on the order in which the contributions from different values of $\vec{r}\,'$ are added. Newton's law of gravity says nothing about the order in which the contributions should be added, since in normal situations vector addition is commutative.

To see how this integral behaves, suppose we first determine the value of $\vec{g}$ at an arbitrary point P by summing the contributions from spherical shells that are centered at P:



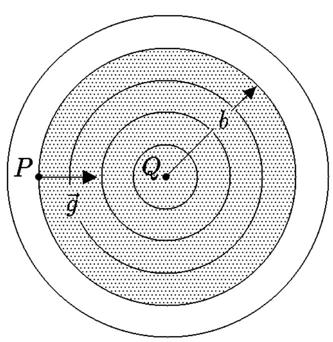
In this case one can argue by symmetry that each shell contributes exactly zero to $\vec{g}$, and hence the sum must be zero.

The integral clearly converged (in fact it vanished!), but it is not absolutely convergent. If we inserted absolute value signs in the integrand, we can evaluate the integral by transforming to polar coordinates with P at the origin. We find a linear divergence:

$$\int G\rho\, d^3\vec{r}\,' \frac{1}{\left|\vec{r}\,' - \vec{r}_P\right|^2} = 4\pi G\rho \int r'^2\, dr' \frac{1}{r'^2} = \infty \ .$$

Thus the integral is convergent but not absolutely convergent, so it is conditionally convergent.

To see the ambiguity that we expect due to the conditional convergence, we need to carry out the integration of Eq. (3.10) with a different ordering. Spherical shells are still very convenient, but suppose we choose spherical shells centered around a different origin. To see what we find, let us calculate $\vec{g}$ at P by summing the contributions from spherical shells which are centered at some other point Q, located a distance b away:



The gravitational field due to a thin spherical shell of mass is well known — inside the shell the field vanishes identically, and outside the shell the field is the same as it would be if the same mass were concentrated at the point in the center of the sphere. For the shells centered at Q, note that the point P will lie inside the shell for all shells with radius $r > b$. These shells will therefore give no contribution to the gravitational field at P. The shells with $r < b$, on the other hand, which are shown with shading in the diagram above, will produce a gravitational field at P. Specifically, they will produce a field at P which is the same as the field that would be produced if the entire mass (for $r < b$) were concentrated at Q. Thus, by this method of summation we find

$$\vec{g} = \frac{GM}{b^2}\hat{e}_{QP} \ , \tag{3.11}$$

where $M = \frac{4\pi}{3}b^3\rho$ is the combined mass of all the shells with $r < b$, and $\hat{e}_{QP}$ is a unit vector pointing from P to Q. So the answer we get depends on the order of summation. Since we could have chosen the point Q to be any distance and in any direction, we could have gotten any answer we wanted.

Thus we can conclude that the integral which determines $\vec{g}$ is ill-defined. By summing the gravitational force using concentric spherical shells centered at different points Q, we can get any answer we want. But what about the simple symmetry argument, which says that the gravitational force must be zero because there is no preferred direction for it to point? It was this argument that Newton found persuasive in the letter to Bentley cited earlier. Newton might phrase the reasoning in the following way: If there is to be a force on the mass located at P, then the force would have to point in some direction. But since all directions are identical in this problem, the force must vanish. To convince Newton that he was wrong, we would have to persuade him that this problem is very special, because there is no way to define an inertial reference frame. Ordinarily one can define an inertial frame by imagining test particles at infinite distances from all others — the inertial frames are those in which these test particles have constant velocities. In the problem of an infinite uniform mass distribution, however, there is no place to put these test particles. Thus, one cannot measure the absolute acceleration of any particle, but instead one must settle for measuring the **relative** acceleration of one particle with respect to another. One can decide, for example, to measure the accelerations of all particles relative to P. One then finds, as we will see later, that all the accelerations point toward P, and that the acceleration of any given particle has a magnitude which is proportional to the distance from P. If one had chosen to measure all accelerations relative to Q, one would have found a similar pattern centered on Q.

THE MATHEMATICAL MODEL:

The approach that I will follow here is a bit more involved than that used in most textbooks, but it also leads to a stronger result. Most textbooks simply assume that a uniform distribution of mass will remain uniform, but here we will show that the inverse square law of gravity leads to this result. Most other force laws would not.

In order to make the problem of an infinite uniform mass distribution well defined, it is necessary to treat the concept of infinity carefully. Specifically, the safest way to think about infinity is to think of it as a **limit** of finite quantities. The easiest approach, which we will use, is to treat the mass distribution as a uniform sphere of radius $R_{\rm max}$. Only at the end of the calculation will we take the limit $R_{\rm max} \to \infty$.

When we choose to use a sphere of mass to define our problem, it is important to ask if the answer would have turned out differently if we had chosen some other shape. I will not try to demonstrate the answer to this question, but I will tell you what it is. Many shapes, such as any of the regular polyhedra (tetrahedron, cube, octahedron, dodecahedron, or icosahedron), would give the same answer in the limit in which the size approaches infinity. If we had used a rectangular solid, however, the answer would have been different. [For those students who have learned about multipole expansions, I mention that it is only the quadrupole moment of the shape that matters in the limit of infinite size.] The solution obtained from the rectangular solid would correspond to an anisotropic model of the universe, in which the gravitational field would be different in different directions. General relativity also allows for the possibility of anisotropic homogeneous solutions, but I have never explored how closely the properties agree. Since our universe is highly isotropic, we are justified in using the sphere to formulate our problem.

We will treat the matter as a nonrelativistic dust of particles which can move freely, with gravity supplying the only significant force. The assumption that the universe is dominated by nonrelativistic matter and that gravity is the only significant force appear to be valid assumptions for our own universe for most of its history, but not for all of it. Recall that in the context of relativity, energy and mass are really the same thing, related by the celebrated formula $E = mc^2$, where c is the speed of light. In the early universe there was a high density of energy in electromagnetic radiation, and this energy density can be expressed as a mass density by dividing it by c^2. For the first approximately 50,000 years of cosmic history, the mass density of the universe was dominated by the electromagnetic radiation and highly relativistic particles, both of which lead to significant pressure forces. These pressure forces, in turn, lead to a contribution to the gravitational force, since general relativity implies that pressures as well as mass or energy densities can serve as the source of a gravitational field. Cosmologists call this early period “radiation-dominated”, and the period in which the universe is dominated by nonrelativistic dust is called “matter-dominated”. Starting in about 1998, astronomers have been gathering evidence that for the past 5 billion years or so the expansion of the universe has not been slowing as it would in a matter-dominated universe, but instead it has been accelerating. These observations were a big surprise to most of us, and they suggest that the universe today is dominated by a nonzero energy density in the vacuum — which is equivalent to what Einstein called the cosmological constant — or some form of peculiar matter that behaves very similarly. The term “dark energy” has been coined to describe this form of energy, which remains rather mysterious as the name suggests.

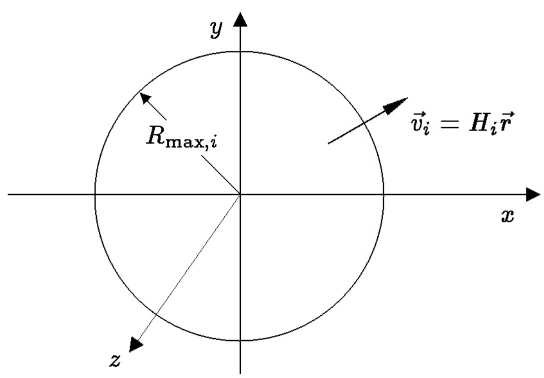
So, it now appears that the universe was radiation-dominated for the first 50,000 years, then became matter-dominated for about 9 billion years, and then "recently" became dark-energy-dominated, for about the last 5 billion years. Later on we will see how the pressure forces of the radiation-dominated period can be incorporated into the model, and we will also learn how to calculate the effects of vacuum energy. For now, however, we will confine our attention to the matter-dominated era.

We will begin the mathematical model of our idealized universe at some arbitrary initial time t_i. At that time we will assume that the universe consists of a sphere of matter with radius $R_{\max,i}$, with uniform mass density ρ_i. For convenience we will introduce an x-y-z coordinate system, with the origin located at the center of the sphere. We will treat the matter as a nonrelativistic dust of particles which can move freely, with gravity supplying the only significant force.

The next step is to specify the initial velocity of each of the particles. In order to agree with the observed properties of the universe, we choose this initial velocity distribution according to Hubble's law: the particles at position $\vec{r}$ are given an initial velocity of the form

$$\vec{v}_i = H_i \vec{r} \ , \tag{3.12}$$

where H_i denotes the initial value of the Hubble "constant". Thus, the initial state of the model is described by the parameters ρ_i, H_i, and $R_{\max,i}$.



The problem now is to calculate the evolution of this model, using Newton's law of gravity. Since each particle is started along a radial trajectory, and since the only forces will be radial, it follows that each particle continues to move along a radial trajectory. Thus we need only keep track of the radius of each particle as a function of time. We will follow an arbitrary particle with initial radius r_i, and we will denote its trajectory by $r(r_i, t)$. To compute the force on this particle due to all the other particles in the model universe, we can divide the mass distribution into thin spherical shells — with each shell centered on the origin and extending from some radius r to $r + dr$. We then use the result quoted earlier for the gravitational field of a thin spherical shell. One concludes that all shells with $r < r_i$ will produce a gravitational field at r_i equivalent to that of a point mass at the origin, while all shells with $r > r_i$ will contribute nothing at all to the gravitational field at r_i. The mass of all the shells with $r < r_i$ is given by

$$M(r_i) = \frac{4\pi}{3} r_i^3 \rho_i \ . \tag{3.13}$$

It is conceivable that at some point in the evolution of the system there could be a crossing of shells — that is, two trajectories $r(r_i, t)$ corresponding to two different values of r_i could cross. However, since initially the Hubble expansion is carrying each shell away from its neighbors, it is clear that a shell crossing will not happen until some nonzero time interval has elapsed. (We will in fact find that shell crossings never occur, but we have no way of knowing this before we start.) As long as no shell crossings have occurred, the mass interior to the shell which began at radius r_i is always equal to the expression for $M(r_i)$ given in Eq. (3.13), since mass is conserved. The gravitational acceleration acting at an arbitrary time t on a particle with initial radius r_i is then given by

$$\vec{g} = -\frac{GM(r_i)}{r^2(r_i, t)} \hat{r} \ , \tag{3.14}$$

where $\hat{r}$ denotes a unit vector in the radial direction. Taking the radial component of this vector equation and using Eq. (3.13), one has

$$\ddot{r} = -\frac{4\pi}{3} \frac{G r_i^3 \rho_i}{r^2} \ , \tag{3.15}$$

where $r \equiv r(r_i, t)$, and an overdot denotes a derivative with respect to t. The initial condition on the velocity given in Eq. (3.12) can be rewritten in this notation as

$$\dot{r}(t{=}t_i) = H_i r_i \ . \tag{3.16}$$

Finally, the initial value of $r(r_i, t)$ is given by

$$r(r_i, t_i) = r_i \ . \tag{3.17}$$

The mathematical problem is then to solve Eq. (3.15), subject to the initial conditions of Eqs. (3.16) and (3.17).

First, note that the dependence on r_i in these equations can be eliminated by a simple rescaling of the as yet unknown function $r(r_i, t)$. That is, define

$$u(r_i, t) \equiv r(r_i, t)/r_i \ . \tag{3.18}$$

Note that r_i does not depend on t, and it can therefore be treated as a constant as far as time derivatives are concerned. Eqs. (3.15)-(3.17) can then be rewritten as

$$\ddot{u} = -\frac{4\pi}{3}\frac{G\rho_i}{u^2} \; , \tag{3.19}$$

$$\dot{u}(t{=}t_i) = H_i \; , \tag{3.20}$$

$$u(r_i, t_i) = 1 \; . \tag{3.21}$$

These equations specify the initial value and time derivative of u, and its acceleration at all times, and they therefore completely determine the function. Since these equations do not involve r_i, it follows that $u(r_i, t)$ does not actually depend on r_i at all. This means that $u(r_i, t)$ is really just an overall scale factor, and we can define

$$a(t) \equiv u(r_i, t) \; . \tag{3.22}$$

Eq. (3.18) then becomes $r(r_i, t) = a(t)r_i$, which means that the particle locations at any given time t are given by a rescaling of their original positions, by the scale factor $a(t)$. Note also that the mean mass density inside a sphere of radius $r(r_i$,t) is given by

$$\rho(t) = \frac{M(r_i)}{\frac{4\pi}{3}r^3} = \frac{\frac{4\pi}{3}r_i^3\rho_i}{\frac{4\pi}{3}r^3} = \frac{\rho_i}{a^3(t)} \; , \tag{3.23}$$

and is also independent of r_i. The mass density thus remains completely uniform. Using Eqs. (3.22) and (3.23), Eq. (3.19) can be rewritten as

$$\boxed{\ddot{a} = -\frac{4\pi}{3}G\rho(t)a \; .} \tag{3.24}$$

Eq. (3.24) describes how the expansion of the scale factor is slowed down by the gravitational effects of the mass density $\rho(t)$.

We can now return to the issue of shell crossing, and see that it never occurs. From Eqs. (3.18) and (3.22) we know that $r(r_i, t) = a(t)r_i$, as long as our equations are valid. Thus, if the first shell crossing occurs at time $t_{\rm shell}$, then the relation $r(r_i, t) = a(t)r_i$ must hold for all t between t_i and $t_{\rm shell}$. But if $r(r_i, t) = a(t)r_i$ holds at time $t_{\rm shell} - \epsilon$, for arbitrarily small $\epsilon > 0$, then there can be no shell crossing at $t = t_{\rm shell}$, since $r(r_i, t) = a(t)r_i$ implies that no two shells with different values of r_i are about to touch.

The limit $R_{\rm max} \to \infty$ is now seen to be trivial. As discussed earlier in the section on "The Homogeneously Expanding Universe" in Lecture Notes 2, this kind of uniform expansion by an overall scale factor $a(t)$ appears to be absolutely homogeneous to the inhabitants of this idealized universe. Looking from the outside we see a sphere with a center and an edge, but someone living anywhere inside the sphere would simply see all of his neighbors receding in a Hubble pattern, with the Hubble expansion rate given by

$$\boxed{H(t) = \dot{a}/a \ .} \tag{3.25}$$

Only someone living so near to the edge that he could actually see it would have any way of knowing that the system was not globally homogeneous. Thus, the limit $R_{\rm max} \to \infty$ which we need to take is trivial. In fact, for observers on the interior of the sphere, nothing whatever depends on $R_{\rm max}$.

A CONSERVATION OF ENERGY EQUATION:

The equations of the last section completely determine the behavior of the model universe, so our only remaining task is to examine the consequences of these equations.

As with most Newtonian systems, conservation of energy is a useful concept. Conservation of energy is of course not an independent statement, but instead follows as a consequence of the Newtonian equations of motion. In this case Eq. (3.19) can easily be used to obtain such an equation. [Eq. (3.24), which gives the deceleration in terms of the mass density ρ, is more useful for most purposes. But it cannot be used by itself to give a conservation of energy equation, since the time dependence is not determined until one adds information about the time dependence of $\rho(t)$. One can of course combine Eq. (3.24) with Eq. (3.23) describing the evolution of $\rho(t)$, but this is equivalent to using Eq. (3.19).] The conservation equation is obtained from Eq. (3.19) by first replacing u by a, then bringing both terms to one side, and then multiplying by $\dot{a}$:

$$\dot{a}\left\{\ddot{a} + \frac{4\pi}{3}\frac{G\rho_i}{a^2}\right\} = 0 \ .$$

Using elementary calculus, the result can be rewritten as

$$\frac{dE}{dt} = 0 \ , \tag{3.26}$$

where

$$E = \frac{1}{2}\dot{a}^2 - \frac{4\pi}{3}\frac{G\rho_i}{a} \ . \tag{3.27}$$

E is not exactly an energy, and does not even have the right units to be an energy. However, if one considers a test particle of mass m that moves with the Hubble expansion starting at radius r_i, then the quantity $E_{\text{phys}} \equiv m r_i^2 E$ is closely related to the energy of that particle. Specifically,

$$\begin{aligned} E_{\text{phys}} &= m r_i^2 \left\{ \frac{1}{2}\dot{a}^2 - \frac{4\pi}{3}\frac{G\rho_i}{a} \right\} \\ &= \frac{1}{2} m \left(\dot{a} r_i\right)^2 - \frac{G m M(r_i)}{a r_i} \ , \end{aligned}$$

where $M(r_i)$ is given by Eq. (3.13). Then, recognizing that $a(t) r_i$ is the radius r of the test particle at time t, we can rewrite E_{phys} as

$$E_{\text{phys}} = \frac{1}{2} m \left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 - \frac{G m M(r_i)}{r} \ . \tag{3.28}$$

This expression is the total energy of a particle of mass m moving radially in the gravitational field of a point particle of mass $M(r_i)$ located at the origin, where we have defined the zero of potential energy at infinity. If the test particle is at the edge of the sphere, with $r = R_{\max}(t)$, then this is the correct expression for the total energy of the test particle, since for all $r \geq R_{\max}(t)$, the gravitational effects of the sphere and a point mass are identical. If the test particle is at a smaller radius, however, then E_{phys} is still conserved, but it is not really the total energy. The mass that is located between r and $R_{\max}(t)$ would affect the amount of energy needed to bring the test particle from infinity, and hence would affect the potential energy of the test particle, but the effect of this mass is not included in E_{phys}. Nonetheless, the mass that is located between r and $R_{\max}(t)$ does not affect the motion of the test particle, so we can define an analogue problem in which this mass is absent. That is, we can define an analogue problem in which $R_{\max,i}$ is chosen so that the test particle is on the edge. For the analogue problem, E_{phys} is truly the total energy of the test particle. The motion of the test particle is the same for the analogue problem and the original problem, so we can understand the conservation of E_{phys} as a consequence of energy conservation for the analogue problem. It can also be shown, and you will have to opportunity to show on Problem Set 3, that E is proportional to the total enery, kinetic plus potential, of the entire sphere.

Using (3.23) to express ρ_i in terms of $\rho(t)$, Eq. (3.27) can be converted to the form

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3} G\rho + \frac{2E}{a^2} \ . \tag{3.29}$$

It is more or less standard notation to introduce the variable k, defined by

$$k = -\frac{2E}{c^2} \ , \tag{3.30}$$

and then to rewrite Eq. (3.29) as

$$\boxed{\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2}\ .} \tag{3.31}$$

Eqs. (3.24) and (3.31) are the two key results of Newtonian cosmology. As long as the mass density is dominated by nonrelativistic matter, as has been the case for most of the history of our universe, these equations are both identical to the corresponding equations obtained from general relativity. They are called the Friedmann equations, after Alexander Friedmann, the Russian meteorologist who first derived the equations from general relativity in 1922. Some authors, however, including Barbara Ryden, use the term Friedmann equation only for Eq. (3.31).

UNITS:

In Lecture Notes 2 I talked about a comoving coordinate system, with coordinates measured in "notches". The scale factor $a(t)$ is then measured in m/notch. The concept of a notch is actually not used, to my knowledge, in any of the standard cosmology texts, but nonetheless I find it a very useful way of thinking — it helps to clarify what exactly the scale factor is, and when it is needed in an equation.

In the last two sections the concept of a notch did not appear, so now I would like to reinstate it. As written, one would infer that the quantity r_i, denoting the radial coordinate of a given particle at time t_i, is to be measured in meters. Note, however, that we used the coordinate r_i not merely to describe the position of the particle at t_i, but also as a permanent label of the trajectory $r(r_i, t)$. The coordinates r_i are in fact being used as comoving coordinates, and only at the special time t_i does the unit of these comoving coordinates correspond to the meter. It thus makes sense to rename the unit of r_i as a notch. The time t_i is then the time at which 1 notch corresponds to 1 m. The trajectory function $r(r_i, t)$ continues to be measured in meters, so by Eqs. (3.18) and (3.22), the scale factor $a(t)$ has the units of m/notch. The variable k then has the units of notch^{-2}.

Note, by the way, that we have still not defined the notch, since the time t_i is completely arbitrary. There are two common conventions. Some books, such as the text by Barbara Ryden, define $a = 1$ m/notch today. Many other books, however, adopt the convention that whenever $k \neq 0$, one defines the notch such that k has the numerical value of ± 1. These books tend to leave the notch arbitrary when discussing the $k = 0$ case.

NATURE OF THE SOLUTIONS:

The equations of Newtonian cosmology have now been written down, and our only remaining task is to examine the behavior of the solutions.

The solutions belong to three different classes, depending on whether E is positive, negative, or zero. The qualitative behavior can be seen most clearly from Eq. (3.27),

$$E = \frac{1}{2}\dot{a}^2 - \frac{4\pi}{3}\frac{G\rho_i}{a} \ .$$

If E is positive ($k < 0$), then one sees that da/dt can never vanish, since it gives the only positive contribution to the right-hand side. Thus an expanding universe with $k < 0$ would continue to expand forever. In a universe of this type there is not enough mass to reverse the expansion of the Hubble flow. Such a universe is called open. On the other hand, if E is negative ($k > 0$), then one sees that da/dt equals zero when

$$a = -\frac{4\pi G\rho_i}{3E} \ . \tag{3.32}$$

This universe reaches a maximum size, and then the pull of gravity overcomes the expansion and causes the universe to collapse into what is sometimes called a "big crunch". A universe of this type is called closed. On the border between these two possibilities is the special case of $E = 0$ ($k = 0$). For reasons that will be discussed in Lecture Notes 5, such a universe is called flat.

The case $k = 0$ implies that the mass density ρ must have a special value ρ_c, which can be found from Eq. (3.31) (remembering that $\dot{a}/a = H$):

$$\boxed{\rho_c = \frac{3H^2}{8\pi G} \ .} \tag{3.33}$$

The quantity ρ_c is called the critical mass density — it is that mass density which puts the universe right on the borderline between eternal expansion and eventual collapse. Numerically, if one takes $H_0 = 100\, h_0$ km-s^{-1}-Mpc^{-1} (as in Eq. (3.3)) and $G = 6.674 \times 10^{-11}$ m^{3}kg^{-1}s^{-2}, one finds that

$$\rho_c = 1.88\, h_0^2 \times 10^{-26} \text{ kg/m}^3 = 1.88\, h_0^2 \times 10^{-29} \text{ g/cm}^3 \ , \tag{3.34}$$

where g is the abbreviation for gram. If $h_0 = 0.677$, which is the estimate from Planck 2018,* currently the best estimate, then $\rho_c = 8.6 \times 10^{-27}$ kg/m^3. The proton mass is

* N. Aghanim et al. (Planck Collaboration), "Planck 2018 results, VI: Cosmological parameters," Table 2, Column 6, arXiv:1807.06209.

1.67×10^{-27} kg, which means that the critical mass density corresponds to about 5.2 protons per cubic meter. Although this mass density seems amazingly small, the actual mass density of the universe is known to be very close to it.

The ratio ρ/ρ_c is standardly denoted by the Greek letter Omega (Ω). If the dark energy is included, then the total Ω for our universe is now known to an accuracy of about half of a percent, and to within this accuracy it is equal to 1* [This is good news for theorists, because the inflationary model of cosmology predicts that $\Omega = 1$, and there is also a theoretical plausibility argument in favor of $\Omega = 1$ that can be made independently of inflation. We will return to these issues later in the course.] According to the Planck 2018 estimates† the dark energy is believed to comprise about 69% of Ω. Only 5% of Ω is due to "ordinary" matter, like the material that we are made out of. "Ordinary" matter is also called *baryonic* matter, since the bulk of its mass is that of protons and neutrons, which belong to a class of particles called baryons. The remaining 26% of the total mass density is *dark matter*. This is matter that is known to exist because of its gravitational effects on other matter, but which is not detected in any other way. The composition of the dark matter is unknown, but it is most likely in the form of a dilute gas of some so-far-undiscovered weakly interacting particle.

The time evolution of the $k = 0$ case is rather easy to calculate. From Eq. (3.27), one sees that $E = 0$ implies that

$$\frac{da}{dt} = \frac{\text{const}}{a^{1/2}} \; . \tag{3.35}$$

The value of the *const* will not be relevant, since it will depend on the arbitrary definition of the notch. One can integrate this equation by rewriting it as

$$a^{1/2} da = \text{const}\; dt \; , \tag{3.36}$$

which integrates to give

$$\frac{2}{3} a^{3/2} = (\text{const}) t + c' \; . \tag{3.37}$$

The ambiguity of the constant of integration c' simply reflects our freedom to redefine the origin of time. It is traditional in big bang cosmology to define the zero of time to be the moment when the scale factor $a(t)$ vanishes — sometimes regarded as the instant of the big bang. One then has the following important result which holds for a flat, matter-dominated universe:

$$\boxed{a(t) \propto t^{2/3} \; .} \tag{3.38}$$

On Problem Set 2 you have explored the consequences of this behavior for the scale factor, and now you know how to derive it.

---

* Planck 2018 VI, op. cit., Table 4, Column 4.

† Planck 2018 VI, op. cit., Table 2, Column 6.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — October 1, 2018
Prof. Alan Guth

Lecture Notes 4

MORE DYNAMICS OF NEWTONIAN COSMOLOGY

THE AGE OF A FLAT UNIVERSE:

We showed last time that the scale factor for a flat, matter-dominated universe behaves as

$$a(t) = bt^{2/3} \ , \tag{4.1}$$

for some constant of proportionality b. This model universe is "flat" in the sense that $k = 0$ (or $\Omega = 1$), and it is "matter-dominated" in the sense that the mass density is dominated by the rest mass of nonrelativistic particles. (The assumption of matter-domination entered the derivation when we assumed that the total mass $M(r_i)$ contained within a given comoving radius r_i does not change as the system evolves. Photons or highly relativistic particles, by contrast, would redshift as the universe expands, and hence they would lose energy. In Lecture Notes 6 we will discuss the evolution of a universe dominated by relativistic particles, and in Lecture Notes 7 we will describe the effects of a cosmological constant, or vacuum energy. Since the energy loss of an expanding gas is proportional to its pressure, the "matter-dominated" case can also be described as the case in which the pressure is negligible.) Using Eq. (4.1), it is easy to calculate the age of such a universe in terms of the Hubble expansion rate. In Lecture Notes 2 it was shown that the Hubble expansion rate is given by

$$H(t) = \dot{a}/a \ , \tag{4.2}$$

where the over-dot has been used to denote a derivative with respect to time t. Thus,

$$H(t) = \frac{\frac{2}{3}bt^{-1/3}}{bt^{2/3}} = \frac{2}{3t} \ . \tag{4.3}$$

Recall that we have defined the origin of time so that the scale factor $a(t)$ vanishes at $t = 0$, so $t = 0$ is the earliest time that exists within the mathematical model. We therefore refer to t as the age of the universe, which can then be expressed in terms of the Hubble expansion rate by

$$\boxed{t = \frac{2}{3}H^{-1} \ .} \tag{4.4}$$

We should keep in mind, however, that we would be foolish to pretend that we actually understand the origin of the universe, so the phrase "age of the universe" is being used loosely. It is almost certain that our universe underwent an extremely hot dense phase at $t \approx 0$, and we can call this phase the big bang. The variable t represents the age of the universe since the big bang, but we can only speculate about whether the big bang actually represents the beginning of time. As we will see near the end of the course, the current understanding of inflationary cosmology suggests that the big bang was very likely not the beginning of time.

Consistency requires that the age of the universe be older than the age of the oldest stars, and this requirement has turned out to be a strong constraint. It was a very serious problem in the time shortly after Edwin Hubble's first measurement, when Hubble's bad estimate for the Hubble expansion rate led to age estimates of only several billion years. More recently, as quoted in Lecture Notes 2, the Planck satellite team, using their own data combined with other data, estimated that $H_0 = 67.66 \pm 0.42$ km-s^{-1}-Mpc^{-1}.* Using Eq. (4.4) and the relationship

$$\frac{1}{10^{10}\ \text{yr}} = 97.8\ \text{km-s}^{-1}\text{-Mpc}^{-1}\ ,$$

we find that a flat matter-dominated universe with a Hubble expansion rate in this range (i.e., 67.24 to 68.08 km-s^{-1}-Mpc^{-1}) must have an age between 9.58 and 9.70 billion years.

Today the oldest stars are believed to be those in globular clusters, which are tightly bound, nearly spherical distributions of stars that are found in the halos of galaxies. A careful study of the globular cluster M4 by Hansen et al.†, using data from 123 orbits of the Hubble Space Telescope, determined an age of 12.7 ± 0.7 billion years. Krauss and Chaboyer‡ argue that Hansen et al. did not take into account all the uncertainties, and claim that the globular clusters in the Milky Way (including M4) have an age of $12.6^{+3.4}_{-2.2}$ billion years. (Both sets of authors are quoting 95% confidence limit errors, also called 2σ errors, meaning that the probability of the true value lying in the quoted range is estimated at 95%.) Krauss and Chaboyer estimate that the stars must take at least 0.8 billion years to form, implying that the universe must be at least 11.2 billion years old.

---

* N. Aghanim et al. (Planck Collaboration), "Planck 2018 results, VI: Cosmological parameters," Table 2, Column 6, arXiv:1807.06209.

† B.M.S. Hansen et al., "The White Dwarf Cooling Sequence of the Globular Cluster Messier 4," *Astrophysical Journal Letters*, vol. 574, p. L155 (2002), http://arXiv.org/abs/astro-ph/0205087.

‡ L.M. Krauss and B. Chaboyer, "Age Estimates of Globular Clusters in the Milky Way: Constraints on Cosmology," *Science,* vol. 299, pp. 65-70 (2003), available with MIT certificates at http://www.sciencemag.org.libproxy.mit.edu/content/299/5603/65.abstract or for purchase at http://www.sciencemag.org/content/299/5603/65.abstract

(That is, the stars are at least 12.6 - 2.2 = 10.4 billion years old, and could not have formed until the universe was already 0.8 billion years old.)

With either of these estimates for the age of the oldest stars, however, the age of the universe is inconsistent with the measured value of the Hubble expansion rate and the assumption of a flat, matter-dominated universe. We will see later in these lecture notes that the age calculation gives a larger answer if we assume an open universe ($\Omega \equiv \rho/\rho_c < 1$), but inflationary cosmology predicts that Ω should be very close to 1, and starting with the first BOOMERANG long duration balloon experiment of 1997,* measurements of the fluctuations in the CMB have indicated a nearly flat universe. The CMB measurements now imply that $\Omega = 1$ to high accuracy. Combining their own data with results of other experiments, the Planck team† concluded that $\Omega = 0.9993 \pm 0.0037$. We will see in Lecture Notes 7, however, that the age discrepancy problem is completely resolved by the inclusion of dark energy. With about 70% of the mass density in the form of dark energy, $H_0 = 67.7$ km-s^{-1}-Mpc^{-1} is consistent with an age of 13.79 ± 0.02 billion years, which is our current best estimate‡ for the age of the universe. Since the age estimates are not consistent if we assume a matter-dominated flat universe, the age calculations help to support the proposition that our universe is dominated by dark energy, which we will discuss in detail in Lecture Notes 7.

THE BIG BANG SINGULARITY:

This mathematical model of the universe starts from a configuration with $a(t) = 0$, which corresponds to infinite density. From Eq. (4.3) we see that the initial value of the Hubble expansion rate is also infinite. We will see shortly that these infinities are not peculiarities of the flat universe model, but occur also in the models for either a closed or open universe. This instant of infinite density is called a singularity.

One should realize, however, that there is no reason to believe that the equations which we have used are valid in the vicinity of this singularity. Thus, although our mathematical models of the universe certainly begin with a singularity, it is an open question whether the universe actually began with a singularity. If we use our equations to follow the history of the universe further and further into the past, the universe becomes denser and denser without limit. At some point one encounters densities that are so far beyond our experience that the equations are no longer to be trusted. We will discuss later in the course where this point may occur, but for now I just want to make it clear that the singularity should not be considered a reliable consequence of the theory.

---

* P. D. Mauskopf et al., "Measurement of a peak in the cosmic microwave background power spectrum from the North American test flight of BOOMERANG," *Astrophysical Journal Letters,* vol. 536, pp. L59–L62 (2000), http://arXiv.org/abs/astro-ph/9911444.

† N. Aghanim et al. (Planck Collaboration), cited above, Table 4, Column 4.

‡ N. Aghanim et al. (Planck Collaboration), cited above, Table 2, Column 6.

THE HORIZON DISTANCE:

Since the age of the universe in the big bang model is finite, it follows that there is a theoretical upper limit to how far we can see. Since light travels at a finite speed, there will be particles in the universe that are so far away that light emitted from these particles will not yet have reached us. The present distance of the furthest particles from which light has had time to reach us is called the horizon distance. If the universe were static and had age t, then the horizon distance would be simply ct. In the real universe, however, everything is constantly in motion, and the value of the horizon distance has to be calculated.

The calculation of the horizon distance can be done most easily by using comoving coordinates. The speed of light rays in a comoving coordinate system was given in Eq. (2.8) as

$$\frac{dx}{dt} = \frac{c}{a(t)} \ . \tag{4.5}$$

(Recall that this formula is based on the statement that the speed of light in meters per second is constant, but the scale factor $a(t)$ is needed to convert from meters per second to notches per second.) One can then calculate the coordinate horizon distance (i.e., the horizon distance in "notches") by calculating the distance which light rays could travel between time zero and some arbitrary final time t. By integrating Eq. (4.5), one sees that

$$\ell_{c,\text{horizon}}(t) = \int_0^t \frac{c}{a(t')} dt' \ . \tag{4.6}$$

The physical horizon distance at time t is then given by

$$\ell_{p,\text{horizon}}(t) = a(t) \int_0^t \frac{c}{a(t')} dt' \ . \tag{4.7}$$

For the special case of a flat, matter-dominated universe, one can use Eq. (4.1) to obtain

$$\ell_{p,\text{horizon}}(t) = bt^{2/3} \int_0^t \frac{c}{bt'^{2/3}} dt' \ . \tag{4.8}$$

Carrying out the integration, the physical horizon distance for a flat matter-dominated universe is found to be

$$\boxed{\ell_{p,\text{horizon}}(t) = 3ct \ .} \tag{4.9}$$

The horizon distance can also be expressed in terms of the Hubble expansion rate, by using Eq. (4.4):

$$\boxed{\ell_{p,\text{horizon}}(t) = 2cH^{-1} \ .} \tag{4.10}$$

Taking $H_0 = 67.7$ km-s^{-1}-Mpc^{-1}, the horizon distance today, under the assumption of a flat matter-dominated universe, is found to be about 29 billion light-years.

The factor of 3 on the right-hand side of Eq. (4.9) implies that the horizon distance is three times larger than it would be in a static universe, and hence it is significantly larger than most of us would probably guess. The reason, of course, is that the horizon distance refers to the ***present*** distance of the most distant matter that can in principle be seen. However, the light that we receive from distant sources was emitted long ago, and the present distance is large because the matter has been moving away from us ever since. In the limiting case of matter exactly at the horizon, the light that we receive today left the object at $t = 0$, at the instant of the big bang.

You might wonder why light from every object did not reach us immediately, since the scale factor $a(t)$ was zero at $t = 0$, so the initial distance between any two objects was zero. To be honest, you are pretty safe in believing whatever you like about what happens at the initial singularity. The classical description certainly breaks down as the mass density approaches infinity, and there does not yet exist a satisfactory quantum description. So, if you want to believe that everything could communicate with everything else at the instant of the singularity, nobody whom I know could prove that you are wrong. But nobody whom I know could prove that you are right, either. If one ignores the possibility of communication at the singularity, however, it is then a well-defined question to ask how far light signals can travel once the classical description becomes valid. As $t \to 0$ the scale factor $a(t)$ approaches zero, but its time derivative $\dot{a}(t)$ approaches infinity. If we think about some object at coordinate distance ℓ_c from us, at early times its physical distance $a(t)\ell_c$ was arbitrarily small, but its velocity of recession $\dot{a}(t)\ell_c$ was arbitrarily large. If such an object emitted a light pulse in our direction at some very early time $t = \epsilon > 0$, even though the light pulse would have traveled toward us at the speed of light, the expansion of the universe was so fast that the distance between us and the light pulse would have initially increased with time. The coordinate distance that the light pulse could travel between then and now is at most equal to $\ell_{c,\text{horizon}}$, as given Eq. (4.6), so the pulse could reach us by now only if the present distance to the object is less than the horizon distance as given by Eq. (4.7).

EVOLUTION OF A CLOSED UNIVERSE:

The time evolution equations are easiest to solve for the case of a flat ($k = 0$) universe, but the equations for a closed or open universe are also soluble. We will first consider the closed universe, for which $k > 0$, $E < 0$, and $\Omega > 1$.

From Lecture Notes 3, we write the Friedmann equation as

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2} \;, \tag{4.11}$$

where the mass density ρ can be written as

$$\rho(t) = \left[\frac{a(t_i)}{a(t)}\right]^3 \rho(t_i) \ . \tag{4.12}$$

(The above equation is a slight generalization of Eq. (3.23), $\rho(t) = \rho_i/a^3(t)$. Eq. (3.23) was derived under the assumption that $a(t_i) = 1$, while Eq. (4.12) makes no such requirement.) Here we will use Eq. (4.12) to conclude that the quantity $\rho(t)a^3(t)$ is independent of time.

To obtain the desired solution in an economical way, it is useful to identify from the beginning the quantities of physical interest. Note that a is measured in units of, for example, meters per notch, while k is measured in units of notch^{-2}. Thus the quantity $a/\sqrt{k}$ has units of physical length (meters, for example), and is therefore independent of the definition of the notch. We will therefore choose

$$\tilde{a}(t) \equiv \frac{a(t)}{\sqrt{k}} \tag{4.13}$$

as the variable to use in our solution of the differential equation. Similarly we will use

$$\tilde{t} \equiv ct \ , \tag{4.14}$$

rather than t, so that all the spacetime variables have units of length.

Multiplying the Friedmann equation (4.11) by $a^2/(kc^2)$, it can be rewritten as

$$\begin{aligned} \frac{1}{kc^2}\left(\frac{da}{dt}\right)^2 &= \frac{8\pi}{3}\frac{G\rho a^2}{kc^2} - 1 \\ &= \frac{8\pi}{3}\frac{G\rho a^3}{k^{3/2}c^2}\frac{\sqrt{k}}{a} - 1 \ . \end{aligned} \tag{4.15}$$

In the second line the factors have been arranged so that we can rewrite it as

$$\left(\frac{d\tilde{a}}{d\tilde{t}}\right)^2 = \frac{2\alpha}{\tilde{a}} - 1 \ , \tag{4.16}$$

where

$$\alpha \equiv \frac{4\pi}{3}\frac{G\rho\tilde{a}^3}{c^2} \ . \tag{4.17}$$

Note that the parameter α also has the units of length. While the above expression for α contains the product $\rho(t)\tilde{a}^3(t)$, Eq. (4.12) guarantees that this quantity is independent of time, so α is a constant. Eq. (4.16) can be solved formally by rewriting it as

$$d\tilde{t} = \frac{d\tilde{a}}{\sqrt{\frac{2\alpha}{\tilde{a}} - 1}} = \frac{\tilde{a}\, d\tilde{a}}{\sqrt{2\alpha\tilde{a} - \tilde{a}^2}} \tag{4.18}$$

and then integrating both sides. One could use indefinite integrals, as we did for the $k = 0$ case in Eq. (3.37). Using indefinite integrals, the arbitrary constant of integration would become an arbitrary constant in the solution to the equation. We found, however, that the arbitrary constant could be eliminated by choosing the zero of time so that $a(t) = 0$ at $t = 0$. Here, mainly for the purpose of demonstrating an alternative method, I will use definite integrals. In this method the arbitrary constant in the solution will be fixed by the specification of the limits of integration. Following the standard convention, I will again choose the zero of time to be the time at which the scale factor is equal to zero. Using a subscript f to denote an arbitrary final time, one has

$$\tilde{t}_f = \int_0^{\tilde{t}_f} d\tilde{t} = \int_0^{\tilde{a}_f} \frac{\tilde{a}\, d\tilde{a}}{\sqrt{2\alpha\tilde{a} - \tilde{a}^2}} \ , \tag{4.19}$$

where $\tilde{a}_f \equiv \tilde{a}(\tilde{t}_f)$. The subscripts f will be dropped when the problem is finished, but for now it is convenient to use them to distinguish the limits of integration from the variables of integration.

The only remaining step is to carry out the integration shown in Eq. (4.19). Completing the square in the denominator, and then replacing the variable of integration by

$$x \equiv \tilde{a} - \alpha \ , \tag{4.20}$$

one finds

$$\begin{aligned} \tilde{t}_f &= \int_0^{\tilde{a}_f} \frac{\tilde{a}\, d\tilde{a}}{\sqrt{\alpha^2 - (\tilde{a} - \alpha)^2}} \\ &= \int_{-\alpha}^{\tilde{a}_f - \alpha} \frac{(x + \alpha)\, dx}{\sqrt{\alpha^2 - x^2}} \ . \end{aligned} \tag{4.21}$$

The integral can now be simplified by the trigonometric substitution

$$x = -\alpha \cos\theta \ , \tag{4.22}$$

which leads to

$$\tilde{t}_f = \alpha \int_0^{\theta_f} (1 - \cos\theta) d\theta = \alpha(\theta_f - \sin\theta_f) \ . \tag{4.23}$$

Since θ_f denotes the final value of θ, we can combine Eqs. (4.20) and (4.22) to find $x_f = \tilde{a}_f - \alpha = -\alpha\cos\theta_f$, so

$$\tilde{a}_f = \alpha(1 - \cos\theta_f) \ . \tag{4.24}$$

Dropping the subscript f and recalling the definitions (4.13) and (4.14), the two equations above provide a solution to our problem:

$$\boxed{\begin{aligned} ct &= \alpha(\theta - \sin\theta)\ , \qquad (4.25a)\\ \frac{a}{\sqrt{k}} &= \alpha(1-\cos\theta)\ . \qquad (4.25b)\end{aligned}}$$

These two equations provide a ***parametric*** description of the function $a(t)$. That is, Eq. (4.25a) determines in principle the function $\theta(t)$, and then Eq. (4.25b) defines the function $a\big(\theta(t)\big)$. (There is, however, no explicit expression for $\theta(t)$, so the function $a(t)$ cannot be constructed explicitly.) Some of you may recognize these equations as the equations for a cycloid. The curve which they trace can be generated by imagining a graph of $a/\sqrt{k}$ vs. ct, with a disk of radius α that rolls along the t-axis, as shown below. As the disk rolls, the point P traces the graph of $a/\sqrt{k}$ vs. ct:

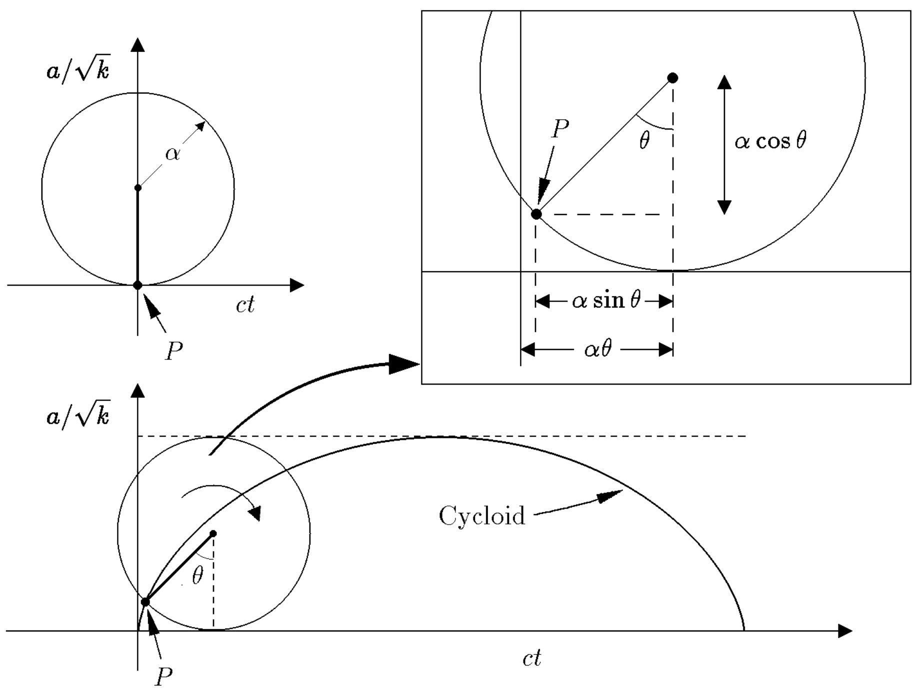


Figure 4.1: Evolution of a closed universe. The figure shows a graph of $a/\sqrt{k}$ vs. ct for a closed matter-dominated universe. If one imagines a circle rolling on the ct axis, a point on the circle traces out a cycloid, which is exactly the equation for $a/\sqrt{k}$ for a closed universe. The insert at the upper right includes labels for various distances, showing the connection with Eq. (4.25).

The relation between Eqs. (4.25) and the rolling disk can be seen in the insert at the top right: after the disk has rolled through an angle θ, the horizontal component of P is given by $\alpha\theta - \alpha\sin\theta$, and the vertical component is given by $\alpha - \alpha\cos\theta$.

The model for the closed universe reaches a maximum scale factor when $\theta = \pi$, which corresponds to a time $ct = \pi\alpha$. The corresponding value of a is given by

$$\frac{a_{\max}}{\sqrt{k}} = 2\alpha \; . \tag{4.26}$$

By that point the pull of gravity has overcome the inertia of the expansion, and the universe begins to contract. The scale factor during the contracting phase reverses the behavior it had during the expansion phase, and the universe ends in a "big crunch". The total lifetime of this universe is then

$$t_{\text{total}} = \frac{2\pi\alpha}{c} = \frac{\pi a_{\max}}{c\sqrt{k}} \; . \tag{4.27}$$

The angle θ is sometimes called the "development angle," because it describes the stage of development of the universe. The universe begins at $\theta = 0$, reaches its maximum expansion at $\theta = \pi$, and then is terminated by a big crunch at $\theta = 2\pi$.

Eqs. (4.25) contain the two parameters α and k, which might lead one to believe that there is a two-parameter family of closed universes. We must remember, however, that these equations still allow us the freedom to define the notch, so the numerical value of k is physically irrelevant. Many books, in fact, ***define*** k to always have the value $+1$ for a closed universe. The parameter α, on the other hand, is physically meaningful, and is related to the total lifetime of the closed universe. (When general relativity effects are described in Lecture Notes 5, we will learn that a closed universe actually has a finite size, and that the maximum size is determined by α.) Thus there is really only a one-parameter class of solutions.

THE AGE OF A CLOSED UNIVERSE:

The formula for the age of a closed universe can be obtained from the formulas in the previous section, but we have to do a little work. Eqs. (4.25) tell how to express the age in terms of α and θ, but this is not the result we want. We would prefer to relate α and θ to other quantities that are in principle measurable. Since we need to determine two variables, α and θ, we will have to imagine measuring two physical quantities. These two measurable quantities can be taken to be the Hubble expansion rate H and the mass density parameter $\Omega \equiv \rho/\rho_c$.

Our goal, then, is to express all the quantities related to the closed universe model in terms of H and Ω. To start, the mass density ρ can be rewritten as $\Omega\rho_c$, where $\rho_c = 3H^2/8\pi G$ (see Eq. (3.33)). So

$$\rho = \frac{3H^2\Omega}{8\pi G} \; , \tag{4.28}$$

or equivalently

$$\frac{8\pi}{3} G\rho = H^2\Omega \ . \tag{4.29}$$

Recalling that $H = \dot{a}/a$ (see Eq. (2.7)), the above formula can be substituted into the Friedmann equation (4.11), yielding

$$H^2 = H^2\Omega - \frac{kc^2}{a^2} \ , \tag{4.30}$$

which can then be solved for a^2 to give

$$\tilde{a}^2 = \frac{a^2}{k} = \frac{c^2}{H^2(\Omega - 1)} \ . \tag{4.31}$$

If we want to get our signs right for the entire evolution of the closed universe, we need to be careful. Note that Eq. (4.31) does not determine the sign of $\tilde{a}$, since it only specifies the value of $\tilde{a}^2$. The scale factor is by definition positive, however, and for the case under consideration $k > 0$. We adopt the standard convention that the square root of a positive number is positive, so $\sqrt{k} > 0$. Thus, the definition $\tilde{a} \equiv a/\sqrt{k}$ implies that $\tilde{a} > 0$, so Eq. (4.31) implies that

$$\tilde{a} = \frac{a}{\sqrt{k}} = \frac{c}{|H|\sqrt{\Omega - 1}} \ . \tag{4.32}$$

We write Eq. (4.32) with absolute value indicators around H, because during the contracting phase H is negative, while we know that only the positive square root of Eq. (4.31) is physically relevant.

We are now ready to evaluate α, using the definition (4.17). Using Eqs. (4.28) and (4.32) to replace ρ and $\tilde{a}$, we find

$$\alpha = \frac{c}{2|H|} \frac{\Omega}{(\Omega - 1)^{3/2}} \ . \tag{4.33}$$

Recall that α has direct physical meaning — if our universe is closed, then the total lifetime of the universe (from big bang to big crunch) would be given by $2\pi\alpha/c$.

The value of θ can now be found from Eq. (4.25b), using Eq. (4.32) to replace $a/\sqrt{k}$ on the left-hand side, and Eq. (4.33) to replace α on the right-hand side. After these substitutions, Eq. (4.25b) becomes

$$\frac{c}{|H|\sqrt{\Omega - 1}} = \frac{c}{2|H|} \frac{\Omega}{(\Omega - 1)^{3/2}} (1 - \cos\theta) \ , \tag{4.34}$$

which can be solved for either $\cos\theta$ or for Ω:

$$\cos\theta = \frac{2 - \Omega}{\Omega} \ , \tag{4.35}$$

$$\Omega = \frac{2}{1+\cos\theta} \ . \tag{4.36}$$

Using

$$\sin\theta = \pm\sqrt{1-\cos^2\theta} = \pm\frac{2\sqrt{\Omega-1}}{\Omega} \ , \tag{4.37}$$

Eq. (4.25a) can now be rewritten to obtain the desired formula for the age of the universe:

$$\boxed{t = \frac{\Omega}{2|H|(\Omega-1)^{3/2}}\left\{\arcsin\left(\pm\frac{2\sqrt{\Omega-1}}{\Omega}\right) \mp \frac{2\sqrt{\Omega-1}}{\Omega}\right\} \quad \text{(for a closed universe)},} \tag{4.38}$$

where the sign choices correspond to Eq. (4.37) (i.e., if the upper sign is used in Eq. (4.37), then both upper signs should be used in Eq. (4.38).) Both the square root and the inverse sine function are multibranched functions, so the evaluation of Eq. (4.38) requires some additional information. It is easy to see which branch to use, however, if one remembers that Eq. (4.38) is a rewriting of Eq. (4.25a), and that in Eq. (4.25a) the variable θ runs monotonically from 0 to 2π over the lifespan of the closed universe.

To describe the branches in detail, it is necessary to divide the full cycle, with θ varying from 0 to 2π, into quadrants. The first quadrant is from $\theta = 0$ to $\theta = \pi/2$, where we see from Eq. (4.36) that $\theta = \pi/2$ corresponds to $\Omega = 2$. Thus, the first quadrant corresponds to the beginning of the expanding phase, with $1 \leq \Omega \leq 2$. For this quadrant $\sin\theta > 0$, so we use the upper signs in Eq. (4.38), and the inverse sine is evaluated in the range 0 to $\pi/2$. The other quadrants are understood in the same way, producing the following table of rules:

Quadrant	Phase	Ω	Sign Choice	$\sin^{-1}()$
1	Expanding	1 to 2	Upper	0 to $\frac{\pi}{2}$
2	Expanding	2 to ∞	Upper	$\frac{\pi}{2}$ to π
3	Contracting	∞ to 2	Lower	π to $\frac{3\pi}{2}$
4	Contracting	2 to 1	Lower	$\frac{3\pi}{2}$ to 2π

Our universe is not currently matter-dominated, but it could be just barely closed. If we considered the hypothesis that our universe was matter-dominated, it would certainly be in the expanding phase, with $\Omega < 2$, and so it would be in the first quadrant. That means that the age would be given by Eq. (4.38), using the upper signs, and evaluating the inverse sine function in the range of 0 to $\pi/2$.

EVOLUTION OF AN OPEN UNIVERSE:

The evolution of an open universe can be calculated in a very similar way, except that one must use hyperbolic trigonometric substitutions in order to carry out the crucial integral. Fortunately, none of the complications with multibranched functions will occur in this case. For the open universe $k < 0$, $E > 0$, and $\Omega < 1$. We will define $\kappa \equiv -k$ to avoid the inconvenience of working with a negative quantity, and we will define

$$\tilde{a}(t) \equiv \frac{a(t)}{\sqrt{\kappa}} \tag{4.39}$$

instead of $\tilde{a} = a(t)/\sqrt{k}$. Eq. (4.16) is then replaced by

$$\left(\frac{d\tilde{a}}{d\tilde{t}}\right)^2 = \frac{2\alpha}{\tilde{a}} + 1 \ , \tag{4.40}$$

while Eq. (4.17),

$$\alpha \equiv \frac{4\pi}{3} \frac{G\rho\tilde{a}^3}{c^2} \ ,$$

continues to be valid, although the meaning of $\tilde{a}$ has changed. Eqs. (4.19) and (4.21) are then replaced by

$$\begin{aligned} \tilde{t}_f &= \int_0^{\tilde{a}_f} \frac{\tilde{a}\, d\tilde{a}}{\sqrt{2\alpha\tilde{a} + \tilde{a}^2}} \\ &= \int_0^{\tilde{a}_f} \frac{\tilde{a}\, d\tilde{a}}{\sqrt{\left(\tilde{a}+\alpha\right)^2 - \alpha^2}} \\ &= \int_\alpha^{\tilde{a}_f+\alpha} \frac{(x-\alpha)\, dx}{\sqrt{x^2-\alpha^2}} \ . \end{aligned} \tag{4.41}$$

In this case one uses the substitution

$$x = \alpha \cosh\theta \ , \tag{4.42}$$

and one obtains the result

$$\boxed{\begin{aligned} ct &= \alpha(\sinh\theta - \theta) \ , \qquad (4.43a) \\ \frac{a}{\sqrt{\kappa}} &= \alpha(\cosh\theta - 1) \ . \qquad (4.43b) \end{aligned}}$$

THE AGE OF AN OPEN UNIVERSE:

The methods are again identical to the ones used previously, so I will just state the results. Eqs. (4.35), (4.36) and (4.37) are replaced by

$$\Omega = \frac{2}{1+\cosh\theta} \;, \quad \cosh\theta = \frac{2-\Omega}{\Omega} \tag{4.44}$$

and

$$\sinh\theta = \sqrt{\cosh^2\theta - 1} = \frac{2\sqrt{1-\Omega}}{\Omega} \;, \tag{4.45}$$

and the final result for the age of the universe becomes

$$\boxed{\begin{gathered} t = \frac{\Omega}{2H(1-\Omega)^{3/2}}\left\{\frac{2\sqrt{1-\Omega}}{\Omega} - \sinh^{-1}\left(\frac{2\sqrt{1-\Omega}}{\Omega}\right)\right\} \\ \text{(for an open universe).} \end{gathered}} \tag{4.46}$$

Below is a graph of Ht versus Ω, using Eqs. (4.38) and (4.46). The graph shows that the curve is actually continuous at $\Omega = 1$, even though the expressions (4.38) and (4.46) look rather different. In fact, these two expressions are really not so different. Although it is not obvious, the two expressions are different ways of writing the same analytic function. You can verify this by using the usual techniques to evaluate square roots of negative numbers, and the trigonometric and hypertrigonometric functions of imaginary arguments.

To summarize, the age the universe can be expressed as a function of H and Ω as

$$|H|\,t = \begin{cases} \dfrac{\Omega}{2(1-\Omega)^{3/2}}\left[\dfrac{2\sqrt{1-\Omega}}{\Omega} - \sinh^{-1}\left(\dfrac{2\sqrt{1-\Omega}}{\Omega}\right)\right] & \text{if } \Omega < 1 \\ 2/3 & \text{if } \Omega = 1 \\ \dfrac{\Omega}{2(\Omega-1)^{3/2}}\left[\sin^{-1}\left(\pm\dfrac{2\sqrt{\Omega-1}}{\Omega}\right) \mp \dfrac{2\sqrt{\Omega-1}}{\Omega}\right] & \text{if } \Omega > 1 \end{cases} \tag{4.47}$$

Graphically,

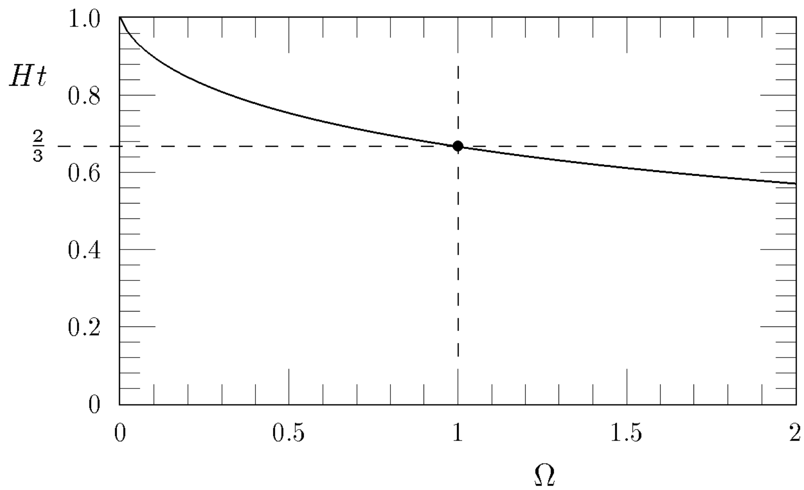


Figure 4.2: The age of a matter-dominated universe, expressed as Ht (where t is the age and H is the Hubble expansion rate), as a function of Ω. The curve describes all three cases of an open ($\Omega < 1$), flat ($\Omega = 1$), and closed ($\Omega > 1$) universe.

THE EVOLUTION OF A MATTER-DOMINATED UNIVERSE:

The following graph shows the evolution of the scale factor for all three cases: open, closed, or flat. The graphs were constructed using Eqs. (4.1) for the flat case, (4.43) for the open case, and (4.25) for the closed case. The parametric equations are plotted by choosing a finely spaced grid of values for θ, using the formulas to determine t and a for each value of θ. Since the only parameter, α, appears merely as an overall factor, a graph that is valid for all values of α can be obtained by plotting the dimensionless ratios $a/(\alpha\sqrt{|k|})$ and ct/α:

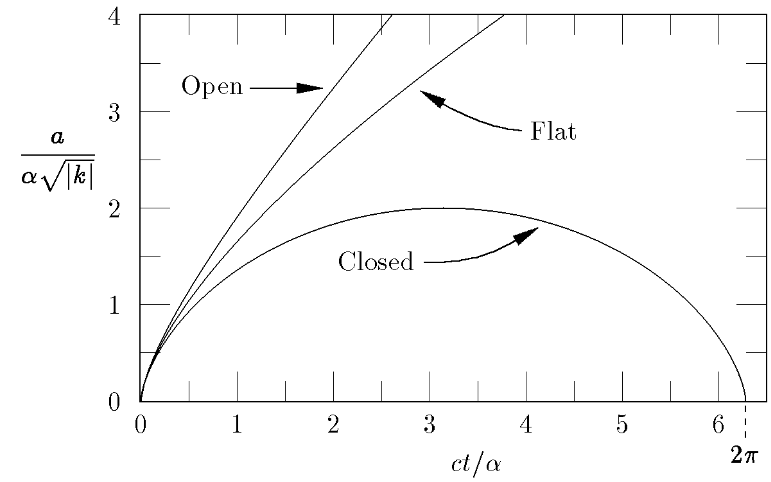


Figure 4.3: The evolution of a matter-dominated universe. Closed and open universes can be characterized by a single parameter α, defined by Eq. (4.17). With the scalings shown on the axis labels, the evolution of a matter-dominated universe is described in all cases by the curves shown in this graph.

Although the graph shows all three cases, it must be remembered that it is still restricted by the assumption that the universe is matter-dominated — that is, the mass density is dominated by nonrelativistic matter, for which pressure forces are negligible.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe October 13, 2018
Prof. Alan Guth

Lecture Notes 5

INTRODUCTION TO NON-EUCLIDEAN SPACES

INTRODUCTION:

The history of non-Euclidean geometry is a fascinating subject, which is described very well in the introductory chapter of *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* by Steven Weinberg. Here I would like to summarize the important points. Although historical in its organization, this section describes some essential mathematics and should be read carefully.

Euclid showed in his *Elements* how geometry could be deduced from a few definitions, axioms, and postulates. One of Euclid's assumptions, however, seemed to generations of mathematicians to be somewhat less obvious than the others. This assumption, known as Euclid's fifth postulate, was stated by Euclid as follows:

> "If a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines if produced indefinitely meet on that side on which the angles are less than two right angles." *[This statement is interpreted to imply that the two straight lines will never meet if extended on the opposite side.]*

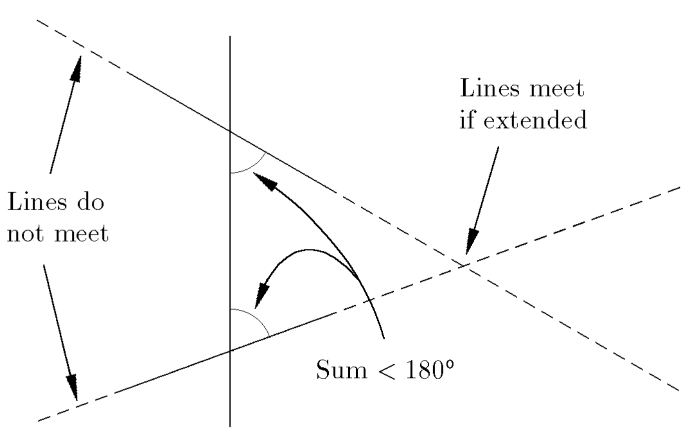


Figure 5.1: Euclid's fifth postulate.

Many mathematicians attempted to prove this postulate from the other assumptions, but all of these attempts ended in failure. It was discovered, however, that the fifth postulate could be replaced by any of a number of equivalent statements, such as:

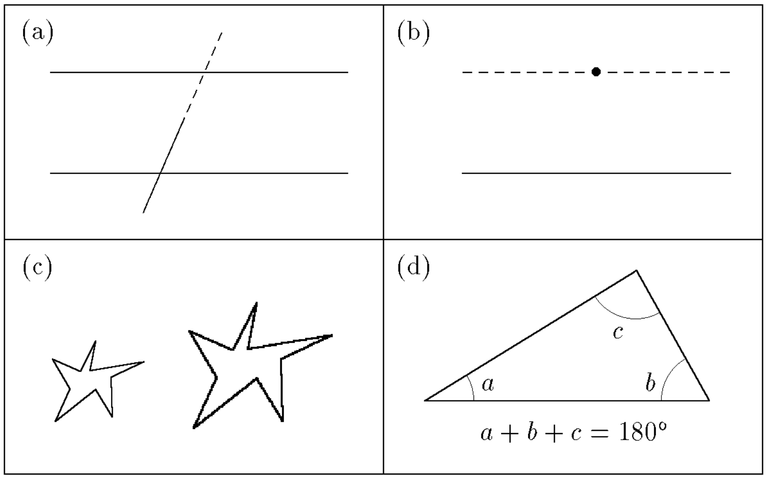


Figure 5.2: Statements equivalent to the fifth postulate.

(a) "If a straight line intersects one of two parallels (i.e, lines which do not intersect however far they are extended), it will intersect the other also."

(b) "There is one and only one line that passes through any given point and is parallel to a given line."

(c) "Given any figure there exists a figure, similar* to it, of any size."

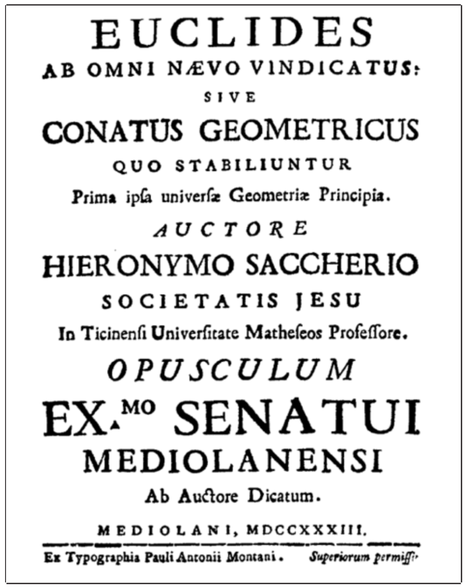
(d) "There is a triangle in which the sum of the three angles is equal to two right angles (i.e., 180°)."

Given Euclid's other assumptions, each of the above statements is equivalent to the fifth postulate.

The attitude of mathematicians toward the fifth postulate underwent a marked change during the eighteenth century, when mathematicians began to consider the possibility of abandoning the fifth postulate. In 1733 the Jesuit Giovanni Geralamo Saccheri (1667–1733) published a study of what geometry would be like if the postulate were false. He, however, was apparently convinced that the fifth postulate must be true, and he pursued this work because he hoped to discover an inconsistency — he didn't.

Carl Friedrich Gauss (1777-1855) seems to have been the first to really take seriously the possibility that the fifth postulate could be false. He, János Bolyai (an Austrian army officer, 1802-1860), and Nikolai Ivanovich Lobachevsky (a Russian mathematician, 1793-1856) independently discovered and explored a geometry which in modern terms is described as a two-dimensional space of constant negative curvature. The space is infinite

* Two polygons are similar if their corresponding angles are equal, and their corresponding sides are proportional.



EUCLIDES
AB OMNI NÆVO VINDICATUS:
SIVE
CONATUS GEOMETRICUS
QUO STABILIUNTUR
Prima ipsa universæ Geometriæ Principia.
AUCTORE
HIERONYMO SACCHERIO
SOCIETATIS JESU
In Ticinensi Universitate Matheseos Professore.
OPUSCULUM
EX.MO SENATUI
MEDIOLANENSI
Ab Auctore Dicatum.
MEDIOLANI, MDCCXXXIII.
Ex Typographia Pauli Antonii Montani. Superiorum permiss

Figure 5.3: The frontispiece of Giovanni Geralamo Saccheri's 1733 book titled *Euclides ab omni naevo vindicatus* (Euclid Freed of Every Flaw). Saccheri pursued the consequences of assuming that the fifth postulate was false, hoping to find a contradiction.

in extent, is homogeneous and isotropic, and satisfies all of Euclid's assumptions except for the fifth postulate. In this space every one of the statements of the fifth postulate and its equivalents listed above are false — through a given point there can be drawn ***infinitely*** many lines parallel to a given line; ***no*** figures of different size are similar; and the sum of the angles of any triangle is ***less*** than 180°.

The surface of a sphere, it should be pointed out, satisfies all the postulates of Euclid except for the fifth and the second, which states that "Any straight line segment can be extended indefinitely in a straight line." From a modern point of view the surface of a sphere provides a perfectly interesting example of a non-Euclidean geometry. Historically, however, this example was not taken very seriously, apparently because it seemed too simple. The great circles would be the objects that play the role of straight lines, but since any two great circles intersect, there could be no such thing as parallel lines.

Despite the work of Gauss, Bolyai, and Lobachevsky, it was still not clear that their non-Euclidean geometry was logically consistent. This problem was not solved until 1870, when Felix Klein (1849-1925) developed an "analytic" description of this geometry. In Klein's description, a "point" of the Gauss-Bolyai-Lobachevsky (G-B-L) geometry can be described by two real number coordinates (x,y), with the restriction

$$x^2 + y^2 < 1 \ . \tag{5.1}$$

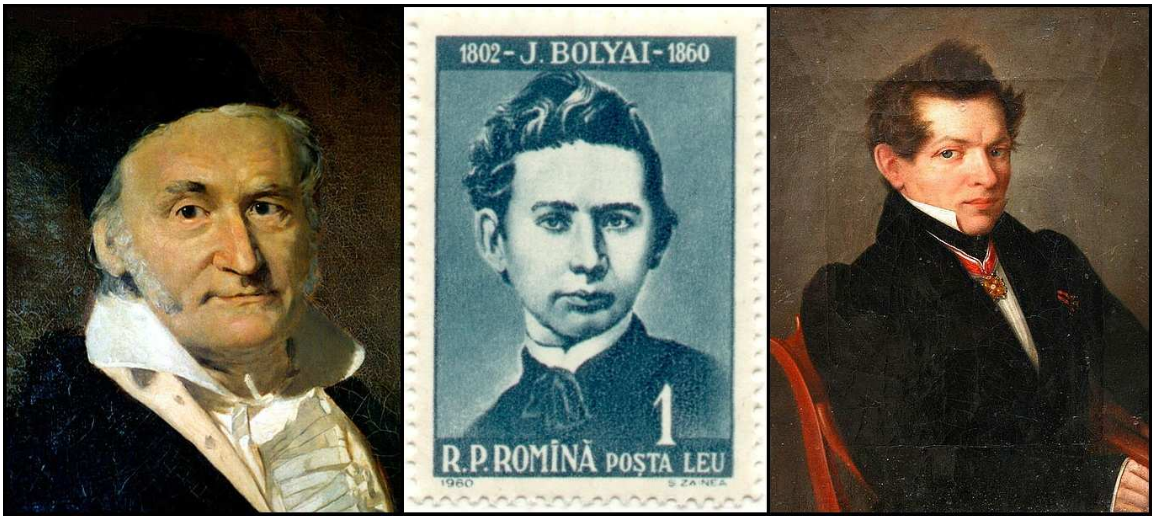


Figure 5.4: Carl Friedrich Gauss, János Bolyai, and Nikolai Ivanovich Lobachevsky independently developed the first example of a mathematical theory in which Euclid's fifth postulate is false, now known as the Gauss–Bolyai–Lobachevsky geometry. Gauss (1777–1855) was the son of poor working-class parents in Germany, but by the time he was 15 his mathematical talents were noticed by the Duke of Brunswick, who sent Gauss to the Collegium Carolinum and then the University of Göttingen. Gauss remained at Göttingen for the rest of his life, becoming Professor of Astronomy and director of the astronomical observatory in 1807. His students included Richard Dedekind, Bernhard Riemann, Peter Gustav Lejeune Dirichlet, Gustav Kirchhoff, August Ferdinand Möbius, and Friedrich Bessel. Bolyai (1802–1860) was the son of Farkas Bolyai, a teacher of mathematics, physics, and chemistry at the Calvinist College in Marosvásárhely, Hungary (now Tirgu-Mures, Romania). Although his father was well-educated, he was nonetheless not well paid, so János attended Marosvásárhely College and later studied military engineering at the Academy of Engineering at Vienna, because that is what they could afford. He then entered the army engineering corps, where he served for 11 years, during which time he carried out his now-famous investigation of non-Euclidean geometry. The work was published in 1831 as an appendix in a book written by his father. Bolyai resigned from the army in 1833 due mainly to health problems, and lived the rest of his life in relative poverty, dying at the age of 57 of pneumonia. The Romanian postage stamp shown here honored the 100th anniversary of Bolyai's death; the picture was apparently fabricated, as no authentic picture of Bolyai is known to exist. Lobachevsky (1792–1856) was the son of Polish parents living in Russia. His father was a clerk in a land-surveying office, who died when Lobachevsky was only seven. His mother relocated the family to Kazan, Russia, where Lobachevsky attended Kazan Gymnasium and later was given a scholarship to Kazan University, where one of his professors was Martin Bartels, who was a teacher and friend of Gauss. Lobachevsky remained at Kazan University for the rest of career, becoming rector of the university in 1827. His work on non-Euclidean geometry was published in the *Kazan Messenger* in 1829, but was rejected for publication by the St. Petersburg Academy of Sciences. Lobachevsky was asked to retire in 1846, and after that his health and financial situation deteriorated, he became blind, and his favorite eldest son died. Lobachevsky himself died before the importance of his work in mathematics was appreciated.

The distance $d(1,2)$ between two points (x_1, y_1) and (x_2, y_2) is then defined to be

$$\cosh\left[\frac{d(1,2)}{a}\right] = \frac{1 - x_1 x_2 - y_1 y_2}{\sqrt{1 - x_1^2 - y_1^2}\,\sqrt{1 - x_2^2 - y_2^2}} \ , \tag{5.2}$$

where a is a fundamental length which sets a scale for the geometry. Note that the space is infinite despite the coordinate restriction of Eq. (5.1), because the distance approaches infinity as either $x_1^2 + y_1^2 \to 1$ or $x_2^2 + y_2^2 \to 1$. Klein showed that with this definition of point and distance the model satisfies all of the assumptions of the G-B-L geometry. Thus, assuming the consistency of the real number system, the consistency of the G-B-L geometry was established. In addition, this work reinforced the important idea of analytic geometry which had been introduced by Descartes. It has since proven to be very useful to describe a geometry not by listing axioms, but instead by giving an explicit description in terms of a coordinate system and distance function.

Gauss went on to develop two very central ideas in non-Euclidean geometry. The first is the distinction between the "inner" and "outer" properties of a surface. The inner properties of a surface are those distance relationships that can be measured within the surface itself, such as in Eq. (5.2). The outer properties refer to the way in which a space might be embedded in a higher dimensional space. For example, the surface of a sphere is a two-dimensional space which we visualize by embedding in a three-dimensional space. Gauss emphasized that the distance relationships within the two-dimensional surface itself provide a complete mathematical system which can be studied independently of any assumptions about the embedding in the three-dimensional space. Gauss wrote in 1827 that it is the inner properties of the surface that are "most worthy of being diligently explored by geometers." Note that the G-B-L geometry cannot be fully embedded in a three-dimensional Euclidean space, although finite patches of it can be so embedded. To describe the whole space, it is necessary to describe it in terms of its inner properties.

Gauss's second central idea had to do with the form of the distance function $d(1,2)$. It turns out that if one allows this function to have any form, then the class of geometries is so unconstrained that nothing very interesting results. Gauss realized first that one need not specify $d(1,2)$ for arbitrary points 1 and 2. It is sufficient to consider only infinitesimal line segments. Such a line segment can be described as extending from the point (x, y) to $(x + \mathrm{d}x, y + \mathrm{d}y)$. The length of a finite segment of a curve is then defined by summing up (integrating) the lengths of the infinitesimal segments that make it up. The distance $d(1,2)$ between two arbitrary points can then be defined as the length of the shortest curve which joins the two points. The concept of a *line* is replaced by a *geodesic,* defined to be any curve that is the shortest path between its endpoints. More precisely, a geodesic is not necessarily the true minimum of the path length — it is only necessary that the path is ***stationary***, in the sense that the first derivative with respect to any variation of the path between the two endpoints must vanish. The path length might then be a minimum, a maximum, or a saddle point.

For the length of the infinitesimal line segment from (x, y) to $(x + \mathrm{d}x, y + \mathrm{d}y)$, Gauss realized that the interesting case is to restrict one's attention to functions for which the squared segment length $\mathrm{d}s^2$ is quadratic in $\mathrm{d}x$ and $\mathrm{d}y$ (i.e., functions for which each term contains two powers of $\mathrm{d}x$ and/or $\mathrm{d}y$). Such functions can be written as

$$\mathrm{d}s^2 = g_{xx}\mathrm{d}x^2 + g_{xy}\mathrm{d}x\,\mathrm{d}y + g_{yx}\mathrm{d}y\,\mathrm{d}x + g_{yy}\mathrm{d}y^2 \ , \tag{5.3}$$

where g_{xx}, g_{xy}, g_{yx}, and g_{yy} are functions of position (x, y) and are together called the metric of the space. (Since g_{xy} and g_{yx} both multiply $\mathrm{d}x\,\mathrm{d}y$, only their sum is relevant. By convention one sets $g_{xy} = g_{yx}$.) Gauss showed that the assumption that $\mathrm{d}s^2$ is quadratic is equivalent to the assumption that in any infinitesimal region it is possible to choose a coordinate system (x', y') in which the distance relation is Euclidean: $\mathrm{d}s^2 = \mathrm{d}x'^2 + \mathrm{d}y'^2$. Today spaces with a metric of this form are generally called either metric spaces or Riemannian spaces.

In Euclidean space one can use any coordinate system one wants, although one usually prefers a Cartesian system in which the metric has the form:

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 \ . \tag{5.4}$$

Any two systems with metrics of this form are related to each other by a translation and/or a rotation. For some purposes, however, it is convenient to use polar coordinates r and θ, for which the metric is given by

$$\mathrm{d}s^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2 \ . \tag{5.5}$$

Thus, the mere fact that the metric does not have the Cartesian form of Eq. (5.4) does ***not*** imply that the underlying space is non-Euclidean — one might simply be using a non-Cartesian coordinate system. It is therefore useful to have some way of describing the inner curvature of a space in a way which is not confused by the choice of a coordinate system. Such a method was developed for two-dimensional spaces by Gauss, who showed that the underlying space is Euclidean if and only if a somewhat complicated expression involving derivatives of the metric is equal to zero. The extension to more than two dimensions was carried out by Georg Friedrich Bernhard Riemann (1826-1866). The details of the Gaussian curvature and the Riemann curvature tensor are beyond the level of this discussion.

GENERAL RELATIVITY:

As I have mentioned before, Einstein's theory of general relativity is nothing more nor less than a theory of gravity. When Einstein invented the special theory of relativity in 1905, he realized immediately that it was inconsistent with Newton's theory of gravity. The inconsistency has nothing in particular to do with the inverse square nature of the force law, and it cannot be remedied by simply modifying the way that the force depends on the distance. Rather, the inconsistency is due to the fact that Newton's law of gravity assumes that the force between two bodies depends instantaneously on the distance between them. That is, to determine the force due to body B acting on body A at time t, one must merely know the position of the two bodies at time t. However, as we discussed in Lecture Notes 1, special relativity implies that the synchronization of clocks depends on the velocity of the observer. Thus, two observers who are moving relative to each other will not agree on what it means to measure the positions of A and B at the same time, and so a physically meaningful quantity like a force cannot be determined by these two positions. If special relativity is correct, then Newton's law of gravity must be modified.

The idea of an action-at-a-distance theory is not completely ruled out by special relativity, but it is very difficult to formulate such a theory. The electromagnetic force of one charged particle acting on another can be expressed by an action-at-a-distance law, but it is rather complicated. (The force law is stated, for example, in *The Feynman Lectures on Physics*, Volume 1, by R.P. Feynman, R.B. Leighton, and M. Sands.) The force on charge A at time t does not depend on the position of charge B at time t, but instead depends on the position (and velocity, and acceleration!) of charge B at a retarded time t'. The time t' is determined by the rule that a light pulse (moving at speed c) can just barely travel from B to A in the time interval from t' to t, as illustrated in the following diagram:

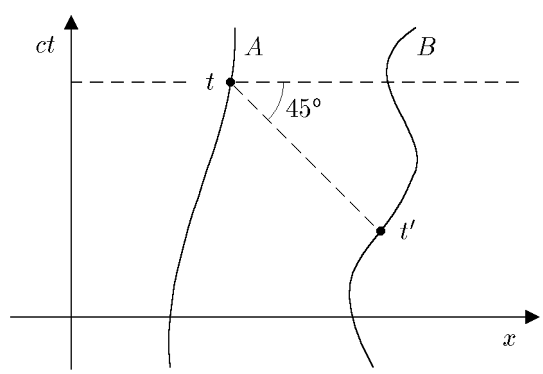


Figure 5.5: Definition of the retarded time t'. The electromagnetic force on particle A at time t, due to particle B, can be expressed in terms of the position, velocity, and acceleration of charge B at the retarded time t'.

Two different observers will agree when this relationship is met, since they agree on what it means for a trajectory to move at the speed of light. However, the two observers will measure different values for the positions, velocities, and accelerations, and it requires a very complicated force law such that both observers will conclude that the law is satisfied.

The simplest way to formulate electromagnetic theory is to avoid action-at-a-distance forces, but instead to use the concept of a field. The electric and magnetic fields are each defined at all points in space, and a charged particle interacts only with the fields at the location of the particle. The evolution of the fields is governed by Maxwell's equations. These equations allow information about the changing position of a particle to propagate in the form of waves which travel at the speed of light.

General relativity is also a theory of fields, similar in type to the Maxwell theory of electromagnetism. In the case of general relativity there is no known action-at-a-distance formalism. The "fields" which are involved in general relativity are of course not the electric and magnetic fields of the Maxwell theory. The fields of general relativity are in fact the metric functions defined earlier. Space and time must be considered together, and it is the metric functions on this "spacetime" which are the fields that general relativity uses to describe gravitation. We will see later that in this curved (i.e., non-Euclidean) spacetime, a freely falling particle is assumed to travel along a geodesic. The attractive effect of gravity then appears simply as a distortion of spacetime.

THE SURFACE OF A SPHERE:

As mentioned above, the surface of a sphere embedded in a three-dimensional Euclidean space is a perfectly good example of a non-Euclidean geometry. In order to develop some of the techniques of non-Euclidean geometry, we begin by studying this familiar system. Since the three-dimensional embedding space is Euclidean, we can use our knowledge of Euclidean geometry to learn about the non-Euclidean two-dimensional geometry of the surface of the sphere. Beware, however, that not all two-dimensional curved surfaces can be embedded in a three-dimensional Euclidean space.

The surface of the sphere can be described by using Cartesian coordinates (x, y, z)

in the three-dimensional space, in which case the surface is given by:

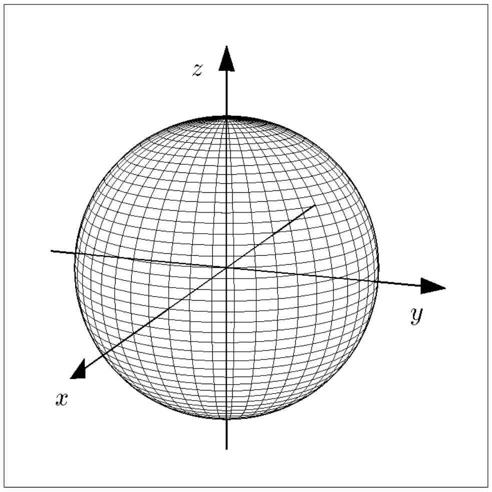


Figure 5.6: A sphere in Cartesian coordinates.

$$x^2 + y^2 + z^2 = R^2 \ , \tag{5.6}$$

where R is the radius of the sphere. We now want to take seriously the notion that the two-dimensional space of the surface defines a two-dimensional geometry with "inner" properties that are independent of the existence of the third dimension. We take the point of view that the third dimension has been introduced only as an aid in visualizing the two-dimensional surface. This third dimension can of course be useful, because in the three-dimensional picture the properties of homogeneity and isotropy are obvious. (Recall here that homogeneity and isotropy refer to properties of the *two-dimensional* space. Homogeneity means that all points on the surface of the sphere are equivalent. Isotropy means that if a two-dimensional creature living in the two-dimensional surface were to look in all directions within the two-dimensional surface, he would see the same thing in all directions.)

In order to describe the two-dimensional world without reference to the third dimension, it is useful to introduce a two-dimensional coordinate system. The most natural choice is to use the usual angular variables θ and ϕ, as shown in Fig. 5.7.

From the diagram we can see that x, y, and z can be expressed as

$$\begin{aligned} x &= R \sin\theta \cos\phi \\ y &= R \sin\theta \sin\phi \\ z &= R \cos\theta \ , \end{aligned} \tag{5.7}$$

where θ runs from 0 to π and ϕ runs from 0 to 2π.

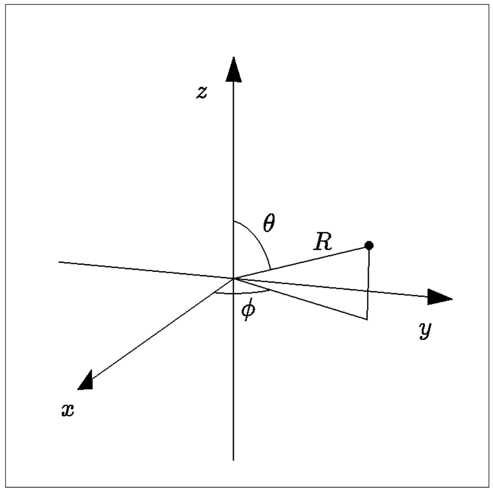


Figure 5.7: Spherical polar coordinates for the surfaces of a sphere.

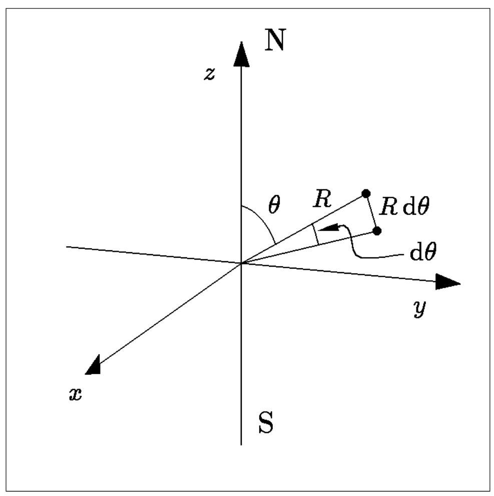


Figure 5.8: Variation of θ in spherical polar coordinates: $\mathrm{d}s = R\mathrm{d}\theta$.

To describe the inner properties of this two-dimensional space, we must write down an expression for the metric. That is, we need an expression for the distance $\mathrm{d}s$ between two points on the surface labelled by (θ, ϕ) and $(\theta + \mathrm{d}\theta, \phi + \mathrm{d}\phi)$. It is helpful to think about varying θ and ϕ one at a time. As θ is increased, the point moves a distance $R\,\mathrm{d}\theta$ toward the south (where I am using the positive z-axis to define a North pole), as can be seen in Fig. 5.8.

When ϕ is increased, the point moves toward the east, tracing out a circle at constant latitude. The radius of the circle is $R \sin\theta$, and so the distance moved is given by $R \sin\theta\,\mathrm{d}\phi$, as shown in the Fig. 5.9.

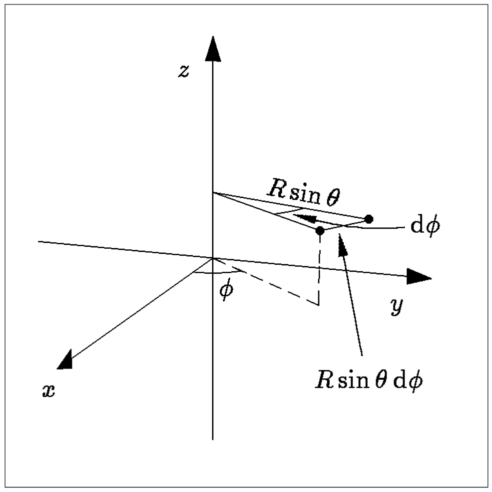


Figure 5.9: Variation of ϕ in spherical polar coordinates: $\mathrm{d}s = R\sin\theta\,\mathrm{d}\phi$.

Since these two displacements are in orthogonal directions, the total distance is given by the Pythagorean theorem:

$$\mathrm{d}s^2 = R^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right) \ . \tag{5.8}$$

Eq. (5.8) describes the metric of the two-dimensional space.

If one wishes to avoid the pictures, one can also derive Eq. (5.8) directly from Eqs. (5.7), by writing

$$\mathrm{d}x = \frac{\partial x}{\partial\theta}\mathrm{d}\theta + \frac{\partial x}{\partial\phi}\mathrm{d}\phi = R\cos\theta\cos\phi\,\mathrm{d}\theta - R\sin\theta\sin\phi\,\mathrm{d}\phi \ ,$$

$$\mathrm{d}y = \frac{\partial y}{\partial\theta}\mathrm{d}\theta + \frac{\partial y}{\partial\phi}\mathrm{d}\phi = R\cos\theta\sin\phi\,\mathrm{d}\theta + R\sin\theta\cos\phi\,\mathrm{d}\phi \ ,$$

and

$$\mathrm{d}z = \frac{\partial z}{\partial\theta}\mathrm{d}\theta + \frac{\partial z}{\partial\phi}\mathrm{d}\phi = -R\sin\theta\,\mathrm{d}\theta \ . \tag{5.9}$$

These expressions can then be substituted into

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \ , \tag{5.10}$$

and after some algebra involving repeated use of the identity $\sin^2\phi + \cos^2\phi = 1$, one again obtains Eq. (5.8).

A CLOSED THREE-DIMENSIONAL SPACE:

The goal here is to use the same techniques to describe a closed three-dimensional space. This space will be homogeneous and isotropic, and it will have a finite volume but no boundary. Since the space is homogeneous and isotropic, it is a candidate for the space in which we live.

To derive a metric for the three-dimensional space, one simply repeats the steps carried out above with one additional dimension. One begins therefore in a Euclidean space with four dimensions, and hence with four Cartesian coordinates which I will call (x, y, z, w). The surface of a sphere in this four-dimensional space is then described by the equation

$$x^2 + y^2 + z^2 + w^2 = R^2 \ . \tag{5.11}$$

Note that the surface of the sphere is a three-dimensional space, since it can be described by three coordinates.

To explicitly describe the surface by three coordinates, one can introduce one more angular variable in addition to θ and ϕ. We therefore introduce ψ, which will represent the angle between the point being described and the w-axis. Since ψ measures the angle from an axis, like θ it ranges from 0 to π. One can then look at the point projected into the x-y-z subspace and define the variables θ and ϕ as we did above. (By "project into the x-y-z subspace", I simply mean to ignore the w-coordinate.) Pictorially one would depict ψ as

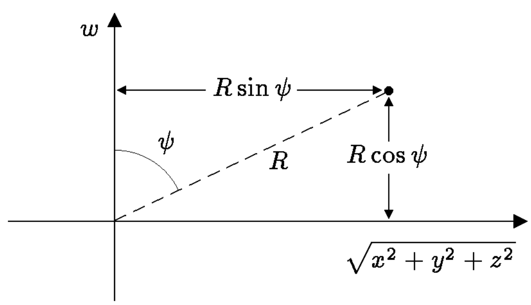


Figure 5.10: The new angular variable ψ, which measures the angle from the w-axis.

and in terms of equations it can be expressed as

$$\begin{aligned} x &= R \sin\psi \sin\theta \cos\phi \\ y &= R \sin\psi \sin\theta \sin\phi \\ z &= R \sin\psi \cos\theta \\ w &= R \cos\psi \ , \end{aligned} \tag{5.12}$$

where

$$0 \le \psi \le \pi \ , \quad 0 \le \theta \le \pi \ , \quad 0 \le \phi \le 2\pi \ , \tag{5.13}$$

and $\phi = 0$ is identified with $\phi = 2\pi$.

Since the coordinate system is to describe the surface, some point on the surface has to be chosen to be the origin of the coordinate system. For the two-dimensional spherical surface of the last section, we can consider the north pole to be the center, and then θ is the radial coordinate that measures the distance from the center. Here we are choosing the center of our coordinate system to be the positive w-axis, which we will also describe as the "north pole". The coordinates of the north pole in the four-dimensional embedding space are $(x=0,\ y=0,\ z=0,\ w=R)$. In the polar coordinate system the north pole is described by $\psi = 0$, and the distance from the north pole is given by $R\psi$. Thus ψ plays the role of the radial coordinate in this system.

To derive the metric, one could proceed purely algebraically along the lines of Eq. (5.9) above, or one could use the geometric arguments which were used to motivate Eq. (5.8). For the geometric approach, one notes that a variation from ψ to $\psi + \mathrm{d}\psi$ results in a displacement by a distance $R\,\mathrm{d}\psi$. A variation in θ or ϕ results in a displacement contained entirely within the x-y-z three-space; $\mathrm{d}s^2$ is given by Eq. (5.8) times an overall factor of $\sin^2\psi$ due to the fact that the radius in the x-y-z space is given by $r\sin\psi$. Assuming that these two displacements are orthogonal to each other, the metric can be written as

$$\mathrm{d}s^2 = R^2\left[\mathrm{d}\psi^2 + \sin^2\psi\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right] \ . \tag{5.14}$$

To complete the justification of Eq. (5.14), we should verify that the infinitesimal displacement of the point when ψ is varied is orthogonal to the displacement caused by infinitesimal variation of θ or ϕ. To see this, let us use vector notation $\vec{r} \equiv (x, y, z, w)$ to describe the four-dimensional space. Then, as ψ is varied from ψ to $\psi + \mathrm{d}\psi$, the vector $\vec{r}$ varies from $\vec{r}$ to $\vec{r} + \mathrm{d}\vec{r}_\psi$, where we can see from Eq. (5.12) that

$$\begin{aligned}\mathrm{d}\vec{r}_\psi &= \left(\frac{\partial x}{\partial\psi}, \frac{\partial y}{\partial\psi}, \frac{\partial z}{\partial\psi}, \frac{\partial w}{\partial\psi}\right)\mathrm{d}\psi \\ &= R\cos\psi(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta, 0)\,\mathrm{d}\psi - R\sin\psi(0,0,0,1)\,\mathrm{d}\psi \ .\end{aligned} \tag{5.15}$$

Note that the components in the x-y-z subspace are proportional to $(x, y, z) = R\sin\psi(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$, so within this subspace the vector points radially outward from the origin. Similarly, as θ is varied from θ to $\theta + \mathrm{d}\theta$, $\vec{r}$ varies from $\vec{r}$ to $\vec{r} + \mathrm{d}\vec{r}_\theta$, where

$$\begin{aligned}\mathrm{d}\vec{r}_\theta &= \left(\frac{\partial x}{\partial\theta}, \frac{\partial y}{\partial\theta}, \frac{\partial z}{\partial\theta}, \frac{\partial w}{\partial\theta}\right)\mathrm{d}\theta \\ &= R\sin\psi(\cos\theta\cos\phi, \cos\theta\sin\phi, -\sin\theta, 0) \ .\end{aligned} \tag{5.16}$$

This time there is no w-component, and we know that varying θ does not change $x^2+y^2+z^2$, and therefore the components within the x-y-z subspace make a tangential vector. Since a tangential vector is orthogonal to a radial vector, it follows that $\mathrm{d}\vec{r}_\psi \cdot \mathrm{d}\vec{r}_\theta = 0$, which is what we wanted to prove. The geometrical argument is easily verified by straightforward calculation:

$$\begin{aligned}\mathrm{d}\vec{r}_\psi \cdot \mathrm{d}\vec{r}_\theta = R^2 \sin\psi\cos\psi[&\sin\theta\cos\theta\cos^2\psi \\ &+ \sin\theta\cos\theta\sin^2\phi - \sin\theta\cos\theta + 0] = 0\ .\end{aligned} \tag{5.17}$$

A similar argument guarantees that $\mathrm{d}\vec{r}_\psi$ is also orthogonal to $\mathrm{d}\vec{r}_\phi$, so the justification of Eq. (5.14) is complete.

Remember that the coordinate system that one uses to describe a curved space is totally arbitrary. Another choice that is frequently used to describe this space is to replace ψ by

$$u \equiv \sin\psi\ . \tag{5.18}$$

Note that u is double-valued: as ψ varies over its range from 0 to π, u varies from 0 to 1 and then decreases back to 0. The new metric can then be found by noting that

$$\mathrm{d}u = \cos\psi\,\mathrm{d}\psi = \sqrt{1-u^2}\,\mathrm{d}\psi\ , \tag{5.19a}$$

and so

$$\mathrm{d}\psi^2 = \frac{\mathrm{d}u^2}{1-u^2}\ , \tag{5.19b}$$

and then

$$\mathrm{d}s^2 = R^2\left\{\frac{\mathrm{d}u^2}{1-u^2} + u^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\}\ . \tag{5.20}$$

In these coordinates it is particularly easy to see that in a small region about the origin, i.e., for $|u| \ll 1$, the u^2 in the denominator can be ignored, and the metric becomes the metric for Euclidean space in spherical polar coordinates. This is just an example of the general principle introduced by Gauss: as long as $\mathrm{d}s^2$ is expressed as a quadratic function of the coordinate differentials, then in any infinitesimal region it is possible to find coordinates for which the metric is Euclidean.

The geometry of this space will be pursued further in the next problem set.

IMPLICATIONS OF GENERAL RELATIVITY:

Eqs. (5.14) or (5.20) describe a curved three-dimensional space which is finite but without boundary. The length scale of this space is described by the parameter R, which can have any value. Since R corresponds to the radius of the sphere as embedded in the four-dimensional space, we will refer to R as the radius of curvature of the space.

Since general relativity describes gravity as a distortion of the spacetime metric, however, one might expect that the dynamics of general relativity would determine the curvature of the space, and hence determine the quantity R. The calculations are beyond these lectures, but the result is simple. General relativity requires that the geometry of the universe be non-Euclidean, except for the special case in which the parameter k defined in Lecture Notes 3 is zero. This is why the $k = 0$ model is called flat. When $k > 0$, which we have been calling a closed universe, general relativity requires that the geometry be a closed three-dimensional space, as described by the metric of Eqs. (5.14) or (5.20). Thus, if gravity is strong enough to cause the universe to recollapse, then it is also strong enough to curve the universe back on itself to create a universe that is finite but unbounded.*

Using Newtonian arguments, we have already calculated how the size of the model universe changes with time, proportional to the scale factor $a(t)$. The Friedmann equations that we obtained are identical to the predictions of general relativity, so the size of the universe will be proportional to the scale factor $a(t)$ that we already calculated. For the closed universe geometry, however, the size of the universe is proportional to the radius of curvature R, so consistency requires that R must be proportional to $a(t)$. Furthermore, we recall that the value of $a(t)$ depends on the size of the "notch." The radius of curvature R, however, is a physical length that must be measured in physical distance units, such as meters. Thus, dimensional consistency requires that $R(t)$ to be proportional to $a(t)/\sqrt{k}$, which also has the units of physical length. The constant of proportionality is fixed by the details of general relativity, but the answer is that the constant of proportionality is 1:

$$R^2(t) = \frac{a^2(t)}{k} \ . \tag{5.21}$$

Although the quantity $a^2(t)/k$ has been described in the context of a purely Newtonian calculation, the speed of light was inserted into the definition of k, which was given by Eq. (3.30) as

$$k = -\frac{2E}{c^2} \ , \quad \text{where} \ \ E = \frac{1}{2}\dot{a}^2 - \frac{4\pi}{3}\frac{G\rho_i}{a} \ .$$

---

* Warning: the simple correspondence between the closure of the universe in time and the closure of the universe in space holds for matter-dominated universes, and even for universes containing arbitrary mixes of matter and radiation. However, when we explore the consequences of a nonzero cosmological constant in Lecture Notes 7, we will find that the relation no longer holds. Universes which are spatially closed might nonetheless expand forever, and universes which are spatially open might nonetheless recollapse.

Thus Eq. (5.21) can be written as

$$R^2(t) = \frac{a^2(t)c^2}{2E} \ ,$$

which shows that curvature is explicitly a relativistic effect. In the nonrelativistic limit where c becomes infinitely large compared to all other velocities, $R(t)$ will approach infinity. Thus in the nonrelativistic limit the radius of curvature of the universe approaches infinity, so the space becomes closer and closer to Euclidean. (Note that the surface of a sphere of infinite radius is actually a plane.)

One can then rewrite the equations of evolution in terms of $R(t)$. Using

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2} \tag{5.22}$$

from Eqs. (3.25) and (3.31), one has

$$\boxed{H^2 = \left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi}{3}G\rho - \frac{c^2}{R^2} \ .} \tag{5.23}$$

To express the value of $R(t)$ in terms of observables, one can replace ρ by $\Omega\rho_c$, where ρ_c is given by $3H^2/(8\pi G)$ as in Eq. (3.33). One then has

$$\boxed{R = \frac{cH^{-1}}{\sqrt{\Omega - 1}} \ ,} \tag{5.24}$$

which is the same as Eq. (4.32). Note that as Ω becomes closer to one (approaching from above), $R(t)$ becomes larger and larger, so the space becomes closer and closer to Euclidean. In addition, Eq. (5.24) shows explicitly that $R(t)$ is proportional to c, as we discussed in the previous paragraph. Thus, if the speed of light is taken to be infinitely larger than all other velocities, then again the space becomes Euclidean. Curvature is therefore a relativistic effect.

THE ROBERTSON-WALKER FORM OF THE METRIC:

When Eq. (5.21) is substituted into Eq. (5.20), the resulting metric is given by

$$\mathrm{d}s^2 = \frac{a^2(t)}{k}\left\{\frac{\mathrm{d}u^2}{1-u^2} + u^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\} \ , \tag{5.25}$$

which is a little more complicated than necessary. It is convenient to replace the radial coordinate u (where $u \equiv \sin\psi$) with a new radial coordinate r defined by

$$r \equiv \frac{u}{\sqrt{k}} \equiv \frac{\sin\psi}{\sqrt{k}} \ . \tag{5.26}$$

Then $\mathrm{d}r = k^{-1/2}\mathrm{d}u$, and the metric can be rewritten as

$$\boxed{\mathrm{d}s^2 = a^2(t)\left\{\frac{\mathrm{d}r^2}{1-kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\} \ .} \tag{5.27}$$

This is the standard form, called the Robertson-Walker metric. Since the coordinate r is proportional to u, and u is double-valued, so is r. That is, $r = 0$ at the center of the coordinate system, which is identified with the north pole of the sphere that describes the closed universe. As r grows the point described by (r, θ, ϕ) moves away from the north pole, and r reaches its maximum value of $1/\sqrt{k}$ when the point reaches the equator of the sphere. If one continues to move the point in the same direction, then r decreases back to zero as the point moves from the equator to the south pole, where r again is zero.

THE OPEN UNIVERSE:

We have seen that when $k > 0$ the universe is spatially closed (finite volume), and that it approaches an infinite volume Euclidean space as $k \to 0$ (i.e., in this limit the radius of the sphere approaches infinity). What happens if $k < 0$?

As you have probably learned from your experience in physics, in many cases the same equations will hold whether the variables that occur in those equations are positive or negative. Thus, we might expect that the formulas derived above would be valid for $k < 0$, and this is indeed the case. However, there is one complication which should be pointed out. Above we made the change of variables given by Eq. (5.26), involving the quantity $\sqrt{k}$. This quantity would be imaginary if k were negative, and thus it would not be possible for both u and r to be real. One can see from Eq. (5.25) that the metric in terms of u is pathological when k is negative, since $\mathrm{d}s^2$ is not positive definite. For $u < 1$ it is in fact negative definite, and for $u > 1$ the sign is indeterminate, since the angular pieces contribute negatively while the radial piece contributes positively. Thus, it seems clear that the u variable must be discarded when $k < 0$. On the other hand, the metric in the form of Eq. (5.27) remains perfectly well behaved for negative values of k. To minimize the possible confusion of dealing with negative quantities, we can define $\kappa = -k$, and rewrite the Robertson-Walker metric (5.27) for open universes as

$$\boxed{\begin{gathered}\mathrm{d}s^2 = a^2(t)\left\{\frac{\mathrm{d}r^2}{1+\kappa r^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\} \ . \\ \text{(Open universe, } \kappa > 0\text{)}\end{gathered}} \tag{5.28}$$

While it is reasonable to assume that Eq. (5.28) is correct, our derivation was certainly far from rigorous. I will not try to give a rigorous derivation, but I will try at least to sketch how a rigorous derivation could be constructed. If we wanted to be more rigorous, we would begin by summarizing the goal: to construct a metric describing a homogeneous and isotropic space. While the θ and ϕ angular coordinates are not very obviously isotropic, we are sufficiently familiar with this construction to be convinced that the angular dependence of the metric above is isotropic. Although the coordinate system makes the north pole ($\theta = 0$) look like a special direction, we know that the coordinates could be redefined to put the north pole of the coordinate system at any angle. The homogeneity of the Robertson-Walker metric is similar, but less familiar to us. For the closed Robertson-Walker metric we know that the space is homogeneous, because we derived the metric by starting with the manifestly homogeneous 3-dimensional sphere embedded in four Euclidean dimensions. But the Robertson-Walker coordinates make the origin ($r = 0$) look special, just as the angular coordinates make the north pole look special. As in the case of the angular coordinates, we know that the origin of the closed Robertson-Walker coordinate system is not really special, and that we could redefine our coordinate system so that the origin can be put at any location.

To show that the open Robertson-Walker metric in Eq. (5.28) is homogeneous, we would start by studying the homogeneity of the closed universe metric in detail, turning the verbal statements in the previous paragraph into an explicit set of coordinate transformations that show how to move the origin to an arbitrary point. The details become rather complicated, as indeed they would if we tried to explicitly show how to construct a coordinate transformation to move the north pole of the (θ, ϕ) angular coordinates. Nonetheless, once the equations are written, it would become clear that they are just a set of algebraic relations: if they hold for all positive k, they will necessarily hold for negative k as well. Thus the same algebra that shows the closed Robertson-Walker universe to be homogeneous also shows that the open metric is homogeneous.

We will not try to show it, but it can be shown that any three-dimensional homogeneous and isotropic space can be described by the Robertson-Walker metric, Eq. (5.27), where k can be positive, negative, or zero. Other coordinate systems are of course possible, but geometrically different spaces are not.

Note that the sign of k affects the question of whether the space is finite or infinite. For $k > 0$, Eq. (5.27) implies that something peculiar happens when $kr^2 = 1$, at which point the metric is singular. Since r is related to the original ψ coordinate by $r = \sin(\psi)/\sqrt{k}$, one sees that this value of the radius variable corresponds to $\psi = \pi/2$, and hence the equator of the original sphere embedded in four dimensions. There is nothing singular about the space, but the metric becomes singular because the coordinate r behaves peculiarly, reaching a maximum value. Beyond the equator, r must get smaller and then approach zero at the "south pole" $(x = 0, y = 0, z = 0, w = -R)$. Thus, the space is finite. However, if $k < 0$ then the metric is given by Eq. (5.28), which remains perfectly well-defined for all values of r, and thus the range of the r-coordinate is infinite. This does not by itself prove that the space is infinite, since the value of a coordinate is not directly measurable. However, one can calculate the physical distance from the origin to a point with radial coordinate r by integrating the metric of Eq. (5.28) along a radial path (with $\mathrm{d}\theta = \mathrm{d}\phi = 0$):

$$\ell_{\text{phys}}(r) = a(t) \int_0^r \frac{\mathrm{d}r'}{\sqrt{1 + \kappa r'^2}} = \frac{\sinh^{-1} \sqrt{\kappa}\, r}{\sqrt{\kappa}} \ , \tag{5.29}$$

where the integration can be carried out by substituting $r' = \sinh(\psi)/\sqrt{\kappa}$. Since the inverse sinh function can become arbitrarily large, the space is infinite.

The G-B-L geometry discussed in the introduction is simply the two-dimensional version of the space of an open universe at some arbitrary fixed time. The realization by Klein described in Eqs. (5.1) and (5.2) represents a somewhat peculiar choice of coordinate system.

THE GENERALIZATION FROM SPACE TO SPACETIME

Eq. (5.27) actually shows only a **spatial** metric, while I said earlier that general relativity describes the gravitational field in terms of a **spacetime** metric. To put the spacetime metric into context, we recall that in special relativity it is possible to define a Lorentz-invariant separation between two events. Specifically, if the coordinates of an event A are (x_A, y_A, z_A, t_A), and the coordinates of an event B are (x_B, y_B, z_B, t_B), then the Lorentz-invariant separation between A and B is defined by

$$s^2 \equiv (x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2 - c^2\,(t_A - t_B)^2 \ . \tag{5.30}$$

By saying that this expression is Lorentz-invariant, we mean that it has the same value in all inertial references frames, even though the individual terms may very well have different values.

While the value of s^2 is the same in all inertial frames, the intuitive meaning of s^2 is easiest to see by considering its value in particular frames. If $s^2 > 0$, then the separation between the events is called *spacelike.* In that case it is always possible to find an inertial reference frame in which the two events are simultaneous, and in that frame s is equal to the spatial distance between the two events. Equivalently, we can say that it is always possible to find an inertial observer to whom the two events appear simultaneous. s is then equal to the distance between these events, as measured by a ruler at rest with respect to this observer. s can be called the *proper distance* between the events. If $s^2 < 0$ then the separation is called *timelike,* and in that case it is always possible to find an inertial observer to whom it appears that the two events occur at the same position. If she defines

$$s^2 = -c^2\tau^2 \ , \tag{5.31}$$

then τ is the time separation between the events when measured on her clock. τ is often called the *proper time* between the two events. Note that if the two events happen to the same object, such as two flashes of the same strobe light, and the object is moving at constant velocity, then the proper time between the flashes is just the time as measured by a clock at rest with respect to the strobe light. If $\mathrm{d}s^2 = 0$, then the separation between the two events is called *lightlike,* and in that case a light pulse leaving the earlier event will arrive at the location of the latter event just as it occurs.

If you are not familiar with the Lorentz-invariant separation, you may want to look at the Appendix at the end of this set of Lecture Notes. There I start with the three basic effects of special relativity, as described in Lecture Notes 1, and show how to construct the Lorentz transformation. The Lorentz transformation is the set of equations that describe how to relate the coordinates of an event in two different inertial coordinate frames, where one frame is moving relative to the other. Using the Lorentz transformation, the Appendix goes on to show that the expression defined by Eq. (5.30) is truly Lorentz-invariant. (For purposes of this course, however, the Appendix can be considered outside the course requirements. It is okay for you to just accept the result that s^2 is Lorentz-invariant.)

The spacetime metric of general relativity is the curved-spacetime generalization of the Lorentz-invariant separation of special relativity. Following the ideas of Gauss discussed near the beginning of these lecture notes, we will restrict our attention to describing the separation between two infinitesimally separated spacetime points (x, y, z, t) and $(x + \mathrm{d}x, y + \mathrm{d}y, z + \mathrm{d}z, t + \mathrm{d}t)$. For special relativity the metric of Eq. (5.30) reduces in the infinitesimal case to

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 - c^2\,\mathrm{d}t^2 \ , \tag{5.32}$$

which is known as the Minkowski metric. Continuing with Gauss's approach, we insist — even when we describe arbitrary curved spacetimes — that $\mathrm{d}s^2$ be expressed as a quadratic expression in the coordinate differentials. This implies (although we will not show it) that for any spacetime point P it is always possible to choose a coordinate system (x', y', z', t') so that the metric reduces to the Minkowski metric in an infinitesimal region around that point. If the spacetime is curved the metric will not have the Minkowski form outside this infinitesimal region, however, so the metric will be called *locally Minkowskian* at the point P.

In curved spacetimes there is no coordinate system in which the metric has the Minkowski form everywhere. Thus, to infer the separation between two points one must know not only the values of the coordinates, but also the metric. The coordinates are then not themselves direct measurements of distance, but instead are just an arbitrary way of labeling points. Since one needs to introduce a metric for any coordinate system, there is nothing that forces us to use any particular coordinate system or set of coordinate systems. This is different from special relativity, where the metric (5.32) is valid only for a special class of coordinate systems, called inertial coordinate systems, which are related to each other by a special class of transformations, called Lorentz transformations. If I were to replace the coordinate x by $x' \equiv \sinh x$, then the metric would no longer look like Eq. (5.32). The coordinate transformation $x' \equiv \sinh x$ is therefore not allowed in the standard formulation of special relativity (although one could use the formalism of general relativity for a special relativity problem if one chose to.) In general relativity, on the other hand, there is usually no coordinate system in which the metric is particularly simple, so the formalism is designed to allow any choice of coordinates, and hence any kind of coordinate transformation. In general relativity, therefore, $x' = \sinh x$ is a perfectly acceptable coordinate transformation. As long as the coordinates allow a unique way to label each point in spacetime, they are acceptable. If I change coordinate systems, I can always change the metric so that the value of $\mathrm{d}s^2$ between any two points remains the same. For this reason $\mathrm{d}s^2$ is said to be *coordinate-invariant.*

When we introduced the two-dimensional spatial metric in Eq. (5.3), we assumed that $\mathrm{d}s^2$ represented the distance between the two points, where the meaning of "distance" was no different from what it would mean in Euclidean geometry — it is what one would measure with a ruler. Here we are trying to generalize this method, so we want to define $\mathrm{d}s^2$ to have the same meaning it would have in special relativity. In special relativity we were able to define $\mathrm{d}s^2$ in terms of the observations made by inertial observers, which means observers for whom the law of inertia is valid, which in turn means observers to whom no net force is applied. In general relativity, forces other than gravity are treated in essentially the same way as in special relativity, so there is no problem defining what it means for the net **nongravitational** force on an observer to vanish. But gravity is trickier. Consider, for example the homogeneously expanding universe that we discussed in Lecture Notes 2, 3, and 4. If I am moving with the expansion of the universe (i.e., if I am at rest with respect to the comoving coordinate system), then I can view myself as being at rest. If I look at the distant galaxies around me, however, they will appear to be slowing in their outward motion, and hence accelerating towards me, under the influence of gravity. But an observer on one of those galaxies would consider himself to be at rest, and I would appear to be accelerating. According to general relativity both points of view are equally valid, so the concept of gravitational acceleration becomes relative.

Another simple and famous example that illustrates the relative nature of gravitational forces is the elevator (thought) experiment. Suppose a man, holding a bag of groceries, is standing in an elevator. Now suppose that the elevator cables are cut, and the elevator free falls downward without friction or air resistance. The man will then accelerate downward with the same acceleration as the elevator, and he will feel no force between his feet and the elevator floor. If he lets go of the bag of groceries, the bag would not move relative to him, but would appear to float in front of him. In the frame of the Earth, all the objects (the elevator, the man, and the groceries) are accelerating downward under the force of gravity. But in the frame of the elevator, everything appears weightless. (Everything is is weightless until the big crunch occurs in the building's basement — but remember, this in only a thought experiment. No living creatures were harmed in the writing of this paragraph.)

We are accustomed to thinking of the frame of the Earth as being the correct "physical" description, because the frame of the Earth is nearly inertial over a large region of space and time. In the context of general relativity, however, both frames are equally correct. Thus, the presence or absence of gravity is determined by which frame of reference we are using. This idea in fact is one of the foundational concepts of general relativity, known as the *equivalence principle.* The physics of the accelerating frame of the elevator, with no gravity, is equivalent to the physics in the rest frame of the Earth, with its gravitational field. The equivalence principle says that it is always possible, in a sufficiently small region, to find a frame of reference in which the force of gravity is absent.

The bottom line here is that if we are trying to generalize the notion of an inertial observer in special relativity, we cannot insist that the gravitational force on the observer vanishes, because this condition will appear to hold in some coordinate systems but not others. So, instead we insist only that the net **nongravitational** force on the observer vanish, and we say that such an observer is *free-falling.* Note that the man in the falling elevator is free-falling, while a man standing in an elevator that is at rest with respect to the Earth is not. In the latter case the floor is pushing upward on the man's feet, so the net nongravitational force is nonzero.

With the replacement of inertial observers by free-falling observers, the meaning of $\mathrm{d}s^2$ in general relativity is the same as what we had in special relativity. If the value of $\mathrm{d}s^2$ calculated between two events is positive, then there is always a free-falling observer to whom the events appear simultaneous. In this case, the proper distance $\mathrm{d}s$ between the events is the distance between them, as measured by a ruler at rest relative to this free-falling observer. If $\mathrm{d}s^2 < 0$, then there is always a free-falling observer for whom the events appear to happen at the same location. One then defines

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 \ , \tag{5.33}$$

as in Eq. (5.31), where $\mathrm{d}\tau$ is again called the proper time interval between the events. It is the time interval between the two events that would be measured by a clock carried by the free-falling observer mentioned above. If $\mathrm{d}s^2 = 0$, then the two events can be connected by a light pulse, which leaves the first event and arrives at the second.*

---

* The concept of a free-falling observer is intimately linked to the concept of a locally Minkowskian coordinate system, so the meaning of $\mathrm{d}s^2$ could also have been explained in terms of these coordinate systems. The free-falling observers are those that are at rest or moving at a constant velocity relative to a coordinate system that is locally Minkowskian at the location of the observer.

INCLUSION OF TIME IN THE ROBERTSON-WALKER METRIC

What happens when we add time to the Robertson-Walker metric of Eq. (5.27)? In general the answer can depend on how we choose to define our time variable, but we will hold with the choice called *cosmic time*, which we discussed in Lecture Notes 2 (in a section called "The Synchronization of Clocks"). We concluded there that it is possible to define a cosmic time variable t which can be measured locally. That is, each observer who is at rest with respect to the matter in her vicinity can measure t on her own wristwatch. The wristwatches throughout the universe can be synchronized, once and for all, by some choice of a cosmic event. For example, we can all agree to set our wristwatches to read 12 billion years when the temperature of the cosmic microwave background radiation reaches 3.0 K, or when the Hubble parameter reaches 85 km-sec^{-1}-Mpc^{-1}. Once the watches are synchronized, we argued that the homogeneity of the universe guarantees that they will stay synchronized: all watches will read the same time when the cosmic background radiation temperature reaches 2.0 K, or when the Hubble parameter reaches 75 km-sec^{-1}-Mpc^{-1}. In practice we usually define the synchronization of cosmic time so that $t = 0$ corresponds to our best estimate of when $a(t)$ was equal to zero, and the Hubble parameter and temperature were infinite. (More precisely, we choose $t = 0$ to correspond to the time when $a(t)$, as extrapolated in our mathematical model, was equal to zero. As discussed in *The Big Bang Singularity* section of Lecture Notes 4, there is no reason for us to have confidence in this extrapolation.)

I think it will be most straightforward for me to write the answer first, and then explain why it could not have been anything different. If the time variable t is taken to be cosmic time, and the metric is to be homogeneous and isotropic, then it can always be written as

$$\boxed{\mathrm{d}s^2 = -c^2\,\mathrm{d}t^2 + a^2(t)\left\{\frac{\mathrm{d}r^2}{1-kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\}\ .} \tag{5.34}$$

So, why does this have to be the answer? Consider first the case in which the separation $\mathrm{d}t = 0$ (i.e., when the two events whose separation we are calculating have the same time coordinate). In that case Eq. (5.34) reduces to our previous expression, Eq. (5.27). Since we have already stated (albeit without proof) that Eq. (5.27) describes the most general possible three-dimensional space that is homogeneous and isotropic, the answer for the $\mathrm{d}t = 0$ case is settled. We could of course choose other coordinates that would make the spatial part of Eq. (5.34) look different, but Eq. (5.34) as written describes the most general possible geometry.

Now consider the interval defined by $\mathrm{d}t \neq 0$, but $\mathrm{d}r = \mathrm{d}\theta = \mathrm{d}\phi = 0$. This represents the motion of a comoving observer for an increment of coordinate time $\mathrm{d}t$. There are no nongravitational forces acting on the comoving observer, so she is also a free-falling observer. This is a timelike separation, so we use the definition $\mathrm{d}s^2 = -c^2\,\mathrm{d}\tau^2$ from Eq. (5.33), and we deduce that $\mathrm{d}t = \mathrm{d}\tau$, where $\mathrm{d}\tau$ is the time measured on the comoving observer's wristwatch. But an interval of cosmic time is defined as the interval measured on the wristwatches of comoving observers, so the metric of Eq. (5.34) implies that t is precisely the time variable that we have called cosmic time. Note that if the coefficient of the $\mathrm{d}t^2$ term in the metric were anything other than $-c^2$, we would have found that the time coordinate interval $\mathrm{d}t$ is proportional to wristwatch time, but not equal to it.

We have now verified that the terms that are present in Eq. (5.34) must have the forms that they have. But what about the possibility of adding other terms. Since the metric is required to be a quadratic function of the coordinate differentials, the only possible new terms that could be added are terms proportional to $\mathrm{d}t\,\mathrm{d}r$, $\mathrm{d}t\,\mathrm{d}\theta$, or $\mathrm{d}t\,\mathrm{d}\phi$. (Recall that terms like $\mathrm{d}r\,\mathrm{d}\theta$ would contribute even when the time is fixed, $\mathrm{d}t = 0$, so such terms have already been ruled out by the statement that Eq. (5.27) is the most general possible homogeneous and isotropic space.) Let us consider first the possibility of adding a term $\mathrm{d}r\,\mathrm{d}t$ to the metric. The claim is that such a term would violate our assumption of isotropy, because it would create a distinction between the direction of increasing and decreasing r. To see this, consider two observers, Tweedledee and Tweedledum, who both start at $r = r_0$ at time $t = t_0$. Tweedledee is moving outward and Tweedledum is moving inward, both with coordinate speed $\mathrm{d}r/\mathrm{d}t = v$ (and with fixed values of θ and ϕ). At $t = t_0 + \mathrm{d}t$, Tweedledee will be located at $r = r_0 + v\,\mathrm{d}t$, while Tweedledum will be located at $r = r_0 - v\,\mathrm{d}t$. Thus the displacement vector of Tweedledee has $\mathrm{d}r > 0$, while that of Tweedledum has $\mathrm{d}r < 0$, and both have the same $\mathrm{d}t$. The hypothetical new term will therefore contribute to $\mathrm{d}s^2$ with opposite signs for the two cases, so the values of $\mathrm{d}s^2$ will be different for Tweedledee and Tweedledum. Since $\mathrm{d}s^2 = -c^2\,\mathrm{d}\tau^2$, and $\mathrm{d}\tau$ is the wristwatch time that each will measure, we conclude that each will have a different wristwatch time at the end of this interval. When they each compare with the comoving observers whose wristwatches read cosmic time, $t = t_0 + \mathrm{d}t$, the two will see different discrepancies. This means that there is a Tweedledee/Tweedledum asymmetry, but the only difference in the setup was their direction of travel. Thus, the addition of such a term would be a violation of isotropy. An identical argument can be made for $\mathrm{d}t\,\mathrm{d}\theta$ or $\mathrm{d}t\,\mathrm{d}\phi$ terms, so we conclude that Eq. (5.34) is necessarily the right answer.

EQUATIONS FOR A GEODESIC

As was stated earlier, in general relativity a freely falling particle is assumed to travel on a geodesic of the curved spacetime. Stated more precisely, the equations of motion in general relativity are derived from the assumption that the path length from the initial point to the final point should have a vanishing derivative with respect to any variation of the path that does not vary the endpoints. If the meaning of this statement is not clear to you at this point, then don't worry yet — it will hopefully become clear once we define some notation.

We will start by deriving the equation for a geodesic in a two-dimensional space with a positive-definite metric (i.e., with all lengths positive). The metric will be assumed to have the general form specified by Gauss, and given earlier as Eq. (5.3):

$$\mathrm{d}s^2 = g_{xx}\mathrm{d}x^2 + g_{xy}\mathrm{d}x\,\mathrm{d}y + g_{yx}\mathrm{d}y\,\mathrm{d}x + g_{yy}\mathrm{d}y^2 \ , \tag{5.3}$$

where g_{xx}, g_{xy}, g_{yx}, and g_{yy} are functions of position (x, y) and are together called the metric of the space. As explained earlier, we take $g_{yx} \equiv g_{xy}$.

The first step will be to simplify the notation, since Eq. (5.3) requires a lot of writing. To start, rename the coordinate x as x^1, and rename y as x^2. Then the two coordinates together can be described as x^i, where i is understood to take on the values 1 and 2. Eq. (5.3) can then be rewritten as

$$\mathrm{d}s^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} g_{ij}(x^k)\,\mathrm{d}x^i\,\mathrm{d}x^j \ , \tag{5.35}$$

where I write the metric as $g_{ij}(x^k)$ to indicate explicitly that it is a function of all of the coordinates x^k. One further simplification is known as the Einstein summation convention. This is no doubt Einstein's most important contribution to ecology, saving barrels of ink and tons of paper each year. The convention stipulates that whenever an index is repeated, it is automatically summed over the standard range (which in this case is from 1 to 2). Using this convention, Eq. (5.35) can be written compactly as

$$\mathrm{d}s^2 = g_{ij}(x^k)\,\mathrm{d}x^i\,\mathrm{d}x^j \ . \tag{5.36}$$

(In using this notation, it is important that the context makes it clear that the superscript i in x^i is to be interpreted as an index, and not a power. You might wonder why people tolerate this confusion, when it could be avoided by writing all indices as subscripts. The reason is that curved space geometers find it useful to use both superscripts and subscripts to denote indices. Quantities with upper indices (superscripts) are called contravariant, and quantities with lower indices (subscripts) are called covariant. These indices can always be arranged so that each summation over a repeated index involves one upper and one lower index, as has been done in Eq. (5.36). To understand fully the meaning of upper and lower indices, one must study how the equations of non-Euclidean geometry are transformed by a redefinition of the coordinate system. We will skip this topic, but I point out that the formalism is constructed so that the rules of transformation are indicated by whether the indices are upper or lower. Furthermore, the transformation rules guarantee that any sum over a repeated index, with one upper and one lower, is invariant under a change of coordinates.)

Now we can state the geodesic problem: given two points x^i_A and x^i_B, what equation determines the geodesic, or shortest path, between the two points? (In this case it will be the shortest path.)

An arbitrary path can be described by a function $x^i(\lambda)$, where λ is a parameter which we take to run between 0 and some final value λ_f. Thus, the statement that the path runs from x^i_A to x^i_B translates into the equations

$$x^i(0) = x^i_A , \qquad x^i(\lambda_f) = x^i_B \ . \tag{5.37}$$

Now focus attention on an infinitesimal segment of the curve, from λ to $\lambda + \mathrm{d}\lambda$. The change in the values of the two coordinates over this segment is given by

$$\mathrm{d}x^i = \frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\mathrm{d}\lambda \ . \tag{5.38}$$

Since $\mathrm{d}\lambda$ is infinitesimal, one need not consider terms in Eq. (5.38) that are higher order in $\mathrm{d}\lambda$. Combining this equation with Eq. (5.36), one has

$$\mathrm{d}s^2 = g_{ij}\big(x^k(\lambda)\big)\,\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}\,\mathrm{d}\lambda^2 \ ,$$

and then

$$\mathrm{d}s = \sqrt{g_{ij}\big(x^k(\lambda)\big)\,\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}}\,\mathrm{d}\lambda \ . \tag{5.39}$$

The total length of the path is then

$$S[x^i(\lambda)] = \int_0^{\lambda_f} \sqrt{g_{ij}\big(x^k(\lambda)\big)\,\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}}\,\mathrm{d}\lambda \ . \tag{5.40}$$

The path length $S[x^i(\lambda)]$ is actually a function of the function $x^i(\lambda)$. A function of a function is usually called a ***functional***, and the argument of the functional is usually enclosed in square brackets.

Next we consider how the path length will vary if the path is changed infinitesimally. To formulate this precisely, we write the equation for a nearby path, with the same endpoints, as

$$\tilde{x}^i(\lambda) = x^i(\lambda) + \alpha w^i(\lambda) \ , \tag{5.41a}$$

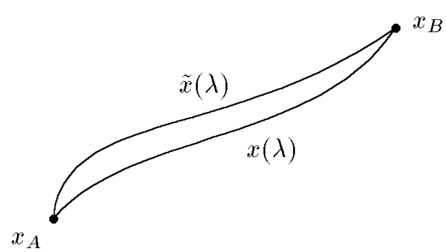


Figure 5.11: A path $x(\lambda)$ and a small variation of it, $\tilde{x}(\lambda)$.

where α is a number (which we will take to be small), and the path variation function $w^i(\lambda)$ is required to satisfy

$$w^i(0) = 0 \ , \qquad w^i(\lambda_f) = 0 \ , \tag{5.41b}$$

so that the new path $\tilde{x}^i(\lambda)$ has the same endpoints as original path $x^i(\lambda)$. The rule for a geodesic is that no matter how the path is varied, the original length is a minimum. This implies that if $w^i(\lambda)$ is held fixed, for any value that satisfies Eq. (5.41b), the path length of $\tilde{x}^i(\lambda)$ should have a minimum at $\alpha = 0$. Thus,

$$\left.\frac{\mathrm{d}\, S\left[\tilde{x}^i(\lambda)\right]}{\mathrm{d}\alpha}\right|_{\alpha=0} = 0 \qquad \text{for all } w^i(\lambda) \quad . \tag{5.42}$$

The problem now is simply to calculate the derivative in Eq. (5.42). To simplify the notation, we define

$$A(\lambda, \alpha) = g_{ij}\left(\tilde{x}^k(\lambda)\right) \frac{\mathrm{d}\tilde{x}^i}{\mathrm{d}\lambda} \frac{\mathrm{d}\tilde{x}^j}{\mathrm{d}\lambda} \ , \tag{5.43}$$

so we can write

$$S\left[\tilde{x}^i(\lambda)\right] = \int_0^{\lambda_f} \sqrt{A(\lambda, \alpha)}\, \mathrm{d}\lambda \ . \tag{5.44}$$

Note that the derivative can be taken inside the integral that defines $S[\tilde{x}^i(\lambda)]$, since the limits of integration do not depend on α. Using the chain rule of differentiation, we find

$$\left.\frac{\mathrm{d}}{\mathrm{d}\alpha} g_{ij}\left(\tilde{x}^k(\lambda)\right)\right|_{\alpha=0} = \left.\frac{\partial g_{ij}}{\partial x^k}\right|_{x^k = x^k(\lambda)} \left.\frac{\partial \tilde{x}^k}{\partial \alpha}\right|_{\alpha=0} = \frac{\partial g_{ij}}{\partial x^k}\left(x^i(\lambda)\right) w^k \ , \tag{5.45}$$

where the Einstein summation convention applies to the sum over k. Differentiating Eq. (5.44), one then finds

$$\begin{aligned} \left.\frac{\mathrm{d}S\left[\tilde{x}^i(\lambda)\right]}{\mathrm{d}\alpha}\right|_{\alpha=0} = \frac{1}{2}\int_0^{\lambda_f} \frac{1}{\sqrt{A(\lambda,0)}} \Bigg\{ & \frac{\partial g_{ij}}{\partial x^k} w^k \frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda} + \\ & + g_{ij} \frac{\mathrm{d}w^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda} + g_{ij}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}w^j}{\mathrm{d}\lambda} \Bigg\} \mathrm{d}\lambda \ , \end{aligned} \tag{5.46}$$

where the metric g_{ij} is to be evaluated at $x^k(\lambda)$.

The expression can be further simplified by recognizing that the summed indices are "dummy" indices, in the sense that their names can be changed without changing the value of the expression. (When one does this, of course, it is essential that the name be changed in the same way for each occurrence of the index.) Suppose then that the third term in curly brackets of the above equation is rewritten by substituting $i \to j$ and $j \to i$. It then becomes identical to the second term, except that the indices on g_{ij} are reversed. But g_{ij} is symmetric in the sense that $g_{ji} = g_{ij}$ (see the remarks following Eq. (5.3)), so the two terms are identical. Thus,

$$\left.\frac{\mathrm{d}S\left[\tilde{x}^i(\lambda)\right]}{\mathrm{d}\alpha}\right|_{\alpha=0} = \frac{1}{2}\int_0^{\lambda_f} \frac{1}{\sqrt{A(\lambda,0)}}\left\{\frac{\partial g_{ij}}{\partial x^k} w^k \frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda} + 2g_{ij}\frac{\mathrm{d}w^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}\right\}\mathrm{d}\lambda\ . \tag{5.47}$$

The next step is to simplify the dependence on $w^i(\lambda)$. The expression above depends explicitly on both the function $w^i(\lambda)$ and its derivative, but the dependence on the derivative can be removed by an integration by parts. Note that the term

$$\int_0^{\lambda_f}\left[\frac{1}{\sqrt{A}} g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}\right]\frac{\mathrm{d}w^i}{\mathrm{d}\lambda}\,\mathrm{d}\lambda$$

can be integrated using

$$\int u\,\mathrm{d}v = -\int v\,\mathrm{d}u + [uv]_{\lambda=0}^{\lambda=\lambda_f}\ ,$$

where

$$u = \frac{1}{\sqrt{A}} g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}\ , \qquad \mathrm{d}u = \frac{\mathrm{d}}{\mathrm{d}\lambda}\left[\frac{1}{\sqrt{A}} g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}\right]\mathrm{d}\lambda$$

$$\mathrm{d}v = \frac{\mathrm{d}w^i}{\mathrm{d}\lambda}\,\mathrm{d}\lambda\ , \qquad v = w^i\ .$$

The surface term $[uv]_{\lambda=0}^{\lambda=\lambda_f}$ then vanishes, since $w^i(0) = w^i(\lambda_f) = 0$. So,

$$\int_0^{\lambda_f}\left[\frac{1}{\sqrt{A}} g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}\right]\frac{\mathrm{d}w^i}{\mathrm{d}\lambda}\,\mathrm{d}\lambda = -\int_0^{\lambda_f}\frac{\mathrm{d}}{\mathrm{d}\lambda}\left[\frac{1}{\sqrt{A}} g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}\right] w^i\,\mathrm{d}\lambda\ . \tag{5.48}$$

Thus, Eq. (5.47) simplifies to

$$\left.\frac{\mathrm{d}S}{\mathrm{d}\alpha}\right|_{\alpha=0} = \frac{1}{2}\int_0^{\lambda_f}\left\{\frac{1}{\sqrt{A}}\frac{\partial g_{ij}}{\partial x^k}\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda} w^k - 2\frac{\mathrm{d}}{\mathrm{d}\lambda}\left[\frac{1}{\sqrt{A}} g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}\right] w^i\right\}\mathrm{d}\lambda\ .$$

If one also renames the indices in the first term by $i \to j, j \to k, k \to i$, one can write

$$\left.\frac{\mathrm{d}S}{\mathrm{d}\alpha}\right|_{\alpha=0} = \int_0^{\lambda_f}\left\{\frac{1}{2\sqrt{A}}\frac{\partial g_{jk}}{\partial x^i}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}\frac{\mathrm{d}x^k}{\mathrm{d}\lambda} - \frac{\mathrm{d}}{\mathrm{d}\lambda}\left[\frac{1}{\sqrt{A}} g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}\right]\right\} w^i(\lambda)\,\mathrm{d}\lambda\ . \tag{5.49}$$

The next step is to set the quantity in curly brackets in the expression above equal to zero. To justify this, one must of course realize that the vanishing of an integral does not in general require that the integrand is zero — that is, it is very easy to find nonzero functions that integrate to zero over some specified range. However, we need to require that the derivative above vanish not merely for some particular value of $w^i(\lambda)$, but rather that it vanish for ***all*** values of $w^i(\lambda)$ that are consistent with Eq. (5.41b). This stronger requirement implies that the integrand must vanish. Note that if the quantity in curly brackets did not vanish, one could choose $w^i(\lambda)$ to equal the quantity in curly brackets, so the integral in Eq. (5.49) becomes the integral of a perfect square. Since then the integrand is nonnegative, the integral can vanish only if the integrand is identically zero. (Technically, the integrand can still be nonzero on a set of measure zero, such as a discrete set of points, since the integral over such a set gives zero in any case. We will restrict ourselves, however, to continuous functions, and then such a quantity must vanish everywhere.) Thus,

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}\lambda}\left[\frac{1}{\sqrt{A}}g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}\right] = \frac{1}{2\sqrt{A}}\frac{\partial g_{jk}}{\partial x^i}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}\frac{\mathrm{d}x^k}{\mathrm{d}\lambda}\ .} \tag{5.50}$$

The above equation is actually quite complicated, since the quantity A defined by Eq. (5.43) is complicated. However, the equation also has more generality than we really need: as we derived it, it will be valid for any parameterization $x^i(\lambda)$ of the path. If we instead make a specific choice about how the path is to be parameterized, then the equation can be simplified. In particular, we can simplify the equation tremendously by choosing λ to be the path length, as measured along the curve. Recalling that

$$\mathrm{d}s = \sqrt{g_{ij}\left(x^k(\lambda)\right)\frac{\mathrm{d}x^i}{\mathrm{d}\lambda}\frac{\mathrm{d}x^j}{\mathrm{d}\lambda}}\,\mathrm{d}\lambda = \sqrt{A}\,\mathrm{d}\lambda\ ,$$

one sees that $\mathrm{d}\lambda = \mathrm{d}s$ requires

$$A = 1 \qquad (\text{for } \lambda = \text{path length}). \tag{5.51}$$

Then the geodesic equation becomes

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}s}\left[g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}s}\right] = \frac{1}{2}\frac{\partial g_{jk}}{\partial x^i}\frac{\mathrm{d}x^j}{\mathrm{d}s}\frac{\mathrm{d}x^k}{\mathrm{d}s}\ ,} \tag{5.52}$$

where I have replaced λ by s to indicate clearly that it is the physical path length.

Eq. (5.52) is in many cases the most convenient form of the geodesic equation, but it is nonetheless not the standard way that the geodesic equation is written in general relativity books. Instead, the standard form is to write an explicit equation for $\mathrm{d}^2x^i/\mathrm{d}s^2$. One begins by expanding the left-hand side of Eq. (5.52), using the chain rule:

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}s}\right] = g_{ij}\frac{\mathrm{d}^2x^j}{\mathrm{d}s^2} + \partial_k g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}s}\frac{\mathrm{d}x^k}{\mathrm{d}s} , \tag{5.53}$$

where I have used the standard abbreviation

$$\partial_k \equiv \frac{\partial}{\partial x^k} . \tag{5.54}$$

The geodesic equation then becomes

$$g_{ij}\frac{\mathrm{d}^2x^j}{\mathrm{d}s^2} = \frac{1}{2}\left(\partial_i g_{jk} - 2\partial_k g_{ij}\right)\frac{\mathrm{d}x^j}{\mathrm{d}s}\frac{\mathrm{d}x^k}{\mathrm{d}s} . \tag{5.55}$$

Using the symmetry of the factor on the right, $-2\partial_k g_{ij}$ can be rewritten more symmetrically as $-\partial_k g_{ij} - \partial_j g_{ik}$. Eq. (5.55) can then be turned into an equation of the desired form by inverting the matrix g_{ij} that appears on the left-hand side. One defines g^{ij} as the matrix inverse of g_{ij}, which in index notation translates into the statement

$$g^{i\ell}g_{\ell j} = \delta^i_j , \tag{5.56}$$

where δ^i_j denotes the Kronecker δ-function (which is defined to be one if $i = j$, and zero otherwise). One can then change the free index in Eq. (5.55) to ℓ, and then multiply by $g^{i\ell}$. The result is written standardly in the form

$$\boxed{\frac{\mathrm{d}^2x^i}{\mathrm{d}s^2} = -\Gamma^i_{jk}\frac{\mathrm{d}x^j}{\mathrm{d}s}\frac{\mathrm{d}x^k}{\mathrm{d}s} ,} \tag{5.57}$$

where

$$\boxed{\Gamma^i_{jk} = \frac{1}{2}g^{i\ell}\left(\partial_j g_{\ell k} + \partial_k g_{\ell j} - \partial_\ell g_{jk}\right) .} \tag{5.58}$$

The quantity Γ^i_{jk} is called the affine connection.

THE SCHWARZSCHILD METRIC

General relativity includes a set of equations known as the Einstein field equations, which describe how a gravitational field is produced by matter. These equations are the analogue of the Maxwell equations of electromagnetism, which describe how an electromagnetic field is produced by charges and currents. The Einstein field equations are beyond the scope of this course, but it will nonetheless be useful to describe some features of the solutions to the field equations.

Of particular interest are the solutions for spherically symmetric objects, such as planets, stars, or black holes. In Newtonian mechanics, you will recall, the gravitational field outside a spherical distribution of matter has the peculiar property that it is independent of the details of the mass distribution. Outside of a spherical distribution, the field is uniquely determined if the total mass is known, independent of how this mass is distributed with radius. In general relativity, it turns out, the same feature is found — the metric is determined solely by the total mass enclosed. The metric for a spherically symmetric distribution of mass, in the region outside the mass, is given by the Schwarzschild metric,

$$\begin{aligned} \mathrm{d}s^2 = -c^2\mathrm{d}\tau^2 = &-\left(1-\frac{2GM}{rc^2}\right)c^2\mathrm{d}t^2 + \left(1-\frac{2GM}{rc^2}\right)^{-1}\mathrm{d}r^2 \\ &+ r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2 \ , \end{aligned} \tag{5.59}$$

where M is the total mass of the object, and θ and ϕ are the usual polar coordinates. Their range is given by $0 \leq \theta \leq \pi$, $0 \leq \phi < 2\pi$, and $\phi = 2\pi$ is identified with $\phi = 0$.

Note that the metric becomes singular at $r = 2GM/c^2$, which is known as the Schwarzschild radius:

$$R_S = \frac{2GM}{c^2} \ . \tag{5.60}$$

A metric is said to be singular if any of the coefficients become infinite, or if any of the coefficients vanish; in this case both happen: the coefficient of the $\mathrm{d}t^2$ term vanishes at the Schwarzschild radius, and the coefficient of $\mathrm{d}r^2$ becomes infinite. The singularity at the Schwarzschild radius, however, does not indicate any true singularity in the structure of space. If a person or instrument fell through the Schwarzschild radius, nothing peculiar would be felt. In this case the singularity is caused only by the choice of the coordinate system, and other coordinate systems can be constructed for which there is no singularity. In this course, however, we will not have time to look at such coordinate systems. The Schwarzschild metric is also singular at $r = 0$; unlike the singularity at $r = R_S$, the singularity at $r = 0$ is a true physical singularity. Physically measurable quantities, such as the tidal forces associated with nonuniform gravitational fields, become infinite at $r = 0$.

Although the singularity at $r = R_S$ is only an artifact of the coordinate system, it can be shown nonetheless that $r = R_S$ represents the point of no return for an object falling into a black hole. If any object (even a photon) falls inside the Schwarzschild radius, then it will never be able to escape. Thus, an object that is contained within its Schwarzschild radius is called a black hole. The sphere at $r = R_S$ is called the "Schwarzschild horizon," meaning that it is impossible, from the outside, to see anything beyond $r = R_S$.

The distinction between a black hole and a star is simply the question of whether this Schwarzschild horizon exists. If the matter extends to radii beyond the value of R_S indicated by Eq. (5.60), then the Schwarzschild metric will not be valid at the Schwarzschild radius. In this case the horizon may or may not exist, depending on the distribution of matter inside the object. However, if the mass distribution is so compact that it is contained within the Schwarzschild radius, then the Schwarzschild metric will describe the space outside of the matter, and the Schwarzschild horizon will be guaranteed to exist.

Just for orientation, we can compute the Schwarzschild radius of the sun, which has a mass of 1.989×10^{30} kg. Thus,

$$\begin{aligned} R_{S,\odot} &= \frac{2 \times 6.673 \times 10^{-11}\ \text{m}^3\text{-kg}^{-1}\text{-s}^{-2} \times 1.989 \times 10^{30}\ \text{kg}}{(2.998 \times 10^{8}\ \text{m-s}^{-1})^2} \\ &= 2.95\ \text{km}\ . \end{aligned} \tag{5.61}$$

So if the sun were compressed to a size smaller than 2.95 km, it would become a black hole.

GEODESICS IN THE SCHWARZSCHILD METRIC

Our purpose in introducing the Schwarzschild metric is mainly to provide an example of the calculation of a geodesic in a realistic general relativity setting.

In this section we will calculate the geodesic, and hence the trajectory, for a particle that is released from rest at $r = r_0$ in the Schwarzschild metric of Eq. (5.59). Note that r is a radial coordinate, in the sense that it provides a measure of how far a spacetime point is from the center of symmetry. However, it would be misleading to call r the radius, since it does not literally measure the distance from the center. If r is varied by an amount $\mathrm{d}r$, the new point is separated from the first not by $\mathrm{d}r$, but instead by $\mathrm{d}r/\sqrt{1 - 2GM/rc^2}$. r is sometimes called the circumferential radius, since the term $r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)$ in the metric implies that the circumference of a circle at a fixed value of r is equal to $2\pi r$, as in Euclidean geometry.

By spherical symmetry, we know that the particle will fall straight toward the center of the sphere, so the coordinates θ and ϕ will remain constant. Thus, the terms in the metric proportional to $\mathrm{d}\theta^2$ and $\mathrm{d}\phi^2$ will give no contribution as the particle moves along the trajectory. Since the spherical symmetry also guarantees that the other terms in the metric are independent of θ and ϕ, these two angles can be completely ignored in solving the problem; the values of the two angles will remain constant at their initial values.

The trajectory of such a particle is timelike, and can be parameterized by the proper time as it would be measured on a clock that moves with the particle. The trajectory can be described by the functions $r(\tau)$ and $t(\tau)$, where the latter function gives the value of the coordinate t as a function of the proper time. The metric (5.59) gives the separation $\mathrm{d}\tau^2$ between two neighboring points along the trajectory. Dividing Eq. (5.59) by $\mathrm{d}\tau^2$, one finds the relation

$$c^2 = \left(1 - \frac{2GM}{rc^2}\right) c^2 \left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 - \left(1 - \frac{2GM}{rc^2}\right)^{-1} \left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 . \tag{5.62}$$

This allows one to determine $\mathrm{d}t/\mathrm{d}\tau$ in terms of $\mathrm{d}r/\mathrm{d}\tau$. To be more compact, we introduce the notation

$$h(r) \equiv 1 - \frac{R_S}{r} = 1 - \frac{2GM}{rc^2} , \tag{5.63}$$

so Eq. (5.62) can be rewritten as

$$c^2 \left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 = c^2 h^{-1}(r) + h^{-2}(r) \left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 . \tag{5.64}$$

To generalize the geodesic equation (5.52) to spacetime trajectories, there is nothing significant that needs to be changed. We are changing the number of dimensions and we are switching to a metric that is not positive definite, but neither of these changes affect the derivation of the geodesic equation in any way. Since the trajectories of particles are timelike, we parameterize the path not by s, which would be imaginary, but instead by τ. This does not change the form of the equation either, since the only place where the parameterization mattered was when we assumed that $A = 1$, in deriving Eq. (5.52) from Eq. (5.50). But the derivation depended only on the prescription that $A = \mathit{constant}$, and not on $A = 1$. In this case we will be using $A = -c^2$, but the geodesic equation will be unaffected. So, we can rewrite the geodesic equation as

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}\tau}\left[g_{\mu\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\right] = \frac{1}{2}\frac{\partial g_{\lambda\sigma}}{\partial x^\mu}\frac{\mathrm{d}x^\lambda}{\mathrm{d}\tau}\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau} ,} \tag{5.65}$$

where I followed a common convention of using Greek letters for spacetime indices. The letters μ, ν, λ, σ, etc. are summed from 0 to 3 when they are repeated, where $x^0 \equiv t$.

Note that of the 4 components of $\mathrm{d}x^\mu/\mathrm{d}\tau$, only two are nonzero: $\mathrm{d}r/\mathrm{d}\tau$ and $\mathrm{d}t/\mathrm{d}\tau$. Since Eq. (5.64) allows us to find $\mathrm{d}t/\mathrm{d}\tau$ in terms of $\mathrm{d}r/\mathrm{d}\tau$, it will be sufficient for us to look at only the geodesic equation for $\mathrm{d}r/\mathrm{d}\tau$. Writing Eq. (5.65) for $\mu = r$, one finds

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left[g_{rr}\frac{\mathrm{d}r}{\mathrm{d}\tau}\right] = \frac{1}{2}\partial_r g_{rr}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 + \frac{1}{2}\partial_r g_{tt}\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 \ , \tag{5.66}$$

where

$$g_{rr} = h^{-1}(r) \ , \tag{5.67}$$

and

$$g_{tt} = -c^2 h(r) \ . \tag{5.68}$$

Using the fact that $\partial_r h(r) = -R_S/r^2$, Eq. (5.66) becomes

$$\begin{aligned} h^{-1}(r)\frac{\mathrm{d}^2 r}{\mathrm{d}\tau^2} - h^{-2}(r)\frac{R_S}{r^2}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 &= \\ -\frac{1}{2}h^{-2}(r)\frac{R_S}{r^2}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 &- \frac{1}{2}c^2\frac{R_S}{r^2}\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 \ . \end{aligned} \tag{5.69}$$

Now use Eq. (5.64) to eliminate $\mathrm{d}t/\mathrm{d}\tau$, and notice that the terms involving $\mathrm{d}r/\mathrm{d}\tau$ cancel against each other. The only remaining terms are proportional to $h^{-1}(r)$, so one can multiply by the inverse of this quantity to obtain

$$\frac{\mathrm{d}^2 r}{\mathrm{d}\tau^2} = -\frac{c^2}{2}\frac{R_S}{r^2} = -\frac{GM}{r^2} \ . \tag{5.70}$$

This equation is identical in form to the corresponding equation in Newtonian mechanics, but the physics is far from identical. In the Newtonian case the time variable denotes a universal time that can be read on any clock, while in the general relativity case the time variable τ represents the proper time that would be measured by a clock that is moving with the falling particle. The time that would be measured on a stationary clock would be different.

Since Eq. (5.70) is a familiar differential equation, we can integrate it without difficulty. The first step is to obtain a conservation of energy equation, which can be done by multiplying the equation by $\mathrm{d}r/\mathrm{d}\tau$. The equation can then be written as

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{\frac{1}{2}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 - \frac{GM}{r}\right\} = 0 \ , \tag{5.71}$$

which implies that the quantity in curly brackets is conserved. If the particle is released from rest at $r = r_0$, then the initial value of this conserved quantity is $-GM/r_0$, so Eq. (5.71) becomes

$$\frac{\mathrm{d}r}{\mathrm{d}\tau} = -\sqrt{2GM\left(\frac{1}{r} - \frac{1}{r_0}\right)} = -\sqrt{\frac{2GM(r_0 - r)}{rr_0}} \ . \tag{5.72}$$

This equation can be reduced to a definite integral by bringing all of the r-dependent factors to one side and integrating:

$$\tau = -\int_{r_0}^{r_f} \mathrm{d}r \sqrt{\frac{rr_0}{2GM(r_0 - r)}} \ . \tag{5.73}$$

This integral can be carried out, so finally we have an expression for the proper time $\tau(r_f)$ at which the particle is at the radius coordinate r_f:

$$\tau(r_f) = \sqrt{\frac{r_0}{2GM}} \left\{ r_0 \tan^{-1}\left(\sqrt{\frac{r_0 - r_f}{r_f}}\right) + \sqrt{r_f(r_0 - r_f)} \right\} \ . \tag{5.74}$$

So, from the point of view of a person riding on the falling particle, the Schwarzschild horizon will be reached in a finite length of time.

However, if we ask how the trajectory evolves as a function of coordinate time t, we will see a very different picture. The velocity with respect to coordinate time can be found by the chain rule:

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \frac{\mathrm{d}r}{\mathrm{d}\tau}\frac{\mathrm{d}\tau}{\mathrm{d}t} = \frac{\mathrm{d}r/\mathrm{d}\tau}{\mathrm{d}t/\mathrm{d}\tau} \ , \tag{5.75}$$

and then Eq. (5.64) can be used to eliminate $\mathrm{d}t/\mathrm{d}\tau$:

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \frac{\mathrm{d}r/\mathrm{d}\tau}{\sqrt{h^{-1}(r) + c^{-2}h^{-2}(r)\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2}} \ . \tag{5.76}$$

It is possible to find an exact solution for t as a function of r, which can be obtained by using Eq. (5.72) to eliminate $\mathrm{d}r/\mathrm{d}\tau$ from the above equation, and then expressing t as an integral over r, similar to Eq. (5.73). The result is very cumbersome, however, and not very illuminating. We are most interested, however, in how Eq. (5.76) behaves when r is near the horizon, and that behavior can be extracted rather easily. Near the horizon $h(r)$ approaches zero so $h^{-1}(r)$ blows up, with

$$h^{-1}(r) = \frac{r}{r - R_S} \approx \frac{R_S}{r - R_S} \ . \tag{5.77}$$

The argument of the square root in the denominator of Eq. (5.76) is then dominated by the second term, which with Eq. (5.77) gives

$$\frac{\mathrm{d}r}{\mathrm{d}t} \approx c\left(\frac{r - R_S}{R_S}\right) \ . \tag{5.78}$$

Rearranging and integrating to some final $r = r_f$, one finds

$$t(r_f) \approx -\frac{R_S}{c}\int^{r_f} \frac{\mathrm{d}r'}{r' - R_S} \approx -\frac{R_S}{c}\ln(r_f - R_S) \ . \tag{5.79}$$

Thus t diverges logarithmically as $r_f \to R_S$, so the object does not reach R_S for any finite value of t. Thus, even though a person falling into a black hole would pass the horizon in a finite amount of time, from the outside the person will never be seen to reach the horizon.

APPENDIX 5A: THE LORENTZ TRANSFORMATION AND THE LORENTZ-INVARIANT INTERVAL

THE LORENTZ TRANSFORMATION:

The kinematic results of special relativity which were discussed in Lecture Notes 1 — time dilation, Lorentz-Fitzgerald contraction, and the relativity of simultaneity — can all be neatly summarized in a set of equations called the Lorentz transformation. These equations relate the coordinates of an event as seen by one inertial observer to the coordinates of the same event as seen by another inertial observer in relative motion.

The Lorentz transformation can be easily derived from the principles that have already been established. Suppose that a space ship observer constructs a physical coordinate system by carrying with him an entire network of measuring rods oriented along his x- and y-axes, as in Fig. 5A.1. He also has a network of clocks. He determines the spatial coordinates of an event by observing where in this network of measuring rods it occurs, and he determines the time by reading it from a clock located at the site of the event. We will refer to these coordinates as x', y', and t', using the primes to distinguish them from our own coordinate system, which we will continue to call x, y, and t. (To simplify the discussion I am assuming that everything happens in the 2-dimensional plane spanned by the x- and y-axes. The z direction can be reinstated very easily, since its properties are the same as those of the y direction.)

Let us suppose that the moving coordinate system is oriented so that its x'-axis moves to the right along our x-axis, and the clocks are synchronized so that the clock at the origin of each system is set to zero at the time when the two origins cross each other.

Notice, that since there is no contraction of the measuring rods that are oriented perpendicular to the motion, the y-coordinate of an event has the same value in either frame. This leads to the first transformation equation,

$$y' = y \ . \tag{5A.1}$$

If there was a third spatial dimension in the problem, one would similarly conclude that $z' = z$.

Suppose now that an event A occurs in our coordinate system at a spacetime point (x, t), where we will set $y \equiv 0$ for simplicity. We now wish to calculate the coordinates as measured by the moving (primed) system. Since $y = y' = 0$, the event will occur on the measuring rod which constitutes the x'-axis of the moving system, so we can for now ignore the existence of the other measuring rods.

Fig. 5A.2 shows the trajectory of the origin of the primed coordinate system, which we will call O'. It starts at the origin of our system at $t = 0$, and then moves to the right at speed v. The diagram also shows that the moving measuring rod which connects

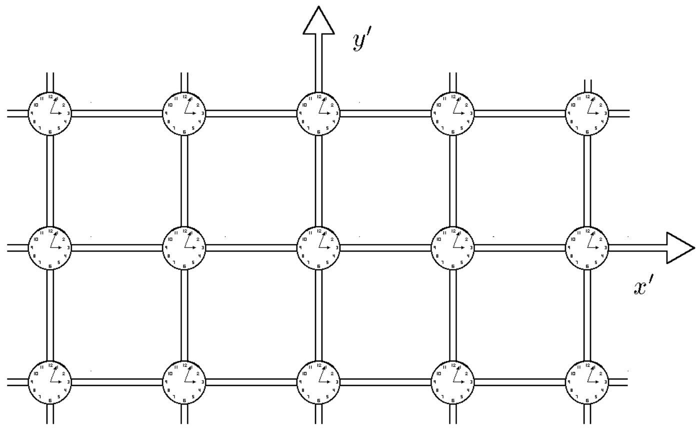


Figure 5A.1: A "physical" coordinate frame, made of clocks and measuring rods.

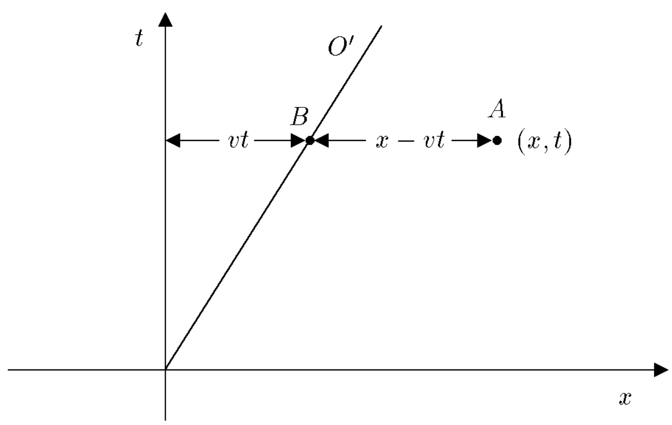


Figure 5A.2: Trajectory of O', the origin of the primed coordinate system. It starts at the origin of our system at $t = 0$, and moves to the right at speed v.

the event A to O' has length $x - vt$, when measured in our frame. However, since the measuring rod is contracted by a factor

$$\gamma \equiv \frac{1}{\sqrt{1 - \beta^2}} , \tag{5A.2}$$

it follows that the length that one would read off from the rod itself must be $\gamma(x - vt)$. Thus,

$$x' = \gamma(x - vt) . \tag{5A.3}$$

To determine t', we must find the time on the moving clock which coincides with the event A. To do this, consider first the event B which occurs at the same time as event A in our frame, but which is located at the origin O' of the moving system. Since the clock at O' is synchronized with ours at $t = 0$ and then runs slowly by a factor of γ, we know that

$$t'(B) = t/\gamma \; . \tag{5A.4}$$

However, the clock at B is trailing the clock at A, and therefore the two clocks will not appear to us to be synchronized. Instead, we have learned that the trailing clock will read a time that is later than the leading clock by an amount $\beta\ell_o/c$, where ℓ_o is rest length of the rod that joins the two clocks. In this case $\ell_o = x' = \gamma(x - vt)$, so

$$\begin{aligned} t'(A) &= t'(B) - \beta\gamma(x - vt)/c \\ &= \left(1 - \beta^2\right)\gamma t - \beta\gamma\left(\frac{x}{c} - \beta t\right) \\ &= \gamma\left(t - \frac{vx}{c^2}\right) \; . \end{aligned} \tag{5A.5}$$

This completes the derivation of the Lorentz transformation equations, which can be summarized as follows:

$$\boxed{\begin{aligned} x' &= \gamma(x - vt) \\ y' &= y \\ z' &= z \\ t' &= \gamma\left(t - \frac{vx}{c^2}\right) \; . \end{aligned}} \tag{5A.6}$$

We have already verified that there is no distinction between the moving reference frame and ours, so that the moving observer observes the same distortion in our measuring devices that we observe in his. In the formalism of the Lorentz transformation, this fact is verified by inverting the transformation. That is, the above equations can be solved to express the unprimed variables in terms of the primed variables. When this exercise is carried out, it is found that the equations have exactly the same form, except that the sign of the relative velocity v is reversed.

THE LORENTZ-INVARIANT INTERVAL:

So far we have considered only pulses of light that move either parallel or perpendicular to the direction of motion. However, the Lorentz transformation allows us to easily verify that the measured speed of light is the same in ***all*** directions. To see this, consider a spherical light pulse that emanates from the origin. In our system, the wave front moves at the speed of light and therefore satisfies the equation

$$x^2 + y^2 + z^2 = c^2 t^2 \ . \tag{5A.7}$$

We need to verify that the same equation holds for the coordinates of the wave front in the primed reference frame. We therefore use the Lorentz transformations to calculate the quantity

$$x'^2 + y'^2 + z'^2 - c^2 t'^2 \ .$$

When we carry out this somewhat complicated but straightforward calculation, we find the following remarkable relation:

$$x'^2 + y'^2 + z'^2 - c^2 t'^2 = x^2 + y^2 + z^2 - c^2 t^2 \ . \tag{5A.8}$$

This quantity,

$$x^2 + y^2 + z^2 - c^2 t^2 \ ,$$

is therefore called the "Lorentz invariant interval" between the event (x, y, z, t) and the origin.

The origin is of course not really a special point, so one can just as well define the Lorentz invariant interval between any two events A and B:

$$\boxed{s^2 \equiv (x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2 - c^2 (t_A - t_B)^2 \ .} \tag{5A.9}$$

Although I am calling the Lorentz invariant interval s^2, I obviously do not mean to imply that it is always positive — it can have either sign. I call it s^2 only because it has the units of cm^2. If s^2 is positive, then the two events are said to be spacelike separated. In that case, it can be shown that there exists a frame of reference in which the two events occur at the same time, and the value of s^2 represents the square of the distance between the events in that frame. If s^2 is negative, the two events are said to be timelike separated. In that case there exists a frame of reference in which the two events occur at the same position, and the value of s^2 represents $-c^2$ times the square of the time separation in that frame. Note also that whenever s^2 is negative one can imagine a clock that moves between the two events at a uniform speed — s^2 is then equal to $-c^2$ times the time interval as measured by the clock. This time interval is sometimes called the proper time between the two events.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — October 25, 2018
Prof. Alan Guth

Lecture Notes 6
BLACK-BODY RADIATION AND THE EARLY HISTORY OF THE UNIVERSE

INTRODUCTION:

In Lecture Notes 3 and 4 we discussed the dynamics of Newtonian cosmology under the assumption that mass is conserved as the universe expands. In that case, since the physical volume is proportional to $a^3(t)$, the mass density $\rho(t)$ is proportional to $1/a^3(t)$. In these lecture notes we will extend our understanding to include the dynamical effects of electromagnetic and other forms of radiation. Electromagnetic radiation is intrinsically relativistic ($v \equiv c$!), so we need to begin by discussing the concepts of mass and energy in the context of relativity.

According to special relativity, mass and energy are equivalent, with the conversion of units given by the famous formula,

$$E = mc^2 \ . \tag{6.1}$$

When one says that mass and energy are equivalent, one is saying that they are just two different ways of expressing precisely the same thing. The total energy of any system is equal to the total mass of the system — sometimes called the relativistic mass — times c^2, the square of the speed of light.

Although c^2 is a large number in conventional units, one can still think of it conceptually as being merely a unit conversion factor. For example, one can imagine measuring the mass/energy of an object in either grams or ergs, with

$$1 \text{ gram} = 8.9876 \times 10^{20} \text{ erg} \ , \tag{6.2}$$

where $c^2 = 8.9876 \times 10^{20}$ cm^2/s^2. So one gram is a ***huge*** number of ergs. For SI units,

$$1 \text{ kg} = 8.9876 \times 10^{16} \text{ joule} = 2.497 \times 10^{10} \text{ kw-hr.} \tag{6.3}$$

To put this number in perspective, we might compare it to the world power supply, which is about 1.8×10^{10} kilowatts, according to the International Energy Agency.* Thus, if we

* *Key World Energy Statistics, 2017*, http://www.iea.org/publications/freepublications/publication/KeyWorld2017.pdf. The 2015 annual "Total Primary Energy Supply" is given as 13,647 million tonnes of oil equivalent (Mtoe), which in more familiar units is 1.587×10^{14} kW-hr. If this energy production were uniformly spread over the year, the average power would be 1.811×10^{10} kW. With a 2015 world population of 7.35 billion people, this corresponds to 2.46 kW per person.

could build a machine that would convert 1 kg per hour entirely into energy, its power output would be about 1.5 times the world's total power supply. A 15 gallon tank of gasoline, if it could be converted entirely to energy, would power the world for two and a half days. Unfortunately, however, it is not so easy: when a uranium-235 nucleus undergoes fission, for example, only about 0.09% of its mass is converted to energy.

Since c is conceptually a unit conversion factor, many physicists (especially nuclear and particle physicists) work in unit systems for which $c \equiv 1$. A common choice is to use the MeV (10^6 eV) or GeV (10^9 eV) as the unit of energy, where

$$1\,\mathrm{eV} = 1 \text{ electron volt} = 1.6022 \times 10^{-19} \text{ J}, \tag{6.4}$$

and then

$$1\,\mathrm{GeV} = 1.7827 \times 10^{-27} \text{ kg}. \tag{6.5}$$

The mass of a proton is 0.938 GeV.

It will be useful to know some basic properties of the energy-momentum four-vector, so I will summarize them here. The energy-momentum four-vector is defined by starting with the momentum three-vector $(p^1, p^2, p^3) \equiv (p^x, p^y, p^z)$, and appending a fourth component

$$p^0 = \frac{E}{c} \,, \tag{6.6}$$

so the four-vector can be written as

$$p^\mu = \left(\frac{E}{c} \,, \vec{p} \right) \,. \tag{6.7}$$

As with the three-vector momentum, the energy-momentum four-vector can be defined for a system of particles as the sum of the vectors for the individual particles. The motivation for putting the four components together is that the four-vector obeys a simple transformation law that describes how to calculate the components measured by an inertial observer in terms of the components measured by another inertial observer who is moving relative to the first. The transformation law is identical to one that describes the transformation of the spacetime coordinate vector, $x^\mu = (ct, \vec{x})$, known as the Lorentz transformation. The mass of a particle in its own rest frame is called its rest mass, which we denote by m_0. If the particle moves with velocity $\vec{v}$, then the relativistic expressions for its momentum and energy are given by

$$\begin{aligned} \vec{p} &= \gamma m_0 \vec{v} \,, \\ E &= \gamma m_0 c^2 = \sqrt{\left(m_0 c^2\right)^2 + \left|\vec{p}\right|^2 c^2} \,, \end{aligned} \tag{6.8}$$

where as usual γ is defined by

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \ . \tag{6.9}$$

Like the Lorentz-invariant interval that we discussed with Eq. (5.30), the energy-momentum four-vector has a Lorentz-invariant square:

$$p^2 \equiv |\vec{p}|^2 - \left(p^0\right)^2 = |\vec{p}|^2 - \frac{E^2}{c^2} = -\left(m_0 c\right)^2 \ . \tag{6.10}$$

For a particle at rest, Eq. (6.8) implies that the energy E_0 is given by

$$E_0 = m_0 c^2 \ , \tag{6.11}$$

since $\vec{p} = 0$. To see the implications of this equation, we can imagine a hydrogen atom, which is composed of a proton and an electron. If the two particles are started an infinite distance apart, then the initial total energy is given by $E_{\rm tot} = (m_p + m_e)c^2$, where we are defining the zero of potential energy so that it vanishes at infinite separation. As the particles come together they attract each other, and therefore accelerate. They gain kinetic energy, and the potential energy becomes negative. If the particles combine to form a hydrogen atom in its ground state (i.e., its lowest energy state), then an energy ΔE is given off. This energy is called the binding energy of the hydrogen, and has a value of 13.6 eV. The energy is most commonly given off in the form of photons. (There is also some kinetic energy associated with the recoil of the hydrogen atom, but the recoil energy is very small when the rest energy of the recoiling object is large compared to the energy given off. Here we will ignore the recoil.) The mass m_H of the resulting hydrogen atom is then given by

$$\boxed{m_H = m_p + m_e - \Delta E/c^2 \ ,} \tag{6.12}$$

where m_p is the mass of the proton, and m_e is the mass of the electron. The rest mass of the system is reduced by the energy given off, divided by c^2. Thus, a small part of the rest mass of the proton and electron has been converted into other forms of energy.

For a particle in motion, one can define a relativistic mass $m_{\rm rel}$ by

$$m_{\rm rel} = \frac{E}{c^2} \ . \tag{6.13}$$

Many authors prefer to never introduce the concept of relativistic mass, and it is certainly not necessary. Since it is defined solely in terms of the energy, anything that can be thought or said in terms of the relativistic mass of a particle can equally well be expressed in terms of its energy. However, when one discusses the gravitational field of a system including relativistic particles, then the concept of relativistic mass can be useful. The gravitational field of a single moving particle, according to general relativity, is anisotropic and rather complicated, but fortunately we will not have to deal with this. However, if one has a gas of relativistic particles with no net momentum in the frame of interest, then the gravitational field can be computed as if the particles were at rest, but using the relativistic mass, as defined by Eq. (6.13). If one adopts the concept of relativistic mass, then the famous equation $E = mc^2$ can be described by saying that energy and mass are equivalent, related in all cases by a factor of c^2. The concept of relativistic mass is also useful when discussing the gravitational force that acts on a body. If a gas of relativistic particles were sealed inside a box, and the box were placed on a scale, then the scale would register the relativistic mass of the particles in the gas.

THE MASS OF RADIATION:

We are perhaps not used to thinking of electromagnetic radiation as having mass, but it is well-known that radiation has an energy density. If the energy density is denoted by u, then the electromagnetic radiation has a relativistic mass density ρ given by

$$\boxed{\rho = u/c^2\ .} \tag{6.14}$$

That is, the formula above describes the amount of relativistic mass ($m_{\rm rel}$) per unit volume. According to general relativity, such a mass density contributes to the gravitational field just like any other mass density.*

To my knowledge nobody has ever actually "weighed" electromagnetic radiation in any way, but the theoretical evidence in favor of Eq. (6.14) is overwhelming — light does have mass. Nonetheless, the photon has zero rest mass, meaning that it cannot be brought to rest. The general relation for the square of the four-momentum reads $p^2 = -(m_0c)^2$, as in Eq. (6.10), so for the photon this becomes $p^2 = 0$. Writing out the square of the four-momentum leads to the following relation for photons:

$$\boxed{|\vec{p}\,|^2 - \frac{E^2}{c^2} = 0\ , \quad \text{or} \quad E = c|\vec{p}\,|\ .} \tag{6.15}$$

In this set of notes we will examine the role which the mass of electromagnetic radiation plays in the early stages of the universe.

---

* Authors who avoid the concept of relativistic mass would reach the same conclusion, but would describe it by saying that the energy density u creates the same gravitational field as a mass density u/c^2.

RADIATION IN AN EXPANDING UNIVERSE:

If we ignore the interactions of photons, then as the universe expands the photons travel on geodesics, and their number is conserved. We will learn later that even when we take into account the emission and absorption of photons by the matter in the universe, their number is still very accurately conserved during the long period after inflation (to be discussed later) and before the formation of the earliest stars. As long as the number is conserved, the number density n_γ of photons varies as $1/a^3(t)$ as the universe expands, just like the number density of nonrelativistic particles:

$$n_\gamma \propto \frac{1}{a^3(t)} \; . \tag{6.16}$$

Note that the Greek letter γ ("gamma") is often used to denote the photon, even when the energy of the photon is far from the range of 10^4–10^7 eV that normally characterizes what are called gamma rays.

Unlike nonrelativistic particles, however, the frequency of each photon is redshifted as the universe expands, as we learned in Lecture Notes 2. The ratio of the period Δt at the time t_2 to the period at the time t_1 is given by the redshift factor

$$\frac{\Delta t(t_2)}{\Delta t(t_1)} \equiv 1 + z = \frac{a(t_2)}{a(t_1)} \; . \tag{6.17}$$

Since the frequency ν (Greek letter "nu") of each photon is related to the period by $\nu = 1/\Delta t$, the frequency of each photon decreases as $1/a(t)$ as the universe expands. According to elementary quantum mechanics, the energy of the photon is related to the frequency by

$$E = h\nu \; , \tag{6.18}$$

where h is Planck's constant ($h = 4.136 \times 10^{-15}$ eV-s). Thus the energy of the photon decreases as $1/a(t)$ as the universe expands. The energy density u_γ of the radiation is given by

$$u_\gamma = n_\gamma \, E_\gamma \; , \tag{6.19}$$

where E_γ is the mean energy per photon, so

$$n_\gamma \propto \frac{1}{a^3(t)} \quad , \quad E_\gamma \propto \frac{1}{a(t)} \qquad \Longrightarrow \qquad \boxed{\rho_\gamma = \frac{u_\gamma}{c^2} \propto \frac{1}{a^4(t)} \; .} \tag{6.20}$$

(Although I have justified this relation with quantum mechanical arguments, it can also be derived from classical electromagnetic theory. However, in this case the quantum argument is simpler.)

THE RADIATION–DOMINATED ERA:

Today the energy density u_r in the cosmic background radiation is given approximately by

$$u_r = 7.01 \times 10^{-14}\ \mathrm{J/m^3}\ . \tag{6.21}$$

(Here I have used the subscript "r" for radiation, rather than "γ" for photons, because I have included both the energy density of photons and the expected density of neutrinos, which we will talk about later.) To find the corresponding mass density, use

$$\begin{aligned}\rho_r = \frac{u}{c^2} &= \frac{7.01 \times 10^{-14}\ \left(\mathrm{kg\text{-}m^2\text{-}s^{-2}}\right)\mathrm{m^{-3}}}{\left(3 \times 10^8\ \mathrm{m\text{-}s^{-1}}\right)^2} \\ &= 7.80 \times 10^{-31}\ \mathrm{kg/m^3} = 7.80 \times 10^{-34}\ \mathrm{g/cm^3}\ .\end{aligned} \tag{6.22}$$

This can be compared with the critical mass density ρ_c, which was calculated in Eq. (3.34):

$$\rho_c = 1.88\, h_0^2\ \times 10^{-29}\ \mathrm{g/cm^3}\ , \tag{3.34}$$

where

$$H_0 = 100\, h_0\, \mathrm{km\text{-}s^{-1}\text{-}Mpc^{-1}}\ .$$

One finds that the fraction Ω_r of closure density in radiation is given by

$$\Omega_r \equiv \frac{\rho_r}{\rho_c} = \frac{7.80 \times 10^{-34}\ \mathrm{g\text{-}cm^{-3}}}{1.88 h_0^2 \times 10^{-29}\ \mathrm{g\text{-}cm^{-3}}} = 4.15 \times 10^{-5}\, h_0^{-2}\ , \tag{6.23}$$

For $h_0 = 0.67$, one finds $\Omega_r = 9.2 \times 10^{-5}$. This is only a very small fraction, but Ω_r was larger in the past. Since $\rho_r \propto 1/a^4$, while the mass density ρ_m of nonrelativistic matter behaves as $1/a^3$, it follows that

$$\rho_r/\rho_m \propto 1/a(t)\ . \tag{6.24}$$

Then density of nonrelativistic matter in our universe (visible and dark matter combined) gives $\Omega_m \approx 0.30$, so today $\rho_r/\rho_m \approx 9.2 \times 10^{-5}/0.30 \approx 3.1 \times 10^{-4}$. The constant of proportionality in Eq. (6.24) is then determined, giving

$$\frac{\rho_r(t)}{\rho_m(t)} = \left[a(t_0)\frac{\rho_r(t_0)}{\rho_m(t_0)}\right]\frac{1}{a(t)} = \frac{a(t_0)}{a(t)} \times 3.1 \times 10^{-4}\ . \tag{6.25}$$

Since $a(t) \to 0$ as $t \to 0$, the right-hand-side approaches infinity in this limit. Thus there was a time at which the value of the right-hand-side went through one, and this time is denoted by t_{eq}, the time of radiation-matter equality. We will assume that the universe is flat, and that for $t > t_{\rm eq}$ we can make the crude approximation that the universe can be treated as if it were dominated by nonrelativistic matter. This approximation ignores the effect of radiation for times shortly after $t_{\rm eq}$, and it also ignores the effect of dark energy (and the consequent acceleration) during the past 5 billion years or so. As discussed in Lecture Notes 3, during the matter-dominated era the scale factor behaves as $a(t) \propto t^{2/3}$. Thus, writing Eq. (6.25) for $t = t_{\rm eq}$ gives

$$\frac{\rho_r(t_{\rm eq})}{\rho_m(t_{\rm eq})} \equiv 1 = \frac{a(t_0)}{a(t_{\rm eq})} \times 3.1 \times 10^{-4} \ . \tag{6.26}$$

Remembering that $a(t_0)/a(t_{\rm eq}) = 1 + z_{\rm eq}$ (see Eq. (2.15)), the redshift $z_{\rm eq}$ of matter-radiation equality is given by

$$z_{\rm eq} = \frac{1}{3.1 \times 10^{-4}} - 1 \approx 3200 \ . \tag{6.27}$$

If we ignore for now the acceleration that our universe has undergone during the last 5 billion years or so, we can approximate it as a flat matter-dominated universe, with $a(t) \propto t^{2/3}$. This gives $t_{\rm eq} = 5.5 \times 10^{-6}\, t_0$, so for $t_0 = 13.8$ Gyr, $t_{\rm eq} \approx 75,000$ years. Our approximations have been crude, but Barbara Ryden quotes a more precise numerical calculation (on p. 97), where she finds $t_{\rm eq} \approx 47,000$ years.

DYNAMICS OF THE RADIATION–DOMINATED ERA:

When we studied the dynamics of a matter-dominated universe (i.e., a universe whose mass density is dominated by nonrelativistic matter) in Lecture Notes 3, we learned that the evolution of such a universe can be described by the two Friedmann equations:

$$\left(\frac{\dot a}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2} \tag{6.28a}$$

$$\ddot a = -\frac{4\pi}{3}G\rho a \ , \tag{6.28b}$$

$$\left(\begin{matrix}\text{matter-dominated}\\ \text{universe}\end{matrix}\right)$$

where $a(t)$ is the scale factor, $\rho(t)$ is the mass density, and an overdot represents differentiation with respect to time t. In such a matter-dominated universe we found that the mass density behaves as

$$\rho(t) \propto \frac{1}{a^3(t)} \qquad \text{(matter-dominated).} \tag{6.29}$$

The three equations above are not independent, but in fact any two of them can be used to derive the third. For example we can derive Eq. (6.28b) by multiplying Eq. (6.28a) by

a^2 and then differentiating it with respect to time. The resulting equation will contain a term proportional to $\dot{\rho}$. Eq. (6.28b) can then be obtained by replacing $\dot{\rho}$ by

$$\dot{\rho} = -3\frac{\dot{a}}{a}\,\rho \qquad \text{(matter-dominated)}, \tag{6.30}$$

which can be derived from Eq. (6.29).

For a universe dominated by radiation, we have already learned (see Eq. (6.20)) that

$$\rho(t) \propto \frac{1}{a^4(t)} \qquad \text{(radiation-dominated)}, \tag{6.31}$$

in contrast to Eq. (6.29). This implies that Eqs. (6.28a) and (6.28b) will no longer be consistent with each other, since the derivation of Eq. (6.28b) described in the previous paragraph will give a different result. To correctly describe a radiation-dominated universe, we will have to reconcile this inconsistency.

While we have not yet used the word, Eq. (6.31) can be viewed as a statement about the *pressure* of radiation. Pressure is relevant, because it is the pressure of a gas that determines how much energy it looses if it expands. Consider, as a thought experiment, a volume of gas contained in a chamber with a movable piston, as shown below:

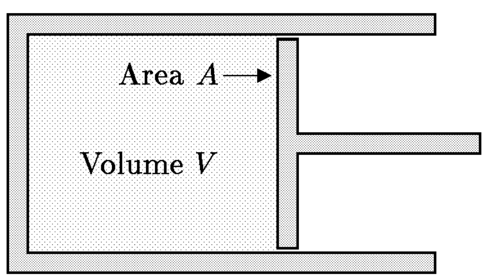


Figure 6.1: A piston chamber, used to discuss the effect of pressure on the rate of change of the energy density of an expanding gas.

We will assume that the piston chamber is small enough so that gravity plays no role in our thought experiment. Let U denote the total energy of the gas, and let p denote the pressure. Suppose that the piston is moved a distance dx to the right. (We suppose that the motion is slow, so that the gas particles have time to respond and to maintain a uniform pressure throughout the volume.) The gas exerts a force pA on the piston, so the gas does work $dW = pA\,dx$ as the piston is moved. The volume increases by an amount $dV = A\,dx$, so $dW = p\,dV$. The energy of the gas decreases by this amount, so

$$dU = -p\,dV\ . \tag{6.32}$$

It can be shown that this relation is valid whenever the volume of a gas is changed, regardless of the shape of the volume.

Now consider a homogeneous, isotropic, expanding universe, described by a scale factor $a(t)$. Let $u = \rho c^2$ denote the energy density of the gas that fills it. We will consider a fixed coordinate volume V_{coord}, so the physical volume will vary as

$$V_{\text{phys}}(t) = a^3(t) V_{\text{coord}} \ , \tag{6.33}$$

and the energy of the gas in this region is given by

$$U = V_{\text{phys}} u \ . \tag{6.34}$$

Using these relations, you will show in Problem Set 6 that

$$\frac{d}{dt}\left(a^3 \rho c^2\right) = -p \frac{d}{dt}(a^3) \ , \tag{6.35}$$

and then that

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) \ . \tag{6.36}$$

By comparing this equation with the matter-dominated relation of Eq. (6.30), we see that nonrelativistic matter has zero pressure. This could have been expected, since nonrelativistic matter means a gas of approximately motionless particles, and we assumed starting in Lecture Notes 3 that there is no loss of energy when the universe filled with nonrelativistic matter expands — the energy spreads out as the volume increases, but otherwise it is not changed. By contrast, you will also show in Problem Set 6 that radiation, with a mass density that falls off as $1/a^4(t)$, has a pressure given by

$$p = \frac{1}{3} u = \frac{1}{3} \rho c^2 \ . \tag{6.37}$$

Thus, the new ingredient that is introduced by radiation, which is causing an inconsistency between Eqs. (6.28a) and (6.28b), is pressure.

The treatment of pressure in general relativity is unambiguous, and the implication for this situation is simple: the $\dot{a}$ equation (6.28a) is not modified, but the $\ddot{a}$ equation (6.28b) needs to be modified. By accepting Eq. (6.28a) and using Eq. (6.36) for $\dot{\rho}$, you will show in Problem Set 6 that Eq. (6.28b) must be modified to read

$$\boxed{\frac{d^2 a}{dt^2} = -\frac{4\pi}{3} G \left(\rho + \frac{3p}{c^2}\right) a \ .} \tag{6.38}$$

While general relativity might be needed to prove the above equation, Newtonian arguments are sufficient to at least make this result seem extremely plausible. We know that when the pressure is non-negligible, $\dot{\rho}$ is given by Eq. (6.36), and that then Eqs. (6.28a) and (6.28b) become incompatible. One or both of these equations, therefore, must be modified by the presence of pressure. The two equations are different from each other, however, in an obvious way. The $\ddot{a}$ equation is a force equation, as in $\vec{F} = m\vec{a}$, and in fact we derived it in our Newtonian model by applying $\vec{F} = m\vec{a}$ to each particle in the model universe. The $\dot{a}$ equation, on the other hand, was derived by finding a first integral of the $\ddot{a}$ equation, and therefore looks like a conservation of energy equation. In fact, we showed in Problem 3, Problem Set 3, that for the Newtonian model with a finite radius $R_{\rm max}$, the $\dot{a}$ equation is precisely equivalent to the statement that the total energy of the Newtonian model universe is fixed. Does it make sense to add a pressure term to a conservation of energy equation? No, it does not. As a toy problem, we can ask what would happen if the universe were filled with TNT, and at a certain pre-arranged time little gremlins throughout the universe ignited the TNT, so the pressure suddenly changed. The pressure change can in principle be very large and fast, but there is no mechanism to cause any of the other quantities in Eq. (6.28a) to change rapidly. We can consider a small region of space, in which the velocities associated with the Hubble expansion are all small, so we can expect that we can trust our Newtonian understanding of how matter should behave. In that case ρ describes an energy density that cannot change discontinuously, and a and $\dot{a}$ describe the positions and velocities of particles, which also cannot change discontinuously. So, our conclusion is that a term depending on the pressure cannot be added to Eq. (6.28a), and then Eq. (6.38) follows as a consequence.

Note that Eq. (6.38) is implying something that is perhaps very surprising: the pressure is contributing to the gravitational acceleration. That is, the pressure as well as the energy density can act as a source for the gravitational field. We will not make much use of Eq. (6.38) in the rest of this chapter, as Eq. (6.28a) will be sufficient for most of our conclusions. But we can keep in mind that Eq. (6.28a) would not be consistent with $\rho(t) \propto 1/a^4(t)$ if Eq. (6.38) were not true. We will learn later that the pressure term in Eq. (6.38) can have dramatically new consequences. In particular, we will learn that pressures, unlike mass densities, can sometimes be negative. Eq. (6.38) implies that a negative pressure can result in a gravitational repulsion. We believe that the current acceleration of the universe, which we mentioned briefly in Lecture Notes 3, can be attributed to the negative pressure of an unidentified material that is called *dark energy*. Many of us also believe that the early universe underwent a very brief period of incredibly rapid acceleration, called *inflation*, which was also driven by a negative pressure. We will return to both of these topics in later sets of lecture notes.

DYNAMICS OF A FLAT RADIATION-DOMINATED UNIVERSE:

As a simple (but important) special case, consider the evolution of a radiation-dominated universe with $k = 0$. From Eqs. (6.20) and (6.28a), one has

$$\frac{1}{a^2}\left(\frac{da}{dt}\right)^2 = \frac{\text{const}}{a^4} \ , \tag{6.39}$$

which leads to

$$\frac{da}{dt} = \frac{\sqrt{\text{const}}}{a} \ . \tag{6.40}$$

This equation can be solved by rewriting it as

$$a\,da = \sqrt{\text{const}}\; dt \tag{6.41}$$

and then integrating both sides to obtain

$$\frac{1}{2}a^2 = \sqrt{\text{const}}\; t + \text{const}' \ . \tag{6.42}$$

The convention is to choose the zero of time so that $a(t) = 0$ for $t = 0$, which implies that $\text{const}' = 0$. Thus, the final result can be written as

$$\boxed{a(t) \propto \sqrt{t} \qquad \text{(radiation-dominated)} \ .} \tag{6.43}$$

The Hubble expansion rate $H(t)$ is given by Eq. (2.8), which says that

$$H(t) = \dot{a}/a \ . \tag{6.44}$$

Combining this equation with Eq. (6.43), one has immediately that

$$\boxed{H(t) = \frac{1}{2t} \qquad \text{(radiation-dominated)} \ .} \tag{6.45}$$

The age of a radiation-dominated universe is therefore related to the Hubble constant by $t = \frac{1}{2}H^{-1}$. (Recall for comparison that for a matter-dominated flat universe with $a(t) \propto t^{2/3}$, the age is $\frac{2}{3}H^{-1}$.) The horizon distance is given by Eq. (4.7), and the result here is

$$\begin{aligned} \ell_{p,\text{horizon}}(t) &= a(t)\int_0^t \frac{c}{a(t')}dt' \\ &= \boxed{2ct \qquad \text{(radiation-dominated)} \ .} \end{aligned} \tag{6.46}$$

(Recall that this answer is to be compared with $3ct$ for the matter-dominated universe.) If one inserts Eq. (6.45) into Eq. (6.28a) (with $k = 0$, still), one obtains a relation for the mass density as a function of time:

$$\rho = \frac{3}{32\pi G t^2} \ . \tag{6.47}$$

Note that the $1/t^2$ behavior in the above equation is consistent with what we already know: $\rho \propto 1/a^4(t)$, and $a(t) \propto \sqrt{t}$.

BLACK–BODY RADIATION:

If a cavity is carved out of any material, and the walls of the cavity are kept at a uniform temperature T, then the cavity will fill with radiation. Assuming that the walls are thick enough so that no radiation can get through them, then the energy density (and also the entire spectrum of the radiation) is determined solely by the temperature T — the composition of the material is entirely irrelevant. The material is serving solely to keep the radiation at a uniform temperature. Radiation of this type is generally called either thermal radiation or black-body radiation.

The motivation for the name "black-body radiation" stems from the fact that a "black" body in empty space can be shown to emit radiation of exactly this intensity and spectrum. Here the word "black" is used to describe an object that absorbs all light that hits it, so there is no reflected light, although there is emission due to thermal effects. Emission is distinguished from reflection by the fact that reflection is an immediate response to the radiation that is currently hitting the material. To understand the radiation emitted by a black body, imagine a block of such material inside the cavity described in the previous paragraph. Since thermal equilibrium has been established, one concludes that the block at temperature T must emit radiation which precisely matches the radiation that it is absorbing — otherwise it would either heat up or cool down, and that would violate the assumption of thermal equilibrium. In fact, not only must the energy densities match, but the entire spectrum must match — otherwise one could imagine introducing a frequency-selecting filter that would cause the black body to heat or cool. That is, if there were any frequency band for which the radiation emitted by the block did not match the radiation hitting the block, then we could surround the block by a filter that transmits only in that frequency band, and we would see the block heat up or cool down. Since objects will never heat up or cool down once thermal equilibrium is reached, the emitted and absorbed radiation must match in every frequency band. Since the block is assumed to be black, none of the emitted radiation is reflection, so all of it is thermal emission that will continue to be emitted even if the block is removed from the cavity. Thus, a black body will emit radiation with an intensity and a spectrum that depends only on the temperature, and not on any property of the material other than the fact that it is black.

The energy density and other properties of the radiation can be derived using the standard principles of statistical mechanics, but the derivation will not be included in this course. However, I will make a few comments about the underlying physics, and then I will state the results. The rule of thumb for classical statistical mechanics is the "equipartition theorem," which says that under certain circumstances (which I will not specify), each degree of freedom of a system at temperature T acquires a mean thermal energy of $\frac{1}{2}kT$. For example, in a gas of point particles each particle acquires a mean thermal energy of $\frac{3}{2}kT$, since motion in the x, y and z directions constitutes three degrees of freedom. For the system of radiation inside a cavity, each possible standing wave pattern corresponds to one degree of freedom. In a rectangular cavity, for example, a standing wave can be described in terms of a polarization, which has two linearly independent values, and a wave vector $\vec{k}$, with the wave amplitude proportional to $\mathrm{Re}\{e^{i\vec{k}\cdot\vec{x}}\}$. For the standing wave to exist, each component of $\vec{k}$ must satisfy the condition that the wave amplitude must vary either an integral or half-integral number of cycles from one side of the cavity to the other. Thus a standing wave pattern exists only for a discrete set of frequencies. The discrete set of frequencies is, however, infinite, since there is no upper limit to the frequency of a standing wave. The number of degrees of freedom is therefore infinite, and the equipartition theorem cannot be applied. This problem is known as the "Jeans catastrophe," and represents an important failure of classical physics. The implications can be stated as follows: if classical physics were correct, then a region of space containing an electromagnetic field could never come into thermal equilibrium — instead it would continue indefinitely to absorb energy from its surroundings, and the energy absorbed would be used to excite higher and higher frequency standing waves of the field. The electromagnetic field would be an infinite heat sink, draining away all thermal energy.

Of course the electromagnetic field does not drain away all thermal energy, and the reason comes from quantum theory. Classically it would be possible to excite a standing wave by an arbitrary amount, but quantum theory requires that the excitations occur only by the addition of discrete photons, each with an energy $h\nu$, where ν is the frequency of the standing wave. For cases in which $h\nu \ll kT$, the classical answer is not changed — such standing waves acquire a mean energy of $\frac{1}{2}kT$ for each polarization. However, for those standing waves with $h\nu \gg kT$, the minimum excitation is much larger than the energy which is classically expected. These modes are then only rarely excited, and the total energy is convergent.

When the calculation is done quantum mechanically, one finds that black-body electromagnetic radiation has an energy density given by

$$\boxed{u = g\frac{\pi^2}{30}\,\frac{(kT)^4}{(\hbar c)^3}\ ,} \tag{6.48}$$

where

$$\begin{aligned} k = \text{Boltzmann's constant} &= 1.381 \times 10^{-16}\,\text{erg/K} \\ &= 8.617 \times 10^{-5}\,\text{eV/K} \quad , \qquad (6.49) \\ \hbar = \frac{h}{2\pi} &= 1.055 \times 10^{-27}\ \text{erg-sec} \\ &= 6.582 \times 10^{-16}\ \text{eV-sec} \ , \end{aligned}$$

and

$$\boxed{g = 2 \quad \text{(for photons)}\ .} \qquad (6.50)$$

The factor of g is introduced to prepare for the discussion below of black body radiation of particles other than photons. g is taken as 2 for photons because the photon has two possible polarization states. The polarization states can be described as linearly polarized, or as circularly polarized, depending on one's choice of basis. In either case, however, there are two polarizations. A photon traveling along the z-axis can be linearly polarized in either the x or y directions, or it can have a circular polarization of left or right. The polarization is related to the intrinsic angular momentum, or spin, of the photon: right circular polarization corresponds to the spin being aligned with the momentum, while left circular polarization is the opposite. Thus one could say that g is taken as 2 because the photon has two spin states.

One also finds that the radiation has a pressure, given by

$$\boxed{p = \frac{1}{3} u\ .} \qquad (6.51)$$

The number density of photons is found to be

$$\boxed{n = g^* \frac{\zeta(3)}{\pi^2} \, \frac{(kT)^3}{(\hbar c)^3}\ ,} \qquad (6.52)$$

where

$$\zeta(3) = \frac{1}{1^3} + \frac{1}{2^3} + \frac{1}{3^3} + \cdots \approx 1.202 \qquad (6.53)$$

is the Riemann zeta function evaluated at 3, and

$$\boxed{g^* = 2 \quad \text{(for photons)}\ .} \tag{6.54}$$

Finally, the radiation has an entropy density s (entropy per unit volume) given by

$$\boxed{s = g\frac{2\pi^2}{45}\ \frac{k^4T^3}{(\hbar c)^3}\ .} \tag{6.55}$$

We will not need to know the precise meaning of entropy, but it will suffice to say that the entropy is a measure of the degree of disorder (or uncertainty) in the statistical system. Entropy is conserved if the system remains in thermal equilibrium, and this assumption appears to be quite accurate for most processes in the early universe. (The inflationary process, to be discussed later, is a colossal exception.) When departures from thermal equilibrium occur, the entropy is monotonically increasing, a principle known as the second law of thermodynamics.

In the laboratory the only kind of thermal radiation that can be achieved is that of photons. The radiation in the early universe, on the other hand, is believed to have also contained neutrinos. During the 20th century these neutrinos were thought to have zero rest mass, like the photon, but that is no longer the case. We now believe that neutrinos have a very small but nonzero mass. Nonetheless, as long as $m_0c^2 \ll kT$, which is certainly the case throughout the history of the universe, the neutrinos contribute to the thermal radiation as if they were massless particles.

Besides having a nonzero rest mass, neutrinos differ from photons in another property which has an important effect on their thermal radiation. The photon belongs to a class of particles called bosons, and these particles have the property that there is no limit to the number of particles that can exist simultaneously in a given quantum state. It is precisely because of this property that the photon can give rise to a classical electromagnetic field. The field behaves classically because it is composed of huge numbers of photons. The neutrino, on the other hand, belongs to a class of particles called fermions. For these particles it is impossible to have more than one particle in a given quantum state at one time. An electron is also a fermion, and the principle of one electron per quantum state is sometimes called the "Pauli Exclusion principle."

In relativistic quantum field theory it is possible to prove the *spin-statistics theorem,* which says that the boson/fermion property of a particle is connected to its intrinsic angular momentum, also called the particle's spin. If the spin is an integer (in units of $\hbar$), then the particle must be a boson. The only other possibility is that the spin is half-integer (more precisely, half-odd-integer, again in units of $\hbar$), in which case the particle is a fermion. The proof requires relativistic invariance, so there is no analogous theorem in nonrelativistic quantum mechanics.

Since fermions obey the Pauli exclusion principle, which is a restriction on the states that they can occupy, the fact that a particle is a fermion leads to a reduction in the number of particles that will be present in black-body radiation. The equations that describe the black-body radiation of fermions have the same form as the equations for bosons, so the energy density u, the pressure p, the number density n, and the entropy density s are again described by Eqs. (6.48), (6.51), (6.52), and (6.55) above. The Pauli exclusion principle, however, causes the factor g to be multiplied by 7/8 if the particle is a fermion, and the factor g^* to be multiplied by 3/4.

To find the values of g and g^* for neutrinos, we must count how many types of neutrinos exist. While there is only one kind of photon, we believe that there are three different species, or *flavors*, of neutrinos: the electron neutrino ν_e, the muon neutrino ν_μ, and the tau neutrino ν_τ. The existence of the three species causes g and g^* to be multiplied by 3. In addition, neutrinos exist as particles and antiparticles, in contrast to the photon which is its own antiparticle. The particle/antiparticle option leads to a factor of 2 for both g and g^*. While the photon has two spin states, the neutrino has only 1: neutrinos are *left-handed*, which means that their spin points in the opposite direction from their momentum, while antineutrinos are *right-handed.* Thus the values of g and g^* for neutrinos are given by

$$\boxed{\begin{aligned} g_\nu &= \underbrace{\frac{7}{8}}_{\text{Fermion factor}} \times \underbrace{3}_{\substack{\text{3 species} \\ \nu_e,\nu_\mu,\nu_\tau}} \times \underbrace{2}_{\text{Particle/antiparticle}} \times \underbrace{1}_{\text{Spin states}} = \frac{21}{4} \;. \qquad (6.56) \\ g^*_\nu &= \underbrace{\frac{3}{4}}_{\text{Fermion factor}} \times \underbrace{3}_{\substack{\text{3 species} \\ \nu_e,\nu_\mu,\nu_\tau}} \times \underbrace{2}_{\text{Particle/antiparticle}} \times \underbrace{1}_{\text{Spin states}} = \frac{9}{2} \;. \qquad (6.57) \end{aligned}}$$

[One might wonder why neutrinos are not produced when a piece of metal is heated until it glows. The answer is that neutrinos interact very weakly at these low energies, and their production rate is totally negligible. Thermal equilibrium neutrino radiation can in principle be seen at any temperature, but it is very difficult to produce. The radiation would reach thermal equilibrium only if it were confined to a box opaque to neutrinos, which means that the walls of the box would have to be much thicker than the diameter of the earth. In the early universe, however, the temperatures were much higher. Neutrino interaction rates increase with energy, so in the early universe they interacted rapidly with the other particles, and were quickly brought to thermal equilibrium.]

As the temperature is increased, more and more types of particles contribute to the thermal radiation. Any particle with $mc^2 \ll kT$ will contribute in essentially the same way as a massless particle. In particular, when kT is much larger than the value of mc^2 for an electron (0.511 MeV), then electron-positron pairs contribute to the thermal radiation. Electrons and positrons each have two spin states, and they are antiparticles of each other. They are again fermions, so

$$\boxed{\begin{aligned} g_{e^+e^-} &= \underbrace{\frac{7}{8}}_{\text{Fermion factor}} \times \underbrace{1}_{\text{Species}} \times \underbrace{2}_{\text{Particle/antiparticle}} \times \underbrace{2}_{\text{Spin states}} = \frac{7}{2}\,. & (6.58)\\ g^*_{e^+e^-} &= \underbrace{\frac{3}{4}}_{\text{Fermion factor}} \times \underbrace{1}_{\text{Species}} \times \underbrace{2}_{\text{Particle/antiparticle}} \times \underbrace{2}_{\text{Spin states}} = 3\,. & (6.59)\end{aligned}}$$

Including photons, three species of neutrinos, and the electron-positron pairs, the total value of g is given by

$$g_{\text{tot}} = 2 + \frac{21}{4} + \frac{7}{2} = 10\frac{3}{4}\,. \tag{6.60}$$

This value is appropriate for values of kT which are larger than 0.511 MeV, but smaller than 106 MeV (where muons begin to be produced).

THE ENERGY DENSITY OF RADIATION

In Eq. (6.21) we stated an estimate for the energy density in radiation of the current universe, which we are now prepared to justify. The value can be calculated in terms of the current temperature T_γ of the cosmic microwave background. The best single measurement of T_γ to date was done by the FIRAS (Far InfraRed Absolute Spectrophotometer) instrument on the COBE (Cosmic Background Explorer) satellite, which released its final analysis in 1999,* reporting a value of $T_\gamma = 2.725 \pm 0.002$ K. In 2009 Fixsen† combined the results of all experiments to date to obtain a value 2.7255 ± 0.0006 K.

The radiation that exists in the universe today consists of photons and neutrinos. The energy density is therefore given by Eq. (6.48), using $g = 2$ for the photon contribution,

* J.C. Mather, D.J. Fixsen, R.A. Shafer, C. Mosier, and D.T. Wilkinson, "Calibrator Design for the COBE Far-Infrared Absolute Spectrophotometer (FIRAS)," *Astrophysical Journal*, vol. 512, pp. 511–520 (1999), http://arxiv.org/abs/astro-ph/9810373.

† D.J. Fixsen, "The Temperature of the Cosmic Microwave Background," *Astrophysical Journal*, vol. 707, pp. 916–920 (2009), http://arxiv.org/abs/arXiv:0911.1955.

and $g = 21/4$ for the neutrino contribution, as given by Eq. (6.56). There is a further complication, which you explore in Problem Set 7: the temperature T_ν of the neutrinos is not the same as the temperature T_γ of the photons, but instead

$$T_\nu = \left(\frac{4}{11}\right)^{1/3} T_\gamma \ , \tag{6.61}$$

This temperature differential is established as the e^+e^- pairs disappear from the thermal equilibrium mix, as kT falls below the electron rest energy of 0.511 MeV. The asymmetry results from the fact that the neutrinos interact too weakly to absorb any significant amount of the energy from the e^+e^- pairs, so all the energy goes into heating the photons relative to the neutrinos. Combining the two contributions to the energy density,

$$\begin{aligned} u_{\text{rad},0} &= \left[2 + \frac{21}{4}\left(\frac{4}{11}\right)^{4/3}\right] \frac{\pi^2}{30} \frac{(kT_\gamma)^4}{(\hbar c)^3} \\ &= 7.01 \times 10^{-14} \text{ J/m}^3 \ , \end{aligned} \tag{6.62}$$

in agreement with Eq. (6.21).

NEUTRINO MASSES:

The fact that neutrinos have mass has become known only relatively recently, and we still do not know what the masses are. The status of particle data is tallied by the Particle Data Group at Lawrence Berkeley Laboratory, which can be found on the web at http://pdg.lbl.gov/. In 1996 the Particle Data Group reported that there is “no direct, unconstested evidence for massive neutrinos,” while in 1998 it added that suggestive evidence had been found. In 2000 the evidence was “rather convincing,” and by 2002 the evidence had become “compelling.”

The evidence remains indirect, however. The mass of a neutrino has never been measured, but instead the existence of a nonzero mass is inferred from the fact that we see neutrinos “oscillate” from one species to another. For many years it was a mystery why we do not detect as many neutrinos from the Sun as is expected, but we are now convinced that the deficit is caused by the fact that the electron neutrinos produced in the Sun can oscillate to become muon or tau neutrinos, which are much harder to detect. The muon and tau neutrinos can now be detected by the Sudbury Neutrino Observatory buried 2100 m underground in a mine near Sudbury, Ontario, and by SuperKamiokande, buried 1000 m in a mine at Hida-city, Gifu prefecture, Japan. In addition, starting in 1998, experiments at SuperKamiokande and other locations have found that muon neutrinos produced by cosmic ray collisions in the upper atmosphere can undergo oscillations into other species before reaching the ground. The 2015 Nobel Prize in Physics was awarded to Takaaki Kajita and Arthur McDonald "for the discovery of neutrino oscillations, which shows that neutrinos have mass."

Such oscillations would not be possible if the neutrinos were massless, essentially because a massless particle experiences an infinite time dilation, so time effectively stops. A massless particle in vacuum cannot do anything except travel at the speed of light. The measurements of the oscillations do not allow a determination of the mass, but instead allow one to infer the differences between the squares of the masses. As of 2016, the Particle Data Group reports

$$\begin{aligned} \Delta m_{21}^2\, c^4 &= (7.53 \pm 0.18) \times 10^{-5}\ \mathrm{eV}^2\ , \\ \Delta m_{32}^2\, c^4 &= (2.44 \pm 0.06) \times 10^{-3}\, \mathrm{eV}^2\ , \end{aligned}$$

or

$$\Delta m_{32}^2\, c^4 = (2.51 \pm 0.06) \times 10^{-3}\, \mathrm{eV}^2\ , \tag{6.63}$$

where the two options for Δm_{32}^2 depend on assumptions about the ordering of the masses. The masses are labeled 1, 2, and 3, which are related to the better-known flavor labels ν_e, ν_μ, and ν_τ in a complicated way. The PDG also reports that the rest energy of each type of neutrino is known to be less than 2 eV. The flavor labels ν_e, ν_μ, and ν_τ indicate how the neutrinos are produced, but in the peculiar context of quantum theory these states do not have a well-defined mass. Instead each state of definite mass is a superposition of different flavor states, and vice versa. Although these issues are fascinating, we will not have cause to pursue them any further. If you have not studied quantum theory you will probably have no idea what the last few sentences mean, and that is okay as far as this course is concerned.

Nonetheless, the presence of any mass for the neutrino, no matter how small, raises an important question about the counting of spin states, which is important in our formulas for the black-body radiation of neutrinos. The bottom line will be that the mass makes no difference, but the reasoning is not simple.

We said above that the neutrino has one spin state, because neutrinos are always left-handed: their spin points in the opposite direction from their momentum. If the neutrino were massless, this statement could be precisely true. It can be shown that for massless particles, if the statement is true for one observer, then the spin and the momentum measured by any other observer would align in the same way. Thus, if the neutrino were massless, its left-handedness would be a relativistically invariant property. While it is difficult to prove this invariance, it is easy to see that the invariance fails if the mass of the neutrino is not zero. For definiteness, consider a left-handed neutrino moving along the z axis in the positive direction, so its spin points in the negative z direction. If it has a nonzero mass then it moves slower than the speed of light, so we can always imagine an observer who moves faster, also along the z axis in the positive direction.

To the moving observer the neutrino will be moving in the negative z direction, but the spin will still point along the negative z direction. Hence, the moving observer will see a right-handed particle. But what is this mysterious right-handed particle? Is this a new spin state that must be counted in our calculations of black-body radiation?

We do not yet have a unique theory of neutrino masses, but there are two possibilities. The neutrino might have a *Majorana* mass, in which case the mysterious right-handed particle in the above thought experiment would be an ordinary antineutrino. Since the antineutrino has already been included in the black-body formulas, they will not be changed. The other possibility is that the neutrino can have a *Dirac* mass, which would be the same type of mass that an electron has. In that case, the mysterious right-handed particle in the thought experiment would be a new spin state of the neutrino. The statement that neutrinos are always left-handed would be blatantly false. Nonetheless, our theories would allow us to calculate the strength of the interactions of these right-handed neutrinos, and they would be incredibly weak. They would be so weak that they would essentially never be produced in the early inverse, so again our black-body formulas would not require modification.

THERMAL HISTORY OF THE UNIVERSE:

We now have all the ingredients necessary to calculate the temperature of the universe as a function of time. Eq. (6.47) gives the mass density as a function of time, and Eq. (6.48) relates the energy density to the temperature. Recalling that $u = \rho c^2$, one can combine these relations and solve for the temperature as a function of time:

$$kT = \left(\frac{45\hbar^3 c^5}{16\pi^3 g G}\right)^{1/4} \frac{1}{\sqrt{t}} \ . \tag{6.64}$$

To find the temperature at 1 sec after the big bang, we now need only plug in numbers:

$$\begin{aligned} kT &= \left[\frac{45\left(1.055\times 10^{-34}\text{ J-s}\right)^3\left(2.998\times 10^{8}\text{ m-s}^{-1}\right)^5}{16\pi^3(10.75)\left(6.673\times 10^{-11}\text{ m}^3\text{-kg}^{-1}\text{s}^{-2}\right)}\right]^{1/4} \\ &\qquad \times \frac{1}{(1\text{ s})^{1/2}} \times \left(\frac{1\text{ J}}{\text{kg-m}^2\text{-s}^{-2}}\right)^{1/4} \\ &= 1.378\times 10^{-13}\text{ J} \ , \end{aligned}$$

where the factor $(1\text{ erg/gm-cm}^2\text{-sec}^{-2})^{1/4}$ is equal to 1, and has been inserted to convert the units to the desired form. Using $1\,\text{eV} = 1.602\times 10^{-12}$ erg, one can convert this result if one wishes to

$$kT = 0.860\text{ MeV} \ .$$

Since one knows that $T \propto t^{-1/2}$, one can write down a general expression for the time-temperature relation, for 0.511 MeV $\ll kT \ll$ 106 MeV, as

$$\boxed{kT = \frac{0.860 \text{ MeV}}{\sqrt{t \text{ (in sec)}}}} \ , \tag{6.65a}$$

or equivalently

$$\boxed{T = \frac{9.98 \times 10^9 \text{ K}}{\sqrt{t \text{ (in sec)}}}} \ . \tag{6.65b}$$

As an example one can use Eq. (6.65b) to calculate the temperature of the universe at the end of the first seven days. (Here we are making a minor error, since the value $g_{\text{tot}} = 10\frac{3}{4}$ is not appropriate when kT falls below 0.5 MeV.) One finds $T \approx 1.3 \times 10^7$ K, which is roughly the temperature which is believed to exist in the core of a bright star.

RELATIONSHIP BETWEEN a AND T:

When a gas of black-body radiation expands in thermal equilibrium, there is a simple relationship between the scale factor a and the temperature T. We have already seen that the energy density $\rho \propto 1/a^4$, and that $\rho \propto T^4$. It follows that the product aT remains constant as the universe expands. The constancy of aT is actually a direct consequence of statistical mechanics, and has nothing to do with the dynamics of the expanding universe. As long as the expansion of the universe is slow enough so that the radiation stays in thermal equilibrium, which it is, then the entropy of the expanding gas remains constant. According to Eq. (6.55) the entropy density is proportional to gT^3, so the total entropy S contained in a fixed region in the comoving coordinate system obeys the relation

$$S = sV_{\text{phys}} = sa^3(t)V_{\text{coord}} \propto ga^3T^3 \ , \tag{6.66}$$

where V_{coord} is the coordinate volume of the region. As long as g does not change, then the conservation of entropy implies that aT remains constant. Eq. (6.66) allows us to also understand what happens when g does change, which happens when there is a change in the kinds of particles that contribute to the black-body radiation. For example, when kT falls below 0.5 MeV and the electron-positron pairs disappear from the thermal equilibrium mix, the entropy that had been contained in the electron-positron component of the gas must be given to the other components. However, at this point the neutrinos have decoupled, which means that they are no longer undergoing significant interactions with the rest of the gas. The entropy from the electron-positron pairs is therefore given entirely to the photons, and essentially none is given to the neutrinos. The photons are heated relative to the neutrinos, and they continue to be hotter than the neutrinos into the present era. On Problem Set 7 you will show that this transfer of entropy from the electron-positron pairs to the photons increases the quantity aT_γ, where T_γ is the photon temperature, by a factor of $(11/4)^{1/3} = 1.40$.

RECOMBINATION AND DECOUPLING:

The observed baryonic matter in the universe — the matter made of protons, neutrons, and electrons — is about 80% hydrogen by mass. Most of the rest is helium, with an almost negligible amount of heavier elements. One can use statistical mechanics to understand the behavior of this hydrogen under the conditions prevalent in the early universe, but I will not attempt such a calculation in this course. As one might guess, hydrogen will ionize (*i.e.* break up into separate protons and electrons) if the temperature is high enough. The temperature necessary to cause ionization depends on the density, but for the history of our universe one can say that the hydrogen is ionized when T is greater than about $4,000\,\mathrm{K}$.

Thus, when the temperature falls below $4,000\,\mathrm{K}$, the ionized hydrogen coalesces into neutral atoms. The process is usually called "recombination," although I am at a loss to explain the significance of the prefix "re-". When recombination occurs, the universe becomes essentially transparent to photons. The photons cease to interact with the other particles, and this process is called "decoupling". Decoupling occurs slightly later than recombination, at a temperature of about $3,000\,\mathrm{K}$, since even a small residual density of free electrons is enough to keep the photons coupled to the other particles. The photons which we observe today in the cosmic background radiation are photons which for the most part have last scattered at the time of decoupling.

We can estimate the time of decoupling by using the constancy of aT. Here T indicates the temperature of the photons, since the neutrinos have decoupled and are not relevant to the current discussion. It is very accurate to assume that aT has remained constant from the time of decoupling to the present, since the photons are not interacting significantly with anything else, so the conservation of photon entropy implies that $a^3 s_\gamma \propto a^3 T^3$ is constant. Using the subscript d to denote quantities evaluated at the time of decoupling, and subscript 0 to denote quantities evaluated at the present time, one has

$$a_d T_d = a_0 T_0 \ , \tag{6.67}$$

from which one has immediately that

$$\frac{a_d}{a_0} = \frac{T_0}{T_d} \ . \tag{6.68}$$

Assuming that the universe is flat, and making the crude approximation that it can be treated as matter-dominated from t_d to the present, one has $a(t) \propto t^{2/3}$ and

$$\left(\frac{t_d}{t_0}\right)^{2/3} = \frac{T_0}{T_d} \ . \tag{6.69}$$

Solving, one has

$$\begin{aligned} t_d &= \left(\frac{T_0}{T_d}\right)^{3/2} t_0 \\ &\approx \left(\frac{2.7\,\mathrm{K}}{3000\,\mathrm{K}}\right)^{3/2} \times \left(13.7\times 10^9 \text{ yr}\right) \approx 370,000 \text{ yr} . \end{aligned} \tag{6.70}$$

On p. 159, Ryden quotes a more accurate numerical calculation, giving $t_d \approx 350,000$ yr.

THE SPECTRUM OF THE COSMIC BACKGROUND RADIATION:

The cosmic background radiation was discovered by Penzias and Wilson in 1965. They measured at one frequency only, but found that the radiation appeared to be coming uniformly from all directions in space. This radiation was quickly identified by Dicke, Peebles, Roll, and Wilkinson as the remnant radiation from the big bang. Since then the measurement of the cosmic background radiation has become a minor industry, and much data has been obtained about the spectrum of the radiation and about its angular distribution in the sky.

The prediction from big bang cosmology is that the spectrum should be thermal, corresponding to black-body radiation that has been redshifted from its initially very high temperature. It is a peculiar feature of the black-body spectrum that it maintains its thermal equilibrium form under uniform redshift, even though the photons in the radiation are noninteracting. That is, if each photon in the black-body probability distribution is redshifted by the same factor, the net effect is to produce a new probability distribution which is again of the black-body form, except that the temperature is modified by a factor of the redshift. Thus, the redshift reduces the temperature, but does not lead to departures from the thermal equilibrium spectrum.

The ideal Planck spectrum for such radiation has an energy density $\rho_\nu(\nu)d\nu$, for radiation in the wavelength interval between ν and $\nu + d\nu$, given by

$$\rho_\nu(\nu)d\nu = \frac{16\pi^2\hbar\nu^3}{c^3}\frac{1}{e^{2\pi\hbar\nu/kT}-1}d\nu . \tag{6.71}$$

The subscript ν on ρ_ν indicates that it is the energy density per frequency interval, while one could alternatively speak of the energy density per wavelength interval, ρ_λ. (As with the other statistical mechanics results in this set of Lecture Notes, we will use Eq. (6.71) without derivation.) Observers usually do not directly measure the energy density, however, but instead measure the intensity of the radiation. It can be shown that the power hitting a detector per frequency interval per area of aperature per solid angle of aperture is given by

$$I_\nu(\nu) = \frac{c}{4\pi}\rho_\nu(\nu) = \frac{4\pi\hbar\nu^3}{c^2}\frac{1}{e^{2\pi\hbar\nu/kT} - 1} \; . \tag{6.72}$$

The data on the spectrum available in 1975 is summarized on the two graphs on the following page. The graphs show measurements of the energy density in the cosmic background radiation at different frequencies (or wavelengths). The lower horizontal axis shows the frequency in gigahertz (10^9 cycles per second), and the upper horizontal axis shows the corresponding wavelength. The solid line is the expected blackbody distribution, shown for the best current determination of the temperature, 2.726 K. Part (a) shows the low frequency measurements, including those of Penzias & Wilson and Roll & Wilkinson (which was published about 6 months after the Penzias & Wilson result). Part (b) includes the full range of interesting frequencies. The circles show the results of each measurement, and the bars indicate the range of the estimated uncertainty. The measurements with small uncertainties are shown with dark shading. A high-frequency broad-band measurement is shown on part (b), labeled "1974 Balloon" — the measured energy density is shown as a solid line, and the estimated uncertainty is indicated by gray shading. The 1971 balloon measurements were taken by the MIT team of Dirk Muehlner and Rainer Weiss. (The energy density on both graphs is measured in electron volts per cubic meter per gigahertz.)

The earth's atmosphere poses a serious problem for measuring the high frequency side of the curve, so the best measurements must be done from balloons, rockets, or satellites. In 1987 a rocket probe was launched by a collaboration between the University of California at Berkeley and Nagoya University in Japan. The resulting paper* included a graph of the remarkable data shown in Figure 6.3.

Note that the points labeled 2 and 3 are much higher than the black body spectrum predicts. Using each of these points individually to determine a temperature, the authors find:

$$\text{Point 2: } T = 2.955 \pm 0.017\,\text{K}$$
$$\text{Point 3: } T = 3.175 \pm 0.027\,\text{K}$$

These numbers correspond to discrepancies of 12 and 16 standard deviations, respectively, from the temperature of $T = 2.74\,\text{K}$ that fits the lower frequency points. In terms of energy, the excess intensity seen at high frequencies in this experiment amounts to about

---

* T. Matsumoto, S. Hayakawa, H. Matsuo, H. Murakami, S. Sato, A.E. Lange, and P.L. Richards, "The Submillimeter Spectrum of the Cosmic Background Radiation," *Astrophysical Journal,* vol. 329, pp. 567–571 (1988), http://adsabs.harvard.edu/abs/1988ApJ...329..567M.

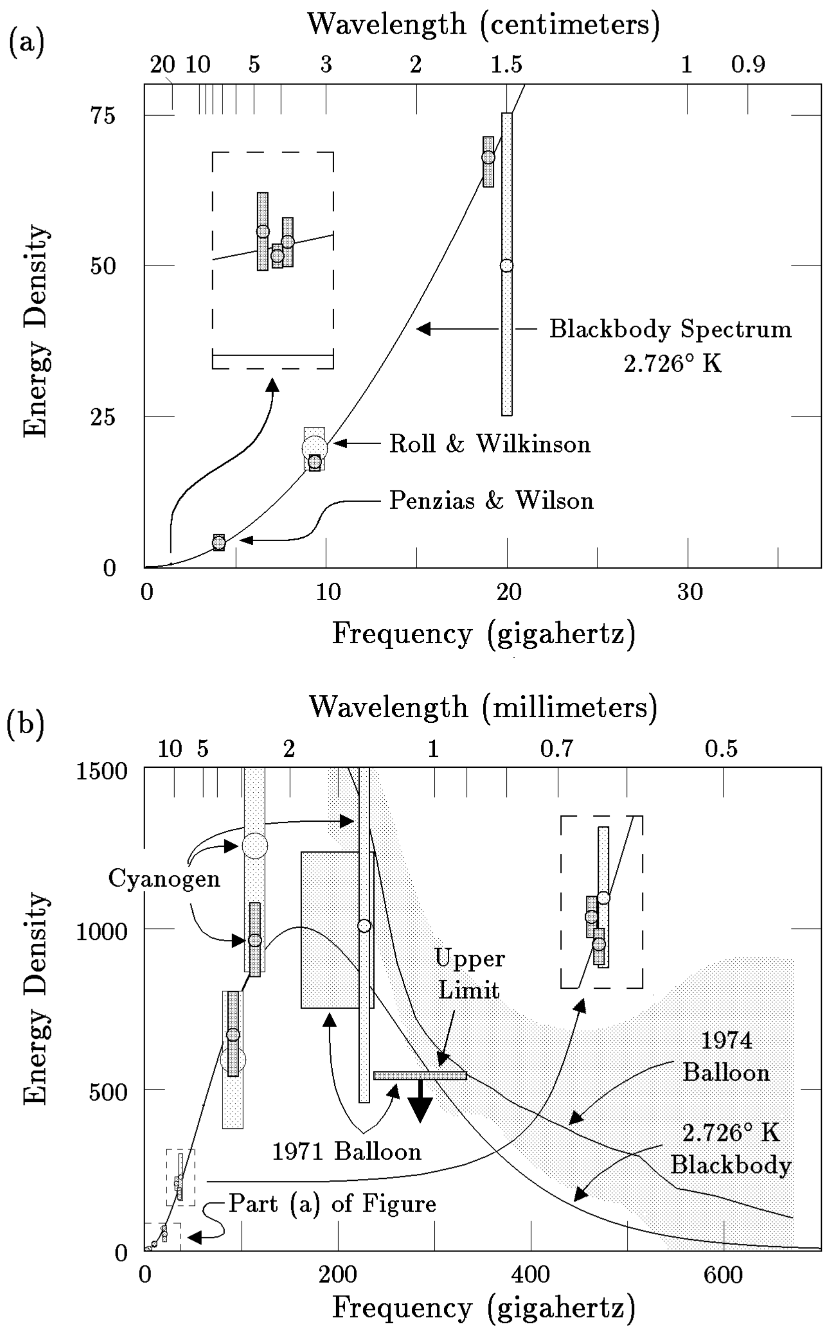


Figure 6.2: The spectrum of the cosmic microwave background as it was known in 1975. Each graph shows the energy density of the radiation, in electron volts per cubic meter per gigahertz, as a function of frequency. Part (a) shows the lowest frequencies, which include the original measurement of Penzias and Wilson, while part (b) includes the full range of interesting frequencies. The curve shows the black-body spectrum for 2.726 K.

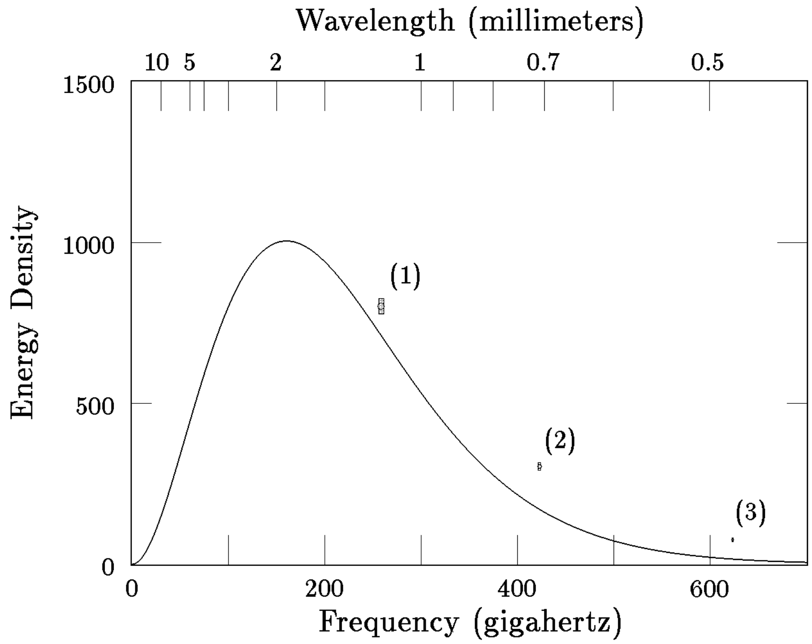


Figure 6.3: Three data points in the CMB spectrum measured by the Berkeley-Nagoya rocket experiment in 1987. Point (3) differs from the theoretically expected curve by 16 standard deviations. The lesson, apparently, is that one should not reject a previously successful theory until the evidence against it is reliably confirmed.

10% of the total energy in the cosmic background radiation. Cosmologists were stunned by the extremely significant disagreement with predictions. Some tried to develop theories to explain the radiation, without much success, while others banked on the theory that it would go away. The experiment looked like a very careful one, however, so it was difficult to dismiss. The most likely source of error in an experiment of this type is the possibility that the detectors were influenced by heat from the exhaust of the launch vehicle — but the experimenters very carefully tracked how the observed radiation varied with time as the detector moved away from the launch rocket, and it seemed clear that the rocket was not a factor.

The same group tried to check their results with a second flight a year later, but the rocket failed and no useful data was obtained.

In the fall of 1989 NASA launched the Cosmic Background Explorer, known as COBE (pronounced "koh-bee"). This marked the first time that a satellite was used to probe the background radiation. Within months, the COBE group announced their first results at a meeting of the American Astronomical Society in Washington, D.C., January 1990. The data was so spectacular that the audience rose to give the speaker, John Mather, a standing ovation. The detailed preprint, with a cover sheet showing a sketch of the satellite, was released the same day, and later published as an *Astrophysical Journal* letter.*

The data showed a perfect fit to the blackbody spectrum, with a temperature of $2.735 \pm 0.06\,\mathrm{K}$, with no evidence whatever for the "submillimeter excess" that had been seen by Matsumoto *et al.* The data was shown with estimated error bars of 1% of the peak intensity, which the group regarded as very conservative. The graph is reproduced here as Fig. 6.5.

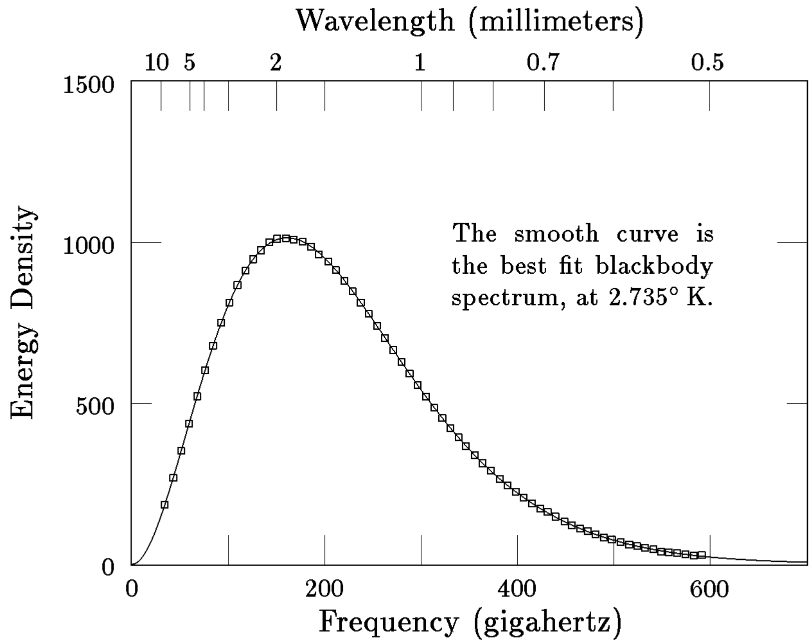


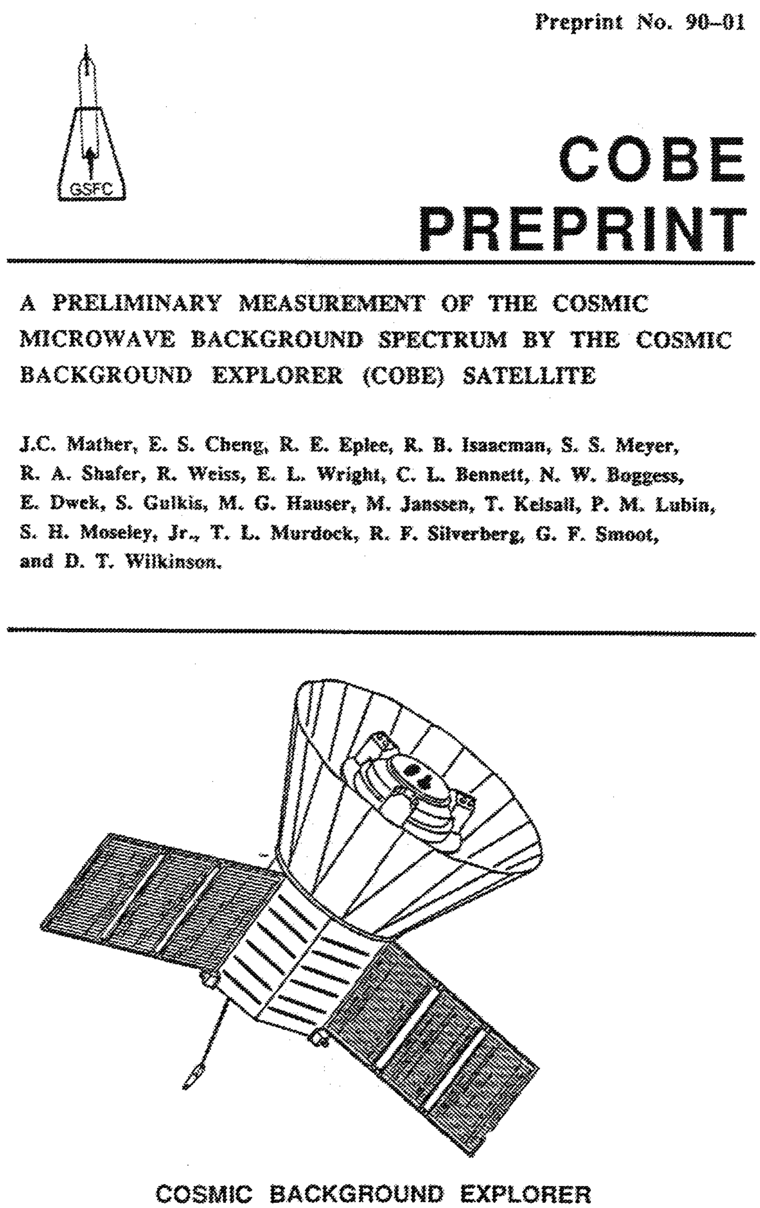
Figure 6.5: The original (1990) COBE measurement of the spectrum of the cosmic microwave background, based on only 9 minutes of data. The vertical axis shows the energy density in units of electron volts per cubic meter per gigahertz.

Once again, the vertical axis is calibrated in electron volts per cubic meter per gigahertz.

Since the COBE instrument is far more precise and has more internal consistency checks, there has been no doubt in the scientific community that the COBE result supercedes the previous one. Despite the 16σ discrepancy of 1988, the cosmic background radiation is now once again believed to have a nearly perfect black-body spectrum.

In January 1993, the COBE team released its final data on the cosmic background radiation spectrum. The first graph had come from just 9 minutes of data, but now the

* J.C. Mather *et al.*, "A preliminary measurement of the cosmic microwave background spectrum by the Cosmic Background Explorer (COBE) satellite," *Astrophysical Journal,* vol. 354, pp. L37–L40 (1990), http://adsabs.harvard.edu/abs/1990ApJ...354L..37M.


Preprint No. 90–01

GSFC

COBE
PREPRINT

A PRELIMINARY MEASUREMENT OF THE COSMIC MICROWAVE BACKGROUND SPECTRUM BY THE COSMIC BACKGROUND EXPLORER (COBE) SATELLITE

J.C. Mather, E. S. Cheng, R. E. Eplee, R. B. Isaacman, S. S. Meyer, R. A. Shafer, R. Weiss, E. L. Wright, C. L. Bennett, N. W. Boggess, E. Dwek, S. Gulkis, M. G. Hauser, M. Janssen, T. Kelsall, P. M. Lubin, S. H. Moseley, Jr., T. L. Murdock, R. F. Silverberg, G. F. Smoot, and D. T. Wilkinson.

COSMIC BACKGROUND EXPLORER

Figure 6.4: The cover page of the original preprint of the COBE cosmic microwave background spectrum measurement.

team had analyzed the data from the entire mission. The error boxes were shrunk beyond visibility to only 0.03%, and the background spectrum was still perfectly blackbody, just as the big bang theory predicted. The new value for the temperature was just a little lower, 2.726 K, with an uncertainty of less than 0.01 K.

The perfection of the spectrum means that the big bang must have been very simple. The COBE team estimated that no more than 0.03% of the energy in the background radiation could have been released anytime after the first year of the life of the universe, since energy released after one year would not have had time to reach such a perfect state of thermal equilibrium. Theories that predict energy release from the decay of turbulent motions or exotic elementary particles, from a generation of exploding or massive stars preceding those already known, or from dozens of other interesting hypothetical objects, were all excluded at once.

Although a few advocates of the steady state universe have not yet given up, the COBE team announced that the theory is ruled out. A nearly perfect blackbody spectrum can be achieved in the steady state theory only by a thick fog of objects that could absorb and re-emit the microwave radiation, allowing the radiation to come to a uniform temperature. Steady state proponents have in the past suggested that interstellar space might be filled by a thin dust of iron whiskers that could create such a fog. However, a fog that is thick enough to explain the new data would be so opaque that distant sources would not be visible.

In this chapter we have discussed mainly the spectrum of the cosmic microwave background (CMB). Starting in 1992, however, with some preliminary results from the COBE satellite, astronomers have also been able to measure the anisotropies of the CMB. This is quite a tour de force, since the radiation is isotropic to an accuracy of about 1 part in 10^5. Since the photons of the CMB have been travelling essentially on straight lines since the time of decoupling, these anisotropies are interpreted as a direct measure of the degree of nonuniformity of the matter in the universe at the time of decoupling, about 380,000 years after the big bang. These non-uniformities are crucially important, because they give us clues about how the universe originated, and because they are believed to be the seeds which led to the formation of the complicated structure that the universe has today. We will return to discuss the physics of these nonuniformities near the end of the course.

The Nobel Prize in Physics 2006 was awarded jointly to John C. Mather and George F. Smoot "for their discovery of the blackbody form and anisotropy of the cosmic microwave background radiation."

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe November 13, 2018
Prof. Alan Guth

Lecture Notes 7
THE COSMOLOGICAL CONSTANT

INTRODUCTION:

Much excitement has been generated since January 1998 over observations that show that the expansion of the universe today is accelerating, rather than decelerating. Two groups of astronomers,* with a total of 52 astronomers in the two groups, have reported evidence for such an acceleration, based on observations of distant ($z \lesssim 1$) Type Ia supernova explosions, which are used as standard candles. (Note that "Ia" is pronounced "one-A," not "eeya.") The first announcement was made at the AAS meeting in January of 1998, leading to news articles in *Science* on January 30 and February 27, 1998, and in *The New York Times* on March 3, March 8, April 21, and May 5, 1998. On May 15 one of the two groups (the The High Z Supernova Search Team) posted a paper on the web titled "Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant".† The other group (The Supernova Cosmology Project) submitted its findings to the web on December 8, 1998.‡ *Science* magazine officially proclaimed this to be the "Breakthrough of the Year" for 1998. In 2011, these discoveries were recognized with the awarding of the Nobel Prize in Physics to Saul Perlmutter, Brian Schmidt, and Adam Riess, and in 2015 the Fundamental Physics Prize was awarded to the same three group leaders, and also to the two entire teams.

The evidence for a cosmological constant has stood up firmly for the twenty years since 1998, and in fact it has gotten significantly stronger. Many cosmologists including me were skeptical in 1998, but now essentially all of us are convinced that the expansion of the universe is accelerating. The simplest explanation is that the universe has a nonzero cosmological constant. An alternative explanation is something called *quintessence,* which has very nearly the same effect. (Quintessence refers to a slowly evolving scalar field that

---

* One group is the Supernova Cosmology Project, based at Lawrence Berkeley Laboratory and headed by Saul Perlmutter. Their web page is

http://www-supernova.lbl.gov/

The other group is the The High Z Supernova Search Team, led by Brian Schmidt, with web page

http://www.cfa.harvard.edu/supernova//HighZ.html.

† http://arXiv.org/abs/astro-ph/9805201, later published as Riess *et al.*, *Astronomical Journal* **116**, No. 3, 1009 (1998).

‡ "Measurements of Ω and Λ from 42 High-Redshift Supernovae," http://arXiv.org/abs/astro-ph/9812133, later published as Perlmutter *et al.*, *Astrophysical Journal* **517**:565–586 (1999).

permeates the universe and fills it with a nearly uniform energy density — we'll get back to that idea when we talk about inflation near the end of the course.) Since no one is sure what exactly is driving this acceleration, the term "dark energy" has been invented to describe the stuff that is driving the acceleration, whatever it might be. A cosmological constant is the simplest explanation, and that is what will be discussed in this set of lecture notes.

BACKGROUND:

The cosmological constant was first proposed by Albert Einstein in 1917, when he was trying for the first time to apply his newly invented theory of general relativity to the universe as a whole.* At the time he believed that the universe was static, since it appeared static and there was no evidence to the contrary. However, when he worked out the consequences of his theory, he discovered the result that we found in Lecture Notes 6, Eq. (6.38):

$$\frac{d^2a}{dt^2} = -\frac{4\pi}{3} G \left(\rho + \frac{3p}{c^2} \right) a \ , \tag{7.1}$$

where a is the scale factor, t is time, G is Newton's gravitational constant, ρ is the mass density, p is the pressure, and c is the speed of light. Taking $\rho > 0$ and $p \approx 0$, Einstein was forced to the conclusion that $d^2a/dt^2 < 0$, so a static ($a = \textit{constant}$) solution did not exist. The problem, essentially, was that gravity is an attractive force, so an initially static universe would collapse.

Einstein's solution was to modify what we call the Einstein field equations — the equations that describe how gravitational fields, in the form of spacetime curvature, are created by matter. He called the new term the cosmological term, because it was motivated by the cosmological argument that it was needed to allow a static universe. The cosmological term could create a repulsive force that could be adjusted in strength so that it could have just the right value to prevent the universe from collapsing. The coefficient of this term was called the cosmological constant and assigned the symbol Λ (capital Greek lambda).

Einstein's static model seemed viable for about a decade, but during the 1920s astronomers discovered that the universe was not static after all. In 1929 Edwin Hubble published his famous paper announcing what we now know as Hubble's law. Einstein was quick to accept Hubble's findings, and discarded his cosmological term as unwarranted.

---

* "Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie" ("Cosmological Considerations on the General Theory of Relativity,") by A. Einstein, *Sitzungsberichte der Preussichen Akad. d. Wissenschaften*, pp. 142–152, 1917. An English translation is available in **The Principle of Relativity,** translated by W. Perrett and G.B. Jeffery, Dover Publications, 1952, and also in **Cosmological Constants,** edited by Jeremy Bernstein and Gerald Feinberg, Columbia University Press, 1986.

COSMOLOGICAL EQUATIONS WITH A COSMOLOGICAL CONSTANT:

Although Einstein did not look at the cosmological constant this way, from a modern perspective the cosmological constant is interpreted as an energy density attributed to the vacuum. That is, the cosmological term in the Einstein field equations is identical to the term that would be added to describe the effect of a nonzero vacuum energy density. Since everything that we see can be described as particles moving through the vacuum, the vacuum energy density becomes a uniform contribution to the total energy, at all points in space and at any time. The relation between Einstein's original symbol Λ and the vacuum energy density $u_{\rm vac}$, or the vacuum mass density $\rho_{\rm vac}$, is given by

$$u_{\rm vac} = \rho_{\rm vac} c^2 = \frac{\Lambda c^4}{8\pi G} \ . \tag{7.2}$$

Einstein's constant Λ has the units of (length)$^{-2}$, while $u_{\rm vac}$ and $\rho_{\rm vac}$ of course have the usual units for energy density and mass density. The pressure that corresponds to this vacuum energy can be obtained by applying the equation of energy conservation, using the fact that the energy density of the vacuum is fixed. On Problem 4 of Problem Set 6 we learned that conservation of energy in a Robertson-Walker universe takes the form

$$\frac{d}{dt}\left(a^3 \rho c^2\right) = -p \frac{d}{dt}(a^3) \ , \tag{7.3}$$

or equivalently

$$\dot\rho = -3\frac{\dot a}{a}\left(\rho + \frac{p}{c^2}\right) \ , \tag{7.4}$$

where the overdot denotes a time derivative. Setting $\dot\rho_{\rm vac} = 0$ gives

$$p_{\rm vac} = -\rho_{\rm vac} c^2 = -\frac{\Lambda c^4}{8\pi G} \ . \tag{7.5}$$

The relation between the pressure and energy density is the same as the relation that we will later discuss for the false vacuum that is responsible for driving the accelerated expansion of the inflationary universe model. From Eq. (7.1), one can see that a negative pressure can drive an acceleration. We must add the contributions of the vacuum energy density and pressure to the right-hand side, so for clarity we will use the symbols ρ_n and p_n to denote the mass density and pressure of normal matter, where *normal* refers to all forms of energy other than the cosmological constant. One then has

$$\begin{aligned} \frac{d^2 a}{dt^2} &= -\frac{4\pi}{3} G \left(\rho_n + \frac{3p_n}{c^2} + \rho_{\rm vac} + \frac{3p_{\rm vac}}{c^2}\right) a \\ &= -\frac{4\pi}{3} G \left(\rho_n + \frac{3p_n}{c^2} - 2\rho_{\rm vac}\right) a \ , \end{aligned} \tag{7.6}$$

where we used Eq. (7.5) to eliminate $p_{\rm vac}$.

We learned in Lecture Notes 6 that the first order Friedmann equation is not modified by pressure, so it is still written as it was first written, as Eq. (3.31):

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2} \ . \tag{7.7}$$

Since the right-hand side depends only on ρ, we find the contribution of the vacuum energy density by replacing ρ by $\rho_n + \rho_{\rm vac}$:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G(\rho_n + \rho_{\rm vac}) - \frac{kc^2}{a^2} \ . \tag{7.8}$$

Using Eqs. (7.6) and (7.8), we can be more precise about what it means to live in an accelerating universe. From Eq. (7.6), we see that $\ddot{a}$ can be positive if the $\rho_{\rm vac}$ term is positive and dominates the right-hand side, so under these circumstances one says that universe accelerates, meaning that the function $a(t)$ accelerates. Since the physical distance ℓ_p to a galaxy at coordinate distance ℓ_c is given by

$$\ell_p(t) = a(t)\ell_c \ ,$$

we see that in an accelerating universe, the relative velocity between galaxies increases with time.

On the other hand, from Eq. (7.8) we see that this acceleration does not necessarily mean that H increases with time. We can more easily see the behavior of Eq. (7.8) if we replace ρ_n by $\rho_m + \rho_{\rm rad}$, where ρ_m is the mass density of nonrelativistic (pressureless) matter and $\rho_{\rm rad}$ is the mass density of radiation:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\,(\underbrace{\rho_m}_{\propto \frac{1}{a^3(t)}} + \underbrace{\rho_{\rm rad}}_{\propto \frac{1}{a^4(t)}} + \rho_{\rm vac}) - \frac{kc^2}{a^2} \ . \tag{7.9}$$

For an open ($k < 0$) or flat ($k = 0$) universe, the right-hand side of Eq. (7.9) contains only positive terms, each of which decreases as the universe expands. Thus H decreases monotonically in such a universe, even if the universe is accelerating. The matter, radiation, and curvature terms all approach zero as $a \to \infty$, so asymptotically

$$H = \frac{\dot{a}}{a} \xrightarrow[a\to\infty]{} \sqrt{\frac{8\pi}{3}G\rho_{\rm vac}} \qquad \text{from above.} \tag{7.10}$$

Note that $H = \dot{a}/a$ can decrease even when $\dot{a}$ is increasing, as long as a is increasing faster. For a closed universe ($k > 0$) it is possible for H to increase as the universe expands, but this happens only if the last term of Eq. (7.9) is large enough in magnitude so that it dominates the rate of change of H. Our universe could be closed, but the last term of Eq. (7.9) is known to be small, so H for our universe is certainly decreasing. The recession velocity of any distant galaxy is accelerating, but $H = v/\ell_p$ can still decrease with time if ℓ_p increases faster than v does. If it is true that the acceleration is caused by vacuum energy density, then Eq. (7.10) describes the asymptotic future of our universe, whether it is open, closed, or precisely flat. However, we should certainly keep in mind that predictions about the infinite future are very dicey. It is possible, for example, that the state that we call the vacuum might not really be stable, but might instead decay into a lower energy state after 10^{1000} years, falsifying our prediction.

THE COSMOLOGICAL CONSTANT AND THE AGE OF THE UNIVERSE:

One effect of a positive cosmological constant is an increase in the age of the universe that is inferred from a given value of the Hubble constant. This effect can be understood qualitatively by remembering that the cosmological constant causes the universe to accelerate. Suppose, then, that we calculated the age of the universe as we learned in Lecture Notes 4, assuming that there was no cosmological constant. Then suppose that we add a vacuum energy term, keeping fixed the current value of the Hubble expansion rate H_0 and the current mass density of nonrelativistic matter and radiation. The new energy contribution adds a positive term to $\ddot{a}$, which means that H has not been falling as fast as it had in the previous $\rho_{\rm vac} = 0$ calculation. Then, as can be seen from the following sketch,

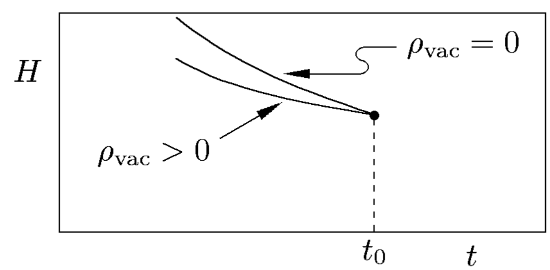


Figure 7.1: Sketch of the Hubble expansion rate H vs. time t, illustrating the difference between a model with and without vacuum energy density.

the Hubble expansion rate in the past would be lower in the new calculation than it was in the first calculation. The slower decrease in H would mean that it takes longer for H to reach its present value, since in both models H starts at infinity at the instant of the big bang. Similarly, the lower value of H in the past would mean that it takes longer for the universe to reach its present mass density. Thus, the new calculation implies a universe which is older than we had calculated in the absence of a cosmological constant.

Quantitatively, we can calculate the age of the universe from Eq. (7.9). To be completely explicit about the time-dependence of each term, we write

$$\begin{aligned}
\rho_m(t) &= \left[\frac{a(t_0)}{a(t)}\right]^3 \rho_{m,0} \\
\rho_{\rm rad}(t) &= \left[\frac{a(t_0)}{a(t)}\right]^4 \rho_{{\rm rad},0} \\
\rho_{\rm vac}(t) &= \rho_{{\rm vac},0} \ .
\end{aligned} \tag{7.11}$$

Here we are using the convention that a subscript 0 denotes the present value of any quantity, so for example $\rho_{m,0}$ denotes the present value of the mass density of nonrelativistic matter. Each of the above equations reflects the known dependence on $a(t)$ for each contribution to the mass density, with the constant of proportionality written so that $\rho_X(t_0) = \rho_{X,0}$, for each type of matter X. Mass densities are usually tabulated as fractions of the critical density,

$$\rho_c = \frac{3H^2}{8\pi G} \ , \tag{7.12}$$

using the convention that for each type of mass density X,

$$\Omega_X \equiv \rho_X/\rho_c \ . \tag{7.13}$$

So, we rewrite Eqs. (7.11) by replacing each $\rho_{X,0}$ by $\Omega_{X,0}\rho_{c,0}$:

$$\begin{aligned}
\rho_m(t) &= \frac{3H_0^2}{8\pi G}\left[\frac{a(t_0)}{a(t)}\right]^3 \Omega_{m,0} \\
\rho_{\rm rad}(t) &= \frac{3H_0^2}{8\pi G}\left[\frac{a(t_0)}{a(t)}\right]^4 \Omega_{{\rm rad},0} \\
\rho_{\rm vac}(t) &= \frac{3H_0^2}{8\pi G}\Omega_{{\rm vac},0} \ .
\end{aligned} \tag{7.14}$$

Defining

$$x \equiv \frac{a(t)}{a(t_0)} \ , \tag{7.15}$$

so that x varies from 0 to 1 as the universe evolves from the big bang to the present, Eq. (7.9) can be rewritten as

$$\left(\frac{\dot a}{a}\right)^2 = H_0^2\left(\frac{\Omega_{m,0}}{x^3} + \frac{\Omega_{{\rm rad},0}}{x^4} + \Omega_{\rm vac}\right) - \frac{kc^2}{a^2} \ . \tag{7.16}$$

It is convenient to rewrite the curvature term in the same form as the other terms, by defining

$$\Omega_{k,0} \equiv -\frac{kc^2}{a^2(t_0)H_0^2} \ , \tag{7.17}$$

so

$$\begin{aligned} H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \left(\frac{\dot{x}}{x}\right)^2 &= H_0^2\left(\frac{\Omega_{m,0}}{x^3} + \frac{\Omega_{\text{rad},0}}{x^4} + \Omega_{\text{vac}} + \frac{\Omega_{k,0}}{x^2}\right) \\ &= \frac{H_0^2}{x^4}\left(\Omega_{m,0}x + \Omega_{\text{rad},0} + \Omega_{\text{vac},0}x^4 + \Omega_{k,0}x^2\right) \ . \end{aligned} \tag{7.18}$$

By specializing this formula to $t = t_0$, for which $x = 1$, one finds $1 = \Omega_{m,0} + \Omega_{\text{rad},0} + \Omega_{\text{vac},0} + \Omega_{k,0}$, so

$$\Omega_{k,0} = 1 - \Omega_{m,0} - \Omega_{\text{rad},0} - \Omega_{\text{vac},0} \ . \tag{7.19}$$

$\Omega_k > 0$ for an open universe, $\Omega_k < 0$ for a closed universe, and $\Omega_k = 0$ for a flat universe. The present age of the universe can then be found by taking the square root of Eq. (7.18),

$$\frac{\dot{x}}{x} = \frac{H_0}{x^2}\sqrt{\Omega_{m,0}x + \Omega_{\text{rad},0} + \Omega_{\text{vac},0}x^4 + \Omega_{k,0}x^2} \ , \tag{7.20}$$

or

$$x\frac{dx}{dt} = H_0\sqrt{\Omega_{m,0}x + \Omega_{\text{rad},0} + \Omega_{\text{vac},0}x^4 + \Omega_{k,0}x^2} \ . \tag{7.21}$$

This equation can be rearranged as

$$dt = \frac{1}{H_0}\frac{x\,dx}{\sqrt{\Omega_{m,0}x + \Omega_{\text{rad},0} + \Omega_{\text{vac},0}x^4 + \Omega_{k,0}x^2}} \ , \tag{7.22}$$

which can be integrated over the range of x from the big bang to the present to give

$$t_0 = \frac{1}{H_0}\int_0^1 \frac{x\,dx}{\sqrt{\Omega_{m,0}x + \Omega_{\text{rad},0} + \Omega_{\text{vac},0}x^4 + \Omega_{k,0}x^2}} \ . \tag{7.23a}$$

The above form is probably the easiest to integrate, but for some purposes it is useful to rewrite it by changing variables of integration to z, where

$$1 + z = \frac{a(t_0)}{a(t)} = \frac{1}{x} \ .$$

The integral then becomes

$$t_0 = \frac{1}{H_0}\int_0^\infty \frac{dz}{(1+z)\sqrt{\Omega_{m,0}(1+z)^3 + \Omega_{\text{rad},0}(1+z)^4 + \Omega_{\text{vac},0} + \Omega_{k,0}(1+z)^2}} \ . \tag{7.23b}$$

In this form one could also find the "look-back time" to any particular redshift z by stopping the integration at that point:

$$t_{\text{look-back}}(z) = \frac{1}{H_0}\int_0^z \frac{dz'}{(1+z')\sqrt{\Omega_{m,0}(1+z')^3+\Omega_{\text{rad},0}(1+z')^4+\Omega_{\text{vac},0}+\Omega_{k,0}(1+z')^2}} \; . \tag{7.24}$$

The look-back time is defined as the time interval between the era that we observe at redshift z and the present.

The general case of the integrals in Eqs. (7.23) and (7.24) can be computed only by numerical integration, but various special cases can be carried out analytically. The case of a matter-dominated universe ($\Omega_{\text{rad}} = \Omega_{\text{vac}} = 0$) was done in Lecture Notes 5. The case of a flat universe composed of nonrelativistic matter and vacuum energy (i.e., $\Omega_{\text{rad}} = \Omega_k = 0$, $\Omega_m + \Omega_{\text{vac}} = 1$) can also be integrated analytically, yielding

$$t_0 = \begin{cases} \dfrac{2}{3H_0}\dfrac{\tan^{-1}\sqrt{\Omega_{m,0}-1}}{\sqrt{\Omega_{m,0}-1}} & \text{if } \Omega_{m,0} > 1,\ \Omega_{\text{vac}} < 0 \\ \dfrac{2}{3H_0} & \text{if } \Omega_{m,0} = 1,\ \Omega_{\text{vac}} = 0 \\ \dfrac{2}{3H_0}\dfrac{\tanh^{-1}\sqrt{1-\Omega_{m,0}}}{\sqrt{1-\Omega_{m,0}}} & \text{if } \Omega_{m,0} < 1,\ \Omega_{\text{vac}} > 0 \; . \end{cases} \tag{7.25}$$

Note that inverse hyperbolic tangents can also be expressed in terms of logarithms, so the answer for the $\Omega_{m,0} < 1$ case can also be written as

$$t_0 = \frac{2}{3H_0}\frac{\ln\left(\sqrt{1-\Omega_{m,0}}+1\right) - \ln\sqrt{\Omega_{m,0}}}{\sqrt{1-\Omega_{m,0}}} \; . \tag{7.26}$$

Although Eq. (7.25) expresses t_0 in terms of three different expressions, the function is actually continuous, so the value for $\Omega_{m,0} = 1$ can be obtained as the limit of either the expression for $\Omega_{m,0} > 1$ or the expression for $\Omega_{m,0} < 1$.

A graph of the age of the universe as a function of the Hubble constant, for a matter-dominated universe **without** a cosmological constant, is given by Eq. (4.47) and is shown in Fig. 7.2. For the parameter choice of $H_0 = 67.7$ km·s^{-1}·Mpc^{-1} (the Planck 2018 value) and $\Omega = 1$, this gives $t_0 = 9.62 \times 10^9$ yr, which is significantly younger than the 11.2 billion year mininum age determined by Krauss and Chaboyer, based on age of the oldest stars, as discussed on pp. 2-3 of Lecture Notes 4.

However, this discrepancy of age estimates goes away when one attributes approximately 70% of Ω to a cosmological constant. A graph of the age of the universe, for

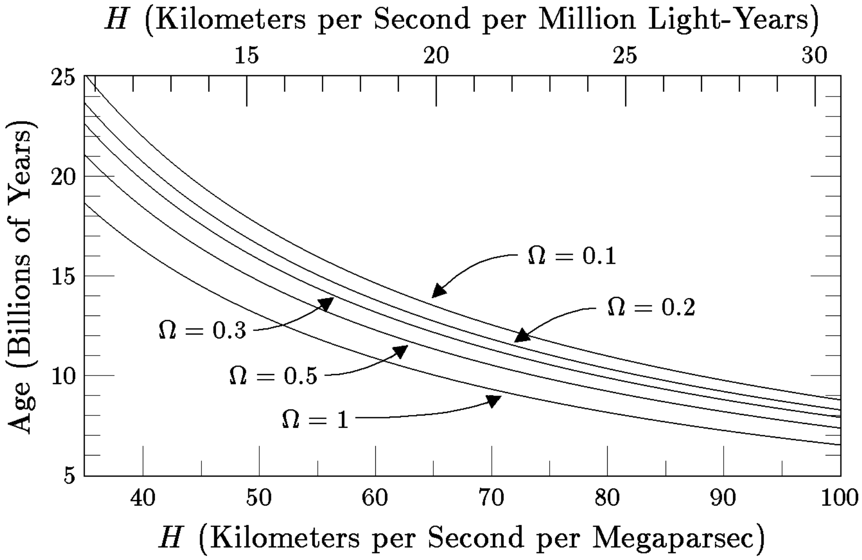


Figure 7.2: The age of an open ($\Omega < 1$), closed ($\Omega > 1$), or flat ($\Omega = 1$) universe containing only nonrelativistic matter.

a flat universe composed of nonrelativistic matter and a cosmological constant, is given by Eq. (7.25) and is shown as Fig. 7.3. Note that Ω_m refers to the mass density of nonrelativistic matter only. For all the model universes shown on this graph, the total Ω (including nonrelativistic matter and vacuum mass density) is one, which is in accord with the predictions of the simplest inflationary models (which will be discussed at the end of the term).

The graph also shows two data points: the point labeled RB refers to the Ryden Benchmark Model (from Barbara Ryden, *Introduction to Cosmology*), and the point labeled Planck2018 is the best fit model to the Planck 2018 data set, combined with a number of other cosmological measurements.* The parameters associated with these two models are shown in Table 7.1.

The Planck 2018 data best fit is generally regarded as the most reliable estimate of cosmological parmeters that we currently have. Using the parameters from this table, Eq. (7.25) gives a current age t_0 of 13.5 billion years for the Ryden Benchmark model,

---

* N. Aghanim et al. (Planck Collaboration), "Planck 2018 results, VI: Cosmological parameters," Table 2, Column 6, arXiv:1807.06209.

* The Planck paper does not give values for Ω_b or $\Omega_{\rm dm}$, but instead gives values for $\Omega_b h^2$ and $\Omega_{\rm dm} h^2$, where $h = H_0/(100\ \text{km}\cdot\text{s}^{-1}\cdot\text{Mpc}^{-1})$. The values shown here were computed from the values for $\Omega_b h^2$, $\Omega_{\rm dm} h^2$, and h, assuming that the uncertainties are uncorrelated.

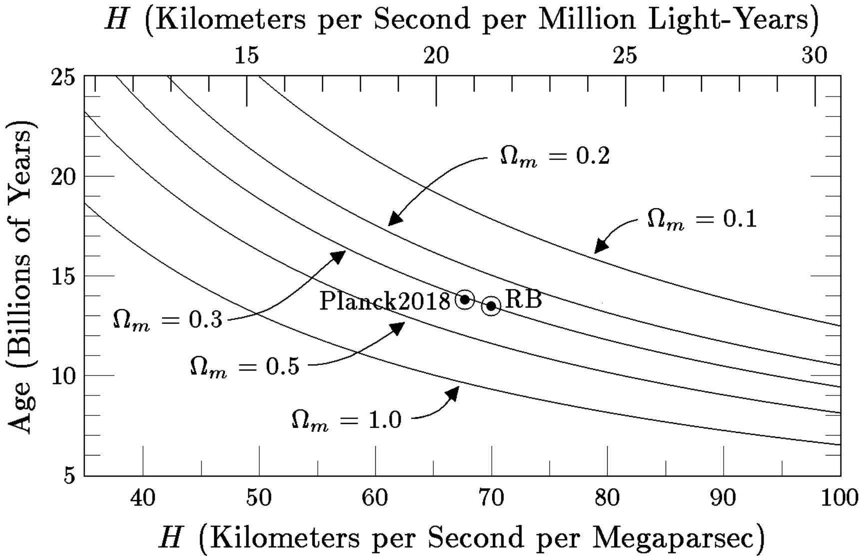


Figure 7.3: The age of a flat universe containing nonrelativistic matter and vacuum energy. The dots indicate the Ryden Benchmark Model and the Planck 2018 best fit.

Parameters	Ryden Benchmark	Planck 2018 Best Fit
$\boldsymbol{H_0}$	70	67.7 ± 0.4 km·s^{-1}·Mpc^{-1}
Baryonic matter $\boldsymbol{\Omega_b}$	0.04	$0.0490 \pm 0.0004^*$
Dark matter $\boldsymbol{\Omega_{\rm dm}}$	0.26	$0.261 \pm 0.003^*$
Total matter $\boldsymbol{\Omega_m}$	0.30	0.311 ± 0.006
Vacuum energy $\boldsymbol{\Omega_{\rm vac}}$	0.70	0.689 ± 0.006

Table 7.1: Cosmological Parameters.

and 13.80 billion years for the Planck 2018 best fit model. The Aghanim et al. Planck "Cosmological Parameters" paper cited above gives a best fit value for the age of the universe of 13.79±0.02 billion years, where the quoted uncertainty of 0.15% is considerably smaller than would be obtained by compounding the uncertainties of the parameters shown in the table: 0.6% for H_0 and 1.8% for Ω_m. Thus, the Planck group is asserting that the uncertainty in H_0 and Ω_m are correlated in just the right way so that they can determine the age of the universe with much greater precision than they can determine either of the input parameters.

(To get some feeling for the stability of these numbers, we can compare with an earlier WMAP data set, the 3-year data.* In that paper the numbers were reported as $H_0 = 73.5 \pm 3.2\ \text{km}{\cdot}\text{s}^{-1}{\cdot}\text{Mpc}^{-1}$, $\Omega_b = 0.041$, $\Omega_{\text{dm}} = 0.196$, $\Omega_m = 0.237 \pm 0.034$, $\Omega_{\text{vac}} = 0.763 \pm 0.034$, and $t_0 = 13.73^{+0.16}_{-0.15}$ billion years.)

We will discuss the physics underlying the Planck and WMAP anisotropy measurements near the end of the term, but for now it is worth mentioning that Planck refers to a European Space Agency satellite experiment that was launched in May 2009 to measure the anisotropies (i.e., nonuniformities) of the cosmic microwave background radiation. WMAP refers to the Wilkinson Microwave Anisotropy Probe, an earlier satellite launched in June 2001, for the same purpose. While the CMB is uniform in all directions to an accuracy of a few parts in 100,000, the nonuniformities can nonetheless be measured to a high degree of accuracy, providing important information about the early universe. Planck and WMAP are still in orbit, but WMAP stopped taking data in August 2010, and Planck stopped taking data in October 2013. WMAP released its first-year data set in February 2003, and later released three-year (March 2006), five-year (March 2008), seven-year (January 2010), and a final nine-year (December 2012) data set. The Planck experiment has released three data sets, in March 2013, in February 2015, and in July 2018. Both WMAP and Planck carried out their observations from a unique location, called the L2 Lagrange Point. L2 is located at a position approximately 1.5 million kilometers from Earth, in a direction opposite to the Sun. It follows the orbit of the Earth around the Sun once per year, always maintaining its position along a radial line drawn from the Sun through the Earth. L2 is an ideal location for astronomy, because the satellite can look outward away from the Sun, so at any time it can view half of the sky with no interference from the Sun, Earth, or Moon. Over the course of one year, the entire sky can be viewed under these ideal conditions.

It may seem strange that a satellite measuring the anisotropies of the CMB can give us values for parameters such as H_0 and Ω_m, but such parameters can be extracted if one has a theoretical prediction for the anisotropies that depends on these parameters. In fact such a theoretical model does exist, and it fits the data extraordinarily well. We will come back to this topic at the very end of the course.

One can include $\Omega_{\text{rad},0}$ in the age calculation of Eq. (7.23) by doing the integral numerically, and using $\Omega_r = 9.2 \times 10^{-5}$, which we found from Eq. (6.23). If one includes it with the Planck parameters, adjusting $\Omega_{\text{vac},0}$ to keep the universe exactly flat, one finds that the age estimate is decreased by 5.7 million years, which is just beyond the level of accuracy of the calculation. So for this calculation, the effect of radiation in the universe is negligible.

---

* D.N. Spergel et al., "Wilkinson Microwave Anisotropy Probe (WMAP) Three Year Observations: Implications for Cosmology," Table 5, 'WMAP Only' column, *Ap. J. Supp.* **170,** 377 (2007), http://arxiv.org/abs/astro-ph/0603449v2.

THE HUBBLE DIAGRAM — RADIATION FLUX VERSUS REDSHIFT:

The claims that the cosmological constant is nonzero are based on the Hubble diagram, the graph which shows the measurements of the radiation flux of sources as a function of their redshift z. To understand how the cosmological constant affects this diagram, we need to derive the formula for the received radiation flux of a specified source, in a model universe which includes a cosmological constant. In principle we need to consider closed, flat, and open universes, but I will show the calculation in detail only for the case of a closed universe. The open-universe case is very similar, so I will merely describe the differences and state the answer for this case. The flat universe is the borderline case between open and closed, so it can be treated as a limiting case of either open or closed universes, or it can be done as a separate calculation.

The Robertson-Walker metric for a closed universe was given as Eq. (5.34):

$$ds^2 = -c^2\,dt^2 + a^2(t)\left\{\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + \sin^2\theta\,d\phi^2\right)\right\} \; . \tag{5.34}$$

The cosmological constant will affect the evolution of $a(t)$, but the form of the metric was determined by the symmetries of homogeneity and isotropy, and will not be changed.

We will be interested in tracing the trajectories of photons traveling along radial lines, so for this purpose it will be useful to introduce the radial coordinate ψ, defined by

$$\sin\psi \equiv \sqrt{k}\,r \; .$$

One finds

$$d\psi = \frac{\sqrt{k}\,dr}{\cos\psi} = \frac{\sqrt{k}\,dr}{\sqrt{1-kr^2}} \; .$$

The metric then simplifies to

$$ds^2 = -c^2\,dt^2 + \tilde{a}^2(t)\left\{d\psi^2 + \sin^2\psi\left(d\theta^2 + \sin^2\theta\,d\phi^2\right)\right\} \; , \tag{7.27}$$

where the new scale factor $\tilde{a}(t)$ is related to the scale factor $a(t)$ by

$$\tilde{a}(t) \equiv \frac{a(t)}{\sqrt{k}} \; . \tag{7.28}$$

This form of the metric is useful for investigating radial motion, because the radial part of the metric is very simple. (You might recall that the closed universe metric was constructed in Lecture Notes 6 by first considering a sphere in 4 Euclidean dimensions. The coordinate ψ defined here is precisely the same as the angle ψ that was used in that

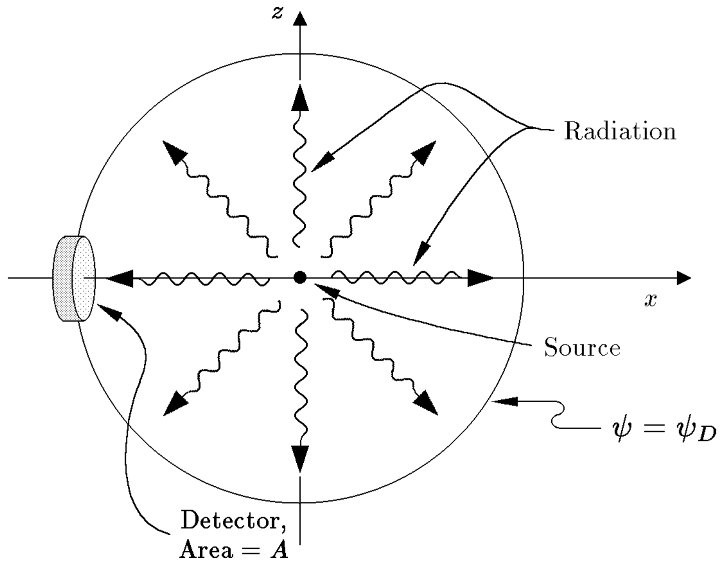


Figure 7.4: Diagram showing how the power of a source is uniformly spread over a sphere that includes the detector on Earth used to measure the energy flux.

construction — it is the angle between the w-axis and a line joining the origin of the 4-dimensional coordinate system to the point in question.)

Fig. 7.4 is a diagram showing how radiation from a distant source reaches a detector on Earth. The diagram shows a comoving coordinate system with the source at the origin, $\psi = 0$. The radial coordinate of the detector, on Earth, is called ψ_D. The diagram also shows a sphere at the same radial coordinate, ψ_D. We assume that the source is spherically symmetric, so that the power emitted by the source is uniformly spread over this sphere. Since the speed of light is independent of angle, all the photons that left the source at some particular time t_S are arriving at the $\psi = \psi_D$ sphere at the present time t_0. To calculate the power received by the detector, we need to know what fraction of those photons hit the detector. The fraction is simply the area of the detector divided by the area of the sphere, or

$$\text{fraction} = \frac{\text{area of detector}}{\text{area of sphere}} = \frac{A}{4\pi\tilde{a}^2(t_0)\sin^2\psi_D} \ .$$

(The area of the sphere at radial coordinate ψ_D is given by $4\pi\tilde{a}^2(t_0)\sin^2\psi_D$, because the part of the metric (7.27) that depends on $d\theta$ and $d\phi$ is equal to $\tilde{a}^2(t)\sin^2\psi$ times the metric for a sphere of unit radius.) The power hitting the detector is further reduced by one factor of

$$1 + z_S = \frac{a(t_0)}{a(t_S)} \ , \tag{7.29}$$

because the frequency, and hence the energy, of each photon is reduced by this factor. In addition, the power is reduced by another factor of $(1+z_S)$ because the rate of arrival of photons is reduced by this factor. Thus, if P is the power that the source was emitting at time t_S, then the power received by the detector today is

$$P_{\text{received}} = \frac{P}{(1+z_S)^2} \frac{A}{4\pi \tilde{a}^2(t_0) \sin^2 \psi_D} \; . \tag{7.30}$$

The flux is given by

$$J = \frac{P_{\text{received}}}{A} = \frac{P}{4\pi (1+z_S)^2 \tilde{a}^2(t_0) \sin^2 \psi_D} \; . \tag{7.31}$$

Eq. (7.31) is the answer to our question, but it is not yet expressed in terms of useful variables — we cannot look up the values of $\tilde{a}(t_0)$ or ψ_D in standard tables, so we need to express them in terms of variables that we can look up. Specifically, we will be able to express the right-hand side of Eq. (7.31) in terms of P, z_S, H_0, and the various contributions to the current value of Ω.

Using the definition of $\tilde{a}(t)$ given by Eq. (7.28), one sees that its present value $\tilde{a}(t_0)$ can be related to $\Omega_{k,0}$, which was defined by Eq. (7.17). With a little rearranging, the relation becomes

$$\tilde{a}(t_0) = \frac{cH_0^{-1}}{\sqrt{-\Omega_{k,0}}} \; . \tag{7.32}$$

(Note that for a closed universe, $\Omega_k < 0$, so the denominator could have been written as $\sqrt{|\Omega_{k,0}|}$.)

Finally, we need to evaluate ψ_D, which we expect to be determined by the redshift z_S and cosmological parameters:

$$\psi_D = \psi(z_S) \; , \tag{7.33}$$

where $\psi(z_S)$ is defined as the ψ coordinate traversed by radial light pulses that are now reaching us with redshift z_S. These light pulses travel along a null trajectory, where the word "null" means that $ds^2 = 0$. Given the metric (7.27), a radial null trajectory is described by

$$0 = -c^2\, dt^2 + \tilde{a}^2(t)\, d\psi^2 \quad \Longrightarrow \quad \frac{d\psi}{dt} = \frac{c}{\tilde{a}(t)} \; . \tag{7.34}$$

The evolution equation for $\tilde{a}(t)$ is identical to the evolution equation for $a(t)$ that was given as Eq. (7.18):

$$H^2 = \left(\frac{\dot{\tilde{a}}}{\tilde{a}}\right)^2 = \frac{H_0^2}{x^4}\left(\Omega_{m,0} x + \Omega_{\text{rad},0} + \Omega_{\text{vac},0} x^4 + \Omega_{k,0} x^2\right) \; , \tag{7.35}$$

where

$$x = \frac{a(t)}{a(t_0)} = \frac{\tilde{a}(t)}{\tilde{a}(t_0)} \ . \tag{7.36}$$

Since the light pulse travels from time $t = t_S$ to $t = t_0$, the radial coordinate that it traverses can be found by integrating Eq. (7.34) to find

$$\psi(z_S) = \int_{t_S}^{t_0} \frac{c}{\tilde{a}(t)}\, dt \ . \tag{7.37}$$

Since we are hoping to express the answer in terms of the redshift of the source z_S, it is useful to change the variable of integration to z, where

$$1 + z = \frac{\tilde{a}(t_0)}{\tilde{a}(t)} \ . \tag{7.38}$$

Then

$$dz = -\frac{\tilde{a}(t_0)}{\tilde{a}(t)^2}\dot{\tilde{a}}(t)\, dt = -\tilde{a}(t_0) H(t)\, \frac{dt}{\tilde{a}(t)} \ . \tag{7.39}$$

The integration becomes

$$\psi(z_S) = \frac{1}{\tilde{a}(t_0)} \int_0^{z_S} \frac{c}{H(z)}\, dz \ . \tag{7.40}$$

In this expression we can replace $\tilde{a}(t_0)$ using Eq. (7.32), and we can replace $H(z)$ using Eq. (7.35), recognizing that $x = 1/(1+z)$. This gives our final expression for $\psi(z_S)$:

$$\boxed{\begin{aligned}\psi(z_S) = &\sqrt{|\Omega_{k,0}|} \\ &\times \int_0^{z_S} \frac{dz}{\sqrt{\Omega_{m,0}(1+z)^3 + \Omega_{\mathrm{rad},0}(1+z)^4 + \Omega_{\mathrm{vac},0} + \Omega_{k,0}(1+z)^2}} \ .\end{aligned}} \tag{7.41}$$

We can now go back to our answer, expressed as Eq. (7.31), and eliminate the unwanted variables. $\tilde{a}(t_0)$ is replaced using Eq. (7.32), and $\sin^2\psi_D$ can be replaced by $\sin^2\psi(z_S)$, giving

$$\boxed{J = \frac{P H_0^2 |\Omega_{k,0}|}{4\pi(1+z_S)^2 c^2 \sin^2\psi(z_S)} \ ,} \tag{7.42}$$

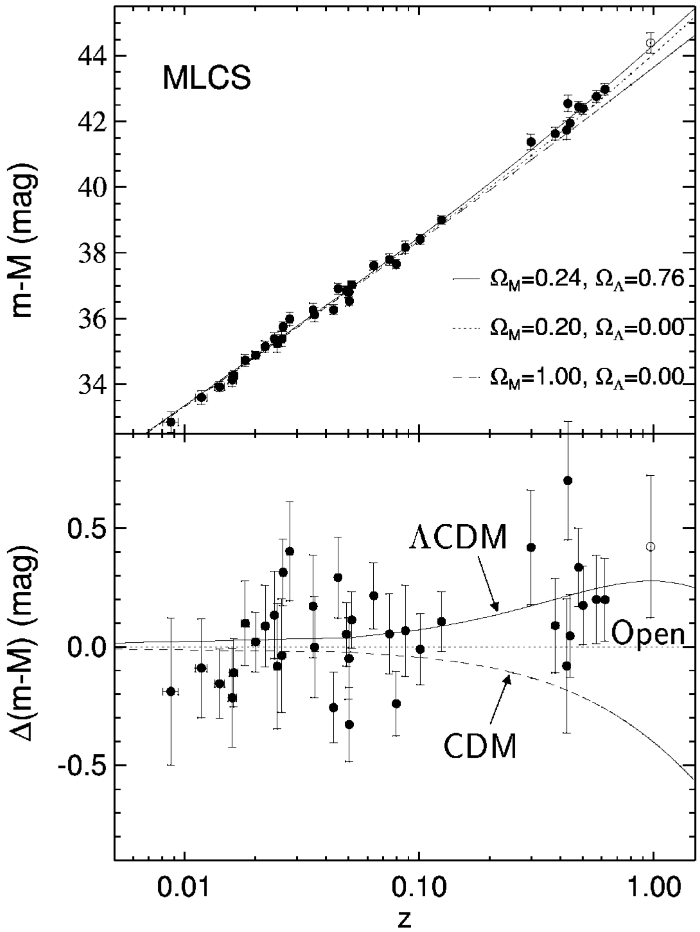


Figure 7.5: Hubble diagram from the paper by Riess *et al.* (1998).

where $\psi(z_S)$ is given by Eq. (7.41).

For a sample of the recent data, I include as Fig. 7.5 a graph of the Hubble diagram from the paper by Riess *et al.* (1998) that was cited at the beginning of these lecture notes. Shown at the top is a graph of magnitude vs. redshift for a sample of supernovae. The vertical axis represents distance as inferred from the brightness, with larger distances at the top.* Each increase of 5 magnitudes corresponds to the brightness decreasing by a

---

* More precisely, $m - M$ is the distance modulus, which is related to the *luminosity distance* d_L by

$$m - M = 5 \log_{10} \left(\frac{d_L}{1 \text{ Mpc}} \right) + 25 \ ,$$

where the luminosity distance is defined as the distance at which the object would have factor of 100, so one magnitude corresponds to a factor of 2.512, and an increase by a tenth of a magnitude corresponds to about a 10% decrease in brightness. Shown on the same graph are three theoretical curves, calculated from Eq. (7.42), using different theoretical parameters. The lowest curve represents a matter-dominated flat universe (CDM = "cold dark matter"), with no cosmological constant. The middle curve represents an open matter-dominated universe, with $\Omega_m = \Omega_{\rm tot} = 0.2$, a value which was observationally plausible before the presence of dark energy became convincing. The uppermost curve, which seems to be the best fit to the data, represents a flat universe which includes nonrelativistic matter and a cosmological constant (ΛCDM = cosmological constant + cold dark matter), with the nonrelativistic matter comprising 0.24 of the critical density, and the vacuum mass density of the cosmological constant comprising 0.76 of the critical density. These ratios were chosen as a best fit to the data, within the class of flat models with these two components. Note that these numbers agree very well with the Planck 2013 best fit model, even though the observations used to determine the parameters are completely different. The initials "MLCS" at the top stand for "Multi-Color Light Curve Shape," a method of analysis that the authors employed to compensate for small differences in the brightness of the supernovae based on the duration of the light output. The graph at the bottom shows the same data, but in a way that visually emphasizes the differences between the three curves. On this graph the middle curve is plotted as a straight line, and the other curves are shown as offsets relative to the middle curve. Note that the curves differ by two or three tenths of a magnitude, indicating that the brightness differences are only 20 to 30%. That is, the measured brightnesses of the distant supernovae are 20 to 30% dimmer than would be expected in the open universe $\Omega_m = \Omega_{\rm tot} = 0.2$ model.

---

to be located to result in the observed brightness, if we were living in a static Euclidean universe. In such a universe the energy flux J at a distance d would be given by

$$J = \frac{P}{4\pi d^2} \ ,$$

so the luminosity distance is given by

$$d_L = \sqrt{\frac{P}{4\pi J}} \ .$$

Thus,

$$m - M = -\frac{5}{2}\log_{10}\left(\frac{4\pi J \times (1\ \text{Mpc})^2}{P}\right) + 25 \ .$$

The connection between this effect and acceleration is a little hard to see, but it can be seen most clearly if one thinks about the appearance of supernovae with a fixed magnitude, and hence a fixed distance as measured by the luminosity. Then the measured points lie to the left of the open universe $\Omega_m = \Omega_{\rm tot} = 0.2$ model, which means that the redshift is lower than expected. Lower redshift means smaller velocities, and hence the universe in the past was expanding more slowly than expected. If the universe in the past was expanding more slowly than expected on the basis of the current expansion rate, it means that some accelerating influence must have been at work.

The graph may not appear to be very conclusive, but nonetheless the data, if taken at face value, is statistically very significant. Especially when this data is combined with the data from the other group, the possibility that we are seeing a statistical fluke is very small. Nonetheless, there are possible systematic errors that are hard to evaluate. The observed effect is simply the fact that distant supernovae, at a given redshift, appear slightly dimmer (by about 20 to 30%) than expected. One alternative explanation might be that there is dust that obscures our view, causing the supernovae to appear dimmer than they really are. The problem with this explanation is that most forms of dust distort the spectrum of the light, absorbing more of the shorter wavelengths, resulting in a "reddening" of the received light. Since this reddening is not observed, the dust must be "gray," the word that is used to describe a filter that absorbs equally across the spectrum. It is physically possible for dust to be gray if the grains are large enough, but such dust is not known to exist. Another difficulty with the dust hypothesis is that if most of the dust is located in the host galaxy of the supernova, as one would expect, then the amount of absorption should depend on where the supernova is located within the galaxy. This would in turn produce scatter in a graph like the one shown above, while the amount of scatter seen is consistent with the known sources of uncertainty. Another totally different explanation for the observations is the possibility that it is caused by galactic evolution. Heavy elements are produced in stars, so 5 billion years ago there was a noticeably lower abundance of heavy elements in galaxies. If the lower abundance of heavy elements could lead to dimmer supernovae explosions, then this evidence for a cosmological constant would disappear. However, astronomers have looked hard to find any visible differences between the early supernovae at large redshift and the recent supernovae nearby, and so far they have found nothing significant. Further, in the nearby universe there are galaxies with a range of abundances of heavy elements, and this has not been observed to produce a difference in the brightness of supernova explosions.

Furthermore a persuasive piece of evidence was uncovered in early 2001 by A.G. Riess, P.E. Nugent, et al.*, who discovered in data from the Hubble Space Telescope

---

* "The Farthest Known Supernova: Support for an Accelerating Universe and a Glimpse of the Epoch of Deceleration," http://arxiv.org/abs/astro-ph/0104455, Riess *et al.*, *Astrophysical Journal* **560**, 49–71 (2001).

a supernova at the colossal redshift of 1.7. This redshift is large enough so that the light left the supernova before the era of acceleration is believed to have started. So this measurement would be expected to show the decelerating behavior expected for earlier times, and indeed it did. By contrast, effects caused by dust or by heavy element abundance would not be expected to reverse at earlier times.

On balance, I think it is fair to say that currently most cosmologists regard the supernova data as persuasive, but not, by itself, irrefutable. However, there is also increasingly strong evidence from observations of the cosmic microwave background radiation, which we have summarized earlier in these lecture notes in terms of the Planck and WMAP results. These observations provide a measurement of the amount of vacuum energy that agrees very well with the supernova results. In addition, they provide very strong evidence that the universe is flat. There is also much evidence from extragalactic astronomy that there is not enough matter in the universe, even including the dark matter, to make up the critical density that is required by general relativity for a flat universe. If this is right, then vacuum energy becomes the most straightforward explanation of where the mass density is hidden. Also, as we have discussed, the inclusion of vacuum energy makes the calculation of the age of the universe from the Hubble expansion rate consistent with the estimated ages of the oldest stars. With all the evidence combined, there seems to be no alternative to the belief that about 70% of the mass density of the universe is in the form of dark energy — a negative pressure material that is either vacuum energy (also called a cosmological constant), or perhaps “quintessence,” which we will discuss later.

THE PARTICLE PHYSICS OF A COSMOLOGICAL CONSTANT:

While the observational evidence for a cosmological constant seems strong, the underlying physics of a cosmological constant remains very mysterious. From the point of view of modern particle physics it is not at all strange that the vacuum should have a nonzero mass density, but it is very hard to imagine any reason why it should have a value anywhere near the value that is being observed.

According to modern particle physics, the vacuum is actually a very complicated state. It is defined as the state with the lowest possible energy density, but it is not “empty” in any conventional sense. For example, the electric and magnetic fields are constantly fluctuating in the vacuum, because the uncertainty principles of quantum theory do not allow them to remain at zero value. These fluctuations give a positive contribution to the vacuum energy. The calculation of this contribution is formally infinite, since each mode of oscillation contributes, and there are an infinite number of modes at arbitrarily short wavelengths. It seems reasonable, however, to truncate this infinite sum at what is called the Planck length,

$$\lambda_{\text{Planck}} = \sqrt{\frac{\hbar G}{c^3}} = 1.6 \times 10^{-33} \text{ cm} . \tag{7.43}$$

This is the scale at which quantum gravity effects are believed to become important, and even the very notion of classical space presumably breaks down. With this cut-off the answer becomes finite, but it is more than 120 orders of magnitude larger than the energy density associated with the observed cosmological constant! You will have a chance to work this out in detail on Problem Set 8.

There are known negative contributions to the vacuum energy density as well, coming from fermions, such as the electron. Fermions give a huge negative contribution to the energy density of the vacuum, an effect that can be understood intuitively in terms of a metaphor known as the "Dirac sea". That is, the Dirac equation which describes relativistic electrons has both positive energy and negative energy solutions. These solutions are viewed as the possible energy levels of particles. When one of the positive energy levels becomes occupied by a particle, the overall energy of the state increases. But the overall energy is lowered whenever one of the negative energy levels becomes occupied by a particle. The vacuum, therefore, is the state in which all the negative enery levels are filled. The occupation of one of the positive energy levels then corresponds to an electron, which can be present in an otherwise vacuum state. The overall energy can also be increased by vacating one of the negative energy levels, leaving behind a "hole in the Dirac sea." Such a hole corresponds to a positron, the antiparticle of the electron.

This "filling of the Dirac sea" gives a negative energy density to the vacuum, since the filling of each negative energy level decreases the overall energy. Like the positive contribution of the electromagnetic field oscillations, the magnitude of this contribution is formally infinite. When it is cut off at the Planck length it becomes finite and comparable in magnitude to the positive contribution of the electromagnetic field.

There is a possibility that these huge positive and negative contributions could somehow cancel each other almost but not quite exactly, but no one knows why. In the absence of any real understanding, physicists had until recently assumed that the positive and negative contributions most likely cancel exactly because of some unknown symmetry principle. Even if that were the case, it would of course be an important challenge to understand why. If there really is a cosmological constant, then it looks like the positive and negative contributions to the vacuum energy density cancel to an accuracy of 120 decimal places, but miss in the 121st decimal place. Or maybe we are just looking at this all wrong.

At the present time, the cosmological constant problem is perhaps the most significant outstanding problem in our understanding of fundamental physics.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe November 29, 2018
Prof. Alan Guth

Lecture Notes 8

PROBLEMS OF THE CONVENTIONAL (NON-INFLATIONARY) HOT BIG BANG MODEL

INTRODUCTION:

By the 1970s the conventional hot big bang model was well established. It was supported strongly by the observation of Hubble expansion, by the existence of a thermal background of microwave radiation, and by the measured abundances of the lightest isotopes of atomic nuclei. Nonetheless the model suffered from at least two serious problems — the horizon/homogeneity problem and the flatness problem — which form the subject of this set of lecture notes. In the next set of lecture notes we will discuss a third problem — the magnetic monopole problem — which arises if one assumes that particle physics at very high temperatures is described by a type of theory called a grand unified theory.

THE HORIZON/HOMOGENEITY PROBLEM:

The horizon problem is the difficulty in explaining the large-scale uniformity of the observed universe. This large-scale uniformity is most evident in the microwave background radiation. This radiation appears slightly hotter in one direction than in the opposite direction, by about one part in a thousand — but this nonuniformity can be attributed to our motion through the background radiation. Once this effect is subtracted out, using best-fit parameters for the velocity, it is found that the residual temperature pattern is uniform to about one part in 10^5. Uniformity in temperature is not necessarily mysterious, as any isolated system will evolve over time towards a state of thermal equilibrium, which is state of uniform temperature. For example, when we take a hot slice of pizza out of the oven, it begins immediately to cool toward room temperature. It turns out, however, that the standard processes of thermal equilibration cannot, in the context of the conventional hot big bang model, explain the uniformity of temperature in the universe. The problem is that in this model the universe evolves much too quickly to allow this uniformity to be achieved by the usual processes by which a system approaches thermal equilibrium.

In order to see this, we will not need to know anything about the details of thermal transport processes. We will use only the fact that no physical process can cause matter, energy, or information to move faster than the speed of light. Thus, no process can carry energy beyond the "horizon distance," which was defined in Lecture Notes 4 as the present distance of the furthest particles from which light has had time to reach us, since the beginning of the universe. The issue of horizons was introduced into cosmology by W.

Rindler in 1956,* and the horizon problem is described (without using the words "horizon problem") in two well-known textbooks: S. Weinberg, *Gravitation and Cosmology*, J. Wiley and Sons (1972), and C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation*, W.H. Freeman & Co. (1973).

In Lecture Notes 4 we found that this horizon distance is given by 3ct for the case of a matter-dominated $k = 0$ universe, and in Lecture Notes 6 we showed that it is 2ct for the case of a radiation-dominated $k = 0$ universe. We also showed in Lecture Notes 6 that the universe became matter-dominated at about 50,000 years after the big bang, and that the cosmic microwave background radiation decoupled from the rest of matter at $t_d \approx 370{,}000$ years after the big bang. Thus, the radiation decoupled well after the universe became matter-dominated, so to a good approximation the horizon distance at this time is given by $\ell_h(t_d) \approx 3ct_d \approx 1{,}100{,}000$ light-years. (I am using the convention that a subscript "d" denotes the value of a given quantity at the time of decoupling, and a subscript "0" will denote its value at the present time.)

For comparison, we would like to calculate the distance, at time t_d, between our own galaxy (or, more precisely, the matter which will later become our galaxy) and the site of emission of the cosmic background radiation that we are now receiving. To do this, we can make use of some previous results. In Lecture Notes 6 we learned that the cosmic microwave background radiation was emitted (or more precisely, decoupled) when the temperature was 3000 °K. Since the current temperature is about 2.7 °K, and since $aT = constant$ as the universe expands, it follows that the redshift at the time of decoupling is given by

$$1 + z = \frac{a(t_0)}{a(t_d)} = \frac{3000°\mathrm{K}}{2.7°\mathrm{K}} \approx 1100 \; , \tag{8.1}$$

where I have also made use of Eq. (2.15) to relate $1 + z$ to the ratio of the scale factors. Knowing $1+z$ we can find the present distance between our galaxy and the site of emission of the radiation, using the result of Problem 3, Problem Set 2. You found there that for a matter-dominated $k = 0$ universe, the present physical distance of an object seen at redshift z is given by

$$\ell_p(t_0) = 2cH_0^{-1}\left[1 - \frac{1}{\sqrt{1+z}}\right] \; . \tag{8.2}$$

Using $H_0 = 67.7$ km-s^{-1}-Mpc^{-1}, one finds that $H_0^{-1} \approx 14.4 \times 10^9$ yr and $\ell_p(t_0) \approx 28.0 \times 10^9$ light-yr. That is, the region of emission of the cosmic background radiation that we are presently observing is a spherical shell of matter at just a little bit less than

* W. Rindler, "Visual horizons in world-models," *Monthly Notices of the Royal Astronomical Society,* Vol. 116, pp. 662–677 (1956), http://adsabs.harvard.edu/abs/1956MNRAS.116..662R

the present horizon distance, $\ell_h(t_0) \approx 2cH_0^{-1}$. But physical distances vary with time as $a(t)$, so the physical radius of this shell of matter at the time of decoupling t_d is given by

$$\begin{aligned} \ell_p(t_d) &= \frac{a(t_d)}{a(t_0)} \ell_p(t_0) \\ &\approx \frac{1}{1100} \times 28.0 \times 10^9 \text{ lt-yr} \approx 2.55 \times 10^7 \text{ lt-yr} . \end{aligned} \tag{8.3}$$

Thus, at the time of emission of the cosmic background radiation, the region of emission was a spherical shell with a radius many times larger than the horizon distance. Specifically, the radius was $\ell_p(t_d)/\ell_h(t_d) \approx 2.55 \times 10^7$ lt-yr$/1.1 \times 10^6$ lt-yr ≈ 23 times larger than the horizon distance.

To state the problem most clearly, suppose that one detects the cosmic microwave background in a certain direction in the sky, and suppose that one also detects the radiation from precisely the opposite direction. At the time of emission, the sources of these two signals were separated from each other by about 46 horizon distances. Thus it is absolutely impossible, within the context of this model, for these two sources to have come into thermal equilibrium by any physical process.

This problem is not a genuine inconsistency of the conventional hot big bang model — if the uniformity is assumed in the initial conditions, then the universe will evolve uniformly. The "problem" is that one of the most salient features of the observed universe — its large-scale uniformity — cannot be explained by the conventional model; it simply must be assumed as an initial condition. The suggestion then is not that the conventional model is wrong, but rather that it is incomplete.

The calculation described above depended on our approximation that the universe was matter-dominated at all relevant times, which is a rather crude approximation. Nonetheless, since 46 is so far from one, we can be confident that this problem will not go away with a more careful calculation.

THE FLATNESS PROBLEM:

A second problem of the conventional hot big bang model is known as the flatness problem — it refers to the difficulty in understanding why the present value of Ω (the ratio of the mass density ρ to the critical mass density ρ_c) is close to 1. Today we know that Ω_0 is equal to 1 at least to within about 1/2% — more precisely, by combining their own data with data from other experiments, the Planck team* concluded that

$$\Omega_0 = 0.9993 \pm 0.0037 \tag{8.4}$$

* N. Aghanim et al. (Planck Collaboration), "Planck 2018 results, VI: Cosmological parameters," Table 4, Column 5, arXiv:1807.06209.

at the 95% confidence level. Historically, however, the flatness problem was already severe even in 1980, when we only knew that Ω_0 was somewhere in the range

$$0.1 < \Omega_0 < 2 \quad \text{(circa 1980)}. \tag{8.5}$$

The key fact is that the value $\Omega = 1$ is a point of unstable equilibrium, something like a pencil balancing on its point. The word "equilibrium" implies that if Ω is ever ***exactly*** equal to one, it will remain equal to one forever — that is, a flat ($k = 0$) universe remains a flat universe. However, if Ω is ever slightly larger than one, it will rapidly grow toward infinity; if Ω is ever slightly smaller than one, it will rapidly fall toward zero. Thus, in order for Ω to be anywhere near 1 today, the value of Ω in the early universe must have been extraordinarily close to one.

Like the horizon problem, this problem is not a genuine inconsistency of the conventional model. If one is willing to assume that the value of Ω in the early universe was extraordinarily close to one, then the model will describe how the universe evolves to have a value of Ω today within the accepted range. The problem is again the lack of explanatory or predictive power of the model — the extraordinary closeness of Ω to unity in the early universe cannot be explained, but must simply be assumed as an initial condition. The mathematics behind the flatness problem was undoubtedly known to almost anyone who has worked on the big bang theory from the 1920's onward, but apparently the first people to consider it a problem in the sense described here were Robert Dicke and P.J.E. Peebles, who published a discussion in 1979.†

To work out the evolution of Ω, we need only recast some relations that we have already derived. The key relation is the first-order Friedmann equation for the evolution of the scale factor, derived in Lecture Notes 3:

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3} G\rho - \frac{kc^2}{a^2} , \tag{8.6}$$

where the overdot represents a derivative with respect to t. Recalling that $\rho_c = 3H^2/8\pi G$ (see Eq. (3.33)), one can divide both sides of the equation by H^2 to give

$$1 = \frac{\rho}{\rho_c} - \frac{kc^2}{a^2 H^2} , \tag{8.7}$$

which can be rewritten as

$$\Omega - 1 = \frac{kc^2}{a^2 H^2} , \tag{8.8}$$

† R.H. Dicke and P.J.E. Peebles, "The big bang cosmology — enigmas and nostrums," in **General Relativity: An Einstein Centenary Survey**, eds: S.W. Hawking and W. Israel, Cambridge University Press (1979).

where we recalled that

$$\Omega \equiv \frac{\rho}{\rho_c} \; . \tag{8.9}$$

But the evolution of a and H are already understood. For a matter-dominated $k = 0$ universe, we know from Lecture Notes 3 that $a \propto t^{2/3}$, and therefore $H = \dot{a}/a = 2/(3t)$. It follows that

$$\Omega - 1 \propto \left(\frac{1}{t^{2/3}}\right)^2 \left(\frac{1}{t^{-1}}\right)^2 \propto t^{2/3} \quad \text{(matter-dominated).} \tag{8.10}$$

For a radiation-dominated $k = 0$ universe, on the other hand, we know from Lecture Notes 6 that $a \propto t^{1/2}$, so $H = 1/(2t)$. This gives

$$\Omega - 1 \propto \left(\frac{1}{t^{1/2}}\right)^2 \left(\frac{1}{t^{-1}}\right)^2 \propto t \quad \text{(radiation dominated).} \tag{8.11}$$

We can now trace the evolution of Ω backward in time. From the Planck limit on Ω_0 listed in Eq. (8.4), we can conclude that

$$|\Omega_0 - 1| < .01 \; . \tag{8.12}$$

We could have written $|\Omega_0 - 1| < 0.005$, but for simplicity we will consider only integer powers of ten. Again for simplicity, we will assume that the universe can be described in terms of a matter-dominated era and a radiation-dominated era, both nearly flat so that Eqs. (8.10) and (8.11) apply, with a sharp transition between the two. The transition occurs at about 50,000 years after the big bang, while we estimate the current age of the universe as 13.8×10^9 years. Using Eq. (8.10), we conclude that the value of Ω at 50,000 years is given by

$$(\Omega - 1)_{t=50{,}000 \text{ yr}} \approx \left(\frac{50{,}000}{13.8 \times 10^9}\right)^{2/3} (\Omega_0 - 1) \approx 2.36 \times 10^{-4} \, (\Omega_0 - 1) \; . \tag{8.13}$$

Let us now calculate the value of Ω at 1 second after the big bang. One second is a particularly interesting time, because it is the earliest time for which we have direct evidence that the conventional hot big bang model seems to be working. The processes which lead to nucleosynthesis begin at about $t = 1$ sec, and the predictions derived from big bang nucleosynthesis calculations are in good agreement with observations.

To find the value of Ω at one second, begin by noting that

$$\frac{1 \text{ sec}}{50{,}000 \text{ yr}} = \frac{1 \text{ sec}}{50{,}000 \text{ yr}} \times \frac{1 \text{ yr}}{3.16 \times 10^7 \text{ sec}} = 6.33 \times 10^{-13} \; .$$

Combining Eqs. (8.11) and (8.13), we find that

$$\begin{aligned}(\Omega-1)_{t=1\ \mathrm{sec}} &\approx\ 6.33\times 10^{-13}\,(\Omega-1)_{t=50{,}000\ \mathrm{yr}}\\ &\approx\ 1.49\times 10^{-16}\,(\Omega_0-1)\ .\end{aligned} \tag{8.14}$$

Using Eq. (8.12), we conclude that

$$|\Omega-1|_{t=1\ \mathrm{sec}} < 10^{-18}\ . \tag{8.15}$$

Thus, at one second after the big bang, the value of Ω must have been equal to one to an accuracy of 18 decimal places! The flatness problem is the statement that the conventional hot big bang model provides no explanation of how the value of Ω came to be tuned so precisely. Note that if we had put ourselves back into the setting of 1980, using Eq. (8.5) instead of (8.4), we still would have reached the extraordinary conclusion that Ω at one second must have equaled 1 to an accuracy of 16 decimal places.

As we will see shortly, the horizon and flatness problems provide much of the motivation of the inflationary universe model, which gives a simple resolution to both of them.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe November 29, 2018
Prof. Alan Guth

Lecture Notes 9
THE MAGNETIC MONOPOLE PROBLEM

INTRODUCTION:

In addition to the horizon and flatness problems discussed in Lecture Notes 8, the conventional (non-inflationary) hot big bang model potentially suffers from another problem, known as the magnetic monopole problem. If one accepts the basic ideas of grand unified theories (GUT's) in addition to those of the conventional cosmological model, then one is led to the conclusion that there is a serious problem with the overproduction of particles called "magnetic monopoles". While a full understanding of the particle physics of grand unified theories is obviously much more than can be accomplished in a single set of lecture notes, the goal here will be to give you a qualitative understanding of what a grand unified theory is and how magnetic monopoles arise in such theories. We will not try to give a solid justification for all the steps along the way, but we will get far enough so that you will be able to estimate for yourself the magnetic monopole production in the early universe, verifying that far too many monopoles are predicted in the context of the conventional hot big bang cosmology.

THE STANDARD MODEL OF PARTICLE PHYSICS:

Before discussing grand unified theories, there are a few things that should be said about the "standard model of particle physics," which is the bedrock of our understanding of particle physics. The standard model, which was developed in the early 1970s, has enjoyed enormous success, giving predictions in agreement with all reliable particle physics experiments so far. The model has been enlarged since its initial discovery, adding a third generation of fundamental fermions, but the form of the standard model has remained unchanged. The original formulation described massless neutrinos, but the model can easily be modified to include neutrino masses (which are now known to be nonzero, due to neutrino oscillations). There is more than one way to add neutrino masses, however, and we are still not sure what is the correct way to do it.

Physicists divide the known interactions in nature into four classes: (1) the strong interactions, which bind quarks together inside protons, neutrons, and other strongly interacting particles, and also provide the residual force responsible for the interactions between these particles; (2) the weak interactions, responsible for example for beta decay ($n \to p + e^- + \bar{\nu}_e$; e.g. neutron $\to$ proton + electron + anti-electron-neutrino); (3) electromagnetic interactions; and (4) gravity. The standard model of particle physics describes the first three of these, omitting gravity. In practice there is no problem ignoring gravity at the level of elementary particle interactions, as the gravitational force between two elementary particles is so weak that it has never been detected. The gravitational force between two protons, for example, is 10^{36} times weaker than the electrostatic force between the same two particles.

The elementary particle content of the standard model of particle physics is shown in Fig. 9.1,* at the right. All particles are classified as either fermions or bosons. Fermions are particles with spins that in units of $\hbar$ are equal to $\frac{1}{2}$, $\frac{3}{2}$, etc., where for the fundamental particles the spin is always $\frac{1}{2}\hbar$. Fermions obey the Pauli exclusion principle. Bosons are particles with spins that are integer multiples of $\hbar$. They obey quantum mechanical rules which are the opposite of the Pauli exclusion principle, so that bosons have an enhanced tendency to fall into the same quantum state — that is the underlying principle behind the workings of a laser. The fermions of the standard model belong to three "generations," where the second and third generations are essentially copies of the first, except they are more massive. (The neutrinos are possibly an exception to this, as we do not know either the values or the ordering of the neutrino masses.) Each generation contains two quarks, with charges 2/3 and -1/3 in units of the magnitude of the electron charge, and also a neutrino and a lepton. Each quark comes in three different "color" states, and all the fermions have associated antiparticles. The color of a quark of course has nothing to do with its visual appearance, but is simply a label which was dubbed "color" because there are three possible values, like the three primary colors.

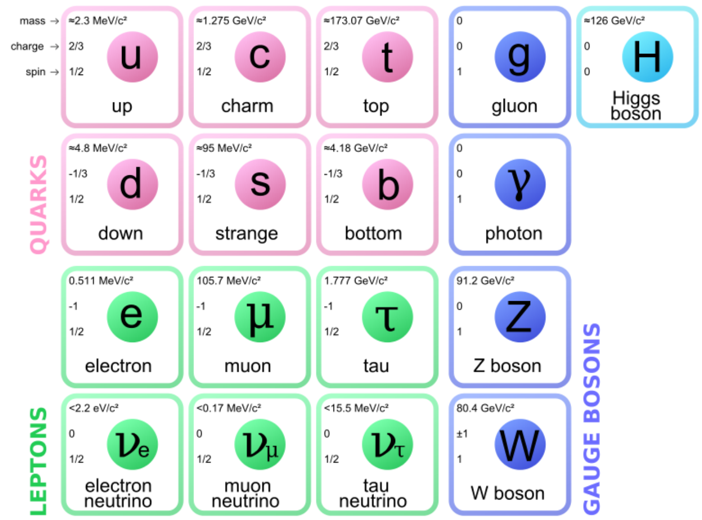


Figure 9.1: The particles of the standard model of particle physics.

The interactions of the standard model are mainly provided by the "gauge bosons" shown in the fourth column of the diagram. These are spin-1 particles, and hence bosons, of which the most familiar is the photon γ. The photon is its own antiparticle, and is said to be the "carrier" of the electromagnetic interactions. The gluons g are the carriers of the strong interaction, and there are eight of them, including antiparticles. The Z^0, W^+, and W^- particles are the carriers of the weak interactions. The Z^0 is its own antiparticle, while the W^+ and W^- are antiparticles of each other.

At the right of the diagram is the Higgs particle (named for Peter W. Higgs of the University of Edinburgh), the existence of which was established in July 2012 at the Large Hadron Collider (LHC) at CERN. The Higgs particle can be seen only in very high energy processes, by modern accelerator standards, and even then only rarely. For example, in

---

* From the Wikimedia Commons. Source: PBS NOVA, Fermilab, Office of Science, United States Department of Energy, Particle Data Group.

the collision of two high energy protons at the LHC, it is possible for two gluons inside the protons to fuse into a Higgs particle, which can then decay to two photons. In addition to its role in describing Higgs particles, the Higgs field is responsible for the masses of the $W^{\pm}$, the Z^0, the quarks, and the $e^{\pm}$, the $\mu^{\pm}$, and the $\tau^{\pm}$. I will come back to the question of what exactly we mean when we say that the Higgs field is responsible for these masses. It may also be responsible for the neutrino masses, but it is not responsible for the mass of the proton; even without the Higgs field, it would be possible for massless quarks to form a massive bound state, such as the proton.

The spin-1 particles are called "gauge" particles because the standard model is an example of what is called a gauge theory. We will not have time to describe in detail what this means, but I will attempt to convey some partial understanding. You already know about one gauge theory — electromagnetism — but electromagnetism is a little too simple to make it obvious how to generalize the idea. The gauge theory aspect of electromagnetism can be seen only if it is written in terms of its potentials: the vector potential $\vec{A}$ and the scalar potential ϕ, which are related to the electric field $\vec{E}$ and the magnetic field $\vec{B}$ by

$$\begin{aligned} \vec{E} &= -\vec{\nabla}\phi - \frac{1}{c}\frac{\partial \vec{A}}{\partial t} \ , \\ \vec{B} &= \vec{\nabla} \times \vec{A} \ . \end{aligned} \tag{9.1}$$

The quantum theory of electromagnetism is always formulated in terms of these potentials. $\vec{A}$ and ϕ can be put together relativistically to define a four-potential A_μ,

$$A_\mu = (-\phi, A_i) \ . \tag{9.2}$$

As you have probably seen, the potentials themselves cannot be measured, but are in fact subject to the symmetry of gauge transformations. That is, given any scalar function $\Lambda(t, \vec{x})$, one can define new potentials ϕ' and $\vec{A}'$ by

$$\begin{aligned} \phi'(t, \vec{x}) &= \phi - \frac{\partial \Lambda}{\partial t} \ , \\ \vec{A}'(t, \vec{x}) &= \vec{A} + \vec{\nabla}\Lambda \ , \end{aligned} \tag{9.3}$$

which can be written in four-vector notation as

$$A'_\mu(x) = A_\mu(x) + \frac{\partial \Lambda}{\partial x^\mu} \ . \tag{9.4}$$

Gauge transformations are a symmetry in the sense that the new fields $A'_\mu(x)$ describe exactly the same physical situation as the original fields: $\vec{E}(t, \vec{x})$ and $\vec{B}(t, \vec{x})$ are unchanged.

If we consider two successive gauge transformations described by functions $\Lambda_1(t,\vec{x})$ and $\Lambda_2(t,\vec{x})$, the combined transformation can be described by a new function $\Lambda_3(t,\vec{x})$ given by

$$\Lambda_3(t,\vec{x}) = \Lambda_1(t,\vec{x}) + \Lambda_2(t,\vec{x}) \ , \tag{9.5}$$

so the combination of gauge transformations is described mathematically by the addition of real numbers. The combination is abelian — it does not matter in what order the gauge transformations of Λ_1 and Λ_2 are performed. The extension of gauge theories to nonabelian (i.e., non-commutative) transformations was invented in 1954 by Chen Ning (Frank) Yang and Robert Mills, and this idea became a key ingredient of the standard model of particle physics and its extensions. The standard model is based on gauge transformations that follow the form of three mathematical groups: SU(3), SU(2), and U(1). SU(3) is defined as the group of complex 3×3 matrices which are special (S) and unitary (U). The group operation is matrix multiplication. A matrix is special if its determinant is 1. A matrix U is unitary if it obeys the relation $U^\dagger U = I$, where $U^\dagger$ is called the adjoint of the matrix U, which means that $U^\dagger$ is the matrix obtained by transposing U (interchange rows and columns) and then taking its complex conjugate. I denotes the identity matrix. The definition of a unitary matrix is equivalent to saying that U has the property that if it multiplies a complex column vector v of the same size, then the norm of v is unchanged: $|Uv| = |v|$, where $|v| \equiv \sqrt{v^\dagger v}$. The group SU(2) is defined analogously, except that it uses 2×2 matrices. The group SU(2) is in fact essentially the same as the group of rotations in three spatial dimensions, although that fact is not obvious if you have not seen it. (It is actually a 2:1 map, with two matrices in SU(2) corresponding to each rotation matrix.) Finally, $U(1)$ is simply the group of complex phases, in the sense that an element of U(1) can be represented as a complex number $z = e^{i\theta}$, where θ is a real number.

In this language standard electromagnetism is a gauge theory based on U(1). From Eq. (9.5) it looks like we should be talking about the group of real numbers under addition, but in fact both descriptions are okay. If we included Dirac fields in our theory, to describe relativistic electrons, the Dirac field ψ would transform under the gauge transformation as

$$\psi'(x) = \exp\{ie_0\Lambda(x)\}\,\psi(x) \ , \tag{9.6}$$

where e_0 is the charge of the electron. So the transformation is fully described by the complex phase $z = \exp\{ie_0\Lambda(x)\}$. Note that this phase does not give us enough information to find Λ, because Λ can be changed by a multiple of $2\pi/e_0$ without changing the phase. But it nonetheless does give us enough information to find the gauge transformation of $A_\mu(x)$, since

$$\frac{\partial\Lambda}{\partial x^\mu} = \frac{1}{ie_0}\, e^{-ie_0\Lambda(x)}\, \frac{\partial}{\partial x^\mu}\, e^{ie_0\Lambda(x)} \ . \tag{9.7}$$

Yang and Mills invented a procedure to construct a field theory based on any of these gauge groups. A gauge transformation is defined by specifying an element of the group at each point in spacetime, and two successive transformations combine according to the definition of multiplication in the mathematical group. Since SU(3) and SU(2) are nonabelian, these are called nonabelian gauge theories. These theories require specific spin-1 fields, which are the gauge fields shown in column 4 of Fig. 9.1. The equations of motion of these gauge fields are completely determined by the gauge symmetry, except for one *coupling constant* g for each gauge symmetry, which describes the size of the nonlinear terms in the equations of motion. Note that linear equations of motion allow any two solutions to be superimposed to obtain a third solution, which means that waves of the field do not interact, and the corresponding particles are free (i.e., do not interact). The coupling constant g therefore describes the strength of the interactions of the gauge particles, both with themselves and with other particles. For electromagnetism the gauge coupling constant is e, the magnitude of the electron charge, which in fact describes the strength of the interaction between photons and any charged particle. Photons are atypical among gauge particles, however, in that photons do not interact with other photons, which is a consequence of the fact that photons obey an abelian gauge theory. In addition to the gauge fields, gauge theories also allow other fields to be present, with interactions that are strongly restricted by the gauge symmetry, but not completely determined by it.

The three gauge symmetries of the standard model are usually described together as a product group, SU(3)×SU(2)×U(1). An element of such a product group is simply an ordered triplet (u_3, u_2, u_1), where u_3 is an element of SU(3), u_2 is an element of SU(2), and u_1 is an element of U(1). Thus the product group provides a compact notation, but really has the same information content as you would get by thinking about the three groups individually.

The SU(3) part describes the strong interactions, while the SU(2) and U(1) together describe the electromagnetic and weak interactions. The two are intertwined in their effect, however, so together they describe the *electroweak* interactions. While electromagnetism is a U(1) gauge theory, the U(1) of electromagnetism is actually a combined transformation that involves the U(1) of the standard model and a rotation about one fixed direction within the SU(2) group.

The Higgs field of the standard model is a complex doublet; i.e.,

$$H(x) \equiv \begin{pmatrix} h_1(x) \\ h_2(x) \end{pmatrix} , \tag{9.8}$$

where $h_1(x)$ and $h_2(x)$ are each complex numbers defined at each point $x \equiv (t, \vec{x})$ in spacetime. The functions $h_1(x)$ and $h_2(x)$ are called the *components* of the Higgs field $H(x)$. The doublet transforms under SU(2) gauge transformations in what is called the fundamental representation. I.e.,

$$H'(x) = u_2(x) H(x) , \tag{9.9}$$

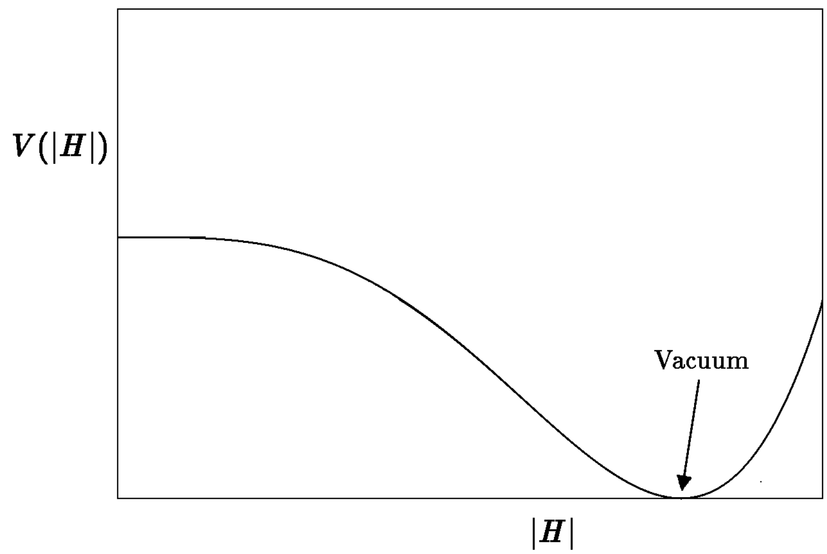


Figure 9.2: The approximate shape of the potential energy function for the Higgs field in the standard model of particle physics.

where u_2 denotes the 2×2 complex unitary matrix that defines the element of the gauge transformation at the spacetime point x. The gauge symmetry implies that the potential energy density of the Higgs field must be gauge invariant, which in turn means that it can depend only on the norm of the field,

$$|H| \equiv \sqrt{|h_1|^2 + |h_2|^2} \; . \tag{9.10}$$

The potential energy function for the Higgs field is assumed to have a peculiar form, as shown in Fig. (9.2). The potential energy function (which actually describes the potential energy density) is chosen to produce a phenomenon called spontaneous symmetry breaking. Spontaneous symmetry breaking is actually a common phenomenon in many branches of physics, including familiar processes such as the freezing of water. In the case of water, the relevant symmetry is rotational invariance. The laws of physics that describe water are completely rotationally invariant, with no direction preferred over any other direction. However, when water freezes, it forms a crystalline lattice which is not rotationally invariant. The crystalline lattice picks out definite directions along which the molecules align. The initial allignment is chosen randomly as the first molecules bind together, and then the rest of the molecules follow the pattern as they join onto the crystal. In general, whenever the ground state of a system has less symmetry than the underlying laws that describe it, it is called spontaneous symmetry breaking.

The equations of the standard model are exactly invariant under the gauge symmetry, but the only value of H that would be invariant under the gauge transformation (9.9) is $H = 0$. But the potential energy function is designed so that the state $H = 0$ has a high energy, and the vacuum state — the state with the lowest possible energy density — has a nonzero value of $|H|$. This means that in the vacuum H must have a value that breaks the symmetry. There are of course an infinite number of directions in (h_1, h_2) space which would have the same value of $|H|$, and all would minimize the energy just as well. Like the direction of the crystal axes, the direction is picked out at some early time, and after that it is "frozen", becoming constant in time and over large regions of space.

The SU(2)×U(1) part of the standard model gauge symmetry implies that in the fundamental equations there is no distinction between electrons and neutrinos. The distinction arises entirely from the spontaneous symmetry breaking. The lepton fields interact with the Higgs fields, and those which interact with the components of the Higgs fields that have nonzero values will behave differently from the components that remain zero. Thus some components of the lepton fields will describe electrons, and some will describe neutrinos.

Before leaving the standard model, I'd like to try to qualitatively explain the connection between Higgs fields and mass. First, when we say that the Higgs field is responsible for the mass of the quarks, leptons, W^+, W^-, and the Z, we are talking about their rest masses. If the Higgs were not included in the theory, all these particles would have zero rest mass, like the photon. To understand how one field can influence the rest mass of another, remember that particles in a quantum field theory are simply the quantized excitations of fields. The rest mass times c^2 is the least energy that a particle can have, so the rest mass is just $1/c^2$ times the energy of the smallest possible excitation. To find this smallest excitation, we imagine describing the field inside a rectangular box, with for example periodic boundary conditions, and expand the field in terms of its normal modes. For small oscillations each normal mode behaves as a harmonic oscillator. When a harmonic oscillator is described quantum mechanically, the lowest energy level is $E = \frac{1}{2}\hbar\omega$, where ω is the (angular) frequency of the oscillator. The excited energy levels are evenly spaced, with the n'th energy level given by

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega \ . \tag{9.11}$$

Thus the smallest excitation is given by

$$\Delta E = E_1 - E_0 = \hbar\omega \ . \tag{9.12}$$

For any field, the mode with the smallest frequency is the homogeneous mode, the mode where the whole field oscillates uniformly. Thus, the mass of the particle is simply

$$m_0 = \frac{\hbar\omega_0}{c^2} \ , \tag{9.13}$$

where ω_0 is the angular frequency for homogeneous oscillations of the field. This formula implies that the rest mass of the photon is zero, since the frequency of a photon is given by $\omega = 2\pi c/\lambda$, so it approaches zero as λ approaches infinity, which is the limit of homogeneous oscillations. This is also the case for the quarks, leptons, W^+, W^-, and the Z of the standard model, if the Higgs field has zero value. But when some components of the Higgs fields have a nonzero value, the equations of the theory imply that these Higgs fields interact with the other fields, in some cases creating a restoring force, proportional to the value of one of the Higgs components, which pushes the other fields towards zero value. This restoring force results in a nonzero frequency for homogeneous oscillations, and hence a rest mass for the corresponding particles.

While the standard model (with some modification for neutrino masses) has been spectacularly successful in explaining all particle physics experiments, few if any physicists regard it as the final story, for at least two types of reasons. First, the theory is incomplete: it does not include gravity, nor does it contain any particle which can account for the dark matter in the universe. Second, the theory is viewed by physicists as being too inelegant (i.e., too ugly) to be the final theory. Specifically, the theory has many more seemingly arbitrary features and free parameters than one would hope. Why should there be three unconnected gauge symmetries, and why should there be three generations of fermions? The theory in its original form has 19 free parameters, such as the masses of each of the fermions and the strengths of the three fundamental interactions, which have values that must be measured, but cannot be deduced from any known principle. To account for neutrino masses, 7 or 8 new parameters must be added. What determines the values of all these parameters? Thus, while the standard model is certainly very accurate over a huge range of phenomena, the field of "beyond-the-standard-model" (BSM) particle physics is burgeoning.

Grand Unified Theories:

Grand unified theories, which were first proposed in the 1970s, are one promising attempt to go beyond the standard model. Grand unified theories are aimed primarily at unifying the three gauge interactions of the standard model, namely the SU(3), SU(2), and U(1) interactions. This is accomplished by embedding all three symmetry groups into a single, larger group, which becomes the gauge symmetry of the full theory. In the context of the full grand unified symmetry, there is no distinction between a neutrino, an electron, or a quark. The distinction is entirely created by the spontaneous breaking of the symmetry. The breaking of the full gauge symmetry down to SU(3)$\times$SU(2)$\times$U(1) is accomplished by introducing Higgs fields to produce the needed spontaneous symmetry breaking. (These fields are different from the Higgs field of the standard model, but they are also called Higgs fields because they play a completely analogous role.)

We are not attempting a full description of any of these topics, but I will explain how the three groups can be embedded in one larger group. There are many ways to do it, but the simplest is to embed them in SU(5), the group of 5×5 unitary matrices with determinant one. This was the original grand unified theory, proposed in 1974 by Howard Georgi and Sheldon L. Glashow of Harvard.* To do this, we can let the SU(3) subgroup of SU(5) be the set of matrices of the form

$$g_3 = \begin{pmatrix} x & x & x & 0 & 0 \\ x & x & x & 0 & 0 \\ x & x & x & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} , \tag{9.14}$$

where the 3×3 block of x's represents an arbitrary SU(3) matrix. Similarly the SU(2) subgroup can be described by matrices of the form

$$g_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & x & x \\ 0 & 0 & 0 & x & x \end{pmatrix} , \tag{9.15}$$

where this time the 2×2 block of x's represents an arbitrary SU(2) matrix. Note that these matrices commute with matrices of the form shown above in Eq. (9.14). Finally, we need to find a set of U(1) matrices, complex phases, which commute with both classes of matrices described above. This can be done by setting

$$g_1 = \begin{pmatrix} e^{2i\theta} & 0 & 0 & 0 & 0 \\ 0 & e^{2i\theta} & 0 & 0 & 0 \\ 0 & 0 & e^{2i\theta} & 0 & 0 \\ 0 & 0 & 0 & e^{-3i\theta} & 0 \\ 0 & 0 & 0 & 0 & e^{-3i\theta} \end{pmatrix} , \tag{9.16}$$

where the factors of 2 and 3 in the exponents were chosen so that the determinant — in this case the product of the diagonal entries — is one, as it must be for an SU(5) matrix.

---

* H. Georgi and S. L. Glashow, "Unity of All Elementary-Particle Forces," *Phys. Rev. Letters*, vol. 32, pp. 438-441 (1974). Available from *Phys. Rev. Letters* at http://prl.aps.org/abstract/PRL/v32/i8/p438_1, from *Phys. Rev. Letters* with an MIT certificate as http://prl.aps.org.libproxy.mit.edu/abstract/PRL/v32/i8/p438_1, or it can be found for example at http://puhep1.princeton.edu/~kirkmcd/examples/EP/georgi_prl_32_438_74.pdf. The paper had a one sentence abstract: "Strong, electromagnetic, and weak forces are conjectured to arise from a single fundamental interaction based on the gauge group SU(5)." According to the INSPIRE database, it has been cited over 4,600 times.

In the SU(5) theory there is only one gauge interaction strength, while in the standard model there are three. The trick of relating one SU(5) interaction strength to the three interaction strengths of SU(3)×SU(2)×U(1) was a key step in the development of grand unified theories. The interaction strengths, it turns out, are not fixed constants, but vary with energy in a calculable way. When the three interaction strengths are extrapolated from the measured values to much higher energies, it is found that to a good approximation they meet, as shown in Figure 9.3.*

In Figure 9.3, α_1, α_2, and α_3 are the coupling strengths of the U(1), SU(2), and SU(3) interactions, respectively, as measured at low energies and extended to high energies according to the theory. (The α_i are related to the coupling constants g_i mentioned earlier by $\alpha_i \equiv g_i^2/4\pi$.) The horizontal axis shows the base-10 logarithm of the energy scale in GeV. The top graph shows the calculation for the standard model, while the bottom graph shows the more promising calculation for the Minimal Supersymmetric Standard Model, an extension of the standard model that incorporates supersymmetry. (Supersymmetry is a proposed, approximate symmetry that connects fermions to bosons and vice versa, which would connect each of the known particles to a partner that is slightly too massive to have yet been seen.) This graph is perhaps one of the strongest pieces of evidence for supersymmetry, and for grand unification. It implies a unification scale of about 10^{16} GeV.

The grand unified theory is constructed so that the spontaneous symmetry breaking gives masses of order 10^{16} GeV to those gauge bosons that represent interactions that are part of SU(5), but not part of the SU(3)×SU(2)×U(1) subgroup of the standard model. Energies of order 10^{16} GeV are totally unattainable; the LHC is designed to reach an energy of 7 TeV = 7,000 GeV per beam, or 14 TeV total. Nonetheless, we can speak theoretically about energies high compared to 10^{16} GeV, and then the spontaneous symmetry breaking of the grand unified theory would become unimportant. At such very high energies, the full gauge symmetry would become apparent. But at energies low compared to 10^{16} GeV, these 10^{16} GeV-scale particles would be too heavy to ever produce, and we would expect to see precisely the particle physics of the SU(3)×SU(2)×U(1) standard model.

The Magnetic Monopole Problem:

A magnetic monopole is a particle with a net North or South magnetic charge. The magnetic field of a monopole points radially outward (or inward), with a magnitude proportional to $1/r^2$, just like the Coulomb field of a point electric charge. Such particles

---

* Taken from the Particle Data Group 2016 Review of Particle Physics, C. Patrignani et al. (Particle Data Group), Chin. Phys. C, **40**, 100001 (2016), Chapter 16, *Grand Unified Theories*, Revised January 2016 by A. Hebecker and J. Hisano, http://www-pdg.lbl.gov/2016/reviews/rpp2016-rev-guts.pdf.

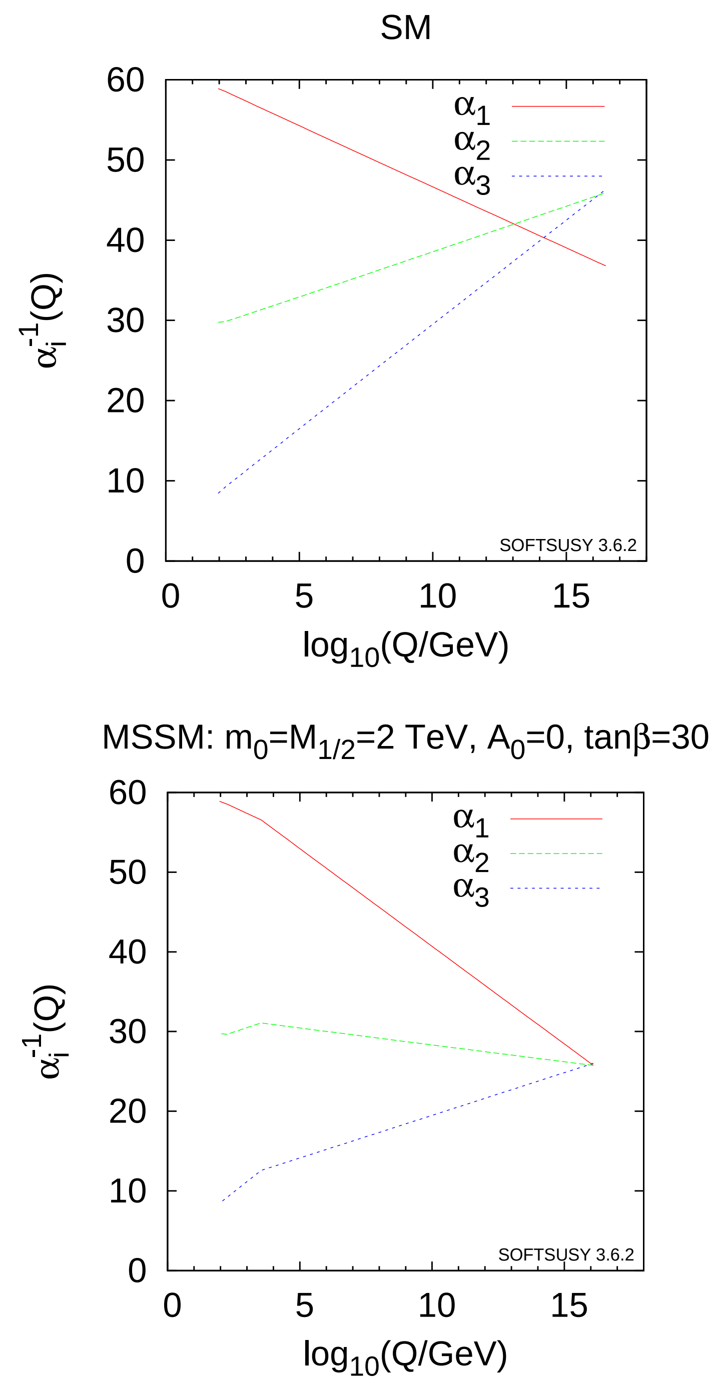


Figure 9.3: Running of the U(1), SU(2), and SU(3) interaction strengths with energy, in the standard model (SM) of particle physics and in the minimal supersymmetric standard model (MSSM). The horizontal axis is $\log_{10} Q$, where Q is the energy in GeV.

do not exist in the usual formulation of electromagnetism, in which all magnetic effects arise from electric currents. An ordinary bar magnet, with internal currents associated with the alignment of electronic orbits, has the form of a dipole, with North and South poles at the two ends. If a bar magnet is cut in half, one obtains two dipoles, each with a North and South pole. Grand unified theories (GUTs), however, imply that magnetic monopoles necessarily exist. They are generally superheavy particles, with mass energies of approximately 10^{18} GeV. That is, they are about two orders of magnitude heavier than the unification scale.

The magnetic monopoles of grand unified theories are constructed as twists, or knots, in the GUT Higgs fields, so the production of magnetic monopoles is closely linked to the behavior of the GUT Higgs fields as the early universe expanded and cooled. More technically, these knots in the GUT Higgs fields are called *topological defects.*

At high temperatures these Higgs fields will undergo large thermal fluctuations. In many grand unified theories the high temperature thermal equilibrium state is one in which the values of the fields average to zero, which means that the GUT symmetry is unbroken. As the system cools a phase transition is encountered. A phase transition is characterized by a specific temperature, called the critical temperature, at which some thermal equilibrium properties of the system change discontinuously. In this case, at temperatures below the critical temperature, some subset of the Higgs fields acquire nonzero mean values in the thermal equilibrium state — the GUT symmetry is thereby spontaneously broken. There may be one or perhaps several such phase transitions before the system reaches the lowest temperature phase — the phase which includes the vacuum. For simplicity, we will discuss the case in which there is only one such phase transition. In any case, the broken symmetry state which exists below the critical temperature is not unique, for precisely the same reason that the vacuum state is not unique.

In the conventional cosmological model, it is assumed that this phase transition occurs quickly once the critical temperature is reached. Thus, in any given region of space the Higgs fields will settle into a broken symmetry state, in which some subset of the Higgs fields acquire nonzero mean values. The choice of this subset is made randomly, just as the orientation of the axes of a crystal are determined randomly when the crystal first starts to condense from a molten liquid. The other particles in the theory, such as the quarks and leptons, are also described by fields, which interact with the Higgs fields in a manner consistent with the GUT symmetry. Through these interactions, the randomly selected combination of nonzero Higgs fields determines what combination of the fields will act like an electron, what combination will act like a u-quark, etc. The same random choice determines what combination of vector boson fields will act like the photon field, and what combinations will act like the W's, Z's, or gluons. In addition, some vector bosons acquire masses of the order of 10^{16} GeV, and these vector bosons are then irrelevant to the low energy physics which we observe in present-day accelerator experiments.

As mentioned above, the magnetic monopoles are examples of defects which form in the phase transition. The defects arise when regions of the high temperature symmetric phase undergo a transition to different broken-symmetry states. In the analogous situation when a liquid crystallizes, different regions may begin to crystallize with different orientations of the crystallographic axes. The domains of different crystal orientation grow and coalesce, and it is energetically favorable for them to smooth the misalignment along their boundaries. The smoothing is often imperfect, however, and localized defects remain.

The detailed nature of these defects is too complicated to explain here, so I will settle for the statement of some general facts. There are three types of defects that can occur. The simplest type is a surfacelike defect called a domain wall. This type of defect arises whenever the broken-symmetry state in one region of space cannot be smoothly deformed into the broken-symmetry state in a neighboring region of space. A domain wall then forms at the interface between the two regions. Some grand unified theories allow for the formation of such domain walls, and others do not. The second type is a linelike defect called a cosmic string. Again, some grand unified theories allow such defects to exist, and others do not. Finally, the third type is a pointlike defect, called a magnetic monopole. In contrast to the first two types of defects, magnetic monopoles exist in ***any*** grand unified theory.

To see how a pointlike defect can arise, let us consider the simplest theory in which they occur. This theory is too simple to describe the real world, but it serves as a "toy" model which is useful to illustrate many features of spontaneously broken gauge theories. The theory has a three-component multiplet of Higgs fields, which I will denote by ϕ_a, where $a = 1, 2$, or 3. The symmetry which operates on this multiplet is identical in its mathematical form to the transformations that describe how the three components of an ordinary vector are modified by a rotation. The potential energy density associated with the Higgs fields is then a function of the three components ϕ_a. The energy density function, however, is an ingredient of the fundamental theory which must be invariant under the symmetry. Thus, the energy density can depend only on

$$|\phi| \equiv \sqrt{\phi_1^2 + \phi_2^2 + \phi_3^2}\ . \tag{9.17}$$

The potential energy density for this field will be assumed to have the general form shown earlier for $|H|$, so that spontaneous symmetry breaking ensues. The energy density will be minimized when $|\phi|$ has some particular nonzero value, which we will call ϕ_v. Now consider the following static configuration of the Higgs field:

$$\phi_a(\vec{r}) = f(r)\hat{r}_a\ , \tag{9.18}$$

where $r \equiv |\vec{r}\,|$, $\hat{r}_a$ denotes the a-component of the unit vector $\hat{r} = \vec{r}/r$, and $f(r)$ is a function which vanishes when $r = 0$ and approaches ϕ_v as $r \to \infty$. This configuration is sketched below in Fig. 9.4. An arrow is drawn at each point in space, and the three vector components of the arrow are used to represent the three components of the Higgs field:

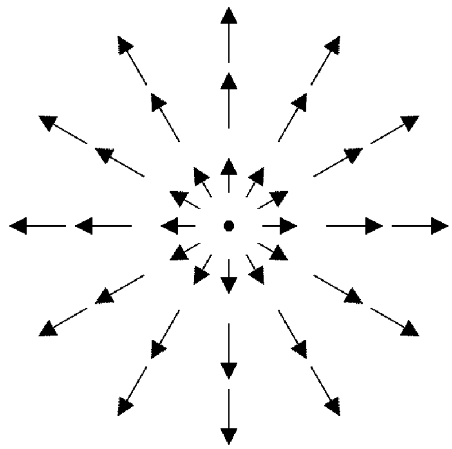


Figure 9.4: Graphical represention of the three-component Higgs field in the vicinity of a magnetic monopole.

If the diagram were constructed as a three-dimensional model, then all of the arrows would point radially outward from the origin. (An antimonopole is described by a similar picture, except that the arrows would point radially inward.) Note that the index a on ϕ_a normally has nothing to do with any direction in physical space — ϕ_1, ϕ_2, and ϕ_3 are just three scalar fields. Their behavior is related by a symmetry of fields — the gauge symmetry — but this symmetry is unrelated to the symmetry of rotations in physical space. Nonetheless, each field ϕ_a is allowed to be an arbitrary function of position, so there is nothing to prevent the fields from assuming the form of Eq. (9.18), as illustrated in Fig. (9.4).

The Higgs fields for the monopole configuration are in a vacuum state at large distances, but the fields differ from their vacuum values in the vicinity of $r = 0$, resulting in a concentration of energy. It can be proven that this configuration is "topologically stable" in the following sense: if the boundary conditions for the fields at infinity are held fixed, and if the fields are required to be continuous functions of position, then there must always be at least one point at which all three components of the Higgs field vanish. I will not attempt to prove this theorem, but I recommend that you stare at the diagram until the theorem becomes believable. Because of this topological property of the magnetic monopole configuration, it is sometimes referred to as a "knot" in the Higgs field. The configuration involves a concentration of energy localized around a point, and it behaves exactly as a particle.

So far I have not mentioned anything about magnetic fields, so the astute reader is no doubt wondering why these particles are called magnetic monopoles. For present purposes the important property is that these objects are topologically stable knots in the Higgs fields, but in fact they must have a net magnetic charge. The reason comes from energy considerations. In the absence of any other fields, the energy of the magnetic monopole Higgs field configuration would be infinite. To understand this infinity, you must accept without proof the fact that the expression for the energy density of a Higgs field contains a term proportional to the square of the gradient. The form of Eq. (9.18) for large r (with $f(r) \to \phi_v$) then implies that the gradient of ϕ_a falls off as $1/r$ at large distances. The total energy within a large sphere is therefore proportional to

$$4\pi \int r^2 \, \mathrm{d}r \left(\frac{1}{r}\right)^2$$

and therefore diverges linearly with the radius of the sphere. However, the expression for the energy density becomes more complicated when the vector boson fields are included. It is beyond what we have time to discuss here, but it can be shown that the total energy of the Higgs field configuration of Eq. (9.18) can be made finite only if the configuration includes vector boson fields that correspond to a net magnetic charge. (The usual $\vec{\nabla} \cdot \vec{B} = 0$ equation holds far away from the center of the monopole, where $|\phi|$ is very close to ϕ_v, but $\vec{\nabla} \cdot \vec{B}$ can be nonzero near the center of the monopole, where the Higgs fields become small and the other gauge fields become part of the dynamics.) Even the magnitude of this magnetic charge is determined uniquely. The magnetic charge must correspond to a value $1/(2\alpha)$ times the electric charge of an electron. Here α denotes the usual fine structure constant of electrodynamics: $\alpha = e^2/\hbar c$ in cgs units, or $\alpha = e^2/4\pi\epsilon_0\hbar c$ in SI (mks) units. In any case, $\alpha \approx 1/137$. This means that the magnetic charge of a monopole is 68.5 times as large as the electric charge of an electron, and the force between two monopoles is then $(68.5)^2$ times as large as the force between electrons at the same distance.

The mass of a monopole can be estimated in these models, and it turns out to be extraordinary. The mass is approximately $1/\alpha$ times the mass scale at which the unification of forces occurs. Since the unification of forces occurs roughly at 10^{16} GeV, it follows that Mc^2 for a monopole is about 10^{18} GeV.

Having gone through the basic physics, we are now in a position to discuss how one estimates the number of magnetic monopoles that would be produced in the GUT phase transition. I will present a crude argument which is probably accurate to within one or two orders of magnitude. Although the argument will be crude, there is really no need to carry out a more accurate calculation. The magnetic monopole problem is so severe than an ambiguity of two orders of magnitude in the estimate is unimportant to the conclusion.

Recall that the monopoles are really knots in the Higgs field, so their number density is related to the misalignment of the Higgs field in different regions of space. This misalignment can be characterized by a “correlation length” ξ. We will need only an approximate definition of this correlation length, so it will suffice to say that ξ is the minimum length such that the Higgs field at a given point in space is almost uncorrelated with the Higgs field a distance ξ away. One then estimates that the number density of magnetic monopoles and antimonopoles is given roughly by

$$n_M \approx 1/\zeta^3 \; . \tag{9.19}$$

In words, we are estimating that every cube with a side of length ξ will have, on the average, approximately one magnetic monopole in it. This estimate was first proposed by T.W.B. Kibble of Imperial College (London).

The remaining problem is to estimate ξ. Here we will be working in the context of conventional cosmology, in which it is assumed that the phase transition occurs quickly once the critical temperature is reached. Under these assumptions the phase transition has no significant effect on the evolution of the early universe. When the universe cools below the critical temperature T_c of the GUT phase transition (with $kT_c \approx 10^{16}$ GeV), it becomes thermodynamically probable for the Higgs field to align uniformly over reasonably large distances. If the system were allowed time to reach thermal equilibrium, then very few monopoles would be present — their abundance would be suppressed by the usual Boltzmann factor

$$e^{-Mc^2/kT}$$

from statistical mechanics. For this case the factor is roughly $e^{-100} \approx 10^{-43}$. However, if the whole process must happen on the time scales at which the early universe evolves, then there is not enough time for this long range correlation of the Higgs field to become established. While we are not prepared to calculate the correlation length in these circumstances, we can safely say that the correlation length must be less than the horizon distance — this statement assumes only that the correlation of the Higgs field requires the transmission of information, and special relativity implies that information cannot propagate faster than the speed of light. It is then a straightforward calculation, which you will do in Problem Set 9, to find a lower bound on the number of monopoles that would have been produced at the GUT phase transition under these assumptions.

If you do the problem right, you should find that the contribution to Ω today, from monopoles, would be bigger than 10^{20}, according to this calculation.

This number is obviously unacceptable, but one way to drive this point home is to consider the age of the universe. As you recall, a large value of Ω implies that the universe slowed down rapidly to its present expansion rate, giving a low predicted age for the universe. The formula for the age of the universe was derived in Lecture Notes 4, assuming that it can be approximated as being matter-dominated throughout its evolution. Since the monopoles would behave as nonrelativistic matter, this would be an excellent approximation here. For $\Omega > 2$, during the expanding phase, the age is given by

$$t = \frac{\Omega}{2H(\Omega-1)^{3/2}} \left\{ \sin^{-1}\left(\frac{2\sqrt{\Omega-1}}{\Omega}\right) - \frac{2\sqrt{\Omega-1}}{\Omega} \right\} ,$$

where the inverse sine function is to be evaluated in the range $\frac{\pi}{2}$ to π. For very large Ω the inverse sine function approaches π, and the age is approximated by

$$t = \frac{\pi}{2H\sqrt{\Omega}} . \tag{9.20}$$

Taking Ω as 10^{20}, the age turns out to be 2.2 years. The prediction that you will find will be even smaller than that, since you should find a value of Ω bigger than 10^{20}.

Thus if grand unified theories are correct — which is plausible but not necessarily true — then we have another serious problem for the conventional hot big bang model. The universe, after all, is certainly more than 2.2 years old!

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe December 6, 2018
Prof. Alan Guth

Lecture Notes 10
THE NEW INFLATIONARY UNIVERSE

INTRODUCTION:

The new inflationary universe is a scenario in which the mass density of at least a small patch of the early universe becomes dominated by the potential energy of a scalar field, in a state which is sometimes called a *false vacuum.* This peculiar form of energy leads to a negative pressure, and hence a repulsive gravitational force, driving the region into a period of exponential expansion, during which it expands by many orders of magnitude — hence the name "inflationary". The word "new" refers to a modification of my original proposal[1] which was suggested independently by Linde[2] and by Albrecht and Steinhardt.[3] They suggested a new mechanism by which the exponential expansion phase could be ended, solving some crucial problems that existed in my original proposal. The inflationary model is very attractive because it offers possible solutions to the horizon/homogeneity problem, the flatness problem, and the magnetic monopole problem, which were discussed in Lecture Notes 8 and 9. It also predicts that the universe should be flat to high accuracy, a fact which has now been verified to an accuracy of 0.4%.[4] In addition, inflationary models give predictions for the properties of the small ripples that are observed in the cosmic microwave background (CMB) radiation. As we will discuss at the end of these notes, the predictions of the simplest inflationary models are beautifully in agreement with what has been measured. If inflation is correct, it would mean that particle physics mechanisms are responsible for the production of essentially all the matter, energy, and entropy in the observed universe.

---

[1] A. H. Guth, "The inflationary universe: A possible solution to the horizon and flatness problems," *Physical Review D*, vol. 23, pp. 347–356 (1981), available at http://prd.aps.org/abstract/PRD/v23/i2/p347_1, or in its original preprint form at http://slac.stanford.edu/pubs/slacpubs/2500/slac-pub-2576.pdf.

[2] A. D. Linde, "A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems," *Physics Letters B*, vol. 108, pp. 389–393 (1982), available at http://www.sciencedirect.com/science/article/pii/0370269382912199.

[3] A. Albrecht and P. J. Steinhardt,"Cosmology for grand unified theories with radiatively induced symmetry breaking," *Physical Review Letters,* vol. 48, pp. 1220–1223 (1982), available at http://prl.aps.org/abstract/PRL/v48/i17/p1220_1, or with an MIT certificate at http://prl.aps.org.libproxy.mit.edu/abstract/PRL/v48/i17/p1220_1.

[4] P. A. R. Ade et al. (Planck Collaboration), "Planck 2015 results, XIII: Cosmological parameters," Table 5, Column 6, Astronomy & Astrophysics vol. 594, article A13 (2016), arXiv:1502.01589.

SCALAR FIELDS AND THE FALSE VACUUM:

The (original) inflationary universe scenario was developed to solve the magnetic monopole problem, but it quickly became clear that the scenario might solve all three of the problems discussed in Lecture Notes 8 and 9. The scenario contained the basic ingredients necessary to eliminate these problems, but unfortunately the scenario also contained one fatal flaw: the exponential expansion was terminated by a phase transition that occurred by the random nucleation of bubbles of the new phase, very similar to the way that water boils. It was found that this violent boiling would lead to a grossly inhomogeneous universe that looks nothing like our universe. This difficulty — which came to be called the *graceful exit problem* — was summarized in the original paper[1] (in a section credited to Erick Weinberg and Harry Kesten as well as me), and was later discussed in detail by Erick Weinberg and me[5] and by Stephen Hawking, Ian Moss, and John Stewart.[6] Fortunately, the graceful exit problem is completely avoided in a variation known as the new inflationary universe, developed independently by Andrei Linde[2] (then at the Lebedev Physical Institute in Moscow, now at Stanford) and by Andreas Albrecht and Paul Steinhardt.[3] (Albrecht and Steinhardt were both at the University of Pennsylvania at the time of their discovery; now Albrecht is at UC Davis, and Steinhardt is at Princeton.)

At the end of these notes I will also briefly describe chaotic inflation, a version of inflation proposed by Linde in 1983.[7] There are now hundreds of versions of inflation, but they are essentially all variants of new or chaotic inflation.

In order for the new inflationary scenario to occur, the underlying particle theory must contain a scalar field ϕ. The potential energy function $V(\phi)$, which represents the potential energy per unit volume, must have a plateau. This plateau is usually taken to be at $\phi \approx 0$, and $\phi = 0$ is usually assumed to be a local maximum of $V(\phi)$. $V(\phi)$ must be very flat in the vicinity of $\phi = 0$. In the example shown below, V is assumed to depend only on $|\phi|$.

---

[5] A. H. Guth and E. J. Weinberg,"Could the universe have recovered from a slow first order phase transition?" *Nuclear Physics B,* vol. 212, pp. 321–364 (1983), available at http://www.sciencedirect.com/science/article/pii/0550321383903073, or with an MIT certificate at http://www.sciencedirect.com.libproxy.mit.edu/science/article/pii/0550321383903073.

[6] S. W. Hawking, I. G. Moss, and J. M. Stewart,"Bubble collisions in the very early universe," *Physical Review D,* vol. 26, pp. 2681–2693 (1982), available at http://prd.aps.org/abstract/PRD/v26/i10/p2681_1, or with an MIT certificate at http://prd.aps.org.libproxy.mit.edu/abstract/PRD/v26/i10/p2681_1.

[7] Andrei D. Linde, "Chaotic Inflation," Physics Letters, vol.~129B, pp.~177-181 (1983).

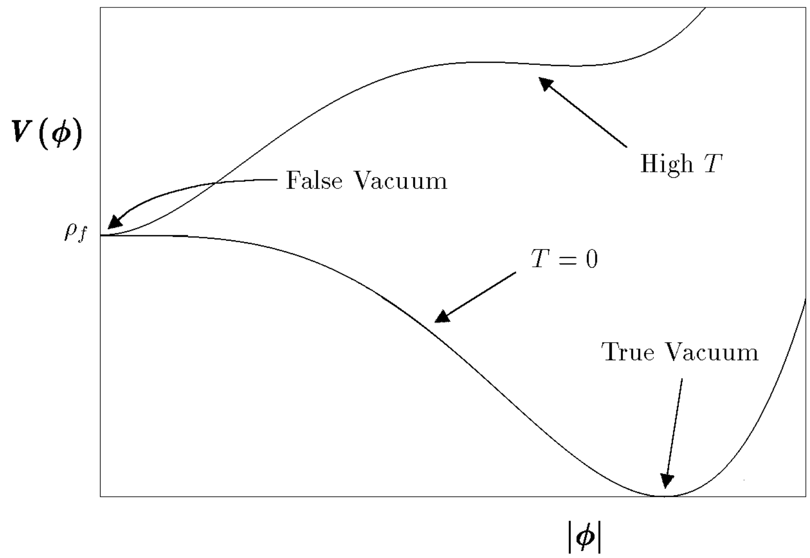


Figure 10.1: The potential energy function (labeled "$T = 0$") and the high-temperature finite-temperature effective potential (labeled "High T") for a scalar field that could drive new inflation.

To discuss the issues one at a time, I will first discuss the physical properties of a scalar field of the type described in the previous paragraph, and then we will consider the role that such a field might play in the early universe.

In most theoretical models of this type, one finds that at high temperature T the thermal equilibrium value of ϕ lies at $\phi = 0$. At high temperatures the field will actually fluctuate wildly, but in most theoretical models the ***average*** value is predicted to be zero. A potential energy function of this general form is shown as Figure 10.1. The curve labeled "High T" is a graph of what is called the finite-temperature effective potential, which is actually a graph of the free energy per unit volume; it will not be important for us to know exactly what free energy is, but to interpret the graph we should keep in mind that the free energy is minimized in the thermal equilibrium state. For these purposes we can treat the energy density (and pressure) of the true vacuum as zero, even though we learned in Lecture Notes 7 that they are apparently not. Taking $\Omega_{\text{vac}} = 0.691$ and $h_0 = 0.677$ from Table 7.1 of Lecture Notes 7 and using Eq. (3.34) from Lecture Notes 3 for the critical mass density, we find that the vacuum energy density of our universe is about

$$\begin{aligned}\rho_{\text{vac}} &= \Omega_{\text{vac}}\rho_c \approx 0.691 \times 1.88 \times (0.677)^2 \times 10^{-26}\ \text{kg/m}^3 \\ &= 5.95 \times 10^{-27}\ \text{kg/m}^3 = 5.95 \times 10^{-30}\ \text{g/cm}^3\ .\end{aligned} \tag{10.1}$$

We will soon see that this number is totally negligible compared to the huge energy densities that we expect for early universe inflation.

The scalar field ϕ that drives the inflation was originally taken to be the Higgs field of a grand unified theory, but it now seems very unlikely that his could work. The Higgs fields are required to have relatively strong interactions in order to induce spontaneous symmetry breaking, which is why the Higgs fields were introduced in the first place. These interactions generically lead to large quantum fluctuations in the evolution of the field, which in turn lead to unacceptably large inhomogeneities in the mass density of the universe. Most inflationary models assume, therefore, the existence of another scalar field, similar to the Higgs field but much more weakly interacting. This field is usually called the *inflaton*.[8]

So, in thermal equilibrium at high temperatures, one expects the scalar field to have a mean value around zero. If the system cools, the thermal excitations will disappear, and the scalar field will find itself in a state of essentially zero temperature, with $\phi \approx 0$. This state is called the false vacuum, and its peculiar properties are the driving force behind the inflationary model.

The false vacuum is clearly unstable, as ϕ will not remain forever at a local maximum of $V(\phi)$. However, if $V(\phi)$ is sufficiently flat, then the time that it takes for ϕ to move away from $\phi = 0$ can be very long compared to the time scale for the evolution of the early universe. Thus, for these purposes the false vacuum can be considered metastable. Furthermore, while ϕ remains near zero, the energy density remains fixed near $V(0)$, and cannot be lowered even if the universe is expanding. It is this property that motivates the name, "false vacuum." To a particle physicist, the vacuum is defined as the state of lowest possible energy density. The adjective "false" is used here to mean "temporary," so a false vacuum is a state which temporarily has the property that its energy density cannot be lowered.[9]

Since the false vacuum has $\phi = 0$ and no other excitations, the mass density has a fixed value which is determined by the potential energy function $V(\phi)$. For a typical

---

[8] While it had been thought for many years that the inflaton could under no circumstances be a Higgs field of any sort, in 2008 Fedor Bezrukov and Mikhail Shaposhnikov proposed that the Higgs field of the standard model of particle physics could serve as the inflaton, if one assumed that in addition to its known interactions, it also has "non-minimal" interactions with gravity — i.e., interactions beyond what is required by the equivalence principle. See F. Berukov and M. Shaposhnikov, "The Standard Model Higgs Boson as the inflaton," Physics Letters B, vol. 659, pp. 703–706 (2008), arXiv:0710.3755 [hep-th].

[9] Historically, the phrase "false vacuum" was first used to refer to a state in which the scalar field was at a local minimum of the potential energy function, so the state could decay only by quantum mechanical tunneling. Here I have stretched the definition a bit, using the phrase to describe a scalar field which, although still quite stable, is near a local maximum of the potential energy function.

grand unified theory, this value can be estimated in terms of the GUT energy scale $E_{\rm GUT} \approx 10^{16}$ GeV by using dimensional analysis:

$$\rho_{\rm f} \approx \frac{E_{\rm GUT}^4}{\hbar^3 c^5} = 2.3 \times 10^{84}\,{\rm kg/m^3} = 2.3 \times 10^{81}\,{\rm g/cm^3}\ . \tag{10.2}$$

(Thus the energy density of our vacuum, estimated in Eq. (10.1), is smaller by more than 100 orders of magnitude.)

The pressure p of the false vacuum is completely determined by the fact that, on the time scales of interest, its energy density cannot be lowered. To see that a constant energy density implies a negative pressure, remember the conservation of energy equation derived in Problem 4 of Problem Set 6:

$$\dot\rho = -3\frac{\dot a}{a}\left(\rho + \frac{p}{c^2}\right)\ . \tag{10.3}$$

If $\dot\rho = 0$, this equation implies immediately that

$$\boxed{p = -\rho_{\rm f} c^2\ .} \tag{10.4}$$

(We used this same argument in Lecture Notes 7, when we were discussing vacuum energy density and the cosmological constant.)

To understand this result from first principles, think of an imaginary piston that is filled with false vacuum and surrounded by ordinary true vacuum, as shown below, in Fig. 10.2:

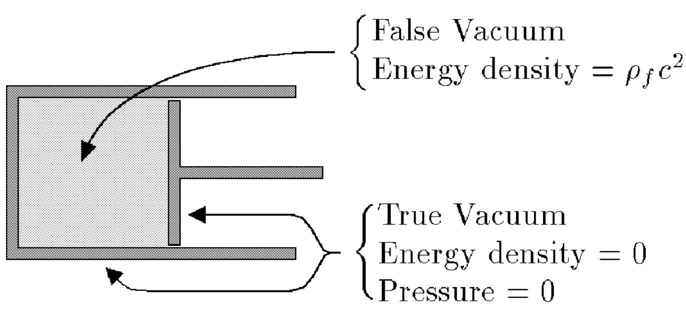
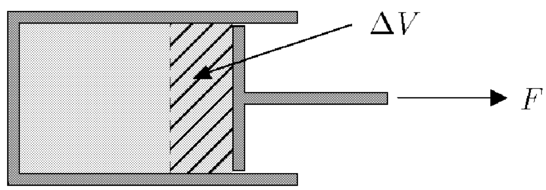


Figure 10.2: A piston used for a thought experiment to show that the pressure of a false vacuum state is the negative of its energy density.

Since this is a thought experiment, we can imagine that the "true vacuum" outside the piston genuinely has zero energy density and zero pressure. If one prefers not to be so imaginative, the energy density and pressure of our vacuum are in any case totally negligible on the scales that are relevant here. Suppose now that the piston is pulled out so that the volume of the chamber increases by ΔV. We assume that the walls of the box are designed to guarantee that the region inside remains completely filled with false vacuum. The energy of the system then increases by $\rho_{\rm f} c^2 \Delta V$, and therefore the agent that moved the piston must have done precisely this amount of work.



$$\Delta W = \rho_f c^2 \, \Delta V = -p \, \Delta V$$

Figure 10.3: The piston of the thought experiment is pulled out, enlarging the chamber. The energy density of the false vacuum inside the chamber is fixed, so the energy in the chamber goes up. The energy must come from the agent that pulled on the piston. For the agent to do positive work, the pressure inside the chamber must be negative.

Since the pressure on the outside is zero, the agent must be pulling against a negative pressure, which would oppose the motion. Quantitatively, since the work done is $-p\Delta V$, it follows that $p = -\rho_{\rm f} c^2$, confirming the previous result.

The large negative pressure creates a gravitational repulsion, exactly as we discussed in Lecture Notes 7 in the context of a cosmological constant. The gravitational repulsion can be seen in the second order differential equation for a, the second order Friedmann equation,

$$\ddot{a} = -\frac{4\pi}{3} G \left(\rho + \frac{3p}{c^2} \right) a \ , \tag{10.5}$$

which implies that both the pressure and the energy density normally contribute to the slowing of the cosmic expansion. For the false vacuum, however, the large negative pressure leads to $\rho + 3p/c^2 < 0$, and it follows that $\ddot{a}$ is ***positive***. The false vacuum creates a gravitational repulsion which causes the growth of the scale factor a to accelerate. It is this repulsion which will drive the colossal expansion of the inflationary scenario. The equations are the same as those for a cosmological constant, except that the false vacuum energy density disappears when the scalar field rolls off the hill in the potential energy diagram, while the vacuum energy associated with a cosmological constant is permanent.

THE NEW INFLATIONARY UNIVERSE:

We can now go through the new inflationary scenario step by step. The starting point of a cosmological scenario is, unfortunately, still somewhat a matter of taste and philosophical prejudice. Some physicists find it plausible to assume that the universe began in some highly symmetrical state. Many others, however, consider it more likely that the universe began in a highly chaotic state, since the number of chaotic configurations is presumably much larger. One advantage of the inflationary scenario, from my point of view, is that it appears to allow a wide variety of starting configurations.

We can begin by discussing what would happen if the early universe were in thermal equilibrium, at least in the sense of having regions of approximately horizon size in which thermal equilibrium held. In that case, inflation could begin if the universe was hot ($kT > 10^{16}$ GeV) in at least some of these regions, and if at least one of these hot regions were expanding rapidly. In the hot regions, thermal equilibrium would imply $<\phi> = 0$, where $<\phi>$ denotes the mean value of the field ϕ as it undergoes its thermal fluctuations. Rapid expansion would cause these regions to cool, and the scalar field would settle down to a cool state in which the field is trapped on the plateau of the potential energy hill. The expansion must be rapid enough so that the cooling of the scalar field occurs before the region recollapses under the influence of gravity.

Thermal equilibrium would make things simple, but we said earlier that the inflaton field must interact very weakly, to avoid generating overly large quantum fluctuations. For such a weakly interacting field, a fairly straightforward calculation of collision rates shows that the mean time between collisions would be long compared to the age of the universe at the onset of inflation. Thus there is no compelling reason to assume thermal equilibrium, although — in the absence of a theory that fixes the initial conditions — one could assume anything one wants. For inflation to start, the minimal assumption would be that there existed at least some regions of high energy density with $<\phi> \approx 0$, and that at least one of these regions was expanding rapidly enough so that ϕ became trapped in the false vacuum.

The above paragraphs describe the new inflationary universe with a hot beginning, but there are certainly other possibilities. Linde has also proposed the idea of chaotic inflation,[7] in which inflation is driven by a scalar field which is initially chaotic but far from thermal equilibrium. In this scenario inflation happens while the scalar field rolls down a gentle hill in the potential energy diagram, so the potential energy diagram need not have a plateau. Alexander Vilenkin[10] (of Tufts University) and Linde[11] have separately investigated speculative but attractive scenarios in which the universe is created by

[10] A. Vilenkin, "The Birth of Inflationary Universes," Physical Review D, vol. 27, p. 2848 (1983). With an MIT certificate, click here.

[11] A. D. Linde, "Quantum creation of the inflationary universe," Lettere al Nuovo Cimento, vol. 39, pp. 401-405 (1984) With an MIT certificate, click here.

a quantum tunneling event, starting from a state of absolutely nothing. In these models the universe enters directly into a de Sitter phase. In a similar spirit James Hartle (of the University of California at Santa Barbara) and Stephen Hawking (of Cambridge University) have proposed a unique quantum wave function for the universe,[12] incorporating dynamics which leads to an inflationary era.

Although a wide variety of scenarios have been proposed to describe the onset of inflation, an important feature of inflation is that all these scenarios lead to similar if not identical predictions. Once inflation starts, the colossal expansion dilutes away the evidence of how it began. Later I will discuss the phenomenon of *eternal inflation,* which carries this idea of dilution to an extreme. We will see that for almost all inflationary models, once inflation starts, it never stops. Instead it goes on producing "pocket universes" forever. This eternal aspect of inflation presumably erases all traces of how inflation began, and it also obviates the question of whether the conditions leading to inflation are likely. As long as the probability that inflation can start is nonzero, and as long as there is no other mechanism that can compete, it appears (at least to this author) that there are no other questions about initial conditions that need to be answered. An ultimate theory of the origin of the universe would still be very interesting, intellectually, but most likely it would not affect in any way the consequences of inflation.

To continue with the description of the new inflationary scenario, we assume that there exists a region which is sufficiently homogeneous, isotropic, and flat to be described by a flat Robertson–Walker metric

$$ds^2 = -c^2 dt^2 + a^2(t) d\vec{x}^{\,2} \ , \tag{10.6}$$

and the equation of motion becomes

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3} G\rho \ . \tag{10.7}$$

The solution is given by

$$\boxed{a(t) = \text{const} \times e^{\chi t} \ ,} \tag{10.8}$$

where

$$\boxed{\chi = \sqrt{\frac{8\pi}{3} G\rho_{\text{f}}} \ .} \tag{10.9}$$

[12] J. B. Hartle and S. W. Hawking, "Wave function of the universe," Physical Review D, vol. 28, p. 2960 (1983). With an MIT certificate, click here.

This exponential expansion is of course the hallmark of the inflationary model. (For our parameters, $\chi^{-1} \approx 10^{-37}$ sec.) Such a space is called a de Sitter space.

We of course cannot expect to find a region of the early universe that is exactly homogenous, isotropic, and flat, so it is important to know that it is enough to come close. As long as a region meets these criteria approximately, the behavior will be governed by what has been called the *cosmological no-hair conjecture*,[13] which holds that the region will evolve so that it locally resembles exact de Sitter space. As long as $p = -\rho c^2 =$ *constant*, which will hold as long as the scalar field ϕ is near its false vacuum value, the space will start to expand and any initial particle density will be diluted. Any initial distortion of the metric is stretched (i.e., redshifted) until it is no longer locally detectable. This behavior can be proven quite generally in a linearized perturbation analysis, and has also been seen to hold in some specific solutions with large perturbations. There is no proof that the early universe must contain regions that start inflating, but it seems very plausible.

An important property of a de Sitter region, which helps to ensure its durability, is the presence of *event horizons*. These are different from the horizons that we have been discussing since Lecture Notes 4, which are technically called *particle horizons*, and refer to the possibility that two objects can be far enough apart so that light from one object would not have had enough time since the big bang to reach the other. The event horizon of de Sitter space can be seen by calculating, in the metric described by Eqs. (10.6) and (10.8), the coordinate distance that light can travel between times t_1 and t_2:

$$\Delta r(t_1, t_2) = \int_{t_1}^{t_2} \frac{c}{a(t)}\,\mathrm{d}t = \frac{c}{\mathrm{const}} \int_{t_1}^{t_2} e^{-\chi t}\,\mathrm{d}t = \frac{c}{\mathrm{const}\,\chi}\left[e^{-\chi t_1} - e^{-\chi t_2}\right] \ . \tag{10.10}$$

The point is that this distance is limited even as $t_2 \to \infty$. Note that

$$\lim_{t_2 \to \infty} a(t_1)\,\Delta r(t_1, t_2) = c\chi^{-1} \ . \tag{10.11}$$

Physically, this means that if two objects at rest in these coordinates are separated by a physical distance more than $c\chi^{-1}$, a light pulse emitted by one object will never reach the other. This in turn means that if a de Sitter region is large compared to $c\chi^{-1}$, then the effect of inhomogeneities from outside the region cannot penetrate into the region any further than a shell of thickness $c\chi^{-1}$. Once the de Sitter region is large compared to $c\chi^{-1}$, it is impervious to outside influences.

---

[13] R. M. Wald, "Asymptotic behavior of homogeneous cosmological models in the presence of a positive cosmological constant," *Physical Review D,* vol. 28, pp. 2118–2120 (1983), available at http://prd.aps.org/abstract/PRD/v28/i8/p2118_1, or with an MIT certificate at http://prd.aps.org.libproxy.mit.edu/abstract/PRD/v28/i8/p2118_1.

As the inflating region continues to exponentially expand, the mass density of the inflaton field is fixed at ρ_f . Thus, the total energy of the inflaton field is increasing! If the inflationary model is right, the energy of the inflaton field is the source of essentially all the matter, energy, and entropy in the observed universe.

This creation of energy seems to violate our naive notions of energy conservation, but we must remember that there is also an energy associated with the cosmic gravitational field— the field by which everything in the universe is attracting everything else, thereby slowing down the cosmic expansion. Even in Newtonian mechanics one can see that the energy density of a gravitational field is ***negative***. To see this, note that the gravitational field is strengthened as one brings masses together from infinity, but the potential energy of the system is lowered as objects are brought together under the influence of the attractive force. Thus the stronger field corresponds to a lower energy. A good analogy is the electrostatic field, since Coulomb's law is very similar to Newton's law. By calculating how much work needs to be done in pushing charges to create a specified configuration of a static electric field, it is possible to show that the energy density stored in an electric field is given by

$$u_{\text{electrostatic}} = \frac{1}{2}\epsilon_0 \left|\vec{E}\right|^2 \tag{10.12a}$$

or

$$u_{\text{electrostatic}} = \frac{1}{8\pi} \left|\vec{E}\right|^2 \ , \tag{10.12b}$$

depending on what units you are using. The calculation for Newtonian gravity is essentially identical, giving

$$\boxed{u_{\text{Newton}} = -\frac{1}{8\pi G} \left|\vec{g}\right|^2 \ .} \tag{10.13}$$

The sign difference arises from the sign difference in the force law: two positive charges repel, while two positive masses attract. In the context of inflation, the energy stored in the gravitational field becomes more and more negative as the universe inflates, while the energy stored in "matter" (everything except gravity) becomes more and more positive. The total energy remains constant, and very small— perhaps it is exactly equal to zero.

After the region has undergone exponential expansion for some time, inflation must somehow end, at least in the region that is going to describe our visible universe. The scalar field is in an unstable configuration, perched at the top of the hill of the potential energy diagram of Fig. 10.1. It will undergo fluctuations due to thermal and/or quantum effects. Some fluctuations begin to grow, and at some point these fluctuations become large enough so that their subsequent evolution can be described by the classical equations of motion. I will use the term "coherence region" to denote a region within which the scalar field is approximately uniform. The coherence regions are irregular in shape, and their initial size is typically of order $c\chi^{-1}$. Note that $c\chi^{-1}$ is only about 10^{-14} proton diameters; the entire observed universe will evolve from a region of this size or smaller.

The scalar field ϕ then "rolls" down the potential energy function shown in Fig. 10.1, obeying the classical equations of motion derived from general relativity. As long as the spatial variations in ϕ are small, these classical equations take the form

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} = -\frac{\partial V}{\partial \phi} \ . \tag{10.14}$$

(The derivation of Eq. (10.14) is a straightforward application of general relativity, but it is a little beyond the scope of this course.) If the initial fluctuation is small, then the flatness of the potential for $\phi \approx 0$ will ensure that the rolling begins very slowly. Note that the second term on the left-hand-side of Eq. (10.14) is a damping term, helping to slow down the speed of rolling. As long as $\phi \approx 0$, the mass density ρ remains about equal to $\rho_{\rm f}$, and the exponential expansion continues. The expansion occurs on a time scale χ^{-1}, while the time scale of the rolling is much slower. This "slow roll" of the scalar field is the crucial new feature in the ***new*** inflationary universe.

For the scenario to work, it is necessary for the length scale of homogeneity to be stretched from $c\chi^{-1}$ to at least about 10 cm before the scalar field ϕ rolls off the plateau of the potential energy diagram. This corresponds to an expansion factor of about 10^{28}, which requires about 65 time constants (χ^{-1}) of expansion. The expected duration of the expansion depends on the precise shape of the scalar field potential, and models have been constructed which yield much more than the minimally required amount of inflation.

When the ϕ field reaches the steep part of the potential, it falls quickly to the bottom and oscillates about the minimum. The time scale of this motion is a typical GUT time of $\hbar/E_{\rm GUT} \approx 7 \times 10^{-41}$ sec, which is very fast compared to the expansion rate. The scalar field oscillations are then quickly damped by the couplings to the other fields, and the energy is rapidly converted into a thermal equilibrium mixture of particles. (From a particle point of view, the scalar field oscillations correspond to a state of spinless particles, just as an oscillating electromagnetic field corresponds to a state of photons. The damping of the scalar field is just the field theory description of the decay of these particles into other kinds of particles.) The release of this energy reheats the region back to a temperature which can be of order $kT \approx 10^{16}$ GeV, or can be much lower, depending on the strength of the interactions. The universe is continuing to expand and cool as the gas of particles approaches a state of thermal equilibrium, so the reheat temperature is low if this process of thermalization is slow, and high if it is quick.

From here on the standard scenario takes over. The era of inflation has set up precisely the initial conditions that had previously been assumed in standard cosmology. You can check that a region of radius ≈ 10 cm, at a temperature $kT \approx 10^{16}$ GeV, will become large enough by the time T falls to 2.7 K to encompass the entire observed universe.

CHAOTIC INFLATION:

While I have described the new inflationary model, because I think it is the simplest version to understand, there are now many variants of inflationary models. One very important variant is known as *chaotic inflation*,[7] invented by Andrei Linde in 1983. Linde realized that in fact inflation does not require a plateau in the potential energy diagram, but can in fact happen with a potential energy function as simple as

$$V(\phi) = \frac{1}{2} m^2 \, \phi^2 \ , \tag{10.15}$$

which in fact describes a non-interacting particle of mass m. If the field ϕ is started at a large enough value, then sufficient inflation can occur as the scalar field rolls towards $\phi = 0$. Linde initially proposed that the scalar field could start at a large value in some places due to "chaotic" initial conditions. Later he showed that quantum fluctuations can cause these models to also undergo eternal inflation, which will be discussed below, so the question of initial conditions is perhaps irrelevant.

SOLUTIONS TO THE COSMOLOGICAL PROBLEMS:

Let me now explain how the three problems of the standard cosmological scenario discussed in Lecture Notes 8 and 9 are avoided in the inflationary scenario. First, let us consider the horizon/homogeneity problem. The problem is clearly avoided in this scenario, since the entire observed universe evolves from a single coherence region. This region had a size of order $c\chi^{-1}$ at the time when the fluctuation began to grow classically. This size is much smaller than the sizes that are relevant in the standard model at these times, and the region therefore had plenty of time to come to a uniform temperature before the onset of inflation. As long as there are about 65 or more time constants of exponential expansion, then the exponential expansion causes this very small region of homogeneity to grow to be large enough to encompass the observed universe.

The flatness problem is avoided by the dynamics of the exponential expansion of the coherence region. As ϕ begins to roll very slowly down the potential, the evolution of the metric is governed by the mass density ρ_{f}. Assuming that the coherence region (or at least a small piece of it) can be approximated by a Robertson-Walker metric, then the scale factor evolves according to the standard Friedmann equation:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3} G\rho - \frac{kc^2}{a^2} \ , \tag{10.16}$$

where $k = +1$, -1, or 0 depending on whether the region approximates a closed, open, or flat universe, respectively. (There could also be perturbations, but the cosmological no-hair theorem guarantees that they would die out quickly.) In this language, the flatness problem is the problem of understanding why the kc^2/a^2 term on the right-hand-side is so extraordinarily small, compared to the other terms. But as the coherence region expands exponentially, the mass density ρ remains very nearly constant at $\rho_{\rm f}$, while the kc^2/a^2 term is suppressed by at least a factor of $\left(10^{28}\right)^2 = 10^{56}$. Since the equation must continue to hold, the term on the left-hand side must remain nearly constant, like the mass density. This provides a "natural" explanation of why the value of the kc^2/a^2 term immediately after the phase transition is smaller than that of the other terms by a tremendous factor.

Except for a very narrow range of parameters, this suppression of the curvature term will vastly exceed that required by present observations. This leads to the prediction that the kc^2/a^2 term of Eq. (10.16) should remain totally negligible until the present era, and even far into the future. This implies that the value of Ω today is expected to be equal to one with a high degree of accuracy.

The inflationary prediction that $\Omega = 1$ seemed to be at odds with observation until 1998, with the discovery of the dark energy. Astronomers never found enough matter to make up a critical mass density, although there was always some room for uncertainty. Some inflationary theorists constructed versions of inflation that could lead to an open universe; this could be arranged by choosing the parameters to that inflation proceeds for just long enough to solve the flatness problem, but not so long that it flattened the universe completely.

But the situation changed dramatically in 1998 with the Supernova Type Ia measurements, which indicated the presence of a cosmological constant or a very slowly evolving scalar field that could simulate a cosmological constant. In either case, the total energy in this new component of the universe is just what is needed to complete the inventory for a flat universe. The best current estimate of Ω_0 is based on the Planck satellite data[4] for the anisotropies of the cosmic microwave background radiation, combined with several other astronomical observations, giving $\Omega_0 = 0.9992 \pm 0.0040$.

Finally, we turn to the monopole problem. Recall that in the standard scenario, the tremendous excess of monopoles was produced by the disorder in the Higgs field (i.e., by the Kibble mechanism). There is no known way to prevent the Kibble mechanism from operating, but as long as inflation occurs after or during the process of monopole formation, the monopoles will be diluted enormously. During inflation the volume of the coherence region increases by a factor of about $\left(10^{28}\right)^3 = 10^{84}$ or more, which is enough to convert the monopole glut into a situation where no monopoles will be seen.

RIPPLES IN THE COSMIC MICROWAVE BACKGROUND

After subtracting a contribution attributed to the motion of the Earth through the cosmic microwave background, the temperature of the CMB appears to be uniform in all directions to an accuracy of about 1 part in 100,000. Nonetheless, at the level of 1 part in

100,000, there are anisotropies (i.e., non-uniformities) that have by now been measured to high precision.

The huge stretching of inflation tends to smooth everything out. Any density of particles that might be present before inflation is diluted away, so that during inflation the energy density is dominated by the energy density of the false vacuum state. If there was any curvature in space itself before inflation began, the effect of inflation is to stretch out those curves. As one stretches a sphere the surface gets flatter and flatter, and the same is actually true for any curved space. So, when inflation is described in the context of classical general relativity, the result of inflation would be an almost completely smooth space. There was a period of about a year, in the very early days of inflation, when this appeared to be a serious problem. If inflation left the universe almost completely smooth, then there would be no way for galaxies to form.

But quantum mechanics came to the rescue. The idea that quantum mechanics might be responsible for the structure of the universe goes back at least to Andrei Sakharov, the Russian nuclear physicist and political activist, who put forward the idea in 1965.[14] In 1981 Mukhanov and Chibisov[15] revived Sakharov's idea in a modern context, studying the density perturbations generated in a closely related model proposed by Alexei Starobinsky[16] in 1980. The original work on density perturbations arising from scalar-field-driven inflation centered around the Nuffield Workshop on the Very Early Universe, Cambridge, U.K., June-July 1982, organized by Gary Gibbons and Stephen Hawking. There was much animated discussion and disagreement during the workshop, but in the end everyone agreed on the answer. There were four papers[17] that came out of the workshop, laying the foundations for calculating density perturbations arising from inflation.

The important feature of quantum mechanics in this context is that it is intrinsically probabilistic. So, while the classical approximation of inflation theory predicts a

---

[14] A. D. Sakharov, "The initial stage of an expanding universe and the appearance of a nonuniform distribution of matter," Zh. Eksp. Teor. Fiz. 49, 345 (1965) [JETP Lett. 22, 241 (1966)].

[15] V. F. Mukhanov and G. V. Chibisov, "Quantum fluctuations and a nonsingular universe," Pis'ma Zh. Eksp. Teor. Fiz. 33, 549 (1981) [JETP Lett. 33, 532 (1981)].

[16] A. A. Starobinsky, "A new type of isotropic cosmological models without singularity," Phys. Lett. B, vol. 91, p. 99 (1980) .

[17] S.W. Hawking, "The development of irregularities in a single bubble inflationary universe," Physics Letters B, vol. 115, p. 295 (1982), (with an MIT certificate, click here); A. A. Starobinsky, "Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations," Physics Letters B, vol. 117, p. 175 (1982); A. H. Guth and S.-Y. Pi, "Fluctuations in the new inflationary universe," Physical Review Letters, vol. 49, p. 1110 (1982) (with an MIT certificate, click here); J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, "Spontaneous creation of almost scale-free density perturbations in an inflationary universe," Physical Review D, vol. 28, p. 679 (1983) (with an MIT certificate, click here).

completely smooth universe, the quantum theory implies that the matter density will be almost uniform, but due to quantum uncertainties the density will be a little higher than average in some places, and a little lower than average in others. These uncertainties are just the ripples that are needed to allow galaxy formation to proceed, and they are just what is needed to compare with observations of the ripples, starting in 1992. Of course quantum uncertainties are not usually significant on macroscropic scales, so it seems bizarre that quantum fluctuations can be responsible for the large-scale structure of the universe. This is made possible, however, by the extremely rapid expansion during inflation, which stretches the quantum fluctuations from very short length scales, where we expect them to be strong, to macroscopic and even astronomical length scales.

Inflation is of course not a unique theory, since we do not know exactly what the inflaton field is, or exactly what equations of motion it obeys. We don't even know that inflation was driven by a single inflaton, as there may have been two or more. Thus, the detailed predictions for density perturbations arising from inflation are model-dependent, meaning that different assumptions about the inflaton will lead to different predictions. Nonetheless, there is a wide class of "simple" inflationary models which give very similar predictions for the spectrum of the density fluctuations. The word "spectrum" here has pretty much the same meaning it would have for sound waves: the perturbations can be broken up into components with definite wavelengths, and the "spectrum" is a description of how the intensity varies with wavelength. (For sound waves we might be more likely to use frequency rather than wavelength, but for cosmological density perturbations we have no choice but to use wavelength — we don't see oscillations, and we expect oscillations only in some cases.) For the ripples on the CMB, the wavelength is measured in degrees, not in meters, since we are seeing a pattern on the sky as a function of polar angles θ amd ϕ.

The "simple" inflationary models that give similar results are more technically called single field slow-roll models, and they are characterized by the facts that there is a single inflaton, and that, during the period when relevant density perturbations are created, both $H = \dot{a}/a$ and $\partial V/\partial\phi$ are nearly constant and the $\ddot{\phi}$ term of Eq. (10.14) is small compared to the other two terms. The overall magnitude of the density perturbations, on the other hand, depends on more of the details of the inflaton potential energy function, so at present there is no inflationary prediction for the magnitude.

The ripples in the CMB are measured most easily from space, although ground-based measurements can also be significant, especially at very short angular wavelengths, for which high angular resolution is needed. So far there have been three satellite experiments that have been completely dedicated to measuring the properties of the CMB. The first was the Cosmic Background Explorer (COBE), launched by NASA in 1989, 15 years after planning began in 1974. In January, 1990, the COBE group announced their first measurements of the CMB spectrum, showing that it agreed beautifully with the expected black-body spectrum (recall Figure 6.5 in Lecture Notes 6). In April of

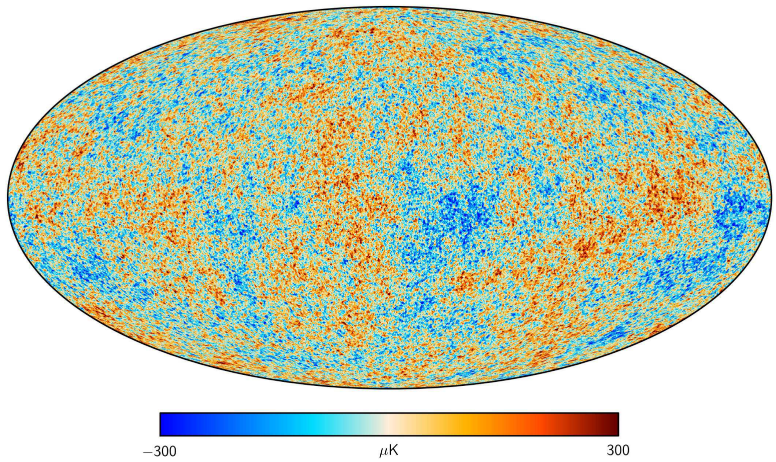


Figure 10.4: The cosmic microwave background radiation as detected by the *Planck* satellite.[18] After correcting for the motion of the Earth, the temperature of the radiation is nearly uniform across the entire sky, with average temperature $T_{\text{cmb}} = 2.726$ K. Tiny deviations from the average temperature have been measured; they are so small that they must be depicted in a color scheme that greatly exaggerates the differences, to make them visible. As shown here, blue spots are slightly colder than T_{cmb} while red spots are slightly warmer than T_{cmb}, across a range of $\Delta T/T_{\text{cmb}} \sim 10^{-4}$.

1992, the team announced the first measurements of anisotropies in the CMB. The 2006 Nobel Prize in Physics was awarded to John Mather and George Smoot for their work on the COBE mission. The second CMB satellite mission was the Wilkinson Microwave Anisotropy Probe (WMAP), launched by NASA in 2001. The WMAP was 45 times more sensitive, with 33 times the angular resolution of its COBE satellite predecessor. The third CMB satellite was Planck, launched in 2009 by the European Space Agency. The resolution of Planck was about $2\frac{1}{2}$ times better than WMAP, with higher sensitivity and also measurements in 9 frequency bands, compared to 5 for WMAP.

Figure 10.4 shows the microwave sky, as seen in the 2015 data release from the Planck satellite. The radiation is almost completely uniform, but the tiny variations are shown in a false-color image, with the temperature color-code shown by the bar at the bottom. This picture is illustrative, but it is hard to learn anything just by looking at it.

Figure 10.5[19] shows a spectrum computed from the 7-year data release of WMAP

[18] R. Adam et al. (Planck Collaboration), "Planck 2015 results, I: Overview of products and scientific results," Figure 9, Astronomy & Astrophysics vol. 594, article A1 (2016), arXiv:1502.01582 [astro-ph.CO]

[19] A. H. Guth and D. I. Kaiser, "Inflationary cosmology: Exploring the Universe from the smallest to the largest scales," Science, 11 Feb 2005, vol. 307, pp. 884-890 (2005). With an MIT certificate, click here. Also available at arXiv:astro-ph/0502328.

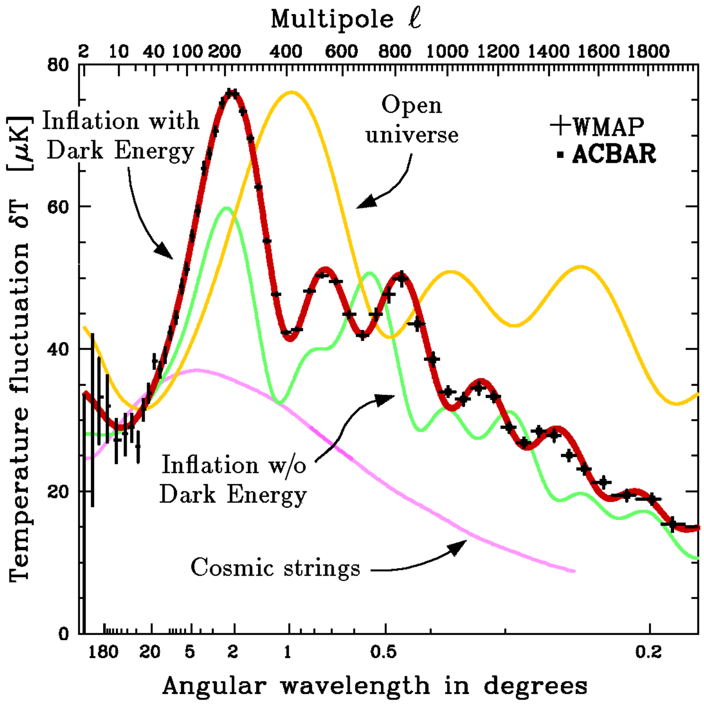


Figure 10.5: Comparison of the WMAP 7-yr and ACBAR observational measurements of the temperature fluctuations in the CMB with several theoretical models, as described in the text.

for long angular wavelengths, and an experiment called ACBAR for shorter angular wavelengths. The graph was constructed by Max Tegmark, to be used in a summary of inflation written by David Kaiser and me.[19] The vertical axis shows the strength of the fluctuations, in microkelvin, and the horizontal axis shows the angular wavelength, with the longest wavelengths on the left. (For those who are familiar with spherical harmonics, the decomposition into angular wavelengths is accomplished by an expansion in spherical harmonics $Y_{\ell m}(\theta, \phi)$, and the vertical axis represents the strength at each ℓ. The angular wavelength is taken as $360°/\ell$.) The graph shows a comparison between different theories. The red line shows the predictions for an inflationary model with $\Omega_{\rm vac} = 0.72$; the yellow line describes an open universe, with $\Omega_m = 0.30$ and $\Omega_{\rm vac} = 0$; the green line describes an inflationary model without dark energy, meaning that $\Omega_m = 1$, $\Omega_{\rm vac} = 0$; the purple line shows the prediction of a completely different mechanism for the generation of density perturbations, called cosmic strings.[20] Cosmic strings were mentioned in passing on

[20] The curve for cosmic strings was taken from U.-L. Pen, U. Seljak, and N. Turok, "Power spectra in global defect theories of cosmic structure formation," Physical Review Letters, vol. 79, pp. 1611–1614 (1997), or with an MIT certificate, click here. Also available at arXiv:astro-ph/9704165.

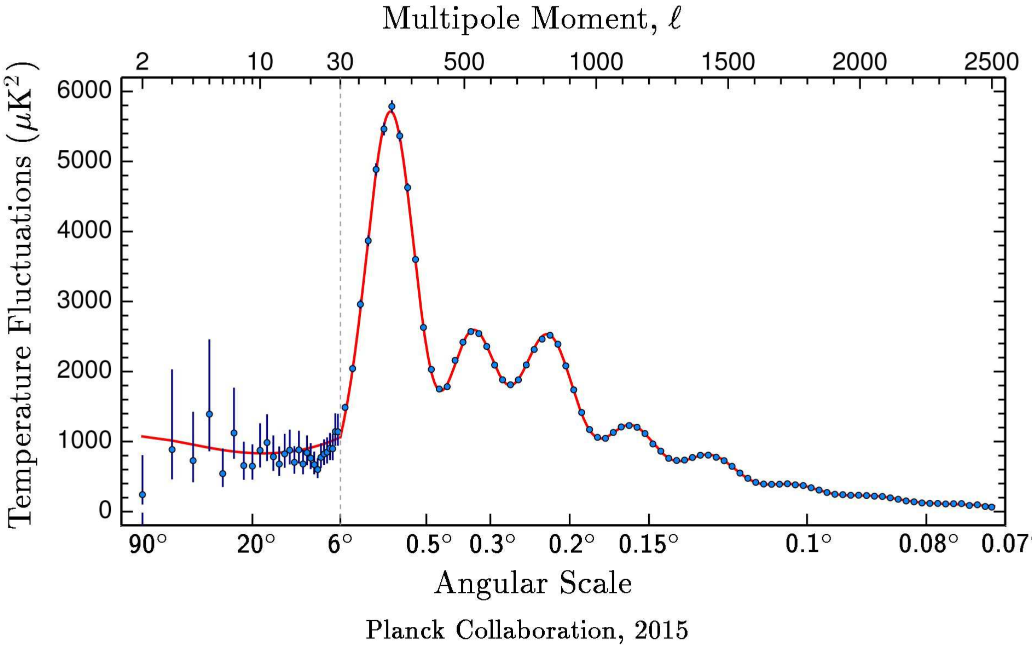


Figure 10.6: The measurements of the CMB temperature fluctuations by the Planck satellite, 2015 data release. Taken from Ref. [4], Figure 1.

p. 13 of Lecture Notes 9; they are linelike topological defects, in contrast to monopoles which are pointlike defects. They could create density perturbations through the random processes involved in their formation, and prior to the careful CMB measurements they were considered a viable theory for the origin of density perturbations. Now, however, they are clearly ruled out.

The error bars on the graph are clearly much larger on the left, at large angular wavelengths, but there is a simple explanation. For perturbations with an angular wavelength of 0.2° there are a huge number of samples on the sky, but for angular wavelengths such as 180° there are very few.

Figure 10.6 shows a more recent graph of the spectrum of the CMB, showing the data from the 2015 data release of the Planck satellite project. The red line shows a theoretical curve from a best-fit simple inflationary model, described in Table 4, Column 1 of Ref. [4]: $\Omega_m = 0.315$, $\Omega_{\rm vac} = 0.685$, and $H_0 = 67.3$ km-s^{-1}-Mpc^{-1}. It is actually a six-parameter fit to the data, where the overall height of the curve is one of the remaining parameters that is fit. There is also a parameter τ that describes a small amount of absorption of CMB photons on the way to the Earth — the fraction that arrive is $e^{-\tau}$, where $\tau = 0.078$; and finally there is a parameter $n_s = 0.966$, which describes a small deviation from the simple approximation that H and $\partial V/\partial\phi$ are constant during the period in which the presently observed density perturbations were created. As one can see, the fit is excellent.

In the words used by the Planck team, "The Planck results offer powerful evidence in favour of simple inflationary models."

ETERNAL INFLATION:

We will not have time to fully discuss the mind-boggling implications of this feature, but the basic facts are pretty straightforward. As the scalar field rolls off the potential energy plateau shown in Fig. 10.1, we must remember that in a full quantum mechanical treatment there will always be some probability that the scalar field will remain at the top of the hill. Approximate calculations show that this probability falls off exponentially with time, with a time constant that is similar to, but maybe a factor of 100 slower than, the time constant of the exponential expansion. This means that if an observer stayed at any one point of the inflating region, it is highly probable that she would see inflation end in a very short amount of time, perhaps 10^{-35} second. However, if we were to calculate how the total volume of false vacuum changes with time, we would find that the growing exponential of the expansion dominates over the falling exponential of the decay, so the total volume of false vacuum grows exponentially in time! Once inflation starts we expect it never to stop, but instead it will continue forever. The decay of the false vacuum (the transition of the scalar field to the true vacuum value) does not happen globally, but instead pieces of the false vacuum undergo the decay and produce huge regions of inhabitable space that can be called *pocket universes.* An infinite number of pocket universes are produced. The collection of the infinite number of pocket universes is called the *multiverse.*

Eternal inflation is easiest to understand for new inflation, but it can happen also in chaotic inflation. In 1986 Linde[21] showed that as the inflaton field "rolls" down a hill in the potential energy diagram, it is possible for quantum fluctuations to drive it up the hill often enough for the volume of the inflating region to increase with time, rather than decrease.

Can we see these other universes? Almost certainly not, although in principle other pocket universes could reveal their presence by colliding with ours. Such collisions could show up as circular distortions in the cosmic microwave background. Astronomers have in fact looked for such patterns,[22] but have not found any persuasive evidence for them.

Is this discussion physics or metaphysics? That's debatable, but in my opinion it is physics, albeit very speculative physics at this stage. First, it seems to be an

---

[21] A. D. Linde, "Eternal Chaotic Inflation," Modern Physics Letters A, vol. 1, issue 2, p. 81 (1986).

[22] S. M. Feeney, M. C. Johnson, D. J. Mortlock, and H. V. Peiris, "First observational tests of eternal inflation," Physical Review Letters, vol. 107, article 071301 (2011). With an MIT certificate, click here. Also available at arXiv:1012.1995 [astro-ph.CO].

almost unavoidable consequence of inflation, which itself makes a number of testable predictions, and has been very successful. Second, it now appears that the possibility of a multiverse may have relevance to perplexing problems in fundamental physics, such as the cosmological constant problem discussed at the end of Lecture Notes 7. The problem, we recall, was that the vacuum energy density of our universe, measured by its acceleration, is vastly smaller (120 orders of magnitude!) than naive estimates from particle physics. The multiverse offers a possible (although controversial) explanation for this situation. According to string theory, there is no unique vacuum state, but instead a colossal number, perhaps 10^{500} or more, of long-lived metastable states, any one of which could serve as the vacuum for a pocket universe.[23] This set of possible vacua is often called the "landscape" of string theory. Even if string theory is not right, it is still possible that nature allows a huge number of different vacuum-like states. Each vacuum-like state would have its own energy density, expected to be typically of the order of the "Planck scale," the energy density that one can construct from the fundamental constants G, $\hbar$, and c. By dimensional analysis, one finds that the only way to construct an energy density from these quantities is

$$\rho_{\text{Planck}} \equiv \frac{c^5}{\hbar G^2} = 5.16 \times 10^{96}\ \text{kg/m}^3 = 5.16 \times 10^{93}\ \text{g/cm}^3\ . \tag{10.17}$$

On Problem Set 8, Problem 5, you found an estimate for the energy density of the vacuum fluctuations of the electromagnetic field, which was of this order of magnitude. Vacuum energy densities can be positive or negative, so a natural expectation is that the energy densities of the possible vacua would range roughly from minus the Planck scale to plus the Planck scale. But if they are anything like evenly spread, there would be a fantastic number (maybe $[(5.95 \times 10^{-30}\ \text{g/cm}^3)/(2 \times 5.16 \times 10^{93}\ \text{g/cm}^3)] \times 10^{500} \approx 6 \times 10^{376}$) of vacua with energy densities as small as what we observe, although they would still be incredibly rare in the full set of $\sim 10^{500}$ vacua.

To explain why we might be living in such a rare type of vacuum, the argument invokes a selection effect associated with the fact that we are living beings. This selection effect is often called the "anthropic principle." We expect most of the pocket universes in the multiverse to have vacuum energies with a magnitude of the order of the Planck scale, but such pocket universes would fly apart (if $\rho_{\text{vac}} > 0$) or implode (if $\rho_{\text{vac}} < 0$) on a time scale of order 10^{-44} sec. (To find the time scale, calculate χ^{-1} for χ given by Eq. (10.9), with ρ_{f} replaced by ρ_{vac}.) It is therefore easy to believe that no life will exist in such typical pocket universes. The complexity of life requires time to evolve, so we expect life

[23] R. Bousso and J. Polchinski, "Quantization of four form fluxes and dynamical neutralization of the cosmological constant," Journal of High Energy Physics, vol. 2000, 06, 006, arXiv:hep-th/0004134. See also R. Bousso and J. Polchinski, "The string theory landscape", Scientific American, vol. 291, p. 78–87 (2004). With an MIT certificate, click here.

to form only in those rare types of vacuum in which the vacuum energy densiy is extremely close to zero. In 1998 H. Martel, P. R. Shapiro, and Steven Weinberg[24] estimated how large the vacuum energy density could be for it to still be possible for matter to condense out of the background into mass concentrations large enough to form observers. They found that under these assumptions, life would form only in those pocket universes in which the vacuum energy density were of the same order of magnitude as the current critical density. This result makes the selection effect explanation look very plausible, but we must keep in mind (1) that the Martel-Shapiro-Weinberg calculation ignored the possibility of life forms very different from ourselves, and (2) the calculation ignored the fact that other parameters of the laws of physics, and not just the cosmological constant, could be different in different pockets.

A PERSONAL SUMMARY:

I hope that you have enjoyed our journey into the current status of cosmology. I personally find it mind-boggling that we can use the big-bang theory to calculate the abundances of the light chemical elements, and even more mind-boggling that we can theorize about the behavior of the universe at 10^{-37} seconds after its beginning. It is mind-boggling that the structure of the universe could have arisen from quantum uncertainties, and astounding that such a wild idea can lead to a fit with the data as good as Figure 10.6.

It is absolutely incredible how far physics has taken us in the quest to understand the universe, but at the same time it is incredible how many key questions remain unanswered. The baryonic matter that we understand comprises only about 5% of the total energy of the universe. What is the dark matter, which makes up 26% of the universe? If the dark energy is really vacuum energy, why is the energy density so much smaller than particle theorists would expect? And if inflation is right, what exactly is the inflaton, and what is the detailed description of its dynamics?

I find it amazing how much we understand about cosmology, and equally amazing how much we don't.

[24] H. Martel, P. R. Shapiro, and S. Weinberg, "Likely Values of the Cosmological Constant," The Astrophysical Journal, vol. 492, pp. 29-40 (1998), arXiv:astro-ph/9701099.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — September 6, 2018
Prof. Alan Guth

PROBLEM SET 1

DUE DATE: Friday, September 14, 2018, 5:00 pm.

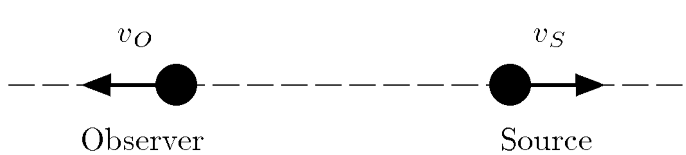
READING ASSIGNMENT: *The First Three Minutes*, Chapters 1 and 2.

NOTE ABOUT EXTRA CREDIT: This problem set contains 40 points of regular problems and 15 points extra credit, so it is probably worthwhile for me to clarify the operational definition of "extra credit". We will keep track of the extra credit grades separately, and at the end of the course I will first assign provisional grades based solely on the regular coursework. I will consult with our teaching assistant, Honggeun Kim, and we will try to make sure that these grades are reasonable. Then I will add in the extra credit, allowing the grades to change upwards accordingly. Finally, Honggeun and I will look at each student's grades individually, and we might decide to give a higher grade to some students who are slightly below a borderline. Students whose grades have improved significantly during the term, students whose average has been pushed down by single low grade, and students who have been affected by adverse personal or medical problems will be the ones most likely to be boosted.

The bottom line is that the extra credit problems are OPTIONAL. You should feel free to skip them, and you will still get an excellent grade in the course if you do well on the regular problems. However, if you have some time and enjoy an extra challenge, then I hope that you will find the extra credit problems interesting and worthwhile.

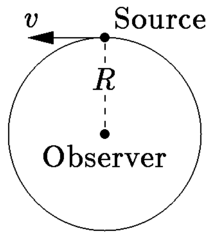
PROBLEM 1: NONRELATIVISTIC DOPPLER SHIFT, SOURCE AND OBSERVER IN MOTION *(15 points)*

Consider the Doppler shift of sound waves, for a case in which both the source and the observer are moving. Suppose the source is moving with a speed v_s relative to the air, while the observer is receding from the source, moving in the opposite direction with speed v_o relative to the air. Calculate the Doppler shift z. (Recall that z is defined by $1+z \equiv \lambda_o/\lambda_s$, where λ_o and λ_s are the wavelengths as measured by the observer and by the source, respectively.) *Hint:* while this problem can be solved directly, you can save time by finding a way to determine the answer by using the cases that are already calculated in Lecture Notes 1.



PROBLEM 2: THE TRANSVERSE DOPPLER SHIFT *(25 points)*

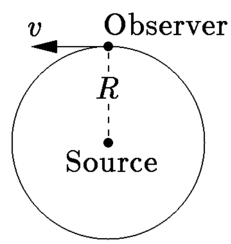
Consider the Doppler shift observed by a stationary observer, from a source that travels in a circular orbit of radius R about the observer. Let the speed of the source be v.



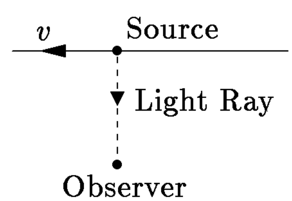
(a) *(5 points)* If the wave in question is sound, and both the source speed v and the wave speed u are very small compared to the speed of light c, what is the Doppler shift z? Assume that the observer is at rest relative to the air.

(b) *(5 points)* If the wave is light, traveling with speed c, and v is not small compared to c, what is the Doppler shift z? This is called the *transverse Doppler shift*, since the velocity of the light ray is perpendicular to the velocity of the source at the time of emission, as seen in the reference frame of the observer.

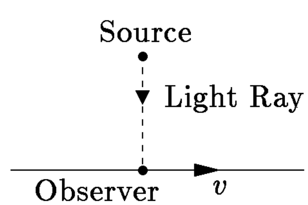
(c) *(5 points)* Still considering light waves and the same pattern of motion as shown in the figure, suppose that the source and the observer were reversed. That is, suppose a light ray is sent from the person at the center of the circle to the person traveling around the circle at speed v. In this case, what would be the Doppler shift z?



(d) *(5 points)* Now suppose that the motion is linear instead of circular. Again we consider light rays, and as in part (b) we assume that the source is moving with a speed v that is not small compared to c. If the light ray is emitted by the source at the moment of its closest approach to the observer, as shown in the diagram, what is the Doppler shift z?



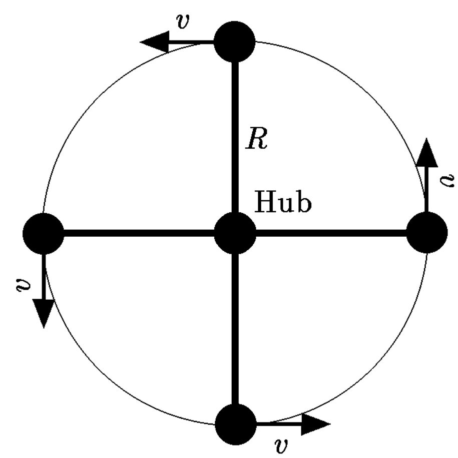
(e) *(5 points)* Again consider linear motion, with light rays. As in part (c), assume that the observer is moving with a speed v that is not small compared to c. If the light ray is received by the observer at the moment of its closest approach to the source, as shown in the diagram, what is the Doppler shift z?



PROBLEM 3: A HIGH-SPEED MERRY-GO-ROUND

(This problem is not required, but can be done for 15 points extra credit.)

Now consider the Doppler shift as it would be observed in a high-speed "merry-go-round." Four evenly-spaced cars travel around a central hub at speed v, all at a distance R from a central hub. Each car is sending waves to all three of the other cars.



(a) If the wave in question is sound, and both the source speed v and the wave speed u are very small compared to the speed of light c, with what Doppler shift z does a given car receive the sound from (i) the car in front of it; (ii) the car behind it; and (iii) the car opposite it?

(b) In the relativistic situation, where the wave is light and the speed v may be comparable to c, what is the answer to the same three parts (i)-(iii) above?

Total points for Problem Set 1: 40, plus 15 points of extra credit.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — September 15, 2018
Prof. Alan Guth

PROBLEM SET 2

DUE DATE: Monday, September 24, 2018, 5:00 pm. The due date has been postponed, since Friday September 21 is a Student Holiday.

SEPTEMBER/OCTOBER				
MON	TUES	WED	THURS	FRI
September 17 Lecture 4	18	19 Lecture 5	20	21
24 Lecture 6 PS 2 due	25	26 Lecture 7	27	28 PS 3 due
October 1 Lecture 8	2	3 Quiz 1 — in class	4	5

READING ASSIGNMENT: Barbara Ryden, *Introduction to Cosmology*, Chapters 1-3.

PLANNING AHEAD: If you want to read ahead, the reading assignment with Problem Set 3 will be Weinberg, *The First Three Minutes,* Chapter 3. Problem Sets 1 through 3, including the reading assignments, will be included in the material covered on Quiz 1, on Wednesday, October 3.

INTRODUCTION TO THE PROBLEM SET

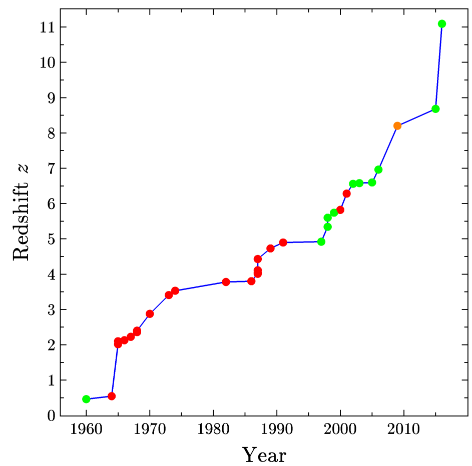
In this problem set we will consider a universe in which the scale factor is given by

$$a(t) = bt^{2/3} \ ,$$

where b is an arbitrary constant of proportionality which should not appear in the answers to any of the questions below. (We will see in Lecture Notes 3 that this is the behavior of a flat universe with a mass density that is dominated by nonrelativistic matter.) We will suppose that a distant galaxy is observed with a redshift z. As a concrete example we will consider the most distant known object with a well-determined redshift, the galaxy

GN-z11, which has a redshift $z = 11.1$. The discovery of this galaxy was announced in March 2016 by an international group of astronomers, using the Hubble Space Telescope*

The rate at which the highest measured redshift has been growing has been dramatic. In 1986 the highest measured redshift was only 3.78. It was 4.01 in 1988, 4.73 in 1992, 4.897 in 1994, and 4.92 in 1998, 5.34 in 2000, 6.28 in 2002, and 6.58 in 2003. In 2006 Iye et al.† discovered a galaxy with a redshift of 6.96. In 2009-2010 astronomers searching the Hubble Space Telescope Ultra Deep Field discovered the galaxy UDFy-38135539, which was claimed to have a redshift of $z = 8.55$. Further observations, however, failed to confirm this number. The diagram at the right shows a graph of the highest confirmed redshift by year of discovery, using the listing in the Wikipedia‡. The red circles represent quasars, the green circles represent galaxies, and the one orange circle at 2009 is a gamma ray burst. The search for high redshift objects continues to be an exciting area of research, as astronomers try to sort out the conditions in the universe when the first galaxies began to form.



PROBLEM 1: DISTANCE TO THE GALAXY *(10 points)*

Let t_0 denote the present time, and let t_e denote the time at which the light that we are currently receiving was emitted by the galaxy. In terms of these quantities, find the present value of the physical distance ℓ_p between this distant galaxy and us.

PROBLEM 2: TIME OF EMISSION *(10 points)*

Express the redshift z in terms of t_0 and t_e. Find the ratio t_e/t_0 for the $z = 11.1$ galaxy.

---

* P. A. Oesh et al., "A Remarkably Luminous Galaxy at $z = 11.1$ Measured with Hubble Space Telescope Grism Spectroscopy," The Astrophysical Journal **819**, 129 (2016), https://arxiv.org/abs/1603.00461.

† Iye et al., "A galaxy at a redshift $z = 6.96$," Nature vol. 443, no. 7108, pp. 186–188 (14 September 14 2006).

‡ "List of the most distant astronomical objects," (2016, April 24). In *Wikipedia, The Free Encyclopedia.* Retrieved 01:03, May 27, 2016, from https://en.wikipedia.org/w/index.php?title=List_of_the_most_distant_astronomical_objects&oldid=716940651

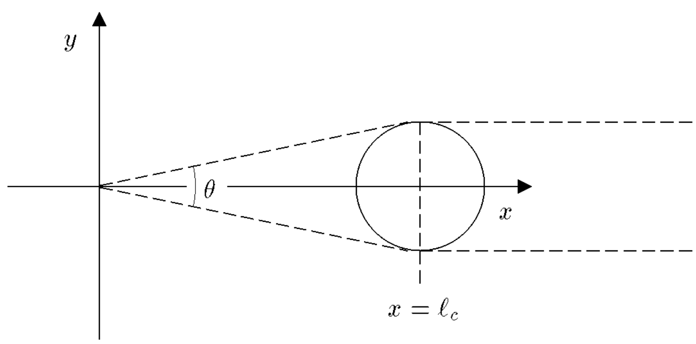
PROBLEM 3: DISTANCE IN TERMS OF REDSHIFT *z* *(10 points)*

Express the present value of the physical distance in terms of the present value of the Hubble expansion rate H_0 and the redshift z. Taking $H_0 \approx 67$ km-sec^{-1}-Mpc^{-1}, how far away is the galaxy? Express your answer both in light-years and in Mpc.

PROBLEM 4: SPEED OF RECESSION *(10 points)*

Find the present rate at which the physical distance ℓ_p between the distant galaxy and us is changing. Express your answer in terms of the redshift z and the speed of light c, and evaluate it numerically for the case $z = 11.1$. Express your answer as a fraction of the speed of light. [If you get it right, this "fraction" is greater than one! Our expanding universe violates special relativity, but is consistent with general relativity.]

PROBLEM 5: APPARENT ANGULAR SIZES *(20 points)*

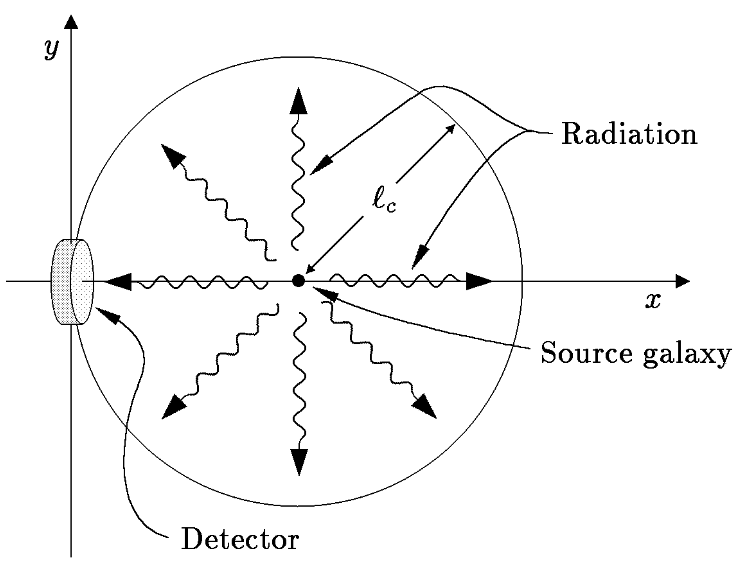


Now suppose for simplicity that the galaxy is spherical, and that its physical diameter was w at the time it emitted the light. (The actual galaxy is seen as an unresolved point source, so we don't know it's actual size and shape.) Find the apparent angular size θ (measured from one edge to the other) of the galaxy as it would be observed from Earth today. Express your answer in terms of w, z, H_0, and c. You may assume that $\theta \ll 1$. Compare your answer to the apparent angular size of a circle of diameter w in a static Euclidean space, at a distance equal to the present value of the physical distance to the galaxy, as found in Problem 1. [Hint: draw diagrams which trace the light rays in the ***comoving*** coordinate system. If you have it right, you will find that θ has a minimum value for $z = 1.25$, and that θ increases for larger z. This phenomenon makes sense if you think about the distance to the galaxy at the time of emission. If the galaxy is ***very*** far away today, then the light that we now see must have left the object very early, when it was rather close to us!]

PROBLEM 6: RECEIVED RADIATION FLUX

(This problem is not required, but can be done for 15 points extra credit.)

At the time of emission, the galaxy had a power output P (measured, say, in ergs/sec) which was radiated uniformly in all directions. This power was emitted in the form of photons. What is the radiation energy flux J from this galaxy at the earth today? Energy flux (which might be measured in ergs-cm^{-2}-sec^{-1}) is defined as the energy per unit area per unit time striking a surface that is orthogonal to the direction of energy flow. The easiest way to solve this problem is to consider the trajectories of the photons, as viewed in comoving coordinates. You must calculate the rate at which photons arrive at the detector, and you must also use the fact that the energy of each photon is proportional to its frequency, and is therefore decreased by the redshift. You may find it useful to think of the detector as a small part of a sphere that is centered on the source, as shown in the following diagram:



Total points for Problem Set 2: 60, plus 15 points of extra credit.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — September 22, 2018
Prof. Alan Guth

PROBLEM SET 3

DUE DATE: Friday, September 28, 2018, 5:00 pm.

READING ASSIGNMENT: Steven Weinberg, *The First Three Minutes*, Chapter 3.

SHORT-TERM CALENDAR:

SEPTEMBER/OCTOBER				
MON	TUES	WED	THURS	FRI
September 17 Lecture 4	18	19 Lecture 5	20	21
24 Lecture 6 PS 2 due	25	26 Lecture 7	27	28 PS 3 due
October 1 Lecture 8	2	3 Quiz 1 — in class	4	5

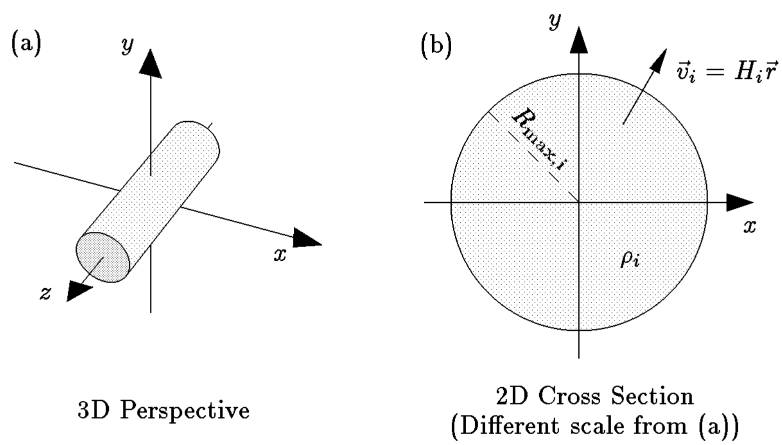
QUIZ DATES FOR THE TERM:
Quiz 1: Wednesday, October 3, 2018
Quiz 2: Monday, November 5, 2018
Quiz 3: Wednesday, December 5, 2018

FIRST QUIZ: The first of three quizzes for the term will be given on Wednesday, October 3, 2018, during the regular class period, in the usual room, 4-163. The quiz material will include Problem Sets 1–3, Lecture Notes 1–3, and the associated readings. The coverage of the quiz will be described in more detail on the class website, http://web.mit.edu/8.286/www/.

PROBLEM 1: A CYLINDRICAL UNIVERSE *(25 points)*

The following problem originated on Quiz 2 of 1994, where it counted 30 points.

The lecture notes showed a construction of a Newtonian model of the universe that was based on a uniform, expanding, sphere of matter. In this problem we will construct a model of a cylindrical universe, one which is expanding in the x and y directions but which has no motion in the z direction. Instead of a sphere, we will describe an infinitely long cylinder of radius $R_{\rm max,i}$, with an axis coinciding with the z-axis of the coordinate system:



We will use cylindrical coordinates, so

$$r = \sqrt{x^2 + y^2}$$

and

$$\vec{r} = x\hat{\imath} + y\hat{\jmath} \; ; \qquad \hat{r} = \frac{\vec{r}}{r} \; ,$$

where $\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$ are the usual unit vectors along the x, y, and z axes. We will assume that at the initial time t_i, the initial density of the cylinder is ρ_i, and the initial velocity of a particle at position $\vec{r}$ is given by the Hubble relation

$$\vec{v}_i = H_i \vec{r} \; .$$

(a) *(5 points)* By using Gauss' law of gravity, it is possible to show that the gravitational acceleration at any point is given by

$$\vec{g} = -\frac{A\mu}{r}\hat{r} \; ,$$

where A is a constant and μ is the total mass per length contained within the radius r. Evaluate the constant A.

(b) *(5 points)* As in the lecture notes, we let $r(r_i, t)$ denote the trajectory of a particle that starts at radius r_i at the initial time t_i. Find an expression for $\ddot{r}(r_i, t)$, expressing the result in terms of r, r_i, ρ_i, and any relevant constants. (Here an overdot denotes a time derivative.)

(c) *(5 points)* Defining

$$u(r_i, t) \equiv \frac{r(r_i, t)}{r_i} ,$$

show that $u(r_i, t)$ is in fact independent of r_i. This implies that the cylinder will undergo uniform expansion, just as the sphere did in the case discussed in the lecture notes. As before, we define the scale factor $a(t) \equiv u(r_i, t)$.

(d) *(5 points)* Express the mass density $\rho(t)$ in terms of the initial mass density ρ_i and the scale factor $a(t)$. Use this expression to obtain an expression for $\ddot{a}$ in terms of a, ρ, and any relevant constants.

(e) *(5 points)* Find an expression for a conserved quantity of the form

$$E = \frac{1}{2}\dot{a}^2 + V(a) .$$

What is $V(a)$? Will this universe expand forever, or will it collapse?

PROBLEM 2: A FLAT UNIVERSE WITH UNUSUAL TIME EVOLUTION *(10 points)*

Consider a ***flat*** universe which is filled with some peculiar form of matter, so that the Robertson–Walker scale factor behaves as

$$a(t) = bt^{3/4} ,$$

where b is a constant.

(a) *(5 points)* For this universe, find the value of the Hubble expansion rate $H(t)$.

(b) *(5 points)* What is the mass density of the universe, $\rho(t)$? (In answering this question, you will need to know that the equation for $\dot{a}/a$ in Lecture Notes 3,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2} ,$$

holds for all forms of matter, while the equation for $\ddot{a}$,

$$\ddot{a} = -\frac{4\pi}{3}G\rho(t)a ,$$

requires modification if the matter has a significant pressure. The $\ddot{a}$ equation is therefore not applicable to this problem.)

PROBLEM 3: ENERGY AND THE FRIEDMANN EQUATION *(30 points)*

The Friedmann equation,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2} \ , \tag{1}$$

was derived in Lecture Notes 3 as a first integral of the equations of motion. The equation was first derived in a different form,

$$E = \frac{1}{2}\dot{a}^2 - \frac{4\pi}{3}\frac{G\rho_i}{a} = \text{constant}, \tag{2}$$

where $k = -2E/c^2$. In this form the equation looks more like a conservation of energy relation, although the constant E does not have the dimensions of energy. There are two ways, however, in which the quantity E can be connected to the conservation of energy. It is related the energy of a test particle that moves with the Hubble expansion, and it is also related to the total energy of the entire expanding sphere of radius R_{max}, which was discussed in Lecture Notes 3 as a method of deriving the Friedmann equations. In this problem you will derive these relations.

First, to see the relation with the energy of a test particle moving with the Hubble expansion, define a physical energy E_{phys} by

$$E_{\text{phys}} \equiv m r_i^2 E \ , \tag{3}$$

where m is the mass of the test particle and r_i is its initial radius. Note that the gravitational force on this particle is given by

$$\vec{F} = -\frac{GmM(r_i)}{r^2}\hat{r} = -\vec{\nabla}V_{\text{eff}}(r) \ , \tag{4}$$

where $M(r_i)$ is the total mass initially contained within a radius r_i of the origin, r is the present distance of the test particle from the origin, and the "effective" potential energy $V_{\text{eff}}(r)$ is given by

$$V_{\text{eff}}(r) = -\frac{GmM(r_i)}{r} \ . \tag{5}$$

The motivation for calling this quantity the "effective" potential energy will be explained below.

(a) *(10 points)* Show that E_{phys} is equal to the "effective" energy of the test particle, defined by

$$E_{\text{eff}} = \frac{1}{2}mv^2 + V_{\text{eff}}(r) \ . \tag{6}$$

We understand that $E_{\rm eff}$ is conserved because it is the energy in an analogue problem in which the test particle moves in the gravitational field of a point particle of mass $M(r_i)$, located at the origin, with potential energy function $V_{\rm eff}(r)$. In this analogue problem the force on the test particle is exactly the same as in the real problem, but in the analogue problem the energy of the test particle is conserved.

We call (6) the "effective" energy because it is really the energy of the analogue problem, and not the real problem. The true potential energy $V(r,t)$ of the test particle is defined to be the amount of work we must supply to move the particle to its present location from some fixed reference point, which we might take to be $r = \infty$. We will not bother to write $V(r,t)$ explicitly, since we will not need it, but we point out that it depends on the time t and on $R_{\rm max}$, and when differentiated gives the correct gravitational force at any radius. By contrast, $V_{\rm eff}(r)$ gives the correct force only at the radius of the test particle, $r = a(t)r_i$. The true potential energy function $V(r,t)$ gives no conservation law, since it is explicitly time-dependent, which is why the quantity $V_{\rm eff}(r)$ is useful.

To relate E to the total energy of the expanding sphere, we need to integrate over the sphere to determine its total energy. These integrals are most easily carried out by dividing the sphere into shells of radius r, and thickness dr, so that each shell has a volume

$$dV = 4\pi r^2 \, dr \ . \tag{7}$$

(b) *(10 points)* Show that the total kinetic energy K of the sphere is given by

$$K = c_K M R^2_{{\rm max},i} \left\{ \frac{1}{2} \dot{a}^2(t) \right\} \ , \tag{8}$$

where c_K is a numerical constant, M is the total mass of the sphere, and $R_{{\rm max},i}$ is the initial radius of the sphere. Evaluate the numerical constant c_K.

(c) *(10 points)* Show that the total potential energy of the sphere can similarly be written as

$$U = c_U M R^2_{{\rm max},i} \left\{ -\frac{4\pi}{3} G \, \frac{\rho_i}{a} \right\} \ . \tag{9}$$

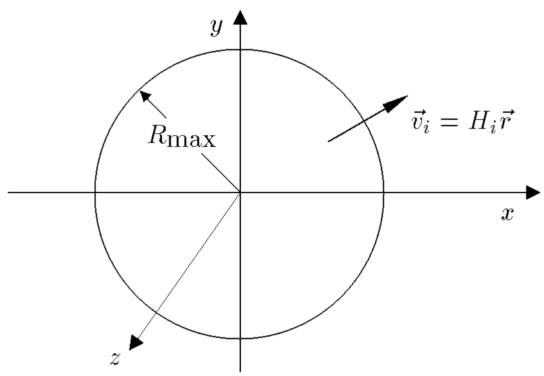
(*Suggestion:* calculate the total energy needed to assemble the sphere by bringing in one shell of mass at a time from infinity.) Show that $c_U = c_K$, so that the total energy of the sphere is given by

$$E_{\rm total} = c_K M R^2_{{\rm max},i} \, E \ . \tag{10}$$

PROBLEM 4: A POSSIBLE MODIFICATION OF NEWTON'S LAW OF GRAVITY *(20 points)*

READ THIS: *This problem was Problem 2 of Quiz 1 of 2011, and the solution is posted as http://web.mit.edu/8.286/www/quiz11/ecqs1-1.pdf. Unlike the situation with other problems, in this case you are encouraged to look at these solutions and benefit from them. When you write your solution, you can even copy it verbatim from these solutions if you wish, although obviously you will learn more if you think about the solution and write your own version.*

In Lecture Notes 3 we developed a Newtonian model of cosmology, by considering a uniform sphere of mass, centered at the origin, with initial mass density ρ_i and an initial pattern of velocities corresponding to Hubble expansion: $\vec{v}_i = H_i\vec{r}$:



We denoted the radius at time t of a particle which started at radius r_i by the function $r(r_i, t)$. Assuming Newton's law of gravity, we concluded that each particle would experience an acceleration given by

$$\vec{g} = -\frac{GM(r_i)}{r^2(r_i,t)}\,\hat{r} \ ,$$

where $M(r_i)$ denotes the total mass contained initially in the region $r < r_i$, given by

$$M(r_i) = \frac{4\pi}{3} r_i^3 \rho_i \ .$$

Suppose that the law of gravity is modified to contain a new, repulsive term, producing an acceleration which grows as the nth power of the distance, with a strength that is independent of the mass. That is, suppose $\vec{g}$ is given by

$$\vec{g} = -\frac{GM(r_i)}{r^2(r_i,t)}\,\hat{r} + \gamma r^n(r_i,t)\,\hat{r} \ ,$$

where γ is a constant. The function $r(r_i, t)$ then obeys the differential equation

$$\ddot{r} = -\frac{GM(r_i)}{r^2(r_i, t)} + \gamma r^n(r_i, t) \ .$$

(a) *(4 points)* As done in the lecture notes, we define

$$u(r_i, t) \equiv r(r_i, t)/r_i \ .$$

Write the differential equation obeyed by u. *(Hint: be sure that u is the only time-dependent quantity in your equation; r, ρ, etc. must be rewritten in terms of u, ρ_i, etc.)*

(b) *(4 points)* For what value of the power n is the differential equation found in part (a) independent of r_i?

(c) *(4 points)* Write the initial conditions for u which, when combined with the differential equation found in (a), uniquely determine the function u.

(d) *(8 points)* If all is going well, then you have learned that for a certain value of n, the function $u(r_i, t)$ will in fact not depend on r_i, so we can define

$$a(t) \equiv u(r_i, t) \ .$$

Show, for this value of n, that the differential equation for a can be integrated once to obtain an equation related to the conservation of energy. The desired equation should include terms depending on a and $\dot{a}$, but not $\ddot{a}$ or any higher derivatives.

Total points for Problem Set 3: 85.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — October 6, 2018
Prof. Alan Guth

PROBLEM SET 4

Revised Version*

DUE DATE: Friday, October 12, 2018, at 5:00 pm.

READING ASSIGNMENT: Steven Weinberg, *The First Three Minutes*, Chapter 4; Barbara Ryden, *Introduction to Cosmology*, Chapters 4 and 5 and Sec. 6.1. In Weinberg's Chapter 4 (and, later, Chapter 5) there are a lot of numbers mentioned. You certainly do not need to learn all these numbers, but you should be familiar with the orders of magnitude. In Ryden's Chapters 4 and 5 (and, later, Chapter 6), the material parallels what we either have done or will be doing in lecture. For these chapters you should consider Ryden's book as an aid to understanding the lecture material, and not as a source of new material. On the upcoming quizzes, there will be no questions based specifically on the material in these chapters.

SHORT-TERM CALENDAR:

OCTOBER/NOVEMBER				
MON	TUES	WED	THURS	FRI
October 8 Columbus Day	9	10 Lecture 9	11	12 **PS 4 due**
October 15 Lecture 10	16	17 Lecture 11	18	19 **PS 5 due**
October 22 Lecture 12	23	24 Lecture 13	25	26
October 29 Lecture 14	30 **PS 6 due**	31 Lecture 15	November 1	2
November 5 Quiz 2 — in class	6	7	8	9

* Revised October 11, 2018: In Problem 1(h), $\mathrm{d}\ell_{p,\gamma B}(t)/dt$ had been mistakenly mistyped (twice) as $\mathrm{d}\ell_{p,\gamma B}(t)$.

QUIZ DATES FOR THE TERM:

Quiz 1: Wednesday, October 3, 2018
Quiz 2: Monday, November 5, 2018
Quiz 3: Wednesday, December 5, 2018

PROBLEM 1: PHOTON TRAJECTORIES AND HORIZONS IN A FLAT UNIVERSE WITH $a(t) = bt^{1/2}$ *(20 points)*

The following questions all pertain to a flat universe, with a scale factor given by

$$a(t) = bt^{1/2} ,$$

where b is a constant and t is the time. We will learn later that this is the behavior of a radiation-dominated flat universe.

(a) *(2 points)* If physical lengths are measured in meters, and coordinate lengths are measured in notches, what are the units of $a(t)$ and the constant b?

(b) *(2 points)* Find the Hubble expansion rate $H(t)$.

(c) *(2 points)* Find the physical horizon distance $\ell_{p,\mathrm{hor}}(t)$. Your answer should give the horizon distance in physical units (e.g., meters) and not coordinate units (e.g., notches).

Consider two pieces of matter, A and B, at a coordinate distance ℓ_c from each other. We will consider a photon that is emitted by A at some early time t_A, traveling toward B. The physical distance between A and B at the time of emission is of course $\ell_{p,AB}(t_A) = bt_A^{1/2}\ell_c$, which approaches zero as $t_A \to 0$.

(d) *(2 points)* What is the rate of change of the physical distance between A and B, $\mathrm{d}\ell_{p,AB}(t)/\mathrm{d}t$, at $t = t_A$? Is the physical distance increasing or decreasing? Does the rate of change approach zero, infinity, negative infinity, or a nonzero finite number as $t_A \to 0$?

(e) *(3 points)* At what time t_B is the photon received by B? As $t_A \to 0$, does t_B approach zero, infinity, or a nonzero finite number?

(f) *(3 points)* Calculate $\ell_{p,\gamma B}(t)$, the physical distance between the photon and B at time t, for $t_A \le t \le t_B$.

(g) *(3 points)* What is the rate of change of the physical distance between the photon and B, $\mathrm{d}\ell_{p,\gamma B}(t)/\mathrm{d}t$, at the instant t_A when the photon is emitted?

(h) *(3 points)* At what value of t_A is this rate of change $\mathrm{d}\ell_{p,\gamma B}(t)/\mathrm{d}t$ equal to zero? For earlier values of t_A, is the physical distance between the photon and B increasing or decreasing at the time of emission? As $t_A \to 0$, does $\mathrm{d}\ell_{p,\gamma B}(t)/\mathrm{d}t$ at the time of emission approach zero, infinity, minus infinity, or a nonzero finite number?

PROBLEM 2: EVOLUTION OF AN OPEN, MATTER-DOMINATED UNIVERSE *(35 points)*

The following problem originated on Quiz 2 of 1992 (ancient history!), where it counted 30 points.

The equations describing the evolution of an open, matter-dominated universe were given in Lecture Notes 4 as

$$ct = \alpha\,(\sinh\theta - \theta)$$

and

$$\frac{a}{\sqrt{\kappa}} = \alpha\,(\cosh\theta - 1) \ ,$$

where α is a constant with units of length. The following mathematical identities, which you should know, may also prove useful on parts (e) and (f):

$$\sinh\theta = \frac{e^{\theta} - e^{-\theta}}{2} \quad , \quad \cosh\theta = \frac{e^{\theta} + e^{-\theta}}{2}$$

$$e^{\theta} = 1 + \frac{\theta}{1!} + \frac{\theta^2}{2!} + \frac{\theta^3}{3!} + \dots \ .$$

a) *(5 points)* Find the Hubble expansion rate H as a function of α and θ.

b) *(5 points)* Find the mass density ρ as a function of α and θ.

c) *(5 points)* Find the mass density parameter Ω as a function of α and θ. As with part (c) of the previous problem, the answer to this part appears in Lecture Notes 4. However, you should show that you get the same answer by combining your answers to parts (a) and (b) of this question.

d) *(6 points)* Find the physical value of the horizon distance, $\ell_{p,\text{horizon}}$, as a function of α and θ.

e) *(7 points)* For very small values of t, it is possible to use the first nonzero term of a power-series expansion to express θ as a function of t, and then a as a function of t. Give the expression for $a(t)$ in this approximation. The approximation will be valid for $t \ll t^*$. Estimate the value of t^*.

f) *(7 points)* Even though these equations describe an open universe, one still finds that Ω approaches one for very early times. For $t \ll t^*$ (where t^* is defined in part (e)), the quantity $1 - \Omega$ behaves as a power of t. Find the expression for $1 - \Omega$ in this approximation.

PROBLEM 3: THE CRUNCH OF A CLOSED, MATTER-DOMINATED UNIVERSE *(25 points)*

This is Problem 6.5 from Barbara Ryden's **Introduction to Cosmology***, with some paraphrasing to make it consistent with the language used in lecture.*

Consider a closed universe containing only nonrelativistic matter. This is the closed universe discussed in Lecture Notes 4, and it is also the "Big Crunch" model discussed in Ryden's section 6.1. At some time during the contracting phase (i.e., when $\theta > \pi$), an astronomer named Elbbuh Niwde discovers that nearby galaxies have blueshifts ($-1 \leq z < 0$) proportional to their distance. He then measures the present values of the Hubble expansion rate, H_0, and the mass density parameter, Ω_0. He finds, of course, that $H_0 < 0$ (because he is in the contracting phase) and $\Omega_0 > 1$ (because the universe is closed). In terms of H_0 and Ω_0, how long a time will elapse between Dr. Niwde's observation at $t = t_0$ and the final Big Crunch at $t = t_{\text{Crunch}} = 2\pi\alpha/c$? Assuming that Dr. Niwde is able to observe all objects within his horizon, what is the most blueshifted (i.e., most negative) value of z that Dr. Niwde is able to see? What is the lookback time to an object with this blueshift? (By lookback time, one means the difference between the time of observation t_0 and the time at which the light was emitted.)

PROBLEM 4: THE AGE OF A MATTER-DOMINATED UNIVERSE AS $\mathbf{\Omega \to 1}$ *(15 points)*

The age t of a matter-dominated universe, for any value of Ω, was given in Lecture Notes 4 as

$$|H|\,t = \begin{cases} \dfrac{\Omega}{2(1-\Omega)^{3/2}}\left[\dfrac{2\sqrt{1-\Omega}}{\Omega} - \operatorname{arcsinh}\left(\dfrac{2\sqrt{1-\Omega}}{\Omega}\right)\right] & \text{if } \Omega < 1 \\ 2/3 & \text{if } \Omega = 1 \\ \dfrac{\Omega}{2(\Omega-1)^{3/2}}\left[\arcsin\left(\pm\dfrac{2\sqrt{\Omega-1}}{\Omega}\right) \mp \dfrac{2\sqrt{\Omega-1}}{\Omega}\right] & \text{if } \Omega > 1 \end{cases} \tag{4.47}$$

It was claimed that this formula is continuous at $\Omega = 1$. In this problem you are asked to show half of this statement. Specifically, you should show that as Ω approaches 1 from below, the expression for $|H|t$ approaches $2/3$. In doing this, you may find it useful to use the Taylor expansion for $\operatorname{arcsinh}(x)$ about $x = 0$:

$$\operatorname{arcsinh}(x) = x - \frac{(1)^2}{3!}x^3 + \frac{(3\cdot 1)^2}{5!}x^5 - \frac{(5\cdot 3\cdot 1)^2}{7!}x^7 + \dots\ .$$

The proof of continuity as $\Omega \to 0$ from above is of course very similar, and you are not asked to show it.

PROBLEM 5: ISOTROPY ABOUT TWO POINTS IN EUCLIDEAN SPACES

(This problem is not required, but can be done for 15 points extra credit. I'd like to give you two weeks to think about it, so you should turn it in with Problem Set 5 on October 19.)

In Steven Weinberg's *The First Three Minutes,* in Chapter 2 on page 24, he gives an argument to show that if a space is isotropic about two distinct points, then it is necessarily homogeneous. He is assuming Euclidean geometry, although he is not explicit about this point. (The statement is simply not true if one allows non-Euclidean spaces — we'll discuss this.) Furthermore, the argument is given in the context of a universe with only two space dimensions, but it could easily be generalized to three, and we will not concern ourselves with remedying this simplification. The statement is true for two-dimensional Euclidean spaces, but Weinberg's argument is not complete. To show that isotropy about two galaxies, 1 and 2, implies that the conditions at any two points A and B must be identical, he constructs two circles. One circle is centered on Galaxy 1 and goes through A, and the other is centered on Galaxy 2 and goes through B. He then argues that the conditions at A and B must both be identical to the conditions at the point C, where the circles intersect. The problem, however, is that the two circles need not intersect. One circle can be completely inside the other, or the two circles can be separated and disjoint. Thus Weinberg's proof is valid for some pairs of points A and B, but cannot be applied to all cases. For 15 points of extra credit, devise a proof that holds in all cases. We have not established axioms for Euclidean geometry, but you may use in your proof any well-known fact about Euclidean geometry.

Total points for Problem Set 4: 95, plus 15 points of extra credit.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — October 13, 2018
Prof. Alan Guth

PROBLEM SET 5

DUE DATE: Friday, October 19, 2018, at 5:00 pm.

READING ASSIGNMENT: Steven Weinberg, *The First Three Minutes,* Chapters 5 and 6, and also Barbara Ryden, *Introduction to Cosmology,* Chapter 10. We are skipping Chapters 7-9 of Ryden for now, but we will come back to them. Chapter 10, about *Nucleosynthesis and the Early Universe,* makes good parallel reading to Weinberg's book, and really has no dependence on the chapters that we are skipping.

SHORT-TERM CALENDAR:

OCTOBER/NOVEMBER				
MON	TUES	WED	THURS	FRI
October 8 Columbus Day	9	10 Lecture 9	11	12 **PS 4 due**
October 15 Lecture 10	16	17 Lecture 11	18	19 **PS 5 due**
October 22 Lecture 12	23	24 Lecture 13	25	26
October 29 Lecture 14	30 **PS 6 due**	31 Lecture 15	November 1	2
November 5 Quiz 2 — in class	6	7	8	9

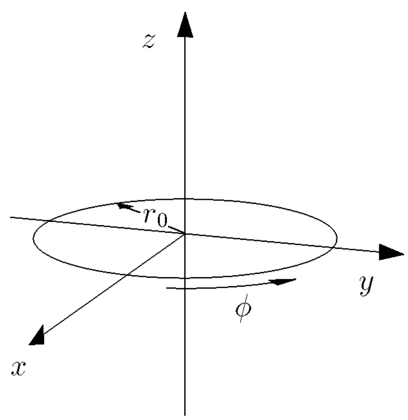
PROBLEM 1: A CIRCLE IN A NON-EUCLIDEAN GEOMETRY
(15 points)

Consider a three-dimensional space described by the following metric:

$$ds^2 = R^2 \left\{ \frac{dr^2}{1-kr^2} + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2 \right) \right\} \ .$$

Here R and k are constants, where k will always have one of the values 1, -1, or 0. θ and ϕ are angular coordinates with the usual properties: $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$, where $\phi = 2\pi$ and $\phi = 0$ are identified. r is a radial coordinate, which runs from 0 to 1 if $k = 1$, and otherwise from 0 to ∞. (This is the Robertson-Walker metric of Eq. (5.27) of Lecture Notes 5, evaluated at some particular time t, with $R \equiv a(t)$. You should be able to work this problem, however, whether or not you have gotten that far. The problem requires only that you understand what a metric means.) Consider a circle described by the equations

$$z = 0$$
$$x^2 + y^2 = r_0^2 \ ,$$



or equivalently by the angular coordinates

$$r = r_0$$
$$\theta = \pi/2 \ .$$

(a) *(5 points)* Find the circumference S of this circle. Hint: break the circle into infinitesimal segments of angular size $d\phi$, calculate the arc length of such a segment, and integrate.

(b) *(5 points)* Find the radius ρ of this circle. Note that ρ is the length of a line which runs from the origin to the circle ($r = r_0$), along a trajectory of $\theta = \pi/2$ and $\phi =$ constant. Hint: Break the line into infinitesimal segments of coordinate length dr, calculate the length of such a segment, and integrate. Consider the case of open and closed universes separately, and take $k = \pm 1$. (If you don't remember why we can take $k = \pm 1$, see the section called "Units" in Lecture Notes 3,). You will want the following integrals:

$$\int \frac{dr}{\sqrt{1-r^2}} = \sin^{-1} r$$

and

$$\int \frac{dr}{\sqrt{1+r^2}} = \sinh^{-1} r \quad .$$

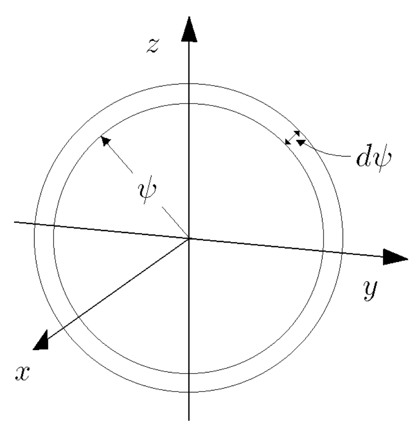
(c) *(5 points)* Express the circumference S in terms of the radius ρ. This result is independent of the coordinate system which was used for the calculation, since S and ρ are both measurable quantities. Since the space described by this metric is homogeneous and isotropic, the answer does not depend on where the circle is located or on how it is oriented. For the two cases of open and closed universes, state whether S is larger or smaller than the value it would have for a Euclidean circle of radius ρ.

PROBLEM 2: VOLUME OF A CLOSED UNIVERSE *(15 points)*

Calculate the total volume of a closed universe, as described by the metric of Eq. (5.14) of Lecture Notes 5:

$$ds^2 = R^2 \left[d\psi^2 + \sin^2\psi \left(d\theta^2 + \sin^2\theta d\phi^2 \right) \right] \quad .$$

Break the volume up into spherical shells of infinitesimal thickness, extending from ψ to $\psi + d\psi$:



By comparing Eq. (5.14) with Eq. (5.8),

$$\mathrm{d}s^2 = R^2 \left(\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\phi^2\right) ,$$

the metric for the surface of a sphere, one can see that as long as ψ is held fixed, the metric for varying θ and ϕ is the same as that for a spherical surface of radius $R\sin\psi$. Thus the area of the spherical surface is $4\pi R^2 \sin^2\psi$. To find the volume, multiply this area by the thickness of the shell (which you can read off from the metric), and then integrate over the full range of ψ, from 0 to π.

PROBLEM 3: SURFACE BRIGHTNESS IN A CLOSED UNIVERSE *(25 points)*

The spacetime metric for a homogeneous, isotropic, closed universe is given by the Robertson-Walker formula:

$$ds^2 = -c^2\,d\tau^2 = -c^2\,dt^2 + a^2(t)\left\{\frac{dr^2}{1-r^2} + r^2\left(d\theta^2 + \sin^2\theta\,d\phi^2\right)\right\} ,$$

where I have taken $k = 1$. To discuss motion in the radial direction, it is more convenient to work with an alternative radial coordinate ψ, related to r by

$$r = \sin\psi .$$

Then

$$\frac{dr}{\sqrt{1-r^2}} = d\psi ,$$

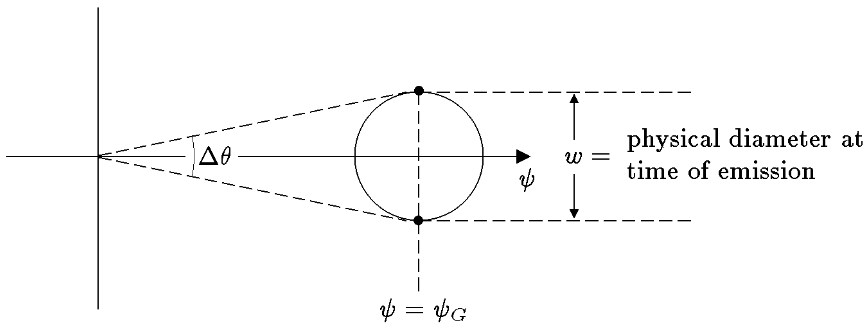
so the metric simplifies to

$$ds^2 = -c^2\,d\tau^2 = -c^2\,dt^2 + a^2(t)\left\{d\psi^2 + \sin^2\psi\left(d\theta^2 + \sin^2\theta\,d\phi^2\right)\right\} .$$

The form of $a(t)$ depends on the nature of the matter in the universe, but for this problem you should consider $a(t)$ to be an arbitrary function. You should simplify your answers as far as it is possible without knowing the function $a(t)$.

(a) *(10 points)* Suppose that the Earth is at the center of these coordinates, and that we observe a spherical galaxy that is located at $\psi = \psi_G$. The light that we see was emitted from the galaxy at time t_G, and is being received today, at a time that we call t_0. At the time of emission, the galaxy had a power output P (which could be measured, for example, in watts, where 1 watt = 1 joule/sec). The power was radiated uniformly in all directions, in the form of photons. What is the radiation energy flux J from this galaxy at the Earth today? Energy flux (which might be measured in joule-m^{-2}-sec^{-1}) is defined as the energy per unit area per unit time striking a surface that is orthogonal to the direction of the energy flow. [Hint: it is easiest to use a comoving coordinate system with the radiating galaxy at the origin.]

(b) *(10 points)* Suppose that the physical diameter of the galaxy at time t_G was w. Find the apparent angular size $\Delta\theta$ (measured from one edge to the other) of the galaxy as it would be observed from Earth today.



(c) *(5 points)* The *surface brightness* σ of the distant galaxy is defined to be the energy flux J per solid angle subtended by the galaxy.* Calculate the surface brightness σ of the galaxy described in parts (a) and (b). [Hint: if you have the right answer, it can be written in terms of P, w, and the redshift z, without any reference to ψ_G. The rapid decrease in σ with z means that high-z galaxies are difficult to distinguish from the night sky.]

PROBLEM 4: TRAJECTORIES AND DISTANCES IN AN OPEN UNIVERSE *(30 points)*

The spacetime metric for a homogeneous, isotropic, open universe is given by the Robertson-Walker formula:

$$ds^2 = -c^2\,d\tau^2 = -c^2\,dt^2 + a^2(t)\left\{\frac{dr^2}{1+r^2} + r^2\left(d\theta^2 + \sin^2\theta\,d\phi^2\right)\right\} ,$$

where I have taken $k = -1$. As in Problem 3, for the discussion of radial motion it is convenient to introduce an alternative radial coordinate ψ, which in this case is related to r by

$$r = \sinh\psi \; .$$

* *Definition of solid angle:* To define the solid angle subtended by the galaxy, imagine surrounding the observer by a sphere of arbitrary radius r. The sphere should be small compared to cosmological distances, so that Euclidean geometry is valid within the sphere. If a picture of the galaxy is painted on the surface of the sphere so that it just covers the real image, then the solid angle, in steradians, is the area of the picture on the sphere, divided by r^2.

Then

$$\frac{dr}{\sqrt{1+r^2}} = d\psi \ ,$$

so the metric simplifies to

$$ds^2 = -c^2\,d\tau^2 = -c^2\,dt^2 + a^2(t)\left\{d\psi^2 + \sinh^2\psi\left(d\theta^2 + \sin^2\theta\,d\phi^2\right)\right\} \ .$$

You should treat the function $a(t)$ as a given function. You should simplify your answers as far as it is possible without knowing explicitly the function $a(t)$.

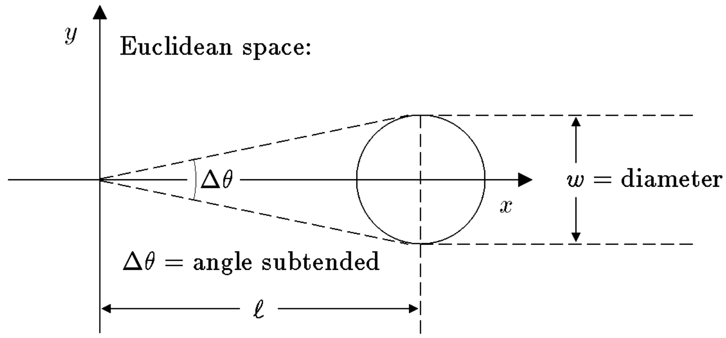
(a) *(5 points)* Suppose that the Earth is at the origin of the coordinate system ($\psi = 0$), and that at the present time, t_0, we receive a light pulse from a distant galaxy G, located at $\psi = \psi_G$. Write down an equation which determines the time t_G at which the light pulse left the galaxy. (You may assume that the light pulse travels on a "null" trajectory, which means that $d\tau = 0$ for any segment of it. Since you don't know $a(t)$ you cannot solve this equation, so please do not try.)

(b) *(5 points)* What is the redshift z_G of the light from galaxy G? (Your answer may depend on t_G, as well as ψ_G, t_0, or any property of the function $a(t)$.)

(c) *(5 points)* To estimate the number of galaxies that one expects to see in a given range of redshifts, it is necessary to know the volume of the region of space that corresponds to this range. Write an expression for the present value of the volume that corresponds to redshifts smaller than that of galaxy G. (You may leave your answer in the form of a definite integral, which may be expressed in terms of ψ_G, t_G, t_0, z_G, or the function $a(t)$.)

(d) *(5 points)* There are a number of different ways of defining distances in cosmology, and generally they are not equal to each other. One choice is called ***proper distance***, which corresponds to the distance that one could in principle measure with rulers. The proper distance is defined as the total length of a network of rulers that are laid end to end from here to the distant galaxy. The rulers have different velocities, because each is at rest with respect to the matter in its own vicinity. They are arranged so that, at the present instant of time, each ruler just touches its neighbors on either side. Write down an expression for the proper distance $\ell_{\rm prop}$ of galaxy G.

(e) *(5 points)* Another common definition of distance is ***angular size distance,*** determined by measuring the apparent size of an object of known physical size. In a static, Euclidean space, a small sphere of diameter w at a distance ℓ will subtend an angle $\Delta\theta = w/\ell$:

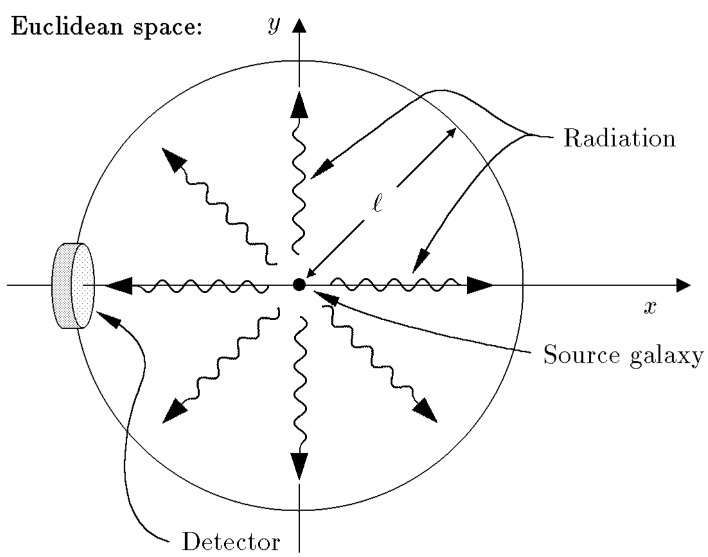


Motivated by this relation, cosmologists define the angular size distance $\ell_{\rm ang}$ of an object by

$$\ell_{\rm ang} \equiv \frac{w}{\Delta\theta} \ .$$

What is the angular size distance $\ell_{\rm ang}$ of galaxy G?

(f) *(5 points)* A third common definition of distance is called ***luminosity distance***, which is determined by measuring the apparent brightness of an object for which the actual total power output is known. In a static, Euclidean space, the energy flux J received from a source of power P at a distance ℓ is given by $J = P/(4\pi\ell^2)$:



Cosmologists therefore define the luminosity distance by

$$\ell_{\rm lum} \equiv \sqrt{\frac{P}{4\pi J}} \ .$$

Find the luminosity distance $\ell_{\rm lum}$ of galaxy G. (Hint: the Robertson-Walker coordinates can be shifted so that the galaxy G is at the origin.)

PROBLEM 5: THE KLEIN DESCRIPTION OF THE G-B-L GEOMETRY

(This problem is not required, but can be done for 15 points extra credit.)

I stated in Lecture Notes 5 that the space invented by Klein, described by the distance relation

$$\cosh\left[\frac{d(1,2)}{a}\right] = \frac{1 - x_1 x_2 - y_1 y_2}{\sqrt{1 - x_1^2 - y_1^2}\,\sqrt{1 - x_2^2 - y_2^2}} \quad ,$$

where

$$x^2 + y^2 < 1 \quad ,$$

is a two-dimensional space of constant negative curvature. In other words, this is just a two-dimensional Robertson–Walker metric, as would be described by a two-dimensional version of Eq. (5.27), with $k = -1$:

$$ds^2 = a^2 \left\{ \frac{dr^2}{1 + r^2} + r^2 d\theta^2 \right\} \quad .$$

The problem is to prove the equivalence.

(a) *(5 points)* As a first step, show that if x and y are replaced by the polar coordinates defined by

$$x = u \cos\theta$$

$$y = u \sin\theta \quad ,$$

then the distance equation can be rewritten as

$$\cosh\left[\frac{d(1,2)}{a}\right] = \frac{1 - u_1 u_2 \cos(\theta_1 - \theta_2)}{\sqrt{1 - u_1^2}\,\sqrt{1 - u_2^2}} \quad .$$

(b) *(5 points)* The next step is to derive the metric from the distance function above. Let

$$u_1 = u \qquad \theta_1 = \theta \quad ,$$

$$u_2 = u + du \qquad \theta_2 = \theta + d\theta \quad ,$$

and

$$d(1,2) = ds \quad .$$

Insert these expressions into the distance function, expand everything to second order in the infinitesimal quantities, and show that

$$ds^2 = a^2 \left\{ \frac{du^2}{(1 - u^2)^2} + \frac{u^2 d\theta^2}{1 - u^2} \right\} \quad .$$

(This part is rather messy, but you should be able to do it.)

(c) *(5 points)* Now find the relationship between r and u and show that the two metric functions are identical. Hint: The coefficients of $d\theta^2$ must be the same in the two cases. Can you now see why Klein had to impose the condition $x^2 + y^2 < 1$?

REMINDER: The following extra credit problem from Problem Set 4 is to be turned in with this problem set, if you choose to do it:

PROBLEM 5 (PROBLEM SET 4): ISOTROPY ABOUT TWO POINTS IN EUCLIDEAN SPACES

(This problem is not required, but can be done for 15 points extra credit. It was first posted with Problem Set 4, but is to be turned in with Problem Set 5.)

In Steven Weinberg's *The First Three Minutes,* in Chapter 2 on page 24, he gives an argument to show that if a space is isotropic about two distinct points, then it is necessarily homogeneous. He is assuming Euclidean geometry, although he is not explicit about this point. (The statement is simply not true if one allows non-Euclidean spaces — we'll discuss this.) Furthermore, the argument is given in the context of a universe with only two space dimensions, but it could easily be generalized to three, and we will not concern ourselves with remedying this simplification. The statement is true for two-dimensional Euclidean spaces, but Weinberg's argument is not complete. To show that isotropy about two galaxies, 1 and 2, implies that the conditions at any two points A and B must be identical, he constructs two circles. One circle is centered on Galaxy 1 and goes through A, and the other is centered on Galaxy 2 and goes through B. He then argues that the conditions at A and B must both be identical to the conditions at the point C, where the circles intersect. The problem, however, is that the two circles need not intersect. One circle can be completely inside the other, or the two circles can be separated and disjoint. Thus Weinberg's proof is valid for some pairs of points A and B, but cannot be applied to all cases. For 15 points of extra credit, devise a proof that holds in all cases. We have not established axioms for Euclidean geometry, but you may use in your proof any well-known fact about Euclidean geometry.

Total points for Problem Set 5: 85, plus up to 15 points extra credit.
Also up to 15 points extra credit for Problem Set 4.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — October 20, 2018
Prof. Alan Guth

PROBLEM SET 6

DUE DATE: Tuesday, October 30, 2018, at 5:00 pm. This problem set is longer than usual, since you have more than one week to do it. It will be the last problem set before Quiz 2. Due to the way holidays fall, Problem Sets 7 and 8 will also be longer than usual.

READING ASSIGNMENT: Steven Weinberg, *The First Three Minutes,* Chapter 7 (*The First One-Hundredth Second*), and also Barbara Ryden, *Introduction to Cosmology,* Chapter 8 (*Dark Matter*).

CALENDAR FOR THE REST OF THE TERM:

OCTOBER/NOVEMBER				
MON	TUES	WED	THURS	FRI
October 15 Lecture 10	16	17 Lecture 11	18	19 **PS 5 due**
October 22 Lecture 12	23	24 Lecture 13	25	26
October 29 Lecture 14	30 **PS 6 due**	31 Lecture 15	November 1	2
November 5 **Quiz 2** — in class	6	7 Lecture 16	8	9
November 12 Veteran's Day	13	14 Lecture 17	15	16 **PS 7 due**
November 19 Lecture 18	20	21 Lecture 19	22 Thanksgiving	23 Thanksgiving
November 26 Lecture 20	27	28 Lecture 21	29	30 **PS 8 due**

DECEMBER				
MON	TUES	WED	THURS	FRI
December 3 Lecture 22	4	5 **Quiz 3**	6	7
December 10 Lecture 23	11	12 Last Class **PS 9 due**	13	14

PROBLEM 1: GEODESICS IN A FLAT UNIVERSE *(25 points)*

According to general relativity, in the absence of any non-gravitational forces a particle will travel along a spacetime geodesic. In this sense, gravity is reduced to a distortion in spacetime.

Consider the case of a flat (*i.e.*, $k = 0$) Robertson–Walker metric, which has the simple form

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

Since the spatial metric is flat, we have the option of writing it in terms of Cartesian rather than polar coordinates. Now consider a particle which moves along the x-axis. (Note that the galaxies are on the average at rest in this system, but one can still discuss the trajectory of a particle which moves through the model universe.)

(a) *(8 points)* Use the geodesic equation to show that the coordinate velocity computed with respect to proper time (*i.e.*, $dx/d\tau$) falls off as $1/a^2(t)$.

(b) *(8 points)* Use the expression for the spacetime metric to relate dx/dt to $dx/d\tau$.

(c) *(9 points)* The physical velocity of the particle relative to the galaxies that it is passing is given by

$$v = a(t)\frac{dx}{dt} \quad .$$

Show that the momentum of the particle, defined relativistically by

$$p = \frac{mv}{\sqrt{1 - v^2/c^2}}$$

falls off as $1/a(t)$. (This implies, by the way, that if the particle were described as a quantum mechanical wave with wavelength $\lambda = h/|\vec{p}\,|$, then its wavelength would stretch with the expansion of the universe, in the same way that the wavelength of light is redshifted.)

PROBLEM 2: METRIC OF A STATIC GRAVITATIONAL FIELD *(25 points)*

In this problem we will consider the metric

$$ds^2 = -\left[c^2 + 2\phi(\vec{x})\right]\, dt^2 + \sum_{i=1}^{3} \left(dx^i\right)^2 \quad ,$$

which describes a static gravitational field. Here i runs from 1 to 3, with the identifications $x^1 \equiv x$, $x^2 \equiv y$, and $x^3 \equiv z$. The function $\phi(\vec{x})$ depends only on the spatial variables $\vec{x} \equiv (x^1, x^2, x^3)$, and not on the time coordinate t.

(a) *(5 points)* Suppose that a radio transmitter, located at $\vec{x}_e$, emits a series of evenly spaced pulses. The pulses are separated by a proper time interval ΔT_e, as measured by a clock at the same location. What is the coordinate time interval Δt_e between the emission of the pulses? (I.e., Δt_e is the difference between the time coordinate t at the emission of one pulse and the time coordinate t at the emission of the next pulse.)

(b) *(5 points)* The pulses are received by an observer at $\vec{x}_r$, who measures the time of arrival of each pulse. What is the ***coordinate*** time interval Δt_r between the reception of successive pulses?

(c) *(5 points)* The observer uses his own clocks to measure the proper time interval ΔT_r between the reception of successive pulses. Find this time interval, and also the redshift z, defined by

$$1 + z = \frac{\Delta T_r}{\Delta T_e} \quad .$$

First compute an exact expression for z, and then expand the answer to lowest order in $\phi(\vec{x})$ to obtain a weak-field approximation. (This weak-field approximation is in fact highly accurate in all terrestrial and solar system applications.)

(d) *(5 points)* A freely falling particle travels on a spacetime geodesic $x^\mu(\tau)$, where τ is the proper time. (I.e., τ is the time that would be measured by a clock moving with the particle.) The trajectory is described by the geodesic equation

$$\frac{d}{d\tau}\left(g_{\mu\nu}\frac{dx^\nu}{d\tau}\right) = \frac{1}{2}\left(\partial_\mu g_{\lambda\sigma}\right)\frac{dx^\lambda}{d\tau}\frac{dx^\sigma}{d\tau} \quad ,$$

where the Greek indices ($\mu, \nu, \lambda, \sigma$, etc.) run from 0 to 3, and are summed over when repeated. Calculate an explicit expression for

$$\frac{d^2 x^i}{d\tau^2} \; ,$$

valid for $i = 1, 2$, or 3. (It is acceptable to leave quantities such as $dt/d\tau$ or $dx^i/d\tau$ in the answer.)

(e) *(5 points)* In the weak-field nonrelativistic-velocity approximation, the answer to the previous part reduces to

$$\frac{d^2x^i}{dt^2} = -\partial_i \phi \ ,$$

so $\phi(\vec{x})$ can be identified as the Newtonian gravitational potential. Use this fact to estimate the gravitational redshift z of a photon that rises from the floor of this room to the ceiling (say 4 meters). (One significant figure will be sufficient.)

PROBLEM 3: CIRCULAR ORBITS IN A SCHWARZSCHILD METRIC *(30 points)*

READ THIS: *This problem was Problem 16 of Review Problems for Quiz 2 of 2011, and the solution is posted as http://web.mit.edu/8.286/www/quiz11/ecqr2-1.pdf. Like Problem 4 of Problem Set 3, but unlike all other homework problems so far, in this case you are encouraged to look at the solutions and benefit from them. When you write your solution, you can even copy it verbatim from these solutions if you wish, although obviously you will learn more if you think about the solution and write your own version.*

The Schwarzschild metric, which describes the external gravitational field of any spherically symmetric distribution of mass (including black holes), is given by

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

where M is the total mass of the object, $0 \leq \theta \leq \pi$, $0 \leq \phi < 2\pi$, and $\phi = 2\pi$ is identified with $\phi = 0$. We will be concerned only with motion outside the Schwarzschild horizon $R_S = 2GM/c^2$, so we can take $r > R_S$. (This restriction allows us to avoid the complications of understanding the effects of the singularity at $r = R_S$.) In this problem we will use the geodesic equation to calculate the behavior of circular orbits in this metric. We will assume a perfectly circular orbit in the x-y plane: the radial coordinate r is fixed, $\theta = 90°$, and $\phi = \omega t$, for some angular velocity ω.

(a) *(7 points)* Use the metric to find the proper time interval $d\tau$ for a segment of the path corresponding to a coordinate time interval dt. Note that $d\tau$ represents the time that would actually be measured by a clock moving with the orbiting body. Your result should show that

$$\frac{d\tau}{dt} = \sqrt{1 - \frac{2GM}{rc^2} - \frac{r^2\omega^2}{c^2}} \ .$$

Note that for $M = 0$ this reduces to the special relativistic relation $d\tau/dt = \sqrt{1 - v^2/c^2}$, but the extra term proportional to M describes an effect that is new with general relativity— the gravitational field causes clocks to slow down, just as motion does.

(b) *(7 points)* Show that the geodesic equation of motion (Eq. (5.65)) for one of the coordinates takes the form

$$0 = \frac{1}{2}\frac{\partial g_{\phi\phi}}{\partial r}\left(\frac{d\phi}{d\tau}\right)^2 + \frac{1}{2}\frac{\partial g_{tt}}{\partial r}\left(\frac{dt}{d\tau}\right)^2 .$$

(c) *(8 points)* Show that the above equation implies

$$r\left(\frac{d\phi}{d\tau}\right)^2 = \frac{GM}{r^2}\left(\frac{dt}{d\tau}\right)^2 ,$$

which in turn implies that

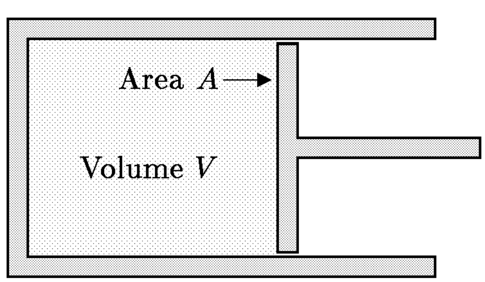
$$r\omega^2 = \frac{GM}{r^2} .$$

Thus, the relation between r and ω is exactly the same as in Newtonian mechanics. *[Note, however, that this does not really mean that general relativity has no effect. First, ω has been defined by $d\phi/dt$, where t is a time coordinate which is not the same as the proper time τ that would be measured by a clock on the orbiting body. Second, r does not really have the same meaning as in the Newtonian calculation, since it is not the measured distance from the center of motion. Measured distances, you will recall, are calculated by integrating the metric, as for example in Problem 1 of Problem Set 5, A Circle in a Non-Euclidean Geometry. Since the angular ($d\theta^2$ and $d\phi^2$) terms in the Schwarzschild metric are unaffected by the mass, however, it can be seen that the circumference of the circle is equal to $2\pi r$, as in the Newtonian calculation.]*

(d) *(8 points)* Show that circular orbits around a black hole have a minimum value of the radial coordinate r, which is larger than R_S. What is it?

PROBLEM 4: GAS PRESSURE AND ENERGY CONSERVATION *(25 points)*

In this problem we will pursue the implications of the conservation of energy. Consider first a gas contained in a chamber with a movable piston, as shown below:



Let U denote the total energy of the gas, and let p denote the pressure. Suppose that the piston is moved a distance dx to the right. (We suppose that the motion is slow, so that the gas particles have time to respond and to maintain a uniform pressure throughout the volume.) The gas exerts a force pA on the piston, so the gas does work $dW = pAdx$ as the piston is moved. Note that the volume increases by an amount $dV = Adx$, so $dW = pdV$. The energy of the gas decreases by this amount, so

$$dU = -pdV \ . \tag{P4.1}$$

It turns out that this relation is valid whenever the volume of a gas is changed, regardless of the shape of the volume.

Now consider a homogeneous, isotropic, expanding universe, described by a scale factor $a(t)$. Let u denote the energy density of the gas that fills it. (Remember that $u = \rho c^2$, where ρ is the mass density of the gas.) We will consider a fixed coordinate volume V_{coord}, so the physical volume will vary as

$$V_{\text{phys}}(t) = a^3(t) V_{\text{coord}} \ . \tag{P4.2}$$

The energy of the gas in this region is then given by

$$U = V_{\text{phys}} u \ . \tag{P4.3}$$

(a) *(9 points)* Using these relations, show that

$$\frac{d}{dt}\left(a^3 \rho c^2\right) = -p \frac{d}{dt}(a^3) \ , \tag{P4.4}$$

and then that

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) \ , \tag{P4.5}$$

where the dot denotes differentiation with respect to t.

(b) *(8 points)* The scale factor evolves according to the relation

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3} G\rho - \frac{kc^2}{a^2} \ . \tag{P4.6}$$

Using Eqs. (P4.5) and (P4.6), show that

$$\ddot{a} = -\frac{4\pi}{3} G \left(\rho + \frac{3p}{c^2}\right) a \ . \tag{P4.7}$$

This equation describes directly the deceleration of the cosmic expansion. Note that there are contributions from the mass density ρ, but also from the pressure p.

(c) *(8 points)* So far our equations have been valid for any sort of a gas, but let us now specialize to the case of black-body radiation. For this case we know that $\rho = bT^4$, where b is a constant and T is the temperature. We also know that as the universe expands, aT remains constant. Using these facts and Eq. (P4.5), find an expression for p in terms of ρ.

PROBLEM 5: THE EFFECT OF PRESSURE ON COSMOLOGICAL EVOLUTION *(25 points)*

A radiation-dominated universe behaves differently from a matter-dominated universe because the pressure of the radiation is significant. In this problem we explore the role of pressure for several fictitious forms of matter.

(a) *(8 points)* For the first fictitious form of matter, the mass density ρ decreases as the scale factor $a(t)$ grows, with the relation

$$\rho(t) \propto \frac{1}{a^6(t)} .$$

What is the pressure of this form of matter? *[Hint: the answer is proportional to the mass density.]*

(b) *(9 points)* Find the behavior of the scale factor $a(t)$ for a flat universe dominated by the form of matter described in part (a). You should be able to determine the function $a(t)$ up to a constant factor.

(c) *(8 points)* Now consider a universe dominated by a different form of fictitious matter, with a pressure given by

$$p = \frac{1}{2}\rho c^2 .$$

As the universe expands, the mass density of this form of matter behaves as

$$\rho(t) \propto \frac{1}{a^n(t)} .$$

Find the power n.

PROBLEM 6: TIME EVOLUTION OF A UNIVERSE WITH MYSTERIOUS STUFF *(15 points)*

Suppose that a model universe is filled with a peculiar form of matter for which

$$\rho \propto \frac{1}{a^5(t)} .$$

Assuming that the model universe is flat, calculate

(a) *(4 points)* The behavior of the scale factor, $a(t)$. You should be able to find $a(t)$ up to an arbitrary constant of proportionality.

(b) *(3 points)* The value of the Hubble parameter $H(t)$, as a function of t.

(c) *(4 points)* The physical horizon distance, $\ell_{p,\mathrm{horizon}}(t)$.

(d) *(4 points)* The mass density $\rho(t)$.

Total points for Problem Set 6: 145.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — November 8, 2018
Prof. Alan Guth

PROBLEM SET 7

DUE DATE: Friday, November 16, 2018, at 4:00 pm.

READING ASSIGNMENT: Steven Weinberg, *The First Three Minutes*, Chapter 8 (*Epilogue: The Prospect Ahead*), and *Afterword: Cosmology Since 1977.* Barbara Ryden, *Introduction to Cosmology*, Chapter 9 (*The Cosmic Microwave Background*).

CALENDAR THROUGH THE END OF THE TERM:

NOVEMBER/DECEMBER				
MON	TUES	WED	THURS	FRI
November 5 **Quiz 2** — in class	6	7 Lecture 16	8	9
November 12 Veteran's Day	13	14 Lecture 17	15	16 **PS 7 due**
November 19 Lecture 18	20	21 Lecture 19	22 Thanksgiving	23 Thanksgiving
November 26 Lecture 20	27	28 Lecture 21	29	30 **PS 8 due**
December 3 Lecture 22	4	5 **Quiz 3** — in class	6	7
December 10 Lecture 23	11	12 Last Class **PS 9 due**	13	14

PROBLEM 1: EFFECT OF AN EXTRA NEUTRINO SPECIES *(15 points)*

According to the standard assumptions (which were used in the lecture notes), there are three species of effectively massless neutrinos. In the temperature range of 1 MeV $< kT <$ 100 MeV, the mass density of the universe is believed to have been dominated by the black-body radiation of photons, electron-positron pairs, and these neutrinos, all of which were in thermal equilibrium.

(a) *(5 points)* Under these assumptions, how long did it take (starting from the instant of the big bang) for the temperature to fall to the value such that $kT = 1$ MeV? (In this part and the next, you may assume that the period when $kT > 100$ MeV was so short that one can calculate as if the value of g that you find for 1 MeV $< kT <$ 100 MeV can be used for earlier times as well.)

(b) *(5 points)* How much time would it have taken if there were one other species of massless neutrino, in addition to the three which we are currently assuming?

(c) *(5 points)* What would be the mass density of the universe when $kT = 1$ MeV under the standard assumptions, and what would it be if there were one other species of massless neutrino?

PROBLEM 2: ENTROPY AND THE BACKGROUND NEUTRINO TEMPERATURE *(15 points)*

The formula for the entropy density of black-body radiation is given in Lecture Notes 6. The derivation of this formula has been left to the statistical mechanics course that you either have taken or hopefully will take. For our purposes, the important point is that the early universe remains very close to thermal equilibrium, and therefore entropy is conserved. The conservation of entropy applies even during periods when particles, such as electron-positron pairs, are "freezing out" of the thermal equilibrium mix. Since total entropy is conserved, the entropy density falls off as $1/a^3(t)$.

When the electron-positron pairs disappear from the thermal equilibrium mixture as kT falls below $m_e c^2 = 0.511$ MeV, the weak interactions have such low cross sections that the neutrinos have essentially decoupled. To a good approximation, all of the energy and entropy released by the annihilation of electrons and positrons is added to the photon gas, and the neutrinos are unaffected. Use the conservation of entropy to show that as electron-positron pair annihilation takes place, aT_γ increases by a factor of $(11/4)^{1/3}$, while aT_ν remains constant. It follows that after the disappearance of the electron-positron pairs, $T_\nu/T_\gamma = (4/11)^{1/3}$. As far as we know, nothing happens that significantly affects this ratio right up to the present day. So we expect today a background of thermal neutrinos which are slightly colder than the 2.7°K background of photons.

Added note: In principle the heating of the photon gas due to electron-positron annihilation can also be calculated by using energy conservation, but it is much more difficult. Since

$$\dot{\rho} = -3H\left(\rho + \frac{p}{c^2}\right)$$

(this was Eq. (6.36) of Lecture Notes 6), one needs to know $p(t)$ to understand the changes in energy density. But as the electron-positron pairs are disappearing, kT is comparable to the electon rest mass $m_e c^2$, and the formula for the thermal equilibrium pressure under these circumstances is complicated.

PROBLEM 3: FREEZE-OUT OF MUONS *(25 points)*

A particle called the muon seems to be essentially identical to the electron, except that it is heavier— the mass/energy of a muon is 106 MeV, compared to 0.511 MeV for the electron. The muon (μ^-) has the same charge as an electron, denoted by $-e$. There is also an antimuon (μ^+), analogous to the positron, with charge $+e$. The muon and antimuon have the same spin as the electron. There is no known particle with a mass between that of an electron and that of a muon.

(a) The formula for the energy density of black-body radiation, as given by Eq. (6.48) of the lecture notes,

$$u = g\frac{\pi^2}{30}\,\frac{(kT)^4}{(\hbar c)^3}\ ,$$

is written in terms of a normalization constant g. What is the value of g for the muons, taking μ^+ and μ^- together?

(b) When kT is just above 106 MeV as the universe cools, what particles besides the muons are contained in the thermal radiation that fills the universe? What is the contribution to g from each of these particles?

(c) As kT falls below 106 MeV, the muons disappear from the thermal equilibrium radiation. At these temperatures all of the other particles in the black-body radiation are interacting fast enough to maintain equilibrium, so the heat given off from the muons is shared among all the other particles. Letting a denote the Robertson-Walker scale factor, by what factor does the quantity aT increase when the muons disappear?

PROBLEM 4: THE REDSHIFT OF THE COSMIC MICROWAVE BACKGROUND *(25 points)*

It was mentioned in Lecture Notes 6 that the black-body spectrum has the peculiar feature that it maintains its form under uniform redshift. That is, as the universe expands, even if the photons do not interact with anything, they will continue to be described by a black-body spectrum, although at a temperature that decreases as the universe expands. Thus, even though the cosmic microwave background (CMB) has not been interacting significantly with matter since 350,000 years after the big bang, the radiation today still has a black-body spectrum. In this problem we will demonstrate this important property of the black-body spectrum.

The spectral energy density $\rho_\nu(\nu, T)$ for the thermal (black-body) radiation of photons at temperature T was stated in Lecture Notes 6 as Eq. (6.71), which we can rewrite as

$$\rho_\nu(\nu, T) = \frac{16\pi^2 \hbar \nu^3}{c^3} \frac{1}{e^{h\nu/kT} - 1} \ , \tag{4.1}$$

where $h = 2\pi\hbar$ is Planck's original constant. $\rho_\nu(\nu, T)\, d\nu$ is the energy per unit volume carried by photons whose frequency is in the interval $[\nu, \nu + d\nu]$. In this problem we will assume that this formula holds at some initial time t_1, when the temperature had some value T_1. We will let $\tilde{\rho}(\nu, t)$ denote the spectral distribution for photons in the universe, which is a function of frequency ν and time t. Thus, our assumption about the initial condition can be expressed as

$$\tilde{\rho}(\nu, t_1) = \rho_\nu(\nu, T_1) \ . \tag{4.2}$$

The photons redshift as the universe expands, and to a good approximation the redshift and the dilution of photons due to the expansion are the only physical effects that cause the distribution of photons to evolve. Thus, using our knowledge of the redshift, we can calculate the spectral distribution $\tilde{\rho}(\nu, t_2)$ at some later time $t_2 > t_1$. It is not obvious that $\tilde{\rho}(\nu, t_2)$ will be a thermal distribution, but in fact we will be able to show that

$$\tilde{\rho}(\nu, t_2) = \rho\big(\nu, T(t_2)\big) \ , \tag{4.3}$$

where in fact $T(t_2)$ will agree with what we already know about the evolution of T in a radiation-dominated universe:

$$T(t_2) = \frac{a(t_1)}{a(t_2)}\, T_1 \ . \tag{4.4}$$

To follow the evolution of the photons from time t_1 to time t_2, we can imagine selecting a region of comoving coordinates with coordinate volume V_c. Within this comoving volume, we can imagine tagging all the photons in a specified infinitesimal range of frequencies, those between ν_1 and $\nu_1 + d\nu_1$. Recalling that the energy of each such photon is $h\nu$, the number dN_1 of tagged photons is then

$$dN_1 = \frac{\tilde{\rho}(\nu_1, t_1)\, a^3(t_1)\, V_c\, d\nu_1}{h\nu_1} \ . \tag{4.5}$$

(a) We now wish to follow the photons in this frequency range from time t_1 to time t_2, during which time each photon redshifts. At the latter time we can denote the range of frequencies by ν_2 to $\nu_2 + d\nu_2$. Express ν_2 and $d\nu_2$ in terms of ν_1 and $d\nu_1$, assuming that the scale factor $a(t)$ is given.

(b) At time t_2 we can imagine tagging all the photons in the frequency range ν_2 to $\nu_2 + d\nu_2$ that are found in the original comoving region with coordinate volume

V_c. Explain why the number dN_2 of such photons, on average, will equal dN_1 as calculated in Eq. (4.5).

(c) Since $\tilde{\rho}(\nu, t_2)$ denotes the spectral energy density at time t_2, we can write

$$dN_2 = \frac{\tilde{\rho}(\nu_2, t_2)\, a^3(t_2)\, V_c\, d\nu_2}{h\nu_2} \;, \tag{4.6}$$

using the same logic as in Eq. (4.5). Use $dN_2 = dN_1$ to show that

$$\tilde{\rho}(\nu_2, t_2) = \frac{a^3(t_1)}{a^3(t_2)}\, \tilde{\rho}(\nu_1, t_1) \;. \tag{4.7}$$

Use the above equation to show that Eq. (4.3) is satisfied, for $T(t)$ given by Eq. (4.4).

Total points for Problem Set 7: 80.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe November 17, 2018
Prof. Alan Guth

PROBLEM SET 8

DUE DATE: Friday, November 30, 2018, at 5:00 pm. This is the last problem set before Quiz 3, which will be Wednesday, December 5. There will also be a Problem Set 10, to be due Wednesday, December 12, the last day of classes.

READING ASSIGNMENT: Barbara Ryden, *Introduction to Cosmology,* Chapter 11 *(Inflation and the Very Early Universe.)* Also read *Inflation and the New Era of High-Precision Cosmology,* by Alan Guth, written for the MIT Physics Department annual newsletter, 2002. It is available at

http://web.mit.edu/physics/news/physicsatmit/physicsatmit_02_cosmology.pdf

The data quoted in the article about the nonuniformities of the cosmic microwave background radiation has since been superceded by much better data, but the conclusions have not changed. They have only gotten stronger.

CALENDAR THROUGH THE END OF THE TERM:

NOVEMBER/DECEMBER				
MON	TUES	WED	THURS	FRI
November 19 Lecture 18	20	21 Lecture 19	22 Thanksgiving	23 Thanksgiving
November 26 Lecture 20	27	28 Lecture 21	29	30 **PS 8 due**
December 3 Lecture 22	4	5 **Quiz 3** — in class	6	7
December 10 Lecture 23	11	12 Last Class **PS 9 due**	13	14

PROBLEM 1: BIG BANG NUCLEOSYNTHESIS *(20 points)*

The calculations of big bang nucleosynthesis depend on a large number of measured parameters. Below you are asked to qualitatively describe the effects of changing some of these parameters. Include a sentence or two to explain each of your answers. (These topics have not been discussed in class, but you are expected to be able to answer the questions on the basis of your readings in Weinberg's and Ryden's books.)

(a) *(5 points)* Suppose an extra neutrino species is added to the calculation. Would the predicted helium abundance go up or down?

(b) *(5 points)* Suppose the weak interactions were stronger than they actually are, so that the thermal equilibrium distribution between neutrons and protons were maintained until $kT \approx 0.25$ MeV. Would the predicted helium abundance be larger or smaller than in the standard model?

(c) *(5 points)* Suppose the proton-neutron mass difference were larger than the actual value of 1.29 MeV/c^2. Would the predicted helium abundance be larger or smaller than in the standard calculation?

(d) *(5 points)* The standard theory of big bang nucleosynthesis assumes that the matter in the universe was distributed homogeneously during the era of nucleosynthesis, but the alternative possibility of inhomogeneous big-bang nucleosynthesis has been discussed since the 1980s. Inhomogeneous nucleosynthesis hinges on the hypothesis that baryons became clumped during a phase transition at $t \approx 10^{-6}$ second, when the hot quark soup converted to a gas of mainly protons, neutrons, and in the early stages, pions. The baryons would then be concentrated in small nuggets, with a comparatively low density outside of these nuggets. After the phase transition but before nucleosynthesis, the neutrons would have the opportunity to diffuse away from these nuggets, becoming more or less uniformly distributed in space. The protons, however, since they are charged, interact electromagnetically with the plasma that fills the universe, and therefore have a much shorter mean free path than the neutrons. Most of the protons, therefore, remain concentrated in the nuggets. Does this scenario result in an increase or a decrease in the expected helium abundance?

PROBLEM 2: BRIGHTNESS VS. REDSHIFT WITH A POSSIBLE COSMOLOGICAL CONSTANT *(25 points)*

In Lecture Notes 7, we derived the relation between the power output P of a source and the energy flux J, for the case of a closed universe:

$$J = \frac{P H_0^2 |\Omega_{k,0}|}{4\pi (1+z_S)^2 c^2 \sin^2 \psi(z_S)} \ ,$$

where

$$\psi(z_S) = \sqrt{|\Omega_{k,0}|} \int_0^{z_S} \frac{dz}{\sqrt{\Omega_{m,0}(1+z)^3 + \Omega_{\text{rad},0}(1+z)^4 + \Omega_{\text{vac},0} + \Omega_{k,0}(1+z)^2}} \ .$$

Here z_S denotes the observed redshift, H_0 denotes the present value of the Hubble expansion rate, $\Omega_{m,0}$, $\Omega_{\text{rad},0}$, and $\Omega_{\text{vac},0}$ denote the present contributions to Ω from nonrelativistic matter, radiation, and vacuum energy, respectively, and $\Omega_{k,0} \equiv 1 - \Omega_{m,0} - \Omega_{\text{rad},0} - \Omega_{\text{vac},0}$.

(a) Derive the corresponding formula for the case of an open universe. You can of course follow the same logic as the derivation in the lecture notes, but the solution you write should be complete and self-contained. (I.e., you should **NOT** say "the derivation is the same as the lecture notes except for ... .")

(b) Derive the corresponding formula for the case of a flat universe. Here there is of course no need to repeat anything that you have already done in part (a). If you wish you can start with the answer for an open or closed universe, taking the limit as $k \to 0$. The limit is delicate, however, because both the numerator and denominator of the equation for J vanish as $\Omega_{k,0} \to 0$.

PROBLEM 3: AGE OF A UNIVERSE WITH MYSTERIOUS STUFF *(20 points)*

READ THIS: *This problem was Problem 8 of Review Problems for Quiz 3 of 2011, and the solution is posted as http://web.mit.edu/8.286/www/quiz11/ecqr3-1.pdf. Like Problem 4 of Problem Set 3 and Problem 3 of Problem Set 6, but unlike all other homework problems so far, in this case you are encouraged to look at the solutions and benefit from them. When you write your solution, you can even copy it verbatim from these solutions if you wish, although obviously you will learn more if you think about the solution and write your own version.*

Consider a universe that contains nonrelativistic matter, radiation, vacuum energy, and the same ***mysterious stuff*** that was introduced in Problem 7 of Review Problems for Quiz 3, from 2011. Since the mass density of mysterious stuff falls off as $1/\sqrt{V}$, where V is the volume, it follows that in an expanding universe the mass density of mysterious stuff falls off as $1/a^{3/2}(t)$.

Suppose that you are given the present value of the Hubble expansion rate H_0, and also the present values of the contributions to $\Omega \equiv \rho/\rho_c$ from each of the constituents: $\Omega_{m,0}$ (nonrelativistic matter), $\Omega_{r,0}$ (radiation), $\Omega_{v,0}$ (vacuum energy density), and $\Omega_{\rm ms,0}$ (mysterious stuff). Our goal is to express the age of the universe t_0 in terms of these quantities.

(a) *(10 points)* Let $x(t)$ denote the ratio

$$x(t) \equiv \frac{a(t)}{a(t_0)}$$

for an arbitrary time t. Write an expression for the total mass density of the universe $\rho(t)$ in terms of $x(t)$ and the given quantities described above.

(b) *(10 points)* Write an integral expression for the age of the universe t_0. The expression should depend only on H_0 and the various contributions to Ω_0 listed above ($\Omega_{m,0}$, $\Omega_{r,0}$, etc.), but it might include x as a variable of integration.

PROBLEM 4: SHARED CAUSAL PAST *(20 points)*

Recently several of my colleagues published a paper (Andrew S. Friedman, David I. Kaiser, and Jason Gallicchio, "The Shared Causal Pasts and Futures of Cosmological Events," http://arxiv.org/abs/arXiv:1305.3943, *Physical Review D,* Vol. 88, article 044038 (2013)) in which they investigated the causal connections in the standard cosmological model. In particular, they calculated the present redshift z of a distant quasar which has the property that a light signal, if sent from our own location at the instant of the big bang, would have just enough time to reach the quasar and return to us, so that we could see the reflection of the signal at the present time. They found $z = 3.65$, using $\Omega_{\rm matter,0} = 0.315$, $\Omega_{\rm rad,0} = 9.29 \times 10^{-5}$, $\Omega_{\rm vac,0} = 0.685 - \Omega_{\rm rad,0}$, and $H_0 = 67.3$ km-s^{-1}-Mpc^{-1}. Feel free to read their paper if you like. Your job, however, is to carry out an independent calculation to find out if they got it right.

(a) *(15 points)* Write an equation that determines this redshift z. The equation may involve one or more integrals which are not evaluated, and the equation itself does not have to be solved.

(b) *(5 points)* The integrals that should appear in your answer to part (a) can be evaluated numerically, and the whole equation you found in part (a) can be solved numerically. Do this, and see how your z compares with 3.65.

PROBLEM 5: MASS DENSITY OF VACUUM FLUCTUATIONS *(25 points)*

The energy density of vacuum fluctuations has been discussed qualitatively in lecture. In this problem we will calculate in detail the energy density associated with quantum fluctuations of the electromagnetic field. To keep the problem finite, we will not consider all of space at once, but instead we will consider the electromagnetic field inside a cube of side L, defined by coordinates

$$0 \leq x \leq L \ ,$$
$$0 \leq y \leq L \ ,$$
$$0 \leq z \leq L \ .$$

Our goal, however, will be to compute the energy density in the limit as the size of the box is taken to infinity.

(a) *(10 points)* The electromagnetic waves inside the box can be decomposed into a Fourier sum of sinusoidal normal modes. Suppose we consider only modes that extend up to a maximum wave number $k_{\rm max}$, or equivalently modes that extend down to a minimum wavelength $\lambda_{\rm min}$, where

$$k_{\rm max} = \frac{2\pi}{\lambda_{\rm min}} \ .$$

How many such modes are there? I do not expect an exact answer, but your approximations should become arbitrarily accurate when $\lambda_{\rm min} \ll L$. (These mode counting techniques are probably familiar to many of you, but in case they are not I have attached an extended hint after part (c).)

(b) *(10 points)* When the electromagnetic field is described quantum mechanically, each normal mode behaves exactly as a harmonic oscillator: if the angular frequency of the mode is ω, then the quantized energy levels have energies given by

$$E_n = \left(n + \tfrac{1}{2}\right) \hbar\omega \ ,$$

where $\hbar$ is Planck's original constant divided by 2π, and n is an integer. The integer n is called the "occupation number," and is interpreted as the number of photons in the specified mode. The minimum energy is not zero, but instead is $\frac{1}{2}\hbar\omega$, which is the energy of the quantum fluctuations of the electromagnetic field. Assuming that the mode sum is cut off at $\lambda_{\rm min}$ equal to the Planck length (as defined in the Lecture Notes), what is the total mass density of these quantum fluctuations?

(c) *(5 points)* How does the mass density of the quantum fluctuations of the electromagnetic field compare with the critical density of our universe?

Extended Hint:

The electromagnetic fields inside a closed box can be expanded as the sum of modes, each of which has a sinusoidally varying time dependence, but the precise form of these modes depends on the nature of the boundary conditions on the walls of the box. Physically reasonable boundary conditions, such as total reflection, are in fact difficult to use in calculations. However, it is known that in the limit of an infinite-sized box, the nature of the boundary conditions will not make any difference. We are therefore free to choose the simplest boundary conditions that we can imagine, and for this purpose we will choose periodic boundary conditions. That is, we will assume that the fields and their normal derivatives on each wall are fixed to precisely match the fields and their normal derivatives on the opposite wall.

To begin, we consider a wave in one dimension, moving at the speed of light. Such waves are most easily described in terms of complex exponentials. If $A(x,t)$ represents the amplitude of the wave, then a sinusoidal wave moving in the positive x-direction can be written as

$$A(x,t) = \mathrm{Re}\left[Be^{ik(x-ct)}\right] \ ,$$

where B is a complex constant and k is a real constant. Defining $\omega = c|k|$, waves in either direction can be written as

$$A(x,t) = \mathrm{Re}\left[Be^{i(kx-\omega t)}\right] \ ,$$

where the sign of k determines the direction. To be periodic with period L, the parameter k must satisfy

$$kL = 2\pi n \ ,$$

where n is an integer. So the spacing between modes is $\Delta k = 2\pi/L$. The density of modes dN/dk (i.e., the number of modes per interval of k) is then one divided by the spacing, or $1/\Delta k$, so

$$\frac{dN}{dk} = \frac{L}{2\pi} \quad \text{(one dimension)} \ .$$

In three dimensions, a sinusoidal wave can be written as

$$A(\vec{x},t) = \mathrm{Re}\left[Be^{i(\vec{k}\cdot\vec{x}-\omega t)}\right] \ ,$$

where $\omega = c|\vec{k}|$, and

$$k_x L = 2\pi n_x \ , \quad k_y L = 2\pi n_y \ , \quad k_z L = 2\pi n_z \ ,$$

where n_x, n_y, and n_z are integers. Thus, in three-dimensional $\vec{k}$-space the allowed values of $\vec{k}$ lie on a cubical lattice, with spacing $2\pi/L$. In counting the modes, one should also remember that for photons there is an extra factor of 2 associated with the fact that electromagnetic waves have two possible polarizations for each allowed value of $\vec{k}$.

PROBLEM 6: PLOTTING THE SUPERNOVA DATA *(EXTRA CREDIT, 20 pts)*

The original data on the Hubble diagram based on Type Ia supernovae are found in two papers. One paper is authored by the High Z Supernova Search Team,* and the other is by the Supernova Cosmology Project.† More recent data from the High Z team, which includes many more data points, can be found in Riess *et al.*, http://arXiv.org/abs/astro-ph/0402512.¶ (By the way, the lead author Adam Riess was an MIT undergraduate physics major, graduating in 1992.)

You are asked to plot the data from either the 2nd or 3rd of these papers, and to include on the graph the theoretical predictions for several cosmological models.

The plot will be similar to the plots contained in these papers, and to the plot on p. 121 of Ryden's book, showing a graph of (corrected) magnitude m vs. redshift z. Your graph should include the error bars. If you plot the Perlmutter *et al.* data, you will be plotting "effective magnitude" m vs. redshift z. The magnitude is related to the flux J of the observed radiation by $m = -\frac{5}{2}\log_{10}(J) + \text{const}$. The value of the constant in this expression will not be needed. The word "corrected" refers both to corrections related to the spectral sensitivity of the detectors and to the brightness of the supernova explosions themselves. That is, the supernova at various distances are observed with different redshifts, and hence one must apply corrections if the detectors used to measure the radiation do not have the same sensitivity at all wavelengths. In addition, to improve the uniformity of the supernova as standard candles, the astronomers apply a correction based on the duration of the light output. Note that our ignorance of the absolute brightness of the supernova, of the precise value of the Hubble constant, and of the constant that appears in the definition of magnitude all combine to give an unknown but constant contribution to the predicted magnitudes. The consequence is that you will be able to move your predicted curves up or down (i.e., translate them by a fixed distance along the m axis). You should choose the vertical positioning of your curve to optimize your fit, either by eyeball or by some more systematic method.

If you choose to plot the data from the 3rd paper, Riess *et al.* 2004, then you should see the note at the end of this problem.

For your convenience, the magnitudes and redshifts for the Supernova Cosmology Project paper, from Tables 1 and 2, are summarized in a text file on the 8.286 web page. The data from Table 5 of the Riess *et al.* 2004 paper, mentioned above, is also posted on the 8.286 web page.

---

* http://arXiv.org/abs/astro-ph/9805201, later published as Riess *et al.*, *Astronomical Journal* **116**, 1009 (1998).

† http://arXiv.org/abs/astro-ph/9812133, later published as Perlmutter *et al.*, *Astrophysical Journal* **517**:565–586 (1999).

¶ Published as *Astrophysical Journal* **607**:665-687 (2004).

For the cosmological models to plot, you should include:

(i) A matter-dominated universe with $\Omega_m = 1$.

(ii) An open universe, with $\Omega_{m,0} = 0.3$.

(iii) A universe with $\Omega_{m,0} = 0.3$ and a cosmological constant, with $\Omega_{\text{vac},0} = 0.7$. (If you prefer to avoid the flat case, you can use $\Omega_{\text{vac},0} = 0.6$. Or, if you want to compare directly with Figure 4 of the Riess *et al.* (2004) paper, you should use $\Omega_{m,0} = 0.29$, $\Omega_{\text{vac},0} = 0.71$.)

You may include any other models if they interest you. You can draw the plot with either a linear or a logarithmic scale in z. I would recommend extending your theoretical plot to $z = 3$, if you do it logarithmically, or $z = 2$ if you do it linearly, even though the data does not go out that far. That way you can see what possible knowledge can be gained by data at higher redshift.

NOTE FOR THOSE PLOTTING DATA FROM RIESS ET AL. 2004:

Unlike the Perlmutter *et al.* data, the Riess *et al.* data is expressed in terms of the distance modulus, which is a direct measure of the luminosity distance. The distance modulus is defined both in the Riess *et al.* paper and in Ryden's book (p. 120) as

$$\mu = 5 \log_{10} \left(\frac{d_L}{1 \text{ Mpc}} \right) + 25 \ ,$$

where Ryden uses the notation $m - M$ for the distance modulus, and d_L is the luminosity distance. The luminosity distance, in turn, is really a measure of the observed brightness of the object. It is defined as the distance that the object would have to be located to result in the observed brightness, if we were living in a static Euclidean universe. More explicitly, if we lived in a static Euclidean universe and an object radiated power P in a spherically symmetric pattern, then the energy flux J at a distance d would be

$$J = \frac{P}{4\pi d^2} \ .$$

That is, the power would be distributed uniformly over the surface of a sphere at radius d. The luminosity distance is therefore defined as

$$d_L = \sqrt{\frac{P}{4\pi J}} \ .$$

Thus, a specified value of the distance modulus μ implies a definite value of the ratio J/P.

In plotting a theoretical curve, you will need to choose a value for H_0. Riess *et al.* do not specify what value they used, but I found that their curve is most closely reproduced if I choose $H_0 = 66$ km-sec^{-1}-Mpc^{-1}. This seems a little on the low side, since the value is usually estimated as 70–72 km-sec^{-1}-Mpc^{-1}, but Riess *et al.* emphasize that they were not concerned with this value. They were concerned with the relative values of the distance moduli, and hence the shape of the graph of the distance modulus vs. z. In their own words, from Appendix A, "The zeropoint, distance scale, absolute magnitude of the fiducial SN Ia or Hubble constant derived from Table 5 are all closely related (or even equivalent) quantities which were arbitrarily set for the sample presented here. Their correct value is not relevant for the analyses presented which only make use of differences between SN Ia magnitudes. Thus the analysis are independent of the aforementioned normalization parameters."

Total points for Problem Set 8: 110, plus an optional 20 points of extra credit.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — December 6, 2018
Prof. Alan Guth

PROBLEM SET 9 (The Last!)

DUE DATE: Wednesday, December 12, 2018, at 5:00 pm.

READING ASSIGNMENT: None.

CALENDAR THROUGH THE END OF THE TERM:

DECEMBER				
MON	TUES	WED	THURS	FRI
December 3 Lecture 22	4	5 **Quiz 3** — in class	6	7
December 10 Lecture 23	11	12 Last Class **PS 9 due**	13	14

PROBLEM 1: THE HORIZON PROBLEM *(20 points)*

The success of the big bang predictions for the abundances of the light elements suggests that the universe was already in thermal equilibrium at one second after the big bang. At this time, the region that later evolves to become the observed universe was, in the context of the conventional (non-inflationary) cosmological model, many horizon distances across. Try to estimate how many. You may assume that the universe is flat, that it was radiation-dominated for $t \lesssim$ 50,000 yr, and for this crude estimate you can also assume that it has been matter-dominated for all $t \gtrsim$ 50,000 yr, and that $a(t)T(t) \approx$ *const* for the whole period from 1 second to the present.

PROBLEM 2: THE FLATNESS PROBLEM *(20 points)*

Although we now know that $\Omega_0 = 1$ to an accuracy of about half a percent, let us pretend that the value of Ω today is 0.1. It nonetheless follows that at 10^{-37} second after the big bang (about the time of the grand unified theory phase transition), Ω must have been extraordinarily close to one. Writing $\Omega = 1 - \delta$, estimate the value of δ at $t = 10^{-37}$ sec (using the standard cosmological model). You may again use any of the approximations mentioned in Problem 1.

PROBLEM 3: THE MAGNETIC MONOPOLE PROBLEM *(20 points)*

In Lecture Notes 9, we learned that Grand Unified Theories (GUTs) imply the existence of magnetic monopoles, which form as "topological defects" (topologically stable knots) in the configuration of the Higgs fields that are responsible for breaking the grand unified symmetry to the SU(3)×SU(2)×U(1) symmetry of the standard model of particle physics. It was stated that if grand unified theories and the conventional (non-inflationary) cosmological model were both correct, then far too many magnetic monopoles would have been produced in the big bang. In this problem we will fill in the mathematical steps of that argument.

At very high temperatures the Higgs fields oscillate wildly, so the fields average to zero. As the temperature T falls, however, the system undergoes a phase transition. The phase transition occurs at a temperature T_c, called the critical temperature, where $kT_c \approx 10^{16}$ GeV. At this phase transition the Higgs fields acquire nonzero expectation values, and the grand unified symmetry is thereby spontaneously broken. The monopoles are twists in the Higgs field expectation values, so the monopoles form at the phase transition. Each monopole is expected to have a mass $M_M c^2 \approx 10^{18}$ GeV, where the subscript "M" stands for "monopole." According to an estimate first proposed by T.W.B. Kibble, the number density n_M of monopoles formed at the phase transition is of order

$$n_M \sim 1/\xi^3 \;, \tag{3.1}$$

where ξ is the correlation length of the field, defined roughly as the maximum distance over which the field at one point in space is correlated with the field at another point in space. The correlation length is certainly no larger than the physical horizon distance at the time of the phase transition, and it is believed to typically be comparable to this upper limit. Note that an upper limit on ξ is a lower limit on n_M — there must be at least of order one monopole per horizon-sized volume.

Assume that the particles of the grand unified theory form a thermal gas of blackbody radiation, as described by Eq. (6.48) of Lecture Notes 6,

$$u = g\frac{\pi^2}{30}\frac{(kT)^4}{(\hbar c)^3} \;, \tag{3.2}$$

with $g_{\rm GUT} \sim 200$. Further assume that the universe is flat and radiation-dominated from its beginning to the time of the GUT phase transition, $t_{\rm GUT}$.

For each of the following questions, first write the answer in terms of physical constants and the parameters T_c, M_M, and $g_{\rm GUT}$, and then evaluate the answers numerically.

(a) *(5 points)* Under the assumptions described above, at what time $t_{\rm GUT}$ does the phase transition occur? Express your answer first in terms of symbols, and then evaluate it in seconds.

(b) *(5 points)* Using Eq. (3.1) and setting ξ equal to the horizon distance, estimate the number density n_M of magnetic monopoles just after the phase transition.

(c) *(5 points)* Calculate the ratio n_M/n_γ of the number of monopoles to the number of photons immediately after the phase transition. Refer to Lecture Notes 6 to remind yourself about the number density of photons. You may assume that the temperature after the phase transition is still approximately T_c.

(d) *(5 points)* For topological reasons monopoles cannot disappear, but they form with an equal number of monopoles and antimonopoles, where the antimonopoles correspond to twists in the Higgs field in the opposite sense. Monopoles and antimonopoles can annihilate each other, but estimates of the rate for this process show that it is negligible. Thus, in the context of the conventional (non-inflationary) hot big bang model, the ratio of monopoles to photons would be about the same today as it was just after the phase transition. Use this assumption to estimate the contribution that these monopoles would make to the value of Ω today.

PROBLEM 4: EXPONENTIAL EXPANSION OF THE INFLATIONARY UNIVERSE *(15 points)*

Recall that the evolution of a Robertson-Walker universe is described by the equation

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2}. \tag{4.1}$$

Suppose that the mass density ρ is given by the constant mass density ρ_f of the false vacuum. For the case $k = 0$, the growing solution is given simply by

$$a(t) = \text{const } e^{\chi t}, \tag{4.2}$$

where

$$\chi = \sqrt{\frac{8\pi}{3}G\rho_f} \tag{4.3}$$

and *const* is an arbitrary constant. Find the growing solution to this equation for an arbitrary value of k. Be sure to consider both possibilities for the sign of k. You may find the following integrals useful:

$$\int \frac{dx}{\sqrt{1+x^2}} = \sinh^{-1} x\ . \tag{4.4a}$$

$$\int \frac{dx}{\sqrt{1-x^2}} = \sin^{-1} x\ . \tag{4.4b}$$

$$\int \frac{dx}{\sqrt{x^2-1}} = \cosh^{-1} x\ . \tag{4.4c}$$

Show that for large times one has

$$a(t) \propto e^{\chi t} \tag{4.5}$$

for all choices of k.

PROBLEM 5: THE HORIZON DISTANCE FOR THE PRESENT UNIVERSE *(25 points)*

We have not discussed horizon distances since the beginning of Lecture Notes 4, when we found that

$$\ell_{p,\text{horizon}}(t) = a(t) \int_0^t \frac{c}{a(t')} dt' \ . \tag{5.1}$$

This formula was derived before we discussed curved spacetimes, but the formula is valid for any Robertson-Walker universe, whether it is open, closed, or flat.

(a) Show that the formula above is valid for closed universes. Hint: write the closed universe metric as it was written in Eq. (7.27):

$$ds^2 = -c^2\, dt^2 + \tilde{a}^2(t) \left\{ d\psi^2 + \sin^2\psi \left(d\theta^2 + \sin^2\theta\, d\phi^2 \right) \right\} \ , \tag{5.2}$$

where

$$\tilde{a}(t) \equiv \frac{a(t)}{\sqrt{k}} \tag{5.3}$$

and ψ is related to the usual Robertson-Walker coordinate r by

$$\sin\psi \equiv \sqrt{k}\, r \ . \tag{5.4}$$

Use the fact that the physical speed of light is c, or equivalently the fact that $ds^2 = 0$ for any segment of the light ray's trajectory.

(b) The evaluation of the formula depends of course on the form of the function $a(t)$, which is governed by the Friedmann equations. For the Planck 2018 best fit to the parameters (see Table 7.1 of Lecture Notes 7, and Eq. (6.23) of Lecture Notes 6),

$$H_0 = 67.7 \text{ km} \cdot \text{s}^{-1} \cdot \text{Mpc}^{-1} \tag{5.5a}$$

$$\Omega_{m,0} = 0.311 \tag{5.5b}$$

$$\begin{aligned} \Omega_{r,0} &= 4.15 \times 10^{-5} h_0^{-2} \quad (T_{\gamma,0} = 2.725\,\text{K}) \\ &= 9.05 \times 10^{-5} \end{aligned} \tag{5.5c}$$

$$\Omega_{\text{vac},0} = 1 - \Omega_{m,0} - \Omega_{r,0} \ , \tag{5.5d}$$

find the current horizon distance, expressed both in light-years and in Mpc. Hint: find an integral expression for the horizon distance, similar to Eq. (7.23a) for the age of the universe. Then do the integral numerically.

Note that the model for which you are calculating does not explicitly include inflation. If it did, the horizon distance would turn out to be vastly larger. By ignoring the inflationary era in calculating the integral of Eq. (5.1), we are finding an effective horizon distance, defined as the present distance of the most distant objects that we can in principle observe by using only photons that have left their sources after the end of inflation. Photons that left their sources earlier than the end of inflation have undergone incredibly large redshifts, so it is reasonable to consider them to be completely unobservable in practice.

PROBLEM 6: A ZERO MASS DENSITY UNIVERSE— GENERAL RELATIVITY DESCRIPTION

(This problem is not required, but can be done for 20 points extra credit.)

In this problem and the next we will explore the connections between special relativity and the standard cosmological model which we have been discussing. Although we have not studied general relativity in detail, the description of the cosmological model that we have been using is precisely that of general relativity. In the limit of zero mass density the effects of gravity will become negligible, and the formulas must then be compatible with the special relativity which we discussed at the beginning of the course. The goal of these two problems is to see exactly how this happens.

These two problems will emphasize the notion that a coordinate system is nothing more than an arbitrary system of designating points in spacetime. A physical object might therefore look very different in two different coordinate systems, but the answer to any well-defined physical question must turn out the same regardless of which coordinate system is used in the calculation.

From the general relativity point of view, the model universe is described by the Robertson-Walker spacetime metric:

$$ds^2 = -c^2 dt^2 + a^2(t) \left\{ \frac{dr^2}{1 - kr^2} + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2 \right) \right\} . \tag{6.1}$$

This formula describes the analogue of the "invariant interval" of special relativity, measured between the spacetime points (t, r, θ, ϕ) and $(t + dt, r + dr, \theta + d\theta, \phi + d\phi)$.

The evolution of the model universe is governed by the general relation

$$\left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi}{3} G\rho - \frac{kc^2}{a^2} , \tag{6.2}$$

except in this case the mass density term is to be set equal to zero.

(a) *(5 points)* Since the mass density is zero, it is certainly less than the critical mass density, so the universe is open. We can then choose $k = -1$. Derive an explicit expression for the scale factor $a(t)$.

(b) *(5 points)* Suppose that a light pulse is emitted by a comoving source at time t_e, and is received by a comoving observer at time t_o. Find the Doppler shift ratio z.

(c) *(5 points)* Consider a light pulse that leaves the origin at time t_e. In an infinitesimal time interval dt the pulse will travel a physical distance $ds = cdt$. Since the pulse is traveling in the radial direction (i.e., with $d\theta = d\phi = 0$), one has

$$cdt = a(t) \frac{dr}{\sqrt{1 - kr^2}} . \tag{6.3}$$

Note that this is a slight generalization of Eq. (2.9), which applies for the case of a Euclidean geometry ($k = 0$). Derive a formula for the trajectory $r(t)$ of the light pulse. You may find the following integral useful:

$$\int \frac{dr}{\sqrt{1+r^2}} = \sinh^{-1} r \ . \tag{6.4}$$

(d) *(5 points)* Use these results to express the redshift z in terms of the coordinate r of the observer. If you have done it right, your answer will be independent of t_e. (In the special relativity description that will follow, it will be obvious why the redshift must be independent of t_e. Can you see the reason now?)

PROBLEM 7: A ZERO MASS DENSITY UNIVERSE— SPECIAL RELATIVITY DESCRIPTION

(This problem is also not required, but can be done for 20 points extra credit.)

In this problem we will describe the same model universe as in the previous problem, but we will use the standard formulation of special relativity. We will therefore use an inertial coordinate system, rather than the comoving system of the previous problem. Please note, however, that in the usual case in which gravity is significant, there is no inertial coordinate system. Only when gravity is absent does such a coordinate system exist.

To distinguish the two systems, we will use primes to denote the inertial coordinates: (t', x', y', z'). Since the problem is spherically symmetric, we will also introduce "polar inertial coordinates" (r', θ', ϕ') which are related to the Cartesian inertial coordinates by the usual relations:

$$\begin{aligned} x' &= r' \sin\theta' \cos\phi' \\ y' &= r' \sin\theta' \sin\phi' \\ z' &= r' \cos\theta' \ . \end{aligned} \tag{7.1}$$

In terms of these polar inertial coordinates, the invariant spacetime interval of special relativity can be written as

$$ds^2 = -c^2 dt'^2 + dr'^2 + r'^2 \left(d\theta'^2 + \sin^2\theta' \, d\phi'^2 \right) \ . \tag{7.2}$$

For purposes of discussion we will introduce a set of comoving observers which travel along with the matter in the universe, following the Hubble expansion pattern. (Although the matter has a negligible mass density, I will assume that enough of it exists to define a velocity at any point in space.) These trajectories must all meet at some spacetime point corresponding to the instant of the big bang, and we will take that spacetime point to be the origin of the coordinate system. Since there are no forces acting in this model universe, the comoving observers travel on lines of constant velocity (all emanating from the origin). The model universe is then confined to the future light-cone of the origin.

(a) *(5 points)* The cosmic time variable t used in the previous problem can be defined as the time measured on the clocks of the comoving observers, starting at the instant of the big bang. Using this definition and your knowledge of special relativity, find the value of the cosmic time t for given values of the inertial coordinates— i.e., find $t(t', r')$. [Hint: first find the velocity of a comoving observer who starts at the origin and reaches the spacetime point (t', r', θ', ϕ'). Note that the rotational symmetry makes θ' and ϕ' irrelevant, so one can examine motion along a single axis.]

(b) *(5 points)* Let us assume that angular coordinates have the same meaning in the two coordinate systems, so that $\theta = \theta'$ and $\phi = \phi'$. We will verify in part (d) below that this assumption is correct. Using this assumption, find the value of the comoving radial coordinate r in terms of the inertial coordinates— i.e., find $r(t', r')$. [Hint: consider an infinitesimal line segment which extends in the θ-direction, with constant values of t, r, and ϕ. Use the fact that this line segment must have the same physical length, regardless of which coordinate system is used to describe it.] Draw a graph of the t'-r' plane, and sketch in lines of constant t and lines of constant r.

(c) *(5 points)* Show that the radial coordinate r of the comoving system is related to the magnitude of the velocity in the inertial system by

$$r = \frac{v/c}{\sqrt{1 - v^2/c^2}} \ . \tag{7.3}$$

Suppose that a light pulse is emitted at the spatial origin ($r' = 0$, $t' = \textit{anything}$) and is received by another comoving observer who is traveling at speed v. With what redshift z is the pulse received? Express z as a function of r, and compare your answer to part (d) of the previous problem.

(d) *(5 points)* In this part we will show that the metric of the comoving coordinate system can be derived from the metric of special relativity, a fact which completely establishes the consistency of the two descriptions. To do this, first write out the equations of transformation in the form:

$$\begin{aligned} t' &= ? \\ r' &= ? \\ \theta' &= ? \\ \phi' &= ? \ , \end{aligned} \tag{7.4}$$

where the question marks denote expressions in t, r, θ, and ϕ. Now consider an infinitesimal spacetime line segment described in the comoving system by its two endpoints: (t, r, θ, ϕ) and $(t + dt, r + dr, \theta + d\theta, \phi + d\phi)$. Calculating to first order in the infinitesimal quantities, find the separation between the coordinates of the two endpoints in the inertial coordinate system— i.e., find dt', dr', $d\theta'$, and $d\phi'$. Now insert these expressions into the special relativity expression for the invariant interval ds^2 , and if you have made no mistakes you will recover the Robertson-Walker metric used in the previous problem.

DISCUSSION OF THE ZERO MASS DENSITY UNIVERSE:

The two problems above demonstrate how the general relativistic description of cosmology can reduce to special relativity when gravity is unimportant, but it provides a misleading picture of the big-bang singularity which I would like to clear up.

First, let me point out that the mass density of the universe increases as one looks backward in time. So, if we imagine a model universe with $\Omega = 0.01$ at a given time, it could be well-approximated by the zero mass density universe at this time. However, no matter how small Ω is at a given time, the mass density will increase as one follows the model to earlier times, and the behavior of the model near $t = 0$ will be very different from the zero mass density model.

In the zero mass density model, the big-bang "singularity" is a single spacetime point which is in fact not singular at all. In the comoving description the scale factor $a(t)$ equals zero at this time, but in the inertial system one sees that the spacetime metric is really just the usual smooth metric of special relativity, expressed in a peculiar set of coordinates. In this model it is unnatural to think of $t = 0$ as really defining the beginning of anything, since the the future light-cone of the origin connects smoothly to the rest of the spacetime.

In the standard model of the universe with a nonzero mass density, the behavior of the singularity is very different. First of all, it really is singular— one can mathematically prove that there is no coordinate system in which the singularity disappears. Thus, the spacetime cannot be joined smoothly onto anything that may have happened earlier.

The differences between the singularities in the two models can also be seen by looking at the horizon distance. We learned in Lecture Notes 4 that light can travel only a finite distance from the time of the big bang to some arbitrary time t, and that this "horizon distance" is given by

$$\ell_p(t) = a(t) \int_0^t \frac{c}{a(t')} dt' \ . \tag{7.5}$$

For the scale factor of the zero mass density universe as found in the problem, one can see that this distance is infinite for any t— for the zero mass density model there is ***no*** horizon. For a radiation-dominated model, however, there is a finite horizon distance given by $2ct$.

Finally, in the zero mass density model the big bang occurs at a single point in spacetime, but for a nonzero mass density model it seems better to think of the big bang as occurring everywhere at once. In terms of the Robertson-Walker coordinates, the singularity occurs at $t = 0$, for all values of r, θ, and ϕ. There is a subtle issue, however, because with $a(t = 0) = 0$, all of these points have zero distance from each other. Mathematically the locus $t = 0$ in a nonzero mass density model is too singular to even be considered part of the space, which consists of all values of $t > 0$. Thus, the question of whether the singularity is a single point is not well defined. For any $t > 0$ the issue is of course clear— the space is homogeneous and infinite (for the case of the open universe). If one wishes to ignore the mathematical subtleties and call the singularity at $t = 0$ a single point, then one certainly must remember that the singularity makes it a very unusual point. Objects emanating from this "point" can achieve an infinite separation in an arbitrarily short length of time.

Total points for Problem Set 9: 100, plus an optional 40 points of extra credit.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — September 28, 2018
Prof. Alan Guth

REVIEW PROBLEMS FOR QUIZ 1

QUIZ DATE: Wednesday, October 3, 2018, during the normal class time.

QUIZ COVERAGE: Lecture Notes 1, 2, and 3; Problem Sets 1, 2, and 3; Weinberg, Chapters 1, 2, and 3; Ryden, Chapters 1, 2, and 3. (While all of Ryden's Chapter 3 has been assigned, questions on the quiz will be limited to Section 3.1. The material in Sections 3.2 and 3.3 will be discussed in lecture later in the course, and you will not be responsible for it until then. Section 3.4 (for the $\kappa = 0$ case) may help you understand the cosmological Doppler shift, also discussed in Lecture Notes 2, but there will be no questions specifically focused on Ryden's discussion.) **One of the problems on the quiz will be taken verbatim (or at least *almost* verbatim) from either the homework assignments, or from the starred problems from this set of Review Problems.** The starred problems are the ones that I recommend that you review most carefully: Problems 2, 4, 7, 12, 15, 17, 19, and 22. The starred problems do not include any reading questions, but parts of the reading questions in these Review Problems may also recur on the upcoming quiz. For the homework problems, extra credit problems are eligible to be the problem used on the quiz.

PURPOSE: These review problems are not to be handed in, but are being made available to help you study. They come mainly from quizzes in previous years. Except for a few parts which are clearly marked, they are all problems that I would consider fair for the coming quiz. In some cases the number of points assigned to the problem on the quiz is listed — in all such cases it is based on 100 points for the full quiz.

In addition to this set of problems, you will find on the course web page the actual quizzes that were given in 1994, 1996, 1998, 2000, 2002, 2004, 2005, 2007, 2009, 2011, 2013, and 2016. The relevant problems from those quizzes have mostly been incorporated into these review problems, but you still may be interested in looking at the original quizzes, just to see how much material has been included in each quiz. Since the schedule and the number of quizzes has varied over the years, the coverage of this quiz will not necessarily be the same as Quiz 1 from all previous years. In fact, however, the first quiz this year covers essentially the same material as the first quiz in either 2009, 2011, 2013, or 2016.

REVIEW SESSION: To help you study for the quiz, there will be a review session led by Honggeun Kim on Sunday, September 30, at 3:30 pm in Room 3-333.

FUTURE QUIZZES: The other quiz dates this term will be Monday, November 5, and Wednesday, December 5, 2018.

INFORMATION TO BE GIVEN ON QUIZ:

Each quiz in this course will have a section of "useful information" at the back of the quiz. For the first quiz, this useful information will be the following:

DOPPLER SHIFT (For motion along a line):

$$z = v/u \quad \text{(nonrelativistic, source moving)}$$

$$z = \frac{v/u}{1 - v/u} \quad \text{(nonrelativistic, observer moving)}$$

$$z = \sqrt{\frac{1+\beta}{1-\beta}} - 1 \quad \text{(special relativity, with } \beta = v/c\text{)}$$

COSMOLOGICAL REDSHIFT:

$$1 + z \equiv \frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}} = \frac{a(t_{\text{observed}})}{a(t_{\text{emitted}})}$$

SPECIAL RELATIVITY:

Time Dilation Factor:

$$\gamma \equiv \frac{1}{\sqrt{1-\beta^2}}\ , \qquad \beta \equiv v/c$$

Lorentz-Fitzgerald Contraction Factor: γ

Relativity of Simultaneity:
Trailing clock reads later by an amount $\beta \ell_0 / c$.

KINEMATICS OF A HOMOGENEOUSLY EXPANDING UNIVERSE:

Hubble's Law: $v = Hr$,
where v = recession velocity of a distant object, H = Hubble expansion rate, and r = distance to the distant object.

Present Value of Hubble Expansion Rate (Planck 2018):

$$H_0 = 67.66 \pm 0.42 \text{ km-s}^{-1}\text{-Mpc}^{-1}$$

Scale Factor: $\ell_p(t) = a(t)\ell_c$,
where $\ell_p(t)$ is the physical distance between any two objects, $a(t)$ is the scale factor, and ℓ_c is the coordinate distance between the objects, also called the comoving distance.

Hubble Expansion Rate: $H(t) = \dfrac{1}{a(t)}\dfrac{\mathrm{d}a(t)}{\mathrm{d}t}$.

Light Rays in Comoving Coordinates: Light rays travel in straight lines with speed $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \dfrac{c}{a(t)}$.

EVOLUTION OF A MATTER-DOMINATED UNIVERSE:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2} \ , \quad \ddot{a} = -\frac{4\pi}{3}G\rho a \ ,$$

$$\rho(t) = \frac{a^3(t_i)}{a^3(t)}\,\rho(t_i)$$

$$\Omega \equiv \rho/\rho_c \ , \quad \text{where} \quad \rho_c = \frac{3H^2}{8\pi G} \ .$$

Flat ($k = 0$): $a(t) \propto t^{2/3}$, $\Omega = 1$

PROBLEM LIST

PROBLEM 1: DID YOU DO THE READING (2000)? *(35 points)*

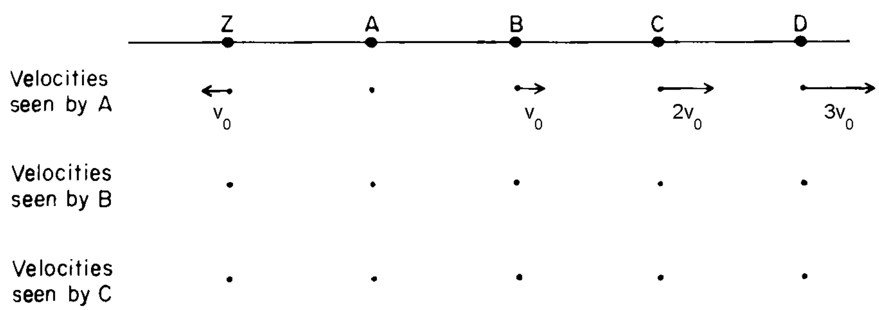
The following problem was Problem 1, Quiz 1, 2000. The parts were each worth 5 points.

a) The Doppler effect for both sound and light waves is named for Johann Christian Doppler, a professor of mathematics at the Realschule in Prague. He predicted the effect for both types of waves in xx42. What are the two digits xx?

b) When the sky is very clear (as it almost never is in Boston), one can see a band of light across the night sky that has been known since ancient times as the Milky Way. Explain in a sentence or two how this band of light is related to the shape of the galaxy in which we live, which is also called the Milky Way.

c) The statement that the distant galaxies are on average receding from us with a speed proportional to their distance was first published by Edwin Hubble in 1929, and has become known as Hubble's law. Was Hubble's original paper based on the study of 2, 18, 180, or 1,800 galaxies?

d) The following diagram, labeled *Homogeneity and the Hubble Law,* was used by Weinberg to explain how Hubble's law is consistent with the homogeneity of the universe:



The arrows and labels from the "Velocities seen by B" and the "Velocities seen by C" rows have been deleted from this copy of the figure, and it is your job to sketch the figure in your exam book with these arrows and labels included. (Actually, in Weinberg's diagram these arrows were not labeled, but the labels are required here so that the grader does not have to judge the precise length of hand-drawn arrows.)

e) The horizon is the present distance of the most distant objects from which light has had time to reach us since the beginning of the universe. The horizon changes with time, but of course so does the size of the universe as a whole. During a time interval in which the linear size of the universe grows by 1%, does the horizon distance

(i) grow by more than 1%, or

(ii) grow by less than 1%, or

(iii) grow by the same 1%?

f) Name the two men who in 1964 discovered the cosmic background radiation. With what institution were they affiliated?

g) At a temperature of 3000 K, the nuclei and electrons that filled the universe combined to form neutral atoms, which interact very weakly with the photons of the background radiation. After this process, known as "recombination," the background radiation expanded freely. Since recombination, how have each of the following quantities varied as the size of the universe has changed? (Your answers should resemble statements such as "proportional to the size of the universe," or "inversely proportional to the square of the size of the universe". The word "size" will be interpreted to mean linear size, not volume.)

(i) the average distance between photons

(ii) the typical wavelength of the radiation

(iii) the number density of photons in the radiation

(iv) the energy density of the radiation

(v) the temperature of the radiation

∗ PROBLEM 2: THE STEADY-STATE UNIVERSE THEORY *(25 points)*

The following problem was Problem 2, Quiz 1, 2000.

The steady-state theory of the universe was proposed in the late 1940s by Hermann Bondi, Thomas Gold, and Fred Hoyle, and was considered a viable model for the universe until the cosmic background radiation was discovered and its properties were confirmed. As the name suggests, this theory is based on the hypothesis that the large-scale properties of the universe do not change with time. The expansion of the universe was an established fact when the steady-state theory was invented, but the steady-state theory reconciles the expansion with a steady-state density of matter by proposing that new matter is created as the universe expands, so that the matter density does not fall. Like the conventional theory, the steady-state theory describes a homogeneous, isotropic, expanding universe, so the same comoving coordinate formulation can be used.

a) *(10 points)* The steady-state theory proposes that the Hubble constant, like other cosmological parameters, does not change with time, so $H(t) = H_0$. Find the most general form for the scale factor function $a(t)$ which is consistent with this hypothesis.

b) *(15 points)* Suppose that the mass density of the universe is ρ_0, which of course does not change with time. In terms of the general form for $a(t)$ that you found in part (a), calculate the rate at which new matter must be created for ρ_0 to remain constant as the universe expands. Your answer should have the units of mass per unit volume per unit time. *[If you failed to answer part (a), you will still receive full credit here if you correctly answer the question for an arbitrary scale factor function $a(t)$.]*

PROBLEM 3: DID YOU DO THE READING (2007)? *(25 points)*

The following problem was Problem 1 on Quiz 1, 2007, where each of the 5 questions was worth 5 points:

(a) In the 1940's, three astrophysicists proposed a "steady state" theory of cosmology, in which the universe has always looked about the same as it does now. State the last name of at least one of these authors. (*Bonus points*: you can earn 1 point each for naming the other two authors, and hence up to 2 additional points, but 1 point will be taken off for each incorrect answer.)

(b) In 1917, a Dutch astronomer named Willem de Sitter did which one of the following accomplishments:

(i) measured the size of the Milky Way galaxy, finding it to be about one billion light-years in diameter.

(ii) resolved Cepheid variable stars in Andromeda and thereby obtained persuasive evidence that Andromeda is not within our own galaxy, but is apparently another galaxy like our own.

(iii) published a catalog, *Nebulae and Star Clusters*, listing 103 objects that astronomers should avoid when looking for comets.

(iv) published a model for the universe, based on general relativity, which appeared to be static but which produced a redshift proportional to the distance.

(v) discovered that the orbital periods of the planets are proportional to the 3/2 power of the semi-major axis of their elliptical orbits.

(c) In 1964–65, Arno A. Penzias and Robert W. Wilson observed a flux of microwave radiation coming from all directions in the sky, which was interpreted by a group of physicists at a neighboring institution as the cosmic background radiation left over from the big bang. Circle the two items on the following list that were **not** part of the story behind this spectacular discovery:

(i) Bell Telephone Laboratory	(ii) MIT	(iii) Princeton University
(iv) pigeons	(v) ground hogs	(vi) Hubble's constant
(vii) liquid helium	(viii) 7.35 cm	

(Grading: 3 pts for 1 correct answer, 5 for 2 correct answers, and -2 for each incorrect answer, but the minimum score is zero.)

(d) Important predictions of the Copernican theory were confirmed by the discovery of the aberration of starlight (which showed that the velocity of the Earth has the time-dependence expected for rotation about the Sun) and by the behavior of the Foucault pendulum (which showed that the Earth rotates). These discoveries were made

(i) during Copernicus' lifetime.

(ii) approximately two and three decades after Copernicus' death, respectively.

(iii) about one hundred years after Copernicus' death.

(iv) approximately two and three centuries after Copernicus' death, respectively.

(e) If one averages over sufficiently large scales, the universe appears to be homogeneous and isotropic. How large must the averaging scale be before this homogeneity and isotropy set in?

(i) 1 AU (1 AU $= 1.496 \times 10^{11}$ m).

(ii) 100 kpc (1 kpc = 1000 pc, 1 pc $= 3.086 \times 10^{16}$ m = 3.262 light-year).

(iii) 1 Mpc (1 Mpc $= 10^6$ pc).

(iv) 10 Mpc.

(v) 100 Mpc.

(vi) 1000 Mpc.

* PROBLEM 4: AN EXPONENTIALLY EXPANDING UNIVERSE *(20 points)*

The following problem was Problem 2, Quiz 2, 1994, and had also appeared on the 1994 Review Problems. As is the case this year, it was announced that one of the problems on the quiz would come from either the homework or the Review Problems. The problem also appeared as Problem 2 on Quiz 1, 2007.

Consider a flat (i.e., a $k = 0$, or a Euclidean) universe with scale factor given by

$$a(t) = a_0 e^{\chi t} ,$$

where a_0 and χ are constants.

(a) *(5 points)* Find the Hubble constant H at an arbitrary time t.

(b) *(5 points)* Let (x, y, z, t) be the coordinates of a comoving coordinate system. Suppose that at $t = 0$ a galaxy located at the origin of this system emits a light pulse along the positive x-axis. Find the trajectory $x(t)$ which the light pulse follows.

(c) *(5 points)* Suppose that we are living on a galaxy along the positive x-axis, and that we receive this light pulse at some later time. We analyze the spectrum of the pulse and determine the redshift z. Express the time t_r at which we receive the pulse in terms of z, χ, and any relevant physical constants.

(d) *(5 points)* At the time of reception, what is the physical distance between our galaxy and the galaxy which emitted the pulse? Express your answer in terms of z, χ, and any relevant physical constants.

PROBLEM 5: DID YOU DO THE READING (1986/1990 COMPOSITE)?

(a) The assumptions of homogeneity and isotropy greatly simplify the description of our universe. We find that there are three possibilities for a homogeneous and isotropic universe: an open universe, a flat universe, and a closed universe. What quantity or condition distinguishes between these three cases: the temperature of the microwave background, the value of $\Omega = \rho/\rho_c$, matter vs. radiation domination, or redshift?

(b) What is the temperature, in Kelvin, of the cosmic microwave background today?

(c) Which of the following supports the hypothesis that the universe is isotropic: the distances to nearby clusters, observations of the cosmic microwave background, clustering of galaxies on large scales, or the age and distribution of globular clusters?

(d) Is the distance to the Andromeda Nebula (roughly) 10 kpc, 5 billion light years, 2 million light years, or 3 light years?

(e) Did Hubble discover the law which bears his name in 1862, 1880, 1906, 1929, or 1948?

(f) When Hubble measured the value of his constant, he found $H^{-1} \approx 100$ million years, 2 billion years, 10 billion years, or 20 billion years?

(g) Cepheid variables are important to cosmology because they can be used to estimate the distances to the nearby galaxies. What property of Cepheid variables makes them useful for this purpose, and how are they used?

(h) Cepheid variable stars can be used as estimators of distance for distances up to about 100 light-years, 10^4 light-years, 10^7 light years, or 10^{10} light-years? *[Note for 2011: this question was based on the reading from Joseph Silk's* **The Big Bang***, and therefore would be not be a fair question for this year.]*

(i) Name the two men who in 1964 discovered the cosmic background radiation. With what institution were they affiliated?

(j) At the time of the discovery of the cosmic microwave background, an active but independent effort was taking place elsewhere. P.J.E. Peebles had estimated that the universe must contain background radiation with a temperature of at least 10°K, and Robert H. Dicke, P.G. Roll, and D.T. Wilkinson had mounted an experiment to look for it. At what institution were these people working?

PROBLEM 6: A FLAT UNIVERSE WITH UNUSUAL TIME EVOLUTION

The following problem was Problem 3, Quiz 2, 1988:

Consider a flat universe filled with a new and peculiar form of matter, with a Robertson–Walker scale factor that behaves as

$$a(t) = bt^{1/3} \ .$$

Here b denotes a constant.

(a) If a light pulse is emitted at time t_e and observed at time t_o, find the physical separation $\ell_p(t_o)$ between the emitter and the observer, at the time of observation.

(b) Again assuming that t_e and t_o are given, find the observed redshift z.

(c) Find the physical distance $\ell_p(t_o)$ which separates the emitter and observer at the time of observation, expressed in terms of c, t_o, and z (i.e., without t_e appearing).

(d) At an arbitrary time t in the interval $t_e < t < t_o$, find the physical distance $\ell_p(t)$ between the light pulse and the observer. Express your answer in terms of c, t, and t_o.

∗ PROBLEM 7: ANOTHER FLAT UNIVERSE WITH AN UNUSUAL TIME EVOLUTION *(40 points)*

The following problem was Problem 3, Quiz 1, 2000.

Consider a ***flat*** universe which is filled with some peculiar form of matter, so that the Robertson–Walker scale factor behaves as

$$a(t) = bt^{\gamma} ,$$

where b and γ are constants. *[This universe differs from the matter-dominated universe described in the lecture notes in that ρ is not proportional to $1/a^3(t)$. Such behavior is possible when pressures are large, because a gas expanding under pressure can lose energy (and hence mass) during the expansion.]* For the following questions, any of the answers may depend on γ, whether it is mentioned explicitly or not.

a) *(5 points)* Let t_0 denote the present time, and let t_e denote the time at which the light that we are currently receiving was emitted by a distant object. In terms of these quantities, find the value of the redshift parameter z with which the light is received.

b) *(5 points)* Find the "look-back" time as a function of z and t_0. The look-back time is defined as the length of the interval in cosmic time between the emission and observation of the light.

c) *(10 points)* Express the present value of the physical distance to the object as a function of H_0, z, and γ.

d) *(10 points)* At the time of emission, the distant object had a power output P (measured, say, in ergs/sec) which was radiated uniformly in all directions, in the form of photons. What is the radiation energy flux J from this object at the earth today? Express your answer in terms of P, H_0, z, and γ. *[Energy flux (which might be measured in erg-cm^{-2}-sec^{-1}) is defined as the energy per unit area per unit time striking a surface that is orthogonal to the direction of energy flow.]*

e) *(10 points)* Suppose that the distant object is a galaxy, moving with the Hubble expansion. Within the galaxy a supernova explosion has hurled a jet of material directly towards Earth with a speed v, measured relative to the galaxy, which is comparable to the speed of light c. Assume, however, that the distance the jet has traveled from the galaxy is so small that it can be neglected. With what redshift z_J would we observe the light coming from this jet? Express your answer in terms of all or some of the variables v, z (the redshift of the galaxy), t_0, H_0, and γ, and the constant c.

PROBLEM 8: DID YOU DO THE READING (1996)? *(25 points)*

The following problem was Problem 1, Quiz 1, 1996:

The following questions are worth 5 points each.

a) In 1814-1815, the Munich optician Joseph Frauenhofer allowed light from the sun to pass through a slit and then through a glass prism. The light was spread into a spectrum of colors, showing lines that could be identified with known elements — sodium, iron, magnesium, calcium, and chromium. Were these lines dark, or bright *(2 points)*? Why *(3 points)*?

b) The Andromeda Nebula was shown conclusively to lie outside our own galaxy when astronomers acquired telescopes powerful enough to resolve the individual stars of Andromeda. Was this feat accomplished by Galileo in 1609, by Immanuel Kant in 1755, by Henrietta Swan Leavitt in 1912, by Edwin Hubble in 1923, or by Walter Baade and Allan Sandage in the 1950s?

c) Some of the earliest measurements of the cosmic background radiation were made indirectly, by observing interstellar clouds of a molecule called cyanogen (CN). State whether each of the following statements is true or false *(1 point each)*:

(i) The first measurements of the temperature of the interstellar cyanogen were made over twenty years before the cosmic background radiation was directly observed.

(ii) Cyanogen helps to measure the cosmic background radiation by reflecting it toward the earth, so that it can be received with microwave detectors.

(iii) One reason why the cyanogen observations were important was that they gave the first measurements of the equivalent temperature of the cosmic background radiation at wavelengths shorter than the peak of the black-body spectrum.

(iv) By measuring the spectrum of visible starlight that passes through the cyanogen clouds, astronomers can infer the intensity of the microwave radiation that bathes the clouds.

(v) By observing chemical reactions in the cyanogen clouds, astronomers can infer the temperature of the microwave radiation in which they are bathed.

d) In about 280 B.C., a Greek philosopher proposed that the Earth and the other planets revolve around the sun. What was the name of this person? *[Note for 2011: this question was based on readings from Joseph Silk's* **The Big Bang***, and therefore is not appropriate for Quiz 1 of this year.]*

e) In 1832 Heinrich Wilhelm Olbers presented what we now know as "Olbers' Paradox," although a similar argument had been discussed as early as 1610 by Johannes Kepler. Olbers argued that if the universe were transparent, static, infinitely old, and was populated by a uniform density of stars similar to our sun, then one of the following consequences would result:

(i) The brightness of the night sky would be infinite.

(ii) Any patch of the night sky would look as bright as the surface of the sun.

(iii) The total energy flux from the night sky would be about equal to the total energy flux from the sun.

(iv) Any patch of the night sky would look as bright as the surface of the moon.

Which one of these statements is the correct statement of Olbers' paradox?

PROBLEM 9: A FLAT UNIVERSE WITH $a(t) \propto t^{3/5}$

The following problem was Problem 3, Quiz 1, 1996:

Consider a ***flat*** universe which is filled with some peculiar form of matter, so that the Robertson–Walker scale factor behaves as

$$a(t) = bt^{3/5} ,$$

where b is a constant.

a) *(5 points)* Find the Hubble constant H at an arbitrary time t.

b) *(5 points)* What is the physical horizon distance at time t?

c) *(5 points)* Suppose a light pulse leaves galaxy A at time t_A and arrives at galaxy B at time t_B. What is the coordinate distance between these two galaxies?

d) *(5 points)* What is the physical separation between galaxy A and galaxy B at time t_A? At time t_B?

e) *(5 points)* At what time is the light pulse equidistant from the two galaxies?

f) *(5 points)* What is the speed of B relative to A at the time t_A? (By "speed," I mean the rate of change of the physical distance with respect to cosmic time, $d\ell_p/dt$.)

g) *(5 points)* For observations made at time t, what is the present value of the physical distance as a function of the redshift z (and the time t)? What physical distance corresponds to $z = \infty$? How does this compare with the horizon distance? (Note that this question does not refer to the galaxies A and B discussed in the earlier parts. In particular, you should not assume that the light pulse left its source at time t_A.)

h) *(5 points)* Returning to the discussion of the galaxies A and B which were considered in parts (c)-(f), suppose the radiation from galaxy A is emitted with total power P. What is the power per area received at galaxy B?

i) *(5 points)* When the light pulse is received by galaxy B, a pulse is immediately sent back toward galaxy A. At what time does this second pulse arrive at galaxy A?

PROBLEM 10: DID YOU DO THE READING (1998)? *(20 points)*

The following questions were taken from Problem 1, Quiz 1, 1998:

The following questions are worth 5 points each.

a) In 1917, Einstein introduced a model of the universe which was based on his newly developed general relativity, but which contained an extra term in the equations which he called the "cosmological term." (The coefficient of this term is called the "cosmological constant.") What was Einstein's motivation for introducing this term?

b) When the redshift of distant galaxies was first discovered, the earliest observations were analyzed according to a cosmological model invented by the Dutch astronomer W. de Sitter in 1917. At the time of its discovery, was this model thought to be static or expanding? From the modern perspective, is the model thought to be static or expanding?

c) The early universe is believed to have been filled with thermal, or black-body, radiation. For such radiation the number density of photons and the energy density are each proportional to powers of the absolute temperature T. Say

$$\text{Number density} \propto T^{n_1}$$
$$\text{Energy density} \propto T^{n_2}$$

Give the values of the exponents n_1 and n_2.

d) At about 3,000 K the matter in the universe underwent a certain chemical change in its form, a change that was necessary to allow the differentiation of matter into galaxies and stars. What was the nature of this change?

PROBLEM 11: ANOTHER FLAT UNIVERSE WITH $a(t) \propto t^{3/5}$ *(40 points)*

The following was Problem 3, Quiz 1, 1998:

Consider a ***flat*** universe which is filled with some peculiar form of matter, so that the Robertson–Walker scale factor behaves as

$$a(t) = bt^{3/5} \ ,$$

where b is a constant.

a) *(5 points)* Find the Hubble constant H at an arbitrary time t.

b) *(10 points)* Suppose a message is transmitted by radio signal (traveling at the speed of light c) from galaxy A to galaxy B. The message is sent at cosmic time t_1, when the physical distance between the galaxies is ℓ_0. At what cosmic time t_2 is the message received at galaxy B? (Express your answer in terms of ℓ_0, t_1, and c.)

c) *(5 points)* Upon receipt of the message, the creatures on galaxy B immediately send back an acknowledgement, by radio signal, that the message has been received. At what cosmic time t_3 is the acknowledgment received on galaxy A? (Express your answer in terms of ℓ_0, t_1, t_2, and c.)

d) *(10 points)* The creatures on galaxy B spend some time trying to decode the message, finally deciding that it is an advertisement for Kellogg's Corn Flakes (whatever that is). At a time Δt after the receipt of the message, as measured on their clocks, they send back a response, requesting further explanation. At what cosmic time t_4 is the response received on galaxy A? In answering this part, you should not assume that Δt is necessarily small. (Express your answer in terms of ℓ_0, t_1, t_2, t_3, Δt, and c.)

e) *(5 points)* When the response is received by galaxy A, the radio waves will be red-shifted by a factor $1 + z$. Give an expression for z. (Express your answer in terms of ℓ_0, t_1, t_2, t_3, t_4, Δt, and c.)

f) *(5 points; No partial credit)* If the time Δt introduced in part (d) is small, the time difference $t_4 - t_3$ can be expanded to first order in Δt. Calculate $t_4 - t_3$ to first order accuracy in Δt. (Express your answer in terms of ℓ_0, t_1, t_2, t_3, t_4, Δt, and c.) [Hint: while this part can be answered by using brute force to expand the answer in part (d), there is an easier way.]

∗ PROBLEM 12: THE DECELERATION PARAMETER

The following problem was Problem 2, Quiz 2, 1992, where it counted 10 points out of 100.

Many standard references in cosmology define a quantity called the ***deceleration parameter*** q, which is a direct measure of the slowing down of the cosmic expansion. The parameter is defined by

$$q \equiv -\ddot{a}(t) \frac{a(t)}{\dot{a}^2(t)} \ .$$

Find the relationship between q and Ω for a matter-dominated universe. *[In case you have forgotten, Ω is defined by*

$$\Omega = \rho/\rho_c \ ,$$

where ρ is the mass density and ρ_c is the critical mass density (i.e., that mass density which corresponds to $k = 0$).]

PROBLEM 13: A RADIATION-DOMINATED FLAT UNIVERSE

We have learned that a matter-dominated homogeneous and isotropic universe can be described by a scale factor $a(t)$ obeying the equation

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

This equation in fact applies to any form of mass density, so we can apply it to a universe in which the mass density is dominated by the energy of photons. Recall that the mass density of nonrelativistic matter falls off as $1/a^3(t)$ as the universe expands; the mass of each particle remains constant, and the density of particles falls off as $1/a^3(t)$ because the volume increases as $a^3(t)$. For the photon-dominated universe, the density of photons falls of as $1/a^3(t)$, but in addition the frequency (and hence the energy) of each photon redshifts in proportion to $1/a(t)$. Since mass and energy are equivalent, the mass density of the gas of photons falls off as $1/a^4(t)$.

For a flat (i.e., $k = 0$) matter-dominated universe we learned that the scale factor $a(t)$ is proportional to $t^{2/3}$. How does $a(t)$ behave for a photon-dominated universe?

PROBLEM 14: DID YOU DO THE READING?

The following problem was taken from Problem 1, Quiz 1, 2004, where each part counted 5 points, for a total of 25 points. The reading assignment included the first three chapters of Ryden, **Introduction to Cosmology***, and the first three chapters of Weinberg,* **The First Three Minutes***.*

(a) In 1826, the astronomer Heinrich Olber wrote a paper on a paradox regarding the night sky. What is Olber's paradox? What is the primary resolution of it?

(b) What is the value of the Newtonian gravitational constant G in Planck units? The Planck length is of the order of 10^{-35} m, 10^{-15} m, 10^{15} m, or 10^{35} m?

(c) What is the Cosmological Principle? Is the Hubble expansion of the universe consistent with it? (For the latter question, a simple "yes" or "no" will suffice.)

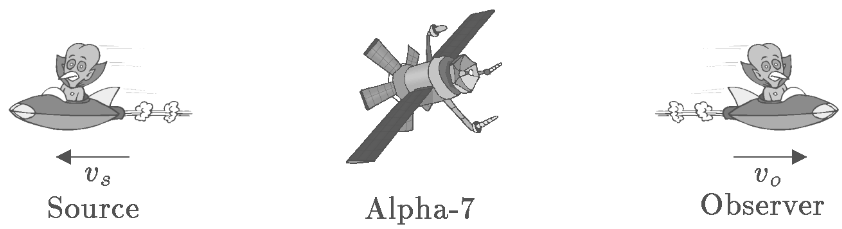
(d) In the "Standard Model" of the universe, when the universe cooled to about 3×10^a K, it became transparent to photons, and today we observe these as the Cosmic Microwave Background (CMB) at a temperature of about 3×10^b K. What are the integers a and b?

(e) What did the universe primarily consist of at about 1/100th of a second after the Big Bang? Include any constituent that is believed to have made up more than 1% of the mass density of the universe.

∗ PROBLEM 15: SPECIAL RELATIVITY DOPPLER SHIFT

The following problem was taken from Problem 2, Quiz 1, 2004, where it counted 20 points.

Consider the Doppler shift of radio waves, for a case in which both the source and the observer are moving. Suppose the source is a spaceship moving with a speed v_s relative to the space station Alpha-7, while the observer is on another spaceship, moving in the opposite direction from Alpha-7 with speed v_o relative to Alpha-7.



(a) *(10 points)* Calculate the Doppler shift z of the radio wave as received by the observer. (Recall that radio waves are electromagnetic waves, just like light except that the wavelength is longer.)

(b) *(10 points)* Use the results of part (a) to determine v_{tot}, the velocity of the source spaceship as it would be measured by the observer spaceship. *(8 points will be given for the basic idea, whether or not you have the right answer for part (a), and 2 points will be given for the algebra.)*

PROBLEM 16: DID YOU DO THE READING?

The following question was taken from Problem 1, Quiz 1, 2005, where it counted 25 points.

(a) *(4 points)* What was the first external galaxy that was shown to be at a distance significantly greater than the most distant known objects in our galaxy? How was the distance estimated?

(b) *(5 points)* What is recombination? Did galaxies begin to form before or after recombination? Why?

(c) *(4 points)* In Chapter IV of his book, Weinberg develops a "recipe for a hot universe," in which the matter of the universe is described as a gas in thermal equilbrium at a very high temperature, in the vicinity of 10^9 K (several thousand million degrees Kelvin). Such a thermal equilibrium gas is completely described by specifying its temperature and the density of the conserved quantities. Which of the following is on this list of conserved quantities? Circle as many as apply.

(i) baryon number (ii) energy per particle (iii) proton number

(iv) electric charge (v) pressure

(d) *(4 points)* The wavelength corresponding to the mean energy of a CMB (cosmic microwave background) photon today is approximately equal to which of the following quantities? (You may wish to look up the values of various physical constants at the end of the quiz.)

(i) 2 fm (2×10^{-15} m)

(ii) 2 microns (2×10^{-6} m)

(iii) 2 mm (2×10^{-3} m)

(iv) 2 m.

(e) *(4 points)* What is the equivalence principle?

(f) *(4 points)* Why is it difficult for Earth-based experiments to look at the small wavelength portion of the graph of CMB energy density per wavelength vs. wavelength?

∗ PROBLEM 17: TRACING A LIGHT PULSE THROUGH A RADIATION-DOMINATED UNIVERSE

The following problem was taken from Problem 3, Quiz 1, 2005, where it counted 25 points.

Consider a flat universe that expands with a scale factor

$$a(t) = bt^{1/2} ,$$

where b is a constant. We will learn later that this is the behavior of the scale factor for a radiation-dominated universe.

(a) *(5 points)* At an arbitrary time $t = t_f$, what is the physical horizon distance? (By “physical,” I mean as usual the distance in physical units, such as meters or centimeters, as measured by a sequence of rulers, each of which is at rest relative to the comoving matter in its vicinity.)

(b) *(3 points)* Suppose that a photon arrives at the origin, at time t_f, from a distant piece of matter that is precisely at the horizon distance at time t_f. What is the time t_e at which the photon was emitted?

(c) *(2 points)* What is the **coordinate** distance from the origin to the point from which the photon was emitted?

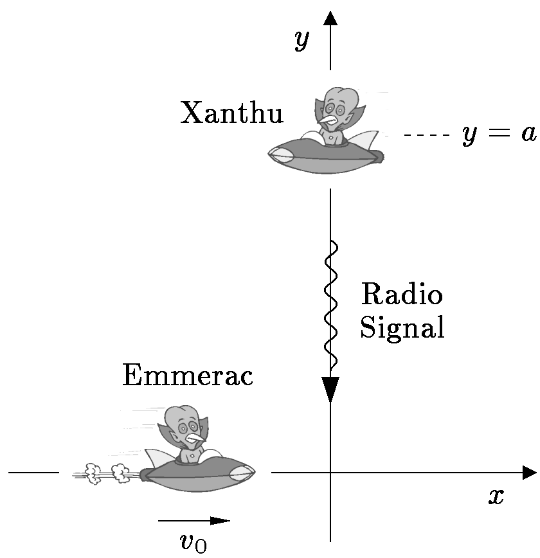
(d) *(10 points)* For an arbitrary time t in the interval $t_e \leq t \leq t_f$, while the photon is traveling, what is the **physical** distance $\ell_p(t)$ from the origin to the location of the photon?

(e) *(5 points)* At what time $t_{\max}$ is the physical distance of the photon from the origin at its largest value?

PROBLEM 18: TRANSVERSE DOPPLER SHIFTS

The following problem was taken from Problem 4, Quiz 1, 2005, where it counted 20 points.

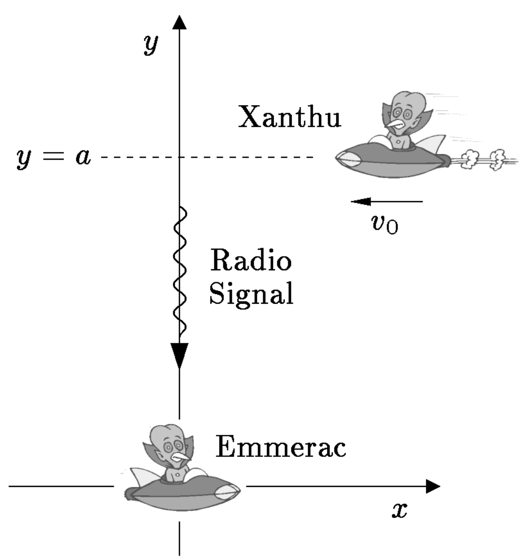
(a) *(8 points)* Suppose the spaceship Xanthu is at rest at location $(x{=}0, y{=}a, z{=}0)$ in a Cartesian coordinate system. (We assume that the space is Euclidean, and that the distance scales in the problem are small enough so that the expansion of the universe can be neglected.) The spaceship Emmerac is moving at speed v_0 along the x-axis in the positive direction, as shown in the diagram, where v_0 is comparable to the speed of light. As the Emmerac crosses the origin, it receives a radio signal that had been sent some time earlier from the Xanthu. Is the radiation received redshifted or blueshifted? What is the redshift z (where negative values of z can be used to describe blueshifts)?



(b) *(7 points)* Now suppose that the Emmerac is at rest at the origin, while the Xanthu is moving in the negative x-direction, at $y = a$ and $z = 0$, as shown in the diagram. That is, the trajectory of the Xanthu can be taken as

$$(x{=}-v_0 t, y{=}a, z{=}0) .$$

At $t = 0$ the Xanthu crosses the y-axis, and at that instant it emits a radio signal along the y-axis, directed at the origin. The radiation is received some time later by the Emmerac. In this case, is the radiation received redshifted or blueshifted? What is the redshift z (where again negative values of z can be used to describe blueshifts)?

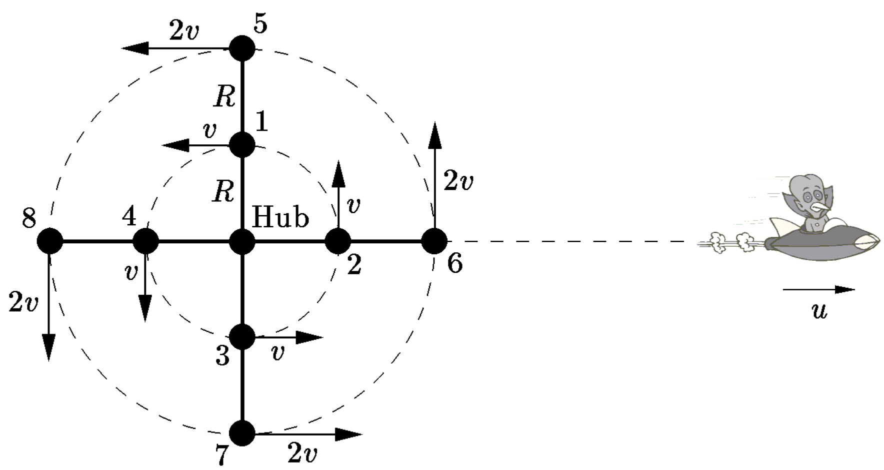


(c) *(5 points)* Is the sequence of events described in (b) physically distinct from the sequence described in (a), or is it really the same sequence of events described in a reference frame that is moving relative to the reference frame used in part (a)? Explain your reasoning in a sentence or two. *(Hint: note that there are three objects in the problem: Xanthu, Emmerac, and the photons of the radio signal.)*

∗ PROBLEM 19: A TWO-LEVEL HIGH-SPEED MERRY-GO-ROUND *(15 points)*

This problem was Problem 3 on Quiz 1, 2007.

Consider a high-speed merry-go-round which is similar to the one discussed in Problem 3 of Problem Set 1, but which has two levels. That is, there are four evenly spaced cars which travel around a central hub at speed v at a distance R from a central hub, and also another four cars that are attached to extensions of the four radial arms, each moving at a speed $2v$ at a distance $2R$ from the center. In this problem we will consider only light waves, not sound waves, and we will assume that v is not negligible compared to c, but that $2v < c$.



We learned in Problem Set 1 that there is no redshift when light from one car at radius R is received by an observer on another car at radius R.

(a) *(5 points)* Suppose that cars 5–8 are all emitting light waves in all directions. If an observer in car 1 receives light waves from each of these cars, what redshift z does she observe for each of the four signals?

(b) *(10 points)* Suppose that a spaceship is receding to the right at a relativistic speed u along a line through the hub, as shown in the diagram. Suppose that an observer in car 6 receives a radio signal from the spaceship, at the time when the car is in the position shown in the diagram. What redshift z is observed?

PROBLEM 20: SIGNAL PROPAGATION IN A FLAT MATTER-DOMINATED UNIVERSE *(55 points)*

The following problem was on Quiz 1, 2009.

Consider a flat, matter-dominated universe, with scale factor

$$a(t) = bt^{2/3} ,$$

where b is an arbitrary constant. For the following questions, the answer to any part may contain symbols representing the answers to previous parts, whether or not the previous part was answered correctly.

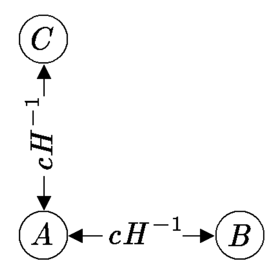
(a) *(10 points)* At time $t = t_1$, a light signal is sent from galaxy A. Let $\ell_{p,sA}(t)$ denote the physical distance of the signal from A at time t. (Note that $t = 0$ corresponds to the origin of the universe, not to the emission of the signal.) (i) Find the speed of separation of the light signal from A, defined as $\mathrm{d}\ell_{p,sA}/\mathrm{d}t$. What is the value of this speed (ii) at the time of emission, t_1, and (iii) what is its limiting value at arbitrarily late times?

(b) *(5 points)* Suppose that there is a second galaxy, galaxy B, that is located at a physical distance cH^{-1} from A at time t_1, where $H(t)$ denotes the Hubble expansion rate and c is the speed of light. (cH^{-1} is called the Hubble length.) Suppose that the light signal described above, which is emitted from galaxy A at time t_1, is directed toward galaxy B. At what time t_2 does it arrive at galaxy B?

(c) *(10 points)* Let $\ell_{p,sB}(t)$ denote the physical distance of the light signal from galaxy B at time t. (i) Find the speed of approach of the light signal towards B, defined as $-\mathrm{d}\ell_{p,sB}/\mathrm{d}t$. What is the value of this speed (ii) at the time of emission, t_1, and (iii) at the time of reception, t_2?

(d) *(10 points)* If an astronomer on galaxy A observes the light arriving from galaxy B at time t_1, what is its redshift z_{BA}?

(e) *(10 points)* Suppose that there is another galaxy, galaxy C, also located at a physical distance cH^{-1} from A at time t_1, but in a direction orthogonal to that of B. If galaxy B is observed from galaxy C at time t_1, what is the observed redshift z_{BC}? Recall that this universe is flat, so Euclidean geometry applies.



(f) *(10 points)* Suppose that galaxy A, at time t_1, emits electromagnetic radiation spherically symmetrically, with power output P. (P might be measured, for example, in watts, where 1 watt = 1 joule/second.) What is the radiation energy flux J that is received by galaxy B at time t_2, when the radiation reaches galaxy B? (J might be measured, for example, in watts per meter^2. Units are mentioned here only to help clarify the meaning of these quantities — your answer should have no explicit units, but should be expressed in terms of any or all of the given quantities t_1, P, and c, plus perhaps symbols representing the answers to previous parts.)

PROBLEM 21: DID YOU DO THE READING? *(25 points)*

The following problem appeared on Quiz 1 of 2011.

(a) *(10 points)* Hubble's law relates the distance of galaxies to their velocity. The Doppler effect provides an accurate tool to measure velocity, while the measure of cosmic distances is more problematic. Explain briefly the method that Hubble used to estimate the distance of galaxies in deriving his law.

(b) *(5 points)* One expects Hubble's law to hold as a consequence of the Cosmological Principle. What does the Cosmological Principle state?

(c) *(10 points)* Give a brief definition for the words *homogeneity* and *isotropy*. Then say for each of the following two statements whether it is true or false. If true explain briefly why. If false give a counter-example. You should assume Euclidean geometry (which Weinberg implicitly assumed in his discussion).

(i) If the universe is isotropic around one point then it has to be homogeneous.

(ii) If the universe is isotropic around two or more distinct points then it has to be homogeneous.

(d) Bonus question: *(2 points extra credit)* If we allow curved (i.e., non-Euclidean) spaces, is it true that a universe which is isotropic around two distinct points has to be homogeneous? If true explain briefly why, and otherwise give a counter-example.

∗ PROBLEM 22: THE TRAJECTORY OF A PHOTON ORIGINATING AT THE HORIZON *(25 points)*

The following problem appeared on Quiz 1 of 2011.

Consider again a flat matter-dominated universe, with a scale factor given by

$$a(t) = bt^{2/3} ,$$

where b is a constant. Let t_0 denote the current time.

(a) *(5 points)* What is the current value of the physical horizon distance $\ell_{p,\text{horizon}}(t_0)$? That is, what is the present distance of the most distant matter that can be seen, limited only by the speed of light.

(b) *(5 points)* Consider a photon that is arriving now from an object that is just at the horizon. Our goal is to trace the trajectory of this object. Suppose that we set up a coordinate system with us at the origin, and the source of the photon along the positive x-axis. What is the coordinate x_0 of the photon at $t = 0$?

(c) *(5 points)* As the photon travels from the source to us, what is its coordinate $x(t)$ as a function of time?

(d) *(5 points)* What is the physical distance $\ell_p(t)$ between the photon and us as a function of time?

(e) *(5 points)* What is the maximum physical distance $\ell_{p,\max}(t)$ between the photon and us, and at what time $t_{\max}$ does it occur?

PROBLEM 23: DID YOU DO THE READING (2016)?

The following problem was taken from Quiz 1, 2016, where it counted 35 points.

(a) *(5 points)* The Milky Way has been known since ancient times as a band of light stretching across the sky. We now recognize the Milky Way as the galaxy of stars in which we live, with a large collection of stars, including our sun, arranged in a giant disk. Since the individual stars are mostly too small for our eyes to resolve, we observe the collective light from these stars, concentrated in the plane of the disk. The idea that the Milky Way is actually a disk of stars was proposed by

(i) Claudius Ptolemy, in the 2nd century AD.

(ii) Johannes Kepler, in 1610.

(iii) Isaac Newton, in 1695.

(iv) Thomas Wright, in 1750.

(v) Immanuel Kant, in 1755.

(vi) Edwin Hubble, in 1923.

(b) *(5 points)* Once it was recognized that we live in a galaxy, it was initially assumed that ours was the only galaxy. The suggestion that some of the patches of light known as nebulae might actually be other galaxies like our own was made by

(i) Claudius Ptolemy, in the 2nd century AD.

(ii) Johannes Kepler, in 1610.

(iii) Isaac Newton, in 1695.

(iv) Thomas Wright, in 1750.

(v) Immanuel Kant, in 1755.

(vi) Edwin Hubble, in 1923.

(c) *(5 points)* The first firm evidence that there is more than one galaxy stemmed from the ability to observe the Andromeda Nebula with high enough resolution to distinguish its individual stars. In particular, the observation of Cepheid variable stars in Andromeda allowed a distance estimate that place it well outside the Milky Way. The observation of Cepheid variable stars in Andromeda was first made by

(i) Johannes Kepler, in 1610.

(ii) Isaac Newton, in 1695.

(iii Thomas Wright, in 1750.

(iv) Immanuel Kant, in 1755.

(v) Henrietta Swan Leavitt and Harlow Shapley in 1915.

(vi) Edwin Hubble, in 1923.

(d) *(5 points)* The first hint that the universe is filled with radiation with an effective temperature near 3 K, although not recognized at the time, was an observation of absorption lines in cyanogen (CN) by Adams and McKellar in 1941. They observed dark spectral lines which they interpreted as absorption by the cyanogen of light coming from the star behind the gas cloud. Explain in a few sentences how these absorption lines can be used to make inferences about the cosmic background radiation bathing the cyanogen gas cloud.

(e) *(5 points)* As the universe expands, the temperature of the cosmic microwave background

(i) goes up in proportion to the scale factor $a(t)$.

(ii) stays constant.

(iii) goes down in proportion to $1/a(t)$.

(iv) goes down in proportion to $1/a^2(t)$.

(f) *(5 points)* When Hubble measured the value of his constant, he found $H^{-1} \approx 100$ million years, 2 billion years, 10 billion years, or 20 billion years?

(g) *(5 points)* Explain in a few sentences what is meant by the equivalence principle?

PROBLEM 24: OBSERVING A DISTANT GALAXY IN A MATTER-DOMINATED FLAT UNIVERSE

The following problem was taken from Quiz 1, 2016, where it counted 40 points.

Suppose that we are living in a matter-dominated flat universe, with a scale factor given by

$$a(t) = bt^{2/3} \ ,$$

where b is a constant. The present time is denoted by t_0.

(a) *(5 points)* If we measure time in seconds, distance in meters, and coordinate distances in notches, what are the units of b?

(b) *(5 points)* Suppose that we observe a distant galaxy which is one half of a "Hubble length" away, which means that the physical distance today is $\ell_p = \frac{1}{2} c H_0^{-1}$, where c

is the speed of light and H_0 is the present value of the Hubble expansion rate. What is the proper velocity $v_p \equiv \frac{\mathrm{d}\ell_p(t)}{\mathrm{d}t}$ of this galaxy relative to us?

(c) *(5 points)* What is the coordinate distance ℓ_c between us and the distant galaxy?

If you did not answer the previous part, you may still continue with the following parts, using the symbol ℓ_c for the coordinate distance to the galaxy.

(d) *(5 points)* At what time t_e was the light that we are now receiving from the galaxy emitted?

(e) *(5 points)* What is the redshift z of the light that we are now receiving from the distant galaxy?

(f) *(10 points)* Consider a light pulse that leaves the distant galaxy at time t_e, as calculated in part (d), and arrives here at the present time, t_0. Calculate the physical distance $r_p(t)$ between the light pulse and us. Find $r_p(t)$ as a function of t for all t between t_e and t_0.

(g) *(5 points)* If we send a radio message now to the distant galaxy, at what time t_r will it be received?

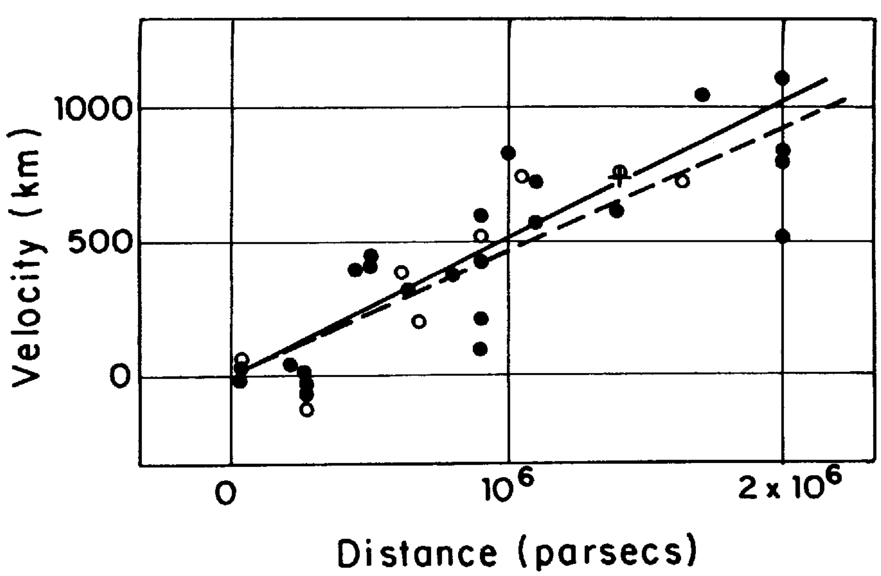
SOLUTIONS

PROBLEM 1: DID YOU DO THE READING (2000)? *(35 points)*

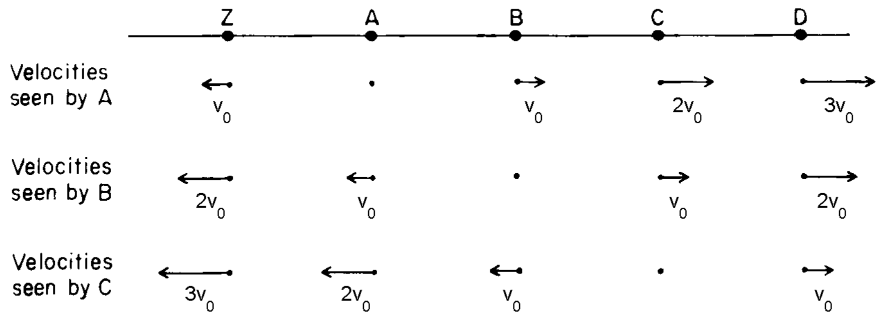
a) Doppler predicted the Doppler effect in $\boxed{\text{1842.}}$

b) Most of the stars of our galaxy, including our sun, lie in a flat disk. We therefore see much more light when we look out from earth along the plane of the disk than when we look in any other direction.

c) Hubble's original paper on the expansion of the universe was based on a study of only $\boxed{\text{18 galaxies.}}$ Well, at least Weinberg's book says 18 galaxies. For my own book I made a copy of Hubble's original graph, which seems to show 24 black dots, each of which represents a galaxy, as reproduced below. The vertical axis shows the recession velocity, in kilometers per second. The solid line shows the best fit to the black dots, each of which represents a galaxy. Each open circle represents a group of the galaxies shown as black dots, selected by their proximity in direction and distance; the broken line is the best fit to these points. The cross shows a statistical analysis of 22 galaxies for which individual distance measurements were not available. I am not sure why Weinberg refers to 18 galaxies, but it is possible that the text of Hubble's article indicated that 18 of these galaxies were measured with more reliability than the rest.



d)



e) During a time interval in which the linear size of the universe grows by 1%, the horizon distance grows by $\boxed{\text{more than 1\%.}}$ To see why, note that the horizon distance is equal to the scale factor times the comoving horizon distance. The scale factor grows by 1% during this time interval, but the comoving horizon distance also grows, since light from the distant galaxies has had more time to reach us.

f) Arno A. Penzias and Robert W. Wilson, Bell Telephone Laboratories.

g) (i) the average distance between photons: $\boxed{\text{proportional to the size of}}$ $\boxed{\text{the universe}}$ (Photons are neither created nor destroyed, so the only effect is that the average distance between them is stretched with the expansion. Since the universe expands uniformly, all distances grow by the same factor.)

(ii) the typical wavelength of the radiation: $\boxed{\text{proportional to the size of}}$ $\boxed{\text{the universe}}$ (See Lecture Notes 3.)

(iii) the number density of photons in the radiation: $\boxed{\text{inversely propor-}}$ $\boxed{\text{tional to the cube of the size of the universe}}$ (From (i), the average distance between photons grows in proportion to the size of the universe. Since the volume of a cube is proportional to the cube of the length of a side, the average volume occupied by a photon grows as the cube of the size of the universe. The number density is the inverse of the average volume occupied by a photon.)

(iv) the energy density of the radiation: $\boxed{\text{inversely proportional to the}}$ $\boxed{\text{fourth power of the size of the universe}}$ (The energy of each photon is proportional to its frequency, and hence inversely proportional to its wavelength. So from (ii) the energy of each photon is inversely proportional to the size of the universe, and from (iii) the number density is inversely proportional to the cube of the size.)

(v) the temperature of the radiation: $\boxed{\text{inversely proportional to the size}}$ $\boxed{\text{of the universe}}$ (The temperature is directly proportional to the average energy of a photon, which according to (iv) is inversely proportional to the size of the universe.)

PROBLEM 2: THE STEADY-STATE UNIVERSE THEORY *(25 points)*

a) *(10 points)* According to Eq. (3.7),

$$H(t) = \frac{1}{a(t)} \frac{da}{dt} .$$

So in this case

$$\frac{1}{a(t)} \frac{da}{dt} = H_0 ,$$

which can be rewritten as

$$\frac{da}{a} = H_0 \, dt .$$

Integrating,

$$\ln a = H_0 \, t + c ,$$

where c is a constant of integration. Exponentiating,

$$\boxed{a = be^{H_0 \, t} ,}$$

where $b = e^c$ is an arbitrary constant.

b) *(15 points)* Consider a cube of side ℓ_c drawn on the comoving coordinate system diagram. The physical length of each side is then $a(t) \, \ell_c$, so the physical volume is

$$V(t) = a^3(t) \, \ell_c^3 .$$

Since the mass density is fixed at $\rho = \rho_0$, the total mass inside this cube at any given time is given by

$$M(t) = a^3(t) \, \ell_c^3 \, \rho_0 .$$

In the absence of matter creation the total mass within a comoving volume would not change, so the increase in mass described by the above equation must be attributed to matter creation. The rate of matter creation per unit time per unit volume is then given by

$$\begin{aligned} \text{Rate} &= \frac{1}{V(t)} \frac{dM}{dt} \\ &= \frac{1}{a^3(t) \, \ell_c^3} \, 3a^2(t) \frac{da}{dt} \, \ell_c^3 \, \rho_0 \\ &= \frac{3}{a} \frac{da}{dt} \, \rho_0 \\ &= \boxed{3H_0 \, \rho_0 \, .} \end{aligned}$$

You were not asked to insert numbers, but it is worthwhile to consider the numerical value after the exam, to see what this answer is telling us. Suppose we take $H_0 = 70$ km-sec^{-1}-Mpc^{-1}, and take ρ_0 to be the critical density, $\rho_c = 3H_0^2/8\pi G$. Then

$$\begin{aligned}
\text{Rate} &= \frac{9H_0^3}{8\pi G} \\
&= \frac{9 \times \left(70\ \text{km-s}^{-1}\text{-Mpc}^{-1}\right)^3}{8\pi \times 6.67260 \times 10^{-11}\ \text{m}^3\text{-kg}^{-1}\text{-s}^{-2}} \\
&= \frac{9 \times \left(70\ \cancel{\text{km}}\text{-s}^{-1}\text{-}\cancel{\text{Mpc}}^{-1}\right)^3}{8\pi \times 6.67260 \times 10^{-11}\ \text{m}^3\text{-kg}^{-1}\text{-s}^{-2}} \\
&\qquad \times \left(\frac{1\ \cancel{\text{Mpc}}}{3.086 \times 10^{22}\ \cancel{\text{m}}}\right)^3 \times \left(\frac{10^3\ \cancel{\text{m}}}{\cancel{\text{km}}}\right)^3 \\
&= 6.26 \times 10^{-44}\ \text{kg-m}^{-3}\text{-s}^{-1}\ .
\end{aligned}$$

To put this number into more meaningful terms, note that the mass of a hydrogen atom is 1.67×10^{-27} kg, and that 1 year $= 3.156 \times 10^7$ s. The rate of matter production required for the steady-state universe theory can then be expressed as roughly one hydrogen atom per cubic meter per billion years! Needless to say, such a rate of matter production is totally undetectable, so the steady-state theory cannot be ruled out by the failure to detect matter production.

PROBLEM 3: DID YOU DO THE READING (2007)? *(25 points)*

The following 5 questions are each worth 5 points:

(a) In the 1940's, three astrophysicists proposed a "steady state" theory of cosmology, in which the universe has always looked about the same as it does now. State the last name of at least one of these authors. (*Bonus points*: you can earn 1 point each for naming the other two authors, and hence up to 2 additional points, but 1 point will be taken off for each incorrect answer.)

Ans: (Weinberg, page 8, or Ryden, page 16): Hermann Bondi, Thomas Gold, and Fred Hoyle.

(b) In 1917, a Dutch astronomer named Willem de Sitter did which one of the following accomplishments:

(i) measured the size of the Milky Way galaxy, finding it to be about one billion light-years in diameter.

(ii) resolved Cepheid variable stars in Andromeda and thereby obtained persuasive evidence that Andromeda is not within our own galaxy, but is apparently another galaxy like our own.

(iii) published a catalog, *Nebulae and Star Clusters*, listing 103 objects that astronomers should avoid when looking for comets.

$\boxed{\text{(iv)}}$ published a model for the universe, based on general relativity, which appeared to be static but which produced a redshift proportional to the distance.

(v) discovered that the orbital periods of the planets are proportional to the 3/2 power of the semi-major axis of their elliptical orbits.

Discussion: (i) is false in part because de Sitter was not involved in the measurement of the size of the Milky Way, but the most obvious error is in the size of the Milky Way. Its actual diameter is reported by Weinberg (p. 16) to be about 100,000 light-years, although now it is believed to be about twice that large. (ii) is an accurate description of an observation by Edwin Hubble in 1923 (Weinberg, pp. 19-20). (iii) describes the work of Charles Messier in 1781 (Weinberg, p. 17). (v) is of course one of Kepler's laws of planetary motion.

(c) In 1964–65, Arno A. Penzias and Robert W. Wilson observed a flux of microwave radiation coming from all directions in the sky, which was interpreted by a group of physicists at a neighboring institution as the cosmic background radiation left over from the big bang. Circle the two items on the following list that were **not** part of the story behind this spectacular discovery:

(i) Bell Telephone Laboratory	(ii) MIT	(iii) Princeton University
(iv) pigeons	$\boxed{\text{(v)}}$ ground hogs	$\boxed{\text{(vi)}}$ Hubble's constant
(vii) liquid helium	(viii) 7.35 cm	

(Grading: 3 pts for 1 correct answer, 5 for 2 correct answers, and -2 for each incorrect answer, but the minimum score is zero.)

Discussion: The discovery of the cosmic background radiation was described in some detail by Weinberg in Chapter 3. The observation was done at Bell Telephone Laboratories, in Holmdel, New Jersey. The detector was cooled with liquid helium to minimize electrical noise, and the measurements were made at a wavelength of 7.35 cm. During the course of the experiment the astronomers had to eject a pair of pigeons who were roosting in the antenna. Penzias and Wilson were not initially aware that the radiation they discovered might have come from the big bang, but Bernard Burke of MIT put them in touch with a group at Princeton University (Robert Dicke, James Peebles, P.G. Roll, and David Wilkinson) who were actively working on this hypothesis.

(d) Important predictions of the Copernican theory were confirmed by the discovery of the aberration of starlight (which showed that the velocity of the Earth has the time-dependence expected for rotation about the Sun) and by the behavior of the Foucault pendulum (which showed that the Earth rotates). These discoveries were made

(i) during Copernicus' lifetime.

(ii) approximately two and three decades after Copernicus' death, respectively.

(iii) about one hundred years after Copernicus' death.

$\boxed{\text{(iv)}}$ approximately two and three centuries after Copernicus' death, respectively.

Ryden discusses this on p. 5. The aberration of starlight was discovered in 1728, while the Foucault pendulum was invented in 1851.

(e) If one averages over sufficiently large scales, the universe appears to be homogeneous and isotropic. How large must the averaging scale be before this homogeneity and isotropy set in?

(i) 1 AU (1 AU $= 1.496 \times 10^{11}$ m).

(ii) 100 kpc (1 kpc = 1000 pc, 1 pc $= 3.086 \times 10^{16}$ m = 3.262 light-year).

(iii) 1 Mpc (1 Mpc $= 10^6$ pc).

(iv) 10 Mpc.

$\boxed{\text{(v)}}$ 100 Mpc.

(vi) 1000 Mpc.

This issue is discussed in Ryden's book on p. 11.

PROBLEM 4: AN EXPONENTIALLY EXPANDING UNIVERSE

(a) According to Eq. (3.7), the Hubble constant is related to the scale factor by

$$H = \dot{a}/a \ .$$

So

$$H = \frac{\chi a_0 e^{\chi t}}{a_0 e^{\chi t}} = \boxed{\ \chi \ .\ }$$

(b) According to Eq. (3.8), the coordinate velocity of light is given by

$$\frac{dx}{dt} = \frac{c}{a(t)} = \frac{c}{a_0} e^{-\chi t} \ .$$

Integrating,

$$\begin{aligned} x(t) &= \frac{c}{a_0} \int_0^t e^{-\chi t'} \, dt' \\ &= \frac{c}{a_0} \left[-\frac{1}{\chi} e^{-\chi t'} \right]_0^t \\ &= \boxed{\ \frac{c}{\chi a_0} \left[1 - e^{-\chi t} \right] \ .\ } \end{aligned}$$

(c) From Eq. (3.11), or from the front of the quiz, one has

$$1 + z = \frac{a(t_r)}{a(t_e)} \ .$$

Here $t_e = 0$, so

$$\begin{aligned} 1 + z &= \frac{a_0 e^{\chi t_r}}{a_0} \\ &\Longrightarrow \quad e^{\chi t_r} = 1 + z \\ &\Longrightarrow \quad \boxed{\ t_r = \frac{1}{\chi} \ln(1+z) \ .\ } \end{aligned}$$

(d) The coordinate distance is $x(t_r)$, where $x(t)$ is the function found in part (b), and t_r is the time found in part (c). So

$$e^{\chi t_r} = 1 + z \ ,$$

and

$$\begin{aligned} x(t_r) &= \frac{c}{\chi a_0}\left[1 - e^{-\chi t_r}\right] \\ &= \frac{c}{\chi a_0}\left[1 - \frac{1}{1+z}\right] \\ &= \frac{cZ}{\chi a_0 (1+z)} \;. \end{aligned}$$

The physical distance at the time of reception is found by multiplying by the scale factor at the time of reception, so

$$\ell_p(t_r) = a(t_r)x(t_r) = \frac{cze^{\chi t_r}}{\chi(1+z)} = \boxed{\frac{cz}{\chi}\;.}$$

PROBLEM 5: DID YOU DO THE READING (1986/1990 composite)?

(a) The distinguishing quantity is $\Omega \equiv \rho/\rho_c$. The universe is open if $\Omega < 1$, flat if $\Omega = 1$, or closed if $\Omega > 1$.

(b) The temperature of the microwave background today is about 3 Kelvin. (The best determination to date* was made by the COBE satellite, which measured the temperature as 2.728 ± 0.004 Kelvin. The error here is quoted with a 95% confidence limit, which means that the experimenters believe that the probability that the true value lies outside this range is only 5%.)

(c) The cosmic microwave background is observed to be highly isotropic.

(d) The distance to the Andromeda nebula is roughly 2 million light years.

(e) 1929.

(f) 2 billion years. Hubble's value for Hubble's constant was high by modern standards, by a factor of 5 to 10.

(g) The absolute luminosity (*i.e.*, the total light output) of a Cepheid variable star appears to be highly correlated with the period of its pulsations. This correlation can be used to estimate the distance to the Cepheid, by measuring the period and the apparent luminosity. From the period one can estimate the absolute luminosity of the star, and then one uses the apparent luminosity and the $1/r^2$ law for the intensity of a point source to determine the distance r.

(h) 10^7 light-years.

(i) Arno A. Penzias and Robert W. Wilson, Bell Telephone Laboratories.

(j) Princeton University.

---

* Astrophysical Journal, vol. **473**, p. 576 (1996): *The Cosmic Microwave Background Spectrum from the Full COBE FIRAS Data Sets*, D.J. Fixsen, E.S. Cheng, J.M. Gales, J.C. Mather, R.A. Shafer, and E.L. Wright.

PROBLEM 6: A FLAT UNIVERSE WITH UNUSUAL TIME EVOLUTION

The key to this problem is to work in comoving coordinates.

[Some students have asked me why one cannot use "physical" coordinates, for which the coordinates really measure the physical distances. In principle one can use any coordinate system on likes, but the comoving coordinates are the simplest. In any other system it is difficult to write down the trajectory of either a particle or a light-beam. In comoving coordinates it is easy to write the trajectory of either a light beam, or a particle which is moving with the expansion of the universe (and hence standing still in the comoving coordinates). Note, by the way, that when one says that a particle is standing still in comoving coordinates, one has not really said very much about it's trajectory. One has said that it is moving with the matter which fills the universe, but one has not said, for example, how the distance between the particle and origin varies with time. The answer to this latter question is then determined by the evolution of the scale factor, $a(t)$.]

(a) The physical separation at t_o is given by the scale factor times the coordinate distance. The coordinate distance is found by integrating the coordinate velocity, so

$$\ell_p(t_o) = a(t_o) \int_{t_e}^{t_o} \frac{c\,dt'}{a(t')} = bt_o^{1/3} \int_{t_e}^{t_o} \frac{c\,dt'}{bt'^{1/3}} = \frac{3}{2} c t_o^{1/3} \left[t_o^{2/3} - t_e^{2/3} \right]$$

$$\boxed{= \frac{3}{2} c t_o \left[1 - (t_e/t_o)^{2/3} \right] .}$$

(b) From the front of the exam,

$$1 + z = \frac{a(t_o)}{a(t_e)} = \left(\frac{t_o}{t_e} \right)^{1/3}$$

$$\Longrightarrow \quad \boxed{z = \left(\frac{t_o}{t_e} \right)^{1/3} - 1 .}$$

(c) By combining the answers to (a) and (b), one has

$$\boxed{\ell_p(t_o) = \frac{3}{2} c t_o \left[1 - \frac{1}{(1+z)^2} \right] .}$$

(d) The physical distance of the light pulse at time t is equal to $a(t)$ times the coordinate distance. The coordinate distance at time t is equal to the starting coordinate distance, $\ell_c(t_e)$, minus the coordinate distance that the light pulse travels between time t_e and time t. Thus,

$$\begin{aligned}
\ell_p(t) &= a(t)\left[\ell_c(t_e) - \int_{t_e}^{t} \frac{c\,dt'}{a(t')}\right] \\
&= a(t)\left[\int_{t_e}^{t_o} \frac{c\,dt'}{a(t')} - \int_{t_e}^{t} \frac{c\,dt'}{a(t')}\right] \\
&= a(t)\int_{t}^{t_o} \frac{c\,dt'}{a(t')} \\
&= bt^{1/3}\int_{t}^{t_o} \frac{c\,dt'}{bt'^{1/3}} = \frac{3}{2}ct^{1/3}\left[t_o^{2/3} - t^{2/3}\right] \\
&= \boxed{\frac{3}{2}ct\left[\left(\frac{t_o}{t}\right)^{2/3} - 1\right]\ .}
\end{aligned}$$

PROBLEM 7: ANOTHER FLAT UNIVERSE WITH AN UNUSUAL TIME EVOLUTION *(40 points)*

a) *(5 points)* The cosmological redshift is given by the usual form,

$$1 + z = \frac{a(t_0)}{a(t_e)}\ .$$

For light emitted by an object at time t_e, the redshift of the received light is

$$1 + z = \frac{a(t_0)}{a(t_e)} = \left(\frac{t_0}{t_e}\right)^{\gamma}\ .$$

So,

$$\boxed{z = \left(\frac{t_0}{t_e}\right)^{\gamma} - 1\ .}$$

b) *(5 points)* The coordinates t_0 and t_e are cosmic time coordinates. The “look-back” time as defined in the exam is then the interval $t_0 - t_e$. We can write this as

$$t_0 - t_e = t_0\left(1 - \frac{t_e}{t_0}\right)\ .$$

We can use the result of part (a) to eliminate t_e/t_0 in favor of z. From (a),

$$\frac{t_e}{t_0} = (1+z)^{-1/\gamma} \quad .$$

Therefore,

$$\boxed{t_0 - t_e = t_0\left[1 - (1+z)^{-1/\gamma}\right] \quad .}$$

c) *(10 points)* The present value of the physical distance to the object, $\ell_p(t_0)$, is found from

$$\ell_p(t_0) = a(t_0)\int_{t_e}^{t_0} \frac{c}{a(t)}dt \quad .$$

Calculating this integral gives

$$\ell_p(t_0) = \frac{ct_0^\gamma}{1-\gamma}\left[\frac{1}{t_0^{\gamma-1}} - \frac{1}{t_e^{\gamma-1}}\right] \quad .$$

Factoring $t_0^{\gamma-1}$ out of the parentheses gives

$$\ell_p(t_0) = \frac{ct_0}{1-\gamma}\left[1 - \left(\frac{t_0}{t_e}\right)^{\gamma-1}\right] \quad .$$

This can be rewritten in terms of z and H_0 using the result of part (a) as well as,

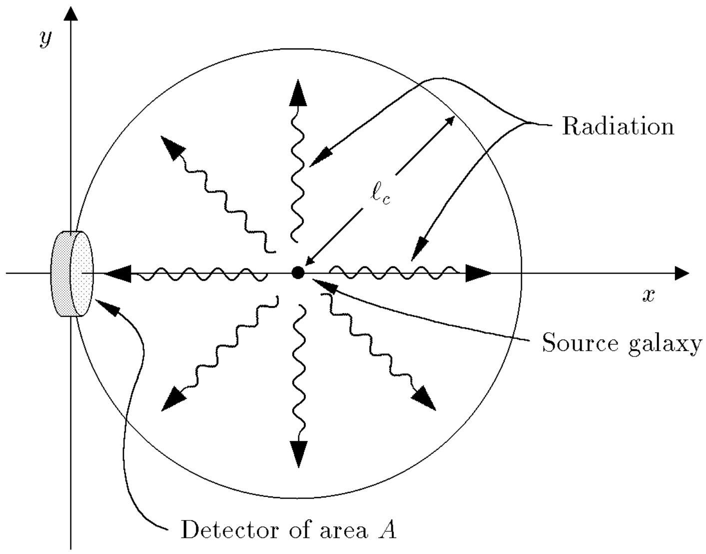
$$H_0 = \frac{\dot{a}(t_0)}{a(t_0)} = \frac{\gamma}{t_0} \quad .$$

Finally then,

$$\boxed{\ell_p(t_0) = cH_0^{-1}\frac{\gamma}{1-\gamma}\left[1 - (1+z)^{\frac{\gamma-1}{\gamma}}\right] \quad .}$$

d) *(10 points)* A nearly identical problem was worked through in Problem 8 of Problem Set 1.

The energy of the observed photons will be redshifted by a factor of $(1+z)$. In addition the rate of arrival of photons will be redshifted relative to the rate of photon emmission, reducing the flux by another factor of $(1+z)$. Consequently, the observed power will be redshifted by two factors of $(1+z)$ to $P/(1+z)^2$.



Imagine a hypothetical sphere in comoving coordinates as drawn above, centered on the radiating object, with radius equal to the comoving distance ℓ_c. Now consider the photons passing through a patch of the sphere with physical area A. In comoving coordinates the present area of the patch is $A/a(t_0)^2$. Since the object radiates uniformly in all directions, the patch will intercept a fraction $(A/a(t_0)^2)/(4\pi\ell_c^2)$ of the photons passing through the sphere. Thus the power hitting the area A is

$$\frac{(A/a(t_0)^2)}{4\pi\ell_c^2}\frac{P}{(1+z)^2} \quad .$$

The radiation energy flux J, which is the received power per area, reaching the earth is then given by

$$J = \frac{1}{4\pi\ell_p(t_0)^2}\frac{P}{(1+z)^2}$$

where we used $\ell_p(t_0) = a(t_0)\ell_c$. Using the result of part (c) to write J in terms of P, H_0, z, and γ gives,

$$\boxed{J = \frac{H_0^2}{4\pi c^2}\left(\frac{1-\gamma}{\gamma}\right)^2 \frac{P}{(1+z)^2\left[1-(1+z)^{\frac{\gamma-1}{\gamma}}\right]^2} \quad .}$$

e) *(10 points)* Following the solution of Problem 1 of Problem Set 1, we can introduce a fictitious relay station that is at rest relative to the galaxy, but located just next to the jet, between the jet and Earth. As in the previous solution, the relay station simply rebroadcasts the signal it receives from the source, at exactly the instant that it receives it. The relay station therefore has no effect on the signal received by the observer, but allows us to divide the problem into two simple parts.

The distance between the jet and the relay station is very short compared to cosmological scales, so the effect of the expansion of the universe is negligible. For this part of the problem we can use special relativity, which says that the period with which the relay station measures the received radiation is given by

$$\Delta t_{\text{relay station}} = \sqrt{\frac{1-\frac{v}{c}}{1+\frac{v}{c}}} \times \Delta t_{\text{source}} \ .$$

Note that I have used the formula from the front of the exam, but I have changed the size of v, since the source in this case is moving toward the relay station, so the light is blue-shifted. To observers on Earth, the relay station is just a source at rest in the comoving coordinate system, so

$$\Delta t_{\text{observed}} = (1+z)\Delta t_{\text{relay station}} \ .$$

Thus,

$$\begin{aligned} 1 + z_J \equiv \frac{\Delta t_{\text{observed}}}{\Delta t_{\text{source}}} &= \frac{\Delta t_{\text{observed}}}{\Delta t_{\text{relay station}}} \frac{\Delta t_{\text{relay station}}}{\Delta t_{\text{source}}} \\ &= (1+z)\big|_{\text{cosmological}} \ \times \ (1+z)\big|_{\text{special relativity}} \\ &= (1+z)\sqrt{\frac{1-\frac{v}{c}}{1+\frac{v}{c}}} \ . \end{aligned}$$

Thus,

$$\boxed{z_J = (1+z)\sqrt{\frac{1-\frac{v}{c}}{1+\frac{v}{c}}} - 1 \ .}$$

Note added: In looking over the solutions to this problem, I found that a substantial number of students wrote solutions based on the incorrect assumption that the Doppler shift could be treated as if it were entirely due to motion. These students used the special relativity Doppler shift formula to convert the redshift z of the galaxy to a velocity of recession, then subtracted from this the speed v of the jet, and then again used the special relativity Doppler shift formula to find the Doppler shift corresponding to this composite velocity. However, as discussed at the end of Lecture Notes 3, the cosmological Doppler shift is given by

$$1 + z \equiv \frac{\Delta t_o}{\Delta t_e} = \frac{a(t_o)}{a(t_e)} \ , \tag{3.11}$$

and is not purely an effect caused by motion. It is really the combined effect of the motion of the distant galaxies and the gravitational field that exists between the galaxies, so the special relativity formula relating z to v does not apply.

PROBLEM 8: DID YOU DO THE READING (1996)?

a) The lines were dark, caused by absorption of the radiation in the cooler, outer layers of the sun.

b) Individual stars in the Andromeda Nebula were resolved by Hubble in 1923.

[The other names and dates are not without significance. In 1609 Galileo built his first telescope; during 1609-10 he resolved the individual stars of the Milky Way, and also discovered that the surface of the moon is irregular, that Jupiter has moons of its own, that Saturn has handles (later recognized as rings), that the sun has spots, and that Venus has phases. In 1755 Immanuel Kant published his *Universal Natural History and Theory of the Heavens*, in which he suggested that at least some of the nebulae are galaxies like our own. In 1912 Henrietta Leavitt discovered the relationship between the period and luminosity of Cepheid variable stars. In the 1950s Walter Baade and Allan Sandage recalibrated the extra-galactic distance scale, reducing the accepted value of the Hubble constant by about a factor of 10.]

c)

(i) True. [In 1941, A. McKellar discovered that cyanogen clouds behave as if they are bathed in microwave radiation at a temperature of about 2.3° K, but no connection was made with cosmology.]

(ii) False. [Any radiation reflected by the clouds is far too weak to be detected. It is the bright starlight shining through the cloud that is detectable.]

(iii) True. [Electromagnetic waves at these wavelengths are mostly blocked by the Earth's atmosphere, so they could not be detected directly until high altitude balloons and rockets were introduced into cosmic background radiation research in the 1970s. Precise data was not obtained until the COBE satellite, in 1990.]

(iv) True. [The microwave radiation can boost the CN molecule from its ground state to a low-lying excited state, a state in which the C and N atoms rotate about each other. The population of this low-lying state is therefore determined by the intensity of the microwave radiation. This population is measured by observing the absorption of starlight passing through the clouds, since there are absorption lines in the visible spectrum caused by transitions between the low-lying state and higher energy excited states.]

(v) False. [No chemical reactions are seen.]

d) Aristarchus. [The heliocentric picture was never accepted by other Greek philosophers, however, and was not revived until the publication of *De Revolutionibus Orbium Coelestium* (*On the Revolutions of the Celestial Spheres*) by Copernicus in 1543.]

e) (ii) Any patch of the night sky would look as bright as the surface of the sun. [Explanation: The crux of the argument is that the brightness of an object, measured for example by the power per area (i.e., flux) hitting the retina of your eye, does not change as the object is moved further away. The power falls off with the square of the distance, but so does the area of the image on your retina — so the power per area is independent of distance. Under the assumptions stated, your line of sight will eventually hit a star no matter what direction you are looking. The energy flux on your retina will therefore be the same as in the image of the sun, so the entire sky will appear as bright as the surface of the sun.]

PROBLEM 9: A FLAT UNIVERSE WITH $a(t) \propto t^{3/5}$

a) In general, the Hubble constant is given by $H = \dot{a}/a$, where the overdot denotes a derivative with respect to cosmic time t. In this case

$$H = \frac{1}{bt^{3/5}} \frac{3}{5} bt^{-2/5} = \boxed{\frac{3}{5t}\ .}$$

b) In general, the (physical) horizon distance is given by

$$\ell_{p,\mathrm{horizon}}(t) = a(t) \int_0^t \frac{c}{a(t')} dt' \ .$$

In this case one has

$$\ell_{p,\mathrm{horizon}}(t) = bt^{3/5} \int_0^t \frac{c}{bt'^{3/5}} dt' = ct^{3/5}\, \frac{5}{2} \left[t^{2/5} - 0^{2/5} \right] = \boxed{\frac{5}{2}\, ct\ .}$$

c) The coordinate speed of light is $c/a(t)$, so the coordinate distance that light travels between t_A and t_B is given by

$$\ell_c = \int_{t_A}^{t_B} \frac{c}{a(t')}\, dt' = \int_{t_A}^{t_B} \frac{c}{bt'^{3/5}}\, dt' = \boxed{\frac{5c}{2b} \left(t_B^{2/5} - t_A^{2/5} \right)\ .}$$

d) The physical separation is just the scale factor times the coordinate separation, so

$$\ell_p(t_A) = a(t_A)\, \ell_c = \boxed{\frac{5}{2} ct_A \left[\left(\frac{t_B}{t_A} \right)^{2/5} - 1 \right]\ .}$$

$$\ell_p(t_B) = a(t_B)\,\ell_c = \boxed{\frac{5}{2}ct_B\left[1 - \left(\frac{t_A}{t_B}\right)^{2/5}\right] .}$$

e) Let t_{eq} be the time at which the light pulse is equidistant from the two galaxies. At this time it will have traveled a coordinate distance $\ell_c/2$, where ℓ_c is the answer to part (c). Since the coordinate speed is $c/a(t)$, the time t_{eq} can be found from:

$$\int_{t_A}^{t_{\text{eq}}} \frac{c}{a(t')}\,dt' = \frac{1}{2}\ell_c$$

$$\frac{5c}{2b}\left(t_{\text{eq}}^{2/5} - t_A^{2/5}\right) = \frac{5c}{4b}\left(t_B^{2/5} - t_A^{2/5}\right)$$

Solving for t_{eq},

$$\boxed{t_{\text{eq}} = \left[\frac{t_A^{2/5} + t_B^{2/5}}{2}\right]^{5/2} .}$$

f) According to Hubble's law, the speed is equal to Hubble's constant times the physical distance. By combining the answers to parts (a) and (d), one has

$$v = H(t_A)\,\ell_p(t_A)$$

$$= \frac{3}{5t_A}\,\frac{5}{2}ct_A\left[\left(\frac{t_B}{t_A}\right)^{2/5} - 1\right] = \boxed{\frac{3}{2}c\left[\left(\frac{t_B}{t_A}\right)^{2/5} - 1\right] .}$$

g) The redshift for radiation observed at time t can be written as

$$1 + z = \frac{a(t)}{a(t_e)} ,$$

where t_e is the time that the radiation was emitted. Solving for t_e,

$$t_e = \frac{t}{(1+z)^{5/3}} .$$

As found in part (d), the physical distance that the light travels between t_e and t, as measured at time t, is given by

$$\ell_p(t) = a(t)\int_{t_e}^{t} \frac{c}{a(t')}\,dt' = \frac{5}{2}ct\left[1 - \left(\frac{t_e}{t}\right)^{2/5}\right] .$$

Substituting the expression for t_e, one has

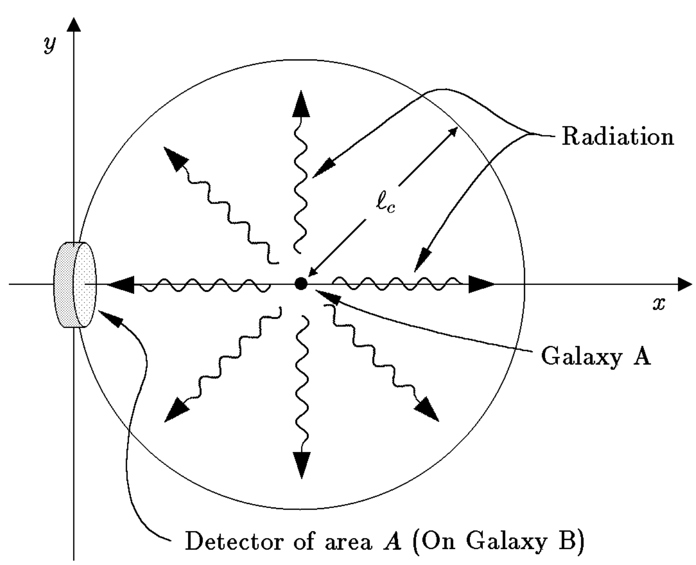
$$\boxed{\ell_p(t) = \frac{5}{2}\, ct \left[1 - \frac{1}{(1+z)^{2/3}}\right] .}$$

As $z \to \infty$, this expression approaches

$$\boxed{\lim_{z\to\infty} \ell_p(t) = \frac{5}{2}\, ct \ ,}$$

which is exactly equal to the horizon distance. It is a general rule that the horizon distance corresponds to infinite redshift z.

h) Again we will view the problem in comoving coordinates. Put galaxy B at the origin, and galaxy A at a coordinate distance ℓ_c along the x-axis. Draw a sphere of radius ℓ_c, centered galaxy A. Also draw a detector on galaxy B, with physical area A (measured at the present time).



The energy from the quasar will radiate uniformly on the sphere. The detector has a physical area A, so in the comoving coordinate picture its area in square notches would be $A/a(t_B)^2$. The detector therefore occupies a fraction of the sphere given by

$$\frac{[A/a(t_B)^2]}{4\pi\ell_c^2} = \frac{A}{4\pi\ell_p(t_B)^2} \ ,$$

so this fraction of the emitted photons will strike the detector.

Next consider the rate of arrival of the photons at the sphere. In lecture we figured out that if a periodic wave is emitted at time t_A and observed at time t_B, then the rate of arrival of the wave crests will be slower than the rate of emission by a redshift factor $1 + z = a(t_B)/a(t_A)$. The same argument will apply to the rate of arrival of photons, so the rate of photon arrival at the sphere will be slower than the rate of emission by the factor $1 + z$, reducing the energy flux by this factor. In addition, each photon is redshifted in frequency by $1 + z$. Since the energy of each photon is proportional to its frequency, the energy flux is reduced by an additional factor of $1 + z$. Thus, the rate at which energy reaches the detector is

$$\text{Power hitting detector} = \frac{A}{4\pi \ell_p(t_B)^2} \frac{P}{(1+z)^2} \ .$$

The red shift z of the light pulse received at galaxy B is given by

$$1 + z = \frac{a(t_B)}{a(t_A)} = \left(\frac{t_B}{t_A}\right)^{3/5} \ .$$

Using once more the expression for $\ell_P(t_B)$ from part (d), one has

$$\boxed{J = \frac{\text{Power hitting detector}}{A} = \frac{P(t_A/t_B)^{6/5}}{25\pi\, c^2\, t_B^2 \left[1 - \left(\frac{t_A}{t_B}\right)^{2/5}\right]^2} \ .}$$

The problem is worded so that t_A, and not z, is the given variable that determines how far galaxy A is from galaxy B. In practice, however, it is usually more useful to express the answer in terms of the redshift z of the received radiation. One can do this by using the above expression for $1 + z$ to eliminate t_A in favor of z, finding

$$J = \frac{P}{25\pi\, c^2\, t_B^2 (1+z)^{2/3} \left[(1+z)^{2/3} - 1\right]^2} \ .$$

i) Let t'_A be the time at which the light pulse arrives back at galaxy A. The pulse must therefore travel a coordinate distance ℓ_c (the answer to part (c)) between time t_B and t'_A, so

$$\int_{t_B}^{t'_A} \frac{c}{a(t')}\, dt' = \ell_c \ .$$

Using the answer from (c) and integrating the left-hand side,

$$\frac{5c}{2b}\left(t_A'^{2/5} - t_B^{2/5}\right) = \frac{5c}{2b}\left(t_B^{2/5} - t_A^{2/5}\right) \ .$$

Solving for t'_A,

$$\boxed{t'_A = \left(2t_B^{2/5} - t_A^{2/5}\right)^{5/2} \ .}$$

PROBLEM 10: DID YOU DO THE READING (1998)?

a) Einstein believed that the universe was static, and the cosmological term was necessary to prevent a static universe from collapsing under the attractive force of normal gravity. [The repulsive effect of a cosmological constant grows linearly with distance, so if the coefficient is small it is important only when the separations are very large. Such a term can be important cosmologically while still being too small to be detected by observations of the solar system or even the galaxy. Recent measurements of distant supernovas ($z \approx 1$), which you may have read about in the newspapers, make it look like maybe there is a cosmological constant after all! Since the cosmological constant is the hot issue in cosmology this season, we will want to look at it more carefully. The best time will be after Lecture Notes 7.]

b) At the time of its discovery, de Sitter's model was thought to be static [although it was known that the model predicted a redshift which, at least for nearby galaxies, was proportional to the distance]. From a modern perspective the model is thought to be expanding.

[It seems strange that physicists in 1917 could not correctly determine if the theory described a universe that was static or expanding, but the mathematical formalism of general relativity can be rather confusing. The basic problem is that when space is not Euclidean there is no simple way to assign coordinates to it. The mathematics of general relativity is designed to be valid for any coordinate system, but the underlying physics can sometimes be obscured by a peculiar choice of coordinates. A change of coordinates can not only distort the apparent geometry of space, but it can also mix up space and time. The de Sitter model was first written down in coordinates that made it look static, so everyone believed it was. Later Arthur Eddington and Hermann Weyl (independently) calculated the trajectories of test particles, discovering that they flew apart.]

c) $n_1 = 3$, and $n_2 = 4$.

d) Above 3,000 K the universe was so hot that the atoms were ionized, dissociated into nuclei and free electrons. At about this temperature, however, the universe was cool enough so that the nuclei and electrons combined to form neutral atoms.

[This process is usually called “recombination,” although the prefix “re-” is totally inaccurate, since in the big bang theory these constituents had never been previously combined. As far as I know the word was first used in this context by P.J.E. Peebles, so I once asked him why the prefix was used. He replied that this word is standard terminology in plasma physics, and was carried over into cosmology.]

[Regardless of its name, recombination was crucial for the clumping of matter into galaxies and stars, because the pressure of the photons in the early universe was enormous. When the matter was ionized, the free electrons interacted strongly with the photons, so the pressure of these photons prevented the matter from clumping. After recombination, however, the matter became very transparent to radiation, and the pressure of the radiation became ineffective.]

[Incidentally, at roughly the same time as recombination (with big uncertainties), the mass density of the universe changed from being dominated by radiation (photons and neutrinos) to being dominated by nonrelativistic matter. There is no known underlying connection between these two events, and it seems to be something of a coincidence that they occurred at about the same time. The transition from radiation-domination to matter-domination also helped to promote the clumping of matter, but the effect was much weaker than the effect of recombination—because of the very high velocity of photons and neutrinos, their pressure remained a significant force even after their mass density became much smaller than that of matter.]

PROBLEM 11: ANOTHER FLAT UNIVERSE WITH $a(t) \propto t^{3/5}$

a) According to Eq. (3.7) of the Lecture Notes,

$$H(t) = \frac{1}{a(t)}\,\frac{da}{dt}\ .$$

For the special case of $a(t) = bt^{3/5}$, this gives

$$H(t) = \frac{1}{bt^{3/5}}\frac{3}{5}bt^{-2/5} = \boxed{\frac{3}{5t}\ .}$$

b) According to Eq. (3.8) of the Lecture Notes, the coordinate velocity of light (in comoving coordinates) is given by

$$\frac{dx}{dt} = \frac{c}{a(t)}\ .$$

Since galaxies A and B have physical separation ℓ_0 at time t_1, their coordinate separation is given by

$$\ell_c = \frac{\ell_0}{b t_1^{3/5}} \ .$$

The radio signal must cover this coordinate distance in the time interval from t_1 to t_2, which implies that

$$\int_{t_1}^{t_2} \frac{c}{a(t)}\, dt = \frac{\ell_0}{b t_1^{3/5}} \ .$$

Using the expression for $a(t)$ and integrating,

$$\frac{5c}{2b}\left(t_2^{2/5} - t_1^{2/5}\right) = \frac{\ell_0}{b t_1^{3/5}} \ ,$$

which can be solved for t_2 to give

$$\boxed{t_2 = \left(1 + \frac{2\ell_0}{5ct_1}\right)^{5/2} t_1 \ .}$$

c) The method is the same as in part (b). The coordinate distance between the two galaxies is unchanged, but this time the distance must be traversed in the time interval from t_2 to t_3. So,

$$\int_{t_2}^{t_3} \frac{c}{a(t)}\, dt = \frac{\ell_0}{b t_1^{3/5}} \ ,$$

which leads to

$$\frac{5c}{2b}\left(t_3^{2/5} - t_2^{2/5}\right) = \frac{\ell_0}{b t_1^{3/5}} \ .$$

Solving for t_3 gives

$$\boxed{t_3 = \left[\left(\frac{t_2}{t_1}\right)^{2/5} + \frac{2\ell_0}{5ct_1}\right]^{5/2} t_1 \ .}$$

The above answer is perfectly acceptable, but one could also replace t_2 by using the answer to part (b), which gives

$$\boxed{t_3 = \left(1 + \frac{4\ell_0}{5ct_1}\right)^{5/2} t_1 \ .}$$

[Alternatively, one could have begun the problem by considering the full round trip of the radio signal, which travels a coordinate distance $2\ell_c$ during the time interval from t_1 to t_3. The problem then becomes identical to part (b), except that the coordinate distance ℓ_c is replaced by $2\ell_c$, and t_2 is replaced by t_3. One is led immediately to the answer in the form of the previous equation.]

d) Cosmic time is defined by the reading of suitably synchronized clocks which are each at rest with respect to the matter of the universe at the same location. (For this problem we will not need to think about the method of synchronization.) Thus, the cosmic time interval between the receipt of the message and the response is the same as what is measured on the galaxy B clocks, which is Δt. The response is therefore sent at cosmic time $t_2 + \Delta t$. The coordinate distance between the galaxies is still $\ell_0/a(t_1)$, so

$$\int_{t_2+\Delta t}^{t_4} \frac{c}{a(t)}\, dt = \frac{\ell_0}{b t_1^{3/5}} \ .$$

Integration gives

$$\frac{5c}{2b}\left[t_4^{2/5} - (t_2 + \Delta t)^{2/5}\right] = \frac{\ell_0}{b t_1^{3/5}} \ ,$$

which can be solved for t_4 to give

$$\boxed{t_4 = \left[\left(\frac{t_2 + \Delta t}{t_1}\right)^{2/5} + \frac{2\ell_0}{5ct_1}\right]^{5/2} t_1 \ .}$$

e) From the formula at the front of the exam,

$$1 + z = \frac{a(t_{\text{observed}})}{a(t_{\text{emitted}})} = \frac{a(t_4)}{a(t_2 + \Delta t)} = \left(\frac{t_4}{t_2 + \Delta t}\right)^{3/5} \ .$$

So,

$$\boxed{z = \frac{a(t_{\text{observed}})}{a(t_{\text{emitted}})} = \frac{a(t_4)}{a(t_2 + \Delta t)} = \left(\frac{t_4}{t_2 + \Delta t}\right)^{3/5} - 1 \ .}$$

f) If Δt is small compared to the time that it takes $a(t)$ to change significantly, then the interval between a signal sent at t_3 and a signal sent at $t_3 + \Delta t$ will be received with a redshift identical to that observed between two successive crests of a wave. Thus, the separation between the receipt of the acknowledgement and the receipt of the response will be a factor $(1+z)$ times longer than the time interval between the sending of the two signals, and therefore

$$\begin{aligned} t_4 - t_3 &= (1+z)\Delta t + \mathcal{O}(\Delta t^2) \\ &= \left(\frac{t_4}{t_2+\Delta t}\right)^{3/5} \Delta t + \mathcal{O}(\Delta t^2) \ . \end{aligned}$$

Since the answer contains an explicit factor of Δt, the other factors can be evaluated to zeroth order in Δt:

$$\boxed{t_4 - t_3 = \left(\frac{t_4}{t_2}\right)^{3/5} \Delta t + \mathcal{O}(\Delta t^2) \ ,}$$

where to first order in Δt the t_4 in the numerator could equally well have been replaced by t_3.

For those who prefer the brute force approach, the answer to part (d) can be Taylor expanded in powers of Δt. To first order one has

$$t_4 = t_3 + \left.\frac{\partial t_4}{\partial \Delta t}\right|_{\Delta t=0} \Delta t + \mathcal{O}(\Delta t^2) \ .$$

Evaluating the necessary derivative gives

$$\frac{\partial t_4}{\partial \Delta t} = \left[\left(\frac{t_2+\Delta t}{t_1}\right)^{2/5} + \frac{2\ell_0}{5ct_1}\right]^{3/2} \left(\frac{t_2+\Delta t}{t_1}\right)^{-3/5} \ ,$$

which when specialized to $\Delta t = 0$ becomes

$$\left.\frac{\partial t_4}{\partial \Delta t}\right|_{\Delta t=0} = \left[\left(\frac{t_2}{t_1}\right)^{2/5} + \frac{2\ell_0}{5ct_1}\right]^{3/2} \left(\frac{t_2}{t_1}\right)^{-3/5} \ .$$

Using the first boxed answer to part (c), this can be simplified to

$$\begin{aligned} \left.\frac{\partial t_4}{\partial \Delta t}\right|_{\Delta t=0} &= \left(\frac{t_3}{t_1}\right)^{3/5} \left(\frac{t_2}{t_1}\right)^{-3/5} \\ &= \left(\frac{t_3}{t_2}\right)^{3/5} \ . \end{aligned}$$

Putting this back into the Taylor series gives

$$t_4 - t_3 = \left(\frac{t_3}{t_2}\right)^{3/5} \Delta t + \mathcal{O}(\Delta t^2) \ ,$$

in agreement with the previous answer.

PROBLEM 12: THE DECELERATION PARAMETER

From the front of the exam, we are reminded that

$$\ddot{a} = -\frac{4\pi}{3} G\rho a$$

and

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3} G\rho - \frac{kc^2}{a^2} \ ,$$

where a dot denotes a derivative with respect to time t. The critical mass density ρ_c is defined to be the mass density that corresponds to a flat ($k = 0$) universe, so from the equation above it follows that

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3} G\rho_c \ .$$

Substituting into the definition of q, we find

$$\begin{aligned} q = -\ddot{a}(t)\frac{a(t)}{\dot{a}^2(t)} &= -\frac{\ddot{a}}{a}\left(\frac{a}{\dot{a}}\right)^2 \\ &= \left(\frac{4\pi}{3} G\rho\right)\left(\frac{3}{8\pi G\rho_c}\right) = \frac{1}{2}\frac{\rho}{\rho_c} = \boxed{\frac{1}{2}\Omega \ .} \end{aligned}$$

PROBLEM 13: A RADIATION-DOMINATED FLAT UNIVERSE

The flatness of the model universe means that $k = 0$, so

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3} G\rho \ .$$

Since

$$\rho(t) \propto \frac{1}{a^4(t)} \ ,$$

it follows that

$$\frac{da}{dt} = \frac{\text{const}}{a} \ .$$

Rewriting this as

$$a\, da = \text{const}\, dt \ ,$$

the indefinite integral becomes

$$\frac{1}{2} a^2 = (\text{const})t + c' \ ,$$

where c' is a constant of integration. Different choices for c' correspond to different choices for the definition of $t = 0$. We will follow the standard convention of choosing $c' = 0$, which sets $t = 0$ to be the time when $a = 0$. Thus the above equation implies that $a^2 \propto t$, and therefore

$$\boxed{a(t) \propto t^{1/2}}$$

for a photon-dominated flat universe.

PROBLEM 14: DID YOU DO THE READING (2004)? *(25 points)*

(a) In 1826, the astronomer Heinrich Olber wrote a paper on a paradox regarding the night sky. What is Olber's paradox? What is the primary resolution of it?

(Ryden, Chapter 2, Pages 6-8)

Ans: Olber's paradox is that the night sky appears to be dark, instead of being uniformly bright. The primary resolution is that the universe has a finite age, and so the light from stars beyond the horizon distance has not reached us yet. (However, even in the steady-state model of the universe, the paradox is resolved because the light from distant stars will be red-shifted beyond the visible spectrum).

(b) What is the value of the Newtonian gravitational constant G in Planck units? The Planck length is of the order of 10^{-35} m, 10^{-15} m, 10^{15} m, or 10^{35} m?

(Ryden, Chapter 1, Page 3)

Ans: $G = 1$ in Planck units, by definition.

The Planck length is of the order of 10^{-35} m. (Note that this answer could be obtained by a process of elimination as long as you remember that the Planck length is much smaller than 10^{-15} m, which is the typical size of a nucleus).

(c) What is the Cosmological Principle? Is the Hubble expansion of the universe consistent with it?

(Weinberg, Chapter 2, Pages 21-23; Ryden, Chapter 2, Page 11)

Ans: The Cosmological Principle states that there is nothing special about our location in the universe, i.e. the universe is homogeneous and isotropic.

Yes, the Hubble expansion is consistent with it (since there is no center of expansion).

(d) In the "Standard Model" of the universe, when the universe cooled to about 3×10^a K, it became transparent to photons, and today we observe these as the Cosmic Microwave Background (CMB) at a temperature of about 3×10^b K. What are the integers a and b?

(Weinberg, Chapter 3; Ryden, Chapter 2, Page 22)

$a = 3$, $b = 0$.

(e) What did the universe primarily consist of at about 1/100th of a second after the Big Bang? Include any constituent that is believed to have made up more than 1% of the mass density of the universe.

(Weinberg, Chapter 1, Page 5)

Ans: Electrons, positrons, neutrinos, and photons.

PROBLEM 15: SPECIAL RELATIVITY DOPPLER SHIFT *(20 points)*

(a) The easiest way to solve this problem is by a double application of the standard special-relativity Doppler shift formula, which was given on the front of the exam:

$$z = \sqrt{\frac{1+\beta}{1-\beta}} - 1 \ , \tag{18.1}$$

where $\beta = v/c$. Remembering that the wavelength is stretched by a factor $1 + z$, we find immediately that the wavelength of the radio wave received at Alpha-7 is given by

$$\lambda_{\text{Alpha}-7} = \sqrt{\frac{1+v_s/c}{1-v_s/c}}\,\lambda_{\text{emitted}} \ . \tag{18.2}$$

The photons that are received by the observer are in fact never received by Alpha-7, but the wavelength found by the observer will be the same as if Alpha-7 acted as a relay station, receiving the photons and retransmitting them at the received wavelength. So, applying Eq. (18.1) again, the wavelength seen by the observer can be written as

$$\lambda_{\text{observed}} = \sqrt{\frac{1+v_o/c}{1-v_o/c}}\,\lambda_{\text{Alpha}-7} \ . \tag{18.3}$$

Combining Eqs. (18.2) and (18.3),

$$\lambda_{\text{observed}} = \sqrt{\frac{1+v_o/c}{1-v_o/c}}\sqrt{\frac{1+v_s/c}{1-v_s/c}}\,\lambda_{\text{emitted}} \ , \tag{18.4}$$

so finally

$$\boxed{z = \sqrt{\frac{1+v_o/c}{1-v_o/c}}\sqrt{\frac{1+v_s/c}{1-v_s/c}} - 1 \ .} \tag{18.5}$$

(b) Although we used the presence of Alpha-7 in determining the redshift z of Eq. (18.5), the redshift is not actually affected by the space station. So the special-relativity

Doppler shift formula, Eq. (18.1), must directly describe the redshift resulting from the relative motion of the source and the observer. Thus

$$\sqrt{\frac{1+v_{\rm tot}/c}{1-v_{\rm tot}/c}}-1=\sqrt{\frac{1+v_o/c}{1-v_o/c}}\sqrt{\frac{1+v_s/c}{1-v_s/c}}-1\ . \tag{18.6}$$

The equation above determines $v_{\rm tot}$ in terms of v_o and v_s, so the rest is just algebra. To simplify the notation, let $\beta_{\rm tot}\equiv v_{\rm tot}/c$, $\beta_o\equiv v_o/c$, and $\beta_s\equiv v_s/c$. Then

$$1+\beta_{\rm tot}=\frac{1+\beta_o}{1-\beta_o}\,\frac{1+\beta_s}{1-\beta_s}\,(1-\beta_{\rm tot})$$

$$\beta_{\rm tot}\left[1+\frac{1+\beta_o}{1-\beta_o}\,\frac{1+\beta_s}{1-\beta_s}\right]=\frac{1+\beta_o}{1-\beta_o}\,\frac{1+\beta_s}{1-\beta_s}-1$$

$$\beta_{\rm tot}\left[\frac{(1-\beta_o-\beta_s+\beta_o\beta_s)+(1+\beta_o+\beta_s+\beta_o\beta_s)}{(1-\beta_o)(1-\beta_s)}\right]=$$

$$\frac{(1+\beta_o+\beta_s+\beta_o\beta_s)-(1-\beta_o-\beta_s+\beta_o\beta_s)}{(1-\beta_o)(1-\beta_s)}$$

$$\beta_{\rm tot}[2(1+\beta_o\beta_s)]=2(\beta_o+\beta_s)$$

$$\beta_{\rm tot}=\frac{\beta_o+\beta_s}{1+\beta_o\beta_s}$$

$$\boxed{v_{\rm tot}=\frac{v_o+v_s}{1+\dfrac{v_ov_s}{c^2}}\ .} \tag{18.7}$$

The final formula is the relativistic expression for the addition of velocities. Note that it guarantees that $|v_{\rm tot}|\le c$ as long as $|v_o|\le c$ and $|v_s|\le c$.

PROBLEM 16: DID YOU DO THE READING (2005)? *(25 points)*

(a) *(4 points)* What was the first external galaxy that was shown to be at a distance significantly greater than the most distant known objects in our galaxy? How was the distance estimated?

Ans: (Weinberg, page 20) The first galaxy shown to be at a distance beyond the size of our galaxy was Andromeda, also known by its Messier number, M31. It is the nearest spiral galaxy to our galaxy. The distance was determined (by Hubble) using Cepheid variable stars, for which the absolute luminosity is proportional to the period. A measurement of a particular Cepheid's period determines the star's absolute luminosity, which, compared to the measured luminosity, determines the distance to the star. (Hubble's initial measurement of the distance to Andromeda used a badly-calibrated version of this period-luminosity relationship and consequently underestimated the distance by more than a factor of two; nonetheless, the initial measurement still showed that the Andromeda Nebula was an order of magnitude more distant than the most distant known objects in our own galaxy.)

(b) *(5 points)* What is recombination? Did galaxies begin to form before or after recombination? Why?

Ans: (Weinberg, pages 64 and 73) Recombination refers to the formation of neutral atoms out of charged nuclei and electrons. Galaxies began to form after recombination. Prior to recombination, the strong electromagnetic interactions between photons and matter produced a high pressure which effectively counteracted the gravitational attraction between particles. Once the universe became transparent to radiation, the matter no longer interacted significantly with the photons and consequently began to undergo gravitational collapse into large clumps.

(c) *(4 points)* In Chapter IV of his book, Weinberg develops a "recipe for a hot universe," in which the matter of the universe is described as a gas in thermal equilbrium at a very high temperature, in the vicinity of 10^9 K (several thousand million degrees Kelvin). Such a thermal equilibrium gas is completely described by specifying its temperature and the density of the conserved quantities. Which of the following is on this list of conserved quantities? Circle as many as apply.

(i) baryon number (ii) energy per particle (iii) proton number

(iv) electric charge (v) pressure

Ans: (Weinberg, page 91) The correct answers are (i) and (iv). A third conserved quantity, lepton number, was not included in the multiple-choice options.

(d) *(4 points)* The wavelength corresponding to the mean energy of a CMB (cosmic microwave background) photon today is approximately equal to which of the following quantities? (You may wish to look up the values of various physical constants at the end of the quiz.)

(i) 2 fm (2×10^{-15} m)

(ii) 2 microns (2×10^{-6} m)

(iii) 2 mm (2×10^{-3} m)

(iv) 2 m.

Ans: (Ryden, page 23) The correct answer is (iii).

If you did not remember this number, you could estimate the answer by remembering that the characteristic temperature of the cosmic microwave background is approximately 3 Kelvin. The typical photon energy is then on the order of kT, from which we can find the frequency as $E = h\nu$. The wavelength of the photon is then

$\lambda = \nu/c$. This approximation gives $\lambda = 5.3$ mm, which is not equal to the correct answer, but it is much closer to the correct answer than to any of the other choices.

(e) *(4 points)* What is the equivalence principle?

Ans: (Ryden, page 27) In its simplest form, the equivalence principle says that the gravitational mass of an object is identical to its inertial mass. This equality implies the equivalent statement that it is impossible to distinguish (without additional information) between an observer in a reference frame accelerating with acceleration $\vec{a}$ and an observer in an inertial reference frame subject to a gravitational force $-m_{obs}\vec{a}$.

(Actually, what the equivalence principle really says is that the ratio of the gravitational to inertial masses m_g/m_i is universal, that is, independent of the material properties of the object in question. The ratio does not necessarily need to be 1. However, once we know that the two types of masses are proportional, we can simply define the gravitational coupling G to make them equal. To see this, consider a theory of gravity where $m_g/m_i = q$. Then the gravitational force law is

$$m_i a = -\frac{GMm_g}{r^2} ,$$

or

$$a = -\frac{GqM}{r^2} .$$

At this point, if we define $G' = Gq$, we have a gravitational theory with gravitational coupling G' and inertial mass equal to gravitational mass.)

(f) *(4 points)* Why is it difficult for Earth-based experiments to look at the small wavelength portion of the graph of CMB energy density per wavelength vs. wavelength?

Ans: (Weinberg, page 67) The Earth's atmosphere is increasingly opaque for wavelength shorter than .3 cm. Therefore, radiation at these wavelengths will be absorbed and rescattered by the Earth's atmosphere; observations of the cosmic microwave background at small wavelengths must be performed above the Earth's atmosphere.

PROBLEM 17: TRACING A LIGHT PULSE THROUGH A RADIATION-DOMINATED UNIVERSE

(a) The physical horizon distance is given in general by

$$\ell_{p,\text{horizon}} = a(t)\int_0^{t_f} \frac{c}{a(t)}\,dt ,$$

so in this case

$$\ell_{p,\text{horizon}} = bt^{1/2}\int_0^{t_f} \frac{c}{bt^{1/2}}\,dt = \boxed{2ct_f\ .}$$

(b) If the source is at the horizon distance, it means that a photon leaving the source at $t = 0$ would just be reaching the origin at t_f. So, $\boxed{t_e = 0\ .}$

(c) The coordinate distance between the source and the origin is the coordinate horizon distance, given by

$$\ell_{c,\text{horizon}} = \int_0^{t_f} \frac{c}{bt^{1/2}}\, dt = \boxed{\frac{2ct_f^{1/2}}{b}\ .}$$

(d) The photon starts at coordinate distance $2c\sqrt{t_f}/b$, and by time t it will have traveled a coordinate distance

$$\int_0^t \frac{c}{bt'^{1/2}}\, dt' = \frac{2c\sqrt{t}}{b}$$

toward the origin. Thus the photon will be at coordinate distance

$$\ell_c = \frac{2c}{b}\left(\sqrt{t_f} - \sqrt{t}\right)$$

from the origin, and hence a physical distance

$$\ell_p(t) = a(t)\ell_c = \boxed{2c\left(\sqrt{t\,t_f} - t\right)\ .}$$

(e) To find the maximum of $\ell_p(t)$, we differentiate it and set the derivative to zero:

$$\frac{d\ell_p}{dt} = \left(\sqrt{\frac{t_f}{t}} - 2\right) c\ ,$$

so the maximum occurs when

$$\sqrt{\frac{t_f}{t_{\text{max}}}} = 2\ ,$$

or

$$\boxed{t_{\text{max}} = \frac{1}{4} t_f\ .}$$

PROBLEM 18: TRANSVERSE DOPPLER SHIFTS

(a) Describing the events in the coordinate system shown, the Xanthu is at rest, so its clocks run at the same speed as the coordinate system time variable, t. The emission of the wavecrests of the radio signal are therefore separated by a time interval equal to the time interval as measured by the source, the Xanthu:

$$\Delta t = \Delta t_s \ .$$

Since the Emmerac is moving perpendicular to the path of the radio waves, at the moment of reception its distance from the Xanthu is at a minimum, and hence its rate of change is zero. Hence successive wavecrests will travel the same distance, as long as $c\Delta t \ll a$. Since the wavecrests travel the same distance, the time separation of their arrival at the Emmerac is Δt, the same as the time separation of their emission. The clocks on the Emmerac, however, and running slowly by a factor of

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \ .$$

The time interval between wave crests as measured by the receiver, on the Emmerac, is therefore smaller by a factor of γ,

$$\Delta t_r = \frac{\Delta t_s}{\gamma} \ .$$

Thus, there is a $\boxed{\text{blueshift.}}$ The redshift parameter z is defined by

$$\frac{\Delta t_r}{\Delta t_s} = 1 + z \ ,$$

so

$$\frac{1}{\gamma} = 1 + z \ ,$$

or

$$\boxed{z = \frac{1-\gamma}{\gamma} \ .}$$

Recall that $\gamma > 1$, so z is negative.

(b) Describing this situation in the coordinate system shown, this time the source on the Xanthu is moving, so the clocks at the source are running slowly. The time between wavecrests, measured in coordinate time t, is therefore larger by a factor of γ than Δt_s, the time as measured by the clock on the source:

$$\Delta t = \gamma \, \Delta t_s \ .$$

Since the radio signal is emitted when the Xanthu is at its minimum separation from the Emmerac, the rate of change of the separation is zero, so each wavecrest travels the same distance (again assuming that $c\Delta t \ll a$). Since the Emmerac is at rest, its clocks run at the same speed as the coordinate time t, and hence the time interval between crests, as measured by the receiver, is

$$\Delta t_r = \Delta t = \gamma\, \Delta t_s \ .$$

Thus the time interval as measured by the receiver is longer than that measured by the source, and hence it is a $\boxed{\text{redshift.}}$ The redshift parameter z is given by

$$1 + z = \frac{\Delta t_r}{\Delta t_s} = \gamma \ ,$$

so

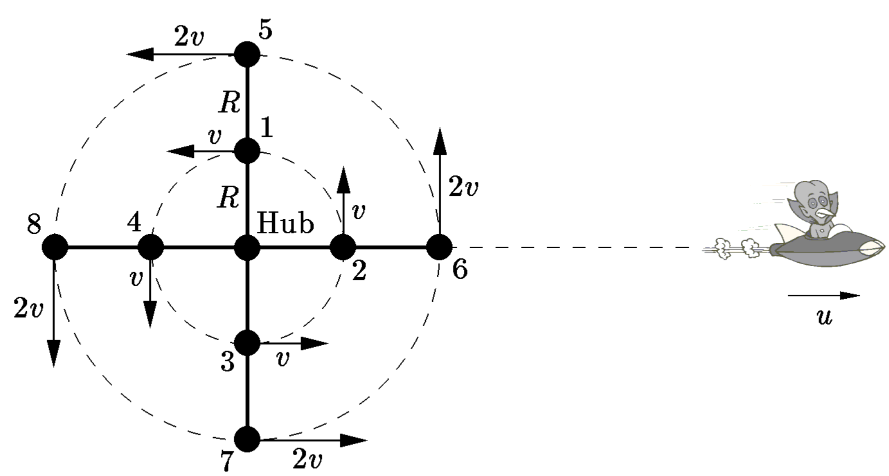
$$\boxed{z = \gamma - 1 \ .}$$

(c) The events described in (a) can be made to look a lot like the events described in (b) by transforming to a frame of reference that is moving to the right at speed v_0 — i.e., by transforming to the rest frame of the Emmerac. In this frame the Emmerac is of course at rest, and the Xanthu is traveling on the trajectory

$$(x{=}-v_0 t, y{=}a, z{=}0) \ ,$$

as in part (b). However, just as the transformation causes the x-component of the velocity of the Xanthu to change from zero to a negative value, so the x-component of the velocity of the radio signal will be transformed from zero to a negative value. Thus in this frame the radio signal will not be traveling along the y-axis, so the events will not match those described in (b). The situations described in (a) and (b) are therefore physically distinct (which they must be if the redshifts are different, as we calculated above).

PROBLEM 19: A TWO-LEVEL HIGH-SPEED MERRY-GO-ROUND *(15 points)*



(a) Since the relative positions of all the cars remain fixed as the merry-go-round rotates, each successive pulse from any given car to any other car takes the same amount of time to complete its trip. Thus there will be no Doppler shift caused by pulses taking different amounts of time; the only Doppler shift will come from time dilation.

We will describe the events from the point of view of an inertial reference frame at rest relative to the hub of the merry-go-round, which we will call the laboratory frame. This is the frame in which the problem is described, in which the inner cars are moving at speed v, and the outer cars are moving at speed $2v$. In the laboratory frame, the time interval between the wave crests emitted by the source $\Delta t_S^{\rm Lab}$ will be exactly equal to the time interval $\Delta t_O^{\rm Lab}$ between two crests reaching the observer:

$$\Delta t_O^{\rm Lab} = \Delta t_S^{\rm Lab} \ .$$

The clocks on the merry-go-round cars are moving relative to the laboratory frame, so they will appear to be running slowly by the factor

$$\gamma_1 = \frac{1}{\sqrt{1 - v^2/c^2}}$$

for the inner cars, and by the factor

$$\gamma_2 = \frac{1}{\sqrt{1 - 4v^2/c^2}}$$

for the outer cars. Thus, if we let Δt_S denote the time between crests as measured by a clock on the source, and Δt_O as the time between crests as measured by a clock moving with the observer, then these quantities are related to the laboratory frame times by

$$\gamma_2 \Delta t_S = \Delta t_S^{\rm Lab} \quad \text{and} \quad \gamma_1 \Delta t_O = \Delta t_O^{\rm Lab} \ .$$

To make sure that the γ-factors are on the right side of the equation, you should keep in mind that any time interval should be measured as shorter on the moving clocks than on the lab clocks, since these clocks appear to run slowly. Putting together the equations above, one has immediately that

$$\Delta t_O = \frac{\gamma_2}{\gamma_1} \Delta t_S \ .$$

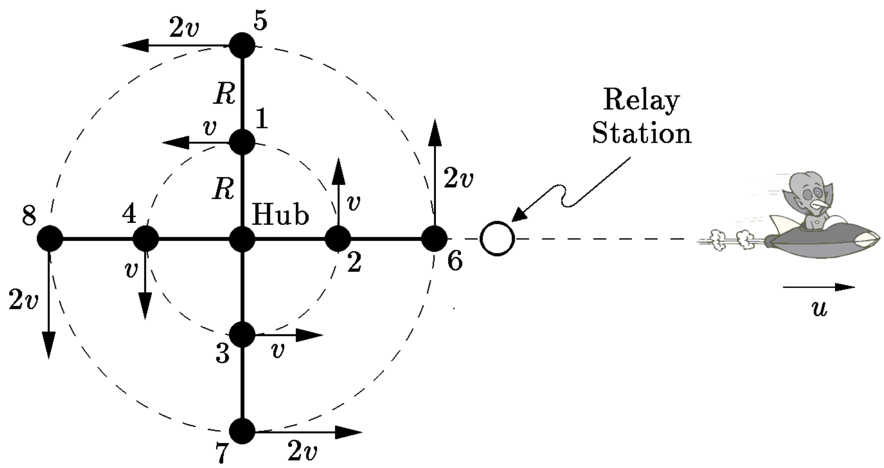
The redshift z is defined by

$$\Delta t_O \equiv (1 + z)\, \Delta t_S \ ,$$

so

$$\boxed{z = \frac{\gamma_2}{\gamma_1} - 1 = \sqrt{\frac{1 - \frac{v^2}{c^2}}{1 - \frac{4v^2}{c^2}}} - 1 \ .}$$

(b) For this part of the problem is useful to imagine a relay station located just to the right of car 6 in the diagram, at rest in the laboratory frame. The relay station rebroadcasts the waves as it receives them, and hence has no effect on the frequency received by the observer, but serves the purpose of allowing us to clearly separate the problem into two parts.



The first part of the discussion concerns the redshift of the signal as measured by the relay station. This calculation would involve both the time dilation and a change in path lengths between successive pulses, but we do not need to do it. It is the standard situation of a source and observer moving directly away from each other, as discussed at the end of Lecture Notes 1. The Doppler shift is given by Eq. (1.33), which was included in the formula sheet. Writing the formula for a recession speed u, it becomes

$$(1+z)|_{\text{relay}} = \sqrt{\frac{1+\frac{u}{c}}{1-\frac{u}{c}}} \ .$$

If we again use the symbol Δt_S for the time between wave crests as measured by a clock on the source, then the time between the receipt of wave crests as measured by the relay station is

$$\Delta t_R = \sqrt{\frac{1+\frac{u}{c}}{1-\frac{u}{c}}} \ \Delta t_S \ .$$

The second part of the discussion concerns the transmission from the relay station to car 6. The velocity of car 6 is perpendicular to the direction from which the pulse is being received, so this is a transverse Doppler shift. Any change in path length between successive pulses is second order in Δt, so it can be ignored. The only effect is therefore the time dilation. As described in the laboratory frame, the time separation between crests reaching the observer is the same as the time separation measured by the relay station:

$$\Delta t_O^{\mathrm{Lab}} = \Delta t_R \ .$$

As in part (a), the time dilation implies that

$$\gamma_2 \Delta t_O = \Delta t_O^{\mathrm{Lab}} \ .$$

Combining the formulas above,

$$\Delta_O = \frac{1}{\gamma_2} \sqrt{\frac{1+\frac{u}{c}}{1-\frac{u}{c}}}\ \Delta t_S \ .$$

Again $\Delta t_O \equiv (1+z)\,\Delta t_S$, so

$$\boxed{z = \frac{1}{\gamma_2}\sqrt{\frac{1+\frac{u}{c}}{1-\frac{u}{c}}} - 1 = \sqrt{\frac{\left(1-\frac{4v^2}{c^2}\right)\left(1+\frac{u}{c}\right)}{1-\frac{u}{c}}} - 1 \ .}$$

PROBLEM 20: SIGNAL PROPAGATION IN A FLAT MATTER-DOMINATED UNIVERSE *(55 points)*

(a)-(i) If we let $\ell_c(t)$ denote the coordinate distance of the light signal from A, then we can make use of Eq. (3.8) from the lecture notes for the coordinate velocity of light:

$$\frac{\mathrm{d}\ell_c}{\mathrm{d}t} = \frac{c}{a(t)} \ . \tag{20.1}$$

Integrating the velocity,

$$\begin{aligned} \ell_c(t) &= \int_{t_1}^{t} \frac{c\,\mathrm{d}t'}{a(t')} = \frac{c}{b}\int_{t_1}^{t} \frac{\mathrm{d}t'}{t'^{2/3}} \\ &= \frac{3c}{b}\left[t^{1/3} - t_1^{1/3}\right] \ . \end{aligned} \tag{20.2}$$

The physical distance is then

$$\begin{aligned}\ell_{p,sA}(t) = a(t)\ell_c(t) &= bt^{2/3}\frac{3c}{b}\left[t^{1/3} - t_1^{1/3}\right] \\ &= 3c\left(t - t^{2/3}t_1^{1/3}\right) \\ &= 3ct\left[1 - \left(\frac{t_1}{t}\right)^{1/3}\right] .\end{aligned} \qquad (20.3)$$

We now need to differentiate, which is done most easily with the middle line of the above equation:

$$\boxed{\frac{\mathrm{d}\ell_{p,sA}}{\mathrm{d}t} = c\left[3 - 2\left(\frac{t_1}{t}\right)^{1/3}\right] .} \qquad (20.4)$$

(ii) At $t = t_1$, the time of emission, the above formula gives

$$\boxed{\frac{\mathrm{d}\ell_{p,sA}}{\mathrm{d}t} = c .} \qquad (20.5)$$

This is what should be expected, since the speed of separation of the light signal at the time of emission is really just a local measurement of the speed of light, which should always give the standard value c.

(iii) At arbitrarily late times, the second term in brackets in Eq. (20.4) becomes negligible, so

$$\boxed{\frac{\mathrm{d}\ell_{p,sA}}{\mathrm{d}t} \to 3c .} \qquad (20.6)$$

Although this answer is larger than c, it does not violate relativity. Once the signal is far from its origin it is carried by the expansion of the universe, and relativity places no speed limit on the expansion of the universe.

(b) This part of the problem involves $H(t_1)$, so we can start by evaluating it:

$$H(t) = \frac{\dot{a}(t)}{a(t)} = \frac{\frac{\mathrm{d}}{\mathrm{d}t}(bt^{2/3})}{bt^{2/3}} = \frac{2}{3t} . \qquad (20.7)$$

Thus, the physical distance from A to B at time t_1 is

$$\ell_{p,BA} = \frac{3}{2}ct_1 . \qquad (20.8)$$

The coordinate distance is the physical distance divided by the scale factor, so

$$\ell_{c,BA} = \frac{cH^{-1}(t_1)}{a(t_1)} = \frac{\frac{3}{2}ct_1}{bt_1^{2/3}} = \frac{3c}{2b}\,t_1^{1/3} \; . \tag{20.9}$$

Since light travels at a coordinate speed $c/a(t)$, the light signal will reach galaxy B at time t_2 if

$$\begin{aligned}\ell_{c,BA} &= \int_{t_1}^{t_2} \frac{c}{bt'^{2/3}}\,\mathrm{d}t' \\ &= \frac{3c}{b}\left[t_2^{1/3} - t_1^{1/3}\right] \; .\end{aligned} \tag{20.10}$$

Setting the expressions (20.9) and (20.10) for $\ell_{c,BA}$ equal to each other, one finds

$$\frac{1}{2}t_1^{1/3} = t_2^{1/3} - t_1^{1/3} \quad \Longrightarrow \quad t_2^{1/3} = \frac{3}{2}t_1^{1/3} \quad \Longrightarrow \quad \boxed{t_2 = \frac{27}{8}t_1 \; .} \tag{20.11}$$

(c)-(i) Physical distances are additive, so if one adds the distance from A and the light signal to the distance from the light signal to B, one gets the distance from A to B:

$$\ell_{p,sA} + \ell_{p,sB} = \ell_{p,BA} \; . \tag{20.12}$$

But $\ell_{p,BA}(t)$ is just the scale factor times the coordinate separation, $a(t)\ell_{c,BA}$. Using the previous relations (20.3) and (20.9) for $\ell_{p,sA}(t)$ and $\ell_{c,BA}$, we find

$$3ct\left[1 - \left(\frac{t_1}{t}\right)^{1/3}\right] + \ell_{p,sB}(t) = \frac{3}{2}ct_1^{1/3}t^{2/3} \; , \tag{20.13}$$

so

$$\ell_{p,sB}(t) = \frac{9}{2}ct_1^{1/3}t^{2/3} - 3ct = 3ct\left[\frac{3}{2}\left(\frac{t_1}{t}\right)^{1/3} - 1\right] \; . \tag{20.14}$$

As a check, one can verify that this expression vanishes for $t = t_2 = (27/8)\,t_1$, and that it equals $(3/2)ct_1$ at $t = t_1$. But we are asked to find the speed of approach, the negative of the derivative of Eq. (20.14):

$$\begin{aligned}\text{Speed of approach} &= -\frac{\mathrm{d}\ell_{p,sB}}{\mathrm{d}t} \\ &= -3ct_1^{1/3}t^{-1/3} + 3c \\ &= \boxed{3c\left[1 - \left(\frac{t_1}{t}\right)^{1/3}\right] .}\end{aligned} \tag{20.15}$$

(ii) At the time of emission, $t = t_1$, Eq. (20.15) gives

$$\boxed{\text{Speed of approach} = 0\ .} \tag{20.16}$$

This makes sense, since at $t = t_1$ galaxy B is one Hubble length from galaxy A, which means that its recession velocity is exactly c. The recession velocity of the light signal leaving A is also c, so the rate of change of the distance from the light signal to B is initially zero.

(iii) At the time of reception, $t = t_2 = (27/8)\,t_1$, Eq. (20.15) gives

$$\boxed{\text{Speed of approach} = c\ ,} \tag{20.17}$$

which is exactly what is expected. As in part (a)-(ii), this is a local measurement of the speed of light.

(d) To find the redshift, we first find the time t_{BA} at which a light pulse must be emitted from galaxy B so that it arrives at galaxy A at time t_1. Using the coordinate distance given by Eq. (20.9), the time of emission must satisfy

$$\frac{3c}{2b}\,t_1^{1/3} = \int_{t_{BA}}^{t_1} \frac{c}{bt'^{2/3}}\,\mathrm{d}t' = \frac{3c}{b}\left(t_1^{1/3} - t_{BA}^{1/3}\right)\ , \tag{20.18}$$

which can be solved to give

$$t_{BA} = \frac{1}{8}\,t_1\ . \tag{20.19}$$

The redshift is given by

$$1 + z_{BA} = \frac{a(t_1)}{a(t_{BA})} = \left(\frac{t_1}{t_{BA}}\right)^{2/3} = 4\ . \tag{20.20}$$

Thus,

$$\boxed{z_{BA} = 3\ .} \tag{20.21}$$

(e) Applying Euclidean geometry to the triangle C-A-B shows that the physical distance from C to B, at time t_1, is $\sqrt{2}cH^{-1}$. The coordinate distance is also larger than the A-B separation by a factor of $\sqrt{2}$. Thus,

$$\ell_{c,BC} = \frac{3\sqrt{2}c}{2b}\,t_1^{1/3}\ . \tag{20.22}$$

If we let t_{BC} be the time at which a light pulse must be emitted from galaxy B so that it arrives at galaxy C at time t_1, we find

$$\frac{3\sqrt{2}c}{2b}\,t_1^{1/3} = \int_{t_{BC}}^{t_1} \frac{c}{bt'^{2/3}}\,\mathrm{d}t' = \frac{3c}{b}\left(t_1^{1/3} - t_{BC}^{1/3}\right) \;, \tag{20.23}$$

which can be solved to find

$$t_{BC} = \left(1 - \frac{\sqrt{2}}{2}\right)^3 t_1 \;. \tag{20.24}$$

Then

$$1 + z_{BC} = \frac{a(t_1)}{a(t_{BC})} = \left(\frac{t_1}{t_{BC}}\right)^{2/3} = \frac{1}{\left(1 - \frac{\sqrt{2}}{2}\right)^2} \;, \tag{20.25}$$

and

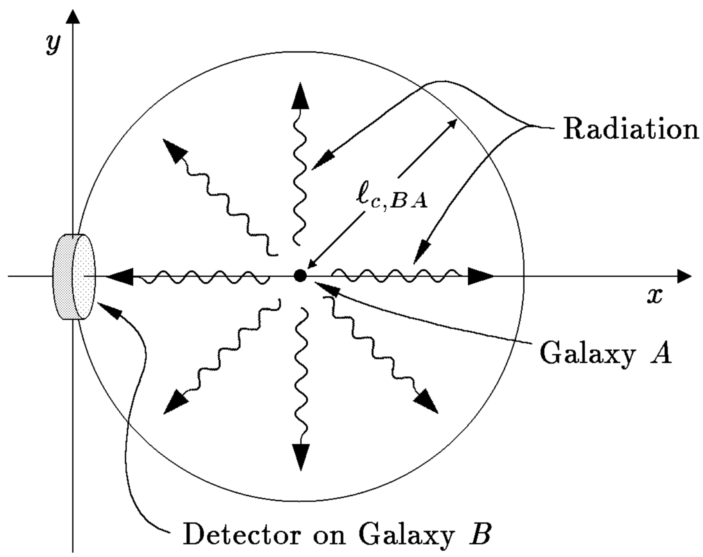
$$\boxed{z_{BC} = \frac{1}{\left(1 - \frac{\sqrt{2}}{2}\right)^2} - 1 \;.} \tag{20.26}$$

Full credit will be given for the answer in the form above, but it can be simplified by rationalizing the fraction:

$$\begin{aligned} z_{BC} &= \frac{1}{\left(1 - \frac{\sqrt{2}}{2}\right)^2} \frac{\left(1 + \frac{\sqrt{2}}{2}\right)^2}{\left(1 + \frac{\sqrt{2}}{2}\right)^2} - 1 \\ &= \frac{1 + \sqrt{2} + \frac{1}{2}}{\frac{1}{4}} - 1 \\ &= \boxed{5 + 4\sqrt{2}\;.} \end{aligned} \tag{20.27}$$

Numerically, $z_{BC} = 10.657$.

(f) Following the solution to Problem 6 of Problem Set 2, we draw a diagram in comoving coordinates, putting the source at the center of a sphere:



The energy from galaxy A will radiate uniformly over the sphere. If the detector has physical area A_D, then in the comoving coordinate picture it has coordinate area $A_D/a^2(t_2)$, since the detection occurs at time t_2 The full coordinate area of the sphere is $4\pi\ell_{c,BA}^2$, so the fraction of photons that hit the detector is

$$\text{fraction} = \frac{\left[A/a(t_2)^2\right]}{4\pi\ell_{c,BA}^2} \ . \tag{20.28}$$

As in Problem 6, the power hitting the detector is reduced by two factors of $(1+z)$: one factor because the energy of each photon is proportional to the frequency, and hence is reduced by the redshift, and one more factor because the rate of arrival of photons is also reduced by the redshift factor $(1+z)$. Thus,

$$\begin{aligned}\text{Power hitting detector} &= P\frac{\left[A/a(t_2)^2\right]}{4\pi\ell_{c,BA}^2}\frac{1}{(1+z)^2} \\ &= P\frac{\left[A/a(t_2)^2\right]}{4\pi\ell_{c,BA}^2}\left[\frac{a(t_1)}{a(t_2)}\right]^2 \\ &= P\frac{A}{4\pi\ell_{c,BA}^2}\frac{a^2(t_1)}{a^4(t_2)} \ .\end{aligned} \tag{20.29}$$

The energy flux is given by

$$J = \frac{\text{Power hitting detector}}{A} \ , \tag{20.30}$$

so

$$J = \frac{P}{4\pi\ell_{c,BA}^2}\frac{a^2(t_1)}{a^4(t_2)} \ . \tag{20.31}$$

From here it is just algebra, using Eqs. (20.9) and (20.11), and $a(t) = bt^{2/3}$:

$$
\begin{aligned}
J &= \frac{P}{4\pi\left[\frac{3c}{2b}t_1^{1/3}\right]^2}\frac{b^2 t_1^{4/3}}{b^4 t_2^{8/3}} \\
&= \frac{P}{4\pi\left[\frac{3c}{2b}t_1^{1/3}\right]^2}\frac{b^2 t_1^{4/3}}{\left(\frac{27}{8}\right)^{8/3} b^4 t_1^{8/3}} \\
&= \frac{P}{4\pi\left[\frac{3c}{2}t_1^{1/3}\right]^2}\frac{t_1^{4/3}}{\left(\frac{3}{2}\right)^{8} t_1^{8/3}} \qquad (20.32)\\
&= \boxed{\frac{2^8}{3^{10}\pi}\frac{P}{c^2 t_1^2}} \\
&= \boxed{\frac{256}{59{,}049\pi}\frac{P}{c^2 t_1^2}\ .}
\end{aligned}
$$

It is debatable which of the last two expressions is the simplest, so I have boxed both of them. One could also write

$$J = 1.380 \times 10^{-3}\frac{P}{c^2 t_1^2}\ . \qquad (20.33)$$

PROBLEM 21: DID YOU DO THE READING (2011)? *(25 points)*†

†Solution written by Daniele Bertolini.

(a) *(10 points)* To determine the distance of the galaxies he was observing Hubble used so called *standard candles*. Standard candles are astronomical objects whose intrinsic luminosity is known and whose distance is inferred by measuring their apparent luminosity. First, he used as standard candles variable stars, whose intrinsic luminosity can be related to the period of variation. Quoting Weinberg's *The First Three Minutes*, chapter 2, pages 19-20:

In 1923 Edwin Hubble was for the first time able to resolve the Andromeda Nebula into separate stars. He found that its spiral arms included a few bright variable stars, with the same sort of periodic variation of luminosity as was already familiar for a class of stars in our galaxy known as Cepheid variables. The reason this was so important was that in the preceding decade the work of Henrietta Swan Leavitt *and Harlow Shapley of the Harvard College Observatory had provided a tight relation between the observed periods of variation of the Cepheids and their absolute luminosities. (Absolute luminosity is the total radiant power emitted by an astronomical object in all directions. Apparent luminosity is the radiant power received by us in each square centimeter of our telescope mirror. It is the apparent rather than the absolute luminosity that determines the subjective degree of brightness of astronomical objects. Of course, the apparent luminosity depends not only on the absolute luminosity, but also on the distance; thus, knowing both the absolute and the apparent luminosities of an astronomical body, we can infer its distance.) Hubble, observing the apparent luminosity of the Cepheids in the Andromeda Nebula, and estimating their absolute luminosity from their periods, could immediately calculate their distance, and hence the distance of the Andromeda Nebula, using the simple rule that apparent luminosity is proportional to the absolute luminosity and inversely proportional to the square of the distance.*

He also used particularly bright stars as standard candles, as we deduce from page 25:

Returning now to 1929: Hubble estimated the distance to 18 galaxies from the apparent luminosity of their brighest stars, and compared these distances with the galaxies' respective velocities, determined spectroscopically from their Doppler shifts.

Note: since from reading just the first part of Weinberg's discussion one could be induced to think that Hubble used just Cepheids as standard candles, students who mentioned only Cepheids got 9 points out of 10. In fact, however, Hubble was able to identify Cepheid variables in only a few galaxies. The Cepheids were crucial, because they served as a calibration for the larger distances, but they were not in themselves sufficient.

(b) *(5 points)* Quoting Weinberg's *The First Three Minutes*, chapter 2, page 21:

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 directions they look. (Here, and below, I 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.*

So the Cosmological principle basically states that the universe appears as homogeneous and isotropic (on scales of distance large enough) to any typical observer, where typical is referred to observers with small local motion compared to the expansion flow. Ryden gives a more general definition of Cosmological Principle, which is valid as well. Quoting Ryden's *Introduction to Cosmology*, chapter 2, page 11 or 14 (depending on which version):

However, modern cosmologists have adopted the **cosmological principle,** *which states: There is nothing special about our location in the universe. The cosmological principle holds true only on large scales (of 100 Mpc or more).*

(c) *(10 points)* Quoting again Ryden's *Introduction to Cosmology*, chapter 2, page 9 or 11:

Saying that the universe is **isotropic** *means that there are no preferred directions in the universe; it looks the same no matter which way you point your telescope. Saying that the universe is* **homogeneous** *means that there are no preferred locations in the universe; it looks the same no matter where you set up your telescope.*

(i) $\boxed{\text{False.}}$ If the universe is isotropic around one point it does not need to be homogeneous. A counter-example is a distribution of matter with spherical symmetry, that is, with a density which is only a function of the radius but does not depend on the direction: $\rho(r, \theta, \phi) \equiv \rho(r)$. In this case for an observer at the center of the distribution the universe looks isotropic but it is not homogeneous.

(ii) $\boxed{\text{True.}}$ For the case of Euclidean geometry isotropy around two or more distinct points does imply homogeneity. Weinberg shows this in chapter 2, page 24. Consider two observers, and two arbitrary points A and B which we would like to prove equivalent. Consider a circle through point A, centered on observer 1, and another circle through point B, centered on observer 2. If C is a point on the intersection of the two circles, then isotropy about the two observers implies that $A = C$ and $B = C$, and hence $A = B$. (This argument was good enough for Weinberg and hence good enough to deserve full credit, but it is actually incomplete: one can find points A and B for which the two circles will not intersect. On your next problem set you will have a chance to invent a better proof.)

(d) *(2 points extra credit)* $\boxed{\text{False.}}$ If we relax the hypothesis of Euclidean geometry, then isotropy around two points does not necessarily imply homogeneity. A counter-example we mentioned in class is a two-dimensional universe consisting of the surface of a sphere. Think of the sphere in three Euclidean dimensions, but the model "universe" consists only of its two-dimensional surface. Imagine latitude and longitude lines to give coordinates to the surface, and imagine a matter distribution that depends only on latitude. This would not be homogeneous, but it would look isotropic to observers at both the north and south poles. While this example describes a two-dimensional universe, which therefore cannot be our universe, we will learn shortly how to construct a three-dimensional non-Euclidean universe with these same properties.

PROBLEM 22: THE TRAJECTORY OF A PHOTON ORIGINATING AT THE HORIZON *(25 points)*

(a) They key idea is that the coordinate speed of light is given by

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{c}{a(t)} ,$$

so the coordinate distance (in notches) that light can travel between $t = 0$ and now ($t = t_0$) is given by

$$\ell_c = \int_0^{t_0} \frac{c\,\mathrm{d}t}{a(t)} .$$

The corresponding physical distance is the horizon distance:

$$\ell_{p,\text{horizon}}(t_0) = a(t_0) \int_0^{t_0} \frac{c\,\mathrm{d}t}{a(t)} .$$

Evaluating,

$$\ell_{p,\text{horizon}}(t_0) = bt_0^{2/3} \int_0^{t_0} \frac{c\,\mathrm{d}t}{bt^{2/3}} = t_0^{2/3}\left[3ct_0^{1/3}\right] = \boxed{3ct_0\ .}$$

(b) As stated in part (a), the coordinate distance that light can travel between $t = 0$ and $t = t_0$ is given by

$$\ell_c = \int_0^{t_0} \frac{c\,\mathrm{d}t}{a(t)} = \frac{3ct_0^{1/3}}{b} .$$

Thus, if we are at the origin, at $t = 0$ the photon must have been at

$$\boxed{x_0 = \frac{3ct_0^{1/3}}{b}\ .}$$

(c) The photon starts at $x = x_0$ at $t = 0$, and then travels in the negative x-direction at speed $c/a(t)$. Thus, it's position at time t is given by

$$x(t) = x_0 - \int_0^t \frac{c\,\mathrm{d}t'}{a(t')} = \frac{3ct_0^{1/3}}{b} - \frac{3ct^{1/3}}{b} = \boxed{\frac{3c}{b}\left(t_0^{1/3} - t^{1/3}\right)\ .}$$

(d) Since the coordinate distance between us and the photon is $x(t)$, measured in notches, the physical distance (in, for example, meters) is just $a(t)$ times $x(t)$. Thus.

$$\ell_p(t) = a(t)x(t) = \boxed{3ct^{2/3}\left(t_0^{1/3} - t^{1/3}\right)\ .}$$

(e) To find the maximum of $\ell_p(t)$, we set the derivative equal to zero:

$$\frac{\mathrm{d}\ell_p(t)}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left[3c\left(t^{2/3}t_0^{1/3} - t\right)\right] = 3c\left[\frac{2}{3}\left(\frac{t_0}{t}\right)^{1/3} - 1\right] = 0\ ,$$

so

$$\left(\frac{t_0}{t_{\rm max}}\right)^{1/3} = \frac{3}{2} \quad \Longrightarrow \quad \boxed{t_{\rm max} = \left(\frac{2}{3}\right)^3 t_0 = \frac{8}{27}\,t_0\ .}$$

The maximum distance is then

$$\ell_{p,\rm max} = \ell_p(t_{\rm max}) = 3c\left(\frac{2}{3}\right)^2 t_0^{2/3}\left[t_0^{1/3} - \left(\frac{2}{3}\right)t_0^{1/3}\right] = 3c\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)t_0$$

$$= \boxed{\frac{4}{9}ct_0\ .}$$

PROBLEM 23: DID YOU DO THE READING (2016)? *(35 points)*

(a) *(5 points)* The Milky Way has been known since ancient times as a band of light stretching across the sky. We now recognize the Milky Way as the galaxy of stars in which we live, with a large collection of stars, including our sun, arranged in a giant disk. Since the individual stars are mostly too small for our eyes to resolve, we observe the collective light from these stars, concentrated in the plane of the disk. The idea that the Milky Way is actually a disk of stars was proposed by

(i) Claudius Ptolemy, in the 2nd century AD.

(ii) Johannes Kepler, in 1610.

(iii) Isaac Newton, in 1695.

(iv) $\boxed{\text{Thomas Wright, in 1750.}}$

(v) Immanuel Kant, in 1755.

(vi) Edwin Hubble, in 1923.

(b) *(5 points)* Once it was recognized that we live in a galaxy, it was initially assumed that ours was the only galaxy. The suggestion that some of the patches of light known as nebulae might actually be other galaxies like our own was made by

(i) Claudius Ptolemy, in the 2nd century AD.

(ii) Johannes Kepler, in 1610.

(iii) Isaac Newton, in 1695.

(iv) Thomas Wright, in 1750.

(v) $\boxed{\text{Immanuel Kant, in 1755.}}$

(vi) Edwin Hubble, in 1923.

(c) *(5 points)* The first firm evidence that there is more than one galaxy stemmed from the ability to observe the Andromeda Nebula with high enough resolution to distinguish its individual stars. In particular, the observation of Cepheid variable stars in Andromeda allowed a distance estimate that placed it well outside the Milky Way. The observation of Cepheid variable stars in Andromeda was first made by

(i) Johannes Kepler, in 1610.

(ii) Isaac Newton, in 1695.

(iii Thomas Wright, in 1750.

(iv) Immanuel Kant, in 1755.

(v) Henrietta Swan Leavitt and Harlow Shapley in 1915.

(vi) $\boxed{\text{Edwin Hubble, in 1923.}}$

(d) *(5 points)* The first hint that the universe is filled with radiation with an effective temperature near 3 K, although not recognized at the time, was an observation of absorption lines in cyanogen (CN) by Adams and McKellar in 1941. They observed dark spectral lines which they interpreted as absorption by the cyanogen of light coming from the star behind the gas cloud. Explain in a few sentences how these absorption lines can be used to make inferences about the cosmic background radiation bathing the cyanogen gas cloud.

Answer:

When an atom absorbs a photon, it is excited from its initial state to some final state, and the energy of the photon must match the energy difference betwen the two states. One of the observed cyanogen lines was associated with a transition starting in the ground state, and two other observed lines were associated with transitions starting from an excited state. By comparing the intensities of these absorption lines, the astronomers could infer the relative abundance of ground state and excited cyanogen molecules, which in turn allowed them to infer the temperature of the gas cloud. They found a temperature of 2.2 K.

(e) *(5 points)* As the universe expands, the temperature of the cosmic microwave background

(i) goes up in proportion to the scale factor $a(t)$.

(ii) stays constant.

(iii) $\boxed{\text{goes down in proportion to } 1/a(t).}$

(iv) goes down in proportion to $1/a^2(t)$.

(f) *(5 points)* When Hubble measured the value of his constant, he found $H^{-1} \approx 100$ million years, $\boxed{\text{2 billion years,}}$ 10 billion years, or 20 billion years?

(g) *(5 points)* Explain in a few sentences what is meant by the equivalence principle?

Answer:

Ryden states that the equivalence principle is the fact that the gravitational mass of any object is equal to its inertial mass. It would also be correct to say that the graviational mass is proportional to the inertial mass. (If they are proportional, there is always a value of G which makes them equal.) The equivalence principle can also be described more generally by saying that gravity is equivalent to acceleration, so that within a small volume the effects of gravity can be removed by describing the system in an accelerating coordinate system.

PROBLEM 24: OBSERVING A DISTANT GALAXY IN A MATTER-DOMINATED FLAT UNIVERSE *(40 points)*

Suppose that we are living in a matter-dominated flat universe, with a scale factor given by

$$a(t) = bt^{2/3} ,$$

where b is a constant. The present time is denoted by t_0.

(a) *(5 points)* If we measure time in seconds, distance in meters, and coordinate distances in notches, what are the units of b?

Answer:

$a(t)$ would be measured in meters/notch, and t would be measured in seconds. So

$$[b] = \frac{[a(t)]}{[t]^{2/3}} = \boxed{\frac{\mathrm{m}}{\text{notch-s}^{2/3}} \; .}$$

(b) *(5 points)* Suppose that we observe a distant galaxy which is one half of a "Hubble length" away, which means that the physical distance today is $\ell_p = \frac{1}{2} c H_0^{-1}$, where c is the speed of light and H_0 is the present value of the Hubble expansion rate. What is the proper velocity $v_p \equiv \frac{\mathrm{d}\ell_p(t)}{\mathrm{d}t}$ of this galaxy relative to us?

Answer:

By Hubble's law, the velocity of recession is equal to H_0 times the physical distance, so

$$v_p = H_0 \left[\frac{1}{2} c H_0^{-1} \right] = \boxed{\frac{1}{2} c \; .}$$

A common error in this part was to use

$$H_0 = \frac{\dot{a}}{a} = \frac{2}{3} \frac{b t_0^{-1/3}}{b t_0^{2/3}} = \frac{2}{3 t_0}$$

to write

$$\ell_p = \frac{3}{4} c t_0 \; ,$$

and then to differentiate this expression with respect to t_0, finding $v_p = 3c/4$. The problem with this approach is that it assumes that the relation $\ell_p = \frac{1}{2} c H^{-1}$ holds for all t, so that one can differentiate it to find the velocity. But an object that is at distance $\frac{1}{2} c H^{-1}$ does not remain at a distance $\frac{1}{2} c H^{-1}$ as time progresses. It is the coordinate distance ℓ_c, and not the physical distance measured in Hubble lengths, that remains constant as the universe expands.

(c) *(5 points)* What is the coordinate distance ℓ_c between us and the distant galaxy? Express your answer in terms of b, t_0, and c (but not H_0).

Answer:

We know that $\ell_p(t) = a(t)\ell_c$, so

$$\ell_c = \frac{\ell_p(t_0)}{a(t_0)} = \frac{c}{2bH_0 t_0^{2/3}} \ .$$

To eliminate H_0, which is not allowed in the answer, we can use

$$H_0 = \frac{1}{a(t_0)}\frac{\mathrm{d}a(t_0)}{\mathrm{d}t_0} = \frac{1}{bt_0^{2/3}}\left[\frac{2}{3}bt_0^{-1/3}\right] = \frac{2}{3t_0} \ .$$

Inserting the result into the line above,

$$\boxed{\ell_c = \frac{3}{4}\frac{ct_0^{1/3}}{b} \ .}$$

If you did not answer the previous part, you may still continue with the following parts, using the symbol ℓ_c for the coordinate distance to the galaxy.

(d) *(5 points)* At what time t_e was the light that we are now receiving from the galaxy emitted?

Answer:

We know that the coordinate velocity of light is

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{c}{a(t)} = \frac{c}{bt^{2/3}}.$$

We can find t_e by the requirement that the coordinate distance that light travels between t_e and t_0 must be equal to ℓ_c found in part (c):

$$\int_{t_e}^{t_0} \frac{c}{bt'^{2/3}}\,\mathrm{d}t' = \frac{3}{4}\frac{ct_0^{1/3}}{b} \ .$$

Integrating,

$$\frac{3c}{b}\left[t_0^{1/3} - t_e^{1/3}\right] = \frac{3}{4}\frac{ct_0^{1/3}}{b} \ .$$

With a little algebra we see

$$t_e^{1/3} = \frac{3}{4}t_0^{1/3} \quad \Longrightarrow \quad \boxed{t_e = \frac{27}{64}t_0 \ .}$$

(e) *(5 points)* What is the redshift z of the light that we are now receiving from the distant galaxy?

Answer:

The redshift is related to the scale factor by

$$1+z=\frac{a(t_0)}{a(t_e)}=\left(\frac{t_0}{t_e}\right)^{2/3}=\left(\frac{64}{27}\right)^{2/3}=\frac{16}{9}\ ,$$

so

$$\boxed{z=\frac{7}{9}\ .}$$

(f) *(10 points)* Consider a light pulse that leaves the distant galaxy at time t_e, as calculated in part (d), and arrives here at the present time, t_0. Calculate the physical distance $r_p(t)$ between the light pulse and us. Find $r_p(t)$ as a function of t for all t between t_e and t_0.

Answer:

We first calculate the coordinate separation $r_c(t)$ between the light pulse and us, as a function of t. At time t_e it is equal to the value of ℓ_c found in part (c), and from that time onward it is reduced by the coordinate distance that light can travel between times t_e and t. Therefore,

$$\begin{aligned}
r_c(t) &= \frac{3}{4}\frac{ct_0^{1/3}}{b}-\int_{t_e}^{t}\frac{c}{b\,t'^{2/3}}\,\mathrm{d}t' \\
&= \frac{3}{4}\frac{ct_0^{1/3}}{b}-\frac{3c}{b}\left[t^{1/3}-t_e^{1/3}\right] \\
&= \frac{3}{4}\frac{ct_0^{1/3}}{b}-\frac{3c}{b}\left[t^{1/3}-\frac{3}{4}t_0^{1/3}\right] \\
&= \frac{3c}{b}\left[t_0^{1/3}-t^{1/3}\right]\ .
\end{aligned}$$

The physical distance is then

$$r_p(t)=bt^{2/3}\,r_c(t)=3c\left[t_0^{1/3}t^{2/3}-t\right]=\boxed{3ct\left[\left(\frac{t_0}{t}\right)^{1/3}-1\right]\ .}$$

(g) *(5 points)* If we send a radio message now to the distant galaxy, at what time t_r will it be received?

Answer:

We calculate the time t_r by which a light ray, starting at t_0, can travel a coordinate distance equal to the value we found in part (c):

$$\int_{t_0}^{t_r} \frac{c}{b\, t'^{2/3}}\, \mathrm{d}t' = \ell_c = \frac{3}{4} \frac{c t_0^{1/3}}{b} \ .$$

Integrating,

$$\frac{3c}{b}\left[t_r^{1/3} - t_0^{1/3}\right] = \frac{3}{4}\frac{c t_0^{1/3}}{b} \ ,$$

from which we find

$$t_r^{1/3} = \frac{5}{4} t_0^{1/3} \quad \Longrightarrow \quad \boxed{t_r = \frac{125}{64}\, t_0 \ .}$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — October 3, 2018
Prof. Alan Guth

QUIZ 1

Reformatted to Remove Blank Pages

Please answer all questions in this stapled booklet.

PROBLEM 1: DID YOU DO THE READING? *(30 points)*

(a) *(5 points)* After telescopes became available, more and more extended objects in the sky, called nebulae, were discovered, but those were thought as members of our galaxy. Who is the person who first proposed that some of the nebulae are galaxies like our own located outside our galaxy?

(i) Isaac Newton

(ii) Immanuel Kant

(iii) Edwin Hubble

(iv) Albert Einstein

(b) *(5 points)* Before 1923, questions of the nature of the spiral and elliptical nebulae could not be settled without some reliable method of determining how far away they are. In 1923, Edwin Hubble was for the first time able to resolve the Andromeda Nebula (galaxy) into separate stars and estimated the distance to the Andromeda Nebula. What observational quantity did he measure to estimate the distance?

(i) the radial velocity of individual stars in the Adromeda Nebula

(ii) the radial velocity of the Andromeda Nebula itself

(iii) the periods of variation of a class of stars in the Andromeda Nebula

(iv) the parallax of bright stars in the Adromeda Nebula

— *Problem 1 continues on next page.* —

(c) *(5 points)* In 1917, a year after the completion of Einstein's general theory of relativity, _________ looked specifically for a solution that would be homogeneous, isotropic, and static, and thus was forced to mutilate the equations by introducing a term, the so-called cosmological constant. In the same year, another solution of the modified theory was found by the Dutch astronomer _________. Although this solution appeared to be static, it had the remarkable property of predicting a redshift proportional to the distance. In 1922, the general homogeneous and isotropic solution of the original Einstein equations was found by the Russian mathematician _________, which provides a mathematical background for the most modern cosmological theories. Which is the right answer to fill in the blanks in turn?

(i) Friedmann — Einstein — de Sitter

(ii) Friedmann — de Sitter — Einstein

(iii) Einstein — Friedmann — de Sitter

(iv) Einstein — de Sitter — Friedmann

(v) de Sitter — Einstein — Friedmann

(vi) de Sitter — Friedmann — Einstein

(d) *(5 points)* After radio noises with the equivalent temperature of about 3.5° K were detected, Penzias, Wilson, Dicke, Peebles, Roll, and Wilkinson decided to publish a pair of companion letters in the Astrophysical Journal, in which Penzias and Wilson would announce their observations, and Dicke, Peebles, Roll, and Wilkinson would explain the cosmological interpretation. What is the title of the paper written by Penzias and Wilson?

(i) "A Measurement of Excess Antenna Temperature at 4,080 Mc/s"

(ii) "Cosmic Black-Body Radiation"

(iii) "Origin of the Microwave Radio Background"

(iv) "Three Degrees Above Zero: Bell Labs in the Information Age"

— Problem 1 continues on next page. —

(e) *(5 points)* The universe contains different types of particles. Which of the following statements is NOT true?

(i) A baryon is defined as a particle made of three quarks.

(ii) Electrons and neutrinos are leptons.

(iii) There are three types of neutrinos and they all have zero charge.

(iv) The component of the universe made of ions, atoms, and molecules is generally referred to as baryonic matter, since only the baryons (protons and neutrons) contribute significantly to the mass density.

(v) About three-fourths of the baryonic matter in the universe is currently in the form of helium.

(f) *(5 points)* If one averages over sufficiently large scales, the universe appears to be homogeneous and isotropic. How large must the averaging scale be before this homogeneity and isotropy set in?*

(i) 1000 Mpc. (1 Mpc $= 10^6$ pc, 1 pc $= 3.086 \times 10^{16}$ m $= 3.262$ light-year).

(ii) 100 Mpc.

(iii) 1 Mpc.

(iv) 100 kpc (1 kpc = 1000 pc).

(v) 1 AU (1 AU $= 1.496 \times 10^{11}$ m).

— *End of Problem 1.* —

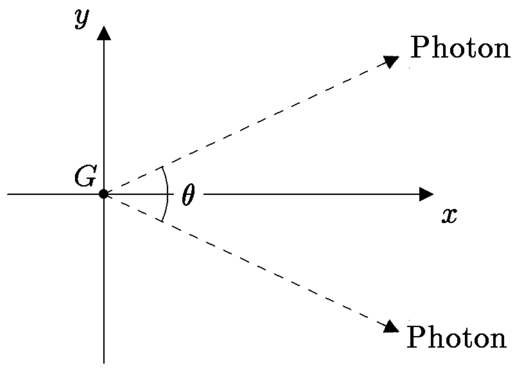
* This question was a replacement, listed on a separate sheet when the quiz was administered.

PROBLEM 2: LIGHT RAYS TRAVELING THROUGH A MATTER-DOMINATED FLAT UNIVERSE *(40 points)*

Consider a flat, matter-dominated universe, with a scale factor given by

$$a(t) = bt^{2/3} ,$$

where b is a constant. Now consider a galaxy G in this universe which at time t_1 emits two photons, with an angular separation θ between their paths, as shown in the diagram:

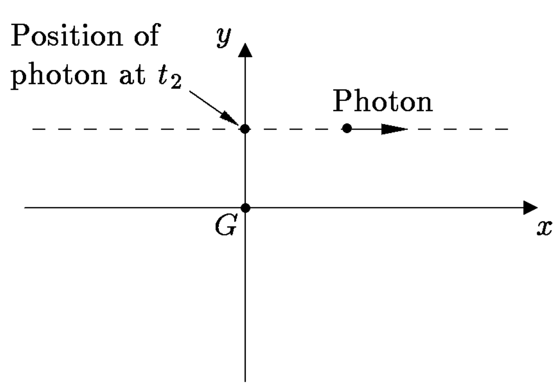


(a) *(10 points)* At cosmic time t (for $t > t_1$), what is the physical distance $\ell_{1,\mathrm{phys}}(t)$ of each of these photons from the galaxy G?

(b) *(5 points)* If the frequency of the photons was ν_1 when they were emitted, what is their frequency $\nu(t)$ at cosmic time t (for $t > t_1$)? $\nu(t)$ should be the frequency as it would be measured by a comoving observer, i.e. an observer at rest with respect to the matter at the same location.

(c) *(10 points)* What is the physical distance $\ell_{2,\mathrm{phys}}(t)$ between the two photons at time t (for $t > t_1$)?

Now consider a different situation, but in the same universe. This time we consider a photon that travels past the galaxy G, traveling in the x direction, in the x-y plane, as shown in the diagram below. We are told that the photon crosses the y axis at time t_2, and at that time the photon is a physical distance h from the galaxy.



(d) *(10 points)* What is the physical distance $\ell_{3,\text{phys}}(t)$ between the photon and the galaxy G at arbitrary time t, which might be earlier or later than t_2?

(e) *(5 points)* At time t_2, what is the recessional speed $d\ell_{3,\text{phys}}(t)/dt$ of the photon from the galaxy. *Hint:* if you are clever, this can be done with very little calculation.

PROBLEM 3: THE STEADY-STATE UNIVERSE THEORY *(30 points)*

The following problem was Problem 2, Quiz 1, 2000. It was also Problem 2 of the Quiz 1 Review Problems, 2018.

Until the discovery of the cosmic microwave background, the steady state theory was considered a viable model of the universe. As the name suggests, this theory is based on the hypothesis that the large-scale properties of the universe do not change with time. The expansion of the universe was an established fact when the steady-state theory was invented, but the steady-state theory reconciles the expansion with a steady-state density of matter by proposing that new matter is created as the universe expands, so that the matter density does not fall. Like the conventional theory, the steady-state theory describes a homogeneous, isotropic, expanding universe, so the same comoving coordinate formulation can be used.

a) *(15 points)* The steady-state theory proposes that the Hubble constant, like other cosmological parameters, does not change with time, so $H(t) = H_0$. Find the most general form for the scale factor function $a(t)$ which is consistent with this hypothesis.

b) *(15 points)* Suppose that the mass density of the universe is ρ_0, which of course does not change with time. In terms of the general form for $a(t)$ that you found in part (a), calculate the rate at which new matter must be created for ρ_0 to remain constant as the universe expands. Your answer should have the units of mass per unit volume per unit time. *[If you failed to answer part (a), you will still receive full credit here if you correctly answer the question for an arbitrary scale factor function $a(t)$.]*

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe October 3, 2018
Prof. Alan Guth

QUIZ 1 FORMULA SHEET

DOPPLER SHIFT (For motion along a line):

$$z = v/u \quad \text{(nonrelativistic, source moving)}$$

$$z = \frac{v/u}{1 - v/u} \quad \text{(nonrelativistic, observer moving)}$$

$$z = \sqrt{\frac{1+\beta}{1-\beta}} - 1 \quad \text{(special relativity, with } \beta = v/c\text{)}$$

COSMOLOGICAL REDSHIFT:

$$1 + z \equiv \frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}} = \frac{a(t_{\text{observed}})}{a(t_{\text{emitted}})}$$

SPECIAL RELATIVITY:

Time Dilation Factor:

$$\gamma \equiv \frac{1}{\sqrt{1-\beta^2}} \; , \qquad \beta \equiv v/c$$

Lorentz-Fitzgerald Contraction Factor: γ

Relativity of Simultaneity:
Trailing clock reads later by an amount $\beta \ell_0 / c$.

KINEMATICS OF A HOMOGENEOUSLY EXPANDING UNIVERSE:

Hubble's Law: $v = Hr$,
where v = recession velocity of a distant object, H = Hubble expansion rate, and r = distance to the distant object.

Present Value of Hubble Expansion Rate (Planck 2018):

$$H_0 = 67.66 \pm 0.42 \text{ km-s}^{-1}\text{-Mpc}^{-1}$$

Scale Factor: $\ell_p(t) = a(t)\ell_c$,

where $\ell_p(t)$ is the physical distance between any two objects, $a(t)$ is the scale factor, and ℓ_c is the coordinate distance between the objects, also called the comoving distance.

Hubble Expansion Rate: $H(t) = \dfrac{1}{a(t)} \dfrac{\mathrm{d}a(t)}{\mathrm{d}t}$.

Light Rays in Comoving Coordinates: Light rays travel in straight lines with speed $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \dfrac{c}{a(t)}$.

EVOLUTION OF A MATTER-DOMINATED UNIVERSE:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3} G\rho - \frac{kc^2}{a^2} \ , \quad \ddot{a} = -\frac{4\pi}{3} G\rho a \ ,$$

$$\rho(t) = \frac{a^3(t_i)}{a^3(t)} \, \rho(t_i)$$

$$\Omega \equiv \rho/\rho_c \ , \quad \text{where} \quad \rho_c = \frac{3H^2}{8\pi G} \ .$$

Flat ($k = 0$): $a(t) \propto t^{2/3}$, $\Omega = 1$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — October 8, 2018
Prof. Alan Guth

QUIZ 1 SOLUTIONS

Quiz Date: October 3, 2018

PROBLEM 1: DID YOU DO THE READING? *(30 points)*

(a) *(5 points)* After telescopes became available, more and more extended objects in the sky, called nebulae, were discovered, but those were thought as members of our galaxy. Who is the person who first proposed that some of the nebulae are galaxies like our own located outside our galaxy?

(i) Isaac Newton

(ii) $\boxed{\text{Immanuel Kant}}$

(iii) Edwin Hubble

(iv) Albert Einstein

(b) *(5 points)* Before 1923, questions of the nature of the spiral and elliptical nebulae could not be settled without some reliable method of determining how far away they are. In 1923, Edwin Hubble was for the first time able to resolve the Andromeda Nebula (galaxy) into separate stars and estimated the distance to the Andromeda Nebula. What observational quantity did he measure to estimate the distance?

(i) the radial velocity of individual stars in the Adromeda Nebula

(ii) the radial velocity of the Andromeda Nebula itself

(iii) $\boxed{\text{the periods of variation of a class of stars in the Andromeda Nebula}}$

(iv) the parallax of bright stars in the Adromeda Nebula

[Comment: Hubble used Cepheid variables to estimate the distance to the Andromeda Nebula (galaxy) with a tight relation between the observed periods of variation of the Cepheids and their absolute luminosities provided by Henrietta Swan Leavitt and Harlow Shapley.]

— Problem 1 continues on next page. —

(c) *(5 points)* In 1917, a year after the completion of Einstein's general theory of relativity, __________ looked specifically for a solution that would be homogeneous, isotropic, and static, and thus was forced to mutilate the equations by introducing a term, the so-called cosmological constant. In the same year, another solution of the modified theory was found by the Dutch astronomer __________. Although this solution appeared to be static, it had the remarkable property of predicting a redshift proportional to the distance. In 1922, the general homogeneous and isotropic solution of the original Einstein equations was found by the Russian mathematician __________, which provides a mathematical background for the most modern cosmological theories. Which is the right answer to fill in the blanks in turn?

(i) Friedmann — Einstein — de Sitter

(ii) Friedmann — de Sitter — Einstein

(iii) Einstein — Friedmann — de Sitter

(iv) $\boxed{\text{Einstein — de Sitter — Friedmann}}$

(v) de Sitter — Einstein — Friedmann

(vi) de Sitter — Friedmann — Einstein

(d) *(5 points)* After radio noises with the equivalent temperature of about 3.5° K were detected, Penzias, Wilson, Dicke, Peebles, Roll, and Wilkinson decided to publish a pair of companion letters in the Astrophysical Journal, in which Penzias and Wilson would announce their observations, and Dicke, Peebles, Roll, and Wilkinson would explain the cosmological interpretation. What is the title of the paper written by Penzias and Wilson?

(i) $\boxed{\text{“A Measurement of Excess Antenna Temperature at 4,080 Mc/s”}}$

(ii) “Cosmic Black-Body Radiation”

(iii) “Origin of the Microwave Radio Background”

(iv) “Three Degrees Above Zero: Bell Labs in the Information Age”

[Comment: (ii) is the title of the companion letter by Dicke, Peebles, Roll, and Wilkinson; (iii) is the title of a paper written by Peebles and Dicke in 1966, in which they refuted a suggestion by Michele Kaufman that the background radiation was emitted by ionized intergalactic hydrogen; and (iv) is the title of a book written by Jeremy Bernstein.]

— Problem 1 continues on next page. —

(e) *(5 points)* The universe contains different types of particles. Which of the following statements is NOT true?

(i) A baryon is defined as a particle made of three quarks.

(ii) Electrons and neutrinos are leptons.

(iii) There are three types of neutrinos and they all have zero charge.

(iv) The component of the universe made of ions, atoms, and molecules is generally referred to as baryonic matter, since only the baryons (protons and neutrons) contribute significantly to the mass density.

(v) $\boxed{\text{About three-fourths of the baryonic matter in the universe is currently in the form of helium.}}$

[Comment: about three-fourths of the baryonic matter in the universe is currently in the form of hydrogen.]

(f) *(5 points)* If one averages over sufficiently large scales, the universe appears to be homogeneous and isotropic. How large must the averaging scale be before this homogeneity and isotropy set in?*

(i) 1000 Mpc. (1 Mpc $= 10^6$ pc, 1 pc $= 3.086 \times 10^{16}$ m $= 3.262$ light-year).

(ii) $\boxed{\text{100 Mpc.}}$

(iii) 1 Mpc.

(iv) 100 kpc (1 kpc $=$ 1000 pc).

(v) 1 AU (1 AU $= 1.496 \times 10^{11}$ m).

— *End of Problem 1.* —

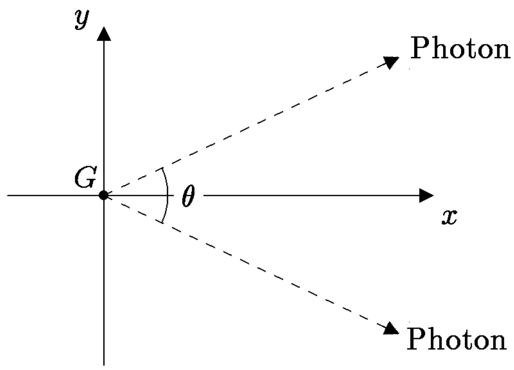
* This question was a replacement, listed on a separate sheet when the quiz was administered.

PROBLEM 2: LIGHT RAYS TRAVELING THROUGH A MATTER-DOMINATED FLAT UNIVERSE *(40 points)*

Consider a flat, matter-dominated universe, with a scale factor given by

$$a(t) = bt^{2/3} ,$$

where b is a constant. Now consider a galaxy G in this universe which at time t_1 emits two photons, with an angular separation θ between their paths, as shown in the diagram:



(a) *(10 points)* At cosmic time t (for $t > t_1$), what is the physical distance $\ell_{1,\text{phys}}(t)$ of each of these photons from the galaxy G?

Answer:

The coordinate speed of light is $c/a(t)$, so the coordinate distance traveled is

$$\begin{aligned}\ell_{1,c}(t) &= \int_{t_1}^{t} \frac{c}{a(t')} \mathrm{d}t' \\ &= \left(\frac{3c}{b}\right) t^{1/3}\Big|_{t_1}^{t} \\ &= \left(\frac{3c}{b}\right) t^{1/3} \left[1 - \left(\frac{t_1}{t}\right)^{1/3}\right] .\end{aligned}$$

The physical distance is then

$$\begin{aligned}\ell_{1,\text{phys}}(t) &= a(t)\ell_{1,c}(t) = bt^{2/3}\ell_{1,c}(t) \\ &= \boxed{3ct\left[1 - \left(\frac{t_1}{t}\right)^{1/3}\right] .}\end{aligned}$$

(b) *(5 points)* If the frequency of the photons was ν_1 when they were emitted, what is their frequency $\nu(t)$ at cosmic time t (for $t > t_1$)? $\nu(t)$ should be the frequency as it would be measured by a comoving observer, i.e. an observer at rest with respect to the matter at the same location.

Answer:

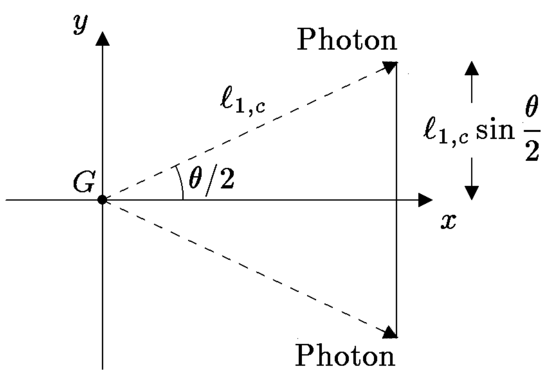
The wavelength of a photon is stretched in proportion to the scale factor, so the frequency is inversely proportional to the scale factor. So

$$\boxed{\nu(t) = \frac{a(t_1)}{a(t)}\nu_1 = \left(\frac{t_1}{t}\right)^{2/3} \nu_1 \ .}$$

(c) *(10 points)* What is the physical distance $\ell_{2,\text{phys}}(t)$ between the two photons at time t (for $t > t_1$)?

Answer:

Since the universe is flat, we can use ordinary Euclidean geometry, as shown in the diagram:



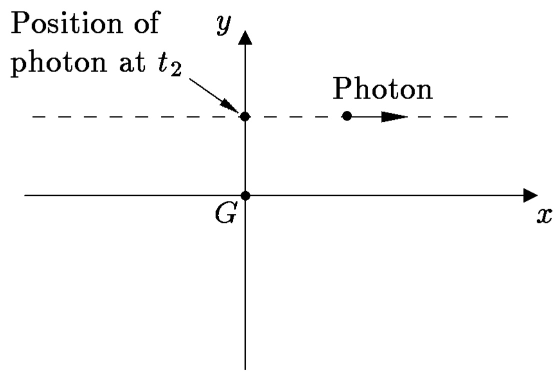
The coordinate distance between the two photons is then given by

$$\ell_{2,c} = 2\ell_{1,c} \sin \frac{\theta}{2} \ .$$

The physical distance is then

$$\begin{aligned}
\ell_{2,\text{phys}} &= 2a(t)\ell_{1,c} \sin \frac{\theta}{2} \\
&= 2bt^{2/3} \left(\frac{3c}{b}\right) t^{1/3} \left[1 - \left(\frac{t_1}{t}\right)^{1/3}\right] \sin \frac{\theta}{2} \\
&= \boxed{6ct \left[1 - \left(\frac{t_1}{t}\right)^{1/3}\right] \sin \frac{\theta}{2} \ .}
\end{aligned}$$

Now consider a different situation, but in the same universe. This time we consider a photon that travels past the galaxy G, traveling in the x direction, in the x-y plane, as shown in the diagram below. We are told that the photon crosses the y axis at time t_2, and at that time the photon is a physical distance h from the galaxy.



(d) *(10 points)* What is the physical distance $\ell_{3,\text{phys}}(t)$ between the photon and the galaxy G at arbitrary time t, which might be earlier or later than t_2?

Answer:

It is important to recognize here that the coordinates shown are comoving coordinates, or map coordinates, so that physical distances are obtained by multiplying by the scale factor. (If these coordinates represented physical distances from the origin, then the Hubble expansion would be driving all particles outward, and the photon trajectory would not be a straight line.) So, if the physical distance between the photon and the galaxy is equal to h at time t_2, then the y coordinate of the photon is equal to

$$y = \frac{h}{a(t_2)} \ .$$

The x coordinate is determined by the fact that it vanishes at time t_2, and then moves toward positive values at the coordinate speed of light,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{c}{a(t)} \ .$$

Thus,

$$x(t) = \int_{t_2}^{t} \frac{c}{bt'^{2/3}} \mathrm{d}t' = \frac{3c}{b} t^{1/3} \left[1 - \left(\frac{t_2}{t} \right)^{1/3} \right] \ .$$

The coordinate distance from the origin is then given by the Pythagorean theorem,

$$\ell_{3,c}(t) = \left[x^2(t) + y^2(t) \right]^{1/2} \ ,$$

so

$$\ell_{3,\mathrm{phys}}(t) = bt^{2/3}\left[x^2(t)+y^2(t)\right]^{1/2}$$

$$= \boxed{\left\{9c^2t^2\left[1-\left(\frac{t_2}{t}\right)^{1/3}\right]^2+\left(\frac{t}{t_2}\right)^{4/3}h^2\right\}^{1/2}} \; .$$

(e) *(5 points)* At time t_2, what is the recessional speed $d\ell_{3,\mathrm{phys}}(t)/dt$ of the photon from the galaxy. *Hint:* if you are clever, this can be done with very little calculation.

Answer:

Note, first of all, that one cannot blindly assume that the photon obeys Hubble's law, since Hubble's law applies only to the comoving matter in the model universe, which is undergoing uniform expansion. It does not apply to objects, such as photons, that are moving relative to the comoving matter. (Motion relative to the comoving matter is called proper motion.)

The answer can be obtained by simply differentiating the above expression for $\ell_{3,\mathrm{phys}}(t)$ with respect to t, and then setting $t = t_2$, but there is a shorter way. If we go back to

$$\ell_{3,\mathrm{phys}}(t) = a(t)\ell_{3,c}(t) \; ,$$

we note that, unlike the the description of uniform Hubble expansion, in this case the coordinate distance $\ell_{3,c}$ depends on time. The coordinate distance between two pieces of comoving matter (i.e., matter expanding with the universe) does not change with time, but here we have the distance between a galaxy (at fixed coordinates) and a photon (which is traveling). However, we can easily see from the diagram that at time t_2, the coordinate distance $\ell_{3,c}(t)$ is at its minimum, and therefore its time derivative at t_2 must be zero. Therefore,

$$\left.\frac{\mathrm{d}\ell_{3,\mathrm{phys}}}{\mathrm{d}t}\right|_{t_2} = \dot a(t_2)\ell_{3,c}(t_2)$$

$$= \left(\frac{\dot a}{a}\right)\left[a\ell_{3,c}(t_2)\right] = H(t_2)\ell_{3,\mathrm{phys}}(t_2)$$

$$= \boxed{\left(\frac{2}{3t_2}\right)h \; .}$$

The average grade on this problem was only 2.8/5, or 55%, which was the lowest for any problem on the quiz. Many students assumed that Hubble's law applied to the photon. This assumption leads to the correct answer at time t_2, when the photon proper velocity is perpendicular to the direction from the photon to the galaxy, but not at other times. Students who gave the correct answer, but attributed it to Hubble's law, were given 4 points out of 5. Another common error was to assert that the speed of light is always measured as c, so $\mathrm{d}\ell_{3,\text{phys}}(t)/\mathrm{d}t = c$. The correct description of the invariance of the speed of light is to say that any inertial observer (which includes all comoving observers) will measure the speed of a photon that passes him as being equal to c. But if the photon is at a different location, then one has to take into account the expansion of the universe, which is done by basing all calculations on the principle that the coordinate speed of light is always equal to $c/a(t)$.

PROBLEM 3: THE STEADY-STATE UNIVERSE THEORY *(30 points)*

The following problem was Problem 2, Quiz 1, 2000. It was also Problem 2 of the Quiz 1 Review Problems, 2018.

Until the discovery of the cosmic microwave background, the steady state theory was considered a viable model of the universe. As the name suggests, this theory is based on the hypothesis that the large-scale properties of the universe do not change with time. The expansion of the universe was an established fact when the steady-state theory was invented, but the steady-state theory reconciles the expansion with a steady-state density of matter by proposing that new matter is created as the universe expands, so that the matter density does not fall. Like the conventional theory, the steady-state theory describes a homogeneous, isotropic, expanding universe, so the same comoving coordinate formulation can be used.

a) *(15 points)* The steady-state theory proposes that the Hubble constant, like other cosmological parameters, does not change with time, so $H(t) = H_0$. Find the most general form for the scale factor function $a(t)$ which is consistent with this hypothesis.

Answer:

The Hubble expansion rate is related to $a(t)$ by

$$H(t) = \frac{1}{a(t)}\,\frac{da}{dt} \ ,$$

so in this case

$$\frac{1}{a(t)}\,\frac{da}{dt} = H_0 \ ,$$

which can be rewritten as

$$\frac{da}{a} = H_0\,dt \ .$$

Integrating,

$$\ln a = H_0\, t + c\ ,$$

where c is a constant of integration. Exponentiating,

$$\boxed{a = b e^{H_0\, t}\ ,}$$

where $b = e^c$ is an arbitrary constant.

b) *(15 points)* Suppose that the mass density of the universe is ρ_0, which of course does not change with time. In terms of the general form for $a(t)$ that you found in part (a), calculate the rate at which new matter must be created for ρ_0 to remain constant as the universe expands. Your answer should have the units of mass per unit volume per unit time. *[If you failed to answer part (a), you will still receive full credit here if you correctly answer the question for an arbitrary scale factor function $a(t)$.]*

Answer:

Consider a cube of side ℓ_c drawn on the comoving coordinate system diagram. The physical length of each side is then $a(t)\,\ell_c$, so the physical volume is

$$V(t) = a^3(t)\,\ell_c^3\ .$$

Since the mass density is fixed at $\rho = \rho_0$, the total mass inside this cube at any given time is given by

$$M(t) = a^3(t)\,\ell_c^3\,\rho_0\ .$$

In the absence of matter creation the total mass within a comoving volume would not change, so the increase in mass described by the above equation must be attributed to matter creation. The rate of matter creation per unit time per unit volume is then given by

$$\begin{aligned}
\text{Rate} &= \frac{1}{V(t)}\frac{dM}{dt} \\
&= \frac{1}{a^3(t)\,\ell_c^3}\,3a^2(t)\frac{da}{dt}\,\ell_c^3\,\rho_0 \\
&= \frac{3}{a}\frac{da}{dt}\,\rho_0 \\
&= \boxed{3H_0\,\rho_0\ .}
\end{aligned}$$

You were not asked to insert numbers, but it is worthwhile to consider the numerical value after the exam, to see what this answer is telling us. Suppose we take $H_0 = 70$ km-sec^{-1}-Mpc^{-1}, and take ρ_0 to be the critical density, $\rho_c = 3H_0^2/8\pi G$. Then

$$\begin{aligned}
\text{Rate} &= \frac{9H_0^3}{8\pi G} \\
&= \frac{9 \times (70 \text{ km-s}^{-1}\text{-Mpc}^{-1})^3}{8\pi \times 6.67260 \times 10^{-11} \text{ m}^3\text{-kg}^{-1}\text{-s}^{-2}} \\
&= \frac{9 \times (70 \cancel{\text{km}}\text{-s}^{-1}\text{-}\cancel{\text{Mpc}}^{-1})^3}{8\pi \times 6.67260 \times 10^{-11} \text{ m}^3\text{-kg}^{-1}\text{-s}^{-2}} \\
&\qquad \times \left(\frac{1 \cancel{\text{Mpc}}}{3.086 \times 10^{22} \cancel{\text{m}}}\right)^3 \times \left(\frac{10^3 \cancel{\text{m}}}{\cancel{\text{km}}}\right)^3 \\
&= 6.26 \times 10^{-44} \text{ kg-m}^{-3}\text{-s}^{-1} \ .
\end{aligned}$$

To put this number into more meaningful terms, note that the mass of a hydrogen atom is 1.67×10^{-27} kg, and that 1 year $= 3.156 \times 10^7$ s. The rate of matter production required for the steady-state universe theory can then be expressed as roughly one hydrogen atom per cubic meter per billion years! Needless to say, such a rate of matter production is totally undetectable, so the steady-state theory cannot be ruled out by the failure to detect matter production.

— End of Problem 3. —

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe October 3, 2018
Prof. Alan Guth

QUIZ 1 FORMULA SHEET

DOPPLER SHIFT (For motion along a line):

$$z = v/u \quad \text{(nonrelativistic, source moving)}$$

$$z = \frac{v/u}{1 - v/u} \quad \text{(nonrelativistic, observer moving)}$$

$$z = \sqrt{\frac{1+\beta}{1-\beta}} - 1 \quad \text{(special relativity, with } \beta = v/c\text{)}$$

COSMOLOGICAL REDSHIFT:

$$1 + z \equiv \frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}} = \frac{a(t_{\text{observed}})}{a(t_{\text{emitted}})}$$

SPECIAL RELATIVITY:

Time Dilation Factor:

$$\gamma \equiv \frac{1}{\sqrt{1-\beta^2}} \,, \qquad \beta \equiv v/c$$

Lorentz-Fitzgerald Contraction Factor: γ

Relativity of Simultaneity:
Trailing clock reads later by an amount $\beta \ell_0 / c$.

KINEMATICS OF A HOMOGENEOUSLY EXPANDING UNIVERSE:

Hubble's Law: $v = Hr$,
where v = recession velocity of a distant object, H = Hubble expansion rate, and r = distance to the distant object.

Present Value of Hubble Expansion Rate (Planck 2018):

$$H_0 = 67.66 \pm 0.42 \text{ km-s}^{-1}\text{-Mpc}^{-1}$$

Scale Factor: $\ell_p(t) = a(t)\ell_c$,
where $\ell_p(t)$ is the physical distance between any two objects, $a(t)$ is the scale factor, and ℓ_c is the coordinate distance between the objects, also called the comoving distance.

Hubble Expansion Rate: $H(t) = \dfrac{1}{a(t)} \dfrac{\mathrm{d}a(t)}{\mathrm{d}t}$.

Light Rays in Comoving Coordinates: Light rays travel in straight lines with speed $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \dfrac{c}{a(t)}$.

EVOLUTION OF A MATTER-DOMINATED UNIVERSE:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3} G\rho - \frac{kc^2}{a^2} \ , \quad \ddot{a} = -\frac{4\pi}{3} G\rho a \ ,$$

$$\rho(t) = \frac{a^3(t_i)}{a^3(t)} \rho(t_i)$$

$$\Omega \equiv \rho/\rho_c \ , \quad \text{where} \quad \rho_c = \frac{3H^2}{8\pi G} \ .$$

Flat ($k = 0$): $a(t) \propto t^{2/3}$, $\Omega = 1$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe October 28, 2018
Prof. Alan Guth

REVIEW PROBLEMS FOR QUIZ 2

Revised Version*

QUIZ DATE: Monday, November 5, 2018, during the normal class time.

COVERAGE: Lecture Notes 4, 5, and through the section on "Dynamics of a Flat Radiation-Dominated Universe" of Lecture Notes 6; Problem Sets 4, 5, and 6; Weinberg, *The First Three Minutes,* Chapters 4 – 7; In Ryden's *Introduction to Cosmology*, we have read Chapters 4, 5, and Sec. 6.1 during this period. These chapters, however, parallel what we have done or will be doing in lecture, so you should take them as an aid to learning the lecture material; there will be no questions explicitly based on these sections from Ryden. But we have also read Chapters 10 (*Nucleosynthesis and the Early Universe*) and 8 (*Dark Matter*) in Ryden, and these are relevant material for the quiz, except for Sec. 10.3 (*Deuterium Synthesis*). We will return to deuterium synthesis later in the course. You can also ignore Ryden's Eqs. (10.11), (10.12), and (10.13) for now. Chapters 4 and 5 of Weinberg's book are packed with numbers; you need not memorize these numbers, but you should be familiar with their orders of magnitude. We will not take off for the spelling of names, as long as they are vaguely recognizable. For dates before 1900, it will be sufficient for you to know when things happened to within 100 years. For dates after 1900, it will be sufficient if you can place events within 10 years. You should expect one 25-point problem based on the readings, and several calculational problems. **One of the problems on the quiz will be taken verbatim (or at least *almost* verbatim) from either the problem sets listed above (extra credit problems included), or from the starred problems from this set of Review Problems.** The starred problems are the ones that I recommend that you review most carefully: Problems 6, 7, 8, 13, 15, 17, 19, and 21.

PURPOSE: These review problems are not to be handed in, but are being made available to help you study. They come mainly from quizzes in previous years. In some cases the number of points assigned to the problem on the quiz is listed — in all such cases it is based on 100 points for the full quiz.

In addition to this set of problems, you will find on the course web page the actual quizzes that were given in 1994, 1996, 1998, 2000, 2002, 2004, 2005, 2007, 2009, 2011, 2013, and 2016. The relevant problems from those quizzes have mostly been incorporated into these review problems, but you still may be interested in looking at the quizzes, just to see how much material has been included in each quiz. The

* Revised November 2, 2018: Problem 23 refers to a table of integrals, which was not included in the original version of the review problems.

coverage of the upcoming quiz will not necessarily match exactly the coverage from previous years, but I believe that all these review problems would be fair problems for the upcoming quiz. The coverage for each quiz in recent years is usually described at the start of the review problems, as I did here. In 2016 we finished Weinberg's book by the time of Quiz 2, but otherwise the coverage was the same as this year.

REVIEW SESSION: To help you study for the quiz, Honggeun Kim will hold a review session, at a time and place to be announced.

FUTURE QUIZ: Quiz 3 will be given on Wednesday, December 5, 2018.

INFORMATION TO BE GIVEN ON QUIZ:

Each quiz in this course will have a section of "useful information" for your reference. For the second quiz, this useful information will be the following:

DOPPLER SHIFT (For motion along a line):

$$z = v/u \quad \text{(nonrelativistic, source moving)}$$

$$z = \frac{v/u}{1 - v/u} \quad \text{(nonrelativistic, observer moving)}$$

$$z = \sqrt{\frac{1+\beta}{1-\beta}} - 1 \quad \text{(special relativity, with } \beta = v/c\text{)}$$

COSMOLOGICAL REDSHIFT:

$$1 + z \equiv \frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}} = \frac{a(t_{\text{observed}})}{a(t_{\text{emitted}})}$$

SPECIAL RELATIVITY:

Time Dilation Factor:

$$\gamma \equiv \frac{1}{\sqrt{1-\beta^2}} \ , \qquad \beta \equiv v/c$$

Lorentz-Fitzgerald Contraction Factor: γ

Relativity of Simultaneity:
Trailing clock reads later by an amount $\beta\ell_0/c$.

Energy-Momentum Four-Vector:

$$p^\mu = \left(\frac{E}{c}\,, \vec{p}\right) \ , \quad \vec{p} = \gamma m_0 \vec{v} \ , \quad E = \gamma m_0 c^2 = \sqrt{\left(m_0 c^2\right)^2 + |\vec{p}|^2\, c^2} \ ,$$

$$p^2 \equiv |\vec{p}|^2 - \left(p^0\right)^2 = |\vec{p}|^2 - \frac{E^2}{c^2} = -\left(m_0 c\right)^2 \ .$$

KINEMATICS OF A HOMOGENEOUSLY EXPANDING UNIVERSE:

Hubble's Law: $v = Hr$,
where v = recession velocity of a distant object, H = Hubble expansion rate, and r = distance to the distant object.

Present Value of Hubble Expansion Rate (Planck 2018):

$$H_0 = 67.66 \pm 0.42 \text{ km-s}^{-1}\text{-Mpc}^{-1}$$

Scale Factor: $\ell_p(t) = a(t)\ell_c$,
where $\ell_p(t)$ is the physical distance between any two objects, $a(t)$ is the scale factor, and ℓ_c is the coordinate distance between the objects, also called the comoving distance.

Hubble Expansion Rate: $H(t) = \dfrac{1}{a(t)}\dfrac{\mathrm{d}a(t)}{\mathrm{d}t}$.

Light Rays in Comoving Coordinates: Light rays travel in straight lines with physical speed c relative to any observer. In Cartesian coordinates, coordinate speed $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \dfrac{c}{a(t)}$. In general, $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = 0$.

Horizon Distance:

$$\begin{aligned}\ell_{p,\text{horizon}}(t) &= a(t)\int_0^t \frac{c}{a(t')}dt' \\ &= \begin{cases} 3ct & \text{(flat, matter-dominated),} \\ 2ct & \text{(flat, radiation-dominated).}\end{cases}\end{aligned}$$

COSMOLOGICAL EVOLUTION:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2} \ , \quad \ddot{a} = -\frac{4\pi}{3}G\left(\rho + \frac{3p}{c^2}\right)a \ ,$$

$$\rho_m(t) = \frac{a^3(t_i)}{a^3(t)}\rho_m(t_i) \ \text{(matter)}, \quad \rho_r(t) = \frac{a^4(t_i)}{a^4(t)}\rho_r(t_i) \ \text{(radiation)}.$$

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) \ , \quad \Omega \equiv \rho/\rho_c \ , \quad \text{where} \quad \rho_c = \frac{3H^2}{8\pi G} \ .$$

EVOLUTION OF A MATTER-DOMINATED UNIVERSE:

Flat ($k=0$): $a(t) \propto t^{2/3}$

$\Omega = 1$.

Closed ($k>0$): $ct = \alpha(\theta - \sin\theta)$, $\dfrac{a}{\sqrt{k}} = \alpha(1-\cos\theta)$,

$$\Omega = \frac{2}{1+\cos\theta} > 1 ,$$

$$\text{where } \alpha \equiv \frac{4\pi}{3}\frac{G\rho}{c^2}\left(\frac{a}{\sqrt{k}}\right)^3 .$$

Open ($k<0$): $ct = \alpha\left(\sinh\theta - \theta\right)$, $\dfrac{a}{\sqrt{\kappa}} = \alpha\left(\cosh\theta - 1\right)$,

$$\Omega = \frac{2}{1+\cosh\theta} < 1 ,$$

$$\text{where } \alpha \equiv \frac{4\pi}{3}\frac{G\rho}{c^2}\left(\frac{a}{\sqrt{\kappa}}\right)^3 ,$$

$$\kappa \equiv -k > 0 .$$

MINKOWSKI METRIC (Special Relativity):

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 .$$

ROBERTSON-WALKER METRIC:

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 = -c^2\,\mathrm{d}t^2 + a^2(t)\left\{\frac{\mathrm{d}r^2}{1-kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\} .$$

Alternatively, for $k>0$, we can define $r = \dfrac{\sin\psi}{\sqrt{k}}$, and then

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 \equiv -c^2\,\mathrm{d}t^2 + \tilde{a}^2(t)\left\{\mathrm{d}\psi^2 + \sin^2\psi\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\} ,$$

where $\tilde{a}(t) = a(t)/\sqrt{k}$. For $k<0$ we can define $r = \dfrac{\sinh\psi}{\sqrt{-k}}$, and then

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 = -c^2\,\mathrm{d}t^2 + \tilde{a}^2(t)\left\{\mathrm{d}\psi^2 + \sinh^2\psi\left(\mathrm{d}\theta^2 + \sin^2\theta\,d\phi^2\right)\right\} ,$$

where $\tilde{a}(t) = a(t)/\sqrt{-k}$. Note that $\tilde{a}$ can be called a if there is no need to relate it to the $a(t)$ that appears in the first equation above.

SCHWARZSCHILD METRIC:

$$\mathrm{d}s^2 \equiv -c^2\mathrm{d}\tau^2 = -\left(1-\frac{2GM}{rc^2}\right)c^2\mathrm{d}t^2 + \left(1-\frac{2GM}{rc^2}\right)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2\ ,$$

GEODESIC EQUATION:

$$\frac{\mathrm{d}}{\mathrm{d}s}\left\{g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}s}\right\} = \frac{1}{2}\left(\partial_i g_{k\ell}\right)\frac{\mathrm{d}x^k}{\mathrm{d}s}\ \frac{\mathrm{d}x^\ell}{\mathrm{d}s}$$

or:

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{g_{\mu\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\right\} = \frac{1}{2}\left(\partial_\mu g_{\lambda\sigma}\right)\frac{\mathrm{d}x^\lambda}{\mathrm{d}\tau}\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau}$$

PROBLEM LIST

PROBLEM 1: DID YOU DO THE READING (2000, 2002)

Parts (a)-(c) of this problem come from Quiz 4, 2000, and parts (d) and (e) come from Quiz 3, 2002.

(a) *(5 points)* By what factor does the lepton number per comoving volume of the universe change between temperatures of $kT = 10$ MeV and $kT = 0.1$ MeV? You should assume the existence of the normal three species of neutrinos for your answer.

(b) *(5 points)* Measurements of the primordial deuterium abundance would give good constraints on the baryon density of the universe. However, this abundance is hard to measure accurately. Which of the following is NOT a reason why this is hard to do?

(i) The neutron in a deuterium nucleus decays on the time scale of 15 minutes, so almost none of the primordial deuterium produced in the Big Bang is still present.

(ii) The deuterium abundance in the Earth's oceans is biased because, being heavier, less deuterium than hydrogen would have escaped from the Earth's surface.

(iii) The deuterium abundance in the Sun is biased because nuclear reactions tend to destroy it by converting it into helium-3.

(iv) The spectral lines of deuterium are almost identical with those of hydrogen, so deuterium signatures tend to get washed out in spectra of primordial gas clouds.

(v) The deuterium abundance is so small (a few parts per million) that it can be easily changed by astrophysical processes other than primordial nucleosynthesis.

(c) *(5 points)* Give three examples of hadrons.

(d) *(6 points)* In Chapter 6 of *The First Three Minutes*, Steven Weinberg posed the question, "Why was there no systematic search for this [cosmic background] radiation, years before 1965?" In discussing this issue, he contrasted it with the history of two different elementary particles, each of which were predicted approximately 20 years before they were first detected. Name one of these two elementary particles. (If you name them both correctly, you will get 3 points extra credit. However, one right and one wrong will get you 4 points for the question, compared to 6 points for just naming one particle and getting it right.)

Answer: ______________________

2nd Answer (optional): ______________________

(e) *(6 points)* In Chapter 6 of *The First Three Minutes*, Steven Weinberg discusses three reasons why the importance of a search for a 3° K microwave radiation background was not generally appreciated in the 1950s and early 1960s. Choose those three reasons from the following list. (2 points for each right answer, circle at most 3.)

(i) The earliest calculations erroneously predicted a cosmic background temperature of only about 0.1° K, and such a background would be too weak to detect.

(ii) There was a breakdown in communication between theorists and experimentalists.

(iii) It was not technologically possible to detect a signal as weak as a 3° K microwave background until about 1965.

(iv) Since almost all physicists at the time were persuaded by the steady state model, the predictions of the big bang model were not taken seriously.

(v) It was extraordinarily difficult for physicists to take seriously *any* theory of the early universe.

(vi) The early work on nucleosynthesis by Gamow, Alpher, Herman, and Follin, et al., had attempted to explain the origin of all complex nuclei by reactions in the early universe. This program was never very successful, and its credibility was further undermined as improvements were made in the alternative theory, that elements are synthesized in stars.

PROBLEM 2: DID YOU DO THE READING (2007)? *(24 points)*

The following problem was Problem 1 of Quiz 2 in 2007.

(a) *(6 points)* In 1948 Ralph A. Alpher and Robert Herman wrote a paper predicting a cosmic microwave background with a temperature of 5 K. The paper was based on a cosmological model that they had developed with George Gamow, in which the early universe was assumed to have been filled with hot neutrons. As the universe expanded and cooled the neutrons underwent beta decay into protons, electrons, and antineutrinos, until at some point the universe cooled enough for light elements to be synthesized. Alpher and Herman found that to account for the observed present abundances of light elements, the ratio of photons to nuclear particles must have been about 10^9. Although the predicted temperature was very close to the actual value of 2.7 K, the theory differed from our present theory in two ways. Circle the two correct statements in the following list. (3 points for each right answer; circle at most 2.)

(i) Gamow, Alpher, and Herman assumed that the neutron could decay, but now the neutron is thought to be absolutely stable.

(ii) In the current theory, the universe started with nearly equal densities of protons and neutrons, not all neutrons as Gamow, Alpher, and Herman assumed.

(iii) In the current theory, the universe started with mainly alpha particles, not all neutrons as Gamow, Alpher, and Herman assumed. (Note: an alpha particle is the nucleus of a helium atom, composed of two protons and two neutrons.)

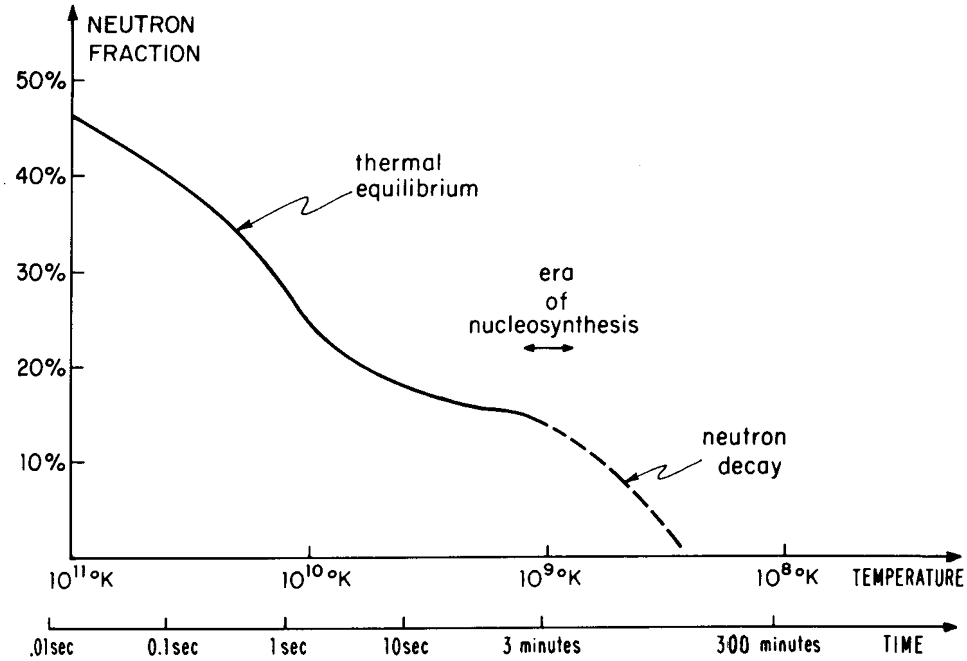
(iv) In the current theory, the conversion of neutrons into protons (and vice versa) took place mainly through collisions with electrons, positrons, neutrinos, and antineutrinos, not through the decay of the neutrons.

(v) The ratio of photons to nuclear particles in the early universe is now believed to have been about 10^3, not 10^9 as Alpher and Herman concluded.

(b) *(6 points)* In Weinberg's "Recipe for a Hot Universe," he described the primordial composition of the universe in terms of three conserved quantities: electric charge, baryon number, and lepton number. If electric charge is measured in units of the electron charge, then all three quantities are integers for which the number density can be compared with the number density of photons. For each quantity, which choice most accurately describes the initial ratio of the number density of this quantity to the number density of photons:

Electric Charge:	(i) $\sim 10^9$	(ii) ~ 1000	(iii) ~ 1
	(iv) $\sim 10^{-6}$	(v) either zero or negligible	
Baryon Number:	(i) $\sim 10^{-20}$	(ii) $\sim 10^{-9}$	(iii) $\sim 10^{-6}$
	(iv) ~ 1	(v) anywhere from 10^{-5} to 1	
Lepton Number:	(i) $\sim 10^9$	(ii) ~ 1000	(iii) ~ 1
	(iv) $\sim 10^{-6}$	(v) could be as high as ~ 1, but is assumed to be very small	

(c) *(12 points)* The figure below comes from Weinberg's Chapter 5, and is labeled *The Shifting Neutron-Proton Balance.*



(i) *(3 points)* During the period labeled "thermal equilibrium," the neutron fraction is changing because (choose one):

(A) The neutron is unstable, and decays into a proton, electron, and antineutrino with a lifetime of about 1 second.

(B) The neutron is unstable, and decays into a proton, electron, and antineutrino with a lifetime of about 15 seconds.

(C) The neutron is unstable, and decays into a proton, electron, and antineutrino with a lifetime of about 15 minutes.

(D) Neutrons and protons can be converted from one into through reactions such as

$$\text{antineutrino} + \text{proton} \longleftrightarrow \text{electron} + \text{neutron}$$
$$\text{neutrino} + \text{neutron} \longleftrightarrow \text{positron} + \text{proton.}$$

(E) Neutrons and protons can be converted from one into the other through reactions such as

$$\text{antineutrino} + \text{proton} \longleftrightarrow \text{positron} + \text{neutron}$$
$$\text{neutrino} + \text{neutron} \longleftrightarrow \text{electron} + \text{proton.}$$

(F) Neutrons and protons can be created and destroyed by reactions such as

$$\text{proton} + \text{neutrino} \longleftrightarrow \text{positron} + \text{antineutrino}$$
$$\text{neutron} + \text{antineutrino} \longleftrightarrow \text{electron} + \text{positron.}$$

(ii) *(3 points)* During the period labeled "neutron decay," the neutron fraction is changing because (choose one):

(A) The neutron is unstable, and decays into a proton, electron, and antineutrino with a lifetime of about 1 second.

(B) The neutron is unstable, and decays into a proton, electron, and antineutrino with a lifetime of about 15 seconds.

(C) The neutron is unstable, and decays into a proton, electron, and antineutrino with a lifetime of about 15 minutes.

(D) Neutrons and protons can be converted from one into the other through reactions such as

$$\begin{gathered}\text{antineutrino} + \text{proton} \longleftrightarrow \text{electron} + \text{neutron} \\ \text{neutrino} + \text{neutron} \longleftrightarrow \text{positron} + \text{proton.}\end{gathered}$$

(E) Neutrons and protons can be converted from one into the other through reactions such as

$$\begin{gathered}\text{antineutrino} + \text{proton} \longleftrightarrow \text{positron} + \text{neutron} \\ \text{neutrino} + \text{neutron} \longleftrightarrow \text{electron} + \text{proton.}\end{gathered}$$

(F) Neutrons and protons can be created and destroyed by reactions such as

$$\begin{gathered}\text{proton} + \text{neutrino} \longleftrightarrow \text{positron} + \text{antineutrino} \\ \text{neutron} + \text{antineutrino} \longleftrightarrow \text{electron} + \text{positron.}\end{gathered}$$

(iii) *(3 points)* The masses of the neutron and proton are not exactly equal, but instead

(A) The neutron is more massive than a proton with a rest energy difference of 1.293 GeV (1 GeV $= 10^9$ eV).

(B) The neutron is more massive than a proton with a rest energy difference of 1.293 MeV (1 MeV $= 10^6$ eV).

(C) The neutron is more massive than a proton with a rest energy difference of 1.293 KeV (1 KeV $= 10^3$ eV).

(D) The proton is more massive than a neutron with a rest energy difference of 1.293 GeV.

(E) The proton is more massive than a neutron with a rest energy difference of 1.293 MeV.

(F) The proton is more massive than a neutron with a rest energy difference of 1.293 KeV.

(iv) *(3 points)* During the period labeled "era of nucleosynthesis," (choose one:)

(A) Essentially all the neutrons present combine with protons to form helium nuclei, which mostly survive until the present time.

(B) Essentially all the neutrons present combine with protons to form deuterium nuclei, which mostly survive until the present time.

(C) About half the neutrons present combine with protons to form helium nuclei, which mostly survive until the present time, and the other half of the neutrons remain free.

(D) About half the neutrons present combine with protons to form deuterium nuclei, which mostly survive until the present time, and the other half of the neutrons remain free.

(E) Essentially all the protons present combine with neutrons to form helium nuclei, which mostly survive until the present time.

(F) Essentially all the protons present combine with neutrons to form deuterium nuclei, which mostly survive until the present time.

PROBLEM 3: DID YOU DO THE READING (2011)? *(20 points)*

The following problem comes from Quiz 2, 2011.

(a) *(8 points)* During nucleosynthesis, heavier nuclei form from protons and neutrons through a series of two particle reactions.

(i) In *The First Three Minutes,* Weinberg discusses two chains of reactions that, starting from protons and neutrons, end up with helium, He^4. Describe at least one of these two chains.

(ii) Explain briefly what is the *deuterium bottleneck*, and what is its role during nucleosynthesis.

(b) *(12 points)* In Chapter 4 of *The First Three Minutes*, Steven Weinberg makes the following statement regarding the radiation-dominated phase of the early universe:

The time that it takes for the universe to cool from one temperature to another is proportional to the difference of the inverse squares of these temperatures.

In this part of the problem you will explore more quantitatively this statement.

(i) For a radiation-dominated universe the scale-factor $a(t) \propto t^{1/2}$. Find the cosmic time t as a function of the Hubble expansion rate H.

(ii) The mass density stored in radiation ρ_r is proportional to the temperature T to the fourth power: i.e., $\rho_r \simeq \alpha T^4$, for some constant α. For a wide range of temperatures we can take $\alpha \simeq 4.52 \times 10^{-32}\,\mathrm{kg} \cdot \mathrm{m}^{-3} \cdot \mathrm{K}^{-4}$. If the temperature is measured in degrees Kelvin (K), then ρ_r has the standard SI units, $[\rho_r] = \mathrm{kg}\cdot\mathrm{m}^{-3}$. Use the Friedmann equation for a flat universe ($k = 0$) with $\rho = \rho_r$ to express the Hubble expansion rate H in terms of the temperature T. You will need the SI value of the gravitational constant $G \simeq 6.67 \times 10^{-11}\,\mathrm{N}\cdot\mathrm{m}^2\cdot\mathrm{kg}^{-2}$. What is the Hubble expansion rate, in inverse seconds, at the start of nucleosynthesis, when $T = T_{\rm nucl} \simeq 0.9 \times 10^9\,\mathrm{K}$?

(iii) Using the results in (i) and (ii), express the cosmic time t as a function of the temperature. Your result should agree with Weinberg's claim above. What is the cosmic time, in seconds, when $T = T_{\rm nucl}$?

PROBLEM 4: DID YOU DO THE READING (2013)? *(25 points)*

The following problem comes from Quiz 2, 2013.

(a) *(6 points)* The primary evidence for dark matter in galaxies comes from measuring their rotation curves, i.e., the orbital velocity v as a function of radius R. If stars contributed all, or most, of the mass in a galaxy, what would we expect for the behavior of $v(R)$ at large radii?

(b) *(5 points)* What is actually found for the behavior of $v(R)$?

(c) *(7 points)* An important tool for estimating the mass in a galaxy is the steady-state virial theorem. What does this theorem state?

(d) *(7 points)* At the end of Chapter 10, Ryden writes "Thus, the very strong asymmetry between baryons and antibaryons today and the large number of photons per baryon are both products of a tiny asymmetry between quarks and anitquarks in the early universe." Explain in one or a few sentences how a tiny asymmetry between quarks and anitquarks in the early universe results in a strong asymmetry between baryons and antibaryons today.

PROBLEM 5: DID YOU DO THE READING (2016)? *(25 points)*

(a) *(5 points)* In Chapter 8 of Barbara Ryden's *Introduction to Cosmology*, she estimates the contribution to Ω from clusters of galaxies as

(i) 0.01 (ii) 0.05 (iii) 0.20 (iv) 0.60 (v) 1.00

(b) *(4 points)* One method of estimating the total mass of a cluster of galaxies is based on the virial theorem. With this method, one estimates the mass by measuring

(i) the radius containing half the luminosity and also the temperature of the X-ray emitting gas at the center of the galaxy.

(ii) the velocity dispersion perpendicular to the line of sight and also the radius containing half of the luminosity of the cluster.

(iii) the velocity dispersion along the line of sight and also the radius containing half of the luminosity of the cluster.

(iv) the velocity dispersion along the line of sight and also the redshift of the cluster.

(v) the velocity dispersion perpendicular to the line of sight and also the redshift of the cluster.

(c) *(4 points)* Another method of estimating the total mass of a cluster of galaxies is to make detailed measurements of the x-rays emitted by the hot intracluster gas.

(i) By assuming that this gas is the dominant component of the mass of the cluster, the mass of the cluster can be estimated.

(ii) By assuming that the hot gas comprises about a third of the mass of the cluster, the total mass of the cluster can be estimated.

(iii) By assuming that the gas is heated by stars and supernovae that make up most of the mass of the cluster, the mass of these stars and supernovae can be estimated.

(iv) By assuming that the gas is heated by interactions with dark matter, which dominates the mass of the cluster, the mass of the cluster can be estimated.

(v) By assuming that this gas is in hydrostatic equilibrium, the temperature, mass density, and even the chemical composition of the cluster can be modeled.

(d) *(6 points)* In Chapter 6 of *The First Three Minutes*, Steven Weinberg discusses three reasons why the importance of a search for a 3° K microwave radiation background was not generally appreciated in the 1950s and early 1960s. Choose those three reasons from the following list. (2 points for each right answer, circle at most 3.)

(i) The earliest calculations erroneously predicted a cosmic background temperature of only about 0.1° K, and such a background would be too weak to detect.

(ii) There was a breakdown in communication between theorists and experimentalists.

(iii) It was not technologically possible to detect a signal as weak as a 3° K microwave background until about 1965.

(iv) Since almost all physicists at the time were persuaded by the steady state model, the predictions of the big bang model were not taken seriously.

(v) It was extraordinarily difficult for physicists to take seriously *any* theory of the early universe.

(vi) The early work on nucleosynthesis by Gamow, Alpher, Herman, and Follin, et al., had attempted to explain the origin of all complex nuclei by reactions in the early universe. This program was never very successful, and its credibility was further undermined as improvements were made in the alternative theory, that elements are synthesized in stars.

(e) *(6 points)* In 1948 Ralph A. Alpher and Robert Herman wrote a paper predicting a cosmic microwave background with a temperature of 5 K. The paper was based on a cosmological model that they had developed with George Gamow, in which the early universe was assumed to have been filled with hot neutrons. As the universe expanded and cooled the neutrons underwent beta decay into protons, electrons, and antineutrinos, until at some point the universe cooled enough for light elements to be synthesized. Alpher and Herman found that to account for the observed present abundances of light elements, the ratio of photons to nuclear particles must have been about 10^9. Although the predicted temperature was very close to the actual value of 2.7 K, the theory differed from our present theory in two ways. Circle the two correct statements in the following list. (3 points for each right answer; circle at most 2.)

(i) Gamow, Alpher, and Herman assumed that the neutron could decay, but now the neutron is thought to be absolutely stable.

(ii) In the current theory, the universe started with nearly equal densities of protons and neutrons, not all neutrons as Gamow, Alpher, and Herman assumed.

(iii) In the current theory, the universe started with mainly alpha particles, not all neutrons as Gamow, Alpher, and Herman assumed. (Note: an alpha particle is the nucleus of a helium atom, composed of two protons and two neutrons.)

(iv) In the current theory, the conversion of neutrons into protons (and vice versa) took place mainly through collisions with electrons, positrons, neutrinos, and antineutrinos, not through the decay of the neutrons.

(v) The ratio of photons to nuclear particles in the early universe is now believed to have been about 10^3, not 10^9 as Alpher and Herman concluded.

∗ PROBLEM 6: EVOLUTION OF AN OPEN UNIVERSE

The following problem was taken from Quiz 2, 1990, where it counted 10 points out of 100.

Consider an open, matter-dominated universe, as described by the evolution equations on the front of the quiz. Find the time t at which $a/\sqrt{\kappa} = 2\alpha$.

∗ PROBLEM 7: ANTICIPATING A BIG CRUNCH

Suppose that we lived in a closed, matter-dominated universe, as described by the equations on the front of the quiz. Suppose further that we measured the mass density parameter Ω to be $\Omega_0 = 2$, and we measured the Hubble "constant" to have some value H_0. How much time would we have before our universe ended in a big crunch, at which time the scale factor $a(t)$ would collapse to 0?

∗ PROBLEM 8: TRACING LIGHT RAYS IN A CLOSED, MATTER-DOMINATED UNIVERSE *(30 points)*

The following problem was Problem 3, Quiz 2, 1998.

The spacetime metric for a homogeneous, isotropic, closed universe is given by the Robertson-Walker formula:

$$ds^2 = -c^2\,d\tau^2 = -c^2\,dt^2 + a^2(t)\left\{\frac{dr^2}{1-r^2} + r^2\left(d\theta^2 + \sin^2\theta\,d\phi^2\right)\right\} \ ,$$

where I have taken $k = 1$. To discuss motion in the radial direction, it is more convenient to work with an alternative radial coordinate ψ, related to r by

$$r = \sin\psi \ .$$

Then

$$\frac{dr}{\sqrt{1-r^2}} = d\psi \ ,$$

so the metric simplifies to

$$ds^2 = -c^2\,d\tau^2 = -c^2\,dt^2 + a^2(t)\left\{d\psi^2 + \sin^2\psi\left(d\theta^2 + \sin^2\theta\,d\phi^2\right)\right\} \ .$$

(a) *(7 points)* A light pulse travels on a null trajectory, which means that $d\tau = 0$ for each segment of the trajectory. Consider a light pulse that moves along a radial line, so $\theta = \phi =$ constant. Find an expression for $d\psi/dt$ in terms of quantities that appear in the metric.

(b) *(8 points)* Write an expression for the physical horizon distance $\ell_{\rm phys}$ at time t. You should leave your answer in the form of a definite integral.

The form of $a(t)$ depends on the content of the universe. If the universe is matter-dominated (*i.e.,* dominated by nonrelativistic matter), then $a(t)$ is described by the parametric equations

$$ct = \alpha(\theta - \sin\theta) \ ,$$
$$a = \alpha(1 - \cos\theta) \ ,$$

where

$$\alpha \equiv \frac{4\pi}{3} \frac{G\rho a^3}{c^2} \ .$$

These equations are identical to those on the front of the exam, except that I have chosen $k = 1$.

(c) *(10 points)* Consider a radial light-ray moving through a matter-dominated closed universe, as described by the equations above. Find an expression for $d\psi/d\theta$, where θ is the parameter used to describe the evolution.

(d) *(5 points)* Suppose that a photon leaves the origin of the coordinate system ($\psi = 0$) at $t = 0$. How long will it take for the photon to return to its starting place? Express your answer as a fraction of the full lifetime of the universe, from big bang to big crunch.

PROBLEM 9: LENGTHS AND AREAS IN A TWO-DIMENSIONAL METRIC *(25 points)*

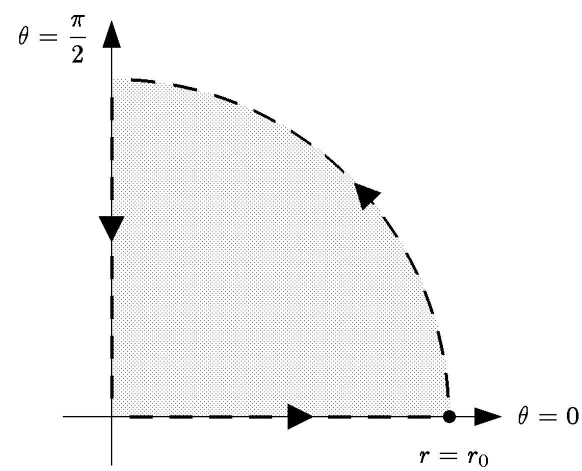
The following problem was Problem 3, Quiz 2, 1994:

Suppose a two dimensional space, described in polar coordinates (r, θ), has a metric given by

$$ds^2 = (1 + ar)^2\, dr^2 + r^2(1 + br)^2\, d\theta^2 \ ,$$

where a and b are positive constants. Consider the path in this space which is formed by starting at the origin, moving along the $\theta = 0$ line to $r = r_0$, then moving at fixed r to

$\theta = \pi/2$, and then moving back to the origin at fixed θ. The path is shown below:



a) *(10 points)* Find the total length of this path.

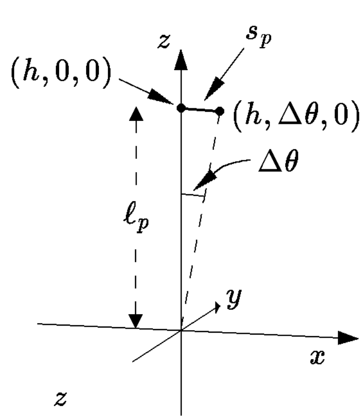
b) *(15 points)* Find the area enclosed by this path.

PROBLEM 10: GEOMETRY IN A CLOSED UNIVERSE *(25 points)*

The following problem was Problem 4, Quiz 2, 1988:

Consider a universe described by the Robertson–Walker metric on the first page of the quiz, with $k = 1$. The questions below all pertain to some fixed time t, so the scale factor can be written simply as a, dropping its explicit t-dependence.

A small rod has one end at the point ($r = h$, $\theta = 0$, $\phi = 0$) and the other end at the point ($r = h$, $\theta = \Delta\theta$, $\phi = 0$). Assume that $\Delta\theta \ll 1$.



(a) Find the physical distance ℓ_p from the origin ($r = 0$) to the first end $(h, 0, 0)$ of the rod. You may find one of the following integrals useful:

$$\int \frac{dr}{\sqrt{1 - r^2}} = \sin^{-1} r$$

$$\int \frac{dr}{1 - r^2} = \frac{1}{2} \ln \left(\frac{1 + r}{1 - r} \right) .$$

(b) Find the physical length s_p of the rod. Express your answer in terms of the scale factor a, and the coordinates h and $\Delta\theta$.

(c) Note that $\Delta\theta$ is the angle subtended by the rod, as seen from the origin. Write an expression for this angle in terms of the physical distance ℓ_p, the physical length s_p, and the scale factor a.

PROBLEM 11: THE GENERAL SPHERICALLY SYMMETRIC METRIC

(20 points)

The following problem was Problem 3, Quiz 2, 1986:

The metric for a given space depends of course on the coordinate system which is used to describe it. It can be shown that for any three dimensional space which is spherically symmetric about a particular point, coordinates can be found so that the metric has the form

$$ds^2 = dr^2 + \rho^2(r) \left[d\theta^2 + \sin^2 \theta \, d\phi^2 \right]$$

for some function $\rho(r)$. The coordinates θ and ϕ have their usual ranges: θ varies between 0 and π, and ϕ varies from 0 to 2π, where $\phi = 0$ and $\phi = 2\pi$ are identified. Given this metric, consider the sphere whose outer boundary is defined by $r = r_0$.

(a) Find the physical radius a of the sphere. (By "radius", I mean the physical length of a radial line which extends from the center to the boundary of the sphere.)

(b) Find the physical area of the surface of the sphere.

(c) Find an explicit expression for the volume of the sphere. Be sure to include the limits of integration for any integrals which occur in your answer.

(d) Suppose a new radial coordinate σ is introduced, where σ is related to r by

$$\sigma = r^2 .$$

Express the metric in terms of this new variable.

PROBLEM 12: VOLUMES IN A ROBERTSON-WALKER UNIVERSE *(20 points)*

The following problem was Problem 1, Quiz 3, 1990:

The metric for a Robertson-Walker universe is given by

$$ds^2 = a^2(t)\left\{\frac{dr^2}{1-kr^2} + r^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right)\right\} \quad .$$

Calculate the volume $V(r_{\max})$ of the sphere described by

$$r \leq r_{\max} \ .$$

You should carry out any angular integrations that may be necessary, but you may leave your answer in the form of a radial integral which is not carried out. Be sure, however, to clearly indicate the limits of integration.

*** PROBLEM 13: THE SCHWARZSCHILD METRIC** *(25 points)*

The follow problem was Problem 4, Quiz 3, 1992:

The space outside a spherically symmetric mass M is described by the Schwarzschild metric, given at the front of the exam. Two observers, designated A and B, are located along the same radial line, with values of the coordinate r given by r_A and r_B, respectively, with $r_A < r_B$. You should assume that both observers lie outside the Schwarzschild horizon.

a) *(5 points)* Write down the expression for the Schwarzschild horizon radius R_S, expressed in terms of M and fundamental constants.

b) *(5 points)* What is the proper distance between A and B? It is okay to leave the answer to this part in the form of an integral that you do not evaluate— but be sure to clearly indicate the limits of integration.

c) *(5 points)* Observer A has a clock that emits an evenly spaced sequence of ticks, with proper time separation $\Delta\tau_A$. What will be the coordinate time separation Δt_A between these ticks?

d) *(5 points)* At each tick of A's clock, a light pulse is transmitted. Observer B receives these pulses, and measures the time separation on his own clock. What is the time interval $\Delta\tau_B$ measured by B.

e) *(5 points)* Suppose that the object creating the gravitational field is a static black hole, so the Schwarzschild metric is valid for all r. Now suppose that one considers the case in which observer A lies on the Schwarzschild horizon, so $r_A \equiv R_S$. Is the proper distance between A and B finite for this case? Does the time interval of the pulses received by B, $\Delta\tau_B$, diverge in this case?

PROBLEM 14: GEODESICS *(20 points)*

The following problem was Problem 4, Quiz 2, 1986:

Ordinary Euclidean two-dimensional space can be described in polar coordinates by the metric

$$ds^2 = dr^2 + r^2\, d\theta^2 \ .$$

(a) Suppose that $r(\lambda)$ and $\theta(\lambda)$ describe a geodesic in this space, where the parameter λ is the arc length measured along the curve. Use the general formula on the front of the exam to obtain explicit differential equations which $r(\lambda)$ and $\theta(\lambda)$ must obey.

(b) Now introduce the usual Cartesian coordinates, defined by

$$x = r\cos\theta \ ,$$

$$y = r\sin\theta \ .$$

Use your answer to (a) to show that the line $y = 1$ is a geodesic curve.

*** PROBLEM 15: AN EXERCISE IN TWO-DIMENSIONAL METRICS** *(30 points)*

(a) *(8 points)* Consider first a two-dimensional space with coordinates r and θ. The metric is given by

$$\mathrm{d}s^2 = \mathrm{d}r^2 + r^2 \mathrm{d}\theta^2 \ .$$

Consider the curve described by

$$r(\theta) = (1 + \epsilon\cos^2\theta)\, r_0 \ ,$$

where ϵ and r_0 are constants, and θ runs from θ_1 to θ_2. Write an expression, in the form of a definite integral, for the length S of this curve.

(b) *(5 points)* Now consider a two-dimensional space with the same two coordinates r and θ, but this time the metric will be

$$\mathrm{d}s^2 = \left(1 + \frac{r}{a}\right) \mathrm{d}r^2 + r^2\, \mathrm{d}\theta^2 \ ,$$

where a is a constant. θ is a periodic (angular) variable, with a range of 0 to 2π, with 2π identified with 0. What is the length R of the path from the origin ($r = 0$) to the point $r = r_0, \theta = 0$, along the path for which $\theta = 0$ everywhere along the path? You can leave your answer in the form of a definite integral. (Be sure, however, to specify the limits of integration.)

(c) *(7 points)* For the space described in part (b), what is the total area contained within the region $r < r_0$. Again you can leave your answer in the form of a definite integral, making sure to specify the limits of integration.

(d) *(10 points)* Again for the space described in part (b), consider a geodesic described by the usual geodesic equation,

$$\frac{\mathrm{d}}{\mathrm{d}s}\left\{g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}s}\right\} = \frac{1}{2}\left(\partial_i g_{k\ell}\right)\frac{\mathrm{d}x^k}{\mathrm{d}s}\frac{\mathrm{d}x^\ell}{\mathrm{d}s} .$$

The geodesic is described by functions $r(s)$ and $\theta(s)$, where s is the arc length along the curve. Write explicitly both (i.e., for i=1=r and i=2=θ) geodesic equations.

PROBLEM 16: GEODESICS ON THE SURFACE OF A SPHERE

In this problem we will test the geodesic equation by computing the geodesic curves on the surface of a sphere. We will describe the sphere as in Lecture Notes 5, with metric given by

$$ds^2 = a^2\left(d\theta^2 + \sin^2\theta\, d\phi^2\right) .$$

(a) Clearly one geodesic on the sphere is the equator, which can be parametrized by $\theta = \pi/2$ and $\phi = \psi$, where ψ is a parameter which runs from 0 to 2π. Show that if the equator is rotated by an angle α about the x-axis, then the equations become:

$$\cos\theta = \sin\psi\sin\alpha$$
$$\tan\phi = \tan\psi\cos\alpha .$$

(b) Using the generic form of the geodesic equation on the front of the exam, derive the differential equation which describes geodesics in this space.

(c) Show that the expressions in (a) satisfy the differential equation for the geodesic. Hint: The algebra on this can be messy, but I found things were reasonably simple if I wrote the derivatives in the following way:

$$\frac{d\theta}{d\psi} = -\frac{\cos\psi\sin\alpha}{\sqrt{1-\sin^2\psi\sin^2\alpha}} , \qquad \frac{d\phi}{d\psi} = \frac{\cos\alpha}{1-\sin^2\psi\sin^2\alpha} .$$

∗ PROBLEM 17: GEODESICS IN A CLOSED UNIVERSE

The following problem was Problem 3, Quiz 3, 2000, where it was worth 40 points plus 5 points extra credit.

Consider the case of closed Robertson-Walker universe. Taking $k = 1$, the spacetime metric can be written in the form

$$ds^2 = -c^2\,d\tau^2 = -c^2\,dt^2 + a^2(t)\left\{\frac{dr^2}{1-r^2} + r^2\left(d\theta^2 + \sin^2\theta\,d\phi^2\right)\right\} .$$

We will assume that this metric is given, and that $a(t)$ has been specified. While galaxies are approximately stationary in the comoving coordinate system described by this metric, we can still consider an object that moves in this system. In particular, in this problem we will consider an object that is moving in the radial direction (r-direction), under the influence of no forces other than gravity. Hence the object will travel on a geodesic.

(a) *(7 points)* Express $d\tau/dt$ in terms of dr/dt.

(b) *(3 points)* Express $dt/d\tau$ in terms of dr/dt.

(c) *(10 points)* If the object travels on a trajectory given by the function $r_p(t)$ between some time t_1 and some later time t_2, write an integral which gives the total amount of time that a clock attached to the object would record for this journey.

(d) *(10 points)* During a time interval dt, the object will move a coordinate distance

$$dr = \frac{dr}{dt}dt .$$

Let $d\ell$ denote the physical distance that the object moves during this time. By "physical distance," I mean the distance that would be measured by a comoving observer (an observer stationary with respect to the coordinate system) who is located at the same point. The quantity $d\ell/dt$ can be regarded as the physical speed v_{phys} of the object, since it is the speed that would be measured by a comoving observer. Write an expression for v_{phys} as a function of dr/dt and r.

(e) *(10 points)* Using the formulas at the front of the exam, derive the geodesic equation of motion for the coordinate r of the object. Specifically, you should derive an equation of the form

$$\frac{d}{d\tau}\left[A\frac{dr}{d\tau}\right] = B\left(\frac{dt}{d\tau}\right)^2 + C\left(\frac{dr}{d\tau}\right)^2 + D\left(\frac{d\theta}{d\tau}\right)^2 + E\left(\frac{d\phi}{d\tau}\right)^2 ,$$

where A, B, C, D, and E are functions of the coordinates, some of which might be zero.

(f) *(5 points EXTRA CREDIT)* On Problem 1 of Problem Set 6 we learned that in a flat Robertson-Walker metric, the relativistically defined momentum of a particle,

$$p = \frac{mv_{\text{phys}}}{\sqrt{1 - \frac{v_{\text{phys}}^2}{c^2}}} ,$$

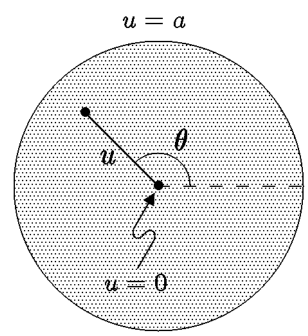
falls off as $1/a(t)$. Use the geodesic equation derived in part (e) to show that the same is true in a closed universe.

PROBLEM 18: A TWO-DIMENSIONAL CURVED SPACE *(40 points)*

The following problem was Problem 3, Quiz 2, 2002.

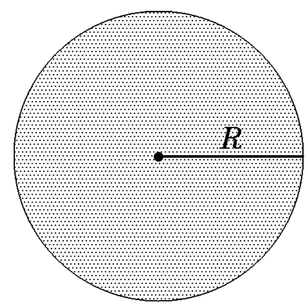
Consider a two-dimensional curved space described by polar coordinates u and θ, where $0 \leq u \leq a$ and $0 \leq \theta \leq 2\pi$, and $\theta = 2\pi$ is as usual identified with $\theta = 0$. The metric is given by

$$ds^2 = \frac{a\,du^2}{4u(a-u)} + u\,d\theta^2\ .$$

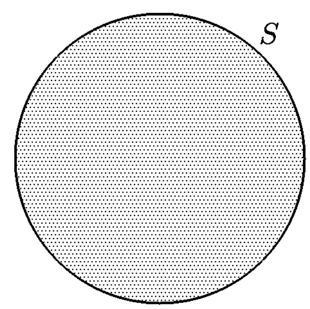


A diagram of the space is shown at the right, but you should of course keep in mind that the diagram does not accurately reflect the distances defined by the metric.

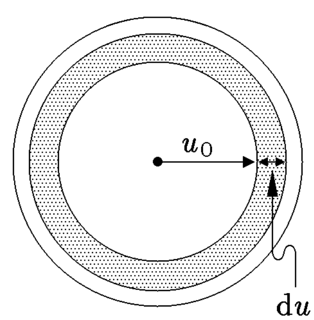
(a) *(6 points)* Find the radius R of the space, defined as the length of a radial (i.e., $\theta = $ *constant*) line. You may express your answer as a definite integral, which you need not evaluate. Be sure, however, to specify the limits of integration.



(b) *(6 points)* Find the circumference S of the space, defined as the length of the boundary of the space at $u = a$.



(c) *(7 points)* Consider an annular region as shown, consisting of all points with a u-coordinate in the range $u_0 \leq u \leq u_0 + du$. Find the physical area dA of this region, to first order in du.



(d) *(3 points)* Using your answer to part (c), write an expression for the total area of the space.

(e) *(10 points)* Consider a geodesic curve in this space, described by the functions $u(s)$ and $\theta(s)$, where the parameter s is chosen to be the arc length along the curve. Find the geodesic equation for $u(s)$, which should have the form

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[F(u,\theta)\,\frac{\mathrm{d}u}{\mathrm{d}s}\right] = \ldots\ ,$$

where $F(u,\theta)$ is a function that you will find. (Note that by writing F as a function of u and θ, we are saying that it *could* depend on either or both of them, but we are not saying that it *necessarily* depends on them.) You need not simplify the left-hand side of the equation.

(f) *(8 points)* Similarly, find the geodesic equation for $\theta(s)$, which should have the form

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[G(u,\theta)\,\frac{\mathrm{d}\theta}{\mathrm{d}s}\right] = \ldots\ ,$$

where $G(u,\theta)$ is a function that you will find. Again, you need not simplify the left-hand side of the equation.

∗ PROBLEM 19: ROTATING FRAMES OF REFERENCE *(35 points)*

The following problem was Problem 3, Quiz 2, 2004.

In this problem we will use the formalism of general relativity and geodesics to derive the relativistic description of a rotating frame of reference.

The problem will concern the consequences of the metric

$$ds^2 = -c^2\,\mathrm{d}\tau^2 = -c^2\,\mathrm{d}t^2 + \left[\mathrm{d}r^2 + r^2\,(\mathrm{d}\phi + \omega\,\mathrm{d}t)^2 + \mathrm{d}z^2\right]\ , \qquad \text{(P19.1)}$$

which corresponds to a coordinate system rotating about the z-axis, where ϕ is the azimuthal angle around the z-axis. The coordinates have the usual range for cylindrical coordinates: $-\infty < t < \infty$, $0 \le r < \infty$, $-\infty < z < \infty$, and $0 \le \phi < 2\pi$, where $\phi = 2\pi$ is identified with $\phi = 0$.

EXTRA INFORMATION

To work the problem, you do not need to know anything about where this metric came from. However, it might (or might not!) help your intuition to know that Eq. (P19.1) was obtained by starting with a Minkowski metric in cylindrical coordinates $\bar{t}$, $\bar{r}$, $\bar{\phi}$, and $\bar{z}$,

$$c^2\,\mathrm{d}\tau^2 = c^2\,\mathrm{d}\bar{t}^{\,2} - \left[\mathrm{d}\bar{r}^2 + \bar{r}^2\,\mathrm{d}\bar{\phi}^2 + \mathrm{d}\bar{z}^2\right]\ ,$$

and then introducing new coordinates t, r, ϕ, and z that are related by

$$\bar{t} = t, \qquad \bar{r} = r, \quad \bar{\phi} = \phi + \omega t, \quad \bar{z} = z\ ,$$

so $\mathrm{d}\bar{t} = \mathrm{d}t$, $\mathrm{d}\bar{r} = \mathrm{d}r$, $\mathrm{d}\bar{\phi} = \mathrm{d}\phi + \omega\,\mathrm{d}t$, and $\mathrm{d}\bar{z} = \mathrm{d}z$.

(a) *(8 points)* The metric can be written in matrix form by using the standard definition

$$ds^2 = -c^2\,\mathrm{d}\tau^2 \equiv g_{\mu\nu}\,dx^\mu\,dx^\nu \ ,$$

where $x^0 \equiv t$, $x^1 \equiv r$, $x^2 \equiv \phi$, and $x^3 \equiv z$. Then, for example, g_{11} (which can also be called g_{rr}) is equal to 1. Find explicit expressions to complete the list of the nonzero entries in the matrix $g_{\mu\nu}$:

$$\begin{aligned} g_{11} &\equiv g_{rr} = 1 \\ g_{00} &\equiv g_{tt} = ? \\ g_{20} &\equiv g_{02} \equiv g_{\phi t} \equiv g_{t\phi} = ? \\ g_{22} &\equiv g_{\phi\phi} = ? \\ g_{33} &\equiv g_{zz} = ? \end{aligned} \tag{P19.2}$$

If you cannot answer part (a), you can introduce unspecified functions $f_1(r)$, $f_2(r)$, $f_3(r)$, and $f_4(r)$, with

$$\begin{aligned} g_{11} &\equiv g_{rr} = 1 \\ g_{00} &\equiv g_{tt} = f_1(r) \\ g_{20} &\equiv g_{02} \equiv g_{\phi t} \equiv g_{t\phi} = f_1(r) \\ g_{22} &\equiv g_{\phi\phi} = f_3(r) \\ g_{33} &\equiv g_{zz} = f_4(r) \ , \end{aligned} \tag{P19.3}$$

and you can then express your answers to the subsequent parts in terms of these unspecified functions.

(b) *(10 points)* Using the geodesic equations from the front of the quiz,

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{g_{\mu\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\right\} = \frac{1}{2}\left(\partial_\mu g_{\lambda\sigma}\right)\frac{\mathrm{d}x^\lambda}{\mathrm{d}\tau}\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau} \ ,$$

explicitly write the equation that results when the free index μ is equal to 1, corresponding to the coordinate r.

(c) *(7 points)* Explicitly write the equation that results when the free index μ is equal to 2, corresponding to the coordinate ϕ.

(d) *(10 points)* Use the metric to find an expression for $\mathrm{d}t/\mathrm{d}\tau$ in terms of $\mathrm{d}r/\mathrm{d}t$, $\mathrm{d}\phi/\mathrm{d}t$, and $\mathrm{d}z/\mathrm{d}t$. The expression may also depend on the constants c and ω. Be sure to note that your answer should depend on the derivatives of t, ϕ, and z with respect to t, not τ. *(Hint: first find an expression for* $\mathrm{d}\tau/\mathrm{d}t$*, in terms of the quantities indicated, and then ask yourself how this result can be used to find* $\mathrm{d}t/\mathrm{d}\tau$*.)*

PROBLEM 20: THE STABILITY OF SCHWARZSCHILD ORBITS *(30 points)*

This problem was Problem 4, Quiz 2 in 2007. I have modified the reference to the homework problem to correspond to the current (2016) context, where it is Problem 3 of Problem Set 6. In 2007 it had also been a homework problem prior to the quiz.

This problem is an elaboration of the Problem 3 of Problem Set 6, for which both the statement and the solution are reproduced at the end of this quiz. This material is reproduced for your reference, but you should be aware that the solution to the present problem has important differences. You can copy from this material, but to allow the grader to assess your understanding, you are expected to present a logical, self-contained answer to this question.

In the solution to that homework problem, it was stated that further analysis of the orbits in a Schwarzschild geometry shows that the smallest *stable* circular orbit occurs for $r = 3R_S$. Circular orbits are possible for $\frac{3}{2}R_S < r < 3R_S$, but they are not stable. In this problem we will explore the calculations behind this statement.

We will consider a body which undergoes small oscillations about a circular orbit at $r(t) = r_0$, $\theta = \pi/2$, where r_0 is a constant. The coordinate θ will therefore be fixed, but all the other coordinates will vary as the body follows its orbit.

(a) *(12 points)* The first step, since $r(\tau)$ will not be a constant in this solution, will be to derive the equation of motion for $r(\tau)$. That is, for the Schwarzschild metric

$$ds^2 = -c^2 d\tau^2 = -h(r)c^2 dt^2 + h(r)^{-1} dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta \, d\phi^2 \ , \qquad \text{(P20.1)}$$

where

$$h(r) \equiv 1 - \frac{R_S}{r} \ ,$$

work out the explicit form of the geodesic equation

$$\frac{d}{d\tau}\left[g_{\mu\nu}\frac{dx^\nu}{d\tau}\right] = \frac{1}{2}\frac{\partial g_{\lambda\sigma}}{\partial x^\mu}\frac{dx^\lambda}{d\tau}\frac{dx^\sigma}{d\tau} \ , \qquad \text{(P20.2)}$$

for the case $\mu = r$. You should use this result to find an explicit expression for

$$\frac{d^2 r}{d\tau^2} \ .$$

You may allow your answer to contain $h(r)$, its derivative $h'(r)$ with respect to r, and the derivative with respect to τ of any coordinate, including $dt/d\tau$.

(b) *(6 points)* It is useful to consider r and ϕ to be the independent variables, while treating t as a dependent variable. Find an expression for

$$\left(\frac{dt}{d\tau}\right)^2$$

in terms of r, $dr/d\tau$, $d\phi/d\tau$, $h(r)$, and c. Use this equation to simplify the expression for $d^2r/d\tau^2$ obtained in part (a). The goal is to obtain an expression of the form

$$\frac{d^2r}{d\tau^2} = f_0(r) + f_1(r)\left(\frac{d\phi}{d\tau}\right)^2 \ . \tag{P20.3}$$

where the functions $f_0(r)$ and $f_1(r)$ might depend on R_S or c, and might be positive, negative, or zero. Note that the intermediate steps in the calculation involve a term proportional to $(dr/d\tau)^2$, but the net coefficient for this term vanishes.

(c) *(7 points)* To understand the orbit we will also need the equation of motion for ϕ. Evaluate the geodesic equation (P20.2) for $\mu = \phi$, and write the result in terms of the quantity L, defined by

$$L \equiv r^2 \frac{d\phi}{d\tau} \ . \tag{P20.4}$$

(d) *(5 points)* Finally, we come to the question of stability. Substituting Eq. (P20.4) into Eq. (P20.3), the equation of motion for r can be written as

$$\frac{d^2r}{d\tau^2} = f_0(r) + f_1(r)\frac{L^2}{r^4} \ .$$

Now consider a small perturbation about the circular orbit at $r = r_0$, and write an equation that determines the stability of the orbit. (That is, if some external force gives the orbiting body a small kick in the radial direction, how can you determine whether the perturbation will lead to stable oscillations, or whether it will start to grow?) You should express the stability requirement in terms of the unspecified functions $f_0(r)$ and $f_1(r)$. You are NOT asked to carry out the algebra of inserting the explicit forms that you have found for these functions.

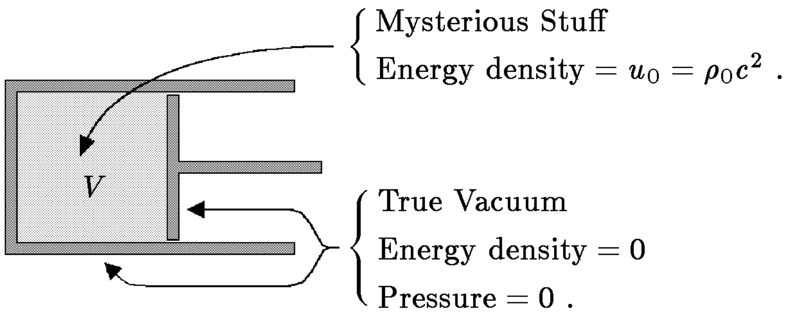
∗ PROBLEM 21: PRESSURE AND ENERGY DENSITY OF MYSTERIOUS STUFF *(25 points)*

The following problem was Problem 3, Quiz 3, 2002.

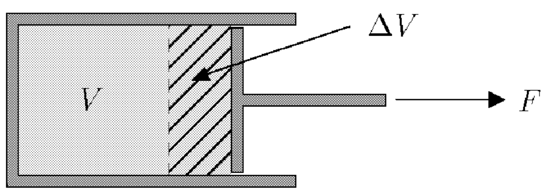
In Lecture Notes 6, with further calculations in Problem 4 of Problem Set 6, a thought experiment involving a piston was used to show that $p = \frac{1}{3}\rho c^2$ for radiation. In this problem you will apply the same technique to calculate the pressure of ***mysterious***

stuff, which has the property that the energy density falls off in proportion to $1/\sqrt{V}$ as the volume V is increased.

If the initial energy density of the mysterious stuff is $u_0 = \rho_0 c^2$, then the initial configuration of the piston can be drawn as



The piston is then pulled outward, so that its initial volume V is increased to $V + \Delta V$. You may consider ΔV to be infinitesimal, so ΔV^2 can be neglected.



(a) *(15 points)* Using the fact that the energy density of mysterious stuff falls off as $1/\sqrt{V}$, find the amount ΔU by which the energy inside the piston changes when the volume is enlarged by ΔV. Define ΔU to be positive if the energy increases.

(b) *(5 points)* If the (unknown) pressure of the mysterious stuff is called p, how much work ΔW is done by the agent that pulls out the piston?

(c) *(5 points)* Use your results from (a) and (b) to express the pressure p of the mysterious stuff in terms of its energy density u. (If you did not answer parts (a) and/or (b), explain as best you can how you would determine the pressure if you knew the answers to these two questions.)

PROBLEM 22: VOLUME OF A CLOSED THREE-DIMENSIONAL SPACE *(15 points)*

This problem is a generalization of Problem 2 of Problem Set 5.

Recall that the spatial part of the metric for a closed universe can be written as

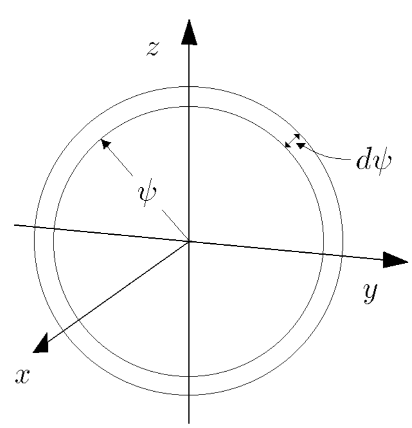
$$\mathrm{d}s^2 = R^2 \left[\mathrm{d}\psi^2 + \sin^2\psi \left(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\phi^2\right)\right] \quad .$$

In this problem we will consider a more general metric, which also describes a closed three-dimensional space, but one that is not homogeneous. The metric will be given by

$$\mathrm{d}s^2 = R^2 \left[\mathrm{d}\psi^2 + f^2(\psi)\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right] \quad ,$$

where $f(\psi)$ is some unspecified function. The coordinates θ and ϕ have the usual range, $0 \leq \theta \leq \pi$, and $0 \leq \phi \leq 2\pi$, and ψ varies in the range $0 \leq \psi \leq \pi$.

Write an integral expression for the volume of this space. The integral should be over a single variable only. Hint: as in Problem 2 of Problem Set 5, you can break the volume up into spherical shells of infinitesimal thickness, extending from ψ to $\psi + d\psi$:



PROBLEM 23: GRAVITATIONAL BENDING OF LIGHT *(30 points)*

When a light ray passes by a massive object, general relativity predicts that it will be bent. Since most celestial objects are nearly spherical, we can use the Schwarzschild metric to calculate the bending. Furthermore, since we are usually interested in objects that are not black holes or anywhere nearly as dense, we can obtain an accurate answer by carrying out the calculation in a weak-field approximation. For a photon that grazes the Sun, for example, the value of $R_{\rm Sch}/R_\odot$, the Schwarzschild radius over the radius of the Sun, is about 4×10^{-6}.

Starting with the Schwarzschild metric,

$$\mathrm{d}s^2 = -\left(1 - \frac{R_{\rm Sch}}{r}\right) c^2 \mathrm{d}t^2 + \left(1 - \frac{R_{\rm Sch}}{r}\right)^{-1} \mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\phi^2) \;, \qquad \text{(P23.1)}$$

where $R_{\rm Sch} = 2GM/c^2$, we can expand in powers of $R_{\rm Sch}/r$ and keep only the first order terms:

$$\mathrm{d}s^2 = -\left(1 - \frac{R_{\rm Sch}}{r}\right) c^2 \mathrm{d}t^2 + \left(1 + \frac{R_{\rm Sch}}{r}\right) \mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\phi^2) \;. \qquad \text{(P23.2)}$$

For this problem it is useful to switch to Cartesian-like coordinates, defined in terms of r, θ, and ϕ by the usual Cartesian formulas,

$$\begin{aligned} x &= r \,\sin\theta\, \cos\phi \;, \\ y &= r \,\sin\theta\, \sin\phi \;, \\ z &= r \cos\theta \;. \end{aligned} \qquad \text{(P23.3)}$$

General relativity allows us to make any coordinate redefinitions that we might want, as long as we calculate the metric in terms of the new coordinates. It is useful to continue to use the quantity r, but now it will be thought of as a function of the coordinates x, y, and z:

$$r = (x^2 + y^2 + z^2)^{1/2} \;. \qquad \text{(P23.4)}$$

The metric can then be rewritten as the Minkowski metric of special relativity, plus small corrections:

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 + \frac{R_{\rm Sch}}{r} c^2 \mathrm{d}t^2 + \frac{R_{\rm Sch}}{r} (\mathrm{d}r)^2 \;, \qquad \text{(P23.5)}$$

where from Eq. (4) one can see that

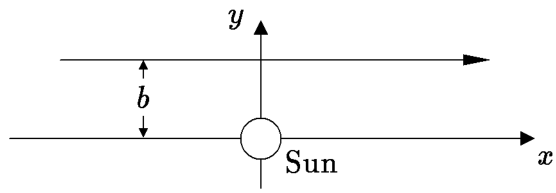
$$\mathrm{d}r = \frac{1}{r}(x\,\mathrm{d}x + y\,\mathrm{d}y + z\,\mathrm{d}z) \;. \qquad \text{(P23.6)}$$

(a) *(6 points)* For the metric as approximated by Eqs. (P23.5) and (P23.6), write the expressions for g_{tt}, g_{xx}, and g_{xy}.

The trajectory of the photon is lightlike, so we cannot use τ to parameterize the trajectory, because proper time intervals along a lightlike trajectory are zero. Nonetheless, it can be shown that one can use an "affine parameter" λ, for which the geodesic equation has the usual form:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left\{g_{\mu\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}\right\} = \frac{1}{2}\left[\partial_\mu g_{\sigma\tau}\right]\frac{\mathrm{d}x^\sigma}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\tau}{\mathrm{d}\lambda} . \tag{P23.7}$$

To obtain an answer that is accurate to first order in G, we begin by considering the *unperturbed* photon trajectory — the trajectory it would have if G were taken as zero, so $R_{\mathrm{Sch}} = 2GM/c^2 = 0$. This would be a straight line in the (x, y, z) coordinates, as shown in the diagram below:



Here b is called the *impact parameter*. We can parameterize this path by

$$x(\lambda) = \lambda , \qquad y(\lambda) = b , \qquad z(\lambda) = 0 , \qquad t(\lambda) = \lambda/c . \tag{P23.8}$$

We will calculate the deflection (to first order in G) by assuming that the photon path is accurately described by Eq. (P23.8), and we will calculate the y-velocity that the photon acquires due to the gravitational attraction of the Sun.

(b) *(9 points)* With the goal of calculating $\mathrm{d}^2y/\mathrm{d}\lambda^2$, we evaluate the geodesic equation for $\mu = y$. Start here by evaluating the left-hand side of Eq. (P23.7) for $\mu = y$, to first order in G. Expand the derivative with respect to λ using the product rule, working out explicitly the derivatives of the relevant $g_{\mu\nu}$ with respect to λ. In parts (b) and (c), you may assume that $x(\lambda)$, $y(\lambda)$, $z(\lambda)$, and $t(\lambda)$, as well as $\mathrm{d}x/\mathrm{d}\lambda$, $\mathrm{d}y/\mathrm{d}\lambda$, $\mathrm{d}z/\mathrm{d}\lambda$, and $\mathrm{d}t/\mathrm{d}\lambda$, are all given to sufficient accuracy by Eq. (P23.8) and its derivatives with respect to λ. (Be careful: it is likely that there are more terms than you will at first notice.)

(c) *(9 points)* Evaluate the right-hand side of Eq. (P23.7) for $\mu = y$, to first order in G. Carry out all derivatives explicitly. (It always pays to be careful.)

(d) *(2 points)* Use your answers to parts (c) and (d) to find an equation for $\mathrm{d}^2y/\mathrm{d}\lambda^2$.

(e) *(4 points)* If the photon starts out on the unperturbed trajectory, its initial value of $\mathrm{d}y/\mathrm{d}\lambda$ will be zero. The final value of $\mathrm{d}y/\mathrm{d}\lambda$ will then be

$$\left.\frac{\mathrm{d}y}{\mathrm{d}\lambda}\right|_{\mathrm{final}} = \int_{-\infty}^{\infty}\frac{\mathrm{d}^2y}{\mathrm{d}\lambda^2}\,\mathrm{d}\lambda . \tag{P23.9}$$

Use this fact to express the deflection angle α, to first order in G, as an explicit integral. You need not carry out the integral, but you may wish to use the table of integrals given below to carry it out so that you can check your answer. The correct final answer is

$$\alpha = \frac{4GM}{c^2 b\ .} \tag{P23.10}$$

TABLE OF INTEGRALS:

$$\int_{-\infty}^{\infty} \frac{1}{(x^2+b^2)}\,\mathrm{d}x = \frac{\pi}{b} \qquad \int_{-\infty}^{\infty} \frac{1}{(x^2+b^2)^{3/2}}\,\mathrm{d}x = \frac{2}{b^2} \qquad \int_{-\infty}^{\infty} \frac{1}{(x^2+b^2)^2}\,\mathrm{d}x = \frac{\pi}{2b^3}$$

$$\int_{-\infty}^{\infty} \frac{x^2}{(x^2+b^2)^2}\,\mathrm{d}x = \frac{\pi}{2b} \qquad \int_{-\infty}^{\infty} \frac{x^2}{(x^2+b^2)^{5/2}}\,\mathrm{d}x = \frac{2}{3b^2} \qquad \int_{-\infty}^{\infty} \frac{x^2}{(x^2+b^2)^3}\,\mathrm{d}x = \frac{\pi}{8b^3}$$

SOLUTIONS

PROBLEM 1: DID YOU DO THE READING?

(a) This is a total trick question. Lepton number is, of course, conserved, so the factor is just 1. See Weinberg chapter 4, pages 91-4.

(b) The correct answer is (i). The others are all real reasons why it's hard to measure, although Weinberg's book emphasizes reason (v) a bit more than modern astrophysicists do: astrophysicists have been looking for other ways that deuterium might be produced, but no significant mechanism has been found. See Weinberg chapter 5, pages 114-7.

(c) The most obvious answers would be proton, neutron, and pi meson. However, there are many other possibilities, including many that were not mentioned by Weinberg. See Weinberg chapter 7, pages 136-8.

(d) The correct answers were the $\boxed{\text{neutrino}}$ and the $\boxed{\text{antiproton.}}$ The neutrino was first hypothesized by Wolfgang Pauli in 1932 (in order to explain the kinematics of beta decay), and first detected in the 1950s. After the positron was discovered in 1932, the antiproton was thought likely to exist, and the Bevatron in Berkeley was built to look for antiprotons. It made the first detection in the 1950s.

(e) The correct answers were $\boxed{\text{(ii), (v) and (vi).}}$ The others were incorrect for the following reasons:

(i) the earliest prediction of the CMB temperature, by Alpher and Herman in 1948, was 5 degrees, not 0.1 degrees.

(iii) Weinberg quotes his experimental colleagues as saying that the 3° K radiation could have been observed "long before 1965, probably in the mid-1950s and perhaps even in the mid-1940s." To Weinberg, however, the historically interesting question is not when the radiation could have been observed, but why radio astronomers did not know that they ought to try.

(iv) Weinberg argues that physicists at the time did not pay attention to either the steady state model or the big bang model, as indicated by the sentence in item (v) which is a direct quote from the book: "It was extraordinarily difficult for physicists to take seriously *any* theory of the early universe".

PROBLEM 2: DID YOU DO THE READING? *(24 points)*

(a) *(6 points)* In 1948 Ralph A. Alpher and Robert Herman wrote a paper predicting a cosmic microwave background with a temperature of 5 K. The paper was based on a cosmological model that they had developed with George Gamow, in which the early universe was assumed to have been filled with hot neutrons. As the universe expanded and cooled the neutrons underwent beta decay into protons, electrons, and antineutrinos, until at some point the universe cooled enough for light elements to be synthesized. Alpher and Herman found that to account for the observed present abundances of light elements, the ratio of photons to nuclear particles must have been about 10^9. Although the predicted temperature was very close to the actual value of 2.7 K, the theory differed from our present theory in two ways. Circle the two correct statements in the following list. (3 points for each right answer; circle at most 2.)

(i) Gamow, Alpher, and Herman assumed that the neutron could decay, but now the neutron is thought to be absolutely stable.

$\boxed{\text{(ii)}}$ In the current theory, the universe started with nearly equal densities of protons and neutrons, not all neutrons as Gamow, Alpher, and Herman assumed.

(iii) In the current theory, the universe started with mainly alpha particles, not all neutrons as Gamow, Alpher, and Herman assumed. (Note: an alpha particle is the nucleus of a helium atom, composed of two protons and two neutrons.)

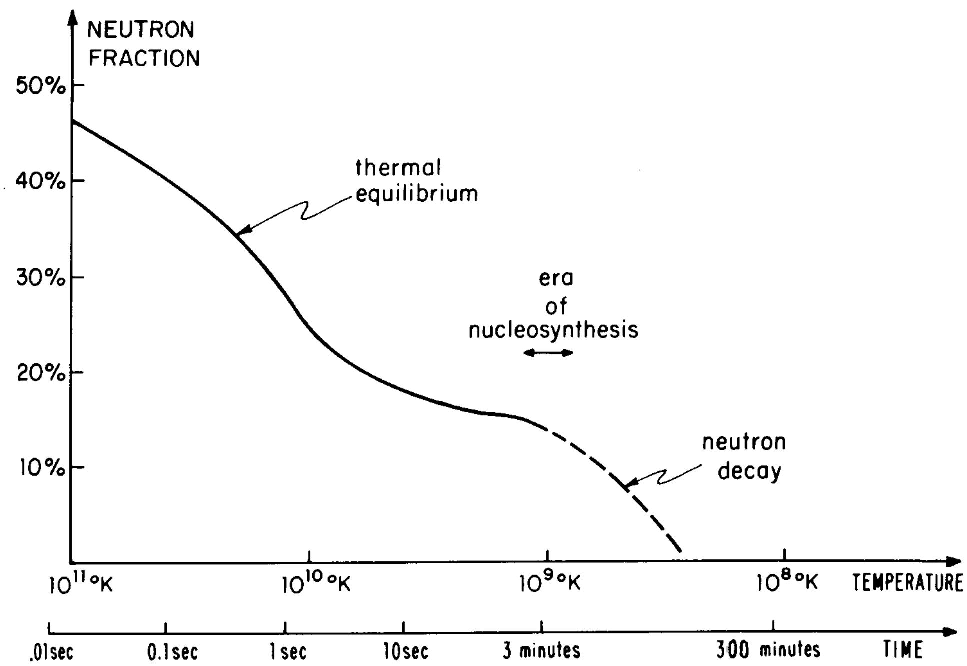
$\boxed{\text{(iv)}}$ In the current theory, the conversion of neutrons into protons (and vice versa) took place mainly through collisions with electrons, positrons, neutrinos, and antineutrinos, not through the decay of the neutrons.

(v) The ratio of photons to nuclear particles in the early universe is now believed to have been about 10^3, not 10^9 as Alpher and Herman concluded.

(b) *(6 points)* In Weinberg's "Recipe for a Hot Universe," he described the primordial composition of the universe in terms of three conserved quantities: electric charge, baryon number, and lepton number. If electric charge is measured in units of the electron charge, then all three quantities are integers for which the number density can be compared with the number density of photons. For each quantity, which choice most accurately describes the initial ratio of the number density of this quantity to the number density of photons:

Electric Charge:	(i) $\sim 10^9$	(ii) ~ 1000	(iii) ~ 1
	(iv) $\sim 10^{-6}$	$\boxed{\text{(v) either zero or negligible}}$	
Baryon Number:	(i) $\sim 10^{-20}$	$\boxed{\text{(ii) } \sim 10^{-9}}$	(iii) $\sim 10^{-6}$
	(iv) ~ 1	(v) anywhere from 10^{-5} to 1	
Lepton Number:	(i) $\sim 10^9$	(ii) ~ 1000	(iii) ~ 1
	(iv) $\sim 10^{-6}$	$\boxed{\text{(v) could be as high as } \sim 1\text{, but is assumed to be very small}}$	

(c) *(12 points)* The figure below comes from Weinberg's Chapter 5, and is labeled *The Shifting Neutron-Proton Balance.*



(i) *(3 points)* During the period labeled "thermal equilibrium," the neutron fraction is changing because (choose one):

(A) The neutron is unstable, and decays into a proton, electron, and antineutrino with a lifetime of about 1 second.

(B) The neutron is unstable, and decays into a proton, electron, and antineutrino with a lifetime of about 15 seconds.

(C) The neutron is unstable, and decays into a proton, electron, and antineutrino with a lifetime of about 15 minutes.

(D) Neutrons and protons can be converted from one into through reactions such as

$$\text{antineutrino} + \text{proton} \longleftrightarrow \text{electron} + \text{neutron}$$
$$\text{neutrino} + \text{neutron} \longleftrightarrow \text{positron} + \text{proton.}$$

$\boxed{\text{(E)}}$ Neutrons and protons can be converted from one into the other through reactions such as

$$\text{antineutrino} + \text{proton} \longleftrightarrow \text{positron} + \text{neutron}$$
$$\text{neutrino} + \text{neutron} \longleftrightarrow \text{electron} + \text{proton.}$$

(F) Neutrons and protons can be created and destroyed by reactions such as

$$\text{proton} + \text{neutrino} \longleftrightarrow \text{positron} + \text{antineutrino}$$
$$\text{neutron} + \text{antineutrino} \longleftrightarrow \text{electron} + \text{positron.}$$

(ii) *(3 points)* During the period labeled "neutron decay," the neutron fraction is changing because (choose one):

(A) The neutron is unstable, and decays into a proton, electron, and antineutrino with a lifetime of about 1 second.

(B) The neutron is unstable, and decays into a proton, electron, and antineutrino with a lifetime of about 15 seconds.

$\boxed{\text{(C)}}$ The neutron is unstable, and decays into a proton, electron, and antineutrino with a lifetime of about 15 minutes.

(D) Neutrons and protons can be converted from one into the other through reactions such as

$$\text{antineutrino} + \text{proton} \longleftrightarrow \text{electron} + \text{neutron}$$
$$\text{neutrino} + \text{neutron} \longleftrightarrow \text{positron} + \text{proton.}$$

(E) Neutrons and protons can be converted from one into the other through reactions such as

$$\text{antineutrino} + \text{proton} \longleftrightarrow \text{positron} + \text{neutron}$$
$$\text{neutrino} + \text{neutron} \longleftrightarrow \text{electron} + \text{proton.}$$

(F) Neutrons and protons can be created and destroyed by reactions such as

$$\text{proton} + \text{neutrino} \longleftrightarrow \text{positron} + \text{antineutrino}$$
$$\text{neutron} + \text{antineutrino} \longleftrightarrow \text{electron} + \text{positron.}$$

(iii) *(3 points)* The masses of the neutron and proton are not exactly equal, but instead

(A) The neutron is more massive than a proton with a rest energy difference of 1.293 GeV (1 GeV = 10^9 eV).

$\boxed{\text{(B)}}$ The neutron is more massive than a proton with a rest energy difference of 1.293 MeV (1 MeV = 10^6 eV).

(C) The neutron is more massive than a proton with a rest energy difference of 1.293 KeV (1 KeV = 10^3 eV).

(D) The proton is more massive than a neutron with a rest energy difference of 1.293 GeV.

(E) The proton is more massive than a neutron with a rest energy difference of 1.293 MeV.

(F) The proton is more massive than a neutron with a rest energy difference of 1.293 KeV.

(iv) *(3 points)* During the period labeled "era of nucleosynthesis," (choose one:)

$\boxed{\text{(A)}}$ Essentially all the neutrons present combine with protons to form helium nuclei, which mostly survive until the present time.

(B) Essentially all the neutrons present combine with protons to form deuterium nuclei, which mostly survive until the present time.

(C) About half the neutrons present combine with protons to form helium nuclei, which mostly survive until the present time, and the other half of the neutrons remain free.

(D) About half the neutrons present combine with protons to form deuterium nuclei, which mostly survive until the present time, and the other half of the neutrons remain free.

(E) Essentially all the protons present combine with neutrons to form helium nuclei, which mostly survive until the present time.

(F) Essentially all the protons present combine with neutrons to form deuterium nuclei, which mostly survive until the present time.

PROBLEM 3: DID YOU DO THE READING? *(20 points)*†

(a) *(8 points)*

(i) *(4 points)* We will use the notation X^A to indicate a nucleus,* where X is the symbol for the element which indicates the number of protons, while A is the mass number, namely the total number of protons and neutrons. With this notation H^1, H^2, H^3, He^3 and He^4 stand for hydrogen, deuterium, tritium, helium-3 and helium-4 nuclei, respectively. Steven Weinberg, in *The First Three Minutes*, chapter V, page 108, describes two chains of reactions that produce helium, starting from protons and neutrons. They can be written as:

$$\boxed{p+n \to H^2+\gamma \qquad H^2+n \to H^3+\gamma \qquad H^3+p \to He^4+\gamma,}$$

$$\boxed{p+n \to H^2+\gamma \qquad H^2+p \to He^3+\gamma \qquad He^3+n \to He^4+\gamma.}$$

These are the two examples given by Weinberg. However, different chains of two particle reactions can take place (in general with different probabilities). For example:

$$p+n \to H^2+\gamma \qquad H^2+H^2 \to He^4+\gamma,$$

$$p+n \to H^2+\gamma \qquad H^2+n \to H^3+\gamma \qquad H^3+H^2 \to He^4+n,$$

$$p+n \to H^2+\gamma \qquad H^2+p \to He^3+\gamma \qquad He^3+H^2 \to He^4+p,$$

$$\ldots$$

Students who described chains different from those of Weinberg, but that can still take place, got full credit for this part. Also, notice that photons in the reactions above carry the additional energy released. However, since the main point was to describe the nuclear reactions, students who didn't include the photons still received full credit.

(ii) *(4 points)* The *deuterium bottleneck* is discussed by Weinberg in *The First Three Minutes*, chapter V, pages 109-110. The key point is that from part (i) it should be clear that deuterium (H^2) plays a crucial role in nucleosynthesis, since it is the starting point for all the chains. However, the deuterium nucleus is extremely loosely bound compared to H^3, He^3, or especially He^4. So, there will be a

---

* Notice that some students talked about atoms, while we are talking about nuclei formation. During nucleosynthesis the temperature is way too high to allow electrons and nuclei to bind together to form atoms. This happens much later, in the process called recombination.

range of temperatures which are low enough for H^3, He^3, and He^4 nuclei to be bound, but too high to allow the deuterium nucleus to be stable. This is the temperature range where the *deuterium bottleneck* is in action: even if H^3, He^3, and He^4 nuclei could in principle be stable at those temperatures, they do not form because deuterium, which is the starting point for their formation, cannot be formed yet. Nucleosynthesis cannot proceed at a significant rate until the temperature is low enough so that deuterium nuclei are stable; at this point the deuterium bottleneck has been passed.

(b) *(12 points)*

(i) *(3 points)* If we take $a(t) = bt^{1/2}$, for some constant b, we get for the Hubble expansion rate:

$$H = \frac{\dot{a}}{a} = \frac{1}{2t} \quad \Longrightarrow \quad \boxed{t = \frac{1}{2H}.}$$

(ii) *(6 points)* By using the Friedmann equation with $k = 0$ and $\rho = \rho_r = \alpha T^4$, we find:

$$H^2 = \frac{8\pi}{3} G\rho_r = \frac{8\pi}{3} G\alpha T^4 \quad \Longrightarrow \quad \boxed{H = T^2 \sqrt{\frac{8\pi}{3} G\alpha}\ .}$$

If we substitute the given numerical values $G \simeq 6.67 \times 10^{-11}\,\mathrm{N} \cdot \mathrm{m}^2 \cdot \mathrm{kg}^{-2}$ and $\alpha \simeq 4.52 \times 10^{-32}\,\mathrm{kg} \cdot \mathrm{m}^{-3} \cdot \mathrm{K}^{-4}$ we get:

$$\boxed{H \simeq T^2 \times 5.03 \times 10^{-21}\,\mathrm{s}^{-1} \cdot \mathrm{K}^{-2}\ .}$$

Notice that the units correctly combine to give H in units of s^{-1} if the temperature is expressed in degrees Kelvin (K). In detail, we see:

$$[G\alpha]^{1/2} = (\mathrm{N} \cdot \mathrm{m}^2 \cdot \mathrm{kg}^{-2} \cdot \mathrm{kg} \cdot \mathrm{m}^{-3} \cdot \mathrm{K}^{-4})^{1/2} = \mathrm{s}^{-1} \cdot \mathrm{K}^{-2}\ ,$$

where we used the fact that $1\ \mathrm{N} = 1\ \mathrm{kg} \cdot \mathrm{m} \cdot \mathrm{s}^{-2}$. At $T = T_{\text{nucl}} \simeq 0.9 \times 10^9\mathrm{K}$ we get:

$$\boxed{H \simeq 4.07 \times 10^{-3}\mathrm{s}^{-1}.}$$

(iii) *(3 points)* Using the results in parts (i) and (ii), we get

$$\boxed{t = \frac{1}{2H} \simeq \left(\frac{9.95 \times 10^{19}}{T^2}\right) \mathrm{s} \cdot \mathrm{K}^2\ .}$$

To good accuracy, the numerator in the expression above can be rounded to 10^{20}. The above equation agrees with Weinberg's claim that, for a radiation dominated universe, time is proportional to the inverse square of the temperature. In particular for $T = T_{\text{nucl}}$ we get:

$$\boxed{t_{\text{nucl}} \simeq 123 \text{ s} \approx 2 \text{ min.}}$$

†Solution written by Daniele Bertolini.

PROBLEM 4: DID YOU DO THE READING? *(25 points)*

(a) *(6 points)* The primary evidence for dark matter in galaxies comes from measuring their rotation curves, i.e., the orbital velocity v as a function of radius R. If stars contributed all, or most, of the mass in a galaxy, what would we expect for the behavior of $v(R)$ at large radii?

Answer: If stars contributed most of the mass, then at large radii the mass would appear to be concentrated as a spherical lump at the center, and the orbits of the stars would be "Keplerian," i.e., orbits in a $1/r^2$ gravitational field. Then $\vec{F} = m\vec{a}$ implies that

$$\frac{1}{R^2} \propto \frac{v^2}{R} \quad \Longrightarrow \quad v \propto \frac{1}{\sqrt{R}} \ .$$

(b) *(5 points)* What is actually found for the behavior of $v(R)$?

Answer: $v(R)$ looks nearly flat at large radii.

(c) *(7 points)* An important tool for estimating the mass in a galaxy is the steady-state virial theorem. What does this theorem state?

Answer: For a gravitationally bound system in equilibrium,

$$\text{Kinetic energy} = -\frac{1}{2}\,(\text{Gravitational potential energy}) \ .$$

(The equality holds whenever $\ddot{I} \approx 0$, where I is the moment of inertia.)

(d) *(7 points)* At the end of Chapter 10, Ryden writes "Thus, the very strong asymmetry between baryons and antibaryons today and the large number of photons per baryon are both products of a tiny asymmetry between quarks and anitquarks in the early universe." Explain in one or a few sentences how a tiny asymmetry between quarks and anitquarks in the early universe results in a strong asymmetry between baryons and antibaryons today.

Answer: When kT was large compared to 150 MeV, the excess of quarks over antiquarks was tiny: only about 3 extra quarks for every 10^9 antiquarks. But there was massive quark-antiquark annihilation as kT fell below 150 MeV, so that today we see the excess quarks, bound into baryons, and almost no sign of antiquarks.

PROBLEM 5: DID YOU DO THE READING (2016)? *(25 points)*

(a) *(5 points)* In Chapter 8 of Barbara Ryden's *Introduction to Cosmology*, she estimates the contribution to Ω from clusters of galaxies as

(i) 0.01 (ii) 0.05 $\boxed{\text{(iii) 0.20}}$ (iv) 0.60 (v) 1.00

(b) *(4 points)* One method of estimating the total mass of a cluster of galaxies is based on the virial theorem. With this method, one estimates the mass by measuring

(i) the radius containing half the luminosity and also the temperature of the X-ray emitting gas at the center of the galaxy.

(ii) the velocity dispersion perpendicular to the line of sight and also the radius containing half of the luminosity of the cluster.

$\boxed{\text{(iii)}}$ the velocity dispersion along the line of sight and also the radius containing half of the luminosity of the cluster.

(iv) the velocity dispersion along the line of sight and also the redshift of the cluster.

(v) the velocity dispersion perpendicular to the line of sight and also the redshift of the cluster.

Explanation: The virial theorem relates the kinetic energy to the potential energy. The key relationship is

$$\frac{1}{2}M\langle v^2\rangle = \frac{\alpha}{2}\frac{GM^2}{r_h} ,$$

where M is the mass of the cluster, $\langle v^2\rangle$ is the average squared velocity of its galaxies, and r_h is the radius containing half the total mass, which is estimated by the radius containing half the luminosity. α is a numerical factor depending on the structure of the cluster, estimated at 0.4 based on observed clusters. Velocities along the line of sight are measured by the spread in Doppler shifts, while velocities perpendicular to the line of sight are essentially impossible to measure, eliminating answers (ii) and (v). Since r_h is needed, neither (i) nor (iv) include enough information. (iii) is exactly right.

(c) *(4 points)* Another method of estimating the total mass of a cluster of galaxies is to make detailed measurements of the x-rays emitted by the hot intracluster gas.

(i) By assuming that this gas is the dominant component of the mass of the cluster, the mass of the cluster can be estimated.

(ii) By assuming that the hot gas comprises about a third of the mass of the cluster, the total mass of the cluster can be estimated.

(iii) By assuming that the gas is heated by stars and supernovae that make up most of the mass of the cluster, the mass of these stars and supernovae can be estimated.

(iv) By assuming that the gas is heated by interactions with dark matter, which dominates the mass of the cluster, the mass of the cluster can be estimated.

$\boxed{\text{(v)}}$ By assuming that this gas is in hydrostatic equilibrium, the temperature, mass density, and even the chemical composition of the cluster can be modeled.

Explanation: The dominant component of the mass is apparently dark matter, so the hot intracluster gas is only a small fraction, and we have no direct way of knowing what fraction. But the gas settles into a state of hydrostatic equilibrium which is determined by pressures and gravitational forces. The gas can be mapped by measuring its x-rays, which allows astronomers to estimate the gravitational forces, and hence the mass.

(d) *(6 points)* In Chapter 6 of *The First Three Minutes*, Steven Weinberg discusses three reasons why the importance of a search for a 3° K microwave radiation background was not generally appreciated in the 1950s and early 1960s. Choose those three reasons from the following list. (2 points for each right answer, circle at most 3.)

(i) The earliest calculations erroneously predicted a cosmic background temperature of only about 0.1° K, and such a background would be too weak to detect.

(ii) There was a breakdown in communication between theorists and experimentalists.

(iii) It was not technologically possible to detect a signal as weak as a 3° K microwave background until about 1965.

(iv) Since almost all physicists at the time were persuaded by the steady state model, the predictions of the big bang model were not taken seriously.

(v) It was extraordinarily difficult for physicists to take seriously *any* theory of the early universe.

(vi) The early work on nucleosynthesis by Gamow, Alpher, Herman, and Follin, et al., had attempted to explain the origin of all complex nuclei by reactions in the early universe. This program was never very successful, and its credibility was further undermined as improvements were made in the alternative theory, that elements are synthesized in stars.

Answer: The correct answers were $\boxed{\text{(ii), (v) and (vi).}}$ The others were incorrect for the following reasons:

(i) the earliest prediction of the CMB temperature, by Alpher and Herman in 1948, was 5 degrees, not 0.1 degrees.

(iii) Weinberg quotes his experimental colleagues as saying that the 3° K radiation could have been observed "long before 1965, probably in the mid-1950s and perhaps even in the mid-1940s." To Weinberg, however, the historically interesting question is not when the radiation could have been observed, but why radio astronomers did not know that they ought to try.

(iv) Weinberg argues that physicists at the time did not pay attention to either the steady state model or the big bang model, as indicated by the sentence in item (v) which is a direct quote from the book: "It was extraordinarily difficult for physicists to take seriously *any* theory of the early universe".

(e) *(6 points)* In 1948 Ralph A. Alpher and Robert Herman wrote a paper predicting a cosmic microwave background with a temperature of 5 K. The paper was based on a cosmological model that they had developed with George Gamow, in which the early universe was assumed to have been filled with hot neutrons. As the universe expanded and cooled the neutrons underwent beta decay into protons, electrons, and antineutrinos, until at some point the universe cooled enough for light elements to be synthesized. Alpher and Herman found that to account for the observed present abundances of light elements, the ratio of photons to nuclear particles must have been about 10^9. Although the predicted temperature was very close to the actual value of 2.7 K, the theory differed from our present theory in two ways. Circle the two correct statements in the following list. (3 points for each right answer; circle at most 2.)

(i) Gamow, Alpher, and Herman assumed that the neutron could decay, but now the neutron is thought to be absolutely stable.

$\boxed{\text{(ii)}}$ In the current theory, the universe started with nearly equal densities of protons and neutrons, not all neutrons as Gamow, Alpher, and Herman assumed.

(iii) In the current theory, the universe started with mainly alpha particles, not all neutrons as Gamow, Alpher, and Herman assumed. (Note: an alpha particle is the nucleus of a helium atom, composed of two protons and two neutrons.)

$\boxed{\text{(iv)}}$ In the current theory, the conversion of neutrons into protons (and vice versa) took place mainly through collisions with electrons, positrons, neutrinos, and antineutrinos, not through the decay of the neutrons.

(v) The ratio of photons to nuclear particles in the early universe is now believed to have been about 10^3, not 10^9 as Alpher and Herman concluded.

PROBLEM 6: EVOLUTION OF AN OPEN UNIVERSE

The evolution of an open, matter-dominated universe is described by the following parametric equations:

$$ct = \alpha(\sinh\theta - \theta)$$
$$\frac{a}{\sqrt{\kappa}} = \alpha(\cosh\theta - 1) \ .$$

Evaluating the second of these equations at $a/\sqrt{\kappa} = 2\alpha$ yields a solution for θ:

$$2\alpha = \alpha(\cosh\theta - 1) \quad \Longrightarrow \quad \cosh\theta = 3 \quad \Longrightarrow \quad \theta = \cosh^{-1}(3) \ .$$

We can use these results in the first equation to solve for t. Noting that

$$\sinh\theta = \sqrt{\cosh^2\theta - 1} = \sqrt{8} = 2\sqrt{2} \ ,$$

we have

$$\boxed{t = \frac{\alpha}{c}\left[2\sqrt{2} - \cosh^{-1}(3)\right] \ .}$$

Numerically, $t \approx 1.06567\,\alpha/c$.

PROBLEM 7: ANTICIPATING A BIG CRUNCH

The critical density is given by

$$\rho_c = \frac{3H_0^2}{8\pi G} \ ,$$

so the mass density is given by

$$\rho = \Omega_0\rho_c = 2\rho_c = \frac{3H_0^2}{4\pi G} \ . \tag{S5.1}$$

Substituting this relation into

$$H_0^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2} \ ,$$

we find

$$H_0^2 = 2H_0^2 - \frac{kc^2}{a^2} \ ,$$

from which it follows that

$$\frac{a}{\sqrt{k}} = \frac{c}{H_0} \ . \tag{S5.2}$$

Now use

$$\alpha = \frac{4\pi}{3}\frac{G\rho a^3}{k^{3/2}c^2} \ .$$

Substituting the values we have from Eqs. (S5.1) and (S5.2) for ρ and $a/\sqrt{k}$, we have

$$\alpha = \frac{c}{H_0} \ . \tag{S5.3}$$

To determine the value of the parameter θ, use

$$\frac{a}{\sqrt{k}} = \alpha(1 - \cos\theta) \ ,$$

which when combined with Eqs. (S5.2) and (S5.3) implies that $\cos\theta = 0$. The equation $\cos\theta = 0$ has multiple solutions, but we know that the θ-parameter for a closed matter-dominated universe varies between 0 and π during the expansion phase of the universe. Within this range, $\cos\theta = 0$ implies that $\theta = \pi/2$. Thus, the age of the universe at the time these measurements are made is given by

$$\begin{aligned} t &= \frac{\alpha}{c}(\theta - \sin\theta) \\ &= \frac{1}{H_0}\left(\frac{\pi}{2} - 1\right) \ . \end{aligned}$$

The total lifetime of the closed universe corresponds to $\theta = 2\pi$, or

$$t_{\rm final} = \frac{2\pi\alpha}{c} = \frac{2\pi}{H_0} \ ,$$

so the time remaining before the big crunch is given by

$$t_{\rm final} - t = \frac{1}{H_0}\left[2\pi - \left(\frac{\pi}{2} - 1\right)\right] = \boxed{\left(\frac{3\pi}{2} + 1\right)\frac{1}{H_0} \ .}$$

PROBLEM 8: TRACING LIGHT RAYS IN A CLOSED, MATTER-DOMINATED UNIVERSE

(a) Since $\theta = \phi =$ constant, $d\theta = d\phi = 0$, and for light rays one always has $d\tau = 0$. The line element therefore reduces to

$$0 = -c^2\,dt^2 + a^2(t)d\psi^2 \ .$$

Rearranging gives

$$\left(\frac{d\psi}{dt}\right)^2 = \frac{c^2}{a^2(t)} ,$$

which implies that

$$\boxed{\frac{d\psi}{dt} = \pm \frac{c}{a(t)} .}$$

The plus sign describes outward radial motion, while the minus sign describes inward motion.

(b) The maximum value of the ψ coordinate that can be reached by time t is found by integrating its rate of change:

$$\psi_{\text{hor}} = \int_0^t \frac{c}{a(t')} dt' .$$

The physical horizon distance is the proper length of the shortest line drawn at the time t from the origin to $\psi = \psi_{\text{hor}}$, which according to the metric is given by

$$\ell_{\text{phys}}(t) = \int_{\psi=0}^{\psi=\psi_{\text{hor}}} ds = \int_0^{\psi_{\text{hor}}} a(t)\, d\psi = \boxed{a(t) \int_0^t \frac{c}{a(t')} dt' .}$$

(c) From part (a),

$$\frac{d\psi}{dt} = \frac{c}{a(t)} .$$

By differentiating the equation $ct = \alpha(\theta - \sin\theta)$ stated in the problem, one finds

$$\frac{dt}{d\theta} = \frac{\alpha}{c}(1 - \cos\theta) .$$

Then

$$\frac{d\psi}{d\theta} = \frac{d\psi}{dt}\frac{dt}{d\theta} = \frac{\alpha(1-\cos\theta)}{a(t)} .$$

Then using $a = \alpha(1 - \cos\theta)$, as stated in the problem, one has the very simple result

$$\boxed{\frac{d\psi}{d\theta} = 1 .}$$

(d) This part is very simple if one knows that ψ must change by 2π before the photon returns to its starting point. Since $d\psi/d\theta = 1$, this means that θ must also change by 2π. From $a = \alpha(1 - \cos\theta)$, one can see that a returns to zero at $\theta = 2\pi$, so this is exactly the lifetime of the universe. So,

$$\boxed{\frac{\text{Time for photon to return}}{\text{Lifetime of universe}} = 1\ .}$$

If it is not clear why ψ must change by 2π for the photon to return to its starting point, then recall the construction of the closed universe that was used in Lecture Notes 5. The closed universe is described as the 3-dimensional surface of a sphere in a four-dimensional Euclidean space with coordinates (x, y, z, w):

$$x^2 + y^2 + z^2 + w^2 = a^2\ ,$$

where a is the radius of the sphere. The Robertson-Walker coordinate system is constructed on the 3-dimensional surface of the sphere, taking the point $(0, 0, 0, 1)$ as the center of the coordinate system. If we define the w-direction as "north," then the point $(0, 0, 0, 1)$ can be called the north pole. Each point (x, y, z, w) on the surface of the sphere is assigned a coordinate ψ, defined to be the angle between the positive w axis and the vector (x, y, z, w). Thus $\psi = 0$ at the north pole, and $\psi = \pi$ for the antipodal point, $(0, 0, 0, -1)$, which can be called the south pole. In making the round trip the photon must travel from the north pole to the south pole and back, for a total range of 2π.

Discussion: Some students answered that the photon would return in the lifetime of the universe, but reached this conclusion without considering the details of the motion. The argument was simply that, at the big crunch when the scale factor returns to zero, all distances would return to zero, including the distance between the photon and its starting place. This statement is correct, but it does not quite answer the question. First, the statement in no way rules out the possibility that the photon might return to its starting point before the big crunch. Second, if we use the delicate but well-motivated definitions that general relativists use, it is not necessarily true that the photon returns to its starting point at the big crunch. To be concrete, let me consider a radiation-dominated closed universe—a hypothetical universe for which the only "matter" present consists of massless particles such as photons or neutrinos. In that case (you can check my calculations) a photon that leaves the north pole at $t = 0$ just reaches the south pole at the big crunch. It might seem that reaching the south pole at the big crunch is not any different from completing the round trip back to the north pole, since the distance between the north pole and the south pole is zero at $t = t_{\text{Crunch}}$, the time of the big crunch. However, suppose we adopt the principle that the instant of the initial singularity and the instant of the final crunch are both too singular to be considered part of the spacetime. We will allow ourselves to mathematically consider times ranging from $t = \epsilon$ to $t = t_{\text{Crunch}} - \epsilon$, where ϵ is arbitrarily small, but we will not try to describe what happens exactly at $t = 0$ or $t = t_{\text{Crunch}}$. Thus, we now consider a photon that starts its journey at $t = \epsilon$, and we follow it until $t = t_{\text{Crunch}} - \epsilon$. For the case of the matter-dominated closed universe, such a photon would traverse a fraction of the full circle that would be almost 1, and would approach 1 as $\epsilon \to 0$. By contrast, for the radiation-dominated closed universe, the photon would traverse a fraction of the full circle that is almost 1/2, and it would approach 1/2 as $\epsilon \to 0$. Thus, from this point of view the two cases look very different. In the radiation-dominated case, one would say that the photon has come only half-way back to its starting point.

PROBLEM 9: LENGTHS AND AREAS IN A TWO-DIMENSIONAL METRIC

a) Along the first segment $d\theta = 0$, so $ds^2 = (1+ar)^2\,dr^2$, or $ds = (1+ar)\,dr$. Integrating, the length of the first segment is found to be

$$S_1 = \int_0^{r_0} (1 + ar)\,dr = r_0 + \frac{1}{2} a r_0^2 \ .$$

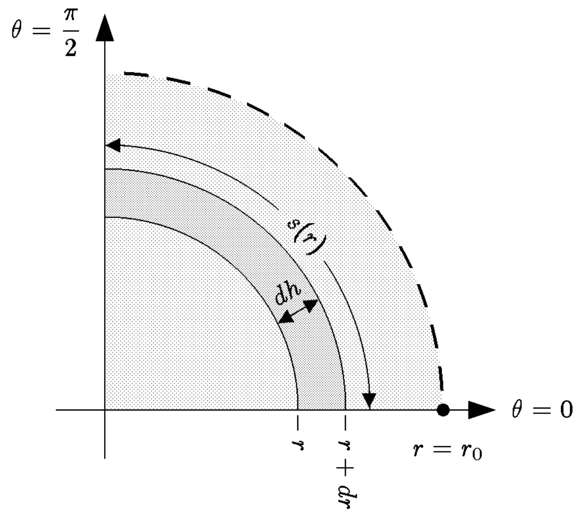
Along the second segment $dr = 0$, so $ds = r(1 + br)\,d\theta$, where $r = r_0$. So the length of the second segment is

$$S_2 = \int_0^{\pi/2} r_0(1 + br_0)\,d\theta = \frac{\pi}{2} r_0(1 + br_0) \ .$$

Finally, the third segment is identical to the first, so $S_3 = S_1$. The total length is then

$$\begin{aligned} S = 2S_1 + S_2 &= 2\left(r_0 + \frac{1}{2} a r_0^2\right) + \frac{\pi}{2} r_0 (1 + br_0) \\ &= \boxed{\left(2 + \frac{\pi}{2}\right) r_0 + \frac{1}{2}(2a + \pi b) r_0^2 \ .} \end{aligned}$$

b) To find the area, it is best to divide the region into concentric strips as shown:



Note that the strip has a coordinate width of dr, but the distance across the width of the strip is determined by the metric to be

$$dh = (1 + ar)\, dr \ .$$

The length of the strip is calculated the same way as S_2 in part (a):

$$s(r) = \frac{\pi}{2} r(1 + br) \ .$$

The area is then

$$dA = s(r)\, dh \ ,$$

so

$$\begin{aligned} A &= \int_0^{r_0} s(r)\, dh \\ &= \int_0^{r_0} \frac{\pi}{2} r(1 + br)(1 + ar)\, dr \\ &= \frac{\pi}{2} \int_0^{r_0} [r + (a + b)r^2 + abr^3]\, dr \\ &= \boxed{\frac{\pi}{2} \left[\frac{1}{2} r_0^2 + \frac{1}{3}(a + b)r_0^3 + \frac{1}{4} abr_0^4 \right]} \end{aligned}$$

PROBLEM 10: GEOMETRY IN A CLOSED UNIVERSE

(a) As one moves along a line from the origin to $(h, 0, 0)$, there is no variation in θ or ϕ. So $d\theta = d\phi = 0$, and

$$ds = \frac{a\,dr}{\sqrt{1-r^2}} \ .$$

So

$$\boxed{\ell_p = \int_0^h \frac{a\,dr}{\sqrt{1-r^2}} = a\sin^{-1} h \ .}$$

(b) In this case it is only θ that varies, so $dr = d\phi = 0$. So

$$ds = ar\,d\theta \ ,$$

so

$$\boxed{s_p = ah\,\Delta\theta \ .}$$

(c) From part (a), one has

$$h = \sin(\ell_p/a) \ .$$

Inserting this expression into the answer to (b), and then solving for $\Delta\theta$, one has

$$\boxed{\Delta\theta = \frac{s_p}{a\sin(\ell_p/a)} \ .}$$

Note that as $a \to \infty$, this approaches the Euclidean result, $\Delta\theta = s_p/\ell_p$.

PROBLEM 11: THE GENERAL SPHERICALLY SYMMETRIC METRIC

(a) The metric is given by

$$ds^2 = dr^2 + \rho^2(r)\left[d\theta^2 + \sin^2\theta\,d\phi^2\right] \ .$$

The radius a is defined as the physical length of a radial line which extends from the center to the boundary of the sphere. The length of a path is just the integral of ds, so

$$a = \int_{\substack{\text{radial path from}\\ \text{origin to } r_0}} ds \ .$$

The radial path is at a constant value of θ and ϕ, so $d\theta = d\phi = 0$, and then $ds = dr$. So

$$a = \int_0^{r_0} dr = \boxed{r_0 \ .}$$

(b) On the surface $r = r_0$, so $dr \equiv 0$. Then

$$ds^2 = \rho^2(r_0)\left[d\theta^2 + \sin^2\theta\, d\phi^2\right] \ .$$

To find the area element, consider first a path obtained by varying only θ. Then $ds = \rho(r_0)\, d\theta$. Similarly, a path obtained by varying only ϕ has length $ds = \rho(r_0)\sin\theta\, d\phi$. Furthermore, these two paths are perpendicular to each other, a fact that is incorporated into the metric by the absence of a $dr\, d\theta$ term. Thus, the area of a small rectangle constructed from these two paths is given by the product of their lengths, so

$$dA = \rho^2(r_0)\sin\theta\, d\theta\, d\phi \ .$$

The area is then obtained by integrating over the range of the coordinate variables:

$$\begin{aligned} A &= \rho^2(r_0)\int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta\, d\theta \\ &= \rho^2(r_0)(2\pi)\left(-\cos\theta\Big|_0^{\pi}\right) \\ \Longrightarrow &\quad \boxed{A = 4\pi\rho^2(r_0) \ .} \end{aligned}$$

As a check, notice that if $\rho(r) = r$, then the metric becomes the metric of Euclidean space, in spherical polar coordinates. In this case the answer above becomes the well-known formula for the area of a Euclidean sphere, $4\pi r^2$.

(c) As in Problem 2 of Problem Set 5, we can imagine breaking up the volume into spherical shells of infinitesimal thickness, with a given shell extending from r to $r + dr$. By the previous calculation, the area of such a shell is $A(r) = 4\pi\rho^2(r)$. (In the previous part we considered only the case $r = r_0$, but the same argument applies for any value of r.) The thickness of the shell is just the path length ds of a radial path corresponding to the coordinate interval dr. For radial paths the metric reduces to $ds^2 = dr^2$, so the thickness of the shell is $ds = dr$. The volume of the shell is then

$$dV = 4\pi\rho^2(r)\, dr \ .$$

The total volume is then obtained by integration:

$$\boxed{V = 4\pi\int_0^{r_0} \rho^2(r)\, dr \ .}$$

Checking the answer for the Euclidean case, $\rho(r) = r$, one sees that it gives $V = (4\pi/3)r_0^3$, as expected.

(d) If r is replaced by a new coordinate $\sigma \equiv r^2$, then the infinitesimal variations of the two coordinates are related by

$$\frac{d\sigma}{dr} = 2r = 2\sqrt{\sigma}\ ,$$

so

$$dr^2 = \frac{d\sigma^2}{4\sigma}\ .$$

The function $\rho(r)$ can then be written as $\rho(\sqrt{\sigma}\,)$, so

$$\boxed{ds^2 = \frac{d\sigma^2}{4\sigma} + \rho^2(\sqrt{\sigma}\,)\left[d\theta^2 + \sin^2\theta\, d\phi^2\right]\ .}$$

PROBLEM 12: VOLUMES IN A ROBERTSON-WALKER UNIVERSE

The product of differential length elements corresponding to infinitesimal changes in the coordinates r, θ and ϕ equals the differential volume element dV. Therefore

$$dV = a(t)\frac{dr}{\sqrt{1-kr^2}} \times a(t)r d\theta \times a(t) r\sin\theta d\phi$$

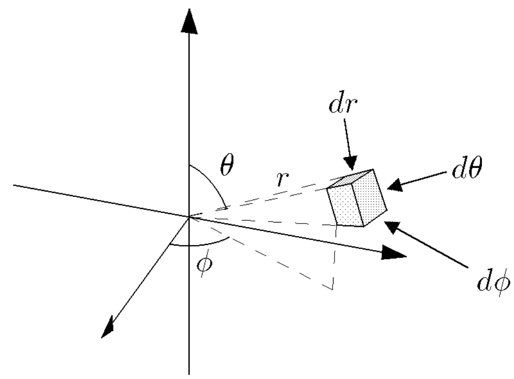
The total volume is then

$$V = \int dV = a^3(t)\int_0^{r_{\max}} dr \int_0^{\pi} d\theta \int_0^{2\pi} d\phi \frac{r^2\sin\theta}{\sqrt{1-kr^2}}$$

We can do the angular integrations immediately:

$$\boxed{V = 4\pi a^3(t)\int_0^{r_{max}} \frac{r^2 dr}{\sqrt{1-kr^2}}\ .}$$

[Pedagogical Note: If you don't see through the solutions above, then note that the volume of the sphere can be determined by integration, after first breaking the volume into infinitesimal cells. A generic cell is shown in the diagram below:



The cell includes the volume lying between r and $r + dr$, between θ and $\theta + d\theta$, and between ϕ and $\phi + d\phi$. In the limit as dr, $d\theta$, and $d\phi$ all approach zero, the cell approaches a rectangular solid with sides of length:

$$\begin{aligned}
ds_1 &= a(t)\frac{dr}{\sqrt{1-kr^2}} \\
ds_2 &= a(t)r\,d\theta \\
ds_3 &= a(t)r\sin\theta\,d\theta \ .
\end{aligned}$$

Here each ds is calculated by using the metric to find ds^2, in each case allowing only one of the quantities dr, $d\theta$, or $d\phi$ to be nonzero. The infinitesimal volume element is then $dV = ds_1 ds_2 ds_3$, resulting in the answer above. The derivation relies on the orthogonality of the dr, $d\theta$, and $d\phi$ directions; the orthogonality is implied by the metric, which otherwise would contain cross terms such as $dr\,d\theta$.]

[Extension: The integral can in fact be carried out, using the substitution

$$\begin{aligned}
\sqrt{k}\,r &= \sin\psi \quad (\text{if } k > 0) \\
\sqrt{-k}\,r &= \sinh\psi \quad (\text{if } k > 0).
\end{aligned}$$

The answer is

$$V = \begin{cases} 2\pi a^3(t)\left[\dfrac{\sin^{-1}\left(\sqrt{k}\,r_{\max}\right)}{k^{3/2}} - \dfrac{\sqrt{1-kr_{\max}^2}}{k}\right] & (\text{if } k > 0) \\[2ex] 2\pi a^3(t)\left[\dfrac{\sqrt{1-kr_{\max}^2}}{(-k)} - \dfrac{\sinh^{-1}\left(\sqrt{-k}\,r_{\max}\right)}{(-k)^{3/2}}\right] & (\text{if } k < 0) \quad .] \end{cases}$$

PROBLEM 13: THE SCHWARZSCHILD METRIC

a) The Schwarzschild horizon is the value of r for which the metric becomes singular. Since the metric contains the factor

$$\left(1-\frac{2GM}{rc^2}\right) ,$$

it becomes singular at

$$\boxed{R_S=\frac{2GM}{c^2} .}$$

b) The separation between A and B is purely in the radial direction, so the proper length of a segment along the path joining them is given by

$$ds^2=\left(1-\frac{2GM}{rc^2}\right)^{-1} dr^2 ,$$

so

$$ds=\frac{dr}{\sqrt{1-\frac{2GM}{rc^2}}} .$$

The proper distance from A to B is obtained by adding the proper lengths of all the segments along the path, so

$$\boxed{s_{AB}=\int_{r_A}^{r_B}\frac{dr}{\sqrt{1-\frac{2GM}{rc^2}}} .}$$

EXTENSION: The integration can be carried out explicitly. First use the expression for the Schwarzschild radius to rewrite the expression for s_{AB} as

$$s_{AB}=\int_{r_A}^{r_B}\frac{\sqrt{r}\,dr}{\sqrt{r-R_S}} .$$

Then introduce the hyperbolic trigonometric substitution

$$r=R_S\cosh^2 u .$$

One then has

$$\sqrt{r-R_S}=\sqrt{R_S}\,\sinh u$$

$$dr = 2R_S \cosh u \sinh u \, du \ ,$$

and the indefinite integral becomes

$$\begin{aligned} \int \frac{\sqrt{r}\,dr}{\sqrt{r-R_S}} &= 2R_S \int \cosh^2 u \, du \\ &= R_S \int (1+\cosh 2u) du \\ &= R_S \left(u + \frac{1}{2}\sinh 2u \right) \\ &= R_S (u + \sinh u \, \cosh u) \\ &= R_S \sinh^{-1}\left(\sqrt{\frac{r}{R_S} - 1} \right) + \sqrt{r(r-R_S)} \ . \end{aligned}$$

Thus,

$$\begin{aligned} s_{AB} = R_S &\left[\sinh^{-1}\left(\sqrt{\frac{r_B}{R_S} - 1} \right) - \sinh^{-1}\left(\sqrt{\frac{r_A}{R_S} - 1} \right) \right] \\ &+ \sqrt{r_B(r_B - R_S)} - \sqrt{r_A(r_A - R_S)} \ . \end{aligned}$$

c) A tick of the clock and the following tick are two events that differ only in their time coordinates. Thus, the metric reduces to

$$-c^2 d\tau^2 = -\left(1 - \frac{2GM}{rc^2} \right) c^2 dt^2 \ ,$$

so

$$d\tau = \sqrt{1 - \frac{2GM}{rc^2}} \, dt \ .$$

The reading on the observer's clock corresponds to the proper time interval $d\tau$, so the corresponding interval of the coordinate t is given by

$$\boxed{\Delta t_A = \frac{\Delta \tau_A}{\sqrt{1 - \frac{2GM}{r_A c^2}}} \ .}$$

d) Since the Schwarzschild metric does not change with time, each pulse leaving A will take the same length of time to reach B. Thus, the pulses emitted by A will arrive at B with a time coordinate spacing

$$\Delta t_B = \Delta t_A = \frac{\Delta \tau_A}{\sqrt{1 - \frac{2GM}{r_A c^2}}} \ .$$

The clock at B, however, will read the proper time and not the coordinate time. Thus,

$$\Delta\tau_B = \sqrt{1 - \frac{2GM}{r_B c^2}}\,\Delta t_B$$

$$= \boxed{\sqrt{\frac{1-\frac{2GM}{r_B c^2}}{1-\frac{2GM}{r_A c^2}}}\,\Delta\tau_A\ .}$$

e) From parts (a) and (b), the proper distance between A and B can be rewritten as

$$s_{AB} = \int_{R_S}^{r_B} \frac{\sqrt{r}\,dr}{\sqrt{r - R_S}}\ .$$

The potentially divergent part of the integral comes from the range of integration in the immediate vicinity of $r = R_{\rm S}$, say $R_S < r < R_S + \epsilon$. For this range the quantity $\sqrt{r}$ in the numerator can be approximated by $\sqrt{R_S}$, so the contribution has the form

$$\sqrt{R_S}\int_{R_S}^{R_S+\epsilon} \frac{dr}{\sqrt{r - R_S}}\ .$$

Changing the integration variable to $u \equiv r - R_S$, the contribution can be easily evaluated:

$$\sqrt{R_S}\int_{R_S}^{R_S+\epsilon} \frac{dr}{\sqrt{r - R_S}} = \sqrt{R_{\rm S}}\int_0^{\epsilon} \frac{du}{\sqrt{u}} = 2\sqrt{R_{\rm S}\epsilon} < \infty\ .$$

So, although the integrand is infinite at $r = R_S$, the integral is still finite.

$$\boxed{\text{The proper distance between } A \text{ and } B \text{ does not diverge.}}$$

Looking at the answer to part (d), however, one can see that when $r_A = R_S$,

$$\boxed{\text{The time interval } \Delta\tau_B \text{ diverges.}}$$

PROBLEM 14: GEODESICS

The geodesic equation for a curve $x^i(\lambda)$, where the parameter λ is the arc length along the curve, can be written as

$$\frac{d}{d\lambda}\left\{g_{ij}\frac{dx^j}{d\lambda}\right\} = \frac{1}{2}\left(\partial_i g_{k\ell}\right)\frac{dx^k}{d\lambda}\,\frac{dx^\ell}{d\lambda} \ .$$

Here the indices j, k, and ℓ are summed from 1 to the dimension of the space, so there is one equation for each value of i.

(a) The metric is given by

$$ds^2 = g_{ij}dx^i dx^j = dr^2 + r^2\, d\theta^2 \ ,$$

so

$$g_{rr} = 1, \qquad g_{\theta\theta} = r^2 \ , \qquad g_{r\theta} = g_{\theta r} = 0 \ .$$

First taking $i = r$, the nonvanishing terms in the geodesic equation become

$$\frac{d}{d\lambda}\left\{g_{rr}\frac{dr}{d\lambda}\right\} = \frac{1}{2}\left(\partial_r g_{\theta\theta}\right)\frac{d\theta}{d\lambda}\,\frac{d\theta}{d\lambda} \ ,$$

which can be written explicitly as

$$\frac{d}{d\lambda}\left\{\frac{dr}{d\lambda}\right\} = \frac{1}{2}\left(\partial_r r^2\right)\left(\frac{d\theta}{d\lambda}\right)^2 \ ,$$

or

$$\boxed{\frac{d^2 r}{d\lambda^2} = r\left(\frac{d\theta}{d\lambda}\right)^2 \ .}$$

For $i = \theta$, one has the simplification that g_{ij} is independent of θ for all (i, j). So

$$\boxed{\frac{d}{d\lambda}\left\{r^2\frac{d\theta}{d\lambda}\right\} = 0 \ .}$$

(b) The first step is to parameterize the curve, which means to imagine moving along the curve, and expressing the coordinates as a function of the distance traveled. (I am calling the locus $y = 1$ a curve rather than a line, since the techniques that are used here are usually applied to curves. Since a line is a special case of a curve, there is nothing wrong with treating the line as a curve.) In Cartesian coordinates, the curve $y = 1$ can be parameterized as

$$x(\lambda) = \lambda \ , \qquad y(\lambda) = 1 \ .$$

(The parameterization is not unique, because one can choose $\lambda = 0$ to represent any point along the curve.) Converting to the desired polar coordinates,

$$r(\lambda) = \sqrt{x^2(\lambda) + y^2(\lambda)} = \sqrt{\lambda^2 + 1} \ ,$$
$$\theta(\lambda) = \tan^{-1} \frac{y(\lambda)}{x(\lambda)} = \tan^{-1}(1/\lambda) \ .$$

Calculating the needed derivatives,*

$$\frac{dr}{d\lambda} = \frac{\lambda}{\sqrt{\lambda^2+1}}$$
$$\frac{d^2r}{d\lambda^2} = \frac{1}{\sqrt{\lambda^2+1}} - \frac{\lambda^2}{(\lambda^2+1)^{3/2}} = \frac{1}{(\lambda^2+1)^{3/2}} = \frac{1}{r^3}$$
$$\frac{d\theta}{d\lambda} = -\frac{1}{1+\left(\frac{1}{\lambda}\right)^2}\frac{1}{\lambda^2} = -\frac{1}{r^2} \ .$$

Then, substituting into the geodesic equation for $i = r$,

$$\frac{d^2r}{d\lambda^2} = r\left(\frac{d\theta}{d\lambda}\right)^2 \Longleftrightarrow \frac{1}{r^3} = r\left(-\frac{1}{r^2}\right)^2 \ ,$$

which checks. Substituting into the geodesic equation for $i = \theta$,

$$\frac{d}{d\lambda}\left\{r^2\frac{d\theta}{d\lambda}\right\} = 0 \Longleftrightarrow \frac{d}{d\lambda}\left\{r^2\left(-\frac{1}{r^2}\right)\right\} = 0 \ ,$$

which also checks.

---

* If you do not remember how to differentiate $\phi = \tan^{-1}(z)$, then you should know how to derive it. Write $z = \tan\phi = \sin\phi/\cos\phi$, so

$$dz = \left(\frac{\cos\phi}{\cos\phi} + \frac{\sin^2\phi}{\cos^2\phi}\right)d\phi = (1+\tan^2\phi)d\phi \ .$$

Then

$$\frac{d\phi}{dz} = \frac{1}{1+\tan^2\phi} = \frac{1}{1+z^2} \ .$$

PROBLEM 15: AN EXERCISE IN TWO-DIMENSIONAL METRICS *(30 points)*

(a) Since

$$r(\theta) = (1 + \epsilon \cos^2\theta)\, r_0 \ ,$$

as the angular coordinate θ changes by $\mathrm{d}\theta$, r changes by

$$\mathrm{d}r = \frac{\mathrm{d}r}{\mathrm{d}\theta}\,\mathrm{d}\theta = -2\epsilon r_0 \cos\theta \sin\theta\, \mathrm{d}\theta \ .$$

$\mathrm{d}s^2$ is then given by

$$\begin{aligned}\mathrm{d}s^2 &= \mathrm{d}r^2 + r^2 \mathrm{d}\theta^2 \\ &= 4\epsilon^2 r_0^2 \cos^2\theta \sin^2\theta\, \mathrm{d}\theta^2 + (1 + \epsilon\cos^2\theta)^2\, r_0^2\, \mathrm{d}\theta^2 \\ &= \left[4\epsilon^2 \cos^2\theta \sin^2\theta + (1 + \epsilon\cos^2\theta)^2\right] r_0^2\, \mathrm{d}\theta^2 \ ,\end{aligned}$$

so

$$\mathrm{d}s = r_0 \sqrt{4\epsilon^2 \cos^2\theta \sin^2\theta + (1 + \epsilon\cos^2\theta)^2}\, \mathrm{d}\theta \ .$$

Since θ runs from θ_1 to θ_2 as the curve is swept out,

$$\boxed{S = r_0 \int_{\theta_1}^{\theta_2} \sqrt{4\epsilon^2 \cos^2\theta \sin^2\theta + (1 + \epsilon\cos^2\theta)^2}\, d\theta \ .}$$

(b) Since θ does not vary along this path,

$$\mathrm{d}s = \sqrt{1 + \frac{r}{a}}\, \mathrm{d}r \ ,$$

and so

$$\boxed{R = \int_0^{r_0} \sqrt{1 + \frac{r}{a}}\, \mathrm{d}r \ .}$$

(c) Since the metric does not contain a term in $\mathrm{d}r\,\mathrm{d}\theta$, the r and θ directions are orthogonal. Thus, if one considers a small region in which r is in the interval r' to $r' + \mathrm{d}r'$, and θ is in the interval θ' to $\theta' + \mathrm{d}\theta'$, then the region can be treated as a rectangle. The side along which r varies has length $\mathrm{d}s_r = \sqrt{1 + (r'/a)}\, \mathrm{d}r'$, while the side along which θ varies has length $\mathrm{d}s_\theta = r'\, \mathrm{d}\theta'$. The area is then

$$\mathrm{d}A = \mathrm{d}s_r\, \mathrm{d}s_\theta = r' \sqrt{1 + (r'/a)}\, \mathrm{d}r'\, \mathrm{d}\theta' \ .$$

To cover the area for which $r < r_0$, r' must be integrated from 0 to r_0, and θ' must be integrated from 0 to 2π:

$$A = \int_0^{r_0} \mathrm{d}r' \int_0^{2\pi} \mathrm{d}\theta'\, r' \sqrt{1 + (r'/a)}\ .$$

But

$$\int_0^{2\pi} \mathrm{d}\theta' = 2\pi\ ,$$

so

$$\boxed{A = 2\pi \int_0^{r_0} \mathrm{d}r'\, r' \sqrt{1 + (r'/a)}\ .}$$

You were not asked to carry out the integration, but it can be done by using the substitution $u = 1 + (r'/a)$, so $\mathrm{d}u = (1/a)\,\mathrm{d}r'$, and $r' = a(u - 1)$. The result is

$$A = \frac{4\pi a^2}{15}\left[2 + \left(\frac{3r_0^2}{a^2} + \frac{r_0}{a} - 2\right)\sqrt{1 + \frac{r_0}{a}}\,\right]\ .$$

(d) The nonzero metric coefficients are given by

$$g_{rr} = 1 + \frac{r}{a}\ , \qquad g_{\theta\theta} = r^2\ ,$$

so the metric is diagonal. For $i = 1 = r$, the geodesic equation becomes

$$\frac{\mathrm{d}}{\mathrm{d}s}\left\{g_{rr}\frac{\mathrm{d}r}{\mathrm{d}s}\right\} = \frac{1}{2}\frac{\partial g_{rr}}{\partial r}\frac{\mathrm{d}r}{\mathrm{d}s}\frac{\mathrm{d}r}{\mathrm{d}s} + \frac{1}{2}\frac{\partial g_{\theta\theta}}{\partial r}\frac{\mathrm{d}\theta}{\mathrm{d}s}\frac{\mathrm{d}\theta}{\mathrm{d}s}\ ,$$

so if we substitute the values from above, we have

$$\frac{\mathrm{d}}{\mathrm{d}s}\left\{\left(1 + \frac{r}{a}\right)\frac{\mathrm{d}r}{\mathrm{d}s}\right\} = \frac{1}{2}\frac{\partial}{\partial r}\left(1 + \frac{r}{a}\right)\left(\frac{\mathrm{d}r}{\mathrm{d}s}\right)^2 + \frac{1}{2}\frac{\partial r^2}{\partial r}\left(\frac{\mathrm{d}\theta}{\mathrm{d}s}\right)^2\ .$$

Simplifying slightly,

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}s}\left\{\left(1 + \frac{r}{a}\right)\frac{\mathrm{d}r}{\mathrm{d}s}\right\} = \frac{1}{2a}\left(\frac{\mathrm{d}r}{\mathrm{d}s}\right)^2 + r\left(\frac{\mathrm{d}\theta}{\mathrm{d}s}\right)^2\ .}$$

The answer above is perfectly acceptable, but one might want to expand the left-hand side:

$$\frac{\mathrm{d}}{\mathrm{d}s}\left\{\left(1 + \frac{r}{a}\right)\frac{\mathrm{d}r}{\mathrm{d}s}\right\} = \frac{1}{a}\left(\frac{\mathrm{d}r}{\mathrm{d}s}\right)^2 + \left(1 + \frac{r}{a}\right)\frac{\mathrm{d}^2r}{\mathrm{d}s^2}\ .$$

Inserting this expansion into the boxed equation above, the first term can be brought to the right-hand side, giving

$$\boxed{\left(1+\frac{r}{a}\right)\frac{\mathrm{d}^2 r}{\mathrm{d}s^2} = -\frac{1}{2a}\left(\frac{\mathrm{d}r}{\mathrm{d}s}\right)^2 + r\left(\frac{\mathrm{d}\theta}{\mathrm{d}s}\right)^2 \ .}$$

The $i = 2 = \theta$ equation is simpler, because none of the g_{ij} coefficients depend on θ, so the right-hand side of the geodesic equation vanishes. One has simply

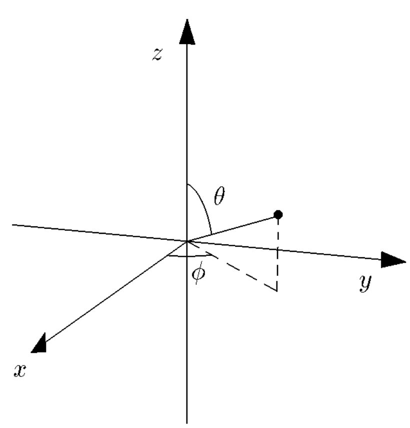
$$\boxed{\frac{\mathrm{d}}{\mathrm{d}s}\left\{r^2\frac{\mathrm{d}\theta}{\mathrm{d}s}\right\} = 0 \ .}$$

For most purposes this is the best way to write the equation, since it leads immediately to $r^2(\mathrm{d}\theta/\mathrm{d}s) = const$. However, it is possible to expand the derivative, giving the alternative form

$$r^2\frac{\mathrm{d}^2\theta}{\mathrm{d}s^2} + 2r\frac{\mathrm{d}r}{\mathrm{d}s}\frac{\mathrm{d}\theta}{\mathrm{d}s} = 0 \ .$$

PROBLEM 16: GEODESICS ON THE SURFACE OF A SPHERE

(a) Rotations are easy to understand in Cartesian coordinates. The relationship between the polar and Cartesian coordinates is given by

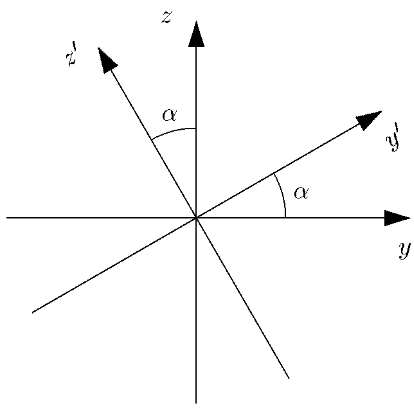


$$x = r\sin\theta\cos\phi$$
$$y = r\sin\theta\sin\phi$$
$$z = r\cos\theta \ .$$

The equator is then described by $\theta = \pi/2$, and $\phi = \psi$, where ψ is a parameter running from 0 to 2π. Thus, the equator is described by the curve $x^i(\psi)$, where

$$x^1 = x = r\cos\psi$$
$$x^2 = y = r\sin\psi$$
$$x^3 = z = 0 \ .$$

Now introduce a primed coordinate system that is related to the original system by a rotation in the y-z plane by an angle α:



$$x = x'$$
$$y = y' \cos\alpha - z' \sin\alpha$$
$$z = z' \cos\alpha + y' \sin\alpha \ .$$

The rotated equator, which we seek to describe, is just the standard equator in the primed coordinates:

$$x' = r\cos\psi \ , \qquad y' = r\sin\psi \ , \qquad z' = 0 \ .$$

Using the relation between the two coordinate systems given above,

$$x = r\cos\psi$$
$$y = r\sin\psi\cos\alpha$$
$$z = r\sin\psi\sin\alpha \ .$$

Using again the relations between polar and Cartesian coordinates,

$$\boxed{\begin{aligned} \cos\theta &= \frac{z}{r} = \sin\psi\sin\alpha \\ \tan\phi &= \frac{y}{x} = \tan\psi\cos\alpha \ . \end{aligned}}$$

(b) A segment of the equator corresponding to an interval $d\psi$ has length $a\,d\psi$, so the parameter ψ is proportional to the arc length. Expressed in terms of the metric, this relationship becomes

$$ds^2 = g_{ij}\frac{dx^i}{d\psi}\frac{dx^j}{d\psi}d\psi^2 = a^2 d\psi^2 \ .$$

Thus the quantity

$$A \equiv g_{ij} \frac{dx^i}{d\psi} \frac{dx^j}{d\psi}$$

is equal to a^2, so the geodesic equation (5.50) reduces to the simpler form of Eq. (5.52). (Note that we are following the notation of Lecture Notes 5, except that the variable used to parameterize the path is called ψ, rather than λ or s. Although A is not equal to 1 as we assumed in Lecture Notes 5, it is easily seen that Eq. (5.52) follows from (5.50) provided only that $A = constant$.) Thus,

$$\frac{d}{d\psi} \left\{ g_{ij} \frac{dx^j}{d\psi} \right\} = \frac{1}{2} \left(\partial_i g_{k\ell} \right) \frac{dx^k}{d\psi} \frac{dx^\ell}{d\psi} .$$

For this problem the metric has only two nonzero components:

$$g_{\theta\theta} = a^2 , \qquad g_{\phi\phi} = a^2 \sin^2\theta .$$

Taking $i = \theta$ in the geodesic equation,

$$\frac{d}{d\psi} \left\{ g_{\theta\theta} \frac{d\theta}{d\psi} \right\} = \frac{1}{2} \partial_\theta g_{\phi\phi} \frac{d\phi}{d\psi} \frac{d\phi}{d\psi} \qquad \Longrightarrow$$

$$\boxed{\frac{d^2\theta}{d\psi^2} = \sin\theta\cos\theta \left(\frac{d\phi}{d\psi} \right)^2 .}$$

Taking $i = \phi$,

$$\frac{d}{d\psi} \left\{ a^2 \sin^2\theta \frac{d\phi}{d\psi} \right\} = 0 \qquad \Longrightarrow$$

$$\boxed{\frac{d}{d\psi} \left\{ \sin^2\theta \frac{d\phi}{d\psi} \right\} = 0 .}$$

(c) This part is mainly algebra. Taking the derivative of

$$\cos\theta = \sin\psi \sin\alpha$$

implies

$$-\sin\theta \, d\theta = \cos\psi \sin\alpha \, d\psi .$$

Then, using the trigonometric identity $\sin\theta = \sqrt{1 - \cos^2\theta}$, one finds

$$\sin\theta = \sqrt{1 - \sin^2\psi \sin^2\alpha} ,$$

so

$$\frac{d\theta}{d\psi} = -\frac{\cos\psi\sin\alpha}{\sqrt{1-\sin^2\psi\sin^2\alpha}} \ .$$

Similarly

$$\tan\phi = \tan\psi\cos\alpha \quad \Longrightarrow \quad \sec^2\phi\, d\phi = \sec^2\psi\, d\psi\cos\alpha \ .$$

Then

$$\begin{aligned}\sec^2\phi = \tan^2\phi + 1 &= \tan^2\psi\cos^2\alpha + 1\\ &= \frac{1}{\cos^2\psi}[\sin^2\psi\cos^2\alpha + \cos^2\psi]\\ &= \sec^2\psi[\sin^2\psi(1-\sin^2\alpha)+\cos^2\psi]\\ &= \sec^2\psi[1-\sin^2\psi\sin^2\alpha] \ ,\end{aligned}$$

So

$$\frac{d\phi}{d\psi} = \frac{\cos\alpha}{1-\sin^2\psi\sin^2\alpha} \ .$$

To verify the geodesic equations of part (b), it is easiest to check the second one first:

$$\begin{aligned}\sin^2\theta\frac{d\phi}{d\psi} &= (1-\sin^2\psi\sin^2\alpha)\frac{\cos\alpha}{1-\sin^2\psi\sin^2\alpha}\\ &= \cos\alpha \ ,\end{aligned}$$

so clearly

$$\frac{d}{d\psi}\left\{\sin^2\theta\frac{d\phi}{d\psi}\right\} = \frac{d}{d\psi}(\cos\alpha) = 0 \ .$$

To verify the first geodesic equation from part (b), first calculate the left-hand side, $d^2\theta/d\psi^2$, using our result for $d\theta/d\psi$:

$$\frac{d^2\theta}{d\psi^2} = \frac{d}{d\psi}\left(\frac{d\theta}{d\psi}\right) = \frac{d}{d\psi}\left\{-\frac{\cos\psi\sin\alpha}{\sqrt{1-\sin^2\psi\sin^2\alpha}}\right\} \ .$$

After some straightforward algebra, one finds

$$\frac{d^2\theta}{d\psi^2} = \frac{\sin\psi\sin\alpha\cos^2\alpha}{\left[1-\sin^2\psi\sin^2\alpha\right]^{3/2}} \ .$$

The right-hand side of the first geodesic equation can be evaluated using the expression found above for $d\phi/d\psi$, giving

$$\sin\theta\cos\theta\left(\frac{d\phi}{d\psi}\right)^2 = \sqrt{1-\sin^2\psi\sin^2\alpha}\,\sin\psi\sin\alpha\,\frac{\cos^2\alpha}{\left[1-\sin^2\psi\sin^2\alpha\right]^2}$$
$$= \frac{\sin\psi\sin\alpha\cos^2\alpha}{\left[1-\sin^2\psi\sin^2\alpha\right]^{3/2}}\ .$$

So the left- and right-hand sides are equal.

PROBLEM 17: GEODESICS IN A CLOSED UNIVERSE

(a) *(7 points)* For purely radial motion, $d\theta = d\phi = 0$, so the line element reduces do

$$-c^2\,d\tau^2 = -c^2\,dt^2 + a^2(t)\left\{\frac{dr^2}{1-r^2}\right\}\ .$$

Dividing by dt^2,

$$-c^2\left(\frac{d\tau}{dt}\right)^2 = -c^2 + \frac{a^2(t)}{1-r^2}\left(\frac{dr}{dt}\right)^2\ .$$

Rearranging,

$$\boxed{\frac{d\tau}{dt} = \sqrt{1-\frac{a^2(t)}{c^2(1-r^2)}\left(\frac{dr}{dt}\right)^2}\ .}$$

(b) *(3 points)*

$$\frac{dt}{d\tau} = \frac{1}{\dfrac{d\tau}{dt}} = \boxed{\frac{1}{\sqrt{1-\dfrac{a^2(t)}{c^2(1-r^2)}\left(\dfrac{dr}{dt}\right)^2}}}\ .$$

(c) *(10 points)* During any interval of clock time dt, the proper time that would be measured by a clock moving with the object is given by $d\tau$, as given by the metric. Using the answer from part (a),

$$d\tau = \frac{d\tau}{dt}\,dt = \sqrt{1-\frac{a^2(t)}{c^2(1-r_p^2)}\left(\frac{dr_p}{dt}\right)^2}\,dt\ .$$

Integrating to find the total proper time,

$$\boxed{\tau = \int_{t_1}^{t_2} \sqrt{1 - \frac{a^2(t)}{c^2(1-r_p^2)} \left(\frac{dr_p}{dt}\right)^2}\, dt \ .}$$

(d) *(10 points)* The physical distance $d\ell$ that the object moves during a given time interval is related to the coordinate distance dr by the spatial part of the metric:

$$d\ell^2 = ds^2 = a^2(t) \left\{ \frac{dr^2}{1-r^2} \right\} \quad \Longrightarrow \quad d\ell = \frac{a(t)}{\sqrt{1-r^2}}\, dr \ .$$

Thus

$$\boxed{v_{\rm phys} = \frac{d\ell}{dt} = \frac{a(t)}{\sqrt{1-r^2}} \frac{dr}{dt} \ .}$$

Discussion: A common mistake was to include $-c^2\, dt^2$ in the expression for $d\ell^2$. To understand why this is not correct, we should think about how an observer would measure $d\ell$, the distance to be used in calculating the velocity of a passing object. The observer would place a meter stick along the path of the object, and she would mark off the position of the object at the beginning and end of a time interval $dt_{\rm meas}$. Then she would read the distance by subtracting the two readings on the meter stick. This subtraction is equal to the physical distance between the two marks, measured at the ***same*** time t. Thus, when we compute the distance between the two marks, we set $dt = 0$. To compute the speed she would then divide the distance by $dt_{\rm meas}$, which is nonzero.

(e) *(10 points)* We start with the standard formula for a geodesic, as written on the front of the exam:

$$\frac{d}{d\tau} \left\{ g_{\mu\nu} \frac{dx^\nu}{d\tau} \right\} = \frac{1}{2} \left(\partial_\mu g_{\lambda\sigma}\right) \frac{dx^\lambda}{d\tau} \frac{dx^\sigma}{d\tau} \ .$$

This formula is true for each possible value of μ, while the Einstein summation convention implies that the indices ν, λ, and σ are summed. We are trying to derive the equation for r, so we set $\mu = r$. Since the metric is diagonal, the only contribution on the left-hand side will be $\nu = r$. On the right-hand side, the diagonal nature of the metric implies that nonzero contributions arise only when $\lambda = \sigma$. The term will vanish unless $dx^\lambda/d\tau$ is nonzero, so λ must be either r or t (i.e., there is no motion in the θ or ϕ directions). However, the right-hand side is proportional to

$$\frac{\partial g_{\lambda\sigma}}{\partial r} \ .$$

Since $g_{tt} = -c^2$, the derivative with respect to r will vanish. Thus, the only nonzero contribution on the right-hand side arises from $\lambda = \sigma = r$. Using

$$g_{rr} = \frac{a^2(t)}{1-r^2} \ ,$$

the geodesic equation becomes

$$\frac{d}{d\tau}\left\{g_{rr}\frac{dr}{d\tau}\right\} = \frac{1}{2}\left(\partial_r g_{rr}\right)\frac{dr}{d\tau}\frac{dr}{d\tau} \ ,$$

or

$$\frac{d}{d\tau}\left\{\frac{a^2}{1-r^2}\frac{dr}{d\tau}\right\} = \frac{1}{2}\left[\partial_r\left(\frac{a^2}{1-r^2}\right)\right]\frac{dr}{d\tau}\frac{dr}{d\tau} \ ,$$

or finally

$$\boxed{\frac{d}{d\tau}\left\{\frac{a^2}{1-r^2}\frac{dr}{d\tau}\right\} = a^2\frac{r}{(1-r^2)^2}\left(\frac{dr}{d\tau}\right)^2 \ .}$$

This matches the form shown in the question, with

$$A = \frac{a^2}{1-r^2} \ , \text{ and } C = a^2\frac{r}{(1-r^2)^2} \ ,$$

with $B = D = E = 0$.

(f) *(5 points EXTRA CREDIT)* The algebra here can get messy, but it is not too bad if one does the calculation in an efficient way. One good way to start is to simplify the expression for p. Using the answer from (d),

$$p = \frac{mv_{\text{phys}}}{\sqrt{1-\frac{v_{\text{phys}}^2}{c^2}}} = \frac{m\frac{a(t)}{\sqrt{1-r^2}}\frac{dr}{dt}}{\sqrt{1-\frac{a^2}{c^2(1-r^2)}\left(\frac{dr}{dt}\right)^2}} \ .$$

Using the answer from (b), this simplifies to

$$p = m\frac{a(t)}{\sqrt{1-r^2}}\frac{dr}{dt}\frac{dt}{d\tau} = m\frac{a(t)}{\sqrt{1-r^2}}\frac{dr}{d\tau} \ .$$

Multiply the geodesic equation by m, and then use the above result to rewrite it as

$$\frac{d}{d\tau}\left\{\frac{ap}{\sqrt{1-r^2}}\right\} = ma^2\,\frac{r}{(1-r^2)^2}\left(\frac{dr}{d\tau}\right)^2 \ .$$

Expanding the left-hand side,

$$LHS = \frac{d}{d\tau}\left\{\frac{ap}{\sqrt{1-r^2}}\right\} = \frac{1}{\sqrt{1-r^2}}\frac{d}{d\tau}\{ap\} + ap\frac{r}{(1-r^2)^{3/2}}\frac{dr}{d\tau}$$
$$= \frac{1}{\sqrt{1-r^2}}\frac{d}{d\tau}\{ap\} + ma^2\frac{r}{(1-r^2)^2}\left(\frac{dr}{d\tau}\right)^2 .$$

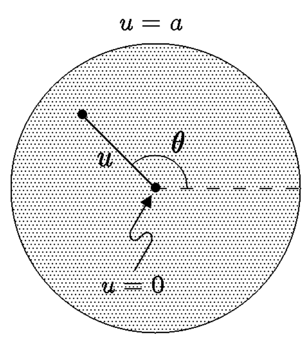
Inserting this expression back into left-hand side of the original equation, one sees that the second term cancels the expression on the right-hand side, leaving

$$\frac{1}{\sqrt{1-r^2}}\frac{d}{d\tau}\{ap\} = 0 .$$

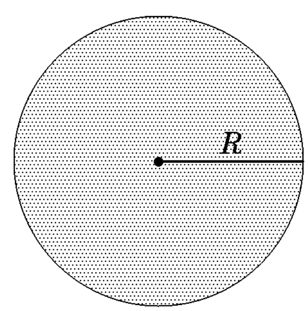
Multiplying by $\sqrt{1-r^2}$, one has the desired result:

$$\frac{d}{d\tau}\{ap\} = 0 \quad \Longrightarrow \quad \boxed{p \propto \frac{1}{a(t)} .}$$

PROBLEM 18: A TWO-DIMENSIONAL CURVED SPACE *(40 points)*



(a) For $\theta = constant$, the expression for the metric reduces to



$$ds^2 = \frac{a\,\mathrm{d}u^2}{4u(a-u)} \quad \Longrightarrow$$
$$ds = \frac{1}{2}\sqrt{\frac{a}{u(a-u)}}\,\mathrm{d}u .$$

To find the length of the radial line shown, one must integrate this expression from the value of u at the center, which is 0, to the value of u at the outer edge, which is a. So

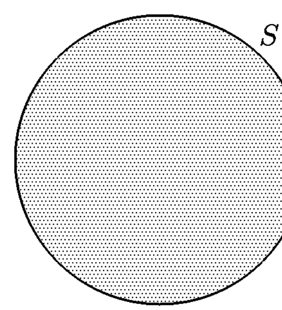
$$\boxed{R = \frac{1}{2}\int_0^a \sqrt{\frac{a}{u(a-u)}}\,\mathrm{d}u\ .}$$

You were not expected to do it, but the integral can be carried out, giving $R = (\pi/2)\sqrt{a}$.

(b) For $u = constant$, the expression for the metric reduces to

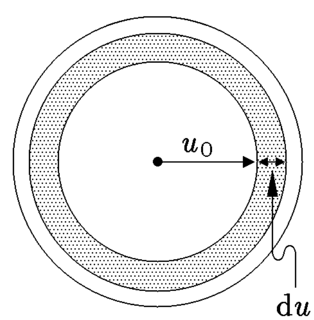
$$ds^2 = u\,\mathrm{d}\theta^2 \quad\Longrightarrow\quad ds = \sqrt{u}\,\mathrm{d}\theta\ .$$

Since θ runs from 0 to 2π, and $u = a$ for the circumference of the space,



$$\boxed{S = \int_0^{2\pi} \sqrt{a}\,\mathrm{d}\theta = 2\pi\sqrt{a}\ .}$$

(c) To evaluate the answer to first order in $\mathrm{d}u$ means to neglect any terms that would be proportional to $\mathrm{d}u^2$ or higher powers. This means that we can treat the annulus as if it were arbitrarily thin, in which case we can imagine bending it into a rectangle without changing its area. The area is then equal to the circumference times the width. Both the circumference and the width must be calculated by using the metric:



$$\begin{aligned} \mathrm{d}A &= \text{circumference} \times \text{width} \\ &= [2\pi\sqrt{u_0}\,] \times \left[\frac{1}{2}\sqrt{\frac{a}{u_0(a-u_0)}}\,\mathrm{d}u\right] \\ &= \boxed{\pi\sqrt{\frac{a}{(a-u_0)}}\,\mathrm{d}u\ .} \end{aligned}$$

(d) We can find the total area by imagining that it is broken up into annuluses, where a single annulus starts at radial coordinate u and extends to $u + \mathrm{d}u$. As in part (a), this expression must be integrated from the value of u at the center, which is 0, to the value of u at the outer edge, which is a.

$$\boxed{A = \pi \int_0^a \sqrt{\frac{a}{(a-u)}}\,\mathrm{d}u\ .}$$

You did not need to carry out this integration, but the answer would be $A = 2\pi a$.

(e) From the list at the front of the exam, the general formula for a geodesic is written as

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}s}\right] = \frac{1}{2}\frac{\partial g_{k\ell}}{\partial x^i}\frac{\mathrm{d}x^k}{\mathrm{d}s}\frac{\mathrm{d}x^\ell}{\mathrm{d}s}\ .$$

The metric components g_{ij} are related to $\mathrm{d}s^2$ by

$$\mathrm{d}s^2 = g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j\ ,$$

where the Einstein summation convention (sum over repeated indices) is assumed. In this case

$$\begin{aligned} g_{11} \equiv g_{uu} &= \frac{a}{4u(a-u)} \\ g_{22} \equiv g_{\theta\theta} &= u \\ g_{12} = g_{21} &= 0\ , \end{aligned}$$

where I have chosen $x^1 = u$ and $x^2 = \theta$. The equation with $\mathrm{d}u/\mathrm{d}s$ on the left-hand side is found by looking at the geodesic equations for $i = 1$. Of course j, k, and ℓ must all be summed, but the only nonzero contributions arise when $j = 1$, and k and ℓ are either both equal to 1 or both equal to 2:

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[g_{uu}\frac{\mathrm{d}u}{\mathrm{d}s}\right] = \frac{1}{2}\frac{\partial g_{uu}}{\partial u}\left(\frac{\mathrm{d}u}{\mathrm{d}s}\right)^2 + \frac{1}{2}\frac{\partial g_{\theta\theta}}{\partial u}\left(\frac{\mathrm{d}\theta}{\mathrm{d}s}\right)^2\ .$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\left[\frac{a}{4u(a-u)}\frac{\mathrm{d}u}{\mathrm{d}s}\right] &= \frac{1}{2}\left[\frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{a}{4u(a-u)}\right)\right]\left(\frac{\mathrm{d}u}{\mathrm{d}s}\right)^2 + \frac{1}{2}\left[\frac{\mathrm{d}}{\mathrm{d}u}(u)\right]\left(\frac{\mathrm{d}\theta}{\mathrm{d}s}\right)^2 \\ &= \frac{1}{2}\left[\frac{a}{4u(a-u)^2} - \frac{a}{4u^2(a-u)}\right]\left(\frac{\mathrm{d}u}{\mathrm{d}s}\right)^2 + \frac{1}{2}\left(\frac{\mathrm{d}\theta}{\mathrm{d}s}\right)^2 \\ &= \boxed{\frac{1}{8}\frac{a(2u-a)}{u^2(a-u)^2}\left(\frac{\mathrm{d}u}{\mathrm{d}s}\right)^2 + \frac{1}{2}\left(\frac{\mathrm{d}\theta}{\mathrm{d}s}\right)^2 .}\end{aligned}$$

(f) This part is solved by the same method, but it is simpler. Here we consider the geodesic equation with $i = 2$. The only term that contributes on the left-hand side is $j = 2$. On the right-hand side one finds nontrivial expressions when k and ℓ are either both equal to 1 or both equal to 2. However, the terms on the right-hand side both involve the derivative of the metric with respect to $x^2 = \theta$, and these derivatives all vanish. So

$$\frac{\mathrm{d}}{\mathrm{d}s}\left[g_{\theta\theta}\frac{\mathrm{d}\theta}{\mathrm{d}s}\right] = \frac{1}{2}\frac{\partial g_{uu}}{\partial\theta}\left(\frac{\mathrm{d}u}{\mathrm{d}s}\right)^2 + \frac{1}{2}\frac{\partial g_{\theta\theta}}{\partial\theta}\left(\frac{\mathrm{d}\theta}{\mathrm{d}s}\right)^2 ,$$

which reduces to

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}s}\left[u\frac{\mathrm{d}\theta}{\mathrm{d}s}\right] = 0 .}$$

PROBLEM 19: ROTATING FRAMES OF REFERENCE *(35 points)*

(a) The metric was given as

$$-c^2\,\mathrm{d}\tau^2 = -c^2\,\mathrm{d}t^2 + \left[\mathrm{d}r^2 + r^2\,(\mathrm{d}\phi + \omega\,\mathrm{d}t)^2 + \mathrm{d}z^2\right] ,$$

and the metric coefficients are then just read off from this expression:

$$\begin{aligned} &g_{11} \equiv g_{rr} = 1 \\ &g_{00} \equiv g_{tt} = \text{coefficient of } \mathrm{d}t^2 = -c^2 + r^2\omega^2 \\ &g_{20} \equiv g_{02} \equiv g_{\phi t} \equiv g_{t\phi} = \frac{1}{2} \times \text{coefficient of } \mathrm{d}\phi\,\mathrm{d}t = r^2\omega^2 \\ &g_{22} \equiv g_{\phi\phi} = \text{coefficient of } \mathrm{d}\phi^2 = r^2 \\ &g_{33} \equiv g_{zz} = \text{coefficient of } \mathrm{d}z^2 = 1 .\end{aligned}$$

Note that the off-diagonal term $g_{\phi t}$ must be multiplied by 1/2, because the expression

$$\sum_{\mu=0}^{3}\sum_{\nu=0}^{3} g_{\mu\nu}\, dx^\mu\, dx^\nu$$

includes the two equal terms $g_{20}\,\mathrm{d}\phi\,\mathrm{d}t + g_{02}\,\mathrm{d}t\,\mathrm{d}\phi$, where $g_{20} \equiv g_{02}$.

(b) Starting with the general expression

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{g_{\mu\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\right\} = \frac{1}{2}\left(\partial_\mu g_{\lambda\sigma}\right)\frac{\mathrm{d}x^\lambda}{\mathrm{d}\tau}\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau}\ ,$$

we set $\mu = r$:

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{g_{r\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\right\} = \frac{1}{2}\left(\partial_r g_{\lambda\sigma}\right)\frac{\mathrm{d}x^\lambda}{\mathrm{d}\tau}\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau}\ .$$

When we sum over ν on the left-hand side, the only value for which $g_{r\nu} \neq 0$ is $\nu = 1 \equiv r$. Thus, the left-hand side is simply

$$\mathrm{LHS} = \frac{\mathrm{d}}{\mathrm{d}\tau}\left(g_{rr}\frac{\mathrm{d}x^1}{\mathrm{d}\tau}\right) = \frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right) = \frac{\mathrm{d}^2 r}{\mathrm{d}\tau^2}\ .$$

The RHS includes every combination of λ and σ for which $g_{\lambda\sigma}$ depends on r, so that $\partial_r\, g_{\lambda\sigma} \neq 0$. This means g_{tt}, $g_{\phi\phi}$, and $g_{\phi t}$. So,

$$\begin{aligned} RHS &= \frac{1}{2}\partial_r(-c^2 + r^2\omega^2)\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 + \frac{1}{2}\partial_r(r^2)\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2 + \partial_r(r^2\omega)\,\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\,\frac{\mathrm{d}t}{\mathrm{d}\tau} \\ &= r\omega^2\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 + r\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2 + 2r\omega\,\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\,\frac{\mathrm{d}t}{\mathrm{d}\tau} \\ &= r\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau} + \omega\,\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2\ . \end{aligned}$$

Note that the final term in the first line is really the sum of the contributions from $g_{\phi t}$ and $g_{t\phi}$, where the two terms were combined to cancel the factor of 1/2 in the general expression. Finally,

$$\boxed{\frac{\mathrm{d}^2 r}{\mathrm{d}\tau^2} = r\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau} + \omega\,\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2\ .}$$

If one expands the RHS as

$$\frac{\mathrm{d}^2 r}{\mathrm{d}\tau^2} = r\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2 + r\omega^2\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 + 2r\omega\,\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\frac{\mathrm{d}t}{\mathrm{d}\tau}\ ,$$

then one can identify the term proportional to ω^2 as the centrifugal force, and the term proportional to ω as the Coriolis force.

(c) Substituting $\mu = \phi$,

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{g_{\phi\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\right\} = \frac{1}{2}\left(\partial_\phi g_{\lambda\sigma}\right)\frac{\mathrm{d}x^\lambda}{\mathrm{d}\tau}\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau} .$$

But none of the metric coefficients depend on ϕ, so the right-hand side is zero. The left-hand side receives contributions from $\nu = \phi$ and $\nu = t$:

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left(g_{\phi\phi}\frac{\mathrm{d}\phi}{\mathrm{d}\tau} + g_{\phi t}\frac{\mathrm{d}t}{\mathrm{d}\tau}\right) = \frac{\mathrm{d}}{\mathrm{d}\tau}\left(r^2\frac{\mathrm{d}\phi}{\mathrm{d}\tau} + r^2\omega\frac{\mathrm{d}t}{\mathrm{d}\tau}\right) = 0 ,$$

so

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}\tau}\left(r^2\frac{\mathrm{d}\phi}{\mathrm{d}\tau} + r^2\omega\frac{\mathrm{d}t}{\mathrm{d}\tau}\right) = 0 .}$$

Note that one cannot "factor out" r^2, since r can depend on τ. If this equation is expanded to give an equation for $\mathrm{d}^2\phi/\mathrm{d}\tau^2$, the term proportional to ω would be identified as the Coriolis force. There is no term proportional to ω^2, since the centrifugal force has no component in the ϕ direction.

(d) If Eq. (P19.1) of the problem is divided by $c^2\mathrm{d}t^2$, one obtains

$$\left(\frac{\mathrm{d}\tau}{\mathrm{d}t}\right)^2 = 1 - \frac{1}{c^2}\left[\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 + r^2\left(\frac{\mathrm{d}\phi}{\mathrm{d}t}+\omega\right)^2 + \left(\frac{\mathrm{d}z}{\mathrm{d}t}\right)^2\right] .$$

Then using

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{1}{\left(\frac{\mathrm{d}\tau}{\mathrm{d}t}\right)} ,$$

one has

$$\boxed{\frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{1}{\sqrt{1 - \frac{1}{c^2}\left[\left(\frac{\mathrm{d}r}{\mathrm{d}t}\right)^2 + r^2\left(\frac{\mathrm{d}\phi}{\mathrm{d}t}+\omega\right)^2 + \left(\frac{\mathrm{d}z}{\mathrm{d}t}\right)^2\right]}} .}$$

Note that this equation is really just

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{1}{\sqrt{1 - v^2/c^2}} ,$$

adapted to the rotating cylindrical coordinate system.

PROBLEM 20: THE STABILITY OF SCHWARZSCHILD ORBITS* *(30 points)*

From the metric:

$$ds^2 = -c^2 d\tau^2 = -h(r)\, c^2 dt^2 + h(r)^{-1} dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2\,, \tag{S20.1}$$

and the convention $ds^2 = g_{\mu\nu} dx^\mu dx^\nu$ we read the nonvanishing metric components:

$$g_{tt} = -h(r)c^2\,, \quad g_{rr} = \frac{1}{h(r)}\,, \quad g_{\theta\theta} = r^2\,, \quad g_{\phi\phi} = r^2 \sin^2\theta\,. \tag{S20.2}$$

We are told that the orbit has $\theta = \pi/2$, so on the orbit $d\theta = 0$ and the relevant metric and metric components are:

$$ds^2 = -c^2 d\tau^2 = -h(r)\, c^2 dt^2 + h(r)^{-1} dr^2 + r^2 d\phi^2\,, \tag{S20.3}$$

$$g_{tt} = -h(r)c^2\,, \quad g_{rr} = \frac{1}{h(r)}\,, \quad g_{\phi\phi} = r^2\,. \tag{S20.4}$$

We also know that

$$h(r) = 1 - \frac{R_S}{r}\,. \tag{S20.5}$$

(a) The geodesic equation

$$\frac{d}{d\tau}\left[g_{\mu\nu}\frac{dx^\nu}{d\tau}\right] = \frac{1}{2}\frac{\partial g_{\lambda\sigma}}{\partial x^\mu}\frac{dx^\lambda}{d\tau}\frac{dx^\sigma}{d\tau}\,, \tag{S20.6}$$

for the index value $\mu = r$ takes the form

$$\frac{d}{d\tau}\left[g_{rr}\frac{dr}{d\tau}\right] = \frac{1}{2}\frac{\partial g_{\lambda\sigma}}{\partial r}\frac{dx^\lambda}{d\tau}\frac{dx^\sigma}{d\tau}\,.$$

Expanding out

$$\frac{d}{d\tau}\left[\frac{1}{h}\frac{dr}{d\tau}\right] = \frac{1}{2}\frac{\partial g_{tt}}{\partial r}\left(\frac{dt}{d\tau}\right)^2 + \frac{1}{2}\frac{\partial g_{rr}}{\partial r}\left(\frac{dr}{d\tau}\right)^2 + \frac{1}{2}\frac{\partial g_{\phi\phi}}{\partial r}\left(\frac{d\phi}{d\tau}\right)^2\,.$$

Using the values in (S20.4) to evaluate the right-hand side and taking the derivatives on the left-hand side:

$$\underline{-\frac{h'}{h^2}\left(\frac{dr}{d\tau}\right)^2} + \frac{1}{h}\frac{d^2r}{d\tau^2} = -\frac{1}{2}c^2 h'\left(\frac{dt}{d\tau}\right)^2 \underline{-\frac{1}{2}\frac{h'}{h^2}\left(\frac{dr}{d\tau}\right)^2} + r\left(\frac{d\phi}{d\tau}\right)^2\,.$$

---

* Solution by Barton Zwiebach.

Here $h' \equiv \frac{dh}{dr}$ and we have supressed the arguments of h and h' to avoid clutter. Collecting the underlined terms to the right and multiplying by h, we find

$$\frac{d^2r}{d\tau^2} = -\frac{1}{2}\,h'\,hc^2\left(\frac{dt}{d\tau}\right)^2 + \frac{1}{2}\frac{h'}{h}\left(\frac{dr}{d\tau}\right)^2 + rh\left(\frac{d\phi}{d\tau}\right)^2 . \tag{S20.7}$$

(b) Dividing the expression (S20.3) for the metric by $d\tau^2$ we readily find

$$-c^2 = -hc^2\left(\frac{dt}{d\tau}\right)^2 + \frac{1}{h}\left(\frac{dr}{d\tau}\right)^2 + r^2\left(\frac{d\phi}{d\tau}\right)^2 ,$$

and rearranging,

$$hc^2\left(\frac{dt}{d\tau}\right)^2 = c^2 + \frac{1}{h}\left(\frac{dr}{d\tau}\right)^2 + r^2\left(\frac{d\phi}{d\tau}\right)^2 . \tag{S20.8}$$

This is the most useful form of the answer. Of course, we also have

$$\left(\frac{dt}{d\tau}\right)^2 = \frac{1}{h} + \frac{1}{h^2c^2}\left(\frac{dr}{d\tau}\right)^2 + \frac{r^2}{hc^2}\left(\frac{d\phi}{d\tau}\right)^2 . \tag{S20.9}$$

We use now (S20.8) to simplify (S20.7):

$$\frac{d^2r}{d\tau^2} = -\frac{1}{2}\,h'\left(c^2 + \frac{1}{h}\left(\frac{dr}{d\tau}\right)^2 + r^2\left(\frac{d\phi}{d\tau}\right)^2\right) + \frac{1}{2}\frac{h'}{h}\left(\frac{dr}{d\tau}\right)^2 + rh\left(\frac{d\phi}{d\tau}\right)^2 .$$

Expanding out, the terms with $(\frac{dr}{d\tau})^2$ cancel and we find

$$\frac{d^2r}{d\tau^2} = -\frac{1}{2}\,h'\,c^2 + \left(rh - \frac{1}{2}\,h'r^2\right)\left(\frac{d\phi}{d\tau}\right)^2 . \tag{S20.10}$$

This is an acceptable answer. One can simplify (S20.10) further by noting that $h' = R_S/r^2$ and $rh = r - R_S$:

$$\frac{d^2r}{d\tau^2} = -\frac{1}{2}\frac{R_Sc^2}{r^2} + \left(r - \frac{3}{2}R_S\right)\left(\frac{d\phi}{d\tau}\right)^2 . \tag{S20.11}$$

In the notation of the problem statement, we have

$$f_0(r) = -\frac{1}{2}\frac{R_Sc^2}{r^2}\,, \quad f_1(r) = r - \frac{3}{2}R_S\,. \tag{S20.12}$$

(c) The geodesic equation (S20.6) for $\mu = \phi$ gives

$$\frac{d}{d\tau}\left[g_{\phi\phi}\,\frac{d\phi}{d\tau}\right] = \frac{1}{2}\frac{\partial g_{\lambda\sigma}}{\partial\phi}\,\frac{dx^\lambda}{d\tau}\frac{dx^\sigma}{d\tau}\,.$$

Since no metric component depends on ϕ, the right-hand side vanishes and we get:

$$\frac{d}{d\tau}\left[r^2\,\frac{d\phi}{d\tau}\right] = 0 \quad \rightarrow \quad \frac{d}{d\tau}L = 0\,, \quad \text{where} \quad L \equiv r^2\,\frac{d\phi}{d\tau}\,. \tag{S20.13}$$

The quantity L is a constant of the motion, namely, it is a number independent of τ.

(d) Using (S20.13) the second-order differential equation (S20.11) for $r(\tau)$ takes the form stated in the problem:

$$\frac{d^2r}{d\tau^2} = f_0(r) + \frac{f_1(r)}{r^4}L^2 \equiv H(r)\,, \tag{S20.14}$$

where we have introduced the function $H(r)$ (recall that L is a constant!). The differential equation then takes the form

$$\frac{d^2r}{d\tau^2} = H(r)\,. \tag{S20.15}$$

Since we are told that a circular orbit with radius r_0 exists, the function $r(\tau) = r_0$ must solve this equation. Being the constant function, the left-hand side vanishes and, consequently, the right-hand side must also vanish:

$$H(r_0) = f_0(r_0) + \frac{f_1(r_0)}{r_0^4}L^2 = 0\,. \tag{S20.16}$$

To investigate stability we consider a small perturbation $\delta r(\tau)$ of the orbit:

$$r(\tau) = r_0 + \delta r(\tau)\,, \quad \text{with} \quad \delta r(\tau) \ll r_0 \ \text{ at some initial } \tau\,.$$

Substituting this into (S20.15) we get, to first nontrivial approximation

$$\frac{d^2\delta r}{d\tau^2} = H(r_0 + \delta r) \simeq H(r_0) + \delta r H'(r_0) = \delta r\, H'(r_0)\,,$$

where $H'(r) = \frac{dH(r)}{dr}$ and we used $H(r_0) = 0$ from (S20.16). The resulting equation

$$\frac{d^2\delta r(\tau)}{d\tau^2} = H'(r_0)\,\delta r(\tau)\,, \tag{S20.17}$$

is familiar because $H'(r_0)$ is just a number. The condition of stability is that this number is negative: $H'(r_0) < 0$. Indeed, in this case (S20.17) is the harmonic oscillator equation

$$\frac{d^2x}{dt^2} = -\omega^2 x\,, \quad \text{with replacements} \quad x \leftrightarrow \delta r, \quad t \leftrightarrow \tau\,, \quad -\omega^2 \leftrightarrow H'(r_0)\,,$$

and the solution describes bounded oscillations. So stability requires:

$$\text{Stability Condition:}\quad H'(r_0) = \frac{d}{dr}\left[f_0(r) + \frac{f_1(r)}{r^4}L^2\right]_{r=r_0} < 0\,. \tag{S20.18}$$

This is the answer to part (d).

– –

For students interested in getting the famous result that orbits are stable for $r > 3R_S$ we complete this part of the analysis below. First we evaluate $H'(r_0)$ in (S20.18) using the values of f_0 and f_1 in (S20.12):

$$H'(r_0) = \frac{d}{dr}\left[-\frac{1}{2}\frac{R_S c^2}{r^2} + \left(\frac{1}{r^3} - \frac{3R_S}{2r^4}\right)L^2\right]_{r=r_0} = \frac{R_S c^2}{r_0^3} - \frac{3L^2}{r_0^5}(r_0 - 2R_S)\,.$$

The inequality in (S20.18) then gives us

$$R_S c^2 - \frac{3L^2}{r_0^2}(r_0 - 2R_S) < 0\,, \tag{S20.19}$$

where we multiplied by $r_0^3 > 0$. To complete the calculation we need the value of L^2 for the orbit with radius r_0. This value is determined by the vanishing of $H(r_0)$:

$$-\frac{1}{2}\frac{R_S c^2}{r_0^2} + (r_0 - \frac{3}{2}R_S)\frac{L^2}{r_0^4} = 0 \quad\rightarrow\quad \frac{L^2}{r_0^2} = \frac{1}{2}\frac{R_S c^2}{(r_0 - \frac{3}{2}R_S)}\,.$$

Note, incidentally, that the equality to the right demands that for a circular orbit $r_0 > \frac{3}{2}R_S$. Substituting the above value of L^2/r_0^2 in (S20.19) we get:

$$R_S c^2 - \frac{3}{2}\frac{R_S c^2}{(r_0 - \frac{3}{2}R_S)}(r_0 - 2R_S) < 0\,.$$

Cancelling the common factors of $R_S c^2$ we find

$$1 - \frac{3}{2}\frac{(r_0 - 2R_S)}{(r_0 - \frac{3}{2}R_S)} < 0\,,$$

which is equivalent to

$$\frac{3}{2}\frac{(r_0 - 2R_S)}{(r_0 - \frac{3}{2}R_S)} > 1\,.$$

For $r_0 > \frac{3}{2}R_S$, we get

$$3(r_0 - 2R_S) > 2(r_0 - \frac{3}{2}R_S) \quad\rightarrow\quad r_0 > 3R_S\,. \tag{S20.20}$$

This is the desired condition for stable orbits in the Schwarzschild geometry.

PROBLEM 21: PRESSURE AND ENERGY DENSITY OF MYSTERIOUS STUFF

(a) If $u \propto 1/\sqrt{V}$, then one can write

$$u(V + \Delta V) = u_0 \sqrt{\frac{V}{V + \Delta V}} \; .$$

(The above expression is proportional to $1/\sqrt{V + \Delta V}$, and reduces to $u = u_0$ when $\Delta V = 0$.) Expanding to first order in ΔV,

$$u = \frac{u_0}{\sqrt{1 + \frac{\Delta V}{V}}} = \frac{u_0}{1 + \frac{1}{2}\frac{\Delta V}{V}} = u_0 \left(1 - \frac{1}{2}\frac{\Delta V}{V}\right) \; .$$

The total energy is the energy density times the volume, so

$$U = u(V + \Delta V) = u_0 \left(1 - \frac{1}{2}\frac{\Delta V}{V}\right) V \left(1 + \frac{\Delta V}{V}\right) = U_0 \left(1 + \frac{1}{2}\frac{\Delta V}{V}\right) \; ,$$

where $U_0 = u_0 V$. Then

$$\boxed{\Delta U = \frac{1}{2}\frac{\Delta V}{V} U_0 \; .}$$

(b) The work done by the agent must be the negative of the work done by the gas, which is $p\,\Delta V$. So

$$\boxed{\Delta W = -p\,\Delta V} \; .$$

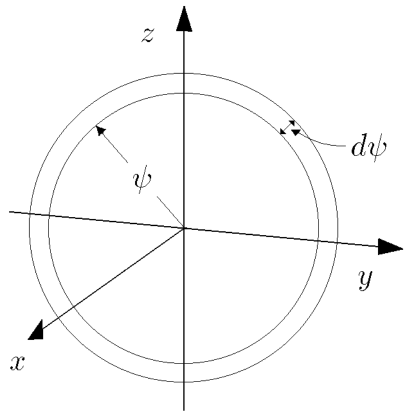
(c) The agent must supply the full change in energy, so

$$\Delta W = \Delta U = \frac{1}{2}\frac{\Delta V}{V} U_0 \; .$$

Combining this with the expression for ΔW from part (b), one sees immediately that

$$p = -\frac{1}{2}\frac{U_0}{V} = \boxed{-\frac{1}{2} u_0 \; .}$$

PROBLEM 22: VOLUME OF A CLOSED THREE-DIMENSIONAL SPACE
(15 points)



The metric for the space that we are considering is

$$ds^2 = R^2 \left[d\psi^2 + f^2(\psi) \left(d\theta^2 + \sin^2\theta d\phi^2 \right) \right] \quad ,$$

For comparison, the metric for the surface of a sphere of radius R is given by

$$ds^2 = R^2 \left(d\theta^2 + \sin^2\theta d\phi^2 \right) \quad .$$

By comparing these two, one sees that the set of points described by $\psi =$ constant (varying θ and ϕ) has the same metric as a sphere of radius $r = Rf(\psi)$. We can save ourselves some trouble in calculating by remembering that the area of such a spherical surface of radius r is $4\pi r^2 = 4\pi R^2 f^2(\psi)$.

The volume of the spherical shell shown in the problem is just the area times the thickness. The thickness is not $d\psi$, since ψ is only a coordinate — remember that in curved space a coordinate and a distance are two different things. The distance is given by the metric. Consider in this case a radial line extending from ψ to $\psi + d\psi$, at constant θ and ϕ. Then

$$ds^2 = R^2 d\psi^2 \quad ,$$

and so the length of the line segment is $ds = R d\psi$.

The volume of the spherical shell is then given by

$$dV = \left[4\pi R^2 f^2(\psi) \right] R d\psi \quad .$$

We must now integrate over the range of ψ, for 0 to π. So,

$$\boxed{V = 4\pi R^3 \int_0^\pi f^2(\psi)\, d\psi \quad .}$$

PROBLEM 23: GRAVITATIONAL BENDING OF LIGHT *(30 points)*

(a) *(6 points)* Note that

$$\begin{aligned}\mathrm{d}r^2 &= \frac{1}{r^2}\left(x\,\mathrm{d}x + y\,\mathrm{d}y + z\,\mathrm{d}z\right)^2 \\ &= \frac{1}{r^2}\left(x^2\,\mathrm{d}x^2 + y^2\,\mathrm{d}y^2 + z^2\,\mathrm{d}z^2 + 2xy\,\mathrm{d}x\,\mathrm{d}y + 2xz\,\mathrm{d}x\,\mathrm{d}z + 2yz\,\mathrm{d}y\,\mathrm{d}z\right) \ .\end{aligned} \tag{S23.1}$$

By using this expression for $(\mathrm{d}r)^2$ in Eq. (P23.5), we have the full expression for $\mathrm{d}s^2$ written out, from which we can read off the components of $g_{\mu\nu}$:

$$\boxed{\begin{aligned} g_{tt} &= \text{coefficient of } \mathrm{d}t^2 = -c^2\left(1 - \frac{R_{\mathrm{Sch}}}{r}\right) \\ g_{xx} &= \text{coefficient of } \mathrm{d}x^2 = 1 + \frac{R_{\mathrm{Sch}}}{r^3}\,x^2 \\ g_{xy} &= \frac{1}{2} \text{ of coefficient of } \mathrm{d}x\,\mathrm{d}y = \frac{R_{\mathrm{Sch}}}{r^3}\,xy \ . \end{aligned}} \tag{S23.2}$$

A number of people missed the factor of 1/2 in the value of g_{xy}. It arises because the general formula is written as $\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu$, which when expanded becomes

$$\mathrm{d}s^2 = g_{xx}\mathrm{d}x^2 + g_{yy}\mathrm{d}y^2 + g_{zz}\mathrm{d}z^2 + g_{tt}\mathrm{d}t^2 + g_{xy}\mathrm{d}x\,\mathrm{d}y + g_{yx}\mathrm{d}y\,\mathrm{d}x + \ldots \ .$$

Since $\mathrm{d}x\,\mathrm{d}y = \mathrm{d}y\,\mathrm{d}x$, the coefficient of $\mathrm{d}x\,\mathrm{d}y$ is $g_{xy} + g_{yx} = 2g_{xy}$.

(b) *(9 points)* It will be useful to know the derivatives of r:

$$\begin{aligned}\frac{\partial r}{\partial x} &= \frac{\partial}{\partial x}(x^2+y^2+z^2)^{1/2} \\ &= \frac{1}{2}(x^2+y^2+z^2)^{-1/2}\frac{\partial}{\partial x}(x^2+y^2+z^2) = \frac{x}{r} \ .\end{aligned} \tag{S23.3}$$

Similarly,

$$\frac{\partial r}{\partial y} = \frac{y}{r} \qquad \text{and} \qquad \frac{\partial r}{\partial z} = \frac{z}{r} \ , \tag{S23.4}$$

and

$$\begin{aligned}\frac{\mathrm{d}r}{\mathrm{d}\lambda} &= \frac{\partial r}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}\lambda} + \frac{\partial r}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}\lambda} + \frac{\partial r}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}\lambda} \\ &= \frac{x}{r} \ .\end{aligned} \tag{S23.5}$$

In the 2nd line I used the value of $\partial r/\partial x$ from Eq. (S23.3), and the derivatives

$$\frac{\mathrm{d}x}{\mathrm{d}\lambda} = 1 \ , \qquad \frac{\mathrm{d}y}{\mathrm{d}\lambda} = \frac{\mathrm{d}z}{\mathrm{d}\lambda} = 0 \tag{S23.6}$$

that can be found from Eq. (P23.8).

Now, to expand the left-hand side of the geodesic equation:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\lambda}\left\{g_{\mu\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}\right\} &= \frac{\mathrm{d}}{\mathrm{d}\lambda}\left\{g_{yy}\frac{\mathrm{d}y}{\mathrm{d}\lambda} + g_{yx}\frac{\mathrm{d}x}{\mathrm{d}\lambda}\right\} \\
&= \frac{\mathrm{d}}{\mathrm{d}\lambda}\left\{\left[1 + \frac{R_{\mathrm{Sch}}}{r^3}y^2\right]\frac{\mathrm{d}y}{\mathrm{d}\lambda} + \frac{R_{\mathrm{Sch}}}{r^3}xy\frac{\mathrm{d}x}{\mathrm{d}\lambda}\right\} \\
&= \frac{\mathrm{d}^2 y}{\mathrm{d}\lambda^2} - 3\frac{R_{\mathrm{Sch}}}{r^4}\frac{x}{r}xy\frac{\mathrm{d}x}{\mathrm{d}\lambda} + \frac{R_{\mathrm{Sch}}}{r^3}\frac{\mathrm{d}x}{\mathrm{d}\lambda}y\frac{\mathrm{d}x}{\mathrm{d}\lambda} \\
&= \boxed{\frac{\mathrm{d}^2 y}{\mathrm{d}\lambda^2} - 3\frac{R_{\mathrm{Sch}}b}{r^5}x^2 + \frac{R_{\mathrm{Sch}}b}{r^3}\ .}
\end{aligned} \tag{S23.7}$$

Note that I dropped a term

$$\frac{R_{\mathrm{Sch}}y^2}{r^3}\frac{\mathrm{d}^2 y}{\mathrm{d}\lambda^2}$$

and a term

$$\frac{R_{\mathrm{Sch}}}{r^3}xy\frac{\mathrm{d}^2 x}{\mathrm{d}\lambda^2}\ ,$$

which is justified because the acceleration $\frac{\mathrm{d}^2 y}{\mathrm{d}\lambda^2}$ will be proportional to G, and R_{Sch} is proportional to G, so this term is 2nd order in G. The problem stated that we are to work to first order in G. No points were taken off, however, from students who retained these or other negligible terms.

Note, however, that $\mathrm{d}^2 y\ \mathrm{d}\lambda^2$ is not negligible, and appears in the answer. This is because $\mathrm{d}y/\mathrm{d}\lambda$ is not actually zero, but is of order G. $\mathrm{d}y/\mathrm{d}\lambda$ is zero for the unperturbed path, but in reality the photon picks up a small velocity in the y-direction, caused by the gravitational attraction of the Sun and proportional to G. $\mathrm{d}^2 y/\mathrm{d}\lambda^2$ will also be proportional to G. When $\mathrm{d}y/\mathrm{d}\lambda$ multiplies a factor proportional to R_{Sch}, the product is of order G^2 and hence negligible. But $\mathrm{d}^2 y/\mathrm{d}\lambda^2$ by itself is of order G and is not negligible.

Note on propagation of errors: I normally do not take off points for propagating errors, so for example a student who forgot the factor of 1/2 in determining g_{xy} would get full credit on part (b), even though the answer would contain terms that are wrong by a factor of 1/2. However, it seems right to me to make an exception to this rule in cases where an error on part (a) causes the consequent answer on a later part to become trivial. For example, if a student described a metric in part (a) which had no dependence on r, then many of the terms in parts (b) and (c) would not be present. In such cases I still took off points in parts (b) and (c), because it didn't seem fair to me to give such a student credit for calculating these terms, when the student exhibited no such capability.

(c) *(9 points)*

$$\begin{aligned}
\frac{1}{2}\frac{\partial}{\partial y}\left(g_{\sigma\tau}\right)\frac{\mathrm{d}x^\sigma}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\tau}{\mathrm{d}\lambda} &= \frac{1}{2}\frac{\partial}{\partial y}(g_{xx})\left(\frac{\mathrm{d}x}{\mathrm{d}\lambda}\right)^2 + \frac{1}{2}\frac{\partial}{\partial y}(g_{tt})\left(\frac{\mathrm{d}t}{\mathrm{d}\lambda}\right)^2 \\
&= \frac{1}{2}\frac{\partial}{\partial y}\left(1 + \frac{R_{\mathrm{Sch}}}{r^3}x^2\right) - \frac{1}{2}c^2\frac{\partial}{\partial y}\left(1 - \frac{R_{\mathrm{Sch}}}{r}\right)\left(\frac{1}{c^2}\right) \\
&= -\frac{1}{2}\left(3\frac{R_{\mathrm{Sch}}}{r^4}\frac{y}{r}x^2\right) - \frac{1}{2}\left(\frac{R_{\mathrm{Sch}}}{r^2}\frac{y}{r}\right) \\
&= \boxed{-\frac{3}{2}\frac{R_{\mathrm{Sch}}b}{r^5}x^2 - \frac{1}{2}\frac{R_{\mathrm{Sch}}b}{r^3}\ .}
\end{aligned} \tag{S23.8}$$

(d) *(2 points)* Combining Eqs. (S23.7) and (S23.8), we find

$$\begin{aligned}
\frac{\mathrm{d}^2y}{\mathrm{d}\lambda^2} &= -\frac{3}{2}\frac{R_{\mathrm{Sch}}b}{r^5}x^2 - \frac{1}{2}\frac{R_{\mathrm{Sch}}b}{r^3} + 3\frac{R_{\mathrm{Sch}}b}{r^5}x^2 - \frac{R_{\mathrm{Sch}}b}{r^3} \\
&= \boxed{\frac{3}{2}R_{\mathrm{Sch}}b\left[\frac{x^2}{r^5} - \frac{1}{r^3}\right]\ .}
\end{aligned} \tag{S23.9}$$

(e) *(4 points)* The final value of $\mathrm{d}y/\mathrm{d}\lambda$ is given by Eq. (P23.9), while the final value of $\mathrm{d}x/\mathrm{d}\lambda$ will be equal to 1, at least up to possible corrections proportional to G. Thus, the final velocity will make an angle α relative to the horizontal, where

$$\begin{aligned}
\tan\alpha &= \frac{\mathrm{d}y/d\lambda|_{\mathrm{final}}}{\mathrm{d}x/d\lambda|_{\mathrm{final}}} \\
&= \int_{-\infty}^{\infty}\frac{\mathrm{d}^2y}{\mathrm{d}\lambda^2}\,\mathrm{d}\lambda\ .
\end{aligned}$$

Since $\tan\alpha$ will be proportional to G, the small angle approximation $\tan\alpha = \alpha$ will apply, and

$$\boxed{\alpha \approx \int_{-\infty}^{\infty}\frac{\mathrm{d}^2y}{\mathrm{d}\lambda^2}\,\mathrm{d}\lambda\ .} \tag{S23.10}$$

Then, using Eqs. (S23.9) and combining with Eqs. (P23.4) and (P23.8),

$$\begin{aligned}\alpha &= \frac{3}{2} R_{\text{Sch}} b \int_{-\infty}^{\infty} \left[\frac{x^2}{r^5} - \frac{1}{r^3} \right] \mathrm{d}\lambda \\ &= \boxed{\frac{3}{2} R_{\text{Sch}} b \int_{-\infty}^{\infty} \left[\frac{\lambda^2}{(\lambda^2+b^2)^{5/2}} - \frac{1}{(\lambda^2+b^2)^{3/2}} \right] \mathrm{d}\lambda \ .}\end{aligned} \tag{S23.11}$$

You were not asked to carry out these integrals, but using the table of integrals given with the problem, one finds

$$\alpha = \frac{3}{2} R_{\text{Sch}} b \left[\frac{2}{3b^2} - \frac{2}{b^2} \right] = -\frac{2R_{\text{Sch}}}{b} = -\frac{4GM}{c^2 b} \ . \tag{S23.12}$$

The minus sign indicates that the deflection is downward, as one would expect.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — Quiz Date: November 5, 2018

Prof. Alan Guth

QUIZ 2

Reformatted to Remove Blank Pages

Please answer all questions in this stapled booklet.

PROBLEM 1: DID YOU DO THE READING? *(25 points)*

(a) *(4 points)* Which of the following statements about deuterium is **NOT** true? Choose one.

(i) The abundance of deuterium in the universe tends to decrease with time, because deuterium is very easily destroyed in stars.

(ii) The most promising way to find the primordial value of deuterium abundance is to look at the spectra of distant quasars to estimate the abundance of deuterium within the quasar itself.

(iii) The Lyman-α transition in deuterium corresponds to a slightly different wavelength than the Lyman-α transition in hydrogen.

(iv) Deuterium plays an important role in forming helium in the early universe mainly by producing tritium or ^{3}He.

(b) *(6 points)* In Chapter 5 of *The First Three Minutes*, Steven Weinberg describes the first three minutes of the history of the universe. Choose two correct statements about the first three minutes. You can assume the fraction by weight of primordial helium is 26 percent. (3 points for each right answer, no penalty for guessing.)

(i) When the temperature of the universe was about 10^{10} °K (t $\sim$ 1 sec), neutrinos and antineutrinos started to behave as free particles, no longer having significant interactions with electrons, positrons, or photons.

(ii) After the neutrinos decoupled from the photons, the temperature of the neutrinos was higher than that of the photons because neutrinos interacted less with other particles as the universe expanded.

(iii) Most of the atoms heavier than helium were made through nucleosynthesis during the first three minutes, and this is why we call this period the era of nucleosynthesis.

(iv) After the first three minutes, there were about 7 times more protons than neutrons, and the ratio of protons to neutrons has been almost preserved until today.

(v) The protons and neutrons became decoupled from the photons after the first three minutes, because the number densities of protons and neutrons were decreased by the formation of helium, and so their interactions with photons became negligible.

(vi) The observed abundance of helium in a galaxy today is much larger than the abundance of primordial helium, because helium is continuously formed inside stars by nuclear fusion.

(c) *(4 points)* The cosmic microwave background radiation was first discovered by Penzias and Wilson in 1964. However, according to Chapter 6 of *The First Three Minutes*, a team at the MIT Radiation Laboratory led by Robert Dicke was able to set an upper limit on any isotropic extraterrestrial radiation background, showing that the equivalent temperature was less than 20 °K at wavelengths of 1.00, 1.25, and 1.50 centimeters. This measurement was made in the

(i) 1920s (ii) 1930s (iii) 1940s (iv) 1950s (v) 1960s

(d) *(5 points)* A free neutron can radioactively decay into a proton, plus two other particles. What are these particles? Give the charge, baryon number, and lepton number for each of these particles, verifying that each of these quantities is conserved in this process.

(e) *(6 points)* In Chapter 8 of Ryden's *Introduction to Cosmology*, she discusses three ways to measure the dark matter in clusters. Give a brief, qualitative description of TWO of them. (If you give three descriptions, only the first two will be graded!)

— End of Problem 1. —

PROBLEM 2: EVOLUTION OF A FRIEDMANN–ROBERTSON–WALKER UNIVERSE *(20 points)*

(a) *(10 points)* The evolution of a homogeneous isotropic model of the universe, known as a Friedmann–Robertson-Walker (FRW) universe, can be described by the following equations:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3} G\rho - \frac{kc^2}{a^2} \;, \tag{2.1}$$

$$\ddot{a} = -\frac{4\pi}{3} G \left(\rho + \frac{3p}{c^2}\right) a \;, \tag{2.2}$$

and

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) \;. \tag{2.3}$$

These equations are not independent, but any two can in fact be used to derive the third. For example, in Problem Set 6, you were asked to use Eqs. (2.1) and (2.3) to derive Eq. (2.2). Here you are asked to show that Eqs. (2.1) and (2.2) can be used to derive Eq. (2.3).

(b) *(6 points)* If the universe were flat, expanding, and filled with a material for which

$$p = -\rho c^2 \;, \tag{2.4}$$

what would be the form of the scale factor $a(t)$?

(c) *(4 points)* In the universe described in part (b), suppose that, at $t = 0$, my friend Bob emits a photon in my direction. Show that if Bob is more than a certain distance away from me at the time of emission, $t = 0$, then the photon will never reach me. What is this distance?

— End of Problem 2. —

PROBLEM 3: EVOLUTION OF AN OPEN, MATTER-DOMINATED UNIVERSE *(35 points)*

The following problem was Problem 2 on Problem Set 4.

The equations describing the evolution of an open, matter-dominated universe were given in Lecture Notes 4 as

$$ct = \alpha\,(\sinh\theta - \theta)$$

and

$$\frac{a}{\sqrt{\kappa}} = \alpha\,(\cosh\theta - 1) \ ,$$

where α is a constant with units of length. The following mathematical identities, which you should know, may also prove useful on parts (e) and (f):

$$\sinh\theta = \frac{e^{\theta} - e^{-\theta}}{2} \quad , \quad \cosh\theta = \frac{e^{\theta} + e^{-\theta}}{2}$$

$$e^{\theta} = 1 + \frac{\theta}{1!} + \frac{\theta^2}{2!} + \frac{\theta^3}{3!} + \ldots \ .$$

(a) *(5 points)* Find the Hubble expansion rate H as a function of α and θ.

(b) *(5 points)* Find the mass density ρ as a function of α and θ.

(c) *(5 points)* Find the mass density parameter Ω as a function of α and θ.

(d) *(6 points)* Find the physical value of the horizon distance, $\ell_{p,\text{horizon}}$, as a function of α and θ.

(e) *(7 points)* For very small values of t, it is possible to use the first nonzero term of a power-series expansion to express θ as a function of t, and then a as a function of t. Give the expression for $a(t)$ in this approximation. The approximation will be valid for $t \ll t^*$. Estimate the value of t^*.

(f) *(7 points)* Even though these equations describe an open universe, one still finds that Ω approaches one for very early times. For $t \ll t^*$ (where t^* is defined in part (e)), the quantity $1 - \Omega$ behaves as a power of t. Find the expression for $1 - \Omega$ in this approximation.

— End of Problem 3. —

PROBLEM 4: RADIAL GEODESICS IN A CLOSED UNIVERSE *(20 points)*

As shown in the formula sheets, we can describe a closed universe by choosing $k = 1$, and then using coordinates (t, r, θ, ϕ), with metric

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 = -c^2\,\mathrm{d}t^2 + a^2(t)\left\{\frac{\mathrm{d}r^2}{1-r^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\} , \tag{4.1}$$

or by using coordinates (t, ψ, θ, ϕ), with metric

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 \equiv -c^2\,\mathrm{d}t^2 + a^2(t)\left\{\mathrm{d}\psi^2 + \sin^2\psi\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\} . \tag{4.2}$$

The connection between the two coordinate systems is given by

$$r = \sin\psi . \tag{4.3}$$

The general spacetime geodesic equation can be written as

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{g_{\mu\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\right\} = \frac{1}{2}\left(\partial_\mu g_{\lambda\sigma}\right)\frac{\mathrm{d}x^\lambda}{\mathrm{d}\tau}\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau} . \tag{4.4}$$

(a) *(7 points)* Using the coordinates (t, ψ, θ, ϕ) and the metric of Eq. (4.2), compute explicitly the geodesic equation for $\mu = \psi$. By "compute explicitly", I mean that $g_{\mu\nu}$ should be replaced by the relevant expressions from Eq. (4.2), and that the sums over indices should be written out, including only the nonzero terms.

(b) *(7 points)* Using instead the coordinates (t, r, θ, ϕ), compute explicitly the geodesic equation for $\mu = r$.

(c) *(6 points)* Are the results from parts (a) and (b) both valid, or is one valid and the other not? If you believe that they are both valid, use Eq. (4.3) to show that they are equivalent. If you believe that only one is valid, state which one is valid, and explain why the other is not. (4 points will be given for showing the correct understanding of this problem, with 2 points allocated to completing the algebra needed to demonstrate your answer.)

— End of Problem 4. —

Problem	Maximum	Score	Initials
1	25		
2	20		
3	35		
4	20		
TOTAL	100		

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — Quiz Date: November 5, 2018
Prof. Alan Guth

QUIZ 2 FORMULA SHEET

DOPPLER SHIFT (For motion along a line):

$$z = v/u \quad \text{(nonrelativistic, source moving)}$$

$$z = \frac{v/u}{1 - v/u} \quad \text{(nonrelativistic, observer moving)}$$

$$z = \sqrt{\frac{1+\beta}{1-\beta}} - 1 \quad \text{(special relativity, with } \beta = v/c\text{)}$$

COSMOLOGICAL REDSHIFT:

$$1 + z \equiv \frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}} = \frac{a(t_{\text{observed}})}{a(t_{\text{emitted}})}$$

SPECIAL RELATIVITY:

Time Dilation Factor:

$$\gamma \equiv \frac{1}{\sqrt{1-\beta^2}} \ , \qquad \beta \equiv v/c$$

Lorentz-Fitzgerald Contraction Factor: γ

Relativity of Simultaneity:
Trailing clock reads later by an amount $\beta\ell_0/c$.

Energy-Momentum Four-Vector:

$$p^\mu = \left(\frac{E}{c}, \vec{p}\right) \ , \quad \vec{p} = \gamma m_0 \vec{v} \ , \quad E = \gamma m_0 c^2 = \sqrt{\left(m_0 c^2\right)^2 + |\vec{p}|^2 c^2} \ ,$$

$$p^2 \equiv |\vec{p}|^2 - \left(p^0\right)^2 = |\vec{p}|^2 - \frac{E^2}{c^2} = -\left(m_0 c\right)^2 \ .$$

KINEMATICS OF A HOMOGENEOUSLY EXPANDING UNIVERSE:

Hubble's Law: $v = Hr$,
where v = recession velocity of a distant object, H = Hubble expansion rate, and r = distance to the distant object.

Present Value of Hubble Expansion Rate (Planck 2018):

$$H_0 = 67.66 \pm 0.42 \text{ km-s}^{-1}\text{-Mpc}^{-1}$$

Scale Factor: $\ell_p(t) = a(t)\ell_c$,
where $\ell_p(t)$ is the physical distance between any two objects, $a(t)$ is the scale factor, and ℓ_c is the coordinate distance between the objects, also called the comoving distance.

Hubble Expansion Rate: $H(t) = \dfrac{1}{a(t)}\dfrac{\mathrm{d}a(t)}{\mathrm{d}t}$.

Light Rays in Comoving Coordinates: Light rays travel in straight lines with physical speed c relative to any observer. In Cartesian coordinates, coordinate speed $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \dfrac{c}{a(t)}$. In general, $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = 0$.

Horizon Distance:

$$\begin{aligned}\ell_{p,\text{horizon}}(t) &= a(t)\int_0^t \frac{c}{a(t')}dt' \\ &= \begin{cases} 3ct & \text{(flat, matter-dominated)}, \\ 2ct & \text{(flat, radiation-dominated)}. \end{cases}\end{aligned}$$

COSMOLOGICAL EVOLUTION:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2} \ , \quad \ddot{a} = -\frac{4\pi}{3}G\left(\rho + \frac{3p}{c^2}\right)a \ ,$$

$$\rho_m(t) = \frac{a^3(t_i)}{a^3(t)}\rho_m(t_i) \ \text{(matter)}, \quad \rho_r(t) = \frac{a^4(t_i)}{a^4(t)}\rho_r(t_i) \ \text{(radiation)}.$$

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) \ , \quad \Omega \equiv \rho/\rho_c \ , \quad \text{where} \quad \rho_c = \frac{3H^2}{8\pi G} \ .$$

EVOLUTION OF A MATTER-DOMINATED UNIVERSE:

Flat ($k = 0$): $a(t) \propto t^{2/3}$

$\Omega = 1$.

Closed ($k > 0$): $ct = \alpha(\theta - \sin\theta)$, $\frac{a}{\sqrt{k}} = \alpha(1 - \cos\theta)$,

$\Omega = \frac{2}{1 + \cos\theta} > 1$,

where $\alpha \equiv \frac{4\pi}{3}\frac{G\rho}{c^2}\left(\frac{a}{\sqrt{k}}\right)^3$.

Open ($k < 0$): $ct = \alpha\left(\sinh\theta - \theta\right)$, $\frac{a}{\sqrt{\kappa}} = \alpha\left(\cosh\theta - 1\right)$,

$\Omega = \frac{2}{1 + \cosh\theta} < 1$,

where $\alpha \equiv \frac{4\pi}{3}\frac{G\rho}{c^2}\left(\frac{a}{\sqrt{\kappa}}\right)^3$,

$\kappa \equiv -k > 0$.

MINKOWSKI METRIC (Special Relativity):

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \ .$$

ROBERTSON-WALKER METRIC:

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 = -c^2\,\mathrm{d}t^2 + a^2(t)\left\{\frac{\mathrm{d}r^2}{1 - kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\} \ .$$

Alternatively, for $k > 0$, we can define $r = \frac{\sin\psi}{\sqrt{k}}$, and then

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 \equiv -c^2\,\mathrm{d}t^2 + \tilde{a}^2(t)\left\{\mathrm{d}\psi^2 + \sin^2\psi\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\} \ ,$$

where $\tilde{a}(t) = a(t)/\sqrt{k}$. For $k < 0$ we can define $r = \frac{\sinh\psi}{\sqrt{-k}}$, and then

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 = -c^2\,\mathrm{d}t^2 + \tilde{a}^2(t)\left\{\mathrm{d}\psi^2 + \sinh^2\psi\left(\mathrm{d}\theta^2 + \sin^2\theta\,d\phi^2\right)\right\} \ ,$$

where $\tilde{a}(t) = a(t)/\sqrt{-k}$. Note that $\tilde{a}$ can be called a if there is no need to relate it to the $a(t)$ that appears in the first equation above.

SCHWARZSCHILD METRIC:

$$\mathrm{d}s^2 \equiv -c^2\mathrm{d}\tau^2 = -\left(1 - \frac{2GM}{rc^2}\right)c^2\mathrm{d}t^2 + \left(1 - \frac{2GM}{rc^2}\right)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2 \ ,$$

GEODESIC EQUATION:

$$\frac{\mathrm{d}}{\mathrm{d}s}\left\{g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}s}\right\} = \frac{1}{2}\left(\partial_i g_{k\ell}\right)\frac{\mathrm{d}x^k}{\mathrm{d}s}\,\frac{\mathrm{d}x^\ell}{\mathrm{d}s}$$

or:
$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{g_{\mu\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\right\} = \frac{1}{2}\left(\partial_\mu g_{\lambda\sigma}\right)\frac{\mathrm{d}x^\lambda}{\mathrm{d}\tau}\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau}$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — November 20, 2018
Prof. Alan Guth

QUIZ 2 SOLUTIONS

Quiz Date: November 5, 2018
Reformatted to Remove Blank Pages

Please answer all questions in this stapled booklet.

PROBLEM 1: DID YOU DO THE READING? *(25 points)*

(a) *(4 points)* Which of the following statements about deuterium is **NOT** true? Choose one.

(i) The abundance of deuterium in the universe tends to decrease with time, because deuterium is very easily destroyed in stars.

$\boxed{\text{(ii)}}$ The most promising way to find the primordial value of deuterium abundance is to look at the spectra of distant quasars to estimate the abundance of deuterium within the quasar itself.

(iii) The Lyman-α transition in deuterium corresponds to a slightly different wavelength than the Lyman-α transition in hydrogen.

(iv) Deuterium plays an important role in forming helium in the early universe mainly by producing tritium or ^{3}He.

[Comment: The most promising way to find the primordial value of the deuterium abundance is to look at the spectra of distant quasars to estimate the abundance of deuterium in ***intergalactic gas clouds*** *that lie between the quasars and us.]*

(b) *(6 points)* In Chapter 5 of *The First Three Minutes*, Steven Weinberg describes the first three minutes of the history of the universe. Choose two correct statements about the first three minutes. You can assume the fraction by weight of primordial helium is 26 percent. (3 points for each right answer, no penalty for guessing.)

$\boxed{\text{(i)}}$ When the temperature of the universe was about 10^{10} °K (t $\sim$ 1 sec), neutrinos and antineutrinos started to behave as free particles, no longer having significant interactions with electrons, positrons, or photons.

(ii) After the neutrinos decoupled from the photons, the temperature of the neutrinos was higher than that of the photons because neutrinos interacted less with other particles as the universe expanded.

(iii) Most of the atoms heavier than helium were made through nucleosynthesis during the first three minutes, and this is why we call this period the era of nucleosynthesis.

$\boxed{\text{(iv)}}$ After the first three minutes, there were about 7 times more protons than neutrons, and the ratio of protons to neutrons has been almost preserved until today.

(v) The protons and neutrons became decoupled from the photons after the first three minutes, because the number densities of protons and neutrons were decreased by the formation of helium, and so their interactions with photons became negligible.

(vi) The observed abundance of helium in a galaxy today is much larger than the abundance of primordial helium, because helium is continuously formed inside stars by nuclear fusion.

[Comment: The statement (i) is described in Chapter 5 of Weinberg's book, and it has also been discussed in class. The correctness of statement (iv) can also be seen from Weinberg's Chapter 5, which says that the fraction of neutrons after the first three minutes is around 14%, and it goes down to around 13% "a little later". Thus there were about 7 times more protons compared to neutrons. Weinberg also explains that most of the neutrons present at this time immdiately combined with protons to form helium, which causes the ratio to be nearly constant until today. If you did not remember Weinberg's numbers, the statement that the fraction by weight of primordial helium is 26% should allow you to determine the neutron to proton ratio, provided that you remember that the helium nucleus consists of 2 protons and two neutrons, that the mass of the proton and neutron are about equal, and that the remaining 74% of the matter is essentially hydrogen, with no neutrons. Thus helium is very nearly half protons and half neutrons by weight, so the neutrons must be about 13% of the matter in the universe. (Note that we are talking about fractions of the total "baryonic" matter, which does not include the dark matter.) (ii) is clearly false, because the temperature of neutrinos becomes lower than that of photons. (iii) is clearly false, because most atoms heavier than helium were made much later in the history of the universe, in the interiors of stars. (v) is false because protons did not decouple from photons until about 350,000 years, and it happened because the plasma of protons and electrons combined to form neutral hydrogen. (vi) is false because most of the helium in the universe today is primordial. Ryden points out, for example, that the abundance of helium in the Sun's atmosphere is only about 28%. Weinberg states, near the end of Chapter 5, that "the 20-30 percent helium abundance could not have been created recently without liberating enormous amounts of radiation that we do not observe."]

(c) *(4 points)* The cosmic microwave background radiation was first discovered by Penzias and Wilson in 1964. However, according to Chapter 6 of *The First Three Minutes*, a team at the MIT Radiation Laboratory led by Robert Dicke was able to set an upper limit on any isotropic extraterrestrial radiation background, showing that the equivalent temperature was less than 20 °K at wavelengths of 1.00, 1.25, and 1.50 centimeters. This measurement was made in the

(i) 1920s (ii) 1930s $\boxed{\text{(iii) 1940s}}$ (iv) 1950s (v) 1960s

(d) *(5 points)* A free neutron can radioactively decay into a proton, plus two other particles. What are these particles? Give the charge, baryon number, and lepton number for each of these particles, verifying that each of these quantities is conserved in this process.

Answer:

The neutron decays through the reaction

$$n \to p + e^- + \bar{\nu}_e \ .$$

The quantum numbers of these particles can be described by the following table:

Particle	Charge	Baryon Number	Lepton Number
Neutron (n)	0	1	0
Proton (p)	$+e$	1	0
Electron (e^-)	-e	0	1
Anti-electron-neutrino ($\bar{\nu}_e$)	0	0	-1

Thus, the total charge of the final state is zero, the total baryon number is 1, and the total lepton number is zero, in all cases matching the initial value of these quantities.

(e) *(6 points)* In Chapter 8 of Ryden's *Introduction to Cosmology*, she discusses three ways to measure the dark matter in clusters. Give a brief, qualitative description of TWO of them. (If you give three descriptions, only the first two will be graded!)

Answer: You should have given two of the following three items.

1) Virial theorem: The virial theorem relates the total kinetic energy of a steady-state cluster to is gravitational potential energy. Since the kinetic energy is proportional to the mass M of the cluster, while the potential energy is proportional to M^2, the relation will hold for only one value of M. By measuring the velocity dispersion (root mean square of the radial galaxy velocities relative to the mean radial velocity) and the size of the cluster, the mass can be inferred.

2) Hot, x-ray emitting gases: By measuring the x-rays emitted by the cluster, it is possible to model the density, temperature, and composition of the hot gas within the cluster (intracluster gas). By assuming that the gas is in hydrostatic equilibrium — i.e., by assuming that the pressure gradients balance the gravitational forces — one can infer the gravitational field, and hence the total mass.

3) Gravitational lensing: If one can find a galaxy behind the cluster, so that the image of the galaxy is gravitationally lensed, then the mass of the galaxy can be inferred by the degree to which the galaxy is lensed.

— End of Problem 1. —

PROBLEM 2: EVOLUTION OF A FRIEDMANN–ROBERTSON–WALKER UNIVERSE *(20 points)*

(a) *(10 points)* The evolution of a homogeneous isotropic model of the universe, known as a Friedmann–Robertson-Walker (FRW) universe, can be described by the following equations:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2} \ , \tag{2.1}$$

$$\ddot{a} = -\frac{4\pi}{3}G\left(\rho + \frac{3p}{c^2}\right)a \ , \tag{2.2}$$

and

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) \ . \tag{2.3}$$

These equations are not independent, but any two can in fact be used to derive the third. For example, in Problem Set 6, you were asked to use Eqs. (2.1) and (2.3) to derive Eq. (2.2). Here you are asked to show that Eqs. (2.1) and (2.2) can be used to derive Eq. (2.3).

Answer:

We can rewrite Eq. (2.1) as

$$\dot{a}^2 = \frac{8\pi}{3}G\rho a^2 - kc^2 \ ,$$

which can then be differentiated to give

$$2\dot{a}\ddot{a} = \frac{16\pi}{3}G\rho a\dot{a} + \frac{8\pi}{3}G\dot{\rho}a^2 \ .$$

The above equation can be solved for $\dot{\rho}$, giving

$$\dot{\rho} = -2\,\frac{\dot{a}}{a}\,\rho + \frac{3}{4\pi G}\frac{\ddot{a}\dot{a}}{a^2} \ .$$

Then if Eq. (2.2) is used to replace $\ddot{a}$, one finds

$$\begin{aligned}\dot{\rho} &= -2\,\frac{\dot{a}}{a}\,\rho - \frac{\dot{a}}{a}\left(\rho + \frac{3p}{c^2}\right)\\ &= -3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) \ .\end{aligned}$$

(b) *(6 points)* If the universe were flat, expanding, and filled with a material for which

$$p = -\rho c^2 \ , \tag{2.4}$$

what would be the form of the scale factor $a(t)$?

Answer:

If $p = -\rho c^2$, then Eq. (2.3) implies that

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) = 0 \ ,$$

so ρ is a constant. Then, since $k = 0$ for a flat universe, Eq. (2.1) implies

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho = \text{const} \ ,$$

so

$$\dot{a} = \pm\sqrt{\frac{8\pi}{3}G\rho}\, a \ .$$

Only the positive option describes an expanding universe, so

$$\dot{a} = \sqrt{\frac{8\pi}{3}G\rho}\, a \quad \Longrightarrow \quad \boxed{a(t) = b\, e^{\sqrt{\frac{8\pi}{3}G\rho}\, t} \ ,}$$

where b is an arbitrary constant.

(c) *(4 points)* In the universe described in part (b), suppose that, at $t = 0$, my friend Bob emits a photon in my direction. Show that if Bob is more than a certain distance away from me at the time of emission, $t = 0$, then the photon will never reach me. What is this distance?

Answer:

The metric for this universe is

$$\mathrm{d}s^2 = -c^2\,\mathrm{d}t^2 + b^2 e^{2Ht}\,\mathrm{d}\vec{x}^2 \ ,$$

where

$$H = \sqrt{\frac{8\pi}{3}G\rho} \ .$$

Light pulses travel with $\mathrm{d}s^2 = 0$, so the coordinate speed of light, for a pulse traveling along the x axis, is given by

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{c}{be^{Ht}} \ .$$

So the coordinate distance that the pulse will travel between time 0 and some arbitrary time t is given by

$$\ell_{\text{coord}}(t) = \int_0^t \frac{\mathrm{d}x}{\mathrm{d}t}(t')\,\mathrm{d}t' = \frac{c}{b}\int_0^t e^{-Ht'}\,\mathrm{d}t' = \frac{c}{bH}\left(1 - e^{-Ht}\right) \ .$$

Therefore, even if we let $t \to \infty$, the coordinate distance traveled by the light pulse will always be less than $c/(bH)$. Since I am stationary in the comoving coordinates, if the initial coordinate distance between Bob and me was more than $c/(bH)$, the light pulse will never reach me. Since the initial time was $t = 0$, with $a(0) = b$,

the light pulse will never reach me if the initial physical distance between Bob and me was more than cH^{-1}, which is called the Hubble length.

[Comment: many students tried to use the formulas that we have learned for the horizon distance, but that is not the same thing. The horizon distance is the present proper distance to the most distant objects from which light has had time to reach us, since $t = 0$. This is often called the "particle horizon," and clearly it is determined solely by the history of the universe, up to the present. The current question concerns whether a photon emitted at the present time by Bob will ever reach me. This question is determined solely by the future evolution of the universe, and is a completely different question from the particle horizon issue. It is also called a horizon, however. The distance beyond which light will never reach me is called the "event horizon."]

— *End of Problem 2.* —

PROBLEM 3: EVOLUTION OF AN OPEN, MATTER-DOMINATED UNIVERSE *(35 points)*

The following problem was Problem 2 on Problem Set 4.

The equations describing the evolution of an open, matter-dominated universe were given in Lecture Notes 4 as

$$ct = \alpha\,(\sinh\theta - \theta)$$

and

$$\frac{a}{\sqrt{\kappa}} = \alpha\,(\cosh\theta - 1) \;,$$

where α is a constant with units of length. The following mathematical identities, which you should know, may also prove useful on parts (e) and (f):

$$\sinh\theta = \frac{e^\theta - e^{-\theta}}{2} \quad , \quad \cosh\theta = \frac{e^\theta + e^{-\theta}}{2}$$

$$e^\theta = 1 + \frac{\theta}{1!} + \frac{\theta^2}{2!} + \frac{\theta^3}{3!} + \dots \;.$$

(a) *(5 points)* Find the Hubble expansion rate H as a function of α and θ.

Answer:

Using the chain rule, the standard formula for the Hubble expansion rate can be rewritten as

$$H(\theta) = \frac{1}{a}\frac{\mathrm{d}a}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}t} \;.$$

The parametric equations for a and t for an open, matter-dominated universe are given by

$$ct = \alpha\,(\sinh\theta - \theta)$$

$$\frac{a}{\sqrt{\kappa}} = \alpha\,(\cosh\theta - 1) \;.$$

Recall that the hyperbolic trigonometric functions are differentiated as

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\sinh\theta = \cosh\theta \;,$$

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\cosh\theta = \sinh\theta \;.$$

So, differentiating the parametric equations,

$$\frac{\mathrm{d}a}{\mathrm{d}\theta} = \alpha\sqrt{k}\sinh\theta \;,$$

$$\frac{\mathrm{d}t}{\mathrm{d}\theta} = \frac{\alpha}{c}(\cosh\theta - 1) = \frac{1}{\mathrm{d}\theta/\mathrm{d}t} \;.$$

Then

$$H(\theta) = \left[\frac{1}{\sqrt{\kappa}\alpha(\cosh\theta - 1)}\right]\left[\alpha\sqrt{\kappa}\sinh\theta\right]\left[\frac{c}{\alpha(\cosh\theta - 1)}\right]$$

$$= \boxed{\frac{c\sinh\theta}{\alpha(\cosh\theta - 1)^2}\ .}$$

(b) *(5 points)* Find the mass density ρ as a function of α and θ.

Answer:

This problem can be attacked by at least three different methods. While you were expected to use only one, we will show all three.

(i) One way to find ρ is to use

$$H^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2}\ .$$

This is usually the safest method to find ρ for a cosmological model, since the above equation is one of the general Friedmann equations. The equation requires that the universe be homogeneous and isotropic, but it is valid for any form of matter. By contrast, the two other methods that will be shown below are valid only for "matter-dominated" universes (i.e., universes that are dominated by nonrelativistic matter, for which the pressure is always negligible). One can rewrite this equation as

$$\frac{8\pi}{3}G\rho = H^2 + \frac{kc^2}{a^2}\ .$$

Recalling that we described open universes by using $\kappa \equiv -k$, this can be rewritten as

$$\frac{8\pi}{3}G\rho = H^2 - \frac{\kappa c^2}{a^2}\ .$$

Replacing H by the answer in part (a) and a by its parametric equation, one finds

$$\begin{aligned}\frac{8\pi}{3}G\rho &= \frac{c^2\sinh^2\theta}{\alpha^2(\cosh\theta - 1)^4} - \frac{\kappa c^2}{\alpha^2\kappa(\cosh\theta - 1)^2} \\ &= \frac{c^2}{\alpha^2(\cosh\theta - 1)^4}\left[\sinh^2\theta - (\cosh\theta - 1)^2\right]\ .\end{aligned}$$

Now make use of the hypertrigonometric identity

$$\cosh^2\theta - \sinh^2\theta = 1$$

to simplify:

$$\begin{aligned}\sinh^2\theta - (\cosh\theta - 1)^2 &= \sinh^2\theta - \cosh^2\theta + 2\cosh\theta - 1 \\ &= 2(\cosh\theta - 1)\ ,\end{aligned}$$

so

$$\frac{8\pi}{3} G\rho = \frac{2c^2}{\alpha^2(\cosh\theta - 1)^3}\ .$$

Dividing both sides of the equation by $(8\pi/3)G$, one finds

$$\boxed{\rho = \frac{3c^2}{4\pi G\alpha^2(\cosh\theta - 1)^3}\ .}$$

(ii) Use the definition of α,

$$\alpha \equiv \frac{4\pi}{3}\frac{G\rho\tilde{a}^3}{c^2}\ ,$$

from Eq. (4.17) of Lecture Notes 4, with Eq. (4.39),

$$\tilde{a}(t) \equiv \frac{a(t)}{\sqrt{\kappa}}\ .$$

One can then solve for ρ, finding

$$\rho = \frac{3}{4\pi}\frac{\alpha\kappa^{3/2}c^2}{Ga^3}\ .$$

By substituting for $a(\theta)$ by using the parametric equation, one finds the final result:

$$\begin{aligned}\rho &= \frac{3}{4\pi}\frac{\alpha\kappa^{3/2}c^2}{G}\frac{1}{\alpha^3\kappa^{3/2}(\cosh\theta - 1)^3} \\ &= \frac{3c^2}{4\pi G\alpha^2(\cosh\theta - 1)^3}\ .\end{aligned}$$

(iii) ρ can also be found from $\ddot{a} = -(4\pi/3)G\rho a$, as long as we know that the universe is matter-dominated. (Be careful, however, about applying this formula in other situations: if the pressure cannot be neglected, then this equation has to be modified.) To evaluate $\ddot{a}$, again use the chain rule. Starting with $\dot{a}$,

$$\dot{a} = \frac{\mathrm{d}a}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}t} = \alpha\sqrt{\kappa}\sinh\theta\frac{c}{\alpha(\cosh\theta - 1)} = \frac{c\sqrt{\kappa}\sinh\theta}{\cosh\theta - 1}\ .$$

Then

$$\ddot{a} = \frac{\mathrm{d}\dot{a}}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}\theta}\left[\frac{c\sqrt{\kappa}\sinh\theta}{\cosh\theta - 1}\right]\frac{c}{\alpha(\cosh\theta-1)}$$
$$= \frac{c^2\sqrt{\kappa}}{\alpha(\cosh\theta-1)}\left[\frac{\cosh\theta}{\cosh\theta-1} - \frac{\sinh^2\theta}{(\cosh\theta-1)^2}\right]$$
$$= \frac{c^2\sqrt{\kappa}}{\alpha(\cosh\theta-1)^3}\left[\cosh\theta(\cosh\theta-1) - \sinh^2\theta\right]$$
$$= \frac{c^2\sqrt{\kappa}}{\alpha(\cosh\theta-1)^3}(1-\cosh\theta) = -\frac{c^2\sqrt{\kappa}}{\alpha(\cosh\theta-1)^2} \ .$$

So

$$\ddot{a} = -\frac{4\pi}{3}G\rho a \quad \Longrightarrow \quad -\frac{c^2\sqrt{\kappa}}{\alpha(\cosh\theta-1)^2} = -\frac{4\pi}{3}G\rho\alpha\sqrt{\kappa}(\cosh\theta-1) \ ,$$

and

$$\rho = \frac{3c^2}{4\pi G\alpha^2(\cosh\theta-1)^3} \ .$$

(c) *(5 points)* Find the mass density parameter Ω as a function of α and θ.

Answer:

The critical mass density satisfies the cosmological evolution equations for $k=0$, so

$$H^2 = \frac{8\pi}{3}G\rho_c \ .$$

Then

$$\Omega \equiv \frac{\rho}{\rho_c} = \frac{8\pi G\rho}{3H^2} \ .$$

Now replace H by the answer to part (a), and ρ by the answer to part (b):

$$\Omega = \frac{8\pi G}{3}\left[\frac{3}{4\pi}\frac{c^2}{G\alpha^2(\cosh\theta-1)^3}\right]\left[\frac{\alpha^2(\cosh\theta-1)^4}{c^2\sinh^2\theta}\right]$$
$$= 2\frac{\cosh\theta-1}{\sinh^2\theta} = 2\frac{\cosh\theta-1}{\cosh^2\theta-1}$$
$$= 2\frac{\cosh\theta-1}{(\cosh\theta+1)(\cosh\theta-1)} = \boxed{\frac{2}{\cosh\theta+1} \ .}$$

The answer can be written even more compactly, if one wishes, by using a further hypertrigonometric identity:

$$\Omega = \frac{2}{\cosh\theta + 1} = \frac{1}{\cosh^2\frac{1}{2}\theta} = \operatorname{sech}^2\frac{1}{2}\theta \ .$$

(d) *(6 points)* Find the physical value of the horizon distance, $\ell_{p,\text{horizon}}$, as a function of α and θ.

Answer:

The basic formula that determines the physical value of the horizon distance is given by Eq. (4.7) of the lecture notes:

$$\ell_{p,\text{horizon}}(t) = a(t)\int_0^t \frac{c}{a(t')}dt' \ .$$

The complication here is that a is given as a function of θ, rather than t. The problem is handled, however, by a simple change of integration variables. One can change the integral over t' to an integral over θ', provided that one replaces

$$\mathrm{d}t' \to \frac{dt'}{\mathrm{d}\theta'}\mathrm{d}\theta' = \frac{\alpha}{c}(\cosh\theta' - 1)\mathrm{d}\theta' \ .$$

One must also re-express the limits of integration in terms of θ. So

$$\begin{aligned}
\ell_{p,\text{horizon}}(t) &= a\big(\theta(t)\big)\int_0^{\theta(t)} \frac{c}{a(\theta')}\frac{dt'}{\mathrm{d}\theta'}\mathrm{d}\theta' \\
&= \alpha\sqrt{\kappa}\big(\cosh\theta(t) - 1\big)\int_0^{\theta(t)} \frac{c}{\alpha\sqrt{\kappa}(\cosh\theta' - 1)}\frac{\alpha}{c}(\cosh\theta' - 1)\,\mathrm{d}\theta' \ . \\
&= \alpha\big(\cosh\theta(t) - 1\big)\int_0^{\theta(t)} \mathrm{d}\theta' = \boxed{\alpha\,\theta(t)\,\big(\cosh\theta(t) - 1\big) \ .}
\end{aligned}$$

(e) *(7 points)* For very small values of t, it is possible to use the first nonzero term of a power-series expansion to express θ as a function of t, and then a as a function of t. Give the expression for $a(t)$ in this approximation. The approximation will be valid for $t \ll t^*$. Estimate the value of t^*.

Answer:

The key to this problem is the use of power series expansions. When this problem appeared as a quiz problem in 1992, I was rather surprised to find that many of the students seemed very inexperienced in this technique. It is a very useful method of approximation, so I strongly urge you to learn it if you don't know it already. In general, any sufficiently smooth function $f(x)$ can be expanded about the point x_0 by the series

$$f(x) = f(x_0) + \frac{1}{1!}f'(x_0)(x - x_0) + \frac{1}{2!}f''(x_0)(x - x_0)^2 + \frac{1}{3!}f'''(x_0)(x - x_0)^3 + \dots ,$$

where the prime is used to denote a derivative. In particular, the exponential, sinh, and cosh functions can be expanded about $\theta = 0$ by the formulas

$$e^\theta = 1 + \frac{\theta}{1!} + \frac{\theta^2}{2!} + \frac{\theta^3}{3!} + \dots$$

$$\sinh\theta = \theta + \frac{\theta^3}{3!} + \frac{\theta^5}{5!} + \frac{\theta^5}{7!} \dots$$

$$\cosh\theta = 1 + \frac{\theta^2}{2!} + \frac{\theta^4}{4!} + \frac{\theta^6}{6!} + \dots .$$

For this problem, we expand the parametric equations for $a(\theta)$ and $t(\theta)$, keeping the first nonvanishing term in the power series expansions:

$$t = \frac{\alpha}{c}(\sinh\theta - \theta) = \frac{\alpha}{c}\left(\frac{\theta^3}{3!} + \dots\right)$$

$$a = \alpha\sqrt{\kappa}(\cosh\theta - 1) = \alpha\sqrt{\kappa}\left(\frac{\theta^2}{2!} + \dots\right) .$$

The first expression can be solved for θ, giving

$$\theta \approx \left(\frac{6ct}{\alpha}\right)^{1/3} ,$$

which can be substituted into the second expression to give

$$\boxed{a \approx \frac{1}{2}\alpha\sqrt{\kappa}\left(\frac{6ct}{\alpha}\right)^{2/3} .}$$

The power series expansions for the sinh and cosh are valid whenever the terms left out are much smaller than the last term kept, which happens when $\theta \ll 1$. Given the above relation between θ and t, this condition is equivalent to

$$t \ll \frac{\alpha}{6c} .$$

Thus,

$$\boxed{t^* \approx \frac{\alpha}{6c}, \text{ or } t^* \approx \frac{\alpha}{c}\ .}$$

Since there is no precise meaning to the statement that an approximation is valid, there is no precise value for t^*.

(f) *(7 points)* Even though these equations describe an open universe, one still finds that Ω approaches one for very early times. For $t \ll t^*$ (where t^* is defined in part (e)), the quantity $1 - \Omega$ behaves as a power of t. Find the expression for $1 - \Omega$ in this approximation.

Answer:

From part (c), the expression for Ω is given by

$$\Omega = \frac{2}{\cosh\theta + 1}\ .$$

So,

$$1 - \Omega = 1 - \frac{2}{\cosh\theta + 1} = \frac{\cosh\theta - 1}{\cosh\theta + 1}\ .$$

Expanding numerator and denominator in power series,

$$1 - \Omega \approx \frac{\frac{\theta^2}{2!} + \frac{\theta^4}{4!} + \dots}{2 + \frac{\theta^2}{2!} + \dots}\ .$$

Keeping only the leading terms,

$$1 - \Omega \approx \frac{\frac{\theta^2}{2}}{2} = \frac{1}{4}\theta^2\ ,$$

so

$$\boxed{1 - \Omega \approx \frac{1}{4}\left(\frac{6ct}{\alpha}\right)^{2/3}\ .}$$

This result shows that the deviation of Ω from 1 is amplified with time. This fact leads to a conundrum called the "flatness problem", which will be discussed later in the course.

A common mistake (very minor) was to keep extra terms, especially in the denominator. Keeping extra terms allows a higher degree of accuracy, so there is nothing wrong with it. However, one should always be sure to keep ***all*** terms of a given order, since keeping only a subset of terms may or may not increase the accuracy. In this case, an extra term in the denominator can be rewritten as a term in the numerator:

$$\frac{\frac{\theta^2}{2!}}{2+\frac{\theta^2}{2!}} = \frac{1}{4}\frac{\theta^2}{1+\frac{\theta^2}{4}} = \frac{1}{4}\theta^2\left(1-\frac{\theta^2}{4}+\ldots\right)$$

$$= \frac{1}{4}\theta^2 - \frac{1}{16}\theta^4 + \ldots\ ,$$

where I used the expansion

$$\frac{1}{1+x} = 1 - x + x^2 - x^3 + x^4 + \ldots\ .$$

Thus, the extra term in the denominator is equivalent to a term in the numerator of order θ^4, but other terms proportional to θ^4 have been dropped. So, it is not worthwhile to keep the 2nd term in the expansion of the denominator.

— *End of Problem 3.* —

PROBLEM 4: RADIAL GEODESICS IN A CLOSED UNIVERSE *(20 points)*

As shown in the formula sheets, we can describe a closed universe by choosing $k = 1$, and then using coordinates (t, r, θ, ϕ), with metric

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 = -c^2\,\mathrm{d}t^2 + a^2(t)\left\{\frac{\mathrm{d}r^2}{1-r^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\} , \tag{4.1}$$

or by using coordinates (t, ψ, θ, ϕ), with metric

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 \equiv -c^2\,\mathrm{d}t^2 + a^2(t)\left\{\mathrm{d}\psi^2 + \sin^2\psi\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\} . \tag{4.2}$$

The connection between the two coordinate systems is given by

$$r = \sin\psi . \tag{4.3}$$

The general spacetime geodesic equation can be written as

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{g_{\mu\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\right\} = \frac{1}{2}\left(\partial_\mu g_{\lambda\sigma}\right)\frac{\mathrm{d}x^\lambda}{\mathrm{d}\tau}\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau} . \tag{4.4}$$

(a) *(7 points)* Using the coordinates (t, ψ, θ, ϕ) and the metric of Eq. (4.2), compute explicitly the geodesic equation for $\mu = \psi$. By "compute explicitly", I mean that $g_{\mu\nu}$ should be replaced by the relevant expressions from Eq. (4.2), and that the sums over indices should be written out, including only the nonzero terms.

Answer:

Since the metric is diagonal, only $\nu = \psi$ contributes to the sum over ν. Similarly λ must equal σ, and the only nonzero values of $\partial_\psi g_{\lambda\sigma}$ are when $\lambda = \sigma = \theta$ and $\lambda = \sigma = \phi$. So Eq. (4.4) becomes

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{g_{\psi\psi}\frac{\mathrm{d}\psi}{\mathrm{d}\tau}\right\} = \frac{1}{2}\left[\frac{\partial g_{\theta\theta}}{\partial\psi}\left(\frac{\mathrm{d}\theta}{\mathrm{d}\tau}\right)^2 + \frac{\partial g_{\phi\phi}}{\partial\psi}\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2\right] .$$

Using $g_{\psi\psi} = a^2(t)$, $g_{\theta\theta} = a^2(t)\sin^2\psi$, and $g_{\phi\phi} = a^2(t)\sin^2\psi\sin^2\theta$, the equation becomes

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{a^2(t)\frac{\mathrm{d}\psi}{\mathrm{d}\tau}\right\} = a^2(t)\,\sin\psi\,\cos\psi\left[\left(\frac{\mathrm{d}\theta}{\mathrm{d}\tau}\right)^2 + \sin^2\theta\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2\right] .} \tag{4.5}$$

You were not asked to expand the left-hand-side, but some of you did. If you do expand the left-hand side, it is important to remember that $a(t)$ depends on t and t depends on τ, so the equation becomes

$$a^2(t)\frac{\mathrm{d}^2\psi}{\mathrm{d}\tau^2} + 2a\dot{a}\frac{\mathrm{d}t}{\mathrm{d}\tau}\frac{\mathrm{d}\psi}{\mathrm{d}\tau} = a^2(t)\,\sin\psi\,\cos\psi\left[\left(\frac{\mathrm{d}\theta}{\mathrm{d}\tau}\right)^2 + \sin^2\theta\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2\right] ,$$

or

$$\frac{\mathrm{d}^2\psi}{\mathrm{d}\tau^2} + 2\left(\frac{\dot{a}}{a}\right)\frac{\mathrm{d}t}{\mathrm{d}\tau}\frac{\mathrm{d}\psi}{\mathrm{d}\tau} = \sin\psi\,\cos\psi\left[\left(\frac{\mathrm{d}\theta}{\mathrm{d}\tau}\right)^2 + \sin^2\theta\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2\right] .$$

(b) *(7 points)* Using instead the coordinates (t, r, θ, ϕ), compute explicitly the geodesic equation for $\mu = r$.

Answer:

Again the equation simplifies significantly, since $g_{\mu\nu}$ is diagonal. On the right-hand side, only 3 of the 4 possible values of $\lambda = \sigma$ contribute, as $\partial_r g_{tt} = 0$. So,

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{g_{rr}\frac{\mathrm{d}r}{\mathrm{d}\tau}\right\} = \frac{1}{2}\left[\frac{\partial g_{rr}}{\partial r}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 + \frac{\partial g_{\theta\theta}}{\partial r}\left(\frac{\mathrm{d}\theta}{\mathrm{d}\tau}\right)^2 + \frac{\partial g_{\phi\phi}}{\partial r}\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2\right] .$$

Now we use

$$g_{rr} = \frac{a^2(t)}{1-r^2} \;,\quad g_{\theta\theta} = a^2(t)r^2 \;,\quad g_{\phi\phi} = a^2(t)r^2\sin^2\theta \;,$$

which allows us to rewrite the equation as

$$\boxed{\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{\frac{a^2(t)}{1-r^2}\frac{\mathrm{d}r}{\mathrm{d}\tau}\right\} = \frac{1}{2}\Bigg[&\frac{2ra^2(t)}{(1-r^2)^2}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 + 2ra^2(t)\left(\frac{\mathrm{d}\theta}{\mathrm{d}\tau}\right)^2 \\ &+ 2ra^2(t)\sin^2\theta\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2\Bigg] ,\end{aligned}}$$

or

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{\frac{a^2(t)}{1-r^2}\frac{\mathrm{d}r}{\mathrm{d}\tau}\right\} = ra^2(t)\left[\frac{1}{(1-r^2)^2}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 + \left(\frac{\mathrm{d}\theta}{\mathrm{d}\tau}\right)^2 + \sin^2\theta\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2\right] .} \tag{4.6}$$

Again you were not asked to expand the left-hand side, but if the left-hand side is expanded, one must remember that $a(t)$ and r both depend on τ. So

$$\frac{a^2(t)}{1-r^2}\frac{\mathrm{d}^2 r}{\mathrm{d}\tau^2} + \frac{2ra^2(t)}{(1-r^2)^2}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 + \frac{2a\dot{a}}{1-r^2}\frac{\mathrm{d}t}{\mathrm{d}\tau}\frac{\mathrm{d}r}{\mathrm{d}\tau}$$
$$= ra^2(t)\left[\frac{1}{(1-r^2)^2}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 + \left(\frac{\mathrm{d}\theta}{\mathrm{d}\tau}\right)^2 + \sin^2\theta\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2\right] .$$

Rearranging terms, the equation can be simplified to

$$\frac{\mathrm{d}^2 r}{\mathrm{d}\tau^2} + 2\left(\frac{\dot{a}}{a}\right)\frac{\mathrm{d}t}{\mathrm{d}\tau}\frac{\mathrm{d}r}{\mathrm{d}\tau}$$
$$= r\left\{-\frac{1}{(1-r^2)}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 + (1-r^2)\left[\left(\frac{\mathrm{d}\theta}{\mathrm{d}\tau}\right)^2 + \sin^2\theta\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2\right]\right\} .$$

(c) *(6 points)* Are the results from parts (a) and (b) both valid, or is one valid and the other not? If you believe that they are both valid, use Eq. (4.3) to show that they are equivalent. If you believe that only one is valid, state which one is valid, and explain why the other is not. (4 points will be given for showing the correct understanding of this problem, with 2 points allocated to completing the algebra needed to demonstrate your answer.)

Answer:

Both answers are valid, since they are both correct forms of the geodesic equation, in different coordinate systems. To see that they are equivalent, we can start with the equation for r, Eq. (4.6), and substitute

$$r = \sin\psi \quad \Longrightarrow \quad \frac{\mathrm{d}r}{\mathrm{d}\tau} = \cos\psi\frac{\mathrm{d}\psi}{\mathrm{d}\tau} \quad \Longrightarrow \quad \frac{1}{1-r^2}\frac{\mathrm{d}r}{\mathrm{d}\tau} = \frac{1}{\cos\psi}\frac{\mathrm{d}\psi}{\mathrm{d}\tau} .$$

So Eq. (4.6) becomes

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{\frac{a^2(t)}{\cos\psi}\frac{\mathrm{d}\psi}{\mathrm{d}t}\right\} = a^2\sin\psi\left[\frac{1}{\cos^2\psi}\left(\frac{\mathrm{d}\psi}{\mathrm{d}\tau}\right)^2 + \left(\frac{\mathrm{d}\theta}{\mathrm{d}\tau}\right)^2 + \sin^2\theta\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2\right] .$$

Partially expanding the left-hand side,

$$\frac{1}{\cos\psi}\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{a^2(t)\frac{\mathrm{d}\psi}{\mathrm{d}t}\right\} + \frac{a^2}{\cos^2\psi}\sin\psi\left(\frac{\mathrm{d}\psi}{\mathrm{d}\tau}\right)^2$$
$$= a^2\sin\psi\left[\frac{1}{\cos^2\psi}\left(\frac{\mathrm{d}\psi}{\mathrm{d}\tau}\right)^2 + \left(\frac{\mathrm{d}\theta}{\mathrm{d}\tau}\right)^2 + \sin^2\theta\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2\right] .$$

The terms proportional to $(\mathrm{d}\psi/\mathrm{d}\tau)^2$ can be seen to cancel, and then multiplication of the equation by $\cos\psi$ gives

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{a^2(t)\frac{\mathrm{d}\psi}{\mathrm{d}\tau}\right\} = a^2(t)\,\sin\psi\,\cos\psi\left[\left(\frac{\mathrm{d}\theta}{\mathrm{d}\tau}\right)^2 + \sin^2\theta\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2\right]\ ,$$

which is identical to Eq. (4.5).

— *End of Problem 4.* —

Problem	Maximum	Score	Initials
1	25		
2	20		
3	35		
4	20		
TOTAL	100		

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — Quiz Date: November 5, 2018
Prof. Alan Guth

QUIZ 2 FORMULA SHEET

DOPPLER SHIFT (For motion along a line):

$$z = v/u \quad \text{(nonrelativistic, source moving)}$$

$$z = \frac{v/u}{1 - v/u} \quad \text{(nonrelativistic, observer moving)}$$

$$z = \sqrt{\frac{1+\beta}{1-\beta}} - 1 \quad \text{(special relativity, with } \beta = v/c\text{)}$$

COSMOLOGICAL REDSHIFT:

$$1 + z \equiv \frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}} = \frac{a(t_{\text{observed}})}{a(t_{\text{emitted}})}$$

SPECIAL RELATIVITY:

Time Dilation Factor:

$$\gamma \equiv \frac{1}{\sqrt{1-\beta^2}} \ , \qquad \beta \equiv v/c$$

Lorentz-Fitzgerald Contraction Factor: γ

Relativity of Simultaneity:
Trailing clock reads later by an amount $\beta \ell_0 / c$.

Energy-Momentum Four-Vector:

$$p^\mu = \left(\frac{E}{c}, \vec{p}\right) \ , \quad \vec{p} = \gamma m_0 \vec{v} \ , \quad E = \gamma m_0 c^2 = \sqrt{\left(m_0 c^2\right)^2 + |\vec{p}|^2 c^2} \ ,$$

$$p^2 \equiv |\vec{p}|^2 - \left(p^0\right)^2 = |\vec{p}|^2 - \frac{E^2}{c^2} = -\left(m_0 c\right)^2 \ .$$

KINEMATICS OF A HOMOGENEOUSLY EXPANDING UNIVERSE:

Hubble's Law: $v = Hr$,
where v = recession velocity of a distant object, H = Hubble expansion rate, and r = distance to the distant object.

Present Value of Hubble Expansion Rate (Planck 2018):

$$H_0 = 67.66 \pm 0.42 \text{ km-s}^{-1}\text{-Mpc}^{-1}$$

Scale Factor: $\ell_p(t) = a(t)\ell_c$,
where $\ell_p(t)$ is the physical distance between any two objects, $a(t)$ is the scale factor, and ℓ_c is the coordinate distance between the objects, also called the comoving distance.

Hubble Expansion Rate: $H(t) = \dfrac{1}{a(t)}\dfrac{\mathrm{d}a(t)}{\mathrm{d}t}$.

Light Rays in Comoving Coordinates: Light rays travel in straight lines with physical speed c relative to any observer. In Cartesian coordinates, coordinate speed $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \dfrac{c}{a(t)}$. In general, $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = 0$.

Horizon Distance:

$$\begin{aligned}\ell_{p,\text{horizon}}(t) &= a(t)\int_0^t \frac{c}{a(t')}dt' \\ &= \begin{cases} 3ct & \text{(flat, matter-dominated),} \\ 2ct & \text{(flat, radiation-dominated).}\end{cases}\end{aligned}$$

COSMOLOGICAL EVOLUTION:

$$H^2 = \left(\frac{\dot a}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2} \ , \quad \ddot a = -\frac{4\pi}{3}G\left(\rho + \frac{3p}{c^2}\right)a \ ,$$

$$\rho_m(t) = \frac{a^3(t_i)}{a^3(t)}\,\rho_m(t_i) \ \ \text{(matter)}, \quad \rho_r(t) = \frac{a^4(t_i)}{a^4(t)}\,\rho_r(t_i) \ \ \text{(radiation)}.$$

$$\dot\rho = -3\frac{\dot a}{a}\left(\rho + \frac{p}{c^2}\right) \ , \quad \Omega \equiv \rho/\rho_c \ , \quad \text{where} \quad \rho_c = \frac{3H^2}{8\pi G} \ .$$

EVOLUTION OF A MATTER-DOMINATED UNIVERSE:

Flat ($k = 0$): $a(t) \propto t^{2/3}$

$\Omega = 1$.

Closed ($k > 0$): $ct = \alpha(\theta - \sin\theta)$, $\dfrac{a}{\sqrt{k}} = \alpha(1 - \cos\theta)$,

$$\Omega = \frac{2}{1 + \cos\theta} > 1 \ ,$$

$$\text{where } \alpha \equiv \frac{4\pi}{3}\frac{G\rho}{c^2}\left(\frac{a}{\sqrt{k}}\right)^3 \ .$$

Open ($k < 0$): $ct = \alpha\left(\sinh\theta - \theta\right)$, $\dfrac{a}{\sqrt{\kappa}} = \alpha\left(\cosh\theta - 1\right)$,

$$\Omega = \frac{2}{1 + \cosh\theta} < 1 \ ,$$

$$\text{where } \alpha \equiv \frac{4\pi}{3}\frac{G\rho}{c^2}\left(\frac{a}{\sqrt{\kappa}}\right)^3 \ ,$$

$$\kappa \equiv -k > 0 \ .$$

MINKOWSKI METRIC (Special Relativity):

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \ .$$

ROBERTSON-WALKER METRIC:

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 = -c^2\,\mathrm{d}t^2 + a^2(t)\left\{\frac{\mathrm{d}r^2}{1 - kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\} \ .$$

Alternatively, for $k > 0$, we can define $r = \dfrac{\sin\psi}{\sqrt{k}}$, and then

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 \equiv -c^2\,\mathrm{d}t^2 + \tilde{a}^2(t)\left\{\mathrm{d}\psi^2 + \sin^2\psi\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\} \ ,$$

where $\tilde{a}(t) = a(t)/\sqrt{k}$. For $k < 0$ we can define $r = \dfrac{\sinh\psi}{\sqrt{-k}}$, and then

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 = -c^2\,\mathrm{d}t^2 + \tilde{a}^2(t)\left\{\mathrm{d}\psi^2 + \sinh^2\psi\left(\mathrm{d}\theta^2 + \sin^2\theta\,d\phi^2\right)\right\} \ ,$$

where $\tilde{a}(t) = a(t)/\sqrt{-k}$. Note that $\tilde{a}$ can be called a if there is no need to relate it to the $a(t)$ that appears in the first equation above.

SCHWARZSCHILD METRIC:

$$\mathrm{d}s^2 \equiv -c^2\mathrm{d}\tau^2 = -\left(1 - \frac{2GM}{rc^2}\right) c^2\mathrm{d}t^2 + \left(1 - \frac{2GM}{rc^2}\right)^{-1} \mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2 \ ,$$

GEODESIC EQUATION:

$$\frac{\mathrm{d}}{\mathrm{d}s}\left\{g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}s}\right\} = \frac{1}{2}\left(\partial_i g_{k\ell}\right)\frac{\mathrm{d}x^k}{\mathrm{d}s}\,\frac{\mathrm{d}x^\ell}{\mathrm{d}s}$$

or: $$\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{g_{\mu\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\right\} = \frac{1}{2}\left(\partial_\mu g_{\lambda\sigma}\right)\frac{\mathrm{d}x^\lambda}{\mathrm{d}\tau}\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau}$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — November 30, 2018
Prof. Alan Guth

REVIEW PROBLEMS FOR QUIZ 3

QUIZ DATE: Wedneday, December 5, 2018, during the normal class time.

COVERAGE: Lecture Notes 6 (pp. 12–end), Lecture Notes 7 and 8. Problem Sets 7 and 8; Steven Weinberg, *The First Three Minutes,* Chapter 8 and the Afterword; Barbara Ryden, *Introduction to Cosmology*, Chapters 9 *(The Cosmic Microwave Background)* and 11 *(Inflation and the Very Early Universe)*; Alan Guth, *Inflation and the New Era of High-Precision Cosmology,*

http://web.mit.edu/physics/news/physicsatmit/physicsatmit_02_cosmology.pdf .

One of the problems on the quiz will be taken verbatim (or at least *almost* verbatim) from either the homework assignments, or from the starred problems from this set of Review Problems. The starred problems are the ones that I recommend that you review most carefully: Problems 2, 6, 8, 11, 12, 15, 17, and 18.

PURPOSE: These review problems are not to be handed in, but are being made available to help you study. They come mainly from quizzes in previous years. In some cases the number of points assigned to the problem on the quiz is listed — in all such cases it is based on 100 points for the full quiz.

In addition to this set of problems, you will find on the course web page the actual quizzes that were given in 1994, 1996, 1998, 2000, 2002, 2004, 2007, 2009, 2011, 2013, and 2016. The relevant problems from those quizzes have mostly been incorporated into these review problems, but you still may be interested in looking at the quizzes, just to see how much material has been included in each quiz. The coverage of the upcoming quiz will not necessarily match the coverage of any of the quizzes from previous years. The coverage for each quiz in recent years is usually described at the start of the review problems, as I did here.

REVIEW SESSION AND OFFICE HOURS: To help you study for the quiz, Honggeun Kim will hold a review session on Sunday, December 2, at 8:00 pm, in Room 4-153. In addition, both Honggeun Kim and I will be moving our office hours for the two last weeks of the term. I (Alan) will hold my office hours on Mondays (December 3 and 10) at 7:30 pm in my office, 6-322, and Honggeun will hold his office hours on Tuesdays (December 4 and 11) in Room 8-308.

INFORMATION TO BE GIVEN ON QUIZ:

For the third quiz, the following information will be made available to you:

DOPPLER SHIFT (For motion along a line):

$$z = v/u \quad \text{(nonrelativistic, source moving)}$$

$$z = \frac{v/u}{1 - v/u} \quad \text{(nonrelativistic, observer moving)}$$

$$z = \sqrt{\frac{1+\beta}{1-\beta}} - 1 \quad \text{(special relativity, with } \beta = v/c\text{)}$$

COSMOLOGICAL REDSHIFT:

$$1 + z \equiv \frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}} = \frac{a(t_{\text{observed}})}{a(t_{\text{emitted}})}$$

SPECIAL RELATIVITY:

Time Dilation Factor:

$$\gamma \equiv \frac{1}{\sqrt{1-\beta^2}} \ , \qquad \beta \equiv v/c$$

Lorentz-Fitzgerald Contraction Factor: γ

Relativity of Simultaneity:
Trailing clock reads later by an amount $\beta \ell_0 / c$.

Energy-Momentum Four-Vector:

$$p^\mu = \left(\frac{E}{c} \, , \vec{p} \right) \ , \quad \vec{p} = \gamma m_0 \vec{v} \ , \quad E = \gamma m_0 c^2 = \sqrt{\left(m_0 c^2\right)^2 + \left|\vec{p}\right|^2 c^2} \ ,$$

$$p^2 \equiv \left|\vec{p}\right|^2 - \left(p^0\right)^2 = \left|\vec{p}\right|^2 - \frac{E^2}{c^2} = -\left(m_0 c\right)^2 \ .$$

KINEMATICS OF A HOMOGENEOUSLY EXPANDING UNIVERSE:

Hubble's Law: $v = Hr$,
where v = recession velocity of a distant object, H = Hubble expansion rate, and r = distance to the distant object.

Present Value of Hubble Expansion Rate (Planck 2018):

$$H_0 = 67.66 \pm 0.42 \text{ km-s}^{-1}\text{-Mpc}^{-1}$$

Scale Factor: $\ell_p(t) = a(t)\ell_c$,
where $\ell_p(t)$ is the physical distance between any two objects, $a(t)$ is the scale factor, and ℓ_c is the coordinate distance between the objects, also called the comoving distance.

Hubble Expansion Rate: $H(t) = \dfrac{1}{a(t)} \dfrac{\mathrm{d}a(t)}{\mathrm{d}t}$.

Light Rays in Comoving Coordinates: Light rays travel in straight lines with physical speed c relative to any observer. In Cartesian coordinates, coordinate speed $\dfrac{\mathrm{d}x}{\mathrm{d}t} = \dfrac{c}{a(t)}$. In general, $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = 0$.

Horizon Distance:

$$\begin{aligned} \ell_{p,\text{horizon}}(t) &= a(t) \int_0^t \frac{c}{a(t')} dt' \\ &= \begin{cases} 3ct & \text{(flat, matter-dominated)}, \\ 2ct & \text{(flat, radiation-dominated)}. \end{cases} \end{aligned}$$

COSMOLOGICAL EVOLUTION:

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}G\rho - \frac{kc^2}{a^2} \ , \quad \ddot{a} = -\frac{4\pi}{3}G\left(\rho + \frac{3p}{c^2}\right)a \ ,$$

$$\rho_m(t) = \frac{a^3(t_i)}{a^3(t)}\,\rho_m(t_i) \ \ \text{(matter)}, \quad \rho_r(t) = \frac{a^4(t_i)}{a^4(t)}\,\rho_r(t_i) \ \ \text{(radiation)}.$$

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) \ , \quad \Omega \equiv \rho/\rho_c \ , \quad \text{where} \quad \rho_c = \frac{3H^2}{8\pi G} \ .$$

EVOLUTION OF A MATTER-DOMINATED UNIVERSE:

Flat ($k = 0$): $a(t) \propto t^{2/3}$

$\Omega = 1$.

Closed ($k > 0$): $ct = \alpha(\theta - \sin\theta)$, $\dfrac{a}{\sqrt{k}} = \alpha(1 - \cos\theta)$,

$$\Omega = \frac{2}{1 + \cos\theta} > 1 \ ,$$

$$\text{where } \alpha \equiv \frac{4\pi}{3}\frac{G\rho}{c^2}\left(\frac{a}{\sqrt{k}}\right)^3 \ .$$

Open ($k < 0$): $ct = \alpha\left(\sinh\theta - \theta\right)$, $\dfrac{a}{\sqrt{\kappa}} = \alpha\left(\cosh\theta - 1\right)$,

$$\Omega = \frac{2}{1 + \cosh\theta} < 1 \ ,$$

$$\text{where } \alpha \equiv \frac{4\pi}{3}\frac{G\rho}{c^2}\left(\frac{a}{\sqrt{\kappa}}\right)^3 \ ,$$

$$\kappa \equiv -k > 0 \ .$$

MINKOWSKI METRIC (Special Relativity):

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \ .$$

ROBERTSON-WALKER METRIC:

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 = -c^2\,\mathrm{d}t^2 + a^2(t)\left\{\frac{\mathrm{d}r^2}{1 - kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\} \ ,$$

where the universe is closed/open/flat if $k > 0/k < 0/k = 0$. k can be taken as $+1$ (-1) if the universe is closed (open).

Alternatively, for $k > 0$, we can define $r = \dfrac{\sin\psi}{\sqrt{k}}$, and then

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 \equiv -c^2\,\mathrm{d}t^2 + \tilde{a}^2(t)\left\{\mathrm{d}\psi^2 + \sin^2\psi\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right\} \ ,$$

where $\tilde{a}(t) = a(t)/\sqrt{k}$. For $k < 0$ we can define $r = \dfrac{\sinh\psi}{\sqrt{-k}}$, and then

$$\mathrm{d}s^2 \equiv -c^2\,\mathrm{d}\tau^2 = -c^2\,\mathrm{d}t^2 + \tilde{a}^2(t)\left\{\mathrm{d}\psi^2 + \sinh^2\psi\left(\mathrm{d}\theta^2 + \sin^2\theta\,d\phi^2\right)\right\} \ ,$$

where $\tilde{a}(t) = a(t)/\sqrt{-k}$. Note that $\tilde{a}$ can be called a if there is no need to relate it to the $a(t)$ that appears in the first equation above.

SCHWARZSCHILD METRIC:

$$\mathrm{d}s^2 \equiv -c^2\mathrm{d}\tau^2 = -\left(1-\frac{2GM}{rc^2}\right)c^2\mathrm{d}t^2 + \left(1-\frac{2GM}{rc^2}\right)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\phi^2 \ ,$$

GEODESIC EQUATION:

$$\frac{\mathrm{d}}{\mathrm{d}s}\left\{g_{ij}\frac{\mathrm{d}x^j}{\mathrm{d}s}\right\} = \frac{1}{2}\left(\partial_i g_{k\ell}\right)\frac{\mathrm{d}x^k}{\mathrm{d}s}\,\frac{\mathrm{d}x^\ell}{\mathrm{d}s}$$

or: $$\frac{\mathrm{d}}{\mathrm{d}\tau}\left\{g_{\mu\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tau}\right\} = \frac{1}{2}\left(\partial_\mu g_{\lambda\sigma}\right)\frac{\mathrm{d}x^\lambda}{\mathrm{d}\tau}\frac{\mathrm{d}x^\sigma}{\mathrm{d}\tau}$$

BLACK-BODY RADIATION:

$$u = g\frac{\pi^2}{30}\,\frac{(kT)^4}{(\hbar c)^3} \qquad \text{(energy density)}$$

$$p = \frac{1}{3}u \qquad \rho = u/c^2 \qquad \text{(pressure, mass density)}$$

$$n = g^*\frac{\zeta(3)}{\pi^2}\,\frac{(kT)^3}{(\hbar c)^3} \qquad \text{(number density)}$$

$$s = g\frac{2\pi^2}{45}\,\frac{k^4T^3}{(\hbar c)^3} \ , \qquad \text{(entropy density)}$$

where

$$g \equiv \begin{cases} 1 \text{ per spin state for bosons (integer spin)} \\ 7/8 \text{ per spin state for fermions (half-integer spin)} \end{cases}$$

$$g^* \equiv \begin{cases} 1 \text{ per spin state for bosons} \\ 3/4 \text{ per spin state for fermions ,} \end{cases}$$

and
$$\zeta(3) = \frac{1}{1^3} + \frac{1}{2^3} + \frac{1}{3^3} + \cdots \approx 1.202 \; .$$

$$g_\gamma = g^*_\gamma = 2 \; ,$$

$$g_\nu = \underbrace{\frac{7}{8}}_{\substack{\text{Fermion}\\ \text{factor}}} \times \underbrace{3}_{\substack{\text{3 species}\\ \nu_e,\nu_\mu,\nu_\tau}} \times \underbrace{2}_{\substack{\text{Particle/}\\ \text{antiparticle}}} \times \underbrace{1}_{\text{Spin states}} = \frac{21}{4} \; ,$$

$$g^*_\nu = \underbrace{\frac{3}{4}}_{\substack{\text{Fermion}\\ \text{factor}}} \times \underbrace{3}_{\substack{\text{3 species}\\ \nu_e,\nu_\mu,\nu_\tau}} \times \underbrace{2}_{\substack{\text{Particle/}\\ \text{antiparticle}}} \times \underbrace{1}_{\text{Spin states}} = \frac{9}{2} \; ,$$

$$g_{e^+e^-} = \underbrace{\frac{7}{8}}_{\substack{\text{Fermion}\\ \text{factor}}} \times \underbrace{1}_{\text{Species}} \times \underbrace{2}_{\substack{\text{Particle/}\\ \text{antiparticle}}} \times \underbrace{2}_{\text{Spin states}} = \frac{7}{2} \; ,$$

$$g^*_{e^+e^-} = \underbrace{\frac{3}{4}}_{\substack{\text{Fermion}\\ \text{factor}}} \times \underbrace{1}_{\text{Species}} \times \underbrace{2}_{\substack{\text{Particle/}\\ \text{antiparticle}}} \times \underbrace{2}_{\text{Spin states}} = 3 \; .$$

EVOLUTION OF A FLAT RADIATION-DOMINATED UNIVERSE:

$$\rho = \frac{3}{32\pi G t^2}$$

$$kT = \left(\frac{45\hbar^3 c^5}{16\pi^3 g G}\right)^{1/4} \frac{1}{\sqrt{t}}$$

For $m_\mu = 106$ MeV $\gg kT \gg m_e = 0.511$ MeV, $g = 10.75$ and then

$$kT = \frac{0.860 \text{ MeV}}{\sqrt{t \text{ (in sec)}}} \left(\frac{10.75}{g}\right)^{1/4}$$

After the freeze-out of electron-positron pairs,

$$\frac{T_\nu}{T_\gamma} = \left(\frac{4}{11}\right)^{1/3} \; .$$

COSMOLOGICAL CONSTANT:

$$u_{\text{vac}} = \rho_{\text{vac}} c^2 = \frac{\Lambda c^4}{8\pi G} \; ,$$

$$p_{\rm vac} = -\rho_{\rm vac} c^2 = -\frac{\Lambda c^4}{8\pi G} \ .$$

GENERALIZED COSMOLOGICAL EVOLUTION:

$$x \frac{\mathrm{d}x}{\mathrm{d}t} = H_0 \sqrt{\Omega_{m,0} x + \Omega_{\rm rad,0} + \Omega_{\rm vac,0} x^4 + \Omega_{k,0} x^2} \ ,$$

where

$$x \equiv \frac{a(t)}{a(t_0)} \equiv \frac{1}{1+z} \ ,$$

$$\Omega_{k,0} \equiv -\frac{kc^2}{a^2(t_0) H_0^2} = 1 - \Omega_{m,0} - \Omega_{\rm rad,0} - \Omega_{\rm vac,0} \ .$$

Age of universe:

$$\begin{aligned} t_0 &= \frac{1}{H_0} \int_0^1 \frac{x \mathrm{d}x}{\sqrt{\Omega_{m,0} x + \Omega_{\rm rad,0} + \Omega_{\rm vac,0} x^4 + \Omega_{k,0} x^2}} \\ &= \frac{1}{H_0} \int_0^\infty \frac{\mathrm{d}z}{(1+z)\sqrt{\Omega_{m,0}(1+z)^3 + \Omega_{\rm rad,0}(1+z)^4 + \Omega_{\rm vac,0} + \Omega_{k,0}(1+z)^2}} \ . \end{aligned}$$

Look-back time:

$$t_{\text{look-back}}(z) =$$

$$\frac{1}{H_0} \int_0^z \frac{\mathrm{d}z'}{(1+z')\sqrt{\Omega_{m,0}(1+z')^3 + \Omega_{\rm rad,0}(1+z')^4 + \Omega_{\rm vac,0} + \Omega_{k,0}(1+z')^2}} \ .$$

PHYSICAL CONSTANTS:

$$G = 6.674 \times 10^{-11}\ \mathrm{m}^3 \cdot \mathrm{kg}^{-1} \cdot \mathrm{s}^{-2} = 6.674 \times 10^{-8}\ \mathrm{cm}^3 \cdot \mathrm{g}^{-1} \cdot \mathrm{s}^{-2}$$

$$\begin{aligned} k = \text{Boltzmann's constant} &= 1.381 \times 10^{-23}\, \mathrm{joule/K} \\ &= 1.381 \times 10^{-16}\, \mathrm{erg/K} \\ &= 8.617 \times 10^{-5}\, \mathrm{eV/K} \end{aligned}$$

$$\begin{aligned} \hbar = \frac{h}{2\pi} &= 1.055 \times 10^{-34}\ \mathrm{joule} \cdot \mathrm{s} \\ &= 1.055 \times 10^{-27}\ \mathrm{erg} \cdot \mathrm{s} \\ &= 6.582 \times 10^{-16}\ \mathrm{eV} \cdot \mathrm{s} \end{aligned}$$

$$c = 2.998 \times 10^8\ \mathrm{m/s}$$

$$= 2.998 \times 10^{10} \text{ cm/s}$$

$$\hbar c = 197.3 \text{ MeV-fm}, \quad 1 \text{ fm} = 10^{-15} \text{ m}$$

$$1 \text{ yr} = 3.156 \times 10^{7} \text{ s}$$

$$1 \text{ eV} = 1.602 \times 10^{-19} \text{ joule} = 1.602 \times 10^{-12} \text{ erg}$$

$$\begin{aligned} 1 \text{ GeV} = 10^9 \text{ eV} &= 1.783 \times 10^{-27} \text{ kg (where } c \equiv 1) \\ &= 1.783 \times 10^{-24} \text{ g} . \end{aligned}$$

Planck Units: The Planck length ℓ_P, the Planck time t_P, the Planck mass m_P, and the Planck energy E_p are given by

$$\begin{aligned} \ell_P = \sqrt{\frac{G\hbar}{c^3}} &= 1.616 \times 10^{-35} \text{ m} , \\ &= 1.616 \times 10^{-33} \text{ cm} , \\ t_P = \sqrt{\frac{\hbar G}{c^5}} &= 5.391 \times 10^{-44} \text{ s} , \\ m_P = \sqrt{\frac{\hbar c}{G}} &= 2.177 \times 10^{-8} \text{ kg} , \\ &= 2.177 \times 10^{-5} \text{ g} , \\ E_P = \sqrt{\frac{\hbar c^5}{G}} &= 1.221 \times 10^{19} \text{ GeV} . \end{aligned}$$

CHEMICAL EQUILIBRIUM:

(This topic was not included in the course in 2018, but the formulas are nonetheless included here for logical completeness. They will not be relevant to Quiz 3. They are relevant to Problem 14 in these Review Problems, which is also not relevant to Quiz 3. Please enjoy looking at these items, or enjoy ignoring them!)

Ideal Gas of Classical Nonrelativistic Particles:

$$n_i = g_i \frac{(2\pi m_i kT)^{3/2}}{(2\pi\hbar)^3} \, e^{(\mu_i - m_i c^2)/kT} .$$

$$\begin{aligned} \text{where } n_i &= \text{number density of particle} \\ g_i &= \text{number of spin states of particle} \\ m_i &= \text{mass of particle} \\ \mu_i &= \text{chemical potential} \end{aligned}$$

For any reaction, the sum of the μ_i on the left-hand side of the reaction equation must equal the sum of the μ_i on the right-hand side. Formula assumes gas is nonrelativistic ($kT \ll m_i c^2$) and dilute ($n_i \ll (2\pi m_i kT)^{3/2}/(2\pi\hbar)^3$).

PROBLEM LIST

PROBLEM 1: DID YOU DO THE READING? *(25 points)*

The following problem was Problem 1, Quiz 3, in 2007. Each part was worth 5 points.

(a) (CMB basic facts) Which one of the following statements about CMB is *not* correct:

(i) After the dipole distortion of the CMB is subtracted away, the mean temperature averaging over the sky is $\langle T \rangle = 2.725K$.

(ii) After the dipole distortion of the CMB is subtracted away, the root mean square temperature fluctuation is $\left\langle \left(\frac{\delta T}{T}\right)^2 \right\rangle^{1/2} = 1.1 \times 10^{-3}$.

(iii) The dipole distortion is a simple Doppler shift, caused by the net motion of the observer relative to a frame of reference in which the CMB is isotropic.

(iv) In their groundbreaking paper, Wilson and Penzias reported the measurement of an excess temperature of about 3.5 K that was isotropic, unpolarized, and free from seasonal variations. In a companion paper written by Dicke, Peebles, Roll and Wilkinson, the authors interpreted the radiation to be a relic of an early, hot, dense, and opaque state of the universe.

(b) (CMB experiments) The current mean energy per CMB photon, about 6×10^{-4} eV, is comparable to the energy of vibration or rotation for a small molecule such as H_2O. Thus microwaves with wavelengths shorter than $\lambda \sim 3$ cm are strongly absorbed by water molecules in the atmosphere. To measure the CMB at $\lambda < 3$ cm, which one of the following methods is *not* a feasible solution to this problem?

(i) Measure CMB from high-altitude balloons, e.g. MAXIMA.

(ii) Measure CMB from the South Pole, e.g. DASI.

(iii) Measure CMB from the North Pole, e.g. BOOMERANG.

(iv) Measure CMB from a satellite above the atmosphere of the Earth, e.g. COBE, WMAP and PLANCK.

(c) (Temperature fluctuations) The creation of temperature fluctuations in CMB by variations in the gravitational potential is known as the Sachs-Wolfe effect. Which one of the following statements is *not* correct concerning this effect?

(i) A CMB photon is redshifted when climbing out of a gravitational potential well, and is blueshifted when falling down a potential hill.

(ii) At the time of last scattering, the nonbaryonic dark matter dominated the energy density, and hence the gravitational potential, of the universe.

(iii) The large-scale fluctuations in CMB temperatures arise from the gravitational effect of primordial density fluctuations in the distribution of nonbaryonic dark matter.

(iv) The peaks in the plot of temperature fluctuation Δ_T vs. multipole l are due to variations in the density of nonbaryonic dark matter, while the contributions from baryons alone would not show such peaks.

(d) (Dark matter candidates) Which one of the following is *not* a candidate of nonbaryonic dark matter?

(i) massive neutrinos

(ii) axions

(iii) matter made of top quarks (a type of quarks with heavy mass of about 171 GeV).

(iv) WIMPs (Weakly Interacting Massive Particles)

(v) primordial black holes

(e) (Signatures of dark matter) By what methods can signatures of dark matter be detected? List two methods. (Grading: 3 points for one correct answer, 5 points for two correct answers. If you give more than two answers, your score will be based on the number of right answers minus the number of wrong answers, with a lower bound of zero.)

* PROBLEM 2: DID YOU DO THE READING? *(25 points)*

This problem was Problem 1, Quiz 3, 2009.

(a) *(10 points)* This question concerns some numbers related to the cosmic microwave background (CMB) that one should never forget. State the values of these numbers, to within an order of magnitude unless otherwise stated. In all cases the question refers to the present value of these quantities.

(i) The average temperature T of the CMB (to within 10%).

(ii) The speed of the Local Group with respect to the CMB, expressed as a fraction v/c of the speed of light. (The speed of the Local Group is found by measuring the dipole pattern of the CMB temperature to determine the velocity of the spacecraft with respect to the CMB, and then removing spacecraft motion, the orbital motion of the Earth about the Sun, the Sun about the galaxy, and the galaxy relative to the center of mass of the Local Group.)

(iii) The intrinsic relative temperature fluctuations $\Delta T/T$, after removing the dipole anisotropy corresponding to the motion of the observer relative to the CMB.

(iv) The ratio of baryon number density to photon number density, $\eta = n_{\text{bary}}/n_\gamma$.

(v) The angular size θ_H, in degrees, corresponding to what was the Hubble distance c/H at the surface of last scattering. This answer must be within a factor of 3 to be correct.

(b) *(3 points)* Because photons outnumber baryons by so much, the exponential tail of the photon blackbody distribution is important in ionizing hydrogen well after kT_γ falls below $Q_H = 13.6$ eV. What is the ratio kT_γ/Q_H when the ionization fraction of the universe is 1/2?

(i) 1/5 (ii) 1/50 (iii) 10^{-3} (iv) 10^{-4} (v) 10^{-5}

(c) *(2 points)* Which of the following describes the Sachs-Wolfe effect?

(i) Photons from fluid which had a velocity toward us along the line of sight appear redder because of the Doppler effect.

(ii) Photons from fluid which had a velocity toward us along the line of sight appear bluer because of the Doppler effect.

(iii) Photons from overdense regions at the surface of last scattering appear redder because they must climb out of the gravitational potential well.

(iv) Photons from overdense regions at the surface of last scattering appear bluer because they must climb out of the gravitational potential well.

(v) Photons traveling toward us from the surface of last scattering appear redder because of absorption in the intergalactic medium.

(vi) Photons traveling toward us from the surface of last scattering appear bluer because of absorption in the intergalactic medium.

(d) *(10 points)* For each of the following statements, say whether it is true or false:

(i) Dark matter interacts through the gravitational, weak, and electromagnetic forces. T or F ?

(ii) The virial theorem can be applied to a cluster of galaxies to find its total mass, most of which is dark matter. T or F ?

(iii) Neutrinos are thought to comprise a significant fraction of the energy density of dark matter. T or F ?

(iv) Magnetic monopoles are thought to comprise a significant fraction of the energy density of dark matter. T or F ?

(v) Lensing observations have shown that MACHOs cannot account for the dark matter in galactic halos, but that as much as 20% of the halo mass could be in the form of MACHOs. T or F ?

PROBLEM 3: DID YOU DO THE READING? *(35 points)*

This was Problem 1 of Quiz 3, 2013.

(a) *(5 points)* Ryden summarizes the results of the COBE satellite experiment for the measurements of the cosmic microwave background (CMB) in the form of three important results. The first was that, in any particular direction of the sky, the spectrum of the CMB is very close to that of an ideal blackbody. The FIRAS instrument on the COBE satellite could have detected deviations from the blackbody spectrum as small as $\Delta\epsilon/\epsilon \approx 10^{-n}$, where n is an integer. To within ± 1, what is n?

(b) *(5 points)* The second result was the measurement of a dipole distortion of the CMB spectrum; that is, the radiation is slightly blueshifted to higher temperatures in one direction, and slightly redshifted to lower temperatures in the opposite direction. To what physical effect was this dipole distortion attributed?

(c) *(5 points)* The third result concerned the measurement of temperature fluctuations after the dipole feature mentioned above was subtracted out. Defining

$$\frac{\delta T}{T}(\theta,\phi) \equiv \frac{T(\theta,\phi) - \langle T \rangle}{\langle T \rangle} \ ,$$

where $\langle T \rangle = 2.725$ K, the average value of T, they found a root mean square fluctuation,

$$\left\langle \left(\frac{\delta T}{T} \right)^2 \right\rangle^{1/2} \ ,$$

equal to some number. To within an order of magnitude, what was that number?

(d) *(5 points)* Which of the following describes the Sachs-Wolfe effect?

(i) Photons from fluid which had a velocity toward us along the line of sight appear redder because of the Doppler effect.

(ii) Photons from fluid which had a velocity toward us along the line of sight appear bluer because of the Doppler effect.

(iii) Photons from overdense regions at the surface of last scattering appear redder because they must climb out of the gravitational potential well.

(iv) Photons from overdense regions at the surface of last scattering appear bluer because they must climb out of the gravitational potential well.

(v) Photons traveling toward us from the surface of last scattering appear redder because of absorption in the intergalactic medium.

(vi) Photons traveling toward us from the surface of last scattering appear bluer because of absorption in the intergalactic medium.

(e) *(5 points)* The flatness problem refers to the extreme fine-tuning that is needed in Ω at early times, in order for it to be as close to 1 today as we observe. Starting with the assumption that Ω today is equal to 1 within about 1%, one concludes that at one second after the big bang,

$$|\Omega - 1|_{t=1 \text{ sec}} < 10^{-m} ,$$

where m is an integer. To within $\pm$ 3, what is m?

(f) *(5 points)* The total energy density of the present universe consists mainly of baryonic matter, dark matter, and dark energy. Give the percentages of each, according to the best fit obtained from the Planck 2013 data. You will get full credit if the first (baryonic matter) is accurate to $\pm 2\%$, and the other two are accurate to within $\pm 5\%$.

(g) *(5 points)* Within the conventional hot big bang cosmology (without inflation), it is difficult to understand how the temperature of the CMB can be correlated at angular separations that are so large that the points on the surface of last scattering was separated from each other by more than a horizon distance. Approximately what angle, in degrees, corresponds to a separation on the surface last scattering of one horizon length? You will get full credit if your answer is right to within a factor of 2.

PROBLEM 4: DID YOU DO THE READING? (2016) *(25 points)*

Except for part (d), you should answer these questions by circling the one statement that is correct.

(a) *(5 points)* In the Epilogue of *The First Three Minutes*, Steve Weinberg wrote: "The more the universe seems comprehensible, the more it also seems pointless." The sentence was qualified, however, by a closing paragraph that points out that

- (i) the quest of the human race to create a better life for all can still give meaning to our lives.
- (ii) if the universe cannot give meaning to our lives, then perhaps there is an afterlife that will.
- (iii) the complexity and beauty of the laws of physics strongly suggest that the universe must have a purpose, even if we are not aware of what it is.
- (iv) the effort to understand the universe gives human life some of the grace of tragedy.

(b) *(5 points)* In the Afterword of *The First Three Minutes*, Weinberg discusses the baryon number of the universe. (The baryon number of any system is the total number of protons and neutrons (and certain related particles known as hyperons) minus the number of their antiparticles (antiprotons, antineutrons, antihyperons) that are contained in the system.) Weinberg concluded that

(i) baryon number is exactly conserved, so the total baryon number of the universe must be zero. While nuclei in our part of the universe are composed of protons and neutrons, the universe must also contain antimatter regions in which nuclei are composed of antiprotons and antineutrons.

(ii) there appears to be a cosmic excess of matter over antimatter throughout the part of the universe we can observe, and hence a positive density of baryon number. Since baryon number is conserved, this can only be explained by assuming that the excess baryons were put in at the beginning.

(iii) there appears to be a cosmic excess of matter over antimatter throughout the part of the universe we can observe, and hence a positive density of baryon number. This can be taken as a positive hint that baryon number is not conserved, which can happen if there exist as yet undetected heavy "exotic" particles.

(iv) it is possible that baryon number is not exactly conserved, but even if that is the case, it is not possible that the observed excess of matter over antimatter can be explained by the very rare processes that violate baryon number conservation.

(c) *(5 points)* In discussing the COBE measurements of the cosmic microwave background, Ryden describes a dipole component of the temperature pattern, for which the temperature of the radiation from one direction is found to be hotter than the temperature of the radiation detected from the opposite direction.

(i) This discovery is important, because it allows us to pinpoint the direction of the point in space where the big bang occurred.

(ii) This is the largest component of the CMB anisotropies, amounting to a 10% variation in the temperature of the radiation.

(iii) In addition to the dipole component, the anisotropies also includes contributions from a quadrupole, octupole, etc., all of which are comparable in magnitude.

(iv) This pattern is interpreted as a simple Doppler shift, caused by the net motion of the COBE satellite relative to a frame of reference in which the CMB is almost isotropic.

(d) *(5 points)* (CMB basic facts) Which one of the following statements about CMB is *not* correct:

(i) After the dipole distortion of the CMB is subtracted away, the mean temperature averaging over the sky is $\langle T \rangle = 2.725K$.

(ii) After the dipole distortion of the CMB is subtracted away, the root mean square temperature fluctuation is $\left\langle \left(\frac{\delta T}{T}\right)^2 \right\rangle^{1/2} = 1.1 \times 10^{-3}$.

(iii) The dipole distortion is a simple Doppler shift, caused by the net motion of the observer relative to a frame of reference in which the CMB is isotropic.

(iv) In their groundbreaking paper, Wilson and Penzias reported the measurement of an excess temperature of about 3.5 K that was isotropic, unpolarized, and free from seasonal variations. In a companion paper written by Dicke, Peebles, Roll and Wilkinson, the authors interpreted the radiation to be a relic of an early, hot, dense, and opaque state of the universe.

(e) *(5 points)* Inflation is driven by a field that is by definition called the *inflaton* field. In standard inflationary models, the field has the following properties:

(i) The inflaton is a scalar field, and during inflation the energy density of the universe is dominated by its potential energy.

(ii) The inflaton is a vector field, and during inflation the energy density of the universe is dominated by its potential energy.

(iii) The inflaton is a scalar field, and during inflation the energy density of the universe is dominated by its kinetic energy.

(iv) The inflaton is a vector field, and during inflation the energy density of the universe is dominated by its kinetic energy.

(v) The inflaton is a tensor field, which is responsible for only a small fraction of the energy density of the universe during inflation.

PROBLEM 5: NUMBER DENSITIES IN THE COSMIC BACKGROUND RADIATION

Today the temperature of the cosmic microwave background radiation is 2.7°K. Calculate the number density of photons in this radiation. What is the number density of thermal neutrinos left over from the big bang?

∗ PROBLEM 6: PROPERTIES OF BLACK-BODY RADIATION *(25 points)*

The following problem was Problem 4, Quiz 3, 1998.

In answering the following questions, remember that you can refer to the formulas at the front of the exam. Since you were not asked to bring calculators, you may leave your answers in the form of algebraic expressions, such as $\pi^{32}/\sqrt{5\zeta(3)}$.

(a) *(5 points)* For the black-body radiation (also called thermal radiation) of photons at temperature T, what is the average energy per photon?

(b) *(5 points)* For the same radiation, what is the average entropy per photon?

(c) *(5 points)* Now consider the black-body radiation of a massless boson which has spin zero, so there is only one spin state. Would the average energy per particle and entropy per particle be different from the answers you gave in parts (a) and (b)? If so, how would they change?

(d) *(5 points)* Now consider the black-body radiation of electron neutrinos. These particles are fermions with spin 1/2, and we will assume that they are massless and have only one possible spin state. What is the average energy per particle for this case?

(e) *(5 points)* What is the average entropy per particle for the black-body radiation of neutrinos, as described in part (d)?

PROBLEM 7: A NEW SPECIES OF LEPTON

The following problem was Problem 2, Quiz 3, 1992, worth 25 points.

Suppose the calculations describing the early universe were modified by including an additional, hypothetical lepton, called an 8.286ion. The 8.286ion has roughly the same properties as an electron, except that its mass is given by $mc^2 = 0.750$ MeV.

Parts (a)-(c) of this question require numerical answers, but since you were not told to bring calculators, you need not carry out the arithmetic. Your answer should be expressed, however, in "calculator-ready" form— that is, it should be an expression involving pure numbers only (no units), with any necessary conversion factors included. (For example, if you were asked how many meters a light pulse in vacuum travels in 5 minutes, you could express the answer as $2.998 \times 10^8 \times 5 \times 60$.)

a) *(5 points)* What would be the number density of 8.286ions, in particles per cubic meter, when the temperature T was given by $kT = 3$ MeV?

b) *(5 points)* Assuming (as in the standard picture) that the early universe is accurately described by a flat, radiation-dominated model, what would be the value of the mass density at $t = .01$ sec? You may assume that $0.75 \text{ MeV} \ll kT \ll 100 \text{ MeV}$, so the particles contributing significantly to the black-body radiation include the photons, neutrinos, e^+-e^- pairs, and 8.286ion-anti8286ion pairs. Express your answer in the units of g/cm^3.

c) *(5 points)* Under the same assumptions as in (b), what would be the value of kT, in MeV, at $t = .01$ sec?

d) *(5 points)* When nucleosynthesis calculations are modified to include the effect of the 8.286ion, is the production of helium increased or decreased? Explain your answer in a few sentences.

e) *(5 points)* Suppose the neutrinos decouple while $kT \gg 0.75$ MeV. If the 8.286ions are included, what does one predict for the value of T_ν/T_γ today? (Here T_ν denotes the temperature of the neutrinos, and T_γ denotes the temperature of the cosmic background radiation photons.)

∗ PROBLEM 8: A NEW THEORY OF THE WEAK INTERACTIONS *(40 points)*

This problem was Problem 3, Quiz 3, 2009.

Suppose a New Theory of the Weak Interactions (NTWI) was proposed, which differs from the standard theory in two ways. First, the NTWI predicts that the weak interactions are somewhat weaker than in the standard model. In addition, the theory implies the existence of new spin-$\frac{1}{2}$ particles (fermions) called the R^+ and R^-, with a rest energy of 50 MeV (where 1 MeV $= 10^6$ eV). This problem will deal with the cosmological consequences of such a theory.

The NTWI will predict that the neutrinos in the early universe will decouple at a higher temperature than in the standard model. Suppose that this decoupling takes place at $kT \approx 200$ MeV. This means that when the neutrinos cease to be thermally coupled to the rest of matter, the hot soup of particles would contain not only photons, neutrinos, and e^+-e^- pairs, but also μ^+, μ^-, π^+, π^-, and π^0 particles, along with the R^+-R^- pairs. (The muon is a particle which behaves almost identically to an electron, except that its rest energy is 106 MeV. The pions are the lightest of the mesons, with zero angular momentum and rest energies of 135 MeV and 140 MeV for the neutral and charged pions, respectively. The π^+ and π^- are antiparticles of each other, and the π^0 is its own antiparticle. Zero angular momentum implies a single spin state.) You may assume that the universe is flat.

(a) *(10 points)* According to the standard particle physics model, what is the mass density ρ of the universe when $kT \approx 200$ MeV? What is the value of ρ at this temperature, according to NTWI? Use either g/cm^3 or kg/m^3. (If you wish, you can save time by not carrying out the arithmetic. If you do this, however, you should give the answer in "calculator-ready" form, by which I mean an expression involving pure numbers (no units), with any necessary conversion factors included, and with the units of the answer specified at the end. For example, if asked how far light travels in 5 minutes, you could answer $2.998 \times 10^8 \times 5 \times 60$ m.)

(b) *(10 points)* According to the standard model, the temperature today of the thermal neutrino background should be $(4/11)^{1/3} T_\gamma$, where T_γ is the temperature of the thermal photon background. What does the NTWI predict for the temperature of the thermal neutrino background?

(c) *(10 points)* According to the standard model, what is the ratio today of the number density of thermal neutrinos to the number density of thermal photons? What is this ratio according to NTWI?

(d) *(10 points)* Since the reactions which interchange protons and neutrons involve neutrinos, these reactions "freeze out" at roughly the same time as the neutrinos decouple. At later times the only reaction which effectively converts neutrons to protons is the free decay of the neutron. Despite the fact that neutron decay is a weak interaction, we will assume that it occurs with the usual 15 minute mean lifetime. Would the helium abundance predicted by the NTWI be higher or lower than the prediction of the standard model? To within 5 or 10%, what would the NTWI predict for the percent abundance (by weight) of helium in the universe? (As in part (a), you can either carry out the arithmetic, or leave the answer in calculator-ready form.)

Useful information: The proton and neutron rest energies are given by $m_p c^2 = 938.27$ MeV and $m_n c^2 = 939.57$ MeV, with $(m_n - m_p)c^2 = 1.29$ MeV. The mean lifetime for the neutron decay, $n \to p + e^- + \bar{\nu}_e$, is given by $\tau = 886$ s.

PROBLEM 9: DOUBLING OF ELECTRONS *(10 points)*

The following was on Quiz 3, 2011 (Problem 4):

Suppose that instead of one species of electrons and their antiparticles, suppose there was also another species of electron-like and positron-like particles. Suppose that the new species has the same mass and other properties as the electrons and positrons. If this were the case, what would be the ratio T_ν/T_γ of the temperature today of the neutrinos to the temperature of the CMB photons.

PROBLEM 10: TIME SCALES IN COSMOLOGY

In this problem you are asked to give the approximate times at which various important events in the history of the universe are believed to have taken place. The times are measured from the instant of the big bang. To avoid ambiguities, you are asked to choose the best answer from the following list:

10^{-43} sec.
10^{-37} sec.
10^{-12} sec.
10^{-5} sec.
1 sec.
4 mins.
10,000 – 1,000,000 years.
2 billion years.
5 billion years.
10 billion years.
13 billion years.
20 billion years.

For this problem it will be sufficient to state an answer from memory, without explanation. The events which must be placed are the following:

(a) the beginning of the processes involved in big bang nucleosynthesis;

(b) the end of the processes involved in big bang nucleosynthesis;

(c) the time of the phase transition predicted by grand unified theories, which takes place when $kT \approx 10^{16}$ GeV;

(d) "recombination", the time at which the matter in the universe converted from a plasma to a gas of neutral atoms;

(e) the phase transition at which the quarks became confined, believed to occur when $kT \approx 300$ MeV.

Since cosmology is fraught with uncertainty, in some cases more than one answer will be acceptable. You are asked, however, to give **ONLY ONE** of the acceptable answers.

* PROBLEM 11: EVOLUTION OF FLATNESS *(15 points)*

The following problem was Problem 3, Quiz 3, 2004.

The "flatness problem" is related to the fact that during the evolution of the standard cosmological model, Ω is always driven away from 1.

(a) *(9 points)* During a period in which the universe is matter-dominated (meaning that the only relevant component is nonrelativistic matter), the quantity

$$\frac{\Omega - 1}{\Omega}$$

grows as a power of t. Show that this is true, and derive the power. (Stating the right power without a derivation will be worth 3 points.)

(b) *(6 points)* During a period in which the universe is radiation-dominated, the same quantity will grow like a different power of t. Show that this is true, and derive the power. (Stating the right power without a derivation will again be worth 3 points.)

In each part, you may assume that the universe was *always* dominated by the specified form of matter.

* PROBLEM 12: THE SLOAN DIGITAL SKY SURVEY $z = 5.82$ QUASAR *(40 points)*

The following problem was Problem 4, Quiz 3, 2004.

On April 13, 2000, the Sloan Digital Sky Survey announced the discovery of what was then the most distant object known in the universe: a quasar at $z = 5.82$. To explain to the public how this object fits into the universe, the SDSS posted on their website an article by Michael Turner and Craig Wiegert titled "How Can An Object We See Today be 27 Billion Light Years Away If the Universe is only 14 Billion Years Old?" Using a model with $H_0 = 65$ km-s^{-1}-Mpc^{-1}, $\Omega_m = 0.35$, and $\Omega_\Lambda = 0.65$, they claimed

(a) that the age of the universe is 13.9 billion years.

(b) that the light that we now see was emitted when the universe was 0.95 billion years old.

(c) that the distance to the quasar, as it would be measured by a ruler today, is 27 billion light-years.

(d) that the distance to the quasar, at the time the light was emitted, was 4.0 billion light-years.

(e) that the present speed of the quasar, defined as the rate at which the distance between us and the quasar is increasing, is 1.8 times the velocity of light.

The goal of this problem is to check all of these conclusions, although you are of course not expected to actually work out the numbers. Your answers can be expressed in terms of H_0, Ω_m, Ω_Λ, and z. Definite integrals need not be evaluated.

Note that Ω_m represents the present density of nonrelativistic matter, expressed as a fraction of the critical density; and Ω_Λ represents the present density of vacuum energy, expressed as a fraction of the critical density. In answering each of the following questions, you may consider the answer to any previous part — whether you answered it or not — as a given piece of information, which can be used in your answer.

(a) *(15 points)* Write an expression for the age t_0 of this model universe?

(b) *(5 points)* Write an expression for the time t_e at which the light which we now receive from the distant quasar was emitted.

(c) *(10 points)* Write an expression for the present physical distance $\ell_{\text{phys},0}$ to the quasar.

(d) *(5 points)* Write an expression for the physical distance $\ell_{\text{phys,e}}$ between us and the quasar at the time that the light was emitted.

(e) *(5 points)* Write an expression for the present speed of the quasar, defined as the rate at which the distance between us and the quasar is increasing.

PROBLEM 13: SECOND HUBBLE CROSSING *(40 points)*

This problem was Problem 3, Quiz 3, 2007. In 2018 we have not yet talked about Hubble crossings and the evolution of density perturbations, so this problem would not be fair as worded. Actually, however, you have learned how to do these calculations, so the problem would be fair if it described in more detail what needs to be calculated.

In Problem Set 9 (2007) we calculated the time $t_{H1}(\lambda)$ of the first Hubble crossing for a mode specified by its (physical) wavelength λ at the present time. In this problem we will calculate the time $t_{H2}(\lambda)$ of the second Hubble crossing, the time at which the growing Hubble length $cH^{-1}(t)$ catches up to the physical wavelength, which is also growing. At the time of the second Hubble crossing for the wavelengths of interest, the universe can be described very simply: it is a radiation-dominated flat universe. However, since λ is defined as the present value of the wavelength, the evolution of the universe between $t_{H2}(\lambda)$ and the present will also be relevant to the problem. We will need to use methods, therefore, that allow for both the matter-dominated era and the onset of the dark-energy-dominated era. As in Problem Set 9 (2007), the model universe that we consider will be described by the WMAP 3-year best fit parameters:

Hubble expansion rate	H_0	$= 73.5\ \text{km}\cdot\text{s}^{-1}\cdot\text{Mpc}^{-1}$
Nonrelativistic mass density	Ω_m	$= 0.237$
Vacuum mass density	Ω_{vac}	$= 0.763$
CMB temperature	$T_{\gamma,0}$	$= 2.725\ \text{K}$

The mass densities are defined as contributions to Ω, and hence describe the mass density of each constituent relative to the critical density. Note that the model is exactly flat, so you need not worry about spatial curvature. Here you are not expected to give a numerical answer, so the above list will serve only to define the symbols that can appear in your answers, along with λ and the physical constants G, $\hbar$, c, and k.

(a) *(5 points)* For a radiation-dominated flat universe, what is the Hubble length $\ell_H(t) \equiv cH^{-1}(t)$ as a function of time t?

(b) *(10 points)* The second Hubble crossing will occur during the interval

$$30\ \text{sec} \ll t \ll 50,000\ \text{years},$$

when the mass density of the universe is dominated by photons and neutrinos. During this era the neutrinos are a little colder than the photons, with $T_\nu = (4/11)^{1/3} T_\gamma$. The total energy density of the photons and neutrinos together can be written as

$$u_{\text{tot}} = g_1 \frac{\pi^2}{30} \frac{(kT_\gamma)^4}{(\hbar c)^3} \ .$$

What is the value of g_1? (For the following parts you can treat g_1 as a given variable that can be left in your answers, whether or not you found it.)

(c) *(10 points)* For times in the range described in part (b), what is the photon temperature $T_\gamma(t)$ as a function of t?

(d) *(15 points)* Finally, we are ready to find the time $t_{H2}(\lambda)$ of the second Hubble crossing, for a given value of the physical wavelength λ today. Making use of the previous results, you should be able to determine $t_{H2}(\lambda)$. If you were not able to answer some of the previous parts, you may leave the symbols $\ell_H(t)$, g_1, and/or $T_\gamma(t)$ in your answer.

PROBLEM 14: NEUTRINO NUMBER AND THE NEUTRON/ PROTON EQUILIBRIUM *(35 points)*

The following problem was 1998 Quiz 4, Problem 4. This would NOT be a fair problem for 2018, as this year we have not discussed big bang nucleosynthesis at this level of detail. But I am including the problem anyway, as you might find it interesting.

In the standard treatment of big bang nucleosynthesis it is assumed that at early times the ratio of neutrons to protons is given by the Boltzmann formula,

$$\frac{n_n}{n_p} = e^{-\Delta E/kT} \ , \tag{1}$$

where k is Boltzmann's constant, T is the temperature, and $\Delta E = 1.29$ MeV is the proton-neutron mass-energy difference. This formula is believed to be very accurate, but it assumes that the chemical potential for neutrons μ_n is the same as the chemical potential for protons μ_p.

(a) *(10 points)* Give the correct version of Eq. (1), allowing for the possibility that $\mu_n \neq \mu_p$.

The equilibrium between protons and neutrons in the early universe is sustained mainly by the following reactions:

$$e^+ + n \longleftrightarrow p + \bar{\nu}_e$$

$$\nu_e + n \longleftrightarrow p + e^- \ .$$

Let μ_e and μ_ν denote the chemical potentials for the electrons (e^-) and the electron neutrinos (ν_e) respectively. The chemical potentials for the positrons (e^+) and the anti-electron neutrinos ($\bar{\nu}_e$) are then $-\mu_e$ and $-\mu_\nu$, respectively, since the chemical potential of a particle is always the negative of the chemical potential for the antiparticle.*

(b) *(10 points)* Express the neutron/proton chemical potential difference $\mu_n - \mu_p$ in terms of μ_e and μ_ν.

The black-body radiation formulas at the beginning of the quiz did not allow for the possibility of a chemical potential, but they can easily be generalized. For example, the formula for the number density n_i (of particles of type i) becomes

$$n_i = g_i^* \frac{\zeta(3)}{\pi^2} \frac{(kT)^3}{(\hbar c)^3} e^{\mu_i/kT} \ .$$

(c) *(10 points)* Suppose that the density of anti-electron neutrinos $\bar{n}_\nu$ in the early universe was higher than the density of electron neutrinos n_ν. Express the thermal

---

* This fact is a consequence of the principle that the chemical potential of a particle is the sum of the chemical potentials associated with its conserved quantities, while particle and antiparticle always have the opposite values of all conserved quantities.

equilibrium value of the ratio n_n/n_p in terms of ΔE, T, and either the ratio $\bar{n}_\nu/n_\nu$ or the antineutrino excess $\Delta n = \bar{n}_\nu - n_\nu$. (Your answer may also contain fundamental constants, such as k, $\hbar$, and c.)

(d) *(5 points)* Would an excess of anti-electron neutrinos, as considered in part (c), increase or decrease the amount of helium that would be produced in the early universe? Explain your answer.

PROBLEM 15: THE EVENT HORIZON FOR OUR UNIVERSE *(25 points)*

The following problem was Problem 3 from Quiz 3, 2013.

We have learned that the expansion history of our universe can be described in terms of a small set of numbers: $\Omega_{m,0}$, the present contribution to Ω from nonrelativistic matter; $\Omega_{\rm rad,0}$, the present contribution to Ω from radiation; $\Omega_{\rm vac}$, the present contribution to Ω from vacuum energy; and H_0, the present value of the Hubble expansion rate. The best estimates of these numbers are consistent with a flat universe, so we can take $k = 0$, $\Omega_{m,0} + \Omega_{\rm rad,0} + \Omega_{\rm vac} = 1$, and we can use the flat Robertson-Walker metric,

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

(a) *(5 points)* Suppose that we are at the origin of the coordinate system, and that at the present time t_0 we emit a spherical pulse of light. It turns out that there is a maximum coordinate radius $r = r_{\rm max}$ that this pulse will ever reach, no matter how long we wait. (The pulse will never actually reach $r_{\rm max}$, but will reach all r such that $0 < r < r_{\rm max}$.) $r_{\rm max}$ is the coordinate of what is called the *event horizon*: events that happen now at $r \geq r_{\rm max}$ will never be visible to us, assuming that we remain at the origin. Assuming for this part that the function $a(t)$ is a known function, write an expression for $r_{\rm max}$. Your answer should be expressed as an integral, which can involve $a(t)$, t_0, and any of the parameters defined in the preamble. [*Advice:* If you cannot answer this, you should still try part (c).]

(b) *(10 points)* Since $a(t)$ is not known explicitly, the answer to the previous part is difficult to use. Show, however, that by changing the variable of integration, you can rewrite the expression for $r_{\rm max}$ as a definite integral involving only the parameters specified in the preamble, without any reference to the function $a(t)$, except perhaps to its present value $a(t_0)$. You are not expected to evaluate this integral. [*Hint:* One method is to use

$$x = \frac{a(t)}{a(t_0)}$$

as the variable of integration, just as we did when we derived the first of the expressions for t_0 shown in the formula sheets.]

(c) *(10 points)* Astronomers often describe distances in terms of redshifts, so it is useful to find the redshift of the event horizon. That is, if a light ray that originated at

$r = r_{\max}$ arrived at Earth today, what would be its redshift $z_{\rm eh}$ (eh = event horizon)? You are not asked to find an explicit expression for $z_{\rm eh}$, but instead an equation that could be solved numerically to determine $z_{\rm eh}$. For this part you can treat $r_{\max}$ as given, so it does not matter if you have done parts (a) and (b). You will get half credit for a correct answer that involves the function $a(t)$, and full credit for a correct answer that involves only explicit integrals depending only on the parameters specified in the preamble, and possibly $a(t_0)$.

PROBLEM 16: THE EFFECT OF PRESSURE ON COSMOLOGICAL EVOLUTION *(25 points)*

The following problem was Problem 2 of Quiz 3, 2016. It was also Problem 2 of Problem Set 7 (2016), except that some numerical constants have been changed, so the answers will not be identical.

A radiation-dominated universe behaves differently from a matter-dominated universe because the pressure of the radiation is significant. In this problem we explore the role of pressure for several fictitious forms of matter.

(a) *(8 points)* For the first fictitious form of matter, the mass density ρ decreases as the scale factor $a(t)$ grows, with the relation

$$\rho(t) \propto \frac{1}{a^8(t)} \ .$$

What is the pressure of this form of matter? *[Hint: the answer is proportional to the mass density.]*

(b) *(9 points)* Find the behavior of the scale factor $a(t)$ for a flat universe dominated by the form of matter described in part (a). You should be able to determine the function $a(t)$ up to a constant factor.

(c) *(8 points)* Now consider a universe dominated by a different form of fictitious matter, with a pressure given by

$$p = \frac{2}{3}\rho c^2 \ .$$

As the universe expands, the mass density of this form of matter behaves as

$$\rho(t) \propto \frac{1}{a^n(t)} \ .$$

Find the power n.

∗ PROBLEM 17: THE FREEZE-OUT OF A FICTITIOUS PARTICLE X *(25 points)*

The following problem was Problem 3 of Quiz 3, 2016.

Suppose that, in addition to the particles that are known to exist, there also existed a family of three spin-1 particles, X^+, X^-, and X^0, all with masses 0.511 MeV/c^2, exactly the same as the electron. The X^- is the antiparticle of the X^+, and the X^0 is its own antiparticle. Since the X's are spin-1 particles with nonzero mass, each particle has three spin states.

The X's do not interact with neutrinos any more strongly than the electrons and positrons do, so when the X's freeze out, all of their energy and entropy are given to the photons, just like the electron-positron pairs.

(a) *(5 points)* In thermal equilibrium when $kT \gg 0.511$ MeV/c^2, what is the total energy density of the X^+, X^-, and X^0 particles?

(b) *(5 points)* In thermal equilibrium when $kT \gg 0.511$ MeV/c^2, what is the total number density of the X^+, X^-, and X^0 particles?

(c) *(10 points)* The X particles and the electron-positron pairs freeze out of the thermal equilibrium radiation at the same time, as kT decreases from values large compared to 0.511 MeV/c^2 to values that are small compared to it. If the X's, electron-positron pairs, photons, and neutrinos were all in thermal equilibrium before this freeze-out, what will be the ratio T_ν/T_γ, the ratio of the neutrino temperature to the photon temperature, after the freeze-out?

(d) *(5 points)* If the mass of the X's was, for example, 0.100 MeV/c^2, so that the electron-positron pairs froze out first, and then the X's froze out, would the final ratio T_ν/T_γ be higher, lower, or the same as the answer to part (c)? Explain your answer in a sentence or two.

∗ PROBLEM 18: THE TIME t_d OF DECOUPLING *(25 points)*

The following problem was Problem 4 of Quiz 3, 2016.

The process by which the photons of the cosmic microwave background stop scattering and begin to travel on straight lines is called *decoupling*, and it happens at a photon temperature of about $T_d \approx 3,000$ K. In Lecture Notes 6 we estimated the time t_d of decoupling, working in the approximation that the universe has been matter-dominated from that time to the present. We found a value of 370,000 years. In this problem we will remove this approximation, although we will not carry out the numerical evaluation needed to compare with the previous answer.

(a) *(5 points)* Let us define

$$x(t) \equiv \frac{a(t)}{a(t_0)} \ ,$$

as on the formula sheets, where t_0 is the present time. What is the value of $x_d \equiv x(t_d)$? Assume that the entropy of photons is conserved from time t_d to the present, and let T_0 denote the present photon temperature.

(b) *(5 points)* Assume that the universe is flat, and that $\Omega_{m,0}$, $\Omega_{\text{rad},0}$, and $\Omega_{\text{vac},0}$ denote the present contributions to Ω from nonrelativistic matter, radiation, and vacuum energy, respectively. Let H_0 denote the present value of the Hubble expansion rate. Write an expression in terms of these quantities for $\mathrm{d}x/\mathrm{d}t$, the derivative of x with respect to t. *Hint: you may use formulas from the formula sheet without derivation, so this problem should require essentially no work. To receive full credit, your answer should include only terms that make a nonzero contribution to the answer.*

(c) *(5 points)* Write an expression for t_d. If your answer involves an integral, you need not try to evaluate it, but you should be sure that the limits of integration are clearly shown.

(d) *(10 points)* Now suppose that in addition to the constituents described in part (b), the universe also contains some of the fictitious material from part (a) of Problem 2, with

$$\rho(t) \propto \frac{1}{a^8(t)} \ .$$

Denote the present contribution to Ω from this fictitious material as $\Omega_{f,0}$. The universe is still assumed to be flat, so the numerical values of $\Omega_{m,0}$, $\Omega_{\text{rad},0}$, and $\Omega_{\text{vac},0}$ must sum to a smaller value than in parts (b) and (c). With this extra contribution to the mass density of the universe, what is the new expression for t_d?

SOLUTIONS

PROBLEM 1: DID YOU DO THE READING? *(25 points)*

The following parts are each worth 5 points.

(a) (CMB basic facts) Which one of the following statements about CMB is *not* correct:

(i) After the dipole distortion of the CMB is subtracted away, the mean temperature averaging over the sky is $\langle T \rangle = 2.725K$.

$\boxed{\text{(ii)}}$ After the dipole distortion of the CMB is subtracted away, the root mean square temperature fluctuation is $\left\langle \left(\frac{\delta T}{T}\right)^2 \right\rangle^{1/2} = 1.1 \times 10^{-3}$.

(iii) The dipole distortion is a simple Doppler shift, caused by the net motion of the observer relative to a frame of reference in which the CMB is isotropic.

(iv) In their groundbreaking paper, Wilson and Penzias reported the measurement of an excess temperature of about 3.5 K that was isotropic, unpolarized, and free from seasonal variations. In a companion paper written by Dicke, Peebles, Roll and Wilkinson, the authors interpreted the radiation to be a relic of an early, hot, dense, and opaque state of the universe.

Explanation: After subtracting the dipole contribution, the temperature fluctuation is about 1.1×10^{-5}.

(b) (CMB experiments) The current mean energy per CMB photon, about 6×10^{-4} eV, is comparable to the energy of vibration or rotation for a small molecule such as H_2O. Thus microwaves with wavelengths shorter than $\lambda \sim 3$ cm are strongly absorbed by water molecules in the atmosphere. To measure the CMB at $\lambda < 3$ cm, which one of the following methods is *not* a feasible solution to this problem?

(i) Measure CMB from high-altitude balloons, e.g. MAXIMA.

(ii) Measure CMB from the South Pole, e.g. DASI.

$\boxed{\text{(iii)}}$ Measure CMB from the North Pole, e.g. BOOMERANG.

(iv) Measure CMB from a satellite above the atmosphere of the Earth, e.g. COBE, WMAP and PLANCK.

Explanation: The North Pole is at sea level. In contrast, the South Pole is nearly 3 kilometers above sea level. BOOMERANG is a balloon-borne experiment launched from Antarctica.

(c) (Temperature fluctuations) The creation of temperature fluctuations in CMB by variations in the gravitational potential is known as the Sachs-Wolfe effect. Which one of the following statements is *not* correct concerning this effect?

(i) A CMB photon is redshifted when climbing out of a gravitational potential well, and is blueshifted when falling down a potential hill.

(ii) At the time of last scattering, the nonbaryonic dark matter dominated the energy density, and hence the gravitational potential, of the universe.

(iii) The large-scale fluctuations in CMB temperatures arise from the gravitational effect of primordial density fluctuations in the distribution of nonbaryonic dark matter.

$\boxed{\text{(iv)}}$ The peaks in the plot of temperature fluctuation Δ_T vs. multipole l are due to variations in the density of nonbaryonic dark matter, while the contributions from baryons alone would not show such peaks.

Explanation: These peaks are due to the acoustic oscillations in the photon-baryon fluid.

(d) (Dark matter candidates) Which one of the following is *not* a candidate of nonbaryonic dark matter?

(i) massive neutrinos

(ii) axions

$\boxed{\text{(iii)}}$ matter made of top quarks (a type of quarks with heavy mass of about 171 GeV).

(iv) WIMPs (Weakly Interacting Massive Particles)

(v) primordial black holes

Explanation: Matter made of top quarks is so unstable that it is seen only fleetingly as a product in high energy particle collisions.

(e) (Signatures of dark matter) By what methods can signatures of dark matter be detected? List two methods. (Grading: 3 points for one correct answer, 5 points for two correct answers. If you give more than two answers, your score will be based on the number of right answers minus the number of wrong answers, with a lower bound of zero.)

Answers:

(i) Galaxy rotation curves. (I.e., measurements of the orbital speed of stars in spiral galaxies as a function of radius R show that these curves remain flat at radii far beyond the visible stellar disk. If most of the matter were contained in the disk, then these velocities should fall off as $1/\sqrt{R}$.)

(ii) Use the virial theorem to estimate the mass of a galaxy cluster. (For example, the virial analysis shows that only 2% of the mass of the Coma cluster consists of stars, and only 10% consists of hot intracluster gas.

(iii) *Gravitational lensing. (For example, the mass of a cluster can be estimated from the distortion of the shapes of the galaxies behind the cluster.)*

(iv) *CMB temperature fluctuations. (I.e., the analysis of the intensity of the fluctuations as a function of multipole number shows that $\Omega_{\rm tot} \approx 1$, and that dark energy contributes $\Omega_\Lambda \approx 0.7$, baryonic matter contributes $\Omega_{\rm bary} \approx 0.04$, and dark matter contributes $\Omega_{\rm dark\ matter} \approx 0.26$.)*

There are other possible answers as well, but these are the ones discussed by Ryden in Chapters 8 and 9.

PROBLEM 2: DID YOU DO THE READING? *(25 points)*

(a) *(10 points)* This question concerns some numbers related to the cosmic microwave background (CMB) that one should never forget. State the values of these numbers, to within an order of magnitude unless otherwise stated. In all cases the question refers to the present value of these quantities.

(i) The average temperature T of the CMB (to within 10%). $\boxed{2.725\,\mathrm{K}}$

(ii) The speed of the Local Group with respect to the CMB, expressed as a fraction v/c of the speed of light. (The speed of the Local Group is found by measuring the dipole pattern of the CMB temperature to determine the velocity of the spacecraft with respect to the CMB, and then removing spacecraft motion, the orbital motion of the Earth about the Sun, the Sun about the galaxy, and the galaxy relative to the center of mass of the Local Group.)

The dipole anisotropy corresponds to a "peculiar velocity" (that is, velocity which is not due to the expansion of the universe) of $630 \pm 20\,\mathrm{km\,s^{-1}}$*, or in terms of the speed of light,* $\boxed{v/c \approx 2 \times 10^{-3}}$.

(iii) The intrinsic relative temperature fluctuations $\Delta T/T$, after removing the dipole anisotropy corresponding to the motion of the observer relative to the CMB. $\boxed{1.1 \times 10^{-5}}$

(iv) The ratio of baryon number density to photon number density, $\eta = n_{\rm bary}/n_\gamma$.

The WMAP 5-year value for $\eta = n_b/n_\gamma = \boxed{(6.225 \pm 0.170) \times 10^{-10}}$*, which to closest order of magnitude is* 10^{-9}.

(v) The angular size θ_H, in degrees, corresponding to what was the Hubble distance c/H at the surface of last scattering. This answer must be within a factor of 3 to be correct. $\boxed{\sim 1^\circ}$

(b) *(3 points)* Because photons outnumber baryons by so much, the exponential tail of the photon blackbody distribution is important in ionizing hydrogen well after kT_γ

falls below $Q_H = 13.6$ eV. What is the ratio kT_γ/Q_H when the ionization fraction of the universe is 1/2?

(i) 1/5 $\boxed{\text{(ii) } 1/50}$ (iii) 10^{-3} (iv) 10^{-4} (v) 10^{-5}

This is not a number one has to commit to memory if one can remember the temperature of (re)combination in eV*, or if only in* K *along with the conversion factor* ($k \approx 10^{-4}\,\text{eV}\,\text{K}^{-1}$)*. One can then calculate that near recombination,* $kT_\gamma/Q_H \approx (10^{-4}\,\text{eV}\,\text{K}^{-1})(3000\,\text{K})/(13.6\,\text{eV}) \approx 1/45$.

(c) *(2 points)* Which of the following describes the Sachs-Wolfe effect?

(i) Photons from fluid which had a velocity toward us along the line of sight appear redder because of the Doppler effect.

(ii) Photons from fluid which had a velocity toward us along the line of sight appear bluer because of the Doppler effect.

$\boxed{\text{(iii)}}$ Photons from overdense regions at the surface of last scattering appear redder because they must climb out of the gravitational potential well.

(iv) Photons from overdense regions at the surface of last scattering appear bluer because they must climb out of the gravitational potential well.

(v) Photons traveling toward us from the surface of last scattering appear redder because of absorption in the intergalactic medium.

(vi) Photons traveling toward us from the surface of last scattering appear bluer because of absorption in the intergalactic medium.

Explanation: Denser regions have a deeper (more negative) gravitational potential. Photons which travel through a spatially varying potential acquire a redshift or blueshift depending on whether they are going up or down the potential, respectively. Photons originating in the denser regions start at a lower potential and must climb out, so they end up being redshifted relative to their original energies.

(d) *(10 points)* For each of the following statements, say whether it is true or false:

(i) Dark matter interacts through the gravitational, weak, and electromagnetic forces. T or $\boxed{\text{F}}$?

(ii) The virial theorem can be applied to a cluster of galaxies to find its total mass, most of which is dark matter. $\boxed{\text{T}}$ or F ?

(iii) Neutrinos are thought to comprise a significant fraction of the energy density of dark matter. T or $\boxed{\text{F}}$?

(iv) Magnetic monopoles are thought to comprise a significant fraction of the energy density of dark matter. T or $\boxed{\text{F}}$?

(v) Lensing observations have shown that MACHOs cannot account for the dark matter in galactic halos, but that as much as 20% of the halo mass could be in the form of MACHOs. $\boxed{\text{T}}$ or F ?

PROBLEM 3: DID YOU DO THE READING? *(35 points)*

(a) *(5 points)* Ryden summarizes the results of the COBE satellite experiment for the measurements of the cosmic microwave background (CMB) in the form of three important results. The first was that, in any particular direction of the sky, the spectrum of the CMB is very close to that of an ideal blackbody. The FIRAS instrument on the COBE satellite could have detected deviations from the blackbody spectrum as small as $\Delta\epsilon/\epsilon \approx 10^{-n}$, where n is an integer. To within ± 1, what is n?

Answer: $n = 4$

(b) *(5 points)* The second result was the measurement of a dipole distortion of the CMB spectrum; that is, the radiation is slightly blueshifted to higher temperatures in one direction, and slightly redshifted to lower temperatures in the opposite direction. To what physical effect was this dipole distortion attributed?

Answer: The large dipole in the CMB is attributed to the motion of the satellite relative to the frame in which the CMB is very nearly isotropic. (The entire Local Group is moving relative to this frame at a speed of about $0.002c$.)

(c) *(5 points)* The third result concerned the measurement of temperature fluctuations after the dipole feature mentioned above was subtracted out. Defining

$$\frac{\delta T}{T}(\theta, \phi) \equiv \frac{T(\theta, \phi) - \langle T \rangle}{\langle T \rangle} ,$$

where $\langle T \rangle = 2.725$ K, the average value of T, they found a root mean square fluctuation,

$$\left\langle \left(\frac{\delta T}{T} \right)^2 \right\rangle^{1/2} ,$$

equal to some number. To within an order of magnitude, what was that number?

Answer:

$$\left\langle \left(\frac{\delta T}{T} \right)^2 \right\rangle^{1/2} = 1.1 \times 10^{-5} .$$

(d) *(5 points)* Which of the following describes the Sachs-Wolfe effect?

(i) Photons from fluid which had a velocity toward us along the line of sight appear redder because of the Doppler effect.

(ii) Photons from fluid which had a velocity toward us along the line of sight appear bluer because of the Doppler effect.

$\boxed{\text{(iii)}}$ Photons from overdense regions at the surface of last scattering appear redder because they must climb out of the gravitational potential well.

(iv) Photons from overdense regions at the surface of last scattering appear bluer because they must climb out of the gravitational potential well.

(v) Photons traveling toward us from the surface of last scattering appear redder because of absorption in the intergalactic medium.

(vi) Photons traveling toward us from the surface of last scattering appear bluer because of absorption in the intergalactic medium.

(e) *(5 points)* The flatness problem refers to the extreme fine-tuning that is needed in Ω at early times, in order for it to be as close to 1 today as we observe. Starting with the assumption that Ω today is equal to 1 within about 1%, one concludes that at one second after the big bang,

$$|\Omega - 1|_{t=1 \text{ sec}} < 10^{-m} ,$$

where m is an integer. To within $\pm$ 3, what is m?

Answer: $m = 18$. (See the derivation in Lecture Notes 8.)

(f) *(5 points)* The total energy density of the present universe consists mainly of baryonic matter, dark matter, and dark energy. Give the percentages of each, according to the best fit obtained from the Planck 2013 data. You will get full credit if the first (baryonic matter) is accurate to $\pm 2\%$, and the other two are accurate to within $\pm 5\%$.

Answer: Baryonic matter: 5%. Dark matter: 26.5%. Dark energy: 68.5%. The Planck 2013 numbers were given in Lecture Notes 7. To the requested accuracy, however, numbers such as Ryden's Benchmark Model would also be satisfactory.

(g) *(5 points)* Within the conventional hot big bang cosmology (without inflation), it is difficult to understand how the temperature of the CMB can be correlated at angular separations that are so large that the points on the surface of last scattering was separated from each other by more than a horizon distance. Approximately what angle, in degrees, corresponds to a separation on the surface last scattering of one horizon length? You will get full credit if your answer is right to within a factor of 2.

Answer: Ryden gives 1° as the angle subtended by the Hubble length on the surface of last scattering. For a matter-dominated universe, which would be a good model for our universe, the horizon length is twice the Hubble length. Any number from 1° to 5° was considered acceptable.

PROBLEM 4: DID YOU DO THE READING? *(25 points)*

Except for part (d), you should answer these questions by circling the one statement that is correct.

(a) *(5 points)* In the Epilogue of *The First Three Minutes*, Steve Weinberg wrote: "The more the universe seems comprehensible, the more it also seems pointless." The sentence was qualified, however, by a closing paragraph that points out that

(i) the quest of the human race to create a better life for all can still give meaning to our lives.

(ii) if the universe cannot give meaning to our lives, then perhaps there is an afterlife that will.

(iii) the complexity and beauty of the laws of physics strongly suggest that the universe must have a purpose, even if we are not aware of what it is.

$\boxed{\text{(iv)}}$ the effort to understand the universe gives human life some of the grace of tragedy.

(b) *(5 points)* In the Afterword of *The First Three Minutes*, Weinberg discusses the baryon number of the universe. (The baryon number of any system is the total number of protons and neutrons (and certain related particles known as hyperons) minus the number of their antiparticles (antiprotons, antineutrons, antihyperons) that are contained in the system.) Weinberg concluded that

(i) baryon number is exactly conserved, so the total baryon number of the universe must be zero. While nuclei in our part of the universe are composed of protons and neutrons, the universe must also contain antimatter regions in which nuclei are composed of antiprotons and antineutrons.

(ii) there appears to be a cosmic excess of matter over antimatter throughout the part of the universe we can observe, and hence a positive density of baryon number. Since baryon number is conserved, this can only be explained by assuming that the excess baryons were put in at the beginning.

$\boxed{\text{(iii)}}$ there appears to be a cosmic excess of matter over antimatter throughout the part of the universe we can observe, and hence a positive density of baryon number. This can be taken as a positive hint that baryon number is not conserved, which can happen if there exist as yet undetected heavy "exotic" particles.

(iv) it is possible that baryon number is not exactly conserved, but even if that is the case, it is not possible that the observed excess of matter over antimatter can be explained by the very rare processes that violate baryon number conservation.

Explanation: All students were given credit for this part, whether they answered it correctly or not. I was in San Francisco when I made up this quiz, and due to poor planning I did not have my copy of *The First Three Minutes.* So I found a version online, but I could only find the British version, published by Flamingo/Fontana Paperbacks, rather than the US version published by Basic Books. I assumed that the "Afterword" in the two versions would be the same, but I was wrong! So this question was based on a different "Afterword" than the one that you read. 55% of you still got it right, but obviously the question was not fair. Apologies.

(c) *(5 points)* In discussing the COBE measurements of the cosmic microwave background, Ryden describes a dipole component of the temperature pattern, for which the temperature of the radiation from one direction is found to be hotter than the temperature of the radiation detected from the opposite direction.

(i) This discovery is important, because it allows us to pinpoint the direction of the point in space where the big bang occurred.

(ii) This is the largest component of the CMB anisotropies, amounting to a 10% variation in the temperature of the radiation.

(iii) In addition to the dipole component, the anisotropies also include contributions from a quadrupole, octupole, etc., all of which are comparable in magnitude.

$\boxed{\text{(iv)}}$ This pattern is interpreted as a simple Doppler shift, caused by the net motion of the COBE satellite relative to a frame of reference in which the CMB is almost isotropic.

Explanation: (i) is nonsense, since the conventional big bang theory descibes a completely homogeneous universe, which has no single point at which the big bang occurred. (ii) is wrong, because the variations in the temperature of the CMB are much smaller than 10%. The dipole term has a magnitude of about 1/1000 of the mean temperature. (iii) is wrong because the dipole is not comparable to the other terms, because they have magnitudes of only about 1/100,000 of the mean.

(d) *(5 points)* (CMB basic facts) Which one of the following statements about CMB is *not* correct:

(i) After the dipole distortion of the CMB is subtracted away, the mean temperature averaging over the sky is $\langle T \rangle = 2.725K$.

$\boxed{\text{(ii)}}$ After the dipole distortion of the CMB is subtracted away, the root mean square temperature fluctuation is $\left\langle \left(\frac{\delta T}{T}\right)^2 \right\rangle^{1/2} = 1.1 \times 10^{-3}$.

(iii) The dipole distortion is a simple Doppler shift, caused by the net motion of the observer relative to a frame of reference in which the CMB is isotropic.

(iv) In their groundbreaking paper, Wilson and Penzias reported the measurement of an excess temperature of about 3.5 K that was isotropic, unpolarized, and free from seasonal variations. In a companion paper written by Dicke, Peebles, Roll and Wilkinson, the authors interpreted the radiation to be a relic of an early, hot, dense, and opaque state of the universe.

Explanation: The right value is

$$\left\langle \left(\frac{\delta T}{T}\right)^2 \right\rangle^{1/2} = 1.1 \times 10^{-5} \ .$$

(e) *(5 points)* Inflation is driven by a field that is by definition called the *inflaton* field. In standard inflationary models, the field has the following properties:

$\boxed{\text{(i)}}$ The inflaton is a scalar field, and during inflation the energy density of the universe is dominated by its potential energy.

(ii) The inflaton is a vector field, and during inflation the energy density of the universe is dominated by its potential energy.

(iii) The inflaton is a scalar field, and during inflation the energy density of the universe is dominated by its kinetic energy.

(iv) The inflaton is a vector field, and during inflation the energy density of the universe is dominated by its kinetic energy.

(v) The inflaton is a tensor field, which is responsible for only a small fraction of the energy density of the universe during inflation.

Explanation: These facts were mentioned in both Section 11.5 (*The Physics of Inflation)* of Ryden's book, and also in the article that you were asked to read called *Inflation and the New Era of High-Precision Cosmology,* written by me for the Physics Department 2002 newsletter.

PROBLEM 5: NUMBER DENSITIES IN THE COSMIC BACKGROUND RADIATION

In general, the number density of a particle in the black-body radiation is given by

$$n = g^* \frac{\xi(3)}{\pi^2} \left(\frac{kT}{\hbar c}\right)^3$$

For photons, one has $g^* = 2$. Then

$$\left.\begin{aligned} k &= 1.381 \times 10^{-16}\text{erg}/{}^\circ\text{K} \\ T &= 2.7\,{}^\circ\text{K} \\ \hbar &= 1.055 \times 10^{-27}\text{erg-sec} \\ c &= 2.998 \times 10^{10}\text{cm/sec} \end{aligned}\right\} \quad \Longrightarrow \quad \left(\frac{kT}{\hbar c}\right)^3 = 1.638 \times 10^3 \text{cm}^{-3} \ .$$

Then using $\xi(3) \simeq 1.202$, one finds

$$\boxed{n_\gamma = 399/\text{cm}^3 \ .}$$

For the neutrinos,

$$g^*_\nu = 2 \times \frac{3}{4} = \frac{3}{2} \quad \text{per species.}$$

The factor of 2 is to account for ν and $\bar{\nu}$, and the factor of 3/4 arises from the Pauli exclusion principle. So for three species of neutrinos one has

$$g^*_\nu = \frac{9}{2} \ .$$

Using the result

$$T^3_\nu = \frac{4}{11} T^3_\gamma$$

from Problem 8 of Problem Set 3 (2000), one finds

$$\begin{aligned} n_\nu &= \left(\frac{g^*_\nu}{g^*_\gamma}\right)\left(\frac{T_\nu}{T_\gamma}\right)^3 n_\gamma \\ &= \left(\frac{9}{4}\right)\left(\frac{4}{11}\right) 399 \text{cm}^{-3} \\ \Longrightarrow &\quad \boxed{n_\nu = 326/\text{cm}^3 \text{ (for all three species combined).}} \end{aligned}$$

PROBLEM 6: PROPERTIES OF BLACK-BODY RADIATION

(a) The average energy per photon is found by dividing the energy density by the number density. The photon is a boson with two spin states, so $g = g^* = 2$. Using the formulas on the front of the exam,

$$E = \frac{g\dfrac{\pi^2}{30}\dfrac{(kT)^4}{(\hbar c)^3}}{g^*\dfrac{\zeta(3)}{\pi^2}\dfrac{(kT)^3}{(\hbar c)^3}}$$

$$= \boxed{\frac{\pi^4}{30\zeta(3)}\,kT\ .}$$

You were not expected to evaluate this numerically, but it is interesting to know that

$$E = 2.701\,kT\ .$$

Note that the average energy per photon is significantly more than kT, which is often used as a rough estimate.

(b) The method is the same as above, except this time we use the formula for the entropy density:

$$S = \frac{g\dfrac{2\pi^2}{45}\dfrac{k^4T^3}{(\hbar c)^3}}{g^*\dfrac{\zeta(3)}{\pi^2}\dfrac{(kT)^3}{(\hbar c)^3}}$$

$$= \boxed{\frac{2\pi^4}{45\zeta(3)}\,k\ .}$$

Numerically, this gives $3.602\,k$, where k is the Boltzmann constant.

(c) In this case we would have $g = g^* = 1$. The average energy per particle and the average entropy particle depends only on the ratio g/g^*, so there would be $\boxed{\text{no difference from the answers given in parts (a) and (b).}}$

(d) For a fermion, g is $7/8$ times the number of spin states, and g^* is $3/4$ times the number of spin states. So the average energy per particle is

$$E = \frac{g\dfrac{\pi^2}{30}\dfrac{(kT)^4}{(\hbar c)^3}}{g^*\dfrac{\zeta(3)}{\pi^2}\dfrac{(kT)^3}{(\hbar c)^3}}$$

$$= \frac{\dfrac{7}{8}\dfrac{\pi^2}{30}\dfrac{(kT)^4}{(\hbar c)^3}}{\dfrac{3}{4}\dfrac{\zeta(3)}{\pi^2}\dfrac{(kT)^3}{(\hbar c)^3}}$$

$$= \boxed{\frac{7\pi^4}{180\zeta(3)}\,kT\ .}$$

Numerically, $E = 3.1514\,kT$.

Warning: the Mathematician General has determined that the memorization of this number may adversely affect your ability to remember the value of π.

If one takes into account both neutrinos and antineutrinos, the average energy per particle is unaffected — the energy density and the total number density are both doubled, but their ratio is unchanged.

Note that the energy per particle is higher for fermions than it is for bosons. This result can be understood as a natural consequence of the fact that fermions must obey the exclusion principle, while bosons do not. Large numbers of bosons can therefore collect in the lowest energy levels. In fermion systems, on the other hand, the low-lying levels can accommodate at most one particle, and then additional particles are forced to higher energy levels.

(e) The values of g and g^* are again 7/8 and 3/4 respectively, so

$$S = \frac{g\,\dfrac{2\pi^2}{45}\,\dfrac{k^4T^3}{(\hbar c)^3}}{g^*\,\dfrac{\zeta(3)}{\pi^2}\,\dfrac{(kT)^3}{(\hbar c)^3}} = \frac{\dfrac{7}{8}\,\dfrac{2\pi^2}{45}\,\dfrac{k^4T^3}{(\hbar c)^3}}{\dfrac{3}{4}\,\dfrac{\zeta(3)}{\pi^2}\,\dfrac{(kT)^3}{(\hbar c)^3}} = \boxed{\frac{7\pi^4}{135\zeta(3)}\,k\ .}$$

Numerically, this gives $S = 4.202\,k$.

PROBLEM 7: A NEW SPECIES OF LEPTON

a) The number density is given by the formula at the start of the exam,

$$n = g^*\frac{\zeta(3)}{\pi^2}\,\frac{(kT)^3}{(\hbar c)^3}\ .$$

Since the 8.286ion is like the electron, it has $g^* = 3$; there are 2 spin states for the particles and 2 for the antiparticles, giving 4, and then a factor of 3/4 because the particles are fermions. So

$$n = 3\frac{\zeta(3)}{\pi^2}\times\left(\frac{3\ \cancel{\text{MeV}}}{6.582\times10^{-16}\ \cancel{\text{eV}}\text{-}\cancel{\text{sec}}\times 2.998\times10^{10}\ \cancel{\text{cm}}\text{-}\cancel{\text{sec}}^{-1}}\right)^3 \times\left(\frac{10^6\ \cancel{\text{eV}}}{1\ \cancel{\text{MeV}}}\right)^3\times\left(\frac{10^2\ \cancel{\text{cm}}}{1\ \text{m}}\right)^3$$
$$= 3\frac{\zeta(3)}{\pi^2}\times\left(\frac{3\times10^6\times10^2}{6.582\times10^{-16}\times2.998\times10^{10}}\right)^3\ \text{m}^{-3}\ .$$

Then

$$\boxed{\text{Answer} = 3\frac{\zeta(3)}{\pi^2}\times\left(\frac{3\times10^6\times10^2}{6.582\times10^{-16}\times2.998\times10^{10}}\right)^3\ .}$$

You were not asked to evaluate this expression, but the answer is 1.29×10^{39}.

b) For a flat cosmology $\kappa = 0$ and one of the Einstein equations becomes

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3} G\rho \quad .$$

During the radiation-dominated era $a(t) \propto t^{1/2}$, as claimed on the front cover of the exam. So,

$$\frac{\dot{a}}{a} = \frac{1}{2t} \quad .$$

Using this in the above equation gives

$$\frac{1}{4t^2} = \frac{8\pi}{3} G\rho \quad .$$

Solve this for ρ,

$$\rho = \frac{3}{32\pi G t^2} \quad .$$

The question asks the value of ρ at $t = 0.01$ sec. With $G = 6.6732 \times 10^{-8}$ cm^3 sec^{-2} g^{-1}, then

$$\boxed{\rho = \frac{3}{32\pi \times 6.6732 \times 10^{-8} \times (0.01)^2}}$$

in units of g/cm^3. You weren't asked to put the numbers in, but, for reference, doing so gives $\rho = 4.47 \times 10^9$ g/cm^3.

c) The mass density $\rho = u/c^2$, where u is the energy density. The energy density for black-body radiation is given in the exam,

$$u = \rho c^2 = g \frac{\pi^2}{30} \frac{(kT)^4}{(\hbar c)^3} \quad .$$

We can use this information to solve for kT in terms of $\rho(t)$ which we found above in part (b). At a time of 0.01 sec, g has the following contributions:

Photons:	$g = 2$
e^+e^-:	$g = 4 \times \frac{7}{8} = 3\frac{1}{2}$
ν_e, ν_μ, ν_τ:	$g = 6 \times \frac{7}{8} = 5\frac{1}{4}$
8.286ion – anti8.286ion	$g = 4 \times \frac{7}{8} = 3\frac{1}{2}$

$$g_{\rm tot} = 14\frac{1}{4} \quad .$$

Solving for kT in terms of ρ gives

$$kT = \left[\frac{30}{\pi^2}\frac{1}{g_{\rm tot}}\hbar^3 c^5 \rho\right]^{1/4} \quad .$$

Using the result for ρ from part (b) as well as the list of fundamental constants from the cover sheet of the exam gives

$$\boxed{kT = \left[\frac{90 \times (1.055 \times 10^{-27})^3 \times (2.998 \times 10^{10})^5}{14.24 \times 32\pi^3 \times 6.6732 \times 10^{-8} \times (0.01)^2}\right]^{1/4} \times \frac{1}{1.602 \times 10^{-6}}}$$

where the answer is given in units of MeV. Putting in the numbers yields $kT = 8.02$ MeV.

d) The production of helium is increased. At any given temperature, the additional particle increases the energy density. Since $H \propto \rho^{1/2}$, the increased energy density speeds the expansion of the universe— the Hubble constant at any given temperature is higher if the additional particle exists, and the temperature falls faster. The weak interactions that interconvert protons and neutrons “freeze out” when they can no longer keep up with the rate of evolution of the universe. The reaction rates at a given temperature will be unaffected by the additional particle, but the higher value of H will mean that the temperature at which these rates can no longer keep pace with the universe will occur sooner. The freeze-out will therefore occur at a higher temperature. The equilibrium value of the ratio of neutron to proton densities is larger at higher temperatures: $n_n/n_p \propto \exp(-\Delta mc^2/kT)$, where n_n and n_p are the number densities of neutrons and protons, and Δm is the neutron-proton mass difference. Consequently, there are more neutrons present to combine with protons to build helium nuclei. In addition, the faster evolution rate implies that the temperature at which the deuterium bottleneck breaks is reached sooner. This implies that fewer neutrons will have a chance to decay, further increasing the helium production.

e) After the neutrinos decouple, the entropy in the neutrino bath is conserved separately from the entropy in the rest of the radiation bath. Just after neutrino decoupling, all of the particles in equilibrium are described by the same temperature which cools as $T \propto 1/a$. The entropy in the bath of particles still in equilibrium just after the neutrinos decouple is

$$S \propto g_{\rm rest} T^3(t) a^3(t)$$

where $g_{\rm rest} = g_{\rm tot} - g_\nu = 9$. By today, the $e^+ - e^-$ pairs and the 8.286ion-anti8.286ion pairs have annihilated, thus transferring their entropy to the photon bath. As a result the temperature of the photon bath is increased relative to that of the neutrino bath. From conservation of entropy we have that the entropy after annihilations is equal to the entropy before annihilations

$$g_\gamma T_\gamma^3 a^3(t) = g_{\text{rest}} T^3(t) a^3(t) \quad .$$

So,

$$\frac{T_\gamma}{T(t)} = \left(\frac{g_{\text{rest}}}{g_\gamma}\right)^{1/3} \quad .$$

Since the neutrino temperature was equal to the temperature before annihilations, we have that

$$\boxed{\frac{T_\nu}{T_\gamma} = \left(\frac{2}{9}\right)^{1/3} \quad .}$$

PROBLEM 8: A NEW THEORY OF THE WEAK INTERACTIONS *(40 points)*

(a) In the standard model, the black-body radiation at $kT \approx 200$ MeV contains the following contributions:

$$\left.\begin{array}{ll} \text{Photons:} & g = 2 \\ e^+e^-: & g = 4 \times \frac{7}{8} = 3\frac{1}{2} \\ \nu_e, \nu_\mu, \nu_\tau: & g = 6 \times \frac{7}{8} = 5\frac{1}{4} \\ \mu^+\mu^-: & g = 4 \times \frac{7}{8} = 3\frac{1}{2} \\ \pi^+\pi^-\pi^0 & g = 3 \end{array}\right\} g_{\text{TOT}} = 17\frac{1}{4}$$

The mass density is then given by

$$\rho = \frac{u}{c^2} = g_{\text{TOT}} \frac{\pi^2}{30} \frac{(kT)^4}{\hbar^3 c^5} \quad .$$

In kg/m^3, one can evaluate this expression by

$$\rho = \left(17\frac{1}{4}\right) \frac{\pi^2}{30} \frac{\left[200 \times 10^6 \,\text{eV} \times \dfrac{1.602 \times 10^{-19}\,\text{J}}{\text{eV}}\right]^4}{\left(1.055 \times 10^{-34}\,\text{J-s}\right)^3 \left(2.998 \times 10^8\,\text{m/s}\right)^5} \quad .$$

Checking the units,

$$[\rho] = \frac{\text{J}^4}{\text{J}^3\text{-s}^3\text{-m}^5\text{-s}^{-5}} = \frac{\text{J-s}^2}{\text{m}^5}$$
$$= \frac{\left(\text{kg-m}^2\text{-s}^{-2}\right)\text{s}^2}{\text{m}^5} = \text{kg/m}^3 \ .$$

So, the final answer would be

$$\boxed{\rho = \left(17\frac{1}{4}\right)\frac{\pi^2}{30}\frac{\left[200\times 10^6 \times 1.602\times 10^{-19}\right]^4}{\left(1.055\times 10^{-34}\right)^3\left(2.998\times 10^8\right)^5}\frac{\text{kg}}{\text{m}^3} \ .}$$

You were not expected to evaluate this, but with a calculator one would find

$$\rho = 2.10\times 10^{18}\ \text{kg/m}^3 \ .$$

In g/cm^3, one would evaluate this expression by

$$\rho = \left(17\frac{1}{4}\right)\frac{\pi^2}{30}\frac{\left[200\times 10^6\,\text{eV}\times\dfrac{1.602\times 10^{-12}\,\text{erg}}{\text{eV}}\right]^4}{\left(1.055\times 10^{-27}\,\text{erg-s}\right)^3\left(2.998\times 10^{10}\,\text{cm/s}\right)^5} \ .$$

Checking the units,

$$[\rho] = \frac{\text{erg}^4}{\text{erg}^3\text{-s}^3\text{-cm}^5\text{-s}^{-5}} = \frac{\text{erg-s}^2}{\text{cm}^5}$$
$$= \frac{\left(\text{g-cm}^2\text{-s}^{-2}\right)\text{s}^2}{\text{cm}^5} = \text{g/cm}^3 \ .$$

So, in this case the final answer would be

$$\boxed{\rho = \left(17\frac{1}{4}\right)\frac{\pi^2}{30}\frac{\left[200\times 10^6 \times 1.602\times 10^{-12}\right]^4}{\left(1.055\times 10^{-27}\right)^3\left(2.998\times 10^{10}\right)^5}\frac{\text{g}}{\text{cm}^3} \ .}$$

No evaluation was requested, but with a calculator you would find

$$\rho = 2.10\times 10^{15}\ \text{g/cm}^3 \ ,$$

which agrees with the answer above.

Note: A common mistake was to leave out the conversion factor 1.602×10^{-19} J/eV (or 1.602×10^{-12} erg/eV), and instead to use $\hbar = 6.582 \times 10^{-16}$ eV-s. But if one works out the units of this answer, they turn out to be eV-sec^2/m^5 (or eV-sec^2/cm^5), which is a most peculiar set of units to measure a mass density.

In the NTWI, we have in addition the contribution to the mass density from R^+-R^- pairs, which would act just like e^+-e^- pairs or μ^+-μ^- pairs, with $g = 3\frac{1}{2}$. Thus $g_{\text{TOT}} = 20\frac{3}{4}$, so

$$\boxed{\rho = \left(20\frac{3}{4}\right)\frac{\pi^2}{30}\frac{\left[200 \times 10^6 \times 1.602 \times 10^{-19}\right]^4}{\left(1.055 \times 10^{-34}\right)^3\left(2.998 \times 10^8\right)^5}\frac{\text{kg}}{\text{m}^3}}$$

or

$$\boxed{\rho = \left(20\frac{3}{4}\right)\frac{\pi^2}{30}\frac{\left[200 \times 10^6 \times 1.602 \times 10^{-12}\right]^4}{\left(1.055 \times 10^{-27}\right)^3\left(2.998 \times 10^{10}\right)^5}\frac{\text{g}}{\text{cm}^3}\ .}$$

Numerically, the answer in this case would be

$$\rho_{\text{NTWI}} = 2.53 \times 10^{18}\ \text{kg/m}^3 = 2.53 \times 10^{15}\ \text{g/cm}^3\ .$$

(b) As long as the universe is in thermal equilibrium, entropy is conserved. The entropy in a given volume of the comoving coordinate system is

$$a^3(t)s\,V_{\text{coord}}\quad ,$$

where s is the entropy density and $a^3 V_{\text{coord}}$ is the physical volume. So

$$a^3(t)s$$

is conserved. After the neutrinos decouple,

$$a^3 s_\nu \quad \text{and} \quad a^3 s_{\text{other}}$$

are separately conserved, where s_{other} is the entropy of everything except neutrinos.

Note that s can be written as

$$s = gAT^3\quad ,$$

where A is a constant. Before the disappearance of the e, μ, R, and π particles from the thermal equilibrium radiation,

$$\begin{aligned} s_\nu &= \left(5\frac{1}{4}\right) AT^3 \\ s_{\text{other}} &= \left(15\frac{1}{2}\right) AT^3 \quad . \end{aligned}$$

So

$$\frac{s_\nu}{s_{\text{other}}} = \frac{5\frac{1}{4}}{15\frac{1}{2}} \quad .$$

If $a^3 s_\nu$ and $a^3 s_{\text{other}}$ are conserved, then so is s_ν / s_{other}. By today, the entropy previously shared among the various particles still in equilibrium after neutrino decoupling has been transfered to the photons so that

$$s_{\text{other}} = s_{\text{photons}} = 2AT_\gamma^3 \quad .$$

The entropy in neutrinos is still

$$s_\nu = \left(5\frac{1}{4}\right) AT_\nu^3 \quad .$$

Since s_ν / s_{other} is constant we know that

$$\frac{\left(5\frac{1}{4}\right) T_\nu^3}{2T_\gamma^3} = \frac{s_\nu}{s_{\text{other}}} = \frac{5\frac{1}{4}}{15\frac{1}{2}}$$

$$\Longrightarrow \quad \boxed{T_\nu = \left(\frac{4}{31}\right)^{1/3} T_\gamma} \quad .$$

(c) One can write

$$n = g^* BT^3 \quad ,$$

where B is a constant. Here $g_\gamma^* = 2$, and $g_\nu^* = 6 \times \frac{3}{4} = 4\frac{1}{2}$. In the standard model, one has today

$$\frac{n_\nu}{n_\gamma} = \frac{g_\nu^* T_\nu^3}{g_\gamma^* T_\gamma^3} = \frac{\left(4\frac{1}{2}\right)}{2} \frac{4}{11} = \boxed{\frac{9}{11}} \quad .$$

In the NTWI,

$$\frac{n_\nu}{n_\gamma} = \frac{\left(4\frac{1}{2}\right)}{2}\,\frac{4}{31} = \boxed{\frac{9}{31}\,.}$$

(d) At $kT = 200$ MeV, the thermal equilibrium ratio of neutrons to protons is given by

$$\frac{n_\mathrm{n}}{n_\mathrm{p}} = e^{-1.29\,\mathrm{MeV}/200\,\mathrm{MeV}} \approx 1 \quad .$$

In the standard theory this ratio would decrease rapidly as the universe cooled and kT fell below the p-n mass difference of 1.29 MeV, but in the NTWI the ratio freezes out at the high temperature corresponding to $kT = 200$ MeV, when the ratio is about 1. When kT falls below 200 MeV in the NTWI, the neutrino interactions

$$n + \nu_e \leftrightarrow p + e^- \qquad \text{and} \qquad n + e^+ \leftrightarrow p + \bar{\nu}_e$$

that maintain the thermal equilibrium balance between protons and neutrons no longer occur at a significant rate, so the ratio $n_/n_p$ is no longer controlled by thermal equilibrium. After kT falls below 200 MeV, the only process that can convert neutrons to protons is the rather slow process of free neutron decay, with a decay time τ_d of about 890 s. Thus, when the deuterium bottleneck breaks at about 200 s, the number density of neutrons will be considerably higher than in the standard model. Since essentially all of these neutrons will become bound into He nuclei, the higher neutron abundance of the NTWI implies a

$$\boxed{\text{higher predicted He abundance.}}$$

To estimate the He abundance, note that if we temporarily ignore free neutron decay, then the neutron-proton ratio would be frozen at about 1 and would remain 1 until the time of nucleosynthesis. At the time of nucleosynthesis essentially all of these neutrons would be bound into He nuclei (each with 2 protons and 2 neutrons). For an initial 1:1 ratio of neutrons to protons, all the neutrons and protons can be bound into He nuclei, with no protons left over in the form of hydrogen, so Y would equal 1. However, the free neutron decay process will cause the ratio n_n/n_p to fall below 1 before the start of nucleosynthesis, so the predicted value of Y would be less than 1.

To calculate how much less, note that Ryden estimates the start of nucleosynthesis at the time when the temperature reaches T_nuc, which is the temperature for which a thermal equilibrium calculation gives $n_D/n_n = 1$. This corresponds to what Weinberg refers to as the breaking of the deuterium bottleneck. The temperature

$T_{\rm nuc}$ is calculated in terms of $\eta = n_B/n_\gamma$ and physical constants, so it would not be changed by the NTWI. The time when this temperature is reached, however, would be changed slightly by the change in the ratio T_ν/T_γ. Since this effect is rather subtle, no points will be taken off if you omitted it. However, to be as accurate as possible, one should recognize that nucleosynthesis occurs during the radiation-dominated era, but long after the e^+-e^- pairs have disappeared, so the black-body radiation consists of photons at temperature T_γ and neutrinos at a lower temperature T_ν. The energy density is given by

$$u = \frac{\pi^2}{30}\frac{(kT_\gamma)^4}{(\hbar c)^3}\left[2 + \left(\frac{21}{4}\right)\left(\frac{T_\nu}{T_\gamma}\right)^4\right] \equiv g_{\rm eff}\frac{\pi^2}{30}\frac{(kT_\gamma)^4}{(\hbar c)^3} \ ,$$

where

$$g_{\rm eff} = 2 + \left(\frac{21}{4}\right)\left(\frac{T_\nu}{T_\gamma}\right)^4 \ .$$

For the standard model

$$g_{\rm eff}^{\rm sm} = 2 + \left(\frac{21}{4}\right)\left(\frac{4}{11}\right)^{4/3} \ ,$$

and for the NTWI

$$g_{\rm eff}^{\rm NTWI} = 2 + \left(\frac{21}{4}\right)\left(\frac{4}{31}\right)^{4/3} \ .$$

The relation between time and temperature in a flat radiation-dominated universe is given in the formula sheets as

$$kT = \left(\frac{45\hbar^3 c^5}{16\pi^3 g G}\right)^{1/4}\frac{1}{\sqrt{t}} \ .$$

Thus,

$$t \propto \frac{1}{g_{\rm eff}^{1/2} T^2} \ .$$

In the standard model Ryden estimates the time of nucleosynthesis as $t_{\rm nuc}^{\rm sm} \approx 200$ s, so in the NTWI it would be longer by the factor

$$t_{\rm nuc}^{\rm NTWI} = \sqrt{\frac{g_{\rm eff}^{\rm sm}}{g_{\rm eff}^{\rm NTWI}}}\, t_{\rm nuc}^{\rm sm} \ .$$

While of coure you were not expected to work out the numerics, this gives

$$t_{\rm nuc}^{\rm NTWI} = 1.20\, t_{\rm nuc}^{\rm sm} \ .$$

Note that Ryden gives $t_{\rm nuc} \approx 200s$, while Weinberg places it at $3\frac{3}{4}$ minutes ≈ 225 s, which is close enough.

To follow the effect of this free decay, it is easiest to do it by considering the ratio neutrons to baryon number, n_n/n_B, since n_B does not change during this period. At freeze-out, when $kT \approx 200$ MeV,

$$\frac{n_n}{n_B} \approx \frac{1}{2} \ .$$

Just before nucleosynthesis, at time $t_{\rm nuc}$, the ratio will be

$$\frac{n_n}{n_B} \approx \frac{1}{2} e^{-t_{\rm nuc}/\tau_d} \ .$$

If free decay is ignored, we found $Y = 1$. Since all the surviving neutrons are bound into He, the corrected value of Y is simply deceased by multiplying by the fraction of neutrons that do not undergo decay. Thus, the prediction of NTWI is

$$\boxed{Y = e^{-t_{\rm nuc}/\tau_d} = \exp\left\{ - \frac{\sqrt{\frac{g_{\rm eff}^{\rm sm}}{g_{\rm eff}^{\rm NTWI}}}\, 200}{890} \right\} \ ,}$$

where $g_{\rm eff}^{\rm sm}$ and $g_{\rm eff}^{\rm NTWI}$ are given above. When evaluated numerically, this would give

$$Y = \text{Predicted He abundance by weight} \approx 0.76 \ .$$

PROBLEM 9: DOUBLING OF ELECTRONS *(10 points)*

The entropy density of black-body radiation is given by

$$\begin{aligned} s &= g \left[\frac{2\pi^2}{45} \frac{k^4}{(\hbar c)^3} \right] T^3 \\ &= g\, C\, T^3 \ , \end{aligned}$$

where C is a constant. At the time when the electron-positron pairs disappear, the neutrinos are decoupled, so their entropy is conserved. All of the entropy from electron-positron pairs is given to the photons, and none to the neutrinos. The same will be true here, for both species of electron-positron pairs.

The conserved neutrino entropy can be described by $S_\nu \equiv a^3 s_\nu$, which indicates the entropy per cubic notch, i.e., entropy per unit comoving volume. We introduce the notation n^- and n^+ for the new electron-like and positron-like particles, and also the convention that

Primed quantities:	values after $e^+e^-n^+n^-$ annihilation
Unprimed quantities:	values before $e^+e^-n^+n^-$ annihilation.

For the neutrinos,

$$S'_\nu = S_\nu \quad \Longrightarrow \quad g_\nu C\left(a'T'_\nu\right)^3 = g_\nu C\left(aT_\nu\right)^3 \quad \Longrightarrow$$

$$\boxed{a'T'_\nu = aT_\nu\ .}$$

For the photons, before $e^+e^-n^+n^-$ annihilation we have

$$T_\gamma = T_{e^+e^-n^+n^-} = T_\nu\ ; \qquad g_\gamma = 2,\ g_{e^+e^-} = g_{n^+n^-} = 7/2\ .$$

When the e^+e^- and n^+n^- pairs annihilate, their entropy is added to the photons:

$$S'_\gamma = S_{e^+e^-} + S_{n^+n^-} + S_\gamma \quad \Longrightarrow \quad 2C\left(a'T'_\gamma\right)^3 = \left(2 + 2\cdot\frac{7}{2}\right)C\left(aT_\gamma\right)^3 \quad \Longrightarrow$$

$$\boxed{a'T'_\gamma = \left(\frac{9}{2}\right)^{1/3} aT_\gamma\ ,}$$

so aT_γ increases by a factor of $(9/2)^{1/3}$.

Before e^+e^- annihilation the neutrinos were in thermal equilibrium with the photons, so $T_\gamma = T_\nu$. By considering the two boxed equations above, one has

$$\boxed{T'_\nu = \left(\frac{2}{9}\right)^{1/3} T'_\gamma\ .}$$

This ratio would remain unchanged until the present day.

PROBLEM 10: TIME SCALES IN COSMOLOGY

(a) 1 sec. *[This is the time at which the weak interactions begin to "freeze out", so that free neutron decay becomes the only mechanism that can interchange protons and neutrons. From this time onward, the relative number of protons and neutrons is no longer controlled by thermal equilibrium considerations.]*

(b) 4 mins. *[By this time the universe has become so cool that nuclear reactions are no longer initiated.]*

(c) 10^{-37} sec. *[We learned in Lecture Notes 7 that kT was about 1 MeV at $t = 1$ sec. Since 1 GeV = 1000 MeV, the value of kT that we want is 10^{19} times higher. In the radiation-dominated era $T \propto a^{-1} \propto t^{-1/2}$, so we get 10^{-38} sec.]*

(d) 10,000 – 1,000,000 years. *[This number was estimated in Lecture Notes 7 as 200,000 years.]*

(e) 10^{-5} sec. *[As in (c), we can use $t \propto T^{-2}$, with $kT \approx 1$ MeV at $t = 1$ sec.]*

PROBLEM 11: EVOLUTION OF FLATNESS *(15 points)*

(a) We start with the Friedmann equation from the formula sheet on the quiz:

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

The critical density is the value of ρ corresponding to $k = 0$, so

$$H^2 = \frac{8\pi}{3}G\rho_c .$$

Using this expression to replace H^2 on the left-hand side of the Friedmann equation, and then dividing by $8\pi G/3$, one finds

$$\rho_c = \rho - \frac{3kc^2}{8\pi G a^2} .$$

Rearranging,

$$\frac{\rho - \rho_c}{\rho} = \frac{3kc^2}{8\pi G a^2 \rho} .$$

On the left-hand side we can divide the numerator and denominator by ρ_c, and then use the definition $\Omega \equiv \rho/\rho_c$ to obtain

$$\frac{\Omega - 1}{\Omega} = \frac{3kc^2}{8\pi G a^2 \rho} . \tag{1}$$

For a matter-dominated universe we know that $\rho \propto 1/a^3(t)$, and so

$$\frac{\Omega - 1}{\Omega} \propto a(t) \; .$$

If the universe is nearly flat we know that $a(t) \propto t^{2/3}$, so

$$\boxed{\frac{\Omega - 1}{\Omega} \propto t^{2/3} \; .}$$

(b) Eq. (1) above is still true, so our only task is to re-evaluate the right-hand side. For a radiation-dominated universe we know that $\rho \propto 1/a^4(t)$, so

$$\frac{\Omega - 1}{\Omega} \propto a^2(t) \; .$$

If the universe is nearly flat then $a(t) \propto t^{1/2}$, so

$$\boxed{\frac{\Omega - 1}{\Omega} \propto t \; .}$$

PROBLEM 12: THE SLOAN DIGITAL SKY SURVEY $z = 5.82$ QUASAR

(40 points)

(a) Since $\Omega_m + \Omega_\Lambda = 0.35 + 0.65 = 1$, the universe is flat. It therefore obeys a simple form of the Friedmann equation,

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3} G(\rho_m + \rho_\Lambda) \; ,$$

where the overdot indicates a derivative with respect to t, and the term proportional to k has been dropped. Using the fact that $\rho_m \propto 1/a^3(t)$ and $\rho_\Lambda = \text{const}$, the energy densities on the right-hand side can be expressed in terms of their present values $\rho_{m,0}$ and $\rho_\Lambda \equiv \rho_{\Lambda,0}$. Defining

$$x(t) \equiv \frac{a(t)}{a(t_0)} \; ,$$

one has

$$\begin{aligned}
\left(\frac{\dot{x}}{x}\right)^2 &= \frac{8\pi}{3} G\left(\frac{\rho_{m,0}}{x^3} + \rho_\Lambda\right) \\
&= \frac{8\pi}{3} G\rho_{c,0}\left(\frac{\Omega_{m,0}}{x^3} + \Omega_{\Lambda,0}\right) \\
&= H_0^2\left(\frac{\Omega_{m,0}}{x^3} + \Omega_{\Lambda,0}\right) \; .
\end{aligned}$$

Here we used the facts that

$$\Omega_{m,0} \equiv \frac{\rho_{m,0}}{\rho_{c,0}}; \qquad \Omega_{\Lambda,0} \equiv \frac{\rho_\Lambda}{\rho_{c,0}} ,$$

and

$$H_0^2 = \frac{8\pi}{3} G\rho_{c,0} .$$

The equation above for $(\dot{x}/x)^2$ implies that

$$\dot{x} = H_0\, x \sqrt{\frac{\Omega_{m,0}}{x^3} + \Omega_{\Lambda,0}} ,$$

which in turn implies that

$$\mathrm{d}t = \frac{1}{H_0} \frac{\mathrm{d}x}{x\sqrt{\frac{\Omega_{m,0}}{x^3} + \Omega_{\Lambda,0}}} .$$

Using the fact that x changes from 0 to 1 over the life of the universe, this relation can be integrated to give

$$t_0 = \int_0^{t_0} \mathrm{d}t = \boxed{\frac{1}{H_0} \int_0^1 \frac{\mathrm{d}x}{x\sqrt{\frac{\Omega_{m,0}}{x^3} + \Omega_{\Lambda,0}}}} .$$

The answer can also be written as

$$\boxed{t_0 = \frac{1}{H_0} \int_0^1 \frac{x\, \mathrm{d}x}{\sqrt{\Omega_{m,0} x + \Omega_{\Lambda,0} x^4}}}$$

or

$$\boxed{t_0 = \frac{1}{H_0} \int_0^\infty \frac{\mathrm{d}z}{(1+z)\sqrt{\Omega_{m,0}(1+z)^3 + \Omega_{\Lambda,0}}}} ,$$

where in the last answer I changed the variable of integration using

$$x = \frac{1}{1+z} ; \qquad \mathrm{d}x = -\frac{\mathrm{d}z}{(1+z)^2} .$$

Note that the minus sign in the expression for $\mathrm{d}x$ is canceled by the interchange of the limits of integration: $x = 0$ corresponds to $z = \infty$, and $x = 1$ corresponds to $z = 0$.

Your answer should look like one of the above boxed answers. You were not expected to complete the numerical calculation, but for pedagogical purposes I will continue. The integral can actually be carried out analytically, giving

$$\int_0^1 \frac{x\,\mathrm{d}x}{\sqrt{\Omega_{m,0}x + \Omega_{\Lambda,0}x^4}} = \frac{2}{3\sqrt{\Omega_{\Lambda,0}}} \ln\left(\frac{\sqrt{\Omega_m + \Omega_{\Lambda,0}} + \sqrt{\Omega_{\Lambda,0}}}{\sqrt{\Omega_m}}\right) .$$

Using

$$\frac{1}{H_0} = \frac{9.778 \times 10^9}{h_0} \text{ yr} ,$$

where $H_0 = 100\,h_0$ km-sec^{-1}-Mpc^{-1}, one finds for $h_0 = 0.65$ that

$$\frac{1}{H_0} = 15.043 \times 10^9 \text{ yr} .$$

Then using $\Omega_m = 0.35$ and $\Omega_{\Lambda,0} = 0.65$, one finds

$$\boxed{t_0 = 13.88 \times 10^9 \text{ yr} .}$$

So the SDSS people were right on target.

(b) Having done part (a), this part is very easy. The dynamics of the universe is of course the same, and the question is only slightly different. In part (a) we found the amount of time that it took for x to change from 0 to 1. The light from the quasar that we now receive was emitted when

$$x = \frac{1}{1+z} ,$$

since the cosmological redshift is given by

$$1 + z = \frac{a(t_{\text{observed}})}{a(t_{\text{emitted}})} .$$

Using the expression for $\mathrm{d}t$ from part (a), the amount of time that it took the universe to expand from $x = 0$ to $x = 1/(1+z)$ is given by

$$t_e = \int_0^{t_e} \mathrm{d}t = \boxed{\frac{1}{H_0}\int_0^{1/(1+z)} \frac{\mathrm{d}x}{x\sqrt{\frac{\Omega_{m,0}}{x^3} + \Omega_{\Lambda,0}}} .}$$

Again one could write the answer other ways, including

$$\boxed{t_0 = \frac{1}{H_0}\int_z^\infty \frac{\mathrm{d}z'}{(1+z')\sqrt{\Omega_{m,0}(1+z')^3+\Omega_{\Lambda,0}}} \ .}$$

Again you were expected to stop with an expression like the one above. Continuing, however, the integral can again be done analytically:

$$\int_0^{x_{\max}} \frac{\mathrm{d}x}{x\sqrt{\frac{\Omega_{m,0}}{x^3}+\Omega_{\Lambda,0}}} = \frac{2}{3\sqrt{\Omega_{\Lambda,0}}}\ln\left(\frac{\sqrt{\Omega_m+\Omega_{\Lambda,0}x_{\max}^3}+\sqrt{\Omega_{\Lambda,0}}\,x_{\max}^{3/2}}{\sqrt{\Omega_m}}\right) \ .$$

Using $x_{\max} = 1/(1+5.82) = .1466$ and the other values as before, one finds

$$\boxed{t_e = \frac{0.06321}{H_0} = 0.9509\times 10^9 \text{ yr} \ .}$$

So again the SDSS people were right.

(c) To find the physical distance to the quasar, we need to figure out how far light can travel from $z = 5.82$ to the present. Since we want the present distance, we multiply the coordinate distance by $a(t_0)$. For the flat metric

$$\mathrm{d}s^2 = -c^2\,\mathrm{d}\tau^2 = -c^2\mathrm{d}t^2 + a^2(t)\left\{\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2+\sin^2\theta\,\mathrm{d}\phi^2)\right\} \ ,$$

the coordinate velocity of light (in the radial direction) is found by setting $\mathrm{d}s^2 = 0$, giving

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \frac{c}{a(t)} \ .$$

So the total coordinate distance that light can travel from t_e to t_0 is

$$\ell_c = \int_{t_e}^{t_0} \frac{c}{a(t)}\,\mathrm{d}t \ .$$

This is not the final answer, however, because we don't explicitly know $a(t)$. We can, however, change variables of integration from t to x, using

$$\mathrm{d}t = \frac{\mathrm{d}t}{\mathrm{d}x}\,\mathrm{d}x = \frac{\mathrm{d}x}{\dot{x}} \ .$$

So

$$\ell_c = \frac{c}{a(t_0)} \int_{x_e}^{1} \frac{\mathrm{d}x}{x\,\dot{x}} \ ,$$

where x_e is the value of x at the time of emission, so $x_e = 1/(1+z)$. Using the equation for $\dot{x}$ from part (a), this integral can be rewritten as

$$\ell_c = \frac{c}{H_0 a(t_0)} \int_{1/(1+z)}^{1} \frac{\mathrm{d}x}{x^2 \sqrt{\frac{\Omega_{m,0}}{x^3} + \Omega_{\Lambda,0}}} \ .$$

Finally, then

$$\boxed{\ell_{\text{phys},0} = a(t_0)\,\ell_c = \frac{c}{H_0} \int_{1/(1+z)}^{1} \frac{\mathrm{d}x}{x^2 \sqrt{\frac{\Omega_{m,0}}{x^3} + \Omega_{\Lambda,0}}} \ .}$$

Alternatively, this result can be written as

$$\boxed{\ell_{\text{phys},0} = \frac{c}{H_0} \int_{1/(1+z)}^{1} \frac{\mathrm{d}x}{\sqrt{\Omega_{m,0}\, x + \Omega_{\Lambda,0}\, x^4}} \ ,}$$

or by changing variables of integration to obtain

$$\boxed{\ell_{\text{phys},0} = \frac{c}{H_0} \int_{0}^{z} \frac{\mathrm{d}z'}{\sqrt{\Omega_{m,0}\,(1+z')^3 + \Omega_{\Lambda,0}}} \ .}$$

Continuing for pedagogical purposes, this time the integral has no analytic form, so far as I know. Integrating numerically,

$$\int_0^{5.82} \frac{\mathrm{d}z'}{\sqrt{0.35\,(1+z')^3 + 0.65}} = 1.8099 \ ,$$

and then using the value of $1/H_0$ from part (a),

$$\boxed{\ell_{\text{phys},0} = 27.23 \text{ light-yr} \ .}$$

Right again.

(d) $\ell_{\text{phys},e} = a(t_e)\ell_c$, so

$$\ell_{\text{phys},e} = \frac{a(t_e)}{a(t_0)}\,\ell_{\text{phys},0} = \boxed{\frac{\ell_{\text{phys},0}}{1+z}}\ .$$

Numerically this gives

$$\boxed{\ell_{\text{phys},e} = 3.992 \times 10^9 \text{ light-yr}\ .}$$

The SDSS announcement is still okay.

(e) The speed defined in this way obeys the Hubble law exactly, so

$$v = H_0\,\ell_{\text{phys},0} = c\int_0^z \frac{\mathrm{d}z'}{\sqrt{\Omega_{m,0}\,(1+z')^3 + \Omega_{\Lambda,0}}}\ .$$

Then

$$\boxed{\frac{v}{c} = \int_0^z \frac{\mathrm{d}z'}{\sqrt{\Omega_{m,0}\,(1+z')^3 + \Omega_{\Lambda,0}}}\ .}$$

Numerically, we have already found that this integral has the value

$$\boxed{\frac{v}{c} = 1.8099\ .}$$

The SDSS people get an A.

PROBLEM 13: SECOND HUBBLE CROSSING *(40 points)*

(a) From the formula sheets, we know that for a flat radiation-dominated universe,

$$a(t) \propto t^{1/2}\ .$$

Since

$$H = \frac{\dot a}{a}\ ,$$

(which is also on the formula sheets),

$$H = \frac{1}{2t} \; .$$

Then

$$\ell_H(t) \equiv cH^{-1}(t) = \boxed{2ct \; .}$$

(b) We are told that the energy density is dominated by photons and neutrinos, so we need to add together these two contributions to the energy density. For photons, the formula sheet reminds us that $g_\gamma = 2$, so

$$u_\gamma = 2\,\frac{\pi^2}{30}\,\frac{(kT_\gamma)^4}{(\hbar c)^3} \; .$$

For neutrinos the formula sheet reminds us that

$$g_\nu = \underbrace{\frac{7}{8}}_{\substack{\text{Fermion}\\ \text{factor}}} \times \underbrace{3}_{\substack{\text{3 species}\\ \nu_e,\nu_\mu,\nu_\tau}} \times \underbrace{2}_{\substack{\text{Particle/}\\ \text{antiparticle}}} \times \underbrace{1}_{\text{Spin states}} = \frac{21}{4} \; ,$$

so

$$u_\nu = \frac{21}{4}\,\frac{\pi^2}{30}\,\frac{(kT_\nu)^4}{(\hbar c)^3} \; .$$

Combining these two expressions and using $T_\nu = (4/11)^{1/3}\,T_\gamma$, one has

$$u = u_\gamma + u_\nu = \left[2 + \frac{21}{4}\left(\frac{4}{11}\right)^{4/3}\right]\frac{\pi^2}{30}\,\frac{(kT_\gamma)^4}{(\hbar c)^3} \; ,$$

so finally

$$\boxed{g_1 = 2 + \frac{21}{4}\left(\frac{4}{11}\right)^{4/3} \; .}$$

(c) The Friedmann equation tells us that, for a flat universe,

$$H^2 = \frac{8\pi}{3}G\rho \; ,$$

where in this case $H = 1/(2t)$ and

$$\rho = \frac{u}{c^2} = g_1\,\frac{\pi^2}{30}\,\frac{(kT_\gamma)^4}{\hbar^3 c^5} \; .$$

Thus

$$\left(\frac{1}{2t}\right)^2 = \frac{8\pi G}{3}\, g_1\, \frac{\pi^2}{30}\, \frac{(kT_\gamma)^4}{\hbar^3 c^5}\ .$$

Solving for T_γ,

$$\boxed{T_\gamma = \frac{1}{k}\left(\frac{45\hbar^3 c^5}{16\pi^3 g_1 G}\right)^{1/4} \frac{1}{\sqrt{t}}\ .}$$

(d) The condition for Hubble crossing is

$$\lambda(t) = cH^{-1}(t)\ ,$$

and the first Hubble crossing always occurs during the inflationary era. Thus any Hubble crossing during the radiation-dominated era must be the second Hubble crossing.

If λ is the present physical wavelength of the density perturbations under discussion, the wavelength at time t is scaled by the scale factor $a(t)$:

$$\lambda(t) = \frac{a(t)}{a(t_0)}\,\lambda\ .$$

Between the second Hubble crossing and now, there have been no freeze-outs of particle species. Today the entropy of the universe is still dominated by photons and neutrinos, so the conservation of entropy implies that aT_γ has remained essentially constant between then and now. Thus,

$$\lambda(t) = \frac{T_{\gamma,0}}{T_\gamma(t)}\,\lambda\ .$$

Using the previous results for $cH^{-1}(t)$ and for $T_\gamma(t)$, the condition $\lambda(t) = cH^{-1}(t)$ can be rewritten as

$$kT_{\gamma,0}\left(\frac{16\pi^3 g_1 G}{45\hbar^3 c^5}\right)^{1/4} \sqrt{t}\,\lambda = 2ct\ .$$

Solving for t, the time of second Hubble crossing is found to be

$$\boxed{t_{H2}(\lambda) = (kT_{\gamma,0}\lambda)^2\left(\frac{\pi^3 g_1 G}{45\hbar^3 c^9}\right)^{1/2}\ .}$$

Extension: You were not asked to insert numbers, but it is of course interesting to know where the above formula leads. If we take $\lambda = 10^6$ lt-yr, it gives

$$t_{H2}(10^6 \text{ lt-yr}) = 1.04 \times 10^7 \text{ s} = 0.330 \text{ year} .$$

For $\lambda = 1$ Mpc,

$$t_{H2}(1 \text{ Mpc}) = 1.11 \times 10^8 \text{ s} = 3.51 \text{ year} .$$

Taking $\lambda = 2.5 \times 10^6$ lt-yr, the distance to Andromeda, the nearest spiral galaxy,

$$t_{H2}(2.5 \times 10^6 \text{ lt-yr}) = 6.50 \times 10^7 \text{ sec} = 2.06 \text{ year} .$$

PROBLEM 14: NEUTRINO NUMBER AND THE NEUTRON/PROTON EQUILIBRIUM

(a) From the chemical equilibrium equation on the front of the exam, the number densities of neutrons and protons can be written as

$$n_n = g_n \frac{(2\pi m_n kT)^{3/2}}{(2\pi\hbar)^3} e^{(\mu_n - m_n c^2)/kT}$$

$$n_p = g_p \frac{(2\pi m_p kT)^{3/2}}{(2\pi\hbar)^3} e^{(\mu_p - m_p c^2)/kT} ,$$

where $g_n = g_p = 2$. Dividing,

$$\frac{n_n}{n_p} = \left(\frac{m_n}{m_p}\right)^{3/2} e^{-(\Delta E + \mu_p - \mu_n)/kT} ,$$

where $\Delta E = (m_n - m_p)c^2$ is the proton-neutron mass-energy difference. Approximating $m_n/m_p \approx 1$, one has

$$\boxed{\frac{n_n}{n_p} = e^{-(\Delta E + \mu_p - \mu_n)/kT} .}$$

The approximation $m_n/m_p \approx 1$ is very accurate (0.14%), but is clearly not necessary. Full credit was given whether or not this approximation was used.

(b) For any allowed chemical reaction, the sum of the chemical potentials on the two sides must be equal. So, from

$$e^+ + n \longleftrightarrow p + \bar{\nu}_e ,$$

we can infer that

$$-\mu_e + \mu_n = \mu_p - \mu_\nu \ ,$$

which implies that

$$\boxed{\mu_n - \mu_p = \mu_e - \mu_\nu \ .}$$

(c) Applying the formula given in the problem to the number densities of electron neutrinos and the corresponding antineutrinos,

$$n_\nu = g_\nu^* \frac{\zeta(3)}{\pi^2} \frac{(kT)^3}{(\hbar c)^3} e^{\mu_\nu/kT}$$

$$\bar{n}_\nu = g_\nu^* \frac{\zeta(3)}{\pi^2} \frac{(kT)^3}{(\hbar c)^3} e^{-\mu_\nu/kT} \ ,$$

since the chemical potential for the antineutrinos ($\bar{\nu}$) is the negative of the chemical potential for neutrinos. A neutrino has only one spin state, so $g_\nu = 3/4$, where the factor of $3/4$ arises because neutrinos are fermions. Setting

$$x \equiv e^{-\mu_\nu/kT}$$

and

$$A \equiv \frac{3}{4} \frac{\zeta(3)}{\pi^2} \frac{(kT)^3}{(\hbar c)^3} \ ,$$

the number density equations can be written compactly as

$$n_\nu = \frac{A}{x} \ , \qquad \bar{n}_\nu = xA \ .$$

To express x in terms of the ratio $\bar{n}_\nu/n_\nu$, divide the second equation by the first to obtain

$$\frac{\bar{n}_\nu}{n_\nu} = x^2 \quad \Longrightarrow \quad x = \sqrt{\frac{\bar{n}_\nu}{n_\nu}} \ .$$

Alternatively, x can be expressed in terms of the difference in number densities $\bar{n}_\nu - n_\nu$ by starting with

$$\Delta n = \bar{n}_\nu - n_\nu = xA - \frac{A}{x} \ .$$

Rewriting the above formula as an explicit quadratic,

$$Ax^2 - \Delta n \, x - A = 0 \ ,$$

one finds

$$x = \frac{\Delta n \pm \sqrt{\Delta n^2 + 4A^2}}{2A} \ .$$

Since the definition of x implies $x > 0$, only the positive root is relevant. Since the number of electrons is still assumed to be equal to the number of positrons, $\mu_e = 0$, so the answer to (b) reduces to $\mu_n - \mu_p = -\mu_\nu$. From (a),

$$\begin{aligned}
\frac{n_n}{n_p} &= e^{-(\Delta E + \mu_p - \mu_n)/kT} \\
&= e^{-(\Delta E + \mu_\nu)/kT} \\
&= x e^{-\Delta E/kT} \\
&= \boxed{\sqrt{\frac{\bar{n}_\nu}{n_\nu}}\, e^{-\Delta E/kT} \ .}
\end{aligned}$$

Alternatively, one can write the answer as

$$\boxed{\frac{n_n}{n_p} = \frac{\sqrt{\Delta n^2 + 4A^2} + \Delta n}{2A}\, e^{-\Delta E/kT} \ ,}$$

where

$$\boxed{A \equiv \frac{3}{4}\frac{\zeta(3)}{\pi^2}\frac{(kT)^3}{(\hbar c)^3} \ .}$$

(d) For $\Delta n > 0$, the answer to (c) implies that the ratio n_n/n_p would be larger than in the usual case ($\Delta n = 0$). This is consistent with the expectation that an excess of antineutrinos will tend to cause p's to turn into n's according to the reaction

$$p + \bar{\nu}_e \longrightarrow e^+ + n \ .$$

Since the amount of helium produced is proportional to the number of neutrons that survive until the breaking of the deuterium bottleneck, starting with a higher equilibrium abundance of neutrons will $\boxed{\text{increase}}$ the production of helium.

PROBLEM 15: THE EVENT HORIZON FOR OUR UNIVERSE *(25 points)*

(a) In a spherical pulse each light ray is moving radially outward, so $\mathrm{d}\theta = \mathrm{d}\phi = 0$. A light ray travels along a null trajectory, meaning that $\mathrm{d}s^2 = 0$, so we have

$$\mathrm{d}s^2 = -c^2\,\mathrm{d}t^2 + a^2(t)\,\mathrm{d}r^2 = 0\ . \tag{3.1}$$

from which it follows that

$$\frac{\mathrm{d}r}{\mathrm{d}t} = \pm\frac{c}{a(t)}\ . \tag{3.2}$$

We are interested in a radial pulse that starts at $r = 0$ at time $t = t_0$, so the limiting value of r is given by

$$\boxed{r_{\max} = \int_{t_0}^{\infty} \frac{c}{a(t)}\,\mathrm{d}t\ .} \tag{3.3}$$

(b) Changing variables of integration to

$$x = \frac{a(t)}{a(t_0)}\ , \tag{3.4}$$

the integral becomes

$$r_{\max} = \int_1^{\infty} \frac{c}{a(t)}\frac{\mathrm{d}t}{\mathrm{d}x}\,\mathrm{d}x = \frac{c}{a(t_0)}\int_1^{\infty}\frac{1}{x}\frac{\mathrm{d}t}{\mathrm{d}x}\,\mathrm{d}x\ , \tag{3.5}$$

where we used the fact that $t = t_0$ corresponds to $x = a(t_0)/a(t_0) = 1$. As given to us on the formula sheet, the first-order Friedmann equation can be written as

$$x\frac{dx}{dt} = H_0\sqrt{\Omega_{m,0}x + \Omega_{\mathrm{rad},0} + \Omega_{\mathrm{vac},0}x^4 + \Omega_{k,0}x^2}\ . \tag{3.6}$$

Using this substitution,

$$\boxed{r_{\max} = \frac{c}{a(t_0)H_0}\int_1^{\infty}\frac{\mathrm{d}x}{\sqrt{\Omega_{m,0}x + \Omega_{\mathrm{rad},0} + \Omega_{\mathrm{vac},0}x^4}}\ ,} \tag{3.7}$$

where we have used $\Omega_{k,0} = 0$, since the universe is taken to be flat.

(c) To find the value of the redshift for the light that we are presently receiving from coordinate distance $r_{\max}$, we can begin by noticing that the time of emission t_e can be determined by the equation which implies that the coordinate distance traveled by a light pulse between times t_e and t_0 must equal $r_{\rm max}$. Using Eq. (3.2) for the coordinate velocity of light, this equation reads

$$\int_{t_e}^{t_0} \frac{c}{a(t)}\,{\rm d}t = r_{\rm max}\ . \tag{3.8}$$

The "half-credit" answer to the quiz problem would include the above equation, followed by the statement that the redshift $z_{\rm eh}$ can be determined from

$$z = \frac{a(t_0)}{a(t_e)} - 1\ . \tag{3.9}$$

The "full-credit" answer is obtained by changing the variable of integration as in part (b), so Eq. (3.8) becomes

$$\begin{aligned} r_{\rm max} &= \int_{x_e}^{1} \frac{c}{a(t)}\,\frac{{\rm d}t}{{\rm d}x}\,{\rm d}x \\ &= \frac{c}{a(t_0)} \int_{x_e}^{1} \frac{1}{x}\,\frac{{\rm d}t}{{\rm d}x}\,{\rm d}x\ , \end{aligned} \tag{3.10}$$

where x_e is the value of x corresponding to $t = t_e$. Then using Eq. (3.6) with $\Omega_{k,0} = 0$, we find

$$\boxed{r_{\rm max} = \frac{c}{a(t_0)H_0} \int_{x_e}^{1} \frac{{\rm d}x}{\sqrt{\Omega_{m,0}x + \Omega_{\rm rad,0} + \Omega_{\rm vac,0}x^4}}\ .} \tag{3.11}$$

To complete the answer in this language, we use

$$\boxed{z = \frac{1}{x_e} - 1\ .} \tag{3.12}$$

Eqs. (3.11) and (3.12) constitute a full answer to the question, but one could go further and replace $r_{\rm max}$ using Eq. (3.7), finding

$$\boxed{\begin{aligned} &\int_{1}^{\infty} \frac{{\rm d}x}{\sqrt{\Omega_{m,0}x + \Omega_{\rm rad,0} + \Omega_{\rm vac,0}x^4}} \\ &\qquad = \int_{x_e}^{1} \frac{{\rm d}x}{\sqrt{\Omega_{m,0}x + \Omega_{\rm rad,0} + \Omega_{\rm vac,0}x^4}}\ . \end{aligned}} \tag{3.13}$$

In this form the answer depends only on the values of $\Omega_{X,0}$.

You were of course not asked to evaluate this formula numerically, but you might be interested in knowing that the Planck 2013 values $\Omega_{m,0} = 0.315$, $\Omega_{\rm vac,0} = 0.685$, and $\Omega_{\rm rad,0} = 9.2 \times 10^{-5}$ lead to $z_{\rm eh} = 1.87$. Thus, no event that is happening now (i.e., at the same value of the cosmic time) in a galaxy at redshift larger than 1.87 will ever be visible to us or our descendants, even in principle.

PROBLEM 16: THE EFFECT OF PRESSURE ON COSMOLOGICAL EVOLUTION *(25 points)*

(a) *(8 points)* This problem is answered most easily by starting from the cosmological formula for energy conservation, which I remember most easily in the form motivated by $dU = -p\,dV$. Using the fact that the energy density u is equal to ρc^2, the energy conservation relation can be written

$$\frac{dU}{dt} = -p\frac{dV}{dt} \quad \Longrightarrow \quad \frac{d}{dt}\left(\rho c^2 a^3\right) = -p\frac{d}{dt}\left(a^3\right) \ .$$

Setting

$$\rho = \frac{\alpha}{a^8}$$

for some constant α, the conservation of energy formula becomes

$$\frac{d}{dt}\left(\frac{\alpha c^2}{a^5}\right) = -p\frac{d}{dt}\left(a^3\right) \ ,$$

which implies

$$-5\frac{\alpha c^2}{a^6}\frac{da}{dt} = -3pa^2\frac{da}{dt} \ .$$

Thus

$$p = \frac{5}{3}\frac{\alpha c^2}{a^8} = \boxed{\frac{5}{3}\rho c^2 \ .}$$

Alternatively, one may start from the equation for the time derivative of ρ,

$$\dot\rho = -3\frac{\dot a}{a}\left(\rho + \frac{p}{c^2}\right) .$$

Since $\rho = \frac{\alpha}{a^8}$, we take the time derivative to find $\dot\rho = -8(\dot a/a)\rho$, and therefore

$$-8\frac{\dot a}{a}\rho = -3\frac{\dot a}{a}\left(\rho + \frac{p}{c^2}\right) ,$$

and therefore

$$p = \frac{5}{3}\,\rho c^2.$$

(b) *(9 points)* For a flat universe, the Friedmann equation reduces to

$$\left(\frac{\dot a}{a}\right)^2 = \frac{8\pi}{3}G\rho \ .$$

Using $\rho \propto 1/a^8$, this implies that

$$\dot{a} = \frac{\beta}{a^3} \ ,$$

for some constant β. Rewriting this as

$$a^3 \, da = \beta \, dt \ ,$$

we can integrate the equation to give

$$\frac{1}{4} a^4 = \beta t + \text{const} \ ,$$

where the constant of integration has no effect other than to shift the origin of the time variable t. Using the standard big bang convention that $a = 0$ when $t = 0$, the constant of integration vanishes. Thus,

$$\boxed{a \propto t^{1/4} \ .}$$

The arbitrary constant of proportionality in this answer is consistent with the wording of the problem, which states that "You should be able to determine the function $a(t)$ up to a constant factor." Note that we could have expressed the constant of proportionality in terms of the constant α that we used in part (a), but there would not really be any point in doing that. The constant α was not a given variable. If the comoving coordinates are measured in "notches," then a is measured in meters per notch, and the constant of proportionality in our answer can be changed by changing the arbitrary definition of the notch.

(c) *(8 points)* We start from the conservation of energy equation in the form

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right) .$$

Substituting $\dot{\rho} = -n(\dot{a}/a)\rho$ and $p = (2/3)\rho c^2$, we have

$$-nH\rho = -3H\left(\frac{5}{3}\rho\right)$$

and therefore

$$\boxed{n = 5.}$$

PROBLEM 17: THE FREEZE-0UT OF A FICTITIOUS PARTICLE X *(25 points)*

(a) *(5 points)* The formula sheet tells us that the energy density of black-body radiation is

$$u = g\frac{\pi^2}{30}\,\frac{(kT)^4}{(\hbar c)^3}\ ,$$

where

$$g \equiv \begin{cases} 1 \text{ per spin state for bosons (integer spin)} \\ 7/8 \text{ per spin state for fermions (half-integer spin)} \end{cases} .$$

Since the X is spin-1, and 1 is an integer, the X particles are bosons and $g = 1$ per spin state. There are 3 species, X^+, X^-, and X^0, and each species we are told has three spin states, so there are a total of 9 spin states, so $g = 9$. Thus,

$$\boxed{u = 9\frac{\pi^2}{30}\,\frac{(kT)^4}{(\hbar c)^3}\ .}$$

Alternatively, one could count the X^+ and X^- as one species with a distinct particle and antiparticle, so $g_{X^+X^-}$ is given by

$$g_{X^+X^-} = \underbrace{1}_{\substack{\text{Fermion}\\ \text{factor}}} \times \underbrace{1}_{\text{Species}} \times \underbrace{2}_{\substack{\text{Particle/}\\ \text{antiparticle}}} \times \underbrace{3}_{\text{Spin states}} = 6\ .$$

The X^0 is its own antiparticle, which means that the particle/antiparticle factor is one, so

$$g_{X^0} = \underbrace{1}_{\substack{\text{Fermion}\\ \text{factor}}} \times \underbrace{1}_{\text{Species}} \times \underbrace{1}_{\substack{\text{Particle/}\\ \text{antiparticle}}} \times \underbrace{3}_{\text{Spin states}} = 3\ ,$$

so the total g for X^+, X^-, and X^0 is again equal to 9.

(b) *(5 points)* The formula sheet tells us that the number density of particles in black-body radiation is

$$n = g^*\frac{\zeta(3)}{\pi^2}\,\frac{(kT)^3}{(\hbar c)^3}\ ,$$

where

$$g^* \equiv \begin{cases} 1 \text{ per spin state for bosons} \\ 3/4 \text{ per spin state for fermions}\ . \end{cases}$$

For bosons $g^* = g$, so g^* for the X particles is 9. Then

$$\boxed{n_X = 9\frac{\zeta(3)}{\pi^2}\,\frac{(kT)^3}{(\hbar c)^3}\ .}$$

(c) *(10 points)* We are told that, when the X particles freeze out, all of their energy and entropy is given to the photons. We use entropy rather than energy to determine the final temperature of the photons, because the entropy in a comoving volume is simply conserved, while the energy density varies as

$$\dot\rho = -3\frac{\dot a}{a}\left(\rho + \frac{p}{c^2}\right) .$$

Thus, to track the energy, we need to know exactly how p behaves, and the behavior of p during freeze-out is complicated, and we have not calculated it in this course.

The formula sheet tells us that the entropy density of a constituent of black-body radiation is given by

$$s = g\frac{2\pi^2}{45}\,\frac{k^4T^3}{(\hbar c)^3}\ .$$

If we consider some fixed coordinate volume $V_{\rm coord}$, the corresponding physical volume is $V_{\rm phys} = V_{\rm coord}\, a^3(t)$, where $a(t)$ is the scale factor. The total entropy of neutrinos in $V_{\rm coord}$ is then

$$S_\nu = g_\nu\frac{2\pi^2}{45}\,\frac{k^4T_\nu^3(t)}{(\hbar c)^3}\,V_{\rm coord}\, a^3(t)\ .$$

The quantities $T_\nu(t)$ and $a(t)$ depend on time, but the expression on the right-hand-side does not, since entropy is conserved. For brevity I will write

$$S_\nu = g_\nu A(t) T_\nu^3(t)\ , \tag{1}$$

where

$$A(t) \equiv \frac{2\pi^2}{45}\,\frac{k^4}{(\hbar c)^3} V_{\rm coord}\, a^3(t)\ .$$

The e^+e^- pairs and the X's contribute to the black-body radiation only before the freeze-out, when $kT \gg 0.511$ MeV$/c^2$. Let t_b denote any time before the freeze-out. Before the freeze-out, the total entropy of photons, e^+e^- pairs, and X particles is given by

$$S_{{\rm before},\gamma e X} = (g_\gamma + g_{e^+e^-} + g_X) A(t_b) T_\gamma^3(t_b)\ . \tag{2}$$

I can call the temperature T_γ, because the e^+e^- pairs and the X's (as well as the neutrinos) are all in thermal equilibrium at this point, so they all have the same temperature.

Using t_a to denote an arbitrary time after the freeze-out, the entropy of the photons during this time period can be written

$$S_{\text{after},\gamma} = g_\gamma A(t_a) T_\gamma^3(t_a) \ . \tag{3}$$

But since the e^+e^- pairs and X particles give all their entropy to the photons, we have

$$S_{\text{after},\gamma} = S_{\text{before},\gamma e X} \ . \tag{4}$$

Then using Eqs. (2) and (3) we find

$$g_\gamma A(t_a) T_\gamma^3(t_a) = (g_\gamma + g_{e^+e^-} + g_X) A(t_b) T_\gamma^3(t_b) \ . \tag{5}$$

We can rewrite the last factor in Eq. (5) by remembering that Eq. (1) holds at all times, and that $T_\nu(t_b) = T_\gamma(t_b)$. So,

$$A(t_b) T_\gamma^3(t_b) = A(t_b) T_\nu^3(t_b) = \frac{S_\nu}{g_\nu} = A(t_a) T_\nu^3(t_a) \ . \tag{6}$$

Substituting Eq. (6) into Eq. (5), we have

$$g_\gamma A(t_a) T_\gamma^3(t_a) = (g_\gamma + g_{e^+e^-} + g_X) A(t_a) T_\nu^3(t_a) \ ,$$

from which we see that

$$T_\gamma^3(t_a) = \frac{g_\gamma + g_{e^+e^-} + g_X}{g_\gamma} T_\nu^3(t_a) \ ,$$

and therefore

$$\begin{aligned} \frac{T_\nu(t_a)}{T_\gamma(t_a)} &= \left(\frac{g_\gamma}{g_\gamma + g_{e^+e^-} + g_X} \right)^{1/3} \\ &= \left(\frac{2}{2 + \frac{7}{2} + 9} \right)^{1/3} = \boxed{\left(\frac{4}{29} \right)^{1/3}} \ . \end{aligned}$$

(d) *(5 points)* The answer would be the same, since it was completely determined by the conservation equation, Eq. (4) in the above answer. Regardless of the order in which the freeze-outs occurred, the total entropy from the e^+e^- pairs and the X's would ultimately be given to the photons, so the amount of heating of the photons would be the same.

PROBLEM 18: THE TIME t_d OF DECOUPLING *(25 points)*

(a) *(5 points)* If the entropy of photons is conserved, then the entropy density falls as

$$s \propto \frac{1}{a^3(t)} \ .$$

Since $s \propto T^3$, it follows that

$$T \propto \frac{1}{a(t)} \ .$$

Thus, the ratio of the scale factors is equal to the inverse of the ratio temperatures:

$$\boxed{x_d = \frac{T_0}{T_d} \ .}$$

(b) *(5 points)* The formula sheet reminds us that

$$x\frac{dx}{dt} = H_0\sqrt{\Omega_{m,0}x + \Omega_{\text{rad},0} + \Omega_{\text{vac},0}x^4 + \Omega_{k,0}x^2} \ ,$$

where

$$\Omega_{k,0} \equiv -\frac{kc^2}{a^2(t_0)H_0^2} = 1 - \Omega_{m,0} - \Omega_{\text{rad},0} - \Omega_{\text{vac},0} \ .$$

So for a flat universe $\Omega_{k,0} = 0$, and we have

$$\boxed{\frac{dx}{dt} = \frac{H_0}{x}\sqrt{\Omega_{m,0}x + \Omega_{\text{rad},0} + \Omega_{\text{vac},0}x^4} \ .}$$

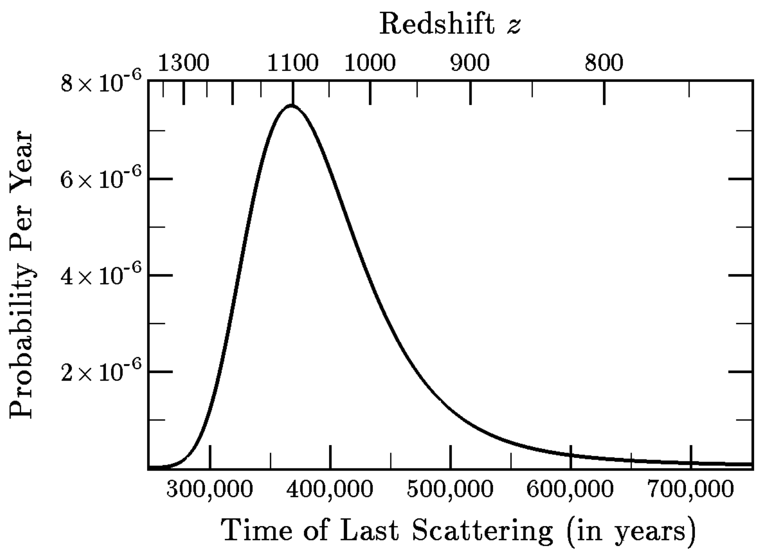
(c) *(5 points)* The answer to part (b) can be rewritten as

$$dt = \frac{x\,dx}{H_0\sqrt{\Omega_{m,0}x + \Omega_{\text{rad},0} + \Omega_{\text{vac},0}x^4}} \ .$$

t_d is the time that elapses from when the universe has $x = 0$ to when it has $x = x_d$, so

$$\boxed{t_d = \frac{1}{H_0}\int_0^{x_d} \frac{x\,dx}{\sqrt{\Omega_{m,0}x + \Omega_{\text{rad},0} + \Omega_{\text{vac},0}x^4}} \ .}$$

You were of course not asked to evaluate this integral numerically, but we will do that now. We take $T_0 = 2.7255$ K from Fixsen et al. (cited in Lecture Notes 6) and the Planck 2015 best fit values of $H_0 = 67.7$ km-s^{-1}-Mpc^{-1}, $\Omega_{m,0} = 0.309$, $\Omega_{\text{vac},0} = 0.691$. The energy density of radiation (photons plus neutrinos) can then be calculated to give $\Omega_{\text{rad},0} = 9.2 \times 10^{-5}$ (see Eq. (6.23) of Lecture Notes 6 and the text of the 2nd paragraph of p. 12 of Lecture Notes 7). To keep our model universe exactly flat, I am modifying $\Omega_{\text{vac},0}$ to set it equal to $0.691 - \Omega_{\text{rad},0}$, which is well within the uncertainties. Numerical integration then gives 366,000 years, very close to our original estimate. Of course this number is still approximate, since we started with $T_d \approx 3000$ K. In any case, the decoupling of the photons in the CMB is actually a gradual process. In 2003 I modified a standard program called CMBFast to calculate the probability distribution of the time of last scattering (published in https://arxiv.org/abs/astro-ph/0306275), with the following results:



The parameters used were $\Omega_{\text{vac},0} = 0.70$, $\Omega_{m,0} = 0.30$, $H_0 = 68$ km-s^{-1}-Mpc^{-1}. The peak of the curve is at 367,000 years, and the median is at 388,000 years.

(d) *(10 points)* The derivation starts with the first-order Friedmann equation. Since we are describing a flat universe, we can start with the Friedmann equation for a flat universe,

$$H^2 = \frac{8\pi}{3} G\rho \ .$$

Now we use the facts that $\rho_m \propto 1/a^3$, $\rho_{\text{rad}} \propto 1/a^4$, $\rho_{\text{vac}} \propto 1$, and $\rho_f \propto 1/a^8$ to write

$$H^2 = \frac{8\pi}{3} G \left[\frac{\rho_{m,0}}{x^3} + \frac{\rho_{\text{rad},0}}{x^4} + \rho_{\text{vac},0} + \frac{\rho_{f,0}}{x^8} \right] \ .$$

Then we use

$$\rho_{m,0} = \rho_c \Omega_{m,0} = \frac{3H_0^2}{8\pi G}\, \Omega_{m,0},$$

with similar relations for the other components of the mass density, to rewrite the Friedmann equation as

$$H^2 = H_0^2 \left[\frac{\Omega_{m,0}}{x^3} + \frac{\Omega_{\text{rad},0}}{x^4} + \Omega_{\text{vac},0} + \frac{\Omega_{f,0}}{x^8} \right] .$$

Next we rewrite H^2 as

$$H^2 = \left(\frac{\dot{a}}{a} \right)^2 = \left(\frac{\dot{x}}{x} \right)^2 ,$$

so

$$\left(\frac{\dot{x}}{x} \right)^2 = H_0^2 \left[\frac{\Omega_{m,0}}{x^3} + \frac{\Omega_{\text{rad},0}}{x^4} + \Omega_{\text{vac},0} + \frac{\Omega_{f,0}}{x^8} \right] ,$$

which can be rewritten as

$$x \frac{dx}{dt} = H_0 \sqrt{\Omega_{m,0} x + \Omega_{\text{rad},0} + \Omega_{\text{vac},0} x^4 + \frac{\Omega_{f,0}}{x^4}} .$$

From here the derivation is identical to that in part (c), leading to

$$\boxed{t_d = \frac{1}{H_0} \int_0^{x_d} \frac{x\, dx}{\sqrt{\Omega_{m,0} x + \Omega_{\text{rad},0} + \Omega_{\text{vac},0} x^4 + \frac{\Omega_{f,0}}{x^4}}} ,}$$

which can also be written more neatly as

$$\boxed{t_d = \frac{1}{H_0} \int_0^{x_d} \frac{x^3\, dx}{\sqrt{\Omega_{m,0} x^5 + \Omega_{\text{rad},0} x^4 + \Omega_{\text{vac},0} x^8 + \Omega_{f,0}}} .}$$

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — December 5, 2018
Prof. Alan Guth

QUIZ 3*

Reformatted to Remove Blank Pages
Please answer all questions in this stapled booklet.

PROBLEM 1: DID YOU DO THE READING? *(20 points)*

(a) *(5 points)* Which one of the following statements about CMB is NOT correct?

(i) The dipole distortion is a simple Doppler shift, caused by the net motion of the observer relative to a frame of reference in which the CMB is isotropic.

(ii) After the dipole distortion of the CMB is subtracted away, the mean temperature averaging over the sky is $<\mathrm{T}> = 2.725\mathrm{K}$.

(iii) After the dipole distortion of the CMB is subtracted away, the temperature of the CMB varies by 0.3 microKelvin across the sky.

(iv) The photons of the CMB have mostly been traveling on straight lines since they were last scattered at $t \approx 370,000$ yr, at a location called the surface of last scattering.

(b) *(5 points)* The nonuniformities in the cosmic microwave background allow us to measure the ripples in the mass density of the universe at the time when the plasma combined to form neutral atoms, about 300,000 - 400,000 years after the big bang. These ripples are crucial for understanding what happened later, since they are the seeds which led to the complicated tapestry of galaxies, clusters of galaxies, and voids. Which of the following sentences describes how these ripples are created in the context of inflationary models:

(i) Magnetic monopoles can form randomly during the grand unified theory phase transition, resulting in nonuniformities in the mass density.

(ii) Cosmic strings, which are linelike topological defects, can form randomly during the grand unified theory phase transition, resulting in nonuniformities in the mass density.

(iii) They are generated by quantum fluctuations during inflation.

(iv) Since the early universe was very hot, there were large thermal fluctuations which ultimately evolved into the ripples in the mass density.

— Problem 1 continues on next page. —

---

* This version includes corrections that were announced during the quiz.

(c) *(5 points)* In Chapter 8 of *The First Three Minutes*, Steven Weinberg describes the future of the universe (assuming, as was thought then to be the case, that the cosmological constant is zero). One possibility that he discusses is that the cosmic matter density could be greater than the critical density. Assuming that we live in such a universe, which of the following statements is NOT true?

(i) The universe is finite and its expansion will eventually cease, giving way to an accelerating contraction.

(ii) Three minutes after the temperature reaches a thousand million degrees (10^9 K), the laws of physics guarantee that the universe will crunch, and time will stop.

(iii) During at least the early part of the contracting phase, we will be able to observe both redshifts and blueshifts.

(iv) When the universe has recontracted to one-hundredth its present size, the radiation background will begin to dominate the sky, with a temperature of about 300 K.

(d) *(5 points)* Which of the following describes the Sachs-Wolfe effect?

(i) Photons from fluid which had a velocity toward us along the line of sight appear redder because of the Doppler effect.

(ii) Photons from fluid which had a velocity toward us along the line of sight appear bluer because of the Doppler effect.

(iii) Photons traveling toward us from the surface of last scattering appear redder because of absorption in the intergalactic medium.

(iv) Photons traveling toward us from the surface of last scattering appear bluer because of absorption in the intergalactic medium.

(v) Photons from overdense regions at the surface of last scattering appear redder because they must climb out of the gravitational potential well.

(vi) Photons from overdense regions at the surface of last scattering appear bluer because they must climb out of the gravitational potential well.

— End of Problem 1. —

PROBLEM 2: TIME EVOLUTION OF A UNIVERSE INCLUDING A HYPOTHETICAL KIND OF MATTER *(30 points)*

Suppose that a flat universe includes nonrelativistic matter, radiation, and also mysticium, where the mass density of mysticium behaves as

$$\rho_{\rm myst} \propto \frac{1}{a^5(t)}$$

as the universe expands. In this problem we will define

$$x(t) \equiv \frac{a(t)}{a(t_0)} ,$$

where t_0 is the present time. For the following questions, you need not evaluate any of the integrals that might arise, but they must be integrals of explicit functions with explicit limits of integration; remember that $a(t)$ is not given. You may express your answers in terms of the present value of the Hubble expansion rate, H_0, and the various contributions to the present value of Ω: $\Omega_{m,0}$, $\Omega_{\rm rad,0}$, and $\Omega_{\rm myst,0}$.

(a) *(7 points)* Write an expression for the Hubble expansion rate $H(x)$.

(b) *(7 points)* Write an expression for the current age of the universe.

(c) *(3 points)* Write an expression for the time $t(x)$ in terms of the value of x.

(d) *(3 points)* Write an expression for the total mass density $\rho(x)$ as a function of x.

(e) *(10 points)* Write an expression for present value of the physical horizon distance, $\ell_{p,\rm hor}(t_0)$.

PROBLEM 3: PROPERTIES OF BLACK-BODY RADIATION *(25 points)*

The following problem was Problem 6 of the Review Problems for Quiz 3.

In answering the following questions, remember that you can refer to the formulas on the formula sheets, circulated separately. Since you were not asked to bring calculators, you may leave your answers in the form of algebraic expressions, such as $\pi^{32}/\sqrt{5\zeta(3)}$.

(a) *(5 points)* For the black-body radiation (also called thermal radiation) of photons at temperature T, what is the average energy per photon?

(b) *(5 points)* For the same radiation, what is the average entropy per photon?

(c) *(5 points)* Now consider the black-body radiation of a massless boson which has spin zero, so there is only one spin state. Would the average energy per particle and entropy per particle be different from the answers you gave in parts (a) and (b)? If so, how would they change?

(d) *(5 points)* Now consider the black-body radiation of electron neutrinos at temperature T. These particles are fermions with spin 1/2, and we will assume that they are massless and have only one possible spin state. What is the average energy per particle for this case?

(e) *(5 points)* What is the average entropy per particle for the black-body radiation of neutrinos, as described in part (d)?

PROBLEM 4: THE CONSEQUENCES OF AN ALT-PHOTON *(25 points)*

Suppose that, in addition to the particles that are known to exist, there also existed an alt-photon, which has exactly the properties of a photon: it is massless, has two spin states (or polarization states), and has the same interactions with other particles that photons do. Like photons, it is its own antiparticle.

(a) *(5 points)* In thermal equilibrium at temperature T, what is the total energy density of alt-photons?

(b) *(5 points)* In thermal equilibrium at temperature T, what is the number density of alt-photons?

(c) *(10 points)* In this situation, what would be the temperature ratios T_ν/T_γ and $T_\nu/T_{\mathrm{alt}\gamma}$ today?

(d) *(5 points)* Would the existence of this particle increase or decrease the abundance of helium, or would it have no effect?

Problem	Maximum	Score	Initials
1	20		
2	30		
3	25		
4	25		
TOTAL	100		

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — December 19, 2018
Prof. Alan Guth

QUIZ 3 SOLUTIONS

Quiz Date: December 5, 2018

Reformatted to Remove Blank Pages
Please answer all questions in this stapled booklet.

PROBLEM 1: DID YOU DO THE READING? *(20 points)*

(a) *(5 points)* Which one of the following statements about CMB is NOT correct?

(i) The dipole distortion is a simple Doppler shift, caused by the net motion of the observer relative to a frame of reference in which the CMB is isotropic.

(ii) After the dipole distortion of the CMB is subtracted away, the mean temperature averaging over the sky is <T> = 2.725K.

$\boxed{\text{(iii)}}$ After the dipole distortion of the CMB is subtracted away, the temperature of the CMB varies by 0.3 microKelvin across the sky.

(iv) The photons of the CMB have mostly been traveling on straight lines since they were last scattered at $t \approx 370,000$ yr, at a location called the surface of last scattering.

[Comment: The actual variation is about 30 microKelvin, or maybe a few times that much. Ryden quotes the COBE root mean square fractional variation of the CMB temperature as

$$\left\langle \left(\frac{\delta T}{T} \right)^2 \right\rangle^{1/2} = 1.1 \times 10^{-5}$$

as Eq. (8.8) (2nd Edition), which gives a value of about 30 microKelvin, given that $T \approx 3\,K$. In Lecture Notes 2 we quoted a value of 4.14×10^{-5} computed from Planck data. The root mean square fluctuations increase with better angular resolution, because fluctuations with small angular wavelengths are not seen unless the resolution is high.

(b) *(5 points)* The nonuniformities in the cosmic microwave background allow us to measure the ripples in the mass density of the universe at the time when the plasma combined to form neutral atoms, about 300,000 - 400,000 years after the big bang. These ripples are crucial for understanding what happened later, since they are the seeds which led to the complicated tapestry of galaxies, clusters of galaxies, and voids. Which of the following sentences describes how these ripples are created in the context of inflationary models:

(i) Magnetic monopoles can form randomly during the grand unified theory phase transition, resulting in nonuniformities in the mass density.

(ii) Cosmic strings, which are linelike topological defects, can form randomly during the grand unified theory phase transition, resulting in nonuniformities in the mass density.

(iii) They are generated by quantum fluctuations during inflation.

(iv) Since the early universe was very hot, there were large thermal fluctuations which ultimately evolved into the ripples in the mass density.

(c) *(5 points)* In Chapter 8 of *The First Three Minutes*, Steven Weinberg describes the future of the universe (assuming, as was thought then to be the case, that the cosmological constant is zero). One possibility that he discusses is that the cosmic matter density could be greater than the critical density. Assuming that we live in such a universe, which of the following statements is NOT true?

(i) The universe is finite and its expansion will eventually cease, giving way to an accelerating contraction.

(ii) Three minutes after the temperature reaches a thousand million degrees (10^9 K), the laws of physics guarantee that the universe will crunch, and time will stop.

(iii) During at least the early part of the contracting phase, we will be able to observe both redshifts and blueshifts.

(iv) When the universe has recontracted to one-hundredth its present size, the radiation background will begin to dominate the sky, with a temperature of about 300 K.

[Comment: Weinberg is very clear no speculations about the end of the universe are guaranteed to be true: "Does time really have to stop some three minutes after the temperature reaches a thousand million degrees? Obviously, we cannot be sure. All the uncertainties that we met in the preceding chapter, in trying to explore the first hundredth of a second, will return to perplex us as we look into the last hundredth of a second."]

(d) *(5 points)* Which of the following describes the Sachs-Wolfe effect?

(i) Photons from fluid which had a velocity toward us along the line of sight appear redder because of the Doppler effect.

(ii) Photons from fluid which had a velocity toward us along the line of sight appear bluer because of the Doppler effect.

(iii) Photons traveling toward us from the surface of last scattering appear redder because of absorption in the intergalactic medium.

(iv) Photons traveling toward us from the surface of last scattering appear bluer because of absorption in the intergalactic medium.

$\boxed{\text{(v)}}$ Photons from overdense regions at the surface of last scattering appear redder because they must climb out of the gravitational potential well.

(vi) Photons from overdense regions at the surface of last scattering appear bluer because they must climb out of the gravitational potential well.

[Comment: Ryden discusses the Sachs-Wolfe effect on pp. 161–162 (2nd Edition).

— End of Problem 1. —

PROBLEM 2: TIME EVOLUTION OF A UNIVERSE INCLUDING A HYPOTHETICAL KIND OF MATTER *(30 points)*

Suppose that a flat universe includes nonrelativistic matter, radiation, and also mysticium, where the mass density of mysticium behaves as

$$\rho_{\rm myst} \propto \frac{1}{a^5(t)}$$

as the universe expands. In this problem we will define

$$x(t) \equiv \frac{a(t)}{a(t_0)} ,$$

where t_0 is the present time. For the following questions, you need not evaluate any of the integrals that might arise, but they must be integrals of explicit functions with explicit limits of integration; remember that $a(t)$ is not given. You may express your answers in terms of the present value of the Hubble expansion rate, H_0, and the various contributions to the present value of Ω: $\Omega_{m,0}$, $\Omega_{\rm rad,0}$, and $\Omega_{\rm myst,0}$.

(a) *(7 points)* Write an expression for the Hubble expansion rate $H(x)$.

(b) *(7 points)* Write an expression for the current age of the universe.

(c) *(3 points)* Write an expression for the time $t(x)$ in terms of the value of x.

(d) *(3 points)* Write an expression for the total mass density $\rho(x)$ as a function of x.

(e) *(10 points)* Write an expression for present value of the physical horizon distance, $\ell_{p,\rm hor}(t_0)$.

(a) Since the universe is flat, the first Friedmann equation becomes

$$H^2 = \frac{8\pi}{3} G\rho ,$$

but then we can write ρ as

$$H^2 = \frac{8\pi}{3} G \left\{ \rho_{m,0} \left[\frac{a(t_0)}{a(t)}\right]^3 + \rho_{\rm rad,0} \left[\frac{a(t_0)}{a(t)}\right]^4 + \rho_{\rm myst,0} \left[\frac{a(t_0)}{a(t)}\right]^5 \right\} .$$

Now use

$$\rho_{c,0} = \frac{3H_0^2}{8\pi G} \text{ and } \Omega \equiv \frac{\rho}{\rho_c} ,$$

so

$$H^2 = \frac{H_0^2}{\rho_{c,0}} \left\{ \rho_{m,0} \left[\frac{a(t_0)}{a(t)}\right]^3 + \rho_{\text{rad},0} \left[\frac{a(t_0)}{a(t)}\right]^4 + \rho_{\text{myst},0} \left[\frac{a(t_0)}{a(t)}\right]^5 \right\}$$

$$= H_0^2 \left\{ \frac{\Omega_{m,0}}{x^3} + \frac{\Omega_{\text{rad},0}}{x^4} + \frac{\Omega_{\text{myst},0}}{x^5} \right\} .$$

Finally,

$$\boxed{H(x) = \frac{H_0}{x^2} \sqrt{\Omega_{m,0} x + \Omega_{\text{rad},0} + \frac{\Omega_{\text{myst},0}}{x}} \ .}$$

(b) To find the current age t_0, we start with

$$H = \frac{\dot{a}}{a} = \frac{\dot{x}}{x} \quad \Longrightarrow \quad \frac{\mathrm{d}x}{\mathrm{d}t} = xH \quad \Longrightarrow \quad \mathrm{d}t = \frac{\mathrm{d}x}{xH} \ .$$

So t_0 can be found by integrating over the range of x, from 0 to 1:

$$t_0 = \int_0^1 \frac{\mathrm{d}x}{xH(x)}$$

$$= \boxed{\frac{1}{H_0} \int_0^1 \frac{x \, \mathrm{d}x}{\sqrt{\Omega_{m,0} x + \Omega_{\text{rad},0} + \frac{\Omega_{\text{myst},0}}{x}}} \ .}$$

(c) To find the time t corresponding to some value of x other than 1, one simply integrates $\mathrm{d}t$ from $x' = 0$ to $x' = x$:

$$t(x) = \int_0^x \frac{\mathrm{d}x'}{x' H(x')}$$

$$= \boxed{\frac{1}{H_0} \int_0^x \frac{x' \, \mathrm{d}x'}{\sqrt{\Omega_{m,0} x' + \Omega_{\text{rad},0} + \frac{\Omega_{\text{myst},0}}{x'}}} \ .}$$

(d) From the first Friedmann equation,

$$H^2 = \frac{8\pi}{3} G\rho \quad \Longrightarrow \quad \rho = \frac{3}{8\pi G} H^2(x) \ .$$

Given the answer in part (a), this becomes

$$\boxed{\rho(x) = \frac{3}{8\pi G}\frac{H_0^2}{x^4}\left[\Omega_{m,0}x + \Omega_{\text{rad},0} + \frac{\Omega_{\text{myst},0}}{x}\right] \ .}$$

(e) The general formula for the physical horizon distance is given on the formula sheet:

$$\ell_{p,\text{hor}}(t) = a(t)\int_0^t \frac{c}{a(t')}dt' \ .$$

Here we are not given the function $a(t)$, but we can change the variable of integration to integrate over x:

$$\mathrm{d}t' = \frac{\mathrm{d}t'}{\mathrm{d}a}\mathrm{d}a = \frac{1}{\dot{a}}\mathrm{d}a = \frac{1}{a}\frac{a}{\dot{a}}\mathrm{d}a = \frac{\mathrm{d}a}{aH(x)} \ .$$

So

$$\begin{aligned}\ell_{p,\text{hor}}(t_0) &= a(t_0)\int_0^{a(t_0)} \frac{c\,\mathrm{d}a}{a^2 H(a)} \\ &= \int_0^1 \frac{c\,\mathrm{d}x}{x^2 H(x)} \\ &= \boxed{\frac{c}{H_0}\int_0^1 \frac{\mathrm{d}x}{\sqrt{\Omega_{m,0}x + \Omega_{\text{rad},0} + \frac{\Omega_{\text{myst},0}}{x}}} \ .}\end{aligned}$$

PROBLEM 3: PROPERTIES OF BLACK-BODY RADIATION *(25 points)*

The following problem was Problem 6 of the Review Problems for Quiz 3.

In answering the following questions, remember that you can refer to the formulas on the formula sheets, circulated separately. Since you were not asked to bring calculators, you may leave your answers in the form of algebraic expressions, such as $\pi^{32}/\sqrt{5\zeta(3)}$.

(a) *(5 points)* For the black-body radiation (also called thermal radiation) of photons at temperature T, what is the average energy per photon?

(b) *(5 points)* For the same radiation, what is the average entropy per photon?

(c) *(5 points)* Now consider the black-body radiation of a massless boson which has spin zero, so there is only one spin state. Would the average energy per particle and entropy per particle be different from the answers you gave in parts (a) and (b)? If so, how would they change?

(d) *(5 points)* Now consider the black-body radiation of electron neutrinos at temperature T. These particles are fermions with spin 1/2, and we will assume that they are massless and have only one possible spin state. What is the average energy per particle for this case?

(e) *(5 points)* What is the average entropy per particle for the black-body radiation of neutrinos, as described in part (d)?

Solution:

(a) The average energy per photon is found by dividing the energy density by the number density. The photon is a boson with two spin states, so $g = g^* = 2$. Using the formulas on the front of the exam,

$$E = \frac{g\dfrac{\pi^2}{30}\dfrac{(kT)^4}{(\hbar c)^3}}{g^*\dfrac{\zeta(3)}{\pi^2}\dfrac{(kT)^3}{(\hbar c)^3}}$$

$$= \boxed{\frac{\pi^4}{30\zeta(3)}\,kT\ .}$$

You were not expected to evaluate this numerically, but it is interesting to know that

$$E = 2.701\,kT\ .$$

Note that the average energy per photon is significantly more than kT, which is often used as a rough estimate.

(b) The method is the same as above, except this time we use the formula for the entropy density:

$$S = \frac{g\dfrac{2\pi^2}{45}\dfrac{k^4T^3}{(\hbar c)^3}}{g^*\dfrac{\zeta(3)}{\pi^2}\dfrac{(kT)^3}{(\hbar c)^3}}$$

$$= \boxed{\frac{2\pi^4}{45\zeta(3)}\,k\ .}$$

Numerically, this gives $3.602\,k$, where k is the Boltzmann constant.

(c) In this case we would have $g = g^* = 1$. The average energy per particle and the average entropy particle depends only on the ratio g/g^*, so there would be $\boxed{\text{no difference from the answers given in parts (a) and (b).}}$

(d) For a fermion, g is $7/8$ times the number of spin states, and g^* is $3/4$ times the number of spin states. So the average energy per particle is

$$E = \frac{g\dfrac{\pi^2}{30}\dfrac{(kT)^4}{(\hbar c)^3}}{g^*\dfrac{\zeta(3)}{\pi^2}\dfrac{(kT)^3}{(\hbar c)^3}}$$

$$= \frac{\dfrac{7}{8}\dfrac{\pi^2}{30}\dfrac{(kT)^4}{(\hbar c)^3}}{\dfrac{3}{4}\dfrac{\zeta(3)}{\pi^2}\dfrac{(kT)^3}{(\hbar c)^3}}$$

$$= \boxed{\frac{7\pi^4}{180\zeta(3)}\,kT\ .}$$

Numerically, $E = 3.1514\,kT$.

Warning: the Mathematician General has determined that the memorization of this number may adversely affect your ability to remember the value of π.

If one takes into account both neutrinos and antineutrinos, the average energy per particle is unaffected — the energy density and the total number density are both doubled, but their ratio is unchanged.

Note that the energy per particle is higher for fermions than it is for bosons. This result can be understood as a natural consequence of the fact that fermions must obey the exclusion principle, while bosons do not. Large numbers of bosons can therefore collect in the lowest energy levels. In fermion systems, on the other hand, the low-lying levels can accommodate at most one particle, and then additional particles are forced to higher energy levels.

(e) The values of g and g^* are again $7/8$ and $3/4$ respectively, so

$$S = \frac{g \dfrac{2\pi^2}{45} \dfrac{k^4 T^3}{(\hbar c)^3}}{g^* \dfrac{\zeta(3)}{\pi^2} \dfrac{(kT)^3}{(\hbar c)^3}} = \frac{\dfrac{7}{8} \dfrac{2\pi^2}{45} \dfrac{k^4 T^3}{(\hbar c)^3}}{\dfrac{3}{4} \dfrac{\zeta(3)}{\pi^2} \dfrac{(kT)^3}{(\hbar c)^3}} = \boxed{\frac{7\pi^4}{135\zeta(3)} k \ .}$$

Numerically, this gives $S = 4.202\, k$.

PROBLEM 4: THE CONSEQUENCES OF AN ALT-PHOTON *(25 points)*

Suppose that, in addition to the particles that are known to exist, there also existed an alt-photon, which has exactly the properties of a photon: it is massless, has two spin states (or polarization states), and has the same interactions with other particles that photons do. Like photons, it is its own antiparticle.

(a) *(5 points)* In thermal equilibrium at temperature T, what is the total energy density of alt-photons?

(b) *(5 points)* In thermal equilibrium at temperature T, what is the number density of alt-photons?

(c) *(10 points)* In this situation, what would be the temperature ratios T_ν/T_γ and $T_\nu/T_{\mathrm{alt}\gamma}$ today?

(d) *(5 points)* Would the existence of this particle increase or decrease the abundance of helium, or would it have no effect?

Solution:

(a) The energy density will be the same as for photons, since there is no difference. The general formula is

$$u = g\frac{\pi^2}{30}\,\frac{(kT)^4}{(\hbar c)^3}\ ,$$

as given on the formula sheets, and $g = 2$ for alt-photons (or photons), since there are two polarization states, and the particles are bosons. So

$$\boxed{u_{\mathrm{alt}\gamma} = \frac{\pi^2}{15}\,\frac{(kT)^4}{(\hbar c)^3}\ .} \tag{4.1}$$

(b) For the number density, the general formula is

$$n = g^*\frac{\zeta(3)}{\pi^2}\,\frac{(kT)^3}{(\hbar c)^3}\ ,$$

where $g^* = 2$ since again the alt-photons are bosons with two polarization states. So

$$\boxed{n_{\mathrm{alt}\gamma} = 2\frac{\zeta(3)}{\pi^2}\,\frac{(kT)^3}{(\hbar c)^3}\ .} \tag{4.2}$$

(c) As in the actual scenario, the event that causes a temperature difference is the disappearance of the electron-positron pairs from the thermal equilibrium mix, which occurs as kT changes from values large compared to $m_e c^2 = 0.511$ MeV to values that are small compared to it. The key point is that this disappearance occurs after the neutrinos have decoupled from the other particles, so all of the entropy from the electron-positron pairs is given to the photons, and none is given to the neutrinos. In this case the entropy is given to both the photons and the alt-photons.

The general formula for entropy density is on the formula sheet, and it can be rewritten as

$$s = AgT^3 \ , \tag{4.3}$$

where

$$A = \frac{2\pi^2}{45}\frac{k^4}{(\hbar c)^3} \ . \tag{4.4}$$

The value of A will in fact not be needed for this problem.

Since the neutrinos have decoupled by the time the e^+e^- pairs disappear, the entropy of neutrinos and the entropy of everything else will be separately conserved. Entropy conservation means that the entropy per comoving volume does not change. During the period before e^+e^- freeze-out, g is constant, so the constancy of entropy per comoving volume implies that

$$S = sV_{\text{phys}} = gT^3 AV_{\text{phys}} = ga^3T^3AV_{\text{coord}} \ , \tag{4.5}$$

so $S/V_{\text{coord}} = \text{const}$ implies that a^3T^3 is constant, and so aT is constant. Here T is the common temperature of photons, alt-photons, electrons and positrons, and neutrinos, all of which were in thermal equilibrium during this period. Since aT is constant during this period, we can give the constant a name,

$$aT = [aT]_{\text{before}} \ . \tag{4.6}$$

For the neutrinos, the formula sheet tells us that

$$g_\nu = \underbrace{\frac{7}{8}}_{\substack{\text{Fermion}\\ \text{factor}}} \times \underbrace{3}_{\substack{\text{3 species}\\ \nu_e,\nu_\mu,\nu_\tau}} \times \underbrace{2}_{\substack{\text{Particle/}\\ \text{antiparticle}}} \times \underbrace{1}_{\text{Spin states}} = \frac{21}{4} \ , \tag{4.7}$$

while

$$g_{e^+e^-} = \underbrace{\frac{7}{8}}_{\substack{\text{Fermion}\\ \text{factor}}} \times \underbrace{1}_{\text{Species}} \times \underbrace{2}_{\substack{\text{Particle/}\\ \text{antiparticle}}} \times \underbrace{2}_{\text{Spin states}} = \frac{7}{2} \ . \tag{4.8}$$

Thus

$$g_{\text{else}} = g_\gamma + g_{\text{alt}\gamma} + g_{e^+e^-} = 2 + 2 + \frac{7}{2} = \frac{15}{2} \ . \tag{4.9}$$

Thus before the e^+e^- freezeout, the two conserved quantities were

$$\frac{S_\nu}{V_{\text{coord}}} = A g_\nu [aT]^3_{\text{before}} \ , \quad \frac{S_{\text{else}}}{V_{\text{coord}}} = A g_{\text{else}} [aT]^3_{\text{before}} \ . \tag{4.10}$$

After e^+e^- freezeout, the temperature of the neutrinos T_ν will no longer be the same as the temperature T_γ of the photons and alt-photons, and of course e^+e^- pairs will no longer be present. But T_γ and $T_{\text{alt}\gamma}$ will be equal to each other, since they have the same interactions; we know that the interactions of the photons keep them in thermal equilibrium until $t_{\text{decoupling}} \sim 380,000$ years, so both the photons and the alt-photons will remain in thermal equilibrium until long after the era of e^+e^- freezeout, which is of order 1–10 seconds. Thus the two conserved quantities will be

$$\frac{S_\nu}{V_{\text{coord}}} = A g_\nu [aT_\nu]^3_{\text{after}} \ , \quad \frac{S_{\text{else}}}{V_{\text{coord}}} = A (g_\gamma + g_{\text{alt}\gamma}) [aT_\gamma]^3_{\text{after}} \ . \tag{4.11}$$

By equating the values of S_ν/V_{coord} before and after, we see that

$$[aT_\nu]_{\text{after}} = [aT]_{\text{before}} \ , \tag{4.12}$$

and then by equating the values of $S_{\text{else}}/V_{\text{coord}}$ before and after, we see that

$$[aT_\gamma]_{\text{after}} = \left(\frac{g_{\text{else}}}{g_\gamma + g_{\text{alt}\gamma}}\right)^{1/3} [aT]_{\text{before}} = \left(\frac{g_{\text{else}}}{g_\gamma + g_{\text{alt}\gamma}}\right)^{1/3} [aT_\nu]_{\text{after}} \ , \tag{4.13}$$

where we used Eq. (4.12) in the last step. It follows that

$$\left[\frac{T_\nu}{T_\gamma}\right]_{\text{after}} = \left(\frac{g_\gamma + g_{\text{alt}\gamma}}{g_{\text{else}}}\right)^{1/3} = \left(\frac{2+2}{\frac{15}{2}}\right)^{1/3} = \boxed{\left(\frac{8}{15}\right)^{1/3} \ .} \tag{4.14}$$

(d) It would increase the abundance of helium. The main effect of the alt-photon would be to increase the expansion rate of the universe, which in turn would cause the neutrinos to decouple earlier from the thermal equilibrium mix, which in turn would mean that the ratio n_n/n_p, the ratio of neutrons to protons, would become frozen at a larger value. The increased expansion rate would also mean less time available for free neutron decay, which further increases the number of neutrons that remain when the temperature falls low enough for helium formation to complete. Essentially all the neutrons become bound into helium, so more neutrons implies more helium.

Problem	Maximum	Score	Initials
1	20		
2	30		
3	25		
4	25		
TOTAL	100		