MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Physics Department

Physics 8.286: The Early Universe — September 22, 2009

Prof. Alan Guth

Lecture Notes 3
THE KINEMATICS OF NEWTONIAN COSMOLOGY

INTRODUCTION:

Observational cosmology is of course a rich and complicated subject. It is described to some degree in the textbooks, and I will not enlarge on that discussion here. I will instead concentrate on the basic results of observational cosmology, which can be summarized in a few statements:

(1) ISOTROPY

On large scales, the universe appears about the same in all directions. The nearby region is rather anisotropic (i.e., looks different in different directions), since it is dominated by the center of a supercluster of galaxies of which our galaxy, the Milky Way, is a part. The center of this supercluster is in the Virgo cluster, approximately 60 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 this isotropy comes from the observation of the cosmic microwave background radiation, which is interpreted as the remnant heat from the big bang itself. Physicists have probed 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 600 km/s (kilometers/second). Once the effect of this motion is subtracted out, then the resulting temperature pattern is uniform in all directions to an accuracy of about 1 part in 100,000. Thus, on the very large scales which are probed by this radiation, the universe is incredibly isotropic.

Note that this fact stands as firm 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 the radiation would appear 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

It is conceivable that the universe appears isotropic because all the galaxies are arranged in concentric spheres about us. Such an assumption, however, is also very contrary to the prejudices that scientists have adopted since Copernicus. So we assume instead that the universe is roughly homogeneous on large scales. That is, we assume that if one observes only large-scale structure, then the universe would look about 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. 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.

(3) HUBBLE'S LAW

Hubble's law, first proposed by Edwin Hubble in 1929, states that all the distant galaxies are receding from us, with a recession velocity given by

$$\boxed{v = Hr\ .} \tag{3.1}$$

Here

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

$$H \equiv \text{Hubble's constant}\ ,$$

and

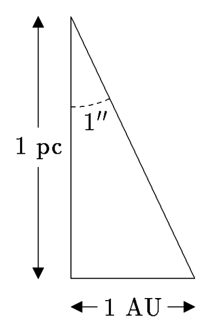
$$r \equiv \text{distance to galaxy}\ .$$

Note that Hubble's constant H was called a "constant" by the astronomers, but it is constant only in the sense that its value changes by very little over the lifetime of an astronomer. Over the lifetime of the universe, H varies considerably. The present value of the Hubble constant is denoted by H_0, following a standard convention in cosmology: the present value of any time-dependent quantity is indicated by a subscript "o". 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. $H(t)$ is sometimes (see for example the table of Astrophysical Constants and Parameters of the Particle Data Group, http://pdg.lbl.gov/2009/reviews/rpp2009-rev-astrophysical-constants.pdf) referred to as the "Hubble expansion rate," which sounds the best to me.

The numerical value of H_0 is difficult to determine, because of the difficulty in measuring distances. During the 1960s, 70s, and 80s, the Hubble constant was merely known to lie somewhere in the range of

$$H_0 = \frac{0.5 - 1.0}{10^{10}\text{ years}} . \tag{3.2}$$

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 (1 pc) is the distance which corresponds to a parallax of 1 second of arc as the Earth moves in its orbit, which is treated for this purpose as a circle whose radius is 1 AU (astronomical unit, $149.597870700(3)\times10^9$ m), as illustrated at the right. One parsec (abbreviated pc) corresponds to 3.2616 light-years.* Astronomers usually quote the value of the Hubble constant in units of km/s per 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. (3.2) corresponds roughly to a Hubble constant



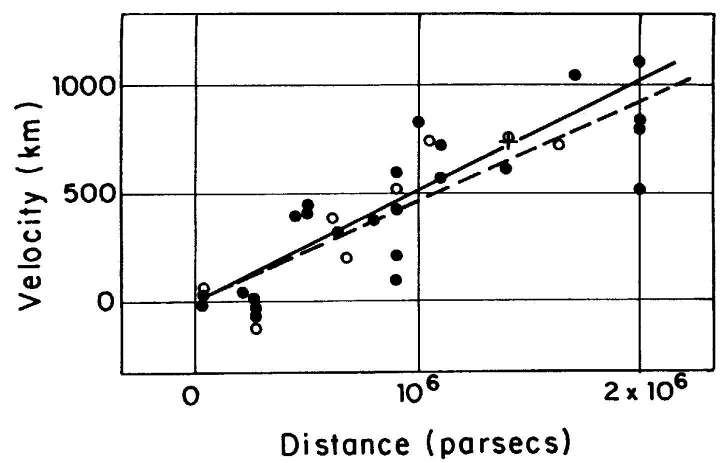
* One drawback in using light-years is that the definition is tied to that of a year, and the International (SI) System of Units does not specify the definition of a year. This is a significant ambiguity, because the tropical year (vernal equinox to vernal equinox) and the sidereal year (full revolution about the Sun, relative to the fixed stars) differ by a 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/static/publications/stylemanual1989.pdf). The day is $24 \times 60 \times 60$ seconds, and the second is defined by atomic standards.

between 50 and 100 km-s^{-1}-Mpc^{-1}. For convenience, astronomers also define the dimensionless quantity h_0 by

$$\boxed{H_0 \equiv h_0 \times (100\text{ km-s}^{-1}\text{-Mpc}^{-1})\ .} \tag{3.3}$$

The range of Eq. (3.2) translates into a value of h_0 between $\frac{1}{2}$ and 1.

While the actual value of the Hubble constant certainly changes very little over the lifetime of an astronomer, the same cannot be said for its measured value. When Hubble first measured it, he found a value of about 500 km-s^{-1}-Mpc^{-1}, due to a very bad estimate of the distance scale. Hubble's original published graph was the following:*



The horizontal axis shows the estimated distance to the galaxies, and the vertical axis shows the recession velocity, 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

* 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).

measurements were not available. The evidence for a straight line is not completely convincing, but of course the data improved tremendously in subsequent years. 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.

As we will see later, a value of the Hubble constant as large as 500 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. (3.2), primarily due to the work of Walter Baade and Allan Sandage. A summary of this history, going up to 2002, was written by Tammann and Reindl.*

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. 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 uncertaintly expressed in Eq. (3.2). 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. In February 2003 the Wilkinson Microwave Anisotropy Probe (WMAP), a satellite dedicated to measuring the cosmic background radiation, announced a fit to its first year of data¶ which gave $H_0 = 72 \pm 5$ km-s^{-1}-Mpc^{-1}. The most precise measurement of H_0 available today is probably the value obtained from the WMAP 5-year data, which was combined by the WMAP team with several other experiments to obtain the value 70.1 ± 1.3 km-s^{-1}-Mpc^{-1}.§

* G.A. Tammann and B. Reindl, to appear 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.

† W.L. Freedman et al., *Astrophys. J.* **553,** 47–72 (2001), available as http://arXiv.org/abs/astro-ph/0012376.

‡ 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, available at http://arXiv.org/abs/astro-ph/0112489.

¶ D.N. Spergel et al., *Astrophys. J. Suppl.* **148,** 175–194 (2003), also available as http://lambda.gsfc.nasa.gov/product/map/dr1/pub_papers/firstyear/parameters/57707.web.pdf.

§ G. Hinshaw et al., "Five-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Data Processing, Sky Maps, and Basic Results," *Astrophys. J. Suppl.* **180,** 225-245 (2009), also available as http://lambda.gsfc.nasa.gov/product/map/dr3/pub_papers/fiveyear/basic_results/wmap5basic_reprint.pdf.

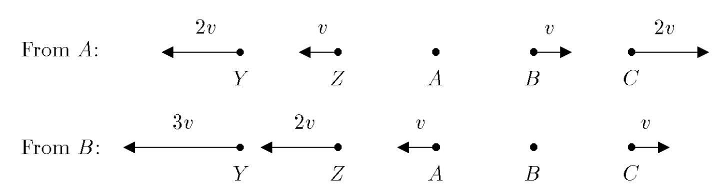
THE HOMOGENEOUSLY EXPANDING UNIVERSE:

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

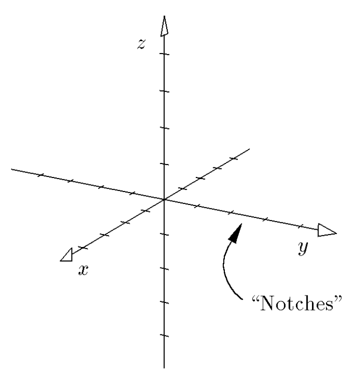
At first thought, one might think that the concept of homogeneity is inconsistent with Hubble's law— if the universe is expanding, there must be a unique point which is at rest. This argument would be valid ***if*** there were some physical way of telling if an object is at rest. However, the basic principle 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. Thus, an object which is at rest in one inertial frame will be moving in another reference frame, and there is no meaning to being absolutely at rest. While special relativity dates from 1905, the basic principle discussed above is in fact a property of Newtonian mechanics as well. (The principle was temporarily abandoned, however, in the 19th century, when the ether was introduced in the description of electromagnetism.)

To see how this works, it is easiest to begin with a one dimensional example. To this end, we will borrow a diagram from Steven Weinberg's book:



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 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 saying that one can draw a map of the universe that does not change with time; the expansion of the universe can be incorporated entirely into the time-variation of the scale factor of the map. At one time a unit distance on the map might correspond to a million light-years, and at a later time a unit distance might correspond to one and a half million light-years. This concept of a map with a time-varying scale factor can be made mathematically concrete by introducing what is called a comoving coordinate system. This coordinate system can be thought of as a coordinate system for the time-independent map, and so the coordinates x, y, and z of a typical galaxy, moving with the Hubble expansion, will be independent of time. The coordinates will be measured in arbitrary units which I will call "notches":



The expansion is described by a scale factor $a(t)$, which gives the physical distance that corresponds to 1 notch at any time t.* $a(t)$ might be measured, for

* Warning: if you look back at versions of the 8.286 notes from earlier years, the scale factor was called $R(t)$. $R(t)$ was the original notation used by Friedmann and Lemaître, and I learned it from Steven Weinberg's *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity,* 1972. Nowadays,

example, in units such as m/notch. The physical distance between any two points at any given time is then given by

$$\boxed{\ell_p(t) = a(t)\ell_c\ .} \tag{3.4}$$

Here ℓ_c denotes the coordinate distance between the two points, measured in notches and independent of time, and ℓ_p denotes the time-dependent physical distance.

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.

We are now in a position to see how the homogeneous expansion implied by Eq. (3.4) leads directly to Hubble's law. To see this, one simply differentiates Eq. (3.4) in order to find the velocities. If ℓ_p denotes the distance between a particular distant galaxy and us, then the recession velocity of that galaxy is given by

$$v = \frac{\mathrm{d}\ell_p}{\mathrm{d}t} = \frac{\mathrm{d}a}{\mathrm{d}t}\ell_c = \left[\frac{1}{a(t)}\ \frac{\mathrm{d}a}{\mathrm{d}t}\right] a(t)\ell_c\ . \tag{3.5}$$

Note that this can be rewritten as

$$v = \frac{\mathrm{d}\ell_p}{\mathrm{d}t} = H\ell_p\ , \tag{3.6}$$

where $H(t)$ is given by

$$\boxed{H(t) = \frac{1}{a(t)}\ \frac{\mathrm{d}a}{\mathrm{d}t}\ .} \tag{3.7}$$

Thus we have not only derived Hubble's law, but Eq. (3.7) gives us an expression for the Hubble expansion rate $H(t)$.

however, almost all authors use $a(t)$, motivated by the fact that the symbol R is used in general relativity to denote a quantity called the curvature scalar. As I edit the notes for this year, beware that an occasional $R(t)$ might slip through.

MOTION OF LIGHT RAYS:

We will of course be interested in tracing the path of light rays through the universe, using the comoving coordinate system. The rule is very simple— light travels at the standard speed c, given by 3.0×10^8 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 $c/a(t)$:

$$\boxed{\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{c}{a(t)}\ .} \tag{3.8}$$

That is, if we use square brackets $[A]$ to denote the units of A, then

$$\left[\frac{c}{a(t)}\right] = \frac{\mathrm{m/s}}{\mathrm{m/notch}} = \frac{\mathrm{notch}}{\mathrm{s}}\ , \tag{3.9}$$

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

The above explanation was more intuitive than rigorous, but a complete formulation of electromagnetic theory in the context of general relativity gives precisely Eq. (3.8).

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. Similarly, we can imagine another race of creatures living in the galaxy M81, who measure a according to their own clocks. We will assume that communication is possible, but it is not practical to use signals to synchronize the clocks, since the signals travel at the speed of light, and the

distance changes with time and is not known. 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— lengths can be measured relative to the wavelength of some sharp line in an atomic spectrum, and similarly time can be measured in the units of some atomic frequency. 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 variation 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.

Note that special relativity still tells us that the synchronization of clocks is not unique, so one must be careful. Clocks that are synchronized by cosmic time will not necessarily appear synchronized by other criteria that one might invoke. For example, suppose we established an inertial frame of reference determined by a network of bases which does not expand with the universe, but instead remains at rest relative to us. In such a frame of reference the clocks on M81 would appear to be moving and hence running slower. These clocks would therefore not remain synchronized with the clocks in the inertial reference frame. Thus, the comoving coordinate system is ***NOT*** an inertial reference frame.

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 it 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? Perhaps it does. 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 model, which we will come to at the end of this course, is characterized by a phase in which the universe is accurately described by a de Sitter space.

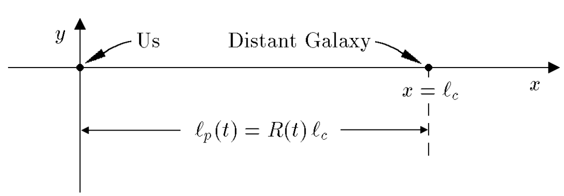
To summarize the important point— if you got lost, just remember this: the time variable t that we are using is called cosmic time, and any observer at rest relative to the galaxies in his vicinity can measure it on his own clock. The clocks throughout the universe can be synchronized by using the Hubble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

Note that the existence of cosmic time makes the universe a simpler spacetime than Minkowski spacetime, the spacetime of special relativity. 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 special relativity, by contrast, the definition of simultaneity depends on the velocity of the observer. (De Sitter space, mentioned above, is much more like Minkowski space.)

THE COSMOLOGICAL REDSHIFT:

Suppose an atom on a distant galaxy is emitting light pulses at a fixed time interval Δt_e ("e" for "emit"). We will receive these pulses at a Doppler shifted interval, which we will call Δt_o ("o" for "observe"). Our goal is to relate the Doppler shift to the behavior of the scale factor $a(t)$.

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:



Let t_e be the cosmic time at which the first pulse is emitted from the distant galaxy, with the second pulse emitted at $t_e + \Delta t_e$. 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 noninertial comoving coordinate system, and the cosmic time of the coordinate system is by defintiion the time as read on the clocks.)

The next step is to understand the relationship between the time interval of emission Δt_e and the time interval of observation Δt_o. Note that after the first pulse is emitted, it travels a physical distance $\lambda_e \equiv c\Delta t_e$ before the second pulse is emitted. If Δt_e is the time between the emission of wave crests, then $\lambda_e \equiv c\Delta t_e$ is the wavelength of the emitted wave. The two pulses are then separated by a coordinate distance $\Delta x = \lambda_e / a(t_e)$. We assume that the period of the wave Δt_e is very short compared to the time scale on which $a(t)$ varies, so it does not matter whether the denominator is written as $a(t_e)$ or $a(t_e + \Delta t_e)$. According to Eq. (3.8), the velocity of light in these coordinates depends on t, but is independent of spatial position. Thus, at any given time the two pulses will travel at the same coordinate velocity $\mathrm{d}x/\mathrm{d}t$, and thus will stay the same distance apart. When they arrive at the observer they will still be separated by the same coordinate distance Δx. The physical separation at the observer will then be given by

$$\lambda_o = a(t_o)\Delta x = \frac{a(t_o)}{a(t_e)}\,\lambda_e \; , \tag{3.10}$$

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

$$\Delta t_o = \frac{\lambda_o}{c} = \frac{a(t_o)}{a(t_e)}\Delta t_e \; . \tag{3.11}$$

Finally, one has

$$\boxed{1 + z \equiv \frac{\Delta t_o}{\Delta t_e} = \frac{a(t_o)}{a(t_e)} \; .} \tag{3.12}$$

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. In fact, Eq. (3.12) is a rigorous consequence of Eq. (3.8) and the construction of the comoving coordinate system. Thus, Eq. (3.12) is an exact result of general relativity, which includes the effects of both special relativity and gravity. It is possible to apply Eq. (3.12) 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. (3.12) differs from the special relativity result in two ways:

(1) The special relativity result holds exactly only in the absence of gravity, while Eq. (3.12) includes the effects of gravity— provided, of course, that one knows the effects of gravity on the scale factor $a(t)$.

(2) Eq. (3.12) expresses the Doppler shift in terms of the behavior of the scale factor $a(t)$ for 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. (3.12) 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.

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 motion by asking what the Doppler shift ***would*** be 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, so it is hard to draw any comparisons between the two situations. Alternatively, you might suggest that we use the special relativity result to determine the Doppler shift caused solely by motion, and then attribute the balance of the Doppler shift as given in Eq. (3.12) to gravity. However, the problem is that in the presence of gravity there is no inertial coordinate system, and thus one has no natural way of defining the v which appears in the special relativity formula.