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 $\Delta T/T = 4.14\times10^{-5}$.

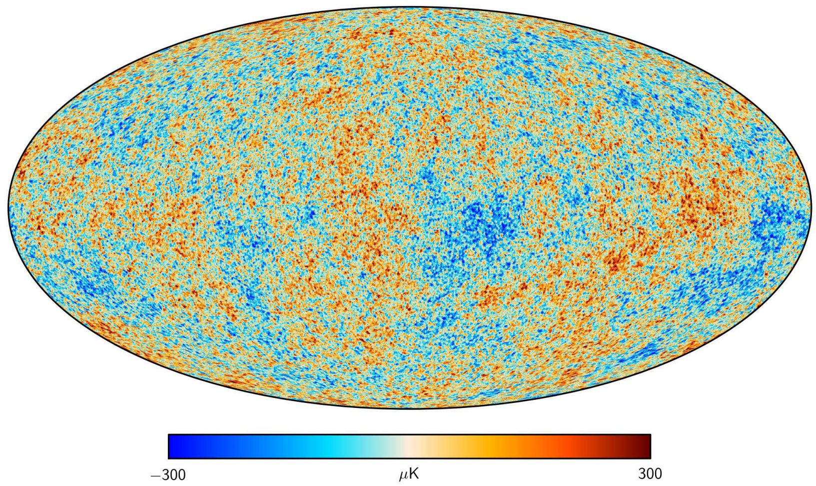

Figure 2.1: The cosmic microwave background radiation as detected by the *Planck* satellite, from the 2015 data release. After correcting for the motion of the Earth, the temperature of the radiation is nearly uniform across the entire sky, with average temperature $T_{\rm cmb} = 2.726$ K. Tiny deviations 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),

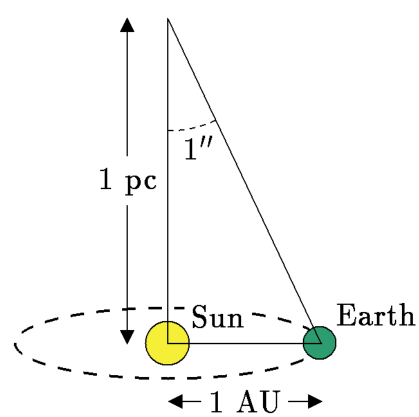

Figure 2.2

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

---

* *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 $1/(10^{10}\ \text{yr})$ 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.

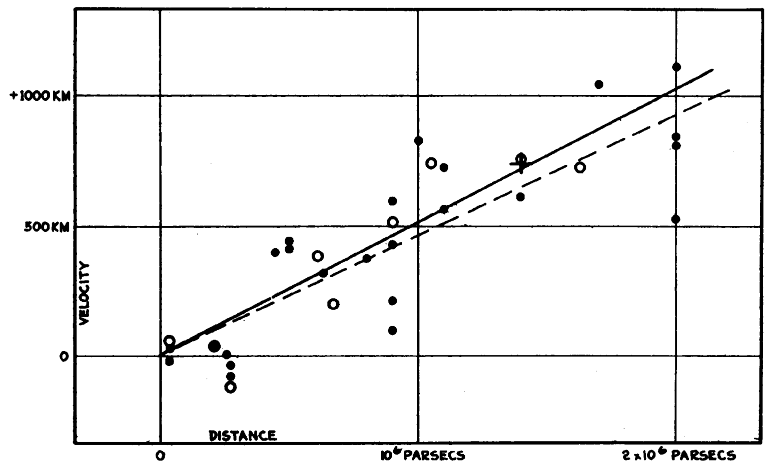

Figure 2.3: Edwin Hubble's original data, published in 1929, which introduced the first observational evidence for Hubble's law and the expansion of the universe.

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:

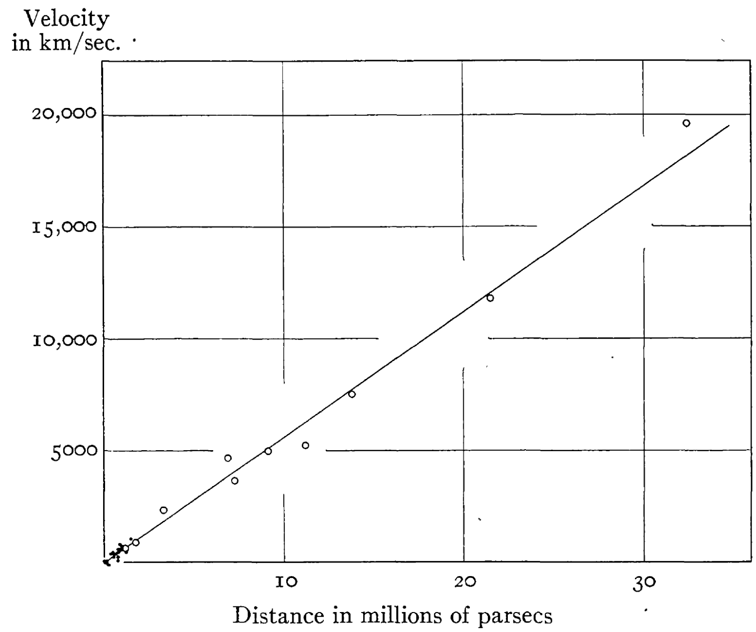

Figure 2.4: Data published by Edwin Hubble and Milton Humason in 1931*, extending Hubble's original measurements to significantly greater distances.

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.

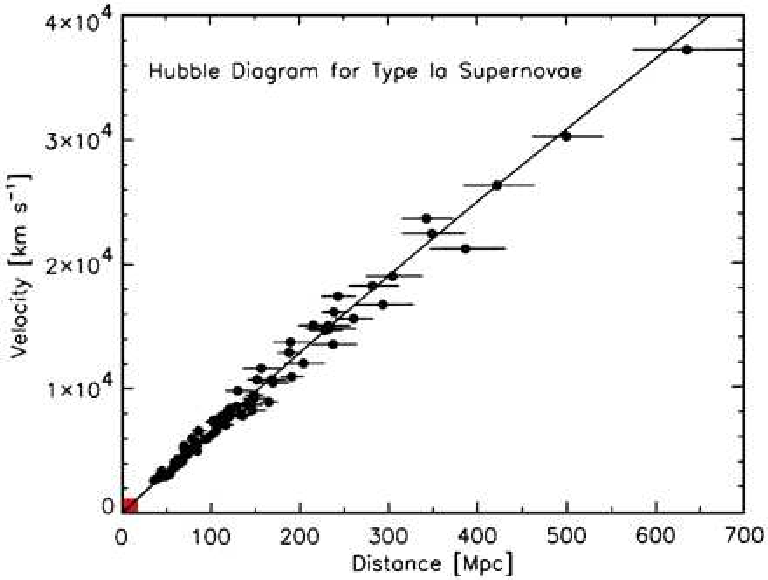

Figure 2.5: An extension of the Hubble diagram, showing observations up to 2002 of Type 1a supernovae. Error bars correspond to uncertainties in determining distances to each object. The small red box near the origin indicates the range covered in Hubble's original plot.

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

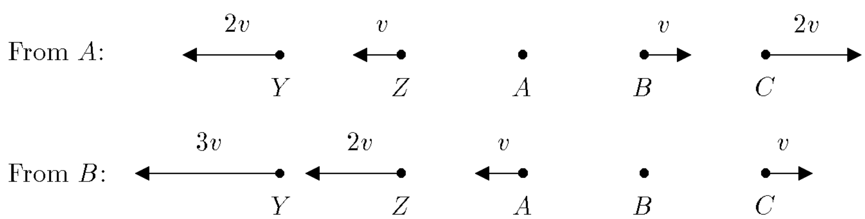

Figure 2.6: Hubble's Law is compatible with homogeneity in space. Each observer can consider herself at rest, and will observe other points moving away from her at speeds proportional to their distance from her.

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:

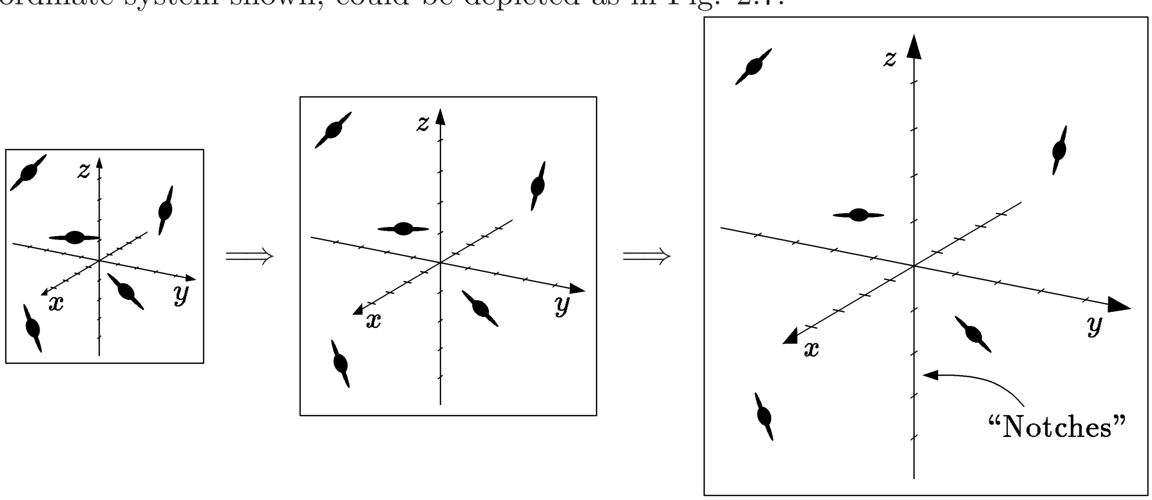

Figure 2.7: By employing "comoving coordinates," a single map can represent the locations of objects in an expanding universe. Distances between objects on the map are measured in "notches," while the relation between notches and physical units (such as centimeters or light-years) changes over time.

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

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

Note that this can be rewritten as

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

where $H(t)$ is given by

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

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 $c/a(t)$:

$$\boxed{\frac{dx}{dt} = \frac{c}{a(t)} \ .} \tag{2.9}$$

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

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

which gives the right units for dx/dt, 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:

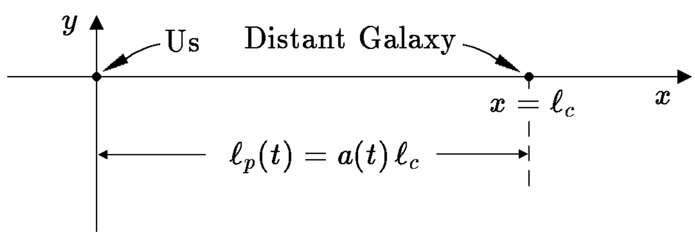

Figure 2.8: Diagram for discussing the transmission of a light signal from a distant galaxy to us. We are at the origin, and the galaxy is along the x-axis, at $x = \ell_c$. The light signal travels to us along the x-axis.

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.