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.)

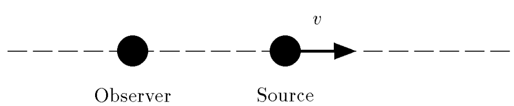
The Doppler shift formula for light requires special relativity, which is not a prerequisite for this course. For this course it will be sufficient for you to know the basic consequences of special relativity, which will be stated in these notes. If you would like to learn more about special relativity, 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.

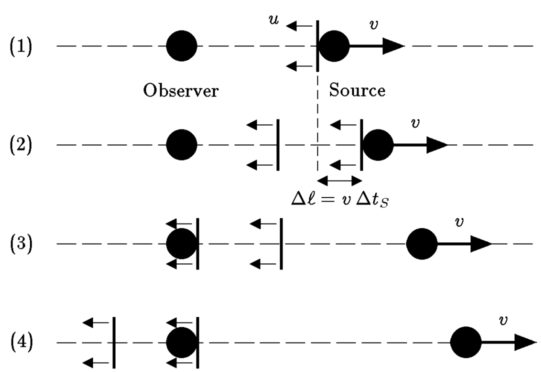
As long as the velocities of the stars in question are small compared to that of light, it is sufficient to use a nonrelativistic analysis. We will begin with the nonrelativistic case, and afterward we will discuss how the calculation is changed by the implications of special relativity. To keep the 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

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

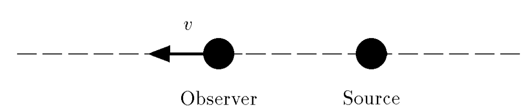
The result is usually described in terms of the "redshift" z, which is defined by the statement that the wavelength is increased by a factor of $(1+z)$. Since the wavelength λ is related to the period Δt by $\lambda = u\Delta t$, we can write the definition of redshift as

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

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

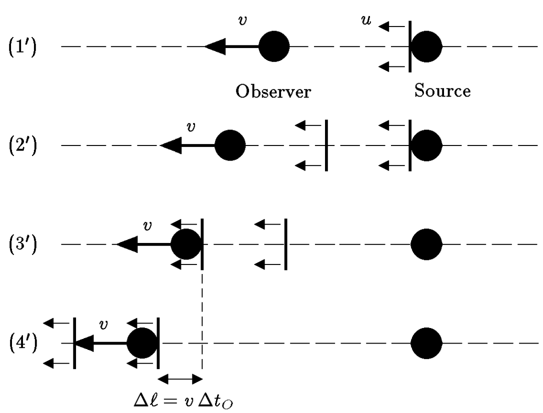
$$\boxed{z = v/u \qquad \text{(nonrelativistic, source moving).}} \tag{1.5}$$

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,

$$z = \frac{\Delta t_O}{\Delta t_S} - 1 = \frac{1}{1-(v/u)} - 1$$

$$= \boxed{\frac{v/u}{1-(v/u)} \qquad \text{(nonrelativistic, observer moving).}} \tag{1.8}$$

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.

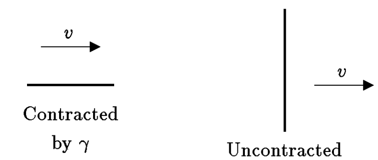
SUMMARY OF SPECIAL RELATIVITY:

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.

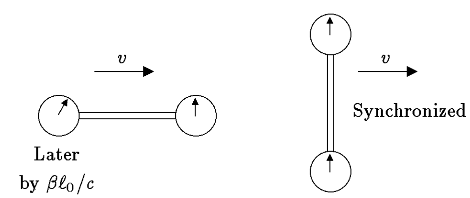
(1) TIME DILATION: Any clock which is moving at speed v relative to a given reference frame will "appear" (to an observer using that reference frame) to run slower than normal by a factor denoted by the Greek letter γ (gamma), and given by

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

(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

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

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.