Cosmology: the "standard model"

Let's start by asking what kind of picture of the universe we might form had we never heard of Einstein and general relativity. We start from a general observation which will play a fundamental role also in post-Einsteinian thinking, namely that as far as we can tell, provided we go to a sufficiently large scale,

(a) the matter in the Universe seems to be roughly constant in density, and

(b) there appears to be nothing special about our own position. (or is this, in fact, a prejudice rather than an experimental fact?).

This gives rise to at least two obvious problems:

(1) Olbers' paradox: light from a source a distance R from us falls off as $1/R^2$ but the number of sources within a shell of radius R increases as R^3. Hence the total light from all sources within a radius R increases as R and tends to infinity as $R \to \infty$. If we assume the light is absorbed in route, then we have a problem with heating. So, why is it dark at night? Only solution: R must be finite!

(2) Newton's dilemma: If gravity continues to operate on the scale of the cosmos, then the Universe must be unstable against gravitational collapse (as is any sphere of finite radius). Two possible solution:

(a) the Universe is in fact infinite, or

(b) it really is unstable and hence time-dependent.

The second solution was as intuitively repugnant to pre-Einsteinian thought as it was originally to Einstein himself.

In fact, we now believe that a correct description of the Universe has to be based on the general concepts of general relativity and must therefore allow for the possibility that space-time is curved, even if we do not believe all the details of Einstein's theory. Although the curvature may be (and almost certainly is!) too small to be easily observable on the scale of the earth, or even the solar system, it may well dominate the behavior of the Universe in the large.

In applying the concepts of general relativity to the Universe as a whole, we first ask what general symmetry arguments we can apply and what they will buy us. One very tempting prejudice is what is called the "perfect cosmological principle" (PCP): The Universe looks the same (on a sufficiently gross scale, of course) as viewed from any poinf in any direction and *at any time*. It turns out that this assumption places extremely strong constraints on the geometry of space-time and in fact allows only a unique kind of space, known technically as De Sitter space; this space is the basis of the so-called steady-state theory of Hoyle, Gold and Bondi. Unfortunately, the general opinion is that the PCP, and hence a fortiori the steady-state theory, has been refuted directly by experiment (cf. below).

The "next best" assumption we can make is the so-called cosmological principle: if we choose our time and space reference frame appropriately, then the Universe at any given time looks the same as viewed from any point in space and in any direction. What is the " appropriate" choice of reference frame? (recall that in GR as such we are entitled to use any frame whatsoever!) Evidently it has to be one in which the local matter of the Universe (averaged over local fluctuations, etc.) is at rest (if nof by definition the Universe would look different in the direction the matter was moving!) So we choose our reference frame to be that of an observer who is at rest with respect to the (averaged) local matter of the Universe. The time measured by such an observer is called "cosmic standard time" (how exactly we define it is of no great importance, provided it does not distinguish one observer from another; e.g. we could define it from the temperature of the blackbody radiation background, see below). Note that in a sense we almost appear to have smuggled back the "ether"! However, the appearance of a "privileged" frame is illusory: although the Universe as a whole certainly looks simpler when viewed from the cosmic reference frame, the basic laws of physics still conform to the postulates of special and general relativity and are valid in any frame.

We still have to choose the coordinates for our space reference frame. Various choices are possible, but the simplest choice is so that *the coordinates of each piece of matter of the Universe are independent of time.* Note carefully that this does not mean that the distance between different galaxies (etc.) is necessarily independent of time (cf. below): it is analogous to defining the coordinates of points marked on a balloon to be relative to the balloon as a whole, the distance between the points still changes as the balloon is blown up. If we make this choice, then it turns out that at any given value of cosmic standard time t the geometry of the Universe is rather simple; it is parametrized, first, by a number k in which is fixed for all time (in all reasonable models!) and can take values $+1$, 0 or -1, and tells us about the overall form of geometrical structure (see below), and secondly by a so-called "scale factor" $R(t)$ which in general may be a function of t. In the special case that $R(t)$ is time-independent it can clearly be absorbed into our definition of the unit of length, but in the (realistic) case where it depends on time it essentially tells us how the average matter in the Universe is expanding (or contracting) in time. That is, if R increases by a factor of 2 this is equivalent to saying that the distance between a pair of neighboring galaxies has doubled, as measured in terms of some standard noncosmological unit such as the wavelength of a particular spectral line (or the centimeter, which is based on this). The question obviously arises, whether this expansion (contraction) is "real", or just a consequence of our arbitrary-looking convention for the coordinate system: we return to this question in the context of Hubble's law.

The quantity k in the expression ("RW metric") for the geometry of the Universe has the following significance: if $k = 0$, then the (space) geometry of the Universe in the large is flat, i.e. it is just that of ordinary 3D Euclidean space. Obviously, in this case the Universe is infinite in spatial extent; such a universe is called "flat". If $k = +1$, the geometry is the 3D generalization of a sphere; that is, if we go far enough in any direction we will return to our starting point. The "radius" of the Universe at any given

time t is just the scale factor $R(t)$, and its volume is finite and of order $R^3(t)$. This is called a "closed" Universe. Finally, if $k = -1$ we have an "open" Universe, which is formally a space of constant *negative* curvature; this is a beast which is very difficult to visualize intuitively, but it may be thought of as somewhat like the behavior of a 2D surface near a saddle point. In this case the scale factor $R(t)$ cannot be interpreted as a "radius", but is in some sense a measure of the (inverse) curvature. The three above possibilities are the *only* ones which are allowed by the cosmological principle.

It should be emphasized that everything said so far is a consequence of quite general considerations of geometry and symmetry: we have nowhere invoked any considerations which are specific to GR or any other theory of gravity. However, to fix the value of the characteristic constant k, and also the all-important time dependence of the scale factor $R(t)$, we have to rely on a specific theory. The class of universes which follow if we combine the cosmological principle with Einstein's GR are called Friedmann universes. It turns out that in such universes application of Einstein's original equations expressing the way in which the presence of mass curves (4D) space-time leads to the following results for the "dynamics" of the scale factor $R(t)$:

$$\left(\frac{dR}{dt}\right)^2 = \frac{8\pi G}{3}\rho(t)R^2(t) - k \qquad (*)$$

In this equation G is the Cavendish (universal gravitational) constant, $R(t)$ is the scale factor of the Universe at time t as above, and $\rho(t)$ is the energy density, which for current conditions (where most of the matter in the Universe is nonrelativistic) is essentially the mass density and hence is proportional to $R^3(t)$. Quite generally, it can be shown that ρ falls off with R faster than $1/R^2$, as a result of which the RHS of the equation can be zero for at most a single value of R; a little further analysis shows* that even in this case d^2R/dt^2 must be nonzero. Hence at least one of the quantities dR/dt and d^2R/dt^2 must be nonzero at any given time. *Thus the original Einstein equations do not tolerate a solution in which the geometry of the Universe is independent of time.* ($R(t) = \text{const.}$).

This conclusion upset Einstein to an extreme degree, since like almost all of his contemporaries he had been brought up to think of the Universe as eternal and unchanging. In fact, it worried him so much that he proposed to remedy the apparent defect of his model by adding to his original equations an extra term containing the so-called cosmological constant. The value of the latter is so tiny that it has no observable effect within the solar system (or even at the level of the binary pulsar, etc.) and hence does not spoil the agreement of the original model with observations, but on the scale of the Universe as a whole the effect of the extra term is (with a suitable choice of the constant) to allow the RHS of the equation to be zero and hence R to be independent of time. In his later (post-Hubble) years Einstein not only rejected the term with the cosmological constant but described its addition as "the worst mistake of my life"; today, even though the original motivation for introducing it has vanished, some people continue to favor its

*This does not follow simply from the above equation itself (which for $k = +1$ is clearly satisfied by $R = 3/8\pi G\rho$, $R(t)$ and $\rho(t)$ independent of t); we need to go back to the analysis which led to it.

addition to "standard" GR in order to solve other technical problems. (In particular, it comes back with a vengeance in the context of "inflation").

Why has the original motivation vanished? Because nowadays, following Hubble's revolutionary discoveries in the early 30's, almost everyone accepts that the Universe is not static but is in fact expanding. Hubble's raw data consisted in the measurement of the frequency of spectral lines emitted by stars which he had already shown (to most people's satisfaction) were in galaxies outside our own. On identifying those lines with those emitted in known atomic transitions have on earth, he found that the frequencies were systematically less than those of the corresponding lines on earth – by an amount which appeared to be roughly proportional to the distance of the galaxy in question from us (which can be independently estimated). This is the famous "redshift". Hubble's redshifts were a small fraction of the original frequency, but in recent years those measurements have been extended to very much more distant objects, including quasars believed to be billions of light years away, which have redshifts of more than 4 (so that, for example, the H_α line, originally in the ultraviolet, is shifted down into the visible spectrum).

The standard interpretation of the redshifts measured by Hubble and subsequent workers is as a Doppler effect. Recall that the sound emitted by the siren of an ambulance moving away from us (towards us) is lower (higher) than when the vehicle is at rest, and that the same is qualitatively true for light even in special relativity. Hence, if the redshift really is a Doppler effecf we can read off from it more or less directly the velocity of the galaxy in question along the "line of sight". Hubble's fundamental result, based on this interpretation is that once local fluctuations are removed the galaxies in our neighborhood are moving away from us, with a velocity which is proportional to their distance from us:

$$v = H_0 r$$

where H_0, the Hubble constant, is (as now believed) approximately 70 km/sec/Mpc, or expressed as a time, $H_0^{-1} \approx 10^{10}\,\mathrm{yr}$. (We will see below what is the significance of this so-called "Hubble time"). In recent decades it has been confirmed that more distant sources also obey this law, at least approximately. The above formula is exactly what we would expect if the scale factor of the Universe, $R(t)$, is increasing at a rate given by the rate $\dot{R}/R = H_0$, in fact apart from (the same) constant $R(t)$ is just the distance of the source from us and $\dot{R}(t)$ the velocify. Note that according to the cosmological principle, all observers who are at rest relative to the local matter of the Universe should see the Hubble expansion no matter where they are: the analogy often used to illustrate this is of a rubber sheet with various points marked on it which is stretched uniformly - the distance between any two points will increase at a rate proportional to their current distance.

The interpretation of the raw data (the redshifts from distant galaxies or quasars) as indicating the expansion of the Universe is so fundamental an element in our current picture of the Universe that it is important to take a moment to think about possible loopholes in the argument leading to that conclusion. The first has to do with the interpretation of the redshift as a Doppler effect. In fact, at least one alternative origin

of redshifts is known in GR, namely the so-called gravitational redshift (cf. lecture 15). To interpret the data in this way we would need to suppose that we are starting at a point of exceptionally high gravitational potential and that neighboring galaxies are systematically at an appreciably lower value. Although this explanation cannot be totally excluded a priori, it seems to violate the spirit of the cosmological principle, in that we would be occupying in some sense a specially privileged position in the Universe, and in addition it seems difficult to reconcile the postulated variation of the gravitational potential with what we know about the distribution of matter in our neighborhood. (An alternative scenario, in which there is nothing special about the other galaxies as such but the stars on which we are observing the redshift are exceptionally compact and hence have large redshifts, seems to put even more weight on coincidence. As to other, currently unknown, sources of redshifting (such as a systematic dependence of the electron charge or mass on position on a cosmic scale), all one can say is that there is no independent evidence for it and hence the principle of Occam's razor would suggest we reject it.

Even supposing that the redshift is indeed a Doppler effect as generally believed, in order to interpret it as indicating an expansion of the Universe we need to be sure that it really is proportional to the distance of the source from us. The question of how we know the distance of an astronomical source from us is itself a very vexed one and is the subject of whole books. In effect, most of what we (think we) know is to one extent or the other dependent on the idea of a "fixed candle": that is, from investigation of stars close enough to us for us to be able to ascertain their distance directly (e.g. by parallax arising from the earth's motion around the sun) we identify particular kinds of star whose absolute brightness ("luminosity") we can correlate with other properties (e.g. in the case of the so-called Cepheid variables, the period of oscillation); then when we see these types of stars further away, we assume that the same relation holds and use it to evaluate their absolute brightness, and hence, from their apparent brightness, their distance from us. We can then iterate this procedure by finding other, more exotic, objects in the distant galaxies in question and using them, in turn, as "standard candles": and so on right out to the most distant objects correctly accessible to our observation. Unfortunately the whole procedure is highly error-prone (which is the main reason for the large current uncertainty in the value of the Hubble constant): because of the sequential nature of the argument, any effort in the estimation of the closest objects to us will propagate out to that of the most distant quasars. In fact, there is a (currently small) minority of astronomers who believe that the observed quasars are actually not at cosmological distances at all but may even lie within our own galaxy; if this were to turn out to be correct, then of course the redshifts from them must have a totally different origin from that usually believed. From now on I shall assume that the "establishment" view is in fact correct, i.e. that the time-variation of the scale factor $R(t)$ inferred from the redshift data is indeed real.

Even given this conclusion, one can of course still ask whether the interpretation of a time variation of $R(t)$ as indicating a physical expansion of the Universe is necessary. Could it not turn out to be merely a matter of our convention for the units of distance,

etc.? The answer, it seems, is that as so often in these matters we have a choice. Either we can make the conventional interpretation, in which we take the speed of light to be constant for an inertial observer by definition and hence (given that we have fixed a unit of time) define the unit of length in terms of the wavelength of a standard spectral line; then the expansion is a real effect, i.e. the distance (measured in units of the wavelength) between neighboring galaxies, really is increasing with time in the way Hubble inferred. Or we can choose our unit of length so that by definition the universe is static (the distance between neighboring galaxies, modulo local fluctuations, is independent of time); but in that case we cannot simultaneously take the wavelength of any spectral line to be independent of time, and given that we have fixed its frequency in terms of some independently defined unit (such as the rate of change of the local CBR temperature, cf. below) the speed of light would therefore also depend on time even for an inertial observer. The situation is reminiscent of the choice between SR and the Lorentz-Fitzerald contraction/time dilation scenario: while the experimental predictions are the same for the two interpretations, the "standard" one is by far the simpler!

Let's now return to our fundamental equation and define the quantify

$$\Omega(t) = \frac{8\pi G\rho(t)R^2(t)}{(dR/dt)^2}$$

While we can discuss this quantity for arbitrary times, the most important conclusions relate to its value "now", which we will call Ω_0. Since the current value of $(dR/dt)/R$ is, according to the standard interpretation just the value H_0 of the Hubble constant, we can write (ρ_0 = mean density of matter (energy) "now")

$$\Omega_0 \equiv \frac{8\pi G\rho_0}{H_0^2}$$

and thus our equation $(*)$ can be written

$$k/(H_0^2R_0^2) = \Omega_0 - 1$$

We see immediately that if $\Omega_0 = 1$ exactly, then $k = 0$ ("flat" Universe); if $\Omega_0 < 1$ then $k = -1$ ("open" Universe), and $\Omega_0 > 1$ then $k = +1$ ("closed" Universe). Furthermore, in the last case, if Ω_0 is not too different from 1 in order of magnitude then the current "radius" of the Universe is of the order of the "Hubble time" H_0^{-1} (times c, in usual units) i.e. it is of the order of 10 billion light years. A similar conclusion follows, for the "open" Universe, for the current scale factor R_0. For a "flat" universe ($k = 0$) the quantify R_0 itself really has no physical meaning (though the quantity H_0 ($= (dR/dT)/R$ "now") of course does); this is not surprising since there is no "characteristic scale" of flat (Euclidean) space.

Friedmann universes, be they flat, open or closed, have one important feature in common. Assuming that the Universe is indeed now expanding we can integrate the equations for $R(t)$ backwards in time and, given any relation between the energy density $\rho(t)$ and the scale factor $R(t)$ which is physically reasonable (see below) we reach the

conclusion that there exists a point some finite time backward in the past at which the scale factor was zero and the energy density (and hence the temperature) infinite. This is the scenario which has become known as the "hot big bang". In any Friedmann model, while the exact value of the "age of the Universe" (i.e. the time elapsed since the hot big bang) depends somewhat on the density, its order of magnitude is always the Hubble time. This, we conclude that within a factor of 2 or so the Universe is approximately 10 billion years old, the exact value depending on the parameters H_0 and (weakly) on Ω_0 and k. It is interesting that this estimated age is comparable to the age of the oldest stars as inferred from the best available astrophysical theory; indeed, it appears that the lower-end estimates of H_0 may actually risk making the Universe younger than its oldest stars! It should be emphasized that uncertainties about the behavior of matter at temperatures and densities very different from those encountered in terrestrial laboratories affect only the first few seconds or minutes of the big bang scenario; thus, the general existence of a big bang, and its data in time, are almost completely insensitive to these unknowns. It might also be asked whether the existence of a "big bang" is simply a pathology of the Friedmann model, and whether, for example, relaxing the cosmological principle so as to allow substantial fluctuations of the matter density on the scale of the "size" (to the extent that it can be defined!) of the Universe might enable one to avoid this conclusion? A classic piece of analysis by Penrose and Hawking around 1970 answered this question definitively in the negative: given only that the Universe is currently expanding, that GR applies at all relevant times and that the relation between energy and mass density has any physically reasonable form, then the hot big bang is inevitable, irrespective of the current degree of inhomogeneity.

So far, everything we have said has depended on pure theory augmented only by Hubble's law. One might reasonably ask: Is there independent experimental evidence for the hot big bang scenario? In response to this question the conventional viewpoint cites three pieces of evidence: the abundance in the Universe of the different isotopes of the light elements, the cosmic background radiation (CBR) and the existence and structure of the galaxies. The reason that the isotopic abundance of light elements is relevant is that, if one extrapolates the laws of particle physics as known from laboratory experiments back to the early Universe (a few minutes after the HBB) one concludes that such elements would have been formed at that time by combination of what were originally free protons and neutrons, and the fraction of each isotope formed can be calculated rather exactly, with essentially no unknowns other than the value of Ω_0 and a related quantity(see below) (which anyway affects the answer only weakly). The comparison of the experimental ratios with the theory is somewhat complicated by the fact that some elements such as He could also have been produced later in the interior of stars, but it is thought one knows how to allow for that, and the general belief is that the agreement between theory and experiment is extremely impressive and (since it is certainly difficult to think up an alternative scenario which would have produced such a good quantitative fit!) leads strong support to the HBB model, at least as it deals with events at a time a few minutes after the "bang" itself.

The second piece of evidence is the cosmic background radiation (CBR), discovered in 1965 by a couple of Bell Labs researchers who were trying to develop microwave detectors, presumably for industrial or military purposes. It had in fact been predicted earlier that if the HBB scenario were correct, a "relic" of it should be left in the form of a uniform back ground radiation with the so-called black body (thermal) spectrum – precisely what was found. (The experimentally observed temperature of the CBR is about 2.7K) This radiation basically tells us about the state of the Universe at the point where the temperature has dropped sufficiently for electrons and protons to combine and form atomic hydrogen; before that the Universe was opaque to radiation because of the strong scattering of the electrons). This is about 100000 years after the HBB, so much later then the "nucleosynthesis" period. The final piece of evidence usually cited has to do with galactic structure, but it has to be said that this (a) refers to a considerably later period ($\sim 3 \times 10^8$ yrs),[†] and (b) rests on much less secure theoretical foundations than the other two.

We have seen that the past history of the Universe, and in particular the occurrence of the HBB, is only fairly weakly sensitive to the value of k (i.e. to whether it is open, flat or closed). With respect to the future things are quite different. Going back to our fundamental equation

$$\left(\frac{dR}{dt}\right)^2 = \frac{8\pi G}{3}\rho(t)R^2(t) - k \qquad (*)$$

we see that if either $k = -1$ (open Universe) or $k = 0$ (flat) the RHS can never become zero, so the Universe will continue to expand forever (though in the flat case the rate of expansion will tend to zero as $t \to \infty$, since $\rho(t) \sim R^{-3}(t)$). On the other hand, if the Universe is closed ($k = +1$) then since the quantify $\frac{8\pi G}{3}\rho(t)R^2(t)$ is decreasing and has a current value $\Omega_0 > 1$, the RHS not only can but must become zero at some time in the future. At this point the expansion reverses and the scale factor $R(t)$ retraces its former path in the backward direction i.e. we get a contraction and the Universe heats up again. Assuming that no new laws of physics come into play at this stage, we expect that the reversal is complete and the original big bang is "reflected" in a Hot Big Crunch.

Thus, if we wish to predict our long-term future it is essential to know the value of k, or equivalently that of the density parameter Ω_0 (since we know that $\Omega_0 >, =, <$ corresponds respectively to $k = -1, 0, +1$). The most obvious way of doing this is to measure the parameters H_0 and ρ_0 directly (G is known to high accuracy). The estimated uncertainty in H_0 is currently about 5%; unfortunately, the uncertainty in ρ, the current mass (or more strictly energy) density of the Universe, is much greater. We can obtain a lower limit on ρ from the observed matter we see in stars, intergalactic space etc. – the so-called "luminous" matter; putting this into the expression for Ω_0 gives $\Omega_0 > 0.005$. However, there are strong suspicions that then must be a large amount of "dark" (non-luminous) matter around, primarily because it seems impossible to understand the way in which the velocity of matter in the outer regions of galaxies like our own depends on radius without postulating such "dark matter". The existence and nature of this "dark

[†]Although the "seeds" may be produced at quite early times.

matter" is one of the major current problems in cosmology. If we assume enough of it to account for the observed rotational properties of galaxies the effect is to raise the lower limit on Ω_0 to $\sim 0.1 - 0.2$. It still only a lower limit because intergalactic space is so vast that it could contain an appreciable contribution to ρ without it giving easily detectable gravitational effects. The most obvious candidate is hydrogen, either atomic or ionized (i.e. free protons and electrons), and fairly tight limits can be put on this from the absence of appreciable absorption effects etc.; but there are all sorts of suggestions for "exotic" intergalactic dark matter (massive neutrinos, axioms etc.) which are much more difficult to exclude.

In principle, there are at least two other ways of determining Ω_0 apart from the direct measurement of ρ_0. First, in any Friedmann universe Ω_0 is simply related to the rate of change of the Hubble parameter $H(t)$ in time, and this is something which we might infer by observation of the redshift from sufficiently distant stellar objects. Unfortunately any detection of this parameter is made uncertain by our lack of knowledge of the history of the distant sources themselves. The second possibility is to use the fact that Ω_0 does enter, albeit weakly, the expressions for the production of light isotopes by cosmic nucleosynthesis in the early stages of the HBB. Because of uncertainties in some of the other parameters occurring in the expansions, this method of determination is also somewhat uncertain but for what it is worth suggests that $\Omega_0 > 1$. Thus at the end of the day all we can say with any degree of confidence is that Ω_0 is certainly greater than 0.005, probably greater than 0.1 and *may* be equal to or somewhat greater than 1.

Actually, this conclusion is itself remarkable, for the following reason. If the parameter $\Omega(t) \equiv 8\pi G\rho(t)/H^2(t)$ started off exactly equal to 1, then it will of course stay exactly equal to 1 at all time ("flat" Universe). If on the other hand it started off even very slightly different from 1, then it is easy to show from our basic equation that it would very rapidly move away from 1. To get a value as close to 1 at the present epoch as thg experiments suggest we would need to have started fantastically close to on at an early stage in the evolution of the Universe - something which would seem to be a pathological coincidence, if $\Omega_0 \neq 1$ exactly.

The relatively simple picture outlined above is complicated by two further considerations which have emerged in recent years. First, it has been appreciated that a number of puzzles concerning the present state of the Universe, and in particular the one mentioned in the last paragraph, can be resolved by postulating that at a very early stage in its history it underwent a period of very rapid (exponential) expansion, with $R(t)$ increasing by a factor of order 10^{30}. This is known as the "inflationary" scenario, and is embraced by most though not all cosmologists. The second complicating factor is that over the last ten years considerable evidence has surfaced which suggests that the expansion of the Universe is actually *accelerating*. (i.e. the Hubble parameter is increasing with time). This discovery was completely unexpected, and is simply inconsistent with the fundamental equation $(*)$ of the Friedmann cosmology as it stands. Ironically, the simplest way of accommodating the acceleration is by re-introducing a version of Einstein's cosmological constant, in the form of a "dark energy" which permeates space but is invisible to our standard probes. Whether this is indeed the solution, or whether

a more radical conceptual revolution is needed, remains to be seen.