The Cosmic Microwave Background and the Big Bang Theory of the Universe

1. Concepts from General Relativity

1.1 Curvature of space

As we discussed earlier, Einstein's equivalence principle states that gravity is equivalent to acceleration. That is, it is impossible to distinguish, by any experiment carried out inside an elevator, the difference between the elevator accelerating in a gravity-free region and the elevator standing still in a gravitational field. This allowed us to describe, for example, the bending of a light beam in a gravitational field and the difference between time observed at different places in a gravitational field. The prediction of light bending as it goes past the sun was tested in 1919 by Eddington. It proved correct, and Einstein's reputation was made. The effect of gravity on time has been tested and confirmed many times, and it is now an important part of the global positioning system.

Einstein, in his general theory of relativity, made an astonishing mental leap and completely changed our point of view. Instead of light bending, he proposed that light travels in straight lines (called geodesics) in a gravitational field and that space itself is curved, thus giving light the appearance of traveling on a curved path. In fact, other bodies moving under the influence of gravity also have their paths deflected by the curvature of space. The curvature actually takes place in 4-dimensional space-time, which makes it even more difficult to visualize, thus also accounting for the differences between clocks in a gravitational field.

The extreme example of curvature of space is the so-called black hole. The space around the black hole is so strongly curved that light, traveling along a "straight" geodesic in spacetime, actually travels along the surface, called the "event horizon," of the black hole, from inside of which nothing – even light – can escape. Moreover, at the event horizon a clock (by which I mean time itself) will actually stop dead. We now believe that we have seen black holes by the deflection of starlight around objects that otherwise are invisible, and we believe that at the center of some distant galaxies there are (or were) enormous black holes, equal to a billion solar masses. These black holes are (or were) sucking in more and more matter, matter that will never escape again. As the matter is sucked in, it gives off tremendous amounts of energy in the form of radio waves, and this death cry can be "heard" now as the radiation from so-called quasars.

1.2 The expanding universe

The concept of curved space also has profound consequences for the shape of the universe itself (no surprise here). Since the universe is composed of matter (and energy, which also has gravitational effects), the universe itself must be curved. But if this is so, Einstein recognized that the universe must be either expanding or contracting as it is pushed and pulled by gravity. However, in 1916, Einstein believed that the universe had been and would remain forever the same. Therefore, he included another term in his equations – the so-called cosmological constant – that would compensate for the matter and make the space of the universe "flat," instead of curved, at least on average over the entire (infinite) universe. Locally, of course, space is curved in the vicinity of stars and galaxies, for example. When Hubble discovered in 1929 that the universe is actually expanding, Einstein called his cosmological constant "his greatest mistake." But more of this later.

The curvature of space depends on the amount of matter that is present. In fact, the curvature can be positive, negative, or zero (flat). In two dimensions, the surface of a sphere has positive curvature, which results from the presence of matter. If we draw a circle on the surface of a sphere, the circumference is less that $2\pi r$, where r is the radius of the circle, as shown in the figure, and the interior angles of a triangle add up to more than π.

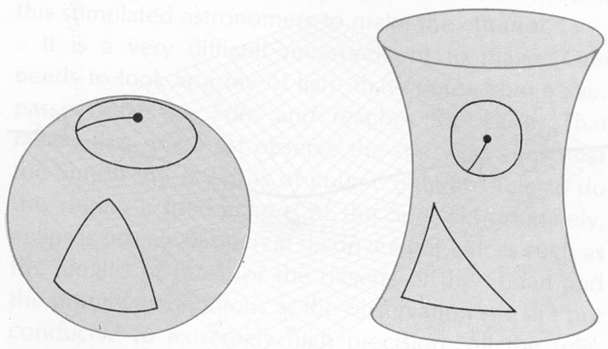


Fig. E.16 (left) The surface of a sphere is a curved two-dimensional surface.

Fig. E.17 (right) The surface of a hyperboloid is another curved two-dimensional surface.

A hyperboloid, also shown, is an example of a surface with negative curvature. The circumference of a circle is larger than $2\pi r$, and the sum of the angles in a triangle is less than π. Don't even try to visualize a 3-dimensional curved space, let alone a 4-dimensional curved spacetime. We can visualize the curvature of a 2-dimensional surface because it curves into the third dimension, as shown above. But a 3-dimensional space curves into the 4^{th} dimension and a 4-dimensional space curves into the 5^{th} dimension, whatever that means. The density of matter that is necessary to exactly cancel the curvature is called the "critical density."

Depending on the total mass (and energy) in the universe, the curvature of the universe can be positive, negative, or flat. If the overall density is greater than the critical density, the universe is curved enough that it closes on itself. A closed universe can be expanding or contracting, but it is finite, and after expanding for a (very long) while, the effect of gravity will eventually overcome the velocity of expansion and the universe will begin to collapse toward the "big crunch," followed perhaps, by another big bang. Don't ask what the universe is expanding into; this is a meaningless question. There is nothing there. Space – that is, the universe - is all there is. Space doesn't expand into anything. But at the beginning, the universe consisted of a single point (whatever that means), at least in Einstein's classical theory, and the universe has expanded from this "big bang" ever since. Ironically, the British astronomer Sir Fred Hoyle coined the term “big bang” to mock a theory that he believed, in 1950, couldn’t be true.

An open universe, on the other hand, has negative curvature, and extends to infinity. Nevertheless, it is expanding everywhere, but in this case the pull of gravity is not enough and the universe continues to expand forever. What does forever mean? What was there before the beginning of time? What is meant by the expansion of an already infinite universe? Try not to think about it.

A flat universe forms the dividing line between a closed universe and an open universe. It extends to infinity and expands forever, but the expansion gradually slows down and eventually just stops. But there is no big crunch. It would be somehow aesthetically pleasing if the universe were precisely flat, on a large enough scale, and it would be nice if we had a theory that said that this is how it must be. But we don’t have this theory yet, and we have to try to find out whether the universe is open, closed, or flat by observation. In fact, we think the universe is at least nearly flat, if not precisely flat. The trouble is, we can’t find enough matter in the universe to make it flat. We are not even close.

The critical density is about $5.7\times10^{-27}\ \text{kg/m}^3$. To put this in perspective, this corresponds to about three hydrogen atoms per cubic meter, average, throughout the universe. Not much, it might seem. But if we count up all the stars we can see, add up all the interstellar gas and dust that we can detect, throw in all the neutrinos, black holes, pulsars, and even light, and divide by the volume we can see, we get about $2\times10^{-28}\ \text{kg/m}^3$, less than one tenth what we need. The rest composes what we call “dark matter,” dark because we can’t see it. There is no lack of fanciful suggestions for what might make up the difference, including exotic particles never observed on earth (why?), such as axions and supersymmetric particles. More recently, we have introduced “dark energy,” whatever that is, and reintroduced Einstein’s cosmological constant to help flatten the curvature. Now we are doing experiments – many having to do with the Cosmic Microwave Background radiation – to try to sort out the various contributions. But in the end, the embarrassing fact is that we don’t seem to have any idea what it is that makes up more than 90 percent of the universe we live in; this is not something we physicists are proud to admit. The good news, in some sense, at least for us physicists, is that there are still careers to be found in the unsolved mysteries of nature; our profession has not yet reached its logical end.

Before finishing the discussion of curved spaces and expanding universes, we should point out a couple of seeming paradoxes. We have already alluded to the fact that it is meaningless to ask the question “what is the universe expanding into?” There is no edge to the universe, even if it is a closed universe. For while a closed universe is finite, it closes on itself, so even if you keep going in a straight line (like light), you will travel in a curved space and never get out. In fact, every point in the universe will, on a large scale, look like every other point. This is called the cosmological principle, and it is what allows us to solve the equations of general relativity at all. A corollary of this is that there is no center of the universe. This isn’t so hard to understand for an infinite (flat or open) universe. After all, what does it mean to be halfway between infinity to the left and infinity to the right. But it is true also for a closed, finite universe; even though the universe has expanded from a single point, there is no center. Think of it this way: visualize a 2-dimensional, closed universe as the surface of a spherical balloon.

Remember, now, that the balloon has positive curvature, but the curvature is into the third dimension, which we 3-dimensional beings can visualize but is not part of the 2-dimensional universe that is the surface of the balloon. As the balloon expands, the 2-dimensional universe starts from a very small area and gets larger. But where is the center of the surface? It is NOT the center of the sphere, since this is in the third dimension! Every point on the surface of the balloon is equivalent to every other (if we ignore the neck of the balloon, where we inflate it; there is no neck on the universe). This is the cosmological principle applied to balloons.

So how do we know - or why do we think – that the universe is expanding? This remarkable idea was discovered by Edwin Hubble in 1929, using the 100-inch telescope at Mount Wilson to observe nearby galaxies. As he peered out into the cosmos, Hubble noticed that every galaxy that he looked at seemed to be moving away from the earth, and the more distant the galaxy the faster it was receding from us. How could he tell how fast they were going away from us, and how could he tell how far away they were located? Actually, the first part – how fast were they going – was the easy part. Each chemical element of which stars are composed has a very particular set of colors that it emits. When Hubble compared these colors to the same ones observed on earth he noticed that they were very slightly shifted toward the red, the long-wavelength (low-frequency) end of the visible spectrum. He recognized that this was due to the so-called Doppler shift caused by the movement of the stars as they emitted the light. Just as the whine of truck tires goes down in frequency as the truck passes by and starts to recede into the distance, the light from the stars shifted to lower frequencies as the stars receded. Figuring out the distance to the stars was harder. Not all stars and galaxies are the same size, so their apparent brightness is not a firm indicator of their distance; in fact, this remains the hardest part of the problem. Hubble used so-called Cepheid stars, identifiable by the characteristic variability of their luminosity, as a sort of "standard candle." Actually, Hubble's data were so scattered in 1929 that it is amazing that he could have made the conclusions he did. More recently scientists here at Vanderbilt, among others, have been using "type I-A supernovas," all of which are exactly the same brightness, as standard candles, and the results have improved.

The bottom line of these observations, called Hubble's law, is that the velocity of recession is proportional to the distance to the star. This can be summarized by one number, called Hubble's constant. If we divide the velocity of recession by the distance to the star, we find that all stars give (about) the same result, namely

$$H_0 = 22 \text{ km/s/million light years}$$

This is called Hubble's constant. The picture that emerges from this observation is that the universe is expanding at the rate $22 \text{ km/s/million light years}$. That is, a star now one million light years away is moving away from us at the speed 22 km/s. A star twice as far away is moving away from us at twice the speed. We are, of course, not at the center of the universe (there IS NO center of the universe!), but all the stars and galaxies that we can observe are moving away from us in all directions. We can understand this by imagining a raisin bread rising in the oven. As the bread rises all the raisins get farther and farther apart. Viewed from any particular raisin, the nearby raisins seem to be receding as though our raisin is the center of the loaf. This suggests that at some time in the (distant!) past, both these stars (or at least the space they occupy now) was nearby, and they have been moving away ever since. How long ago were they at the same place where we are? Well, since a light year (the distance light travels in one year) is 9.5×10^{15} m, the time of the big bang must have been

$$t=\frac{1}{H_0}=\frac{9.5\times10^{15}\times10^{3}\ \text{km/million light years}}{22\ \text{km/s/million light years}}=4.3\times10^{17}\ \text{s}=14\ \text{billion years}$$

This is the roughly the age of the universe, and is slightly (a billion years or so) older than the estimated age of the oldest stars. More recently, this same result, and much more, is being revealed by measurements of the cosmic microwave background radiation. Which brings us, finally, to why we are here.

2. The CMB

2.1 First 300,000 years

The origin of what we now call the cosmic microwave background lies in the origin of the universe itself, and to understand it we need to understand what happened at and shortly (within the first 300,000 years) after the big bang. Conversely, as we struggle to learn about the basic principles of physics (such as what makes up the 90 percent of the universe that we can't explain!) we can test our ideas against what we can observe of the big bang as seen through, among other things, the cosmic microwave background. In fact, the cosmic microwave background is now our best window on what happened back then, our "cosmic Rosetta stone," as someone has called it.

At the very beginning of time, the universe, what there was of it, was a primordial soup of energy, in a form we don't yet fully understand. As this soup expanded, it cooled, and some of the energy condensed into particles. By one second after the big bang, the universe had cooled to 10^{10} K, and by the end of three minutes it had cooled to 10^{9} K. This was cool enough for protons and neutrons to condense into nuclei of hydrogen and helium, but the universe was still mostly light (photons), neutrinos, and antineutrinos. It would be another 300,000 years before the universe was cool enough (about 3000 K) for the electrons to attach themselves to the nuclei and form hydrogen and helium atoms. But this was a critical time for the cosmic background radiation. It was only much later (a billion years, or so) that the universe became cool enough for the helium and hydrogen to coalesce into stars, and longer still for the first stars to die and explode as supernovae to form the chemical elements heavier than helium, of which we, and the earth, are mostly composed. But more of this later.

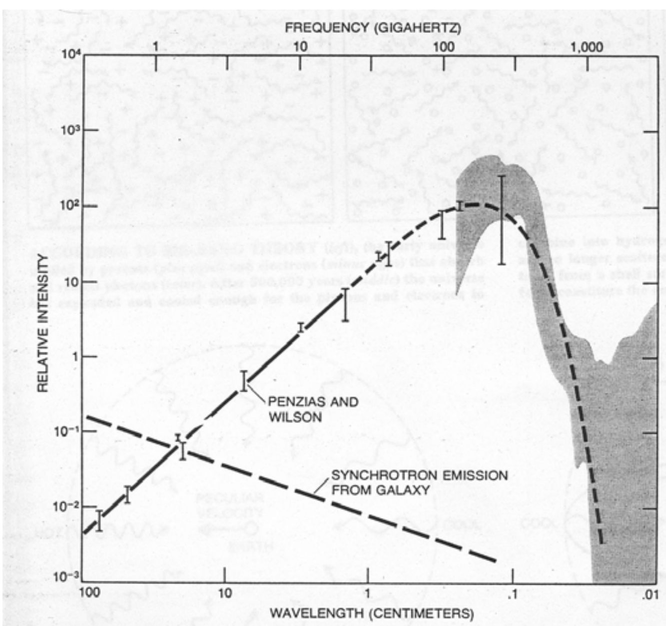
2.2 Blackbody radiation

As the universe cooled, following those first few minutes, the photons interacted strongly with the electrons, being scattered by a process called Thomson scattering. Thus, the light was confined by the electrons and as the universe expanded the light and the electrons cooled together, in thermal equilibrium. The spectrum of the light was a nearly perfect blackbody spectrum, as we would call it, with most of the light at longer wavelengths and a decreasing amount of light at shorter wavelengths. We would obtain the same spectrum, but shifted to much longer wavelengths, if we looked at the radiation inside a closed box at a certain temperature. This spectrum is therefore well known to physicists, and well understood.

2.3 Formation of the CMB

This process continued for about 300,000 years, as long as the photons interacted strongly with the electrons and remained in thermodynamic equilibrium. But things didn't stay this way forever. Eventually, the universe cooled enough for the electrons to remain attached to the hydrogen and helium nuclei and form neutral hydrogen and helium atoms. At this point, which occurred rather suddenly, at least compared with the time scale of 300,000 years, the universe became transparent to the background radiation. Thereafter, the radiation in the universe and the matter stopped interacting and went their separate ways. The universe still cooled, but the radiation and the matter cooled differently and no longer looked the same. In fact, the matter started to condense into clouds, and the clouds into interstellar dust, stars, galaxies, and clusters of galaxies, all the objects that we see in the universe today (and things we can't see but that we know must be there).

But what happened to the radiation was much simpler. Basically, once the radiation became decoupled from the matter in the universe, it simply stretched as space itself expanded. All the wavelengths of the light stretched in direct proportion as space itself stretched. Therefore, since the energy of a photon is in inverse proportion to the wavelength, and the temperature of the blackbody radiation is in direct proportion to the energy of the photons, the temperature of the cosmic background radiation dropped in inverse proportion to the size of the universe. When the universe was a mere 300,000 years old, the temperature of the radiation was 3000 K, so now, after 14 billion years, the universe is 1000 times larger, and the temperature of the radiation has dropped to 3 K. In fact, the cosmic background radiation is still a nearly perfect example of blackbody radiation, as shown in the figure.



But it is not a *perfect* example of blackbody radiation. Since a nonuniformity cannot form in an otherwise perfectly uniform universe, all the coalescing of matter into stars and galaxies must have been initiated by some initial nonuniformity left over after the big bang. Even raindrops cannot form from moisture in the air without something – a speck of dust or an ion left behind by a cosmic ray – to nucleate the condensation. There is now a large scientific enterprise underway to interpret what we see in the universe in terms of the initial fluctuations; from this we hope to learn about the fundamental forces of nature that caused the primordial fluctuations. The details of the cosmic background are a therefore a window into the mysteries of our origins.

2.4 Anisotropy

Some of the nonuniformities in the cosmic background are easy to understand just in terms of our motion through the universe. Although there is no center of the universe with respect to which we could plot our position and movement, it is nevertheless possible for us to have motion relative to the absolute reference frame of the universe. In our balloon model, there is no center of the surface of the balloon, but we can move around the surface and even circumnavigate that universe. So, too, we can move relative to the real universe. In fact, of course, the earth moves relative to the sun, and the sun relative to the galaxy. But most of the velocity of our motion relative to the absolute universe is due to the motion of our galaxy. The sum of all these motions is about 371 km/s, but you probably never noticed. In fact, this motion shows up in the spectrum of the cosmic background not as a shift in the spectrum but only as a slightly higher temperature (about one part in 1000) in the direction we are moving.

But the nonuniformities of the universe that eventually caused matter to coalesce into stars and galaxies also left their mark on the cosmic background radiation. This is simple to understand in the following way. If, for example, a certain region of the universe was slightly denser than other regions, then electrons in that region would have persisted slightly longer, until a slightly lower temperature was reached, and the cosmic background radiation in that region would have decoupled from matter at a slightly lower temperature. But these fluctuations must have had their origin when the universe was much smaller, almost certainly when the entire universe was smaller than a single atom, and since the fluctuations can't arise in an absolutely uniform universe they must be imbedded in the laws of the universe themselves. Thus, if we can understand these fluctuations we may learn something about the fundamental forces of physics. There is certainly no hope of doing an experiment anything like this ever again!

2.5 What more can we learn?

The difficulty, of course, is that we cannot repeat, or even mimic the origin of the universe in the laboratory, and the one experiment that was done (unless the universe is closed and has been repeatedly banging and crunching forever) is pretty remote, now, and all we can do is comb the debris for information. But there is obviously much to learn, since we don't have any idea what 90 percent of the universe is even made of. So the way we proceed is to construct theories and compare the predictions of these theories with what we observe. For example, Einstein postulated the so-called cosmological constant. When he found out that the universe is actually expanding, he removed the constant from his theory, but now, as we examine the debris in greater detail, we find that to explain what we see we need the cosmological constant back again, although with a different value than Einstein gave it. Similarly, it has been postulated that when the universe was very young (less than 10^{-30} s old) it underwent a sudden expansion, which we call inflation. This hypothesis of inflation has several ramifications, among them the fact that the universe must be nearly flat (which it is). The inflation hypothesis also makes predictions about details of the fluctuations in the universe that we are just becoming able to measure. In the same way, we can’t explain the “missing mass” of the universe, yet, and in fact we can’t explain all the properties of the universe in terms of matter of any kind, so we have to postulate “dark energy” as well as “dark matter.” It seems now that when we finally come up with the “theory of everything” (if we ever do), there will be only one experiment to which we can compare it, the origin and existence of the universe itself.