First Light: Revealing the Early Universe

Professor Chris Lintott

29 May 2024

From our perspective on the surface on a small planet around an ordinary star, it may appear to be something of a surprise that we can make any sensible conjecture about the past, present and future of the Universe. Yet it turns out that – as long as we seek to describe only the evolution of the cosmos as a whole, the physics involved is simple enough to fit on a single sheet of paper; the calculation of, for example, how a cricket ball propelled through the air on a damp morning will move, is much more intricate and includes more variables than writing down the equations that govern how the Universe is changing. The fate of the Universe is determined by what Astronomer Royal Martin Rees called 'just six numbers'.

Any theory needs constraining by observation, and in this case our observations are determined by the fundamental fact that the light moves at a fixed but finite speed. As noted, several times in this series, this means that when we look out into the cosmos we are looking at the past. The 150 million kilometres that separate us from the Sun mean that we see it as it was eight minutes ago; were it to disappear right now we would remain oblivious for the time it takes to hard boil an egg[1]. Most of the stars that can be seen with the naked eye are seen as they were a few hundred years ago, with the most distant – like brilliant Deneb, in Cygnus, seen as they were a few thousand years ago. The light we receive from the most distant object visible to the naked eye, the Andromeda galaxy, has taken just over two million years to reach us[2].

The cosmic distance ladder proceeds from here, through the nearest large cluster of galaxies, which we see as it was sixty-five million years ago, to the systems seen in recent images from JWST (discussed at the end of lecture one) whose light has been travelling towards us for more than thirteen billion years, at a time when the Universe was very different from that around us today – galaxies were smaller, stars younger, and star formation much more common than in the present day.

JWST is a feat of engineering, with a sunshield the size of a lawn tennis court designed to keep the telescope cool[3], and hexagonal mirrors coated in gold to reflect light to its instruments. These innovations were necessary because the telescope sees in infrared light, at longer wavelengths than our eyes can, and this choice was driven by the desire to see those distant galaxies, which are brightest in this wavelength region.

Why is this? The light we receive from galaxies comes from their constituent stars, and most stars – like our Sun – are brightest in the visible part of the spectrum. Indeed, this is most likely why this is the *visible* part of the spectrum, evolution having done its best to make use of the light that we receive (a decent argument, incidentally, that aliens we meet will also see in roughly the same range of light we do, though they may also be equipped, as some animals are, to see further into the infrared or ultraviolet). Galaxies, including those in the early Universe, should therefore be brightest in the visible, but they are clearly not.

Even a cursory look at the spectra of light we see from them shows that it has been redshifted, with the same familiar features caused by the presence of elements including hydrogen and helium there, but

---

[1] The speed of light is approximately one foot per nanosecond, which means that you see your feet as they were five or six nanoseconds ago.

[2] Some people, with exceptionally acute vision, claim to be able to see the Triangulum Galaxy, the third large system in our Local Group. It is 2.7 million light years away

[3] The effectiveness of the shield can be quoted in the same units as used for sunscreen – it has a SPF of 1 million.

moved towards the red. Work by Vesto Slipher, at Lowell Observatory, in the 1920s showed that different galaxies show different degrees of redshift, a cosmological fact in need of explanation.

Between the world wars, telescopes developed to the point where the brightest individual stars could be discerned in distant galaxies. The presence of a class of luminous variable star known as Cepheids, which brighten and fade with a period that is related to their luminosity[4], allowed the distance to their host galaxies to be accurately measured for the first time.

Edwin Hubble combined observations of these ‘standard candles’ with careful observations of the host galaxy’s spectra, producing what we now know as the Hubble-Lemaître Law[5]: the further away a galaxy is from us, the greater the observed redshift. This is an astounding result, which seems to connect the properties of distant galaxies to our position in the Universe, but it has been confirmed by nearly a century of observations, including a large Cepheid-hunting program with the *Hubble Space Telescope*, named after Edwin with this purpose in mind.

The explanation for what is seen has profound consequences. We can make an analogy between the observed redshift and the change in pitch of an ambulance siren as it goes past, caused by the compression and expansion of the sound waves emitted by a moving source (the Doppler effect), and express the redshift as a velocity. The Hubble Law then becomes this: the faster away a galaxy is, the faster it is receding from us.

This result is naturally explained by our residing in a universe in which space itself is expanding: the more distant a galaxy is, the more it will be affected by the expansion of the space that sits between us and it, and it is this expansion which produces the redshift. We can measure the rate with which space is currently expanding, a value we call the Hubble constant, which lies around 70 km/s for every Megaparsec. A megaparsec is more than three-thousand light-years, a scale on which seventy kilometers per second is hardly noticeable, and so this explains why the cosmic expansion is visible only on cosmological scales.

It is important to note that, as the expansion of space is universal, any observer in any other galaxy would obtain the same result. Our apparent status at the centre of a cosmic explosion is an illusion, and we do not exist in a special place in the Universe. These results do, however, imply that there is a special time in the Universe’s history; run the film backwards, so space is contracting, rather than expanding, and we get to a point where everything converges. This moment was given the rather disdainful moniker of ‘the Big Bang’ by cosmologist Fred Hoyle, who felt the existence of such a moment was so implausible as to rule out the theory completely.

The physics that explains what was happening in that ‘first’ moment is still little understood, and the theories that we do have are untested, and perhaps, at least for the foreseeable future, untestable. We can make useful predictions about the immediate aftermath, however, and use them to test what has become known as ‘the Big Bang theory’.

One early set of ideas, much in vogue amongst theoretical cosmologists in the 1950s, suggested that our Big Bang may merely have been one of a series. If the observed expansion was followed by a period of contraction, then a Bang could be followed by a Crunch[6]. This neatly allows us to escape from explaining why our Big Bang happened but creates a problem. During the course of the Universe’s existence between bangs, things happen – stars, for example, convert hydrogen into first helium, and then heavier elements. Yet there is no sign in our Universe of the debris of stellar activity in prior cycles of activity. Each Big Bang seems to erase evidence of its predecessor, producing a universe each time that is essentially a blank slate.

This could put paid to the cyclic universe idea, but it can be saved if the state of the cosmos just after (or just before) a Big Bang is hot, hot enough to break up atomic nuclei and reduce everything back down to protons and electrons. The aftermath of the Big Bang would thus be a hot, dense environment, with cooling happening only as a consequence of later expansion. In this hot Big Bang scenario, it was felt, it would be possible for the sequence of cosmic expansion and contraction to repeat forever. What was missing was

---

[4] A relationship now known as the Leavitt law, after Henrietta Swan Leavitt who found it.

[5] Lemaître being the theorist who predicted it, and the International Astronomical Union recently feeling that theorists don’t get enough love.

[6] The resulting period where the Universe is once again reduced to a dense state is known as the Big Crunch, though I like an alternative name: the Gnab Gib

any evidence that this happened.

Bell laboratories have already featured in this year's lectures, as host to the engineers and pioneers of radio astronomers Karl Jansky and, a little later, Grote Reber. In the 1960s, Arno Penzias and Robert Wilson were following in their path, using a new 'Big Horn' antenna to characterize the sources of noise which interfered with radio communications. Their work was meticulous; dealing with thunderstorms and tracking down interference, they were left with a faint background hiss, accounting for a few percent of the noise their powerful instrument could pick up.

Determined to work out what this remaining, pesky source of noise is, Penzias and Wilson eventually blamed local pigeons, who, nesting in the Big Horn itself, had coated their instruments in what is described in their paper as a 'white dilectric substance'. A pigeon trap – now stored in the collections of the Smithsonian Museum in Washington DC – was obtained, and put to use, but despite the removal of these avian inhabitants, the faint hiss of noise in the antenna's systems could still be heard.

As it was persistent and unchanging, the two researchers decided that it must be extraterrestrial, coming from the sky. They had discovered what we now call the Cosmic Microwave Background, a short wave radio hum that pervades the Universe. Down the road at Princeton, the group led by cosmologist Robert Dicke were preparing to design an experiment to look for the echoes of their hot Big Bang – a signal they felt should manifest as a background hum of microwaves.

The two groups got in touch, and the discovery by Penzias and Wilson was published in the same issue of the journal that contained the theory, worked out by Dicke's group. But why should a hot big bang produce an afterglow of microwaves?

The answer lies in conditions just a few minutes after the Big Bang itself. At this time, the Universe was hot enough that neutral atoms did not exist; electrons had enough energy to escape the atomic nuclei (mostly single protons – so hydrogen – but with some helium and a tiny amount of lithium too) that normally bind them. Any light emitted would, before travelling more than a centimetre, almost certainly encounter a freely moving electron and be scattered in a random direction, at which point it would hit another electron and so on and so on.

Essentially, space at this time was not empty, but rather comprised of a broiling sea of electrons. Were you to be transported there, you would (assuming we kept you safe from being boiled alive etc) find yourself embedded in a luminous fog, unable to see further than a centimetre or so in front of your face.

As the Universe expanded, things changed. It cooled, to the point, reached something like 400,000 years after the Big Bang itself, when electrons could be captured by atomic nuclei. Very suddenly, in the course of maybe a year, the cosmos transformed from being filled with electrons to filled, as it is today, with empty space. Light emitted just before this event can likely travel across the whole visible Universe without hitting anything – it will fill space as a reminder of the early, dense, hot phase with which everything began.

The physics of this early era is well understood, and we can predict the spectrum of light that would be emitted in such conditions. It should be brightest in the infrared, but, as space as expanded over the last fourteen billion years or so, it will be redshifted into the microwave region of the spectrum, just where Penzias and Wilson found it. The microwave background they found is a snapshot of the Universe as it was at the moment the electrons were captured – frozen in time like a crowd of nightclub dancers captured by strobe lighting.

This immediately caused a problem. The radiation that Penzias and Wilson had found was uniform – it appeared the same brightness and temperature across the whole sky. It thus stood in stark contrast to the Universe we observe today, which is decidedly lumpy, with regions packed with galaxies and clusters of galaxies different from regions of space, often called voids, where they do not exist. How do we produce a lumpy Universe from smooth beginnings?

The answer is that we don't. The first space-based mission to study the microwave background, COBE[7], produced a map that showed tiny fluctuations in the otherwise uniform background. They were small, perhaps amounting to changes in temperature of one part in ten thousand, but they were there, and their announcement in 1992 was widely seen as being final confirmation that the hot Big Bang picture was right. Cosmologist George Smoot, who later won the Nobel Prize with John Mather for his work leading the

---

[7] Appropriately enough, the COsmic Background Explorer.

COBE mission, in a fit of hyperbole, described the result as ‘like seeing the face of God.’

Well, maybe. But simulations since have gone on to underline the importance of these tiny fluctuations. Gravity acts to exaggerate them, with those parts of the Universe which have, in the beginning, slightly more material than the average, tending to attract more material over time. This processes, visualised in endless supercomputer simulations, produce a cosmos that matches the one we observe around us from relatively simple beginnings. The degree of lumpiness depends on the total amount of mass in the Universe, and so combined with observations of the initial conditions revealed in the CMB these simulations provide powerful constraints on the amount of matter in the Universe.

Other cosmological traits can be derived from observing the CMB too. We can look at the scale of the fluctuations we observe, which are about a degree across, and expect them to correspond to the size of the largest features we observe in the Universe today. The details of this method are complex, involving observing what are known as baryon acoustic oscillations, but the principle is sound. Comparing the two scales gives us a measure of the rate of expansion in the more than thirteen billion years that the light from the CMB has been travelling towards us.

This effort has led cosmologists into trouble, as it produces values which are a few percent different from those generated when we try to measure the Hubble constant by observing our local Universe. These methods use more luminous standard candles[8] than the Cepheids mentioned earlier, such as type 1a supernovae, objects whose intrinsic brightness is known, and which can then be compared with their apparent brightness, giving us the distance to their host systems. With this data in hand, we can make a grander version of the plot that Hubble used to reveal the expansion of the cosmos, but the speed of expansion does not match that derived from the CMB.

The resulting discrepancy between methods which would be expected to agree with each other has charmingly been labelled not a disagreement or argument, but rather a ‘tension’. It has refused to be resolved as new measurements have been made, or by the high-resolution study of the CMB made recently by the European Space Agency’s *Planck* satellite.

What does this tension mean? It may be that we are simply not understanding something about the objects we are using standard candles, but much of the excitement comes from the hope that it reveals something is fundamentally wrong with our cosmological picture. Results from the upcoming Vera Rubin Observatory, which will discover more supernovae in its first two months of operation than have been seen in the whole of human history, will go a long way to showing if the ‘tension’ is real.

Meanwhile, there are new efforts to study the Microwave Background itself, testing an idea posed to solve a mystery inherent to the early Universe. I mentioned that it was uniform, but it is not at all clear why that should be the case. One possibility is to blame the Big Bang – to say there must be something in the physics of that primordial moment that inevitably produces a mostly (but not entirely) smooth Universe – but most cosmologists subscribe to an as yet untested theory the posits an early period of inflation.

According to this idea, in the first tiny fraction of a second there was a sudden period of rapid expansion. Its effects will have been to wash out any large scale variations – post-inflation, our observable Universe will be only a tiny fraction of the whole, and so it is not surprising that it is uniform. The idea also has several other benefits, including explaining the small fluctuations we do see, and which are the seeds for the structure we see around us today, as previously tiny quantum fluctuations, frozen by the rapid expansion into the size we see today.

There is good reason to believe that such a period of inflation would have left its mark on the cosmic microwave background. Telescopes designed to detect this signal have been operating at the South Pole for the last decade or so; despite a false alarm in 2015, no such signal has yet been detected.

New telescopes in Chile are under construction, which should help. One of the reasons these measurements are difficult is that dust from our own galaxy shines brightly at the same wavelengths as the CMB, creating a foreground which has to be subtracted. Or, if we are interested in our galaxy, we can treasure this interference, using it to understand the nature of the dust which fills the Milky Way, and which provides the raw material for future episodes of star and planet formation.

In attempting to study the distant Universe, we have returned to where I began the lecture series – in

---

[8] More precisely, ‘standardizable candles’ – objects whose luminosity we can work out based on their observed properties, even if it’s not always exactly the same.

thinking about our own galaxy's place in the cosmos. In between, I've tried to describe the modern practice of astronomy – making measurements which are often simple in conception, but with extreme precision – revealing a grand and marvellous Universe which we are only now beginning to understand, and which is full of surprises.

Further reading

The original discovery of the CMB is reported in A. Penzias and R. Wilson, "A Measurement of Excess Antenna Temperature at 4080 Mc/s," *ApJ* 142, no. 419 (1965), https://ui.adsabs.harvard.edu/abs/1965ApJ...142..419P

The explanation by Dicke's group, published in the same issue as the discovery paper, is given in R. Dicke et al., "Cosmic Black-Body Radiation," *ApJ* 142, no. 414 (1965)

https://articles.adsabs.harvard.edu/full/1965ApJ...142..419P

The description of the cosmology of the early Universe given here is now out of date in some aspects, but still unparalleled in reach and clarity: S. Weinberg, *The First Three Minutes: A Modern View of the Origin of the Universe,* 2nd ed. (Basic Books, 1993).

I would also recommend: P. Coles, *Cosmology: A Very Short Introduction* (Oxford University Press, 2001), and the venerable online tutorial by Ned Wright, here: https://astro.ucla.edu/~wright/cosmolog.htm

For those looking for something more technical, there's a good review here: R. Durrer, "The Cosmic Microwave Background: The History of Its Experimental Investigation and Its Significance for Cosmology," *Classical and Quantum Gravity*, 32, no. 12 (2015), id. 124007, available at arxiv.org/abs/1506.01907.