# 

# 

# 

23 September 2015

**The Primeval Fireball**

Professor Joseph Silk

The universe underwent an inflationary period of expansion. From an extremely hot beginning, it cooled down, and then reheated to form the expanding universe that we observe today. The relic radiation generated the fossil glow of the cosmic microwave background radiation, a testament to its fiery origin. Infinitesimal fluctuations in temperature were laid down very soon after the beginning. Gravity triumphed in slightly over-dense regions where galaxies eventually formed. Microwave telescopes map out the structure of the universe long before the galaxies were present. I will describe the current challenges in cosmic microwave background astronomy.

The universe began in a state of extreme density and temperature, some 14 billion years ago. Our best estimate is that its density was once higher than its present value by 120 powers of 10. This is at the limit of extrapolation of our current theories of gravity and quantum physics.

And it was very hot, hotter than the radiation that presently fills the universe by some 28 powers of 10 in temperature. How do we arrive at such incredible extrapolations from the present?

**The Expansion of the Universe**

The story begins with the discovery of the expansion of the universe in 1929 by American astronomer Edwin Hubble. Now confirmed at immense distances from us, over many billions of light years, the extrapolation to an ultra-dense beginning is essentially inevitable. Of course, most of this extrapolation is based on theory. But direct measurement provides a solid foundation.

We measure the expansion directly back in time to when galaxy separations were a third of their present value, and we infer the associated increase in temperature of the radiation to when the observable universe was a thousandth of its present size, 380,000 years after the beginning of the expansion. The origin of the light elements occurred after only a few minutes, as the high temperature enabled protons to slam together and convert into helium nuclei by overcoming their electromagnetic repulsion as nuclear forces took over. The origin of known particles and their constituents such as quarks occurred at a nanosecond after the beginning. Let us convert to Planck time. This is the unit of time that demarcates the emergence of the gravitational force and is as far back as we can extrapolate. The fundamental forces, including the nuclear forces, emerged at a million Planck times. All the presently visible universe was then about the size of a grapefruit.

These are seemingly speculative statements but are based on harsh reality. There are three pillars to the Big Bang: the expansion of the universe, the light element abundances and the primeval fireball radiation. The ratio of light element abundances probes the physics of the first second, and the abundances of helium, deuterium and lithium are now measured to high accuracy. The formation of matter takes us to the first nanosecond, and the unification of the fundamental forces takes us back to the Planck time. There is not much wriggle room. Let us consider these three pillars.

**The Light Elements**

The light elements were synthesized in the first three minutes of the Big Bang. How did this come about? And why weren’t the heavier chemical elements also produced?

The story develops like a crime scene. It begins with brilliant deductions. It ends with compelling evidence. Russian immigrant physicist George Gamow, based at George Washington University in Washington DC, believed that the Friedmann-Lemaitre expanding universe was the ideal environment for synthesizing the chemical elements. This was in the 1940s. He overcame a key theoretical obstacle by realizing that at high enough energies, protons could overcome their mutual electromagnetic repulsive force and fuse together into helium nuclei. He had to assume a hot universe, full of radiation, at a temperature of billions of degrees. This occurred naturally in the first seconds of the expansion. He, along with colleague Robert Hermann and especially his student Ralph Alpher, calculated that one helium nucleus would be synthesized for every ten protons. Within less than a decade, astronomers had measured the helium abundance in a variety of sites, from massive nearby stars to distant galaxies. They were able to subtract off the small contribution from stars, which radiate via thermonuclear fusion of hydrogen into helium, and confirmed that there was a universal value in accord with the Big Bang prediction.

Gamow had hoped to synthesize all of the chemical elements in this way. But nature forbade this: it turns out that the way to synthesize carbon, via combination of two helium nuclei each of mass 4, into beryllium of mass 8, to combine with another helium nucleus to form a carbon nucleus of mass 12 is unachievable in nature. Beryllium of mass 8 is an extremely unstable isotope! Synthesis in massive stars, by combining three helium nuclei directly into a carbon nucleus, was eventually to provide the solution for the heavy elements. There was not enough time for this three-body encounter to occur in the Big Bang. But we observe far too much helium, the second most abundant element in the universe, to have formed in stars. The Big Bang gave the solution for the helium abundance. There was one helium nucleus for every ten hydrogen protons. The Big Bang also predicted formation of deuterium and other light element abundances in the observed proportions.

The legacy of Gamow’s brilliant deduction about the necessity for a hot beginning to the universe was the realization, first quantified in Alpher’s doctoral thesis in 1947, that the universe today is filled with relic thermal radiation, today at microwave frequencies. Fred Hoyle, who first coined the term “Big Bang” in a 1951 BBC radio broadcast, reminisced much later about his encounter in 1953 in Ann Arbor, Michigan with Gamow:

“I recall George driving me around in the white Cadillac, explaining his conviction that the Universe must have a microwave background…”

With hindsight, the discovery could have even been made a decade earlier in 1941, when anomalous excitations of interstellar molecules seemed to require an ambient microwave background radiation field. However Hoyle, along with all of his contemporaries, did not take this result sufficiently seriously. Indeed, Hoyle went on to write,

“I recall my telling George that it was impossible for the Universe to have a microwave background with a temperature as high as he was claiming because observations of the CH and CN radicals by Andrew McKellar had set an upper limit of 3 K for any such background.” The sad truth is that Gamow’s prediction was forgotten, and only a decade and a half after the pioneering work of Alpher and Gamow was the cosmic microwave background discovered.

**The Cosmic Microwave Background**

The cosmic microwave background was discovered in 1964. Arno Penzias and Robert Wilson were radio astronomers employed by Bell Laboratories in New Jersey. This was at a time when giant commercial companies could set up giant research laboratories to do blue skies research. In Bell’s case, this led to the discovery of transistors in 1947 and lasers in 1960, of immense commercial significance, and of the cosmic microwave background in 1964, the ultimate evidence for the hot Big Bang.

Penzias and Wilson were assigned a microwave telescope that had been used for satellite communications, but was recently made redundant as technology improved. Their initial goal was to make high precision microwave maps of the Milky Way. They were disconcerted to find an irreducible background. This noise in their receiver came from all over the sky, with no dependence on the ecliptic plane, traced by the planets, or the Milky Way. Wherever they pointed the telescope, the microwave noise remained constant, at an equivalent temperature of about 3 degrees Kelvin.

Meanwhile cosmologists led by Robert Dicke at nearby Princeton University were searching for the relic radiation from the Big Bang. They deduced it would appear at microwave frequencies and be isotropic in the sky. Dicke was unaware of George Gamow’s work on the fossil radiation: Gamow had actually predicted such a radiation field. His argument was based on the fact that the Big Bang must have been hot in order to produce the large amount of helium in the universe, and the temperature of the radiation would have cooled by now to a few degrees Kelvin. It just took a phone call with Dicke, and Penzias and Wilson were able to announce the discovery of the cosmic microwave background, and realize that they had accidentally discovered the fossil glow from the Big Bang. They had measured its intensity at a single frequency. Over the next two decades, increasingly precise measurements established that the microwave glow was blackbody radiation, that expected in a perfect furnace, the Big Bang, and cooled down to microwave frequencies as the universe expanded. The observational studies of the spectrum culminated in the COBE satellite’s 1990 measurement of the most perfect blackbody ever measured, in the sky.

How is blackbody radiation generated? It requires an ideal oven, where the radiation scatters around, and is absorbed and remitted by the particles present. In the case of the early universe, the temperature was at millions of degrees when this occurred. With so much interaction between particles and radiation, there had to be a perfect balance between the temperature of the radiation and the matter. But as the universe expanded, the temperature cooled, and eventually the ability of particles to change the numbers of photons, by absorbing and reradiating, stopped, because the density had become too low. This happened about a year after the Big Bang. At this point in time, a slight difference in temperature developed between radiation and matter, and the nature of the radiation changed. It was no longer blackbody. The hotter particles heated the colder photons, and they remained slightly hotter than the typical photons. Thermalization ceased, there was no longer precisely one temperature for both matter and radiation The net effect on the radiation is that we have a distorted blackbody distribution, one whose imperfections could reveal the story of the thermal history of the universe between one year, the end of thermalization, and 380 thousand years, when the universe first became transparent to the radiation. But the expected distortions are very small, amounting to a few parts in a billion.

The epoch of transparency began when the last scatterings of photons occurred. The universe was a thousand times smaller, or a billion times denser, than it is today. Galaxies had not yet formed. The photons were previously scattered by free electrons, as the hydrogen was fully ionized. Virtually all of the electrons recombined into atoms as the universe entered the atomic era, the dominant constituents being atoms of hydrogen and helium. Much later, only when the first galaxies formed, would the universe again predominantly consist of ionized plasma, as the energetic photons that were produced by the first massive stars ionized essentially all the hydrogen of the intergalactic medium. From the epoch of hydrogen recombination, the cosmic blackbody radiation reaches us unscathed, apart from tiny heating effects due to the few residual electrons. Far more distortion of the initial blackbody spectrum occurs when the photons pass through the denser ionized gas in groups and clusters of galaxies, but even there, the distortions remain small, at the level of parts in a million. Still, these are large enough to give a foreground that most likely masks any possible signal from the very early universe.

**The Search for Fluctuations in the Primeval Fireball**

In their 1965 discovery paper, Penzias and Wilson wrote:

“This excess temperature is, within the limits of our observations, isotropic…”

Within two years they showed that the cosmic microwave background revealed no angular variations on degree scales at a level of about 3 percent. However, theory dictated that there had to be fluctuations somewhere on the sky below this level. The universe must have contained density fluctuations in order to seed the growth of large-scale structure. The idea, well known since the writings of Georges Lemaitre, was that the excess self-gravity associated with slight over-densities would gradually initiate the local collapse of matter relative to the expanding universe. In this way, the galaxies would form.

I was a graduate student at Harvard in 1967, and my PhD thesis was to do precisely these calculations. One of my first papers was titled Fluctuations in the Primordial Fireball, and made the case that the theory of galaxy formation required temperature fluctuations of a few tenths of a per cent. This number proved to be remarkably optimistic. It was to take three decades of searching before the elusive fluctuations were finally detected.

The delay was due to a combination of reasons. The theoretical predictions were refined and decreased in strength as more precise calculations were done. Perhaps the major ingredient added in the 1980s was the realization that the universe was most likely dominated by cold dark matter. The dark matter has to be very weakly interacting, and in particular, does not interact with photons or electrons. Hence it does not suffer from the viscosity between electrons and photons, and consequently protons that suppresses density and temperature fluctuations on small angular scales. However dark matter fluctuations are not suppressed and do supply gravity, so that once the electrons and protons recombine into atoms as the universe cools, the atoms can now feel the gravity of the dark matter fluctuations. This boost in growth to the density fluctuations that are the seeds of galaxy formation means that the relic temperature fluctuations are reduced in strength. Along with more precise calculations of the decoupling of matter and radiation, this meant that the predictions were now reduced from a few parts in a thousand to a few parts in ten thousand.

**Discovery**

The first detection of primordial temperature fluctuations came with the COBE satellite in 1992. These were on angular scales of seven degrees or more, the beam of the microwave telescope. This was a breakthrough in cosmology. It demonstrated that the early universe was not completely smooth and homogeneous. But it was a big extrapolation to get to the scales on which galaxies formed. One could see the mountain ranges but not resolve the individual peaks.

The pace quickened over the following decade. The most important result came with a balloon-borne telescope called BOOMERANG that flew in 1998. This experiment made a definitive discovery from the first high precision detections of temperature fluctuations on degree scales in the cosmic microwave background. The geometry of the universe is Euclidean, or more prosaically, space is flat. The first peak detection was at a level of 0.1 milliKelvin, compared to the 3 degrees Kelvin of the cosmic microwave background and was visible as degree-scale hot and cold spots in the first maps.

This was first precisely predicted in 1984. The logic was as follows. Fluctuations in density grow according to the expansion factor of the universe, which we usually call the redshift. At redshift 3000, the universe first was dominated by matter as opposed to radiation. This meant fluctuation could subsequently grow. But the scales where growth is completed by the present time, galaxies and galaxy clusters are small enough that the radiation leaks before last scattering, so the strength is reduced. The fluctuations on degree scales are about a factor of 3 smaller that comes from the near perfect fit to the later data. The connection between temperature and density fluctuations reduces the temperature fluctuations by another factor of 3. So we end up with a prediction of around 0.1 milliKelvin, as observed. In hindsight, it seems amazing that it took a decade and a half to verify the prediction.

Over the first decade of the 21st century, a new satellite began producing remarkable maps of the CMB. The NASA-launched telescope was initially named as the Microwave Anisotropy Probe. The premature death of its key proponent David Wilkinson, a CMB anisotropy pioneer, led to its being renamed as WMAP when the first scientific results were reported in 2003. WMAP was 45 times more sensitive than COBE and had 30 times finer angular resolution. The new era of precision cosmology was launched with highly detailed maps of the microwave sky. Satellites were crucial as one had to examine the whole sky in order to seek out the cosmic background that is buried in many parts of the sky below galactic foreground emission. Imagine driving with a very dirty wind screen: one has to clean the windscreen to see clearly. One has to clean up the sky by subtracting off the galactic grit.

WMAP instrumentation was based on radiometer technology, and operated from 20 to 90 GHz. This was ideal for removing the radio background, like an obscuring fog produced by relativistic electrons traversing interstellar magnetic fields and generating diffuse synchrotron radiation.

**Confirmation**

The European Space Agency joined the CMB competition in 2009 with the launch of the Planck satellite. Planck included radiometers and bolometers, the combination being designed to operate from 30 GHz to 860 GHz. The addition of the higher frequency channels allowed better angular resolution than achieved by the rival WMAP telescope, and, most importantly, meant that one could do better at removing the galactic foregrounds, especially the dust. Interstellar dust is the dominant galactic emission at the higher frequencies. Planck took data until 2012, when its coolant supply was exhausted.

The WMAP telescope inaugurated the precision era of cosmology. The experiment provided some nine years of data. The Planck telescope continued and refined the search for the primeval fireball radiation. From analyzing the maps of temperature fluctuations across the sky, we can deduce the strength of fluctuations on different angular scales. This distribution of fluctuation strengths can be visualized as a curve with decreasing peaks towards smaller angular scales. Theory tells us that the radiation in the early universe interacted with matter and the effects of gravity generated a series of compression and rarefaction waves, like sound waves. The strongest waves occurred when the radiation last interacted with the matter, ever since, it streams freely towards us. By detecting a series of peaks and troughs on ever decreasing scales, we infer that the waves were generated coherently in the first instants of the Big Bang. The waves eventually peter out in strength towards smaller scales because the radiation leaks out from the smallest fluctuations.

Remarkably, we can express all of this measured information in terms of a few numbers. These are what we call the parameters of cosmology. From WMAP, and then PLANCK, the cosmological parameters were measured to within a few percent. What are these parameters? In fact there are just seven numbers. We are confident in our conclusions because there are several independent ways of measuring some of the key parameters, independently of the cosmic microwave background. These other methods use ground-based and space telescopes to survey the universe at optical and infrared wavelengths.

First there is the average density of baryons, which is of ordinary matter, as measured today. We can directly see the baryons: they are in stars and in diffuse gas. Next there is the density of dark matter. Here we use dynamics: this gives us the total mass of a self-gravitating system such as a galaxy. The difference between luminous matter and total matter gives the density of dark matter. A second technique uses gravitational lensing of distant galaxies; this enables us to make dark matter maps.

The dark energy is the component of the universe which acts like antigravity: it accounts for the measured acceleration of the most distant galaxies. Dark energy is also equivalent to a mass density. The sum of all three components of the density, baryons, dark matter and dark energy, adds up to give us the total density of the universe. There is a critical value, for which there is exactly enough gravity to balance the kinetic energy of the expansion. This critical value, named after its original proponents, Albert Einstein and Willem de Sitter, tells us whether the universe will expand forever if its density is subcritical, or is destined to collapse in the future if supercritical. So far, we cannot distinguish the difference. The future of the universe is uncertain!

The fourth and fifth numbers describe the primordial fluctuations in density that seeded the formation of large-scale structure in the young universe. One is the strength of the fluctuations, and the other describes their distribution, that is the relative numbers of small and large-scale fluctuations. The sixth number controls the visibility of the temperature fluctuations that are generated some 380,000 years after the Big Bang. This was when the radiation last scattered off of electrons. The atomic era then began and the radiation propagated freely to us. In fact, there is subsequently some residual scattering, since the universe became re-ionised about a billion years later. The degree of this scattering controlled by the residual electrons, and we use the term “optical depth” to describe the number of scatterings. If this is too large, the temperature fluctuations would be erased.

The last parameter is the expansion rate of the universe, or the Hubble constant. It is also a measure of the age of the universe and enables us to calculate distances to distant galaxies once we measure their recession velocity.

In summary, our seven numbers are: the density of matter, the density of dark energy, the total density, the amplitude of density fluctuations, the slope of the distribution of density fluctuations, the optical depth, and the expansion rate of the universe. Armed with seven numbers, and the Big Bang model, one has a superb fit to the strength of the temperature fluctuations over a huge range of angular scales, from 90 degrees down to a small fraction of a degree. And the model fit agrees perfectly with what has now been dubbed the standard model of cosmology: an Einstein-de Sitter expanding universe at early times that at late times morphs into an accelerating universe dominated by the cosmological constant. As the density of the standard components of matter and radiation drop, the cosmological constant begins to dominate. This is happening now, within the past ten billion years, or within one space doubling time.

**Polarization**

The cosmic microwave background radiation is very weakly polarized. Polarization is due to twisting of the electric and magnetic fields that define a propagating light-wave. The main cause of polarization is scattering by interstellar dust. The effect amounts to about 1 percent of the fluctuation strength, and increasing measurement precision in cosmology motivates mapping the polarization, and eventually cleaning the sky maps.

However, there is a cosmological polarization signal that amounts to, at most, about 0.1 percent of the fluctuation strength that is generated by inflation in the first instants of the universe. This is caused by gravitational waves. In principle, this polarization occurs everywhere, and in particular in regions so far apart today, many degrees apart on the sky, that there has been no causal contact since inflation. It is a particular type of stretching and shearing, that stretches without compressing and is unique to the effect of gravitational waves as they traverse matter.

Other effects produce a combination of both shearing and compressive modes, and these can, in principle be subtracted to reveal the elusive primordial signal. Detection would be a remarkable verification of inflation, and provide a glimpse of conditions in the universe at an unprecedentedly early epoch.

A detection was tentatively reported in 2014 by the BICEP-2 experiment at the South Pole, but subsequently retracted. It turned out that the polarization signal they measured was due to interstellar dust, an effect that had not been predicted beforehand.

As of 2015, many experiments are competing to detect the gravity wave signal in the cosmic microwave background. The sensitivity needed will need to be at least a hundred times better than WMAP, measuring a few billionths of a degree Kelvin rather than the microKelvin sensitivity of the beautiful WMAP/PLANCK maps. The next decade promises to be an exciting period of discovery. Over the next five years, we will see ground-based and balloon experiments.

The new ground-based experiments are under construction at the South Pole, the high Atacama Desert in Chile, and in Tibet. The balloon telescopes are designed to spend weeks or even months at an altitude of 40 kilometers, and to make circumpolar flights so that the payloads can easily be recovered. A new period of satellite experiments will begin in 2022 with Japan’s LITEBIRD Telescope. It is likely that NASA and ESA will be launching their own missions on a similar time frame.

© Professor Joseph Silk, 2015

Gresham College

Barnard’s Inn Hall

Holborn

London

EC1N 2HH

www.gresham.ac.uk