2.2 The Discovery of the CMB

Scientific discoveries seem to have a natural time to emerge, which is proved by the way that they are often made independently by several people at nearly the same time. Sometimes a new theory makes a clean prediction which is easy to test; sometimes a new technology makes obvious what was previously hidden; and sometimes the ebb and flow of scientific fashion channels research into an area where something wholly unexpected will, in retrospect inevitably, turn up (we call this serendipity). Discoveries made *before* their time, like Einstein's General Relativity, are celebrated as epic feats of prescience. What, then, are we to make of discoveries made *after* their time, discoveries that were banging on the door of science for years but somehow were left out in the cold? The biggest puzzle about the discovery of the Cosmic Microwave Background is why it did not happen ten or fifteen years earlier.

2.2.1 A temperature with a very restricted meaning

In 1935 a massive new spectrograph was commissioned on the 100-inch Mt. Wilson telescope, at that time the largest in the world. The spectrograph was designed for the highest dispersion, spreading a small range of wavelengths over a long photographic plate, so that very fine structure in the spectrum could be recorded. It soon discovered a number of narrow absorption lines in the spectra of bright stars, which were attributed to interstellar gas. Some of these lines could not be identified with atoms of any element, and it was suggested that they were due to simple molecules. This was confirmed in 1941 when a long observation of the star Zeta Ophiuchi by Walter S. Adams found several new, faint lines from the molecules CH and CN (cyanogen), that had been predicted by Andrew McKellar in 1940.

One of the most basic (and counter-intuitive) features of quantum mechanics is that the angular momentum of any object is restricted to multiples of $h/2\pi$, where h is Planck's constant. McKellar had noted that all the lines known in 1940 were from molecules in the **ground state**, with zero angular momentum, which implied that the "rotational" temperature (his quotes) must be very low, < 2.7 K, "if indeed, the concept of such a temperature in a region with so low a density of both matter and radiation has any meaning". But in addition to the ground state lines, Adams' 1941 spectra revealed a faint line from cyanogen molecules with one unit of rotation, and from its the relative intensity McKellar was able to work out the relative number of molecules in each state; in turn this depends only on the rotational temperature. A short note in 1941 gave the result: CN had a rotational temperature of 2.3 K.

As the comment I have quoted shows, McKellar did not take this number very seriously, and published nothing more on interstellar molecules (he was more interested in comets). Adams continued with his observations, but was mislead by the coincidence of this figure with the "temperature of space" recorded in Eddington's book (Section 2.1.4). Adams commented on the "excellent agreement" with theory. Nobody noticed at the time that equivalence in energy density is not enough. CN molecules in their ground state are an example of the kind we discussed earlier, where absorption is restricted to effectively a single wavelength, λ2.64 mm in this case. Eddington's model of the interstellar radiation field predicts a brightness temperature of around 10^{-8} K at this wavelength (see Ned Wright's web site); even by astronomical standards, this is not in good agreement with 2.3 K!

McKellar's rotational temperature became quite well known, as the discovery of interstellar molecules had many implications in astrophysics. A standard text (Herzberg 1950) quoted the value of 2.3 K, "which has of course a very restricted meaning".

2.2.2 Ylem and after

In 1941 astronomers could be forgiven for ignoring cosmology, which had developed into a rather dry subject with little contact with the rest of astrophysics.

The situation changed after the war, when the early universe caught the imagination of George Gamow. Gamow! If you had heard only a quarter of what I have heard about him, and I have only heard a very little of all there is to hear, you would be prepared for any sort of remarkable tale. He was a big Russian who had been on holiday (he said) in the USA since 1934 (he never did get around to going home). If a bowler hat shattered (after being filled with liquid air), or mysterious coats of arms appeared on office doors at Los Alamos, or a paper on nuclear physics was posted to *Nature* apparently from a mountain-top, you could be sure Gamow was at the bottom of it. He was the first person to explain α-radiation in terms of quantum physics, and to find some of the crucial steps in the way stars generate energy. His physical insight was legendary, as were his practical jokes, his disregards for trivial details (such as getting sums right), and his drinking. He also wrote the *Mr Tompkins* books.

Figure 2.3: The genie Gamow emerges from a bottle of Ylem, flanked by Robert Herman and Ralph Alpher. The original bottle is now in the National Air and Space Museum, Washington DC.

Gamow and his small band of associates, notably Robert Herman and Ralph Alpher, saw the early universe as the ultimate nuclear reactor. Their aim was nothing short of explaining the origin of all the chemical elements in a series of nuclear reactions in the first few minutes, starting from a soup of neutrons they called **ylem** (pronounced "eye-lem", a medieval word for primordial matter). The time was ripe: nuclear physics had just become a well-tested science and could now be applied to "cosmic creation", as Gamow called it (his rival, Fred Hoyle, christened Gamow's theory the "Big Bang"). As Gamow's team studied the big bang in more detail, they realised that the ylem could not have been pure neutrons; in fact it mostly consisted of gamma-rays, forming a heat bath with a temperature of about 10^9 K at the time the elements were formed. In 1948 Alpher and Herman had the crucial insight that these photons would still be around today, cooled by the expansion of the universe to, they estimated, 5 K.

Over the next few years the group's estimate for the temperature fluctuated between 6 and 50 K, partly due to revised astronomical data and partly because they (especially Gamow) kept trying out different approaches to deriving the temperature, some of which were very rough and ready. Unfortunately they never proposed an experimental search for these photons. One reason was that Gamow had heard about Eddington's "temperature of space" and thought that this would be difficult to disentangle from the relic photons from primordial nucleosynthesis. Indeed, Fred Hoyle once told Gamow that his theory was ruled out by the low interstellar temperature discovered by McKellar! Like Adams and McKellar before them, Gamow and Hoyle failed to spot the enormous difference between starlight with energy density equivalent to a 3 K blackbody, and a genuine blackbody spectrum. It is also true that Alpher and Hermann had privately consulted radio astronomers and experimental physicists, and were told that the measurement was impossible; an unduly pessimistic assessment as it turned out.

Gamow popularised his theory with public lectures, books, TV debates with Hoyle, and had even written to the Pope to recommend it. But in the mid-'50s things went wrong. The dream of making all the elements in the big bang proved impossible: the process could make only the lightest, helium and lithium. Meanwhile Hoyle and his collaborators counterattacked with a spectacular paper showing how nearly all the elements could be made in stars. They hoped the one exception, helium, could soon be included as well. Gamow's theory had been taken as far as it could go and his group dispersed. The general attitude of astronomers and physicists at the time seems to have been that a highly speculative theory of "creation" had failed. It was not widely appreciated that the early universe described by Gamow was a more or less inevitable implication of the basic expanding universe picture, which never went out of style despite strong competition from the Steady State theory pushed by Hoyle.

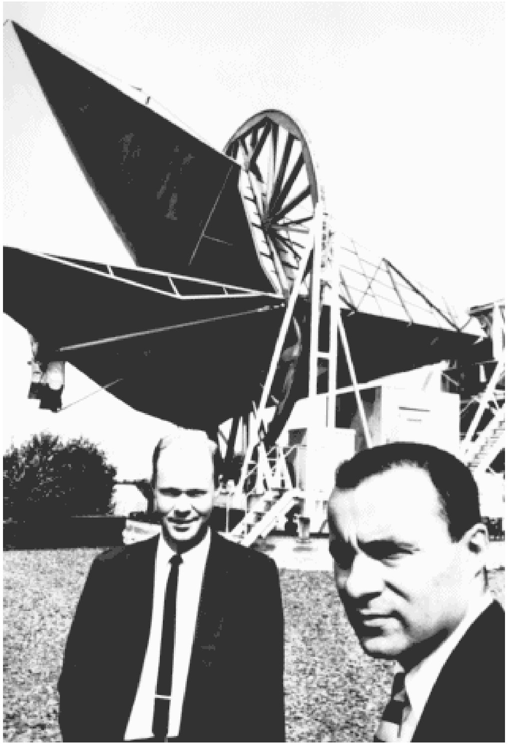
2.2.3 The horn

The CMB may have been detected several times in the 1950s without anybody noticing. This was the pioneering era of radio astronomy, when the sky was being surveyed for the first time at many frequencies in the radio band. In order to make this sort of measurement it is conventional to use a "cold load", typically a pot of liquid nitrogen or helium simmering at its boiling point (77 K or 4 K respectively) to give a constant-brightness reference. The input to the receiver is switched rapidly between the main antenna and a small one looking at the load, so that the noise generated by the receiver itself (typically 100 K in those days) would be the same for both. In this way small differences between sky and load temperature can easily be seen, which would otherwise have been lost in the uncertainty of measuring the receiver noise. This technique is known as Dicke switching, after its inventor.

In 1955 E. Le Roux surveyed the sky at λ33 cm wavelength from the radio observatory of Nançay, near Paris. He reported highly isotropic emission with brightness temperature 3±2 K. In 1957 T. A. Shmaonov in the Soviet Union found a similar result, T_B = 4±3 K at λ3.2 cm. Neither of these results are formally significant according to the quoted errors, but both are suspiciously close to the CMB temperature. It seems likely that the error bars were increased because of the `obvious' 3 K systematic error! Probably neither observer knew of the other's work, or of McKellar's.

By now the Space Age had arrived. In 1960 NASA launched its first communications satellite, Echo, a large aluminised balloon that just acted as a reflector, allowing signals to be bounced between ground stations on the East and West coasts of the USA. The West coast station was NASA's Goldstone facility; the one on the East coast was provided by Bell Telephone Laboratories, which built a large microwave horn antenna at its experimental station near Holmdel, New Jersey. While calibrating this system, Ed Ohm found an excess temperature of 3.3 K at λ11 cm, *without* using a cryogenic cold load; by 1961, technology had far surpassed what was necessary to detect the background radiation! This time, the result did not pass unnoticed by astronomers. Soviet scientists were keeping a close eye on goings-on at Bell Labs, and Doroshkevich and Novikov cited Ohm's paper in a 1964 review of Gamow theory. Unfortunately, they misunderstood this key result: they thought Ohm attributed the excess temperature to emission from the atmosphere, whereas in fact it was a residual *after* atmospheric emission of nearly the same amount had been subtracted from the signal. Yet again, a detection of the CMB was cited as evidence against its existence!

Figure 2.4: Arno Penzias (right) and Robert Wilson with the The Bell Labs Horn. Credit: Penzias' home page at Bell Labs.



When the Echo and the follow-up Telstar projects were completed, the horn was taken over by two new recruits, Arno Penzias and Robert Wilson, to do some radio astronomy at λ7 cm. They *did* use a cryogenic (liquid helium) cold load and there could be no doubt at all about the excess temperature. They spent an entire year trying to eliminate it on the assumption that it was a problem with the horn. This included scrubbing out pigeon droppings that had accumulated over the years. None of this had any significant effect on the excess temperature, and by 1965 they were at a loss.

Meanwhile, at Princeton University, a few miles away from Holmdel, Robert H. Dicke had re-invented the cosmic background radiation. Dicke was the leading experimental physicist of his day. His career had started with radio science; Dicke switching was just one of his inventions.

In the early 1960s he developed his own cosmological theory, in which the universe oscillated in size, passing through a very compact phase (although not the infinitely compact "singularity" predicted by General Relativity). On this theory, light emitted in previous cycles would be compressed at the bounce to high energies, essentially duplicating the conditions of Gamow's Big Bang (Dicke always claimed that he had forgotten about Gamow's work, despite having heard Gamow lecture on the subject. Gamow did not believe him).

Dicke's group had began to build a receiver system to look for the radiation when Penzias heard about their prediction through the radio astronomy grapevine, and phoned to find out more. It happened that Dicke's team were having their weekly `brown bag' lunch in Dicke's office when the call came through. Dicke mostly listened but occasionally repeated crucial phrases - `horn antenna,' `liquid He calibrator,' `excess noise,'... At the end Dicke hung up, turned to the others and said `Well boys, we've been scooped.'[2.1]

In July 1965 the *Astrophysical Journal* published back to back papers by the two groups. The Princeton group outlined the implications of a 3.5 K background for big bang cosmology, and mentioned that they had been about to start their own search when they heard about the Bell Labs result. Penzias and Wilson's contribution, "A measurement of excess antenna temperature at 4080 Mc/s" was just over a page long and focussed entirely on convincing the reader that the detection was real. They laconically noted: `A possible explanation for the observed excess noise temperature is the one given by Dicke, Peebles, Roll, and Wilkinson'. This paper won Penzias and Wilson the 1978 Nobel Prize for Physics.[2.2]

2.2.3.1 Microwaves

Figure 2.5: The Post Office Tower in London is the heart of the UK network of microwave telecom links. Older readers may remember that originally the antennas on the tower were smaller versions of the Bell Labs horn.

The words used to describe portions of the radio spectrum reveal a quest for ever-shorter wavelengths and so higher frequencies. Shorter wavelengths make it possible to build highly directional, and hence efficient, antennas: from the laws of diffraction, an antenna of size *D* will produce a beam with angular width $\theta \approx \lambda/D$. Short-wave, used for ham radio and the BBC world service, has wavelengths of tens of metres. Beyond that we have Very High Frequency and then Ultra High Frequency (used for TV broadcasts), with wavelengths of tens of centimetres. Microwaves, developed during World War II for radar, were the next step: wavelengths of a few centimetres. They are also used for narrow-beam telecommunication links, both cross country (Fig. 2.5) and to satellites. The Bell Labs horn operated at λ7 cm, smack in the middle of the microwave region (for comparison, microwave ovens use λ= 12 cm), and so naturally its famous discovery was christened the microwave background. But these original observations were in the long-wavelength tail of the CMB spectrum, which peaks at about λ= 1 mm. To radio engineers this qualifies as mm-wave rather than microwave, so CMB is perhaps a misnomer. It has certainly confused a number of cosmologists into thinking that "microwave" refers to millimetre rather than centimetre wavelengths.