From issue 2593 of New Scientist magazine, 03 March 2007, page 33-37

Inside inflation: after the big bang

Peter Coles

MASSACHUSETTS, 1981. A young physicist comes up with what seems to be an absurd idea: the universe went through a period of ultra-fast expansion just after the big bang. Alan Guth cannot prove that this "inflation" actually happened nor can he suggest a compelling physical reason why it should have, but the idea seems nevertheless to solve several major problems in cosmology.

Fast forward to today. Guth is a professor at the Massachusetts Institute of Technology and inflation is now well established as an essential component of cosmology. But should it be?

There is little direct evidence that inflation actually took place. Observations of the cosmic microwave background - "fossil" radiation from the big bang - are consistent with the idea that inflation took place, but that doesn't mean it actually happened. What's more, we still don't know what would have caused it if it did. So how confident can we be that inflation is really a part of the universe's history?

A quarter of a century ago, our understanding of the universe was much less precise than it is today. In those days it was a domain in which theoretical speculation ruled over measurement and observation. Technology simply wasn't up to the task of performing large-scale galaxy surveys or detecting the all-important details in the cosmic microwave background (see "Shadow of the big bang").

The lack of stringent experimental constraints made cosmology a theorists' paradise in which many imaginative and esoteric ideas blossomed. Not all survived to be included in the standard model of cosmology, but inflation has proved to be one of the hardiest, and indeed most beautiful, flowers in the cosmological garden.

Although some of the concepts involved had been formulated in the 1970s by Russian physicist Alexei Starobinsky, it was Guth's 1981 paper that first crystallised the picture of the inflationary universe. At this time cosmologists didn't know that the universe was as flat as we now believe it is, but it was still a puzzle why it was even anywhere near flat. After all, the great breakthrough of Einstein's general theory of relativity was the realisation that space could be curved. Of all the possible initial conditions, isn't it very improbable that our universe should be flat?

What's more, the distribution of stuff in our universe is also astonishingly smooth. Although it contains galaxies that cluster into immense chains more than a 100 million light years long, on scales of billions of light years it is almost uniform. This also seems surprising. Why is the celestial tablecloth so immaculately ironed?

Guth grappled with these questions and realised that they could be resolved rather elegantly if only the force of gravity could be persuaded to change from pull to push for a very short time just after the big bang. The expansion of the universe would then speed up rather than slow down. The universe could then inflate by an enormous factor (1030 or more) in a fraction of a millisecond. Even if it were initially curved and wrinkled, all memory of this messy starting configuration would be wiped out. The present-day universe would be very flat and very smooth no matter how it had started out.

So how could this bizarre period of anti-gravity be possible? Guth hit upon a simple physical mechanism by which inflation might just work in practice. It relied on the fact that in the extreme conditions just after the big bang, matter would not adhere to the classical laws describing gases and liquids but must instead be described by quantum field theory. The simplest type of quantum field is called a scalar field; such objects are associated with particles that have no spin, the quantum parallel of angular momentum. Modern particle theory involves many scalar fields that are not observed in low-energy interactions, but which may well dominate affairs at the extreme energies of the primordial fireball.

Just as classical fluids can undergo a "phase transition" if they are heated or cooled, such as the transition from steam to liquid water, a similar thing happens with scalar fields: their configuration is expected to change as the universe expands and cools. Phase transitions do not happen instantaneously, however, and sometimes a bubble of the substance involved can get trapped in an uncomfortable state in between where it was and where it wants to be, like a bubble of gas trapped in a liquid.

Guth realised that if a scalar field got stuck in such a false state, it could free up energy - in a form known as vacuum energy - to drive a small piece of the universe into accelerated expansion. In the process, the tiny bubble can inflate to the size of the entire universe. We don't know which scalar field of the many that may theoretically exist is responsible for generating inflation, but whatever it is, it is now dubbed the inflaton.

This mechanism is an echo of a much earlier idea introduced to the world of cosmology by Einstein in 1916, although he didn't use the term vacuum energy. He called it a cosmological constant, and also considered it to be a modification of the law of gravity rather than something arising from quantum fields. Nevertheless, Einstein's idea was incorporated by Dutch mathematician Willem de Sitter into a theoretical model of an accelerating universe. This is essentially the same mathematics that is used in modern inflationary cosmology.

The connection between scalar fields and the cosmological constant may also eventually explain why our universe's expansion seems to be accelerating now - something that has been attributed to a mysterious force called dark energy. However, that would require a scalar field with a much lower energy than that required to drive inflation. Perhaps dark energy is some kind of shadow of the inflaton.

Guth wasn't the sole creator of inflation. About the same time, many others including Andy Albrecht, Paul Steinhardt, Andrei Linde and Starobinsky, produced different and in some cases more compelling variations on the basic theme. It was almost as if it was an idea whose time had come. Within just a few years inflation had become an indispensable part of cosmological theory.

Literally hundreds of versions appeared in the leading scientific journals: old inflation, new inflation, chaotic inflation, extended inflation, and so on. Out of this activity came the realisation that a phase transition wasn't really necessary, all that mattered was that the scalar field should find itself in a state where the vacuum energy dominated.

It also became clear that even theories that didn't involve scalar fields could behave as if they did. Modified gravity or theories with extra space-time dimensions provided ways of mimicking scalar fields with rather different physics. And if inflation could work with one scalar field, why not have inflation with two or more? The only problem was that there wasn't a shred of evidence that inflation had actually happened.

This episode provides a fascinating glimpse into the historical and sociological development of cosmology in the 1980s and 1990s. Inflation is undoubtedly a beautiful idea, but the problems it solves are theoretical, not observational. For example, the flatness of the universe only appears to require this explanation because we don't have a theory of initial conditions for the universe that might provide a better reason. Inflation turns an initially curved universe into a flat one, but the fact that the universe appears to be flat doesn't prove that inflation happened.

Certain initial conditions could lead to present-day flatness without the intervention of an inflationary epoch. One might argue that these are special cases and therefore "improbable", making it more probable that inflation happened than that it didn't. On the other hand, without a theory of the initial conditions how can we say which are more probable? Based on this kind of argument alone, we could probably never work out whether we live in an inflationary universe or not.

But there is another thread in the story that makes it a much more compelling scientific theory, because it makes direct contact with observations. Although it was not the original motivation for the idea, Guth and others realised very early on that if a scalar field were responsible for inflation then it should be governed by the usual rules of quantum fields.

One of the things that quantum physics tells us is that no quantum process evolves entirely smoothly. The famous Heisenberg uncertainty principle imposes a degree of unpredictability on the behaviour of the inflaton. The upshot of this is that although inflation smoothes away any primordial wrinkles in the fabric of space-time, in the process it lays down others of its own.

The inflationary wrinkles are really ripples, wave-like density fluctuations vibrating through the matter of the early universe like sound waves travelling through air. Without these fluctuations the cosmos would be smooth and featureless, containing no variations in density or pressure, and therefore no sound waves. Even if it began in a fireball, such a universe would be silent. Inflation puts the "bang" in big bang.

The acoustic oscillations generated by inflation comprise a wide range of wavelengths. Most importantly of all, they are formed "coherently". That is, because inflation happens so rapidly, all of the acoustic wavelengths are excited at the same time, just as hitting a metal pipe with a hammer generates a wide range of frequencies, all starting at the same time. The result is not just random noise but something more tuneful. The big bang wasn't exactly melodic, but there is a discernible relic of the coherent nature of the sound waves in the pattern of temperature fluctuations in the cosmic microwave background as seen by NASA's Wilkinson Microwave Anisotropy Probe (WMAP). The hot and cold spots of the microwave background reflect denser or rarefied regions, and the acoustic peaks seen by WMAP offer compelling proof that whatever generated the pattern did so coherently.

There are very few alternative theories capable of reproducing the WMAP results. Some interesting ideas have emerged recently from string theory. Since this theory requires more space-time dimensions than the four we are used to, something has to be done with the extra ones we don't observe. For example, in so-called braneworld cosmologies our four-dimensional universe exists as a subset (called a brane) of a more multi-dimensional space.

This idea may one day lead to a viable alternative to inflation. But it is early days and not all the calculations needed to establish this theory have yet been done. In any case, not every cosmologist feels the urge to make cosmology consistent with string theory, which has even less evidence in favour of it than inflation. So does WMAP prove inflation happened? If not, will we ever know?

It is difficult to talk sensibly about scientific proof of phenomena that are so far removed from everyday experience. At what level can we prove anything in astronomy? We all accept that the Earth goes around the sun, but do we really know for sure that the universe is expanding? I would say that the latter hypothesis has survived so many tests and is consistent with so many other aspects of cosmology that it has become, for pragmatic reasons, an indispensable part of our world view. But I would hesitate to say that it was proven beyond all reasonable doubt.

The same goes for inflation. It is a beautiful idea that fits snugly with standard cosmology and binds many parts of it together but that doesn't necessarily make it true. Many theories are beautiful, but that is not sufficient to prove them right. When generating theoretical ideas scientists should be fearlessly radical, but when it comes to interpreting evidence we should all be deeply conservative.

As for the future of cosmology, WMAP has provided a tantalising glimpse of further evidence and paved the way for even more stringent tests of the standard framework. Primordial density fluctuations produce not only a pattern of temperature variations over the sky, but also a corresponding pattern of polarisation. This is fiendishly difficult to measure, partly because it is such a weak signal (only a few per cent of the temperature signal) and partly because the microwaves are heavily polluted by polarised radiation from our own galaxy. Although WMAP did indeed detect polarisation, the published map is heavily corrupted by foreground noise.

Future generations of experiments, such as the European Space Agency's Planck Surveyor, due for launch in 2008, will have to grapple with the thorny issue of foreground subtraction if it is to make progress. But there is a crucial means of cross-checking the results that would justify these endeavours. The key is that inflation does not just produce acoustic waves, it also generates twisting deformations of space-time called gravitational waves.

Gravitational waves produce a very particular form of polarisation pattern that cannot be generated by acoustic oscillations, so hunting for this signal seems a promising way to test inflation. Even though it is a very weak signal, and the experience of WMAP suggests it might be swamped by foreground noise, it is definitely worth a go. Finding it would add considerably to the evidence in favour of inflation as an element of physical reality.

Besides providing strong evidence for the standard model of cosmology, WMAP has also provided tantalising evidence that we may be missing something. Not all the properties of the microwave sky seem consistent with the model. For example, the pattern of hot and cold spots should be structureless, mirroring the random fluctuations of the primordial density perturbations. In reality, certain components of the pattern are inexplicably aligned, as in the so-called "axis of evil" discovered in 2005 by Kate Land and João Magueijo of Imperial College London (New Scientist, 22 October 2005, p 19). These anomalies could be systematic errors in the data, or perhaps residual foreground effects that need to be subtracted, but they could equally indicate the presence of things that can't be described within the standard model.

Cosmology is now a mature and respectable science. Yet there are still many gaps in our knowledge. We don't know the form of the "dark matter" responsible for unexplained extra gravity. Nor do we have any real understanding of dark energy. We don't know for sure if inflation happened, and we are certainly a long way from being able to identify the inflaton. In a way we are still as confused as ever about how the universe began. But perhaps now we are confused on a higher level and for better reasons.

Our ability to reconstruct the history of the universe, or at least to attempt this feat, depends on the fact that light travels with a finite speed. The further away we see a light source, the further back in time its light was emitted. We can now observe light emitted from stars in distant galaxies when the universe was less than a sixth of its current size. In fact we can see even further back than this using microwave radiation rather than optical light.

Our universe is bathed in a faint glow of microwaves produced when it was about one-thousandth of its current size and had a temperature of thousands of kelvin, rather than the chilly 3 K that prevails today. The existence of this cosmic background radiation is one of the key pieces of evidence in favour of the big bang model. It was discovered in 1965 by Arno Penzias and Robert Wilson, for which they subsequently won the Nobel prize.

It is not just the cosmic microwave background that has helped us construct the standard model of cosmology. Observations of distant supernovae and the pattern seen in the large-scale distribution of galaxies have also offered hints. The picture that has emerged from these disparate clues is of a universe dominated by dark energy and dark matter, in which the early stages of cosmic evolution involved an episode of accelerated expansion called inflation.

Assembling the standard model of cosmology has been a gradual process, reaching its latest form with recent results from NASA's Wilkinson Microwave Anisotropy Probe (WMAP). For several years this satellite has been mapping the properties of the cosmic microwave background and how it varies across the sky. Small variations in the temperature of the background reflect sound waves excited in the hot plasma of the primordial fireball (see Diagram). Various telltale properties of these waves allow us to probe the early universe in much the same way that solar astronomers use observations of the surface of the sun to understand its inner structure.

The detection of the primeval sound waves is one of the triumphs of modern cosmology, not least because the amplitude of the waves tells us precisely how loud the big bang really was. The fundamental tone tells us that the universe is very nearly flat, while the overtones pin down a dozen or so important cosmological parameters to unprecedented accuracy - a truly remarkable achievement.

BRIEF HISTORY OF TIME

Inflation lasted less than a millisecond, yet it is responsible for most of the growth of the universe and left its imprint on the microwave background (inset)