From issue 2601 of New Scientist magazine, 28 April 2007, page 28-33

The cosmos - before the big bang

How did the universe begin? The question is as old as humanity. Sure, we know that something like the big bang happened, but the theory doesn't explain some of the most important bits: why it happened, what the conditions were at the time, and other imponderables.

Many cosmologists think our standard picture of how the universe came to be is woefully incomplete or even plain wrong, and they have been dreaming up a host of strange alternatives to explain how we got here. For the first time, they are trying to pin down the initial conditions of the big bang. In particular, they want to solve the long-standing mystery of how the universe could have begun in such a well-ordered state, as fundamental physics implies, when it seems utter chaos should have reigned.

Several models have emerged that propose intriguing answers to this question. One says the universe began as a dense sea of black holes. Another says the big bang was sparked by a collision between two membranes floating in higher-dimensional space. Yet another says our universe was originally ripped from a larger entity, and that in turn countless baby universes will be born from the wreckage of ours. Crucially, each scenario makes unique and testable predictions; observations coming online in the next few years should help us to decide which, if any, is correct.

Not that modelling the origin of the universe is anything new. The conventional approach is to take the laws of physics and extrapolate backwards from the present. From observations dating back to the 1920s, we can see that galaxies are moving farther and farther apart: the universe is expanding. By reversing that expansion, researchers concluded that 13.7 billion years ago the universe was in a very small, dense and hot state. The big bang theory, first proposed in 1927 by Georges Lemaître, was bolstered in 1964 by the discovery of the cosmic microwave background - the radiation filling the universe that is thought to be a relic of the big bang - and has ruled ever since.

In 1981, a major addition was made to the big bang picture. Alan Guth of the Massachusetts Institute of Technology and others proposed that the expansion of the early universe happened much faster than originally thought. This theory, called cosmic inflation, explained the surprising uniformity of the visible universe by saying that it grew exponentially from a patch that was extremely tiny to start with (New Scientist, 3 March, p 33). Though highly successful in this regard, inflation still doesn't explain the initial conditions of the universe.

That's because inflation would have taken place between 10-35 and 10-32 seconds after the big bang. Going back further in time, we hit a brick wall because the two pillars of modern physics - quantum field theory and general relativity - break down. Physicists don't have a complete recipe with which to concoct the behaviour of matter, energy and space-time under such extreme conditions, and it's hard to blame them.

To get around this, some are basing their ideas around an age-old tenet. The second law of thermodynamics dictates that the entropy of the universe - a measure of its disorder - increases with time. So the universe began in its most orderly state and has been getting messier ever since. The problem is, it would have been more likely to be chaotic and disordered, so what was this initial state? "It's tremendously important that any respectable model of the early universe explains why entropy is so low near the big bang," says Sean Carroll, a cosmologist at the California Institute of Technology in Pasadena.

Enter the first of the new models. The entropy question has led Thomas Banks of the University of California, Santa Cruz, and Willy Fischler of the University of Texas at Austin to conclude that the universe in its earliest moments - when it was less than 10-35 seconds old - was a sea of black holes. They call this scenario "holographic cosmology".

The idea is based on the holographic principle, which was proposed in 1993 by Gerard't Hooft of Utrecht University in the Netherlands and developed by Leonard Susskind of Stanford University in California. Although it is unproven, many physicists think the holographic principle is right: all the information in a given volume of space can be represented by physical laws that exist on its surface. Entropy can be thought of as a measure of information content - the more disordered a system, the more information it takes to describe it. Cast in these terms, the holographic principle says the entropy in a given volume is limited by its surface area, and maximised in the case of a black hole.

Now imagine turning back the clock towards the big bang. Matter and energy get packed together more densely into each shrinking region of space until we reach the entropy density limit, which corresponds to filling up these regions with a sea of microscopic black holes.

According to Banks and Fischler, the universe began as this black hole "fluid" (see Diagram). From any vantage point, black holes would fill the entire space around, but how densely they fill it would fluctuate according to the uncertainty principle of quantum mechanics. A fluctuation towards lower density would mean that in that region the black hole event horizons would not fill every last bit of volume, but would have some ordinary space between them, free of black holes and filled with radiation.

This creates the conditions for our observable universe to come into existence. If the black holes in the region where ordinary space opens up are densely packed and moving fast, their collisions and mergers make them grow until they fill the space, pulling it back into the black hole fluid. But if the black holes are far enough apart and moving slowly, mergers won't happen fast enough. In such a region the ordinary space filled with hot radiation would quickly expand, pushing the black holes further apart.

About 10-35 seconds after the beginning of time, this bubble of ordinary space joins up with the conventional picture, in which inflation expands our universe to more than 1 kilometre across in a tiny fraction of a millisecond. Eventually, particles condense out of the radiation to produce the building blocks of stars, galaxies, planets and life.

So how do Banks and Fischler explain the low entropy of the early universe? Many bubbles of ordinary space could have emerged from the black hole fluid, but to avoid collapsing back into the fluid, they need to have low entropy (www.arxiv.org/hep-th/0701146). That's because higher entropy corresponds to faster-moving black holes that are prone to colliding and merging. If our bubble of space had begun with higher entropy, it would not have survived. "There wouldn't have been a universe to live in," Banks says.

Other researchers are still debating the merits of holographic cosmology. "It's a very interesting speculation that is neither obviously true nor obviously false. Time will tell," says Susskind. After all, he says, "there is an enormous gulf separating the earliest origin from observational cosmology".

The model raises the controversial issue of whether time began at the big bang. "There's no necessity in the rules of quantum mechanics for time to extend out to the infinite past, or for that matter, the infinite future," Banks says. An origin of time has its own problems, though. "If there were no beginning, I would sleep better at night. I think it would be more elegant," says Max Tegmark, a cosmologist at MIT. A beginning of time raises the question of "why certain things come into existence and others don't".

In other words, this approach does not explain the origin of the big bang, says Paul Steinhardt of Princeton University. In 2002 he and Neil Turok of the University of Cambridge proposed a scenario in which the big bang is not the beginning of time, but just the start of another cosmological cycle (New Scientist, 16 March 2002, p 26). Their model, which has withstood some recent challenges, provides a different mechanism for the low entropy of the early universe.

Steinhardt and Turok's model is motivated by string theory, an approach to unifying relativity and quantum mechanics in which there are extra dimensions of space beyond the three we can see. In their model, our visible universe is a 3D sheet called a membrane, or brane, floating in four-dimensional space (see Diagram). Another 3D brane, with possibly very different physics, hovers nearby. The branes collide every so often, making ours heat up to an astronomical 1023 kelvin and expand, with some energy eventually condensing into matter. From our point of view, confined to our brane, it would look like a big bang - even though the universe was already there.

After the branes collide they separate and stretch out, causing the expansion of space within them to speed up. This corresponds to the accelerated expansion of the universe that researchers observe today and explain by invoking a repulsive force known as dark energy (New Scientist, 17 February, p 28). The branes will eventually slow down, stop and start hurtling towards one another again. Whenever the next collision occurs, new matter and radiation will be injected into our brane, as if a new big bang has gone off.

One potential problem with this "cyclic brane" model is that small differences in the distribution of matter and energy within our brane could get amplified during a collision, leading to a lumpy universe that looks nothing like ours. Steinhardt and Turok have argued, however, that dark energy becomes stronger as the branes approach one another, and that this overwhelms the small fluctuations, keeping the universe smooth.

The cyclic brane model might seem radically different from Banks and Fischler's black hole fluid scenario - what's more, it does not invoke conventional inflation - but remarkably they share some common ground. Black holes would be produced in copious amounts under the extreme conditions of a brane collision, Steinhardt says. "Maybe it's not so different from the state that Banks and Fischler have in mind," he says.

Yet its explanation of the low-entropy question is quite different. The second law of thermodynamics makes it hard for a given cosmological cycle to start with low entropy: you'd think entropy would have accumulated in previous cycles. The brane scenario solves this problem. The stretching of each brane means that matter, radiation and entropy all get enormously diluted before a collision. By the time of the "big bang" that follows, the entropy density - and therefore the total entropy that any observer can see - is very low. To get enough dilution, the universe must go at least a trillion years between collisions.

Though intriguing, the model has yet to gain widespread support. "It's quite specific, and it does try to be an alternative to inflation, which is absolutely a good thing to have," says Carroll, but he is still unconvinced. "It's not very clear to many people why this would be considered an improvement [on inflation]."

As for the beginning of time, there is no way to tell whether the cycling has been going on forever. "We don't know yet how to make that into a scientifically decidable question," Steinhardt says. The problem is that information about previous cycles tends to get scrambled. Even if the cycling had a beginning, there may be no way to detect it. Nevertheless Steinhardt remains optimistic. "We addressed a lot of the show-stopper problems that might have stopped people from thinking about cyclic models," he says. "That's really opened the door for people to come up with other imaginative ideas that take us back to the big bang and beyond."

One such model that has emerged says our universe began as a fragment of a mother universe shattered by dark energy, and that our universe will in turn give rise to countless others. Developed by Lauris Baum and Paul Frampton, both from the University of North Carolina in Chapel Hill, the scenario also manages to get around the problem of accumulating entropy, but in a different way (Physical Review Letters, vol 98, 071301).

Their model starts with the assumption that the amount of dark energy in a given volume increases as the universe expands. This is plausible, as measurements to date of dark energy are imprecise. A slowly increasing density would lead the repulsive force to destroy galaxies, stars and even individual atoms, culminating in an irreversible disaster called the "big rip" in which the universe's expansion rate becomes infinite. So Baum and Frampton designed the model's dark energy to have an attractive force as well that starts out negligible but later grows quickly; the repulsive aspect dominates when the universe is young and small, which is still the case now.

According to their scenario, the universe is expanding faster and faster, diluting matter and radiation enormously. Eventually, each patch of the universe moves away from other regions faster than the speed of light. This does not violate the speed limit dictated by relativity, since the expansion of space itself is happening faster than light, rather than the motion of particles through that space. Since no particle or force can travel faster than light, each patch is cut off from the others and becomes an island universe.

Left just a bit longer, this process would lead to the end of the universe, but at the last instant, less than 10-27 seconds before a would-be big rip, the attractive aspect of the dark energy finally overtakes the repulsive part. This causes each island universe to contract, but eventually it gets so dense that its radiation reverses the contraction. We are left with innumerable expanding little universes - of which ours may have been one (see Diagram). At this point, the model joins up with the standard inflation scenario, and matter eventually clumps together to form the stars and galaxies we see around us.

What about the low-entropy question? As in the cyclic brane model, the fragmenting universe manages to avoid being hobbled by the accumulation of entropy from cycle to cycle. At the end of each cycle, the entropy that has been produced is divided among the huge number of new universes spawned from the fragmentation of the old one. As a result, the baby universes each begin with a clean slate.

Far in the future, the whole process will repeat itself, spawning countless new universes from the wreckage of ours. This suggests that the number of universes was smaller in the past. If we go back far enough, was there an original universe that started it all? In other words, would time still have a beginning? No, says Frampton. "I would say the number of universes is and always has been and always will be infinite," he says.

Others find the scenario fascinating but incomplete. "It's a kind of new idea," says Steinhardt. The model pushes entropy outside the borders of our early universe, he says. "But how do you get this turnaround? That remains to be explained." Some are more dubious of cyclic models in general. "I have not seen any theory that's convinced me that it really works forever into the past," says Tegmark.

Any kind of conviction will require new experimental evidence. Fortunately, the two cyclic models make very different predictions that should allow researchers to choose between them. Dark energy appears in both, but its behaviour is different. In order for the fragmenting universe scenario to work, the dark energy first has to grow stronger - more and more dense - as the universe expands. Physicists denote different behaviours of dark energy using a parameter they call w, which describes how dark energy varies with time.

Dark energy that stays the same as the universe expands corresponds to a w of -1, and is sometimes called a cosmological constant. Dark energy that increases with time, as in the fragmenting universe, corresponds to a w with a more negative value, for example, -1.05. By contrast, in the cyclic brane model, dark energy results from the potential energy between the two branes, which depends on how far apart they are. As the branes move apart, as they would be now, dark energy's strength decreases. This corresponds to a w that is greater than -1, for example, -0.95.

Since dark energy affects the universe's expansion, researchers can look for changes in its strength by measuring the rate of expansion at different times in the universe's history. Astronomers do this by using supernova explosions; these allow them to measure the speed of receding galaxies at different points in time. Of course, this method can only tell us about dark energy after stars formed, but the cosmic microwave background can be used to chart its strength back to a much earlier time, 380,000 years after the big bang, when the universe first became transparent to light. Looking nearly 13.7 billion light-years away in any direction, we see the radiation emitted by the hot gas that filled the early universe. From this background radiation, astronomers can measure the recession speed of the gas, which tells us how fast the universe was expanding at the time.

Combining the two methods suggests that dark energy is constant or nearly constant, with w close to -1. That is where new measurements come in. The European Space Agency (ESA) Planck satellite, scheduled to launch in 2008, will measure the microwave background with the greatest precision to date, allowing w to be calculated to within about 1 per cent. If Planck shows w to be definitively on one side or the other of -1, then one of the two cyclic models would be ruled out. If it is very nearly -1, both would suffer. "Let's hope it's not too close," says Frampton.

Testing holographic cosmology and its sea of black holes is likely to be more difficult. One piece of evidence is potentially observable: black holes from the early universe, some of which should have survived to the present day. "That would be something to look for," Banks says.

Primordial black holes are also produced in the cyclic brane scenario, but they would be tiny and would be expected to evaporate a fraction of a second after their birth through a process called Hawking radiation.

The largest black holes from holographic cosmology, though less than 100 grams, might survive to the present day because of a strange property: they would possess a magnetic field with just one pole. All magnets observed to date come in north-south pairs, but physicists believe that "monopoles" - magnetic particles with only one pole - would have been produced in the early universe. The relic black holes would have sucked in large numbers of monopoles, which, crucially, could be as big as 1016 times the mass of a proton. Particles that large tend to resist being ejected as radiation, so some black holes would retain their contents and might still exist nearby, perhaps at the centre of our galaxy where the gravitational field is strong. Their small size, however, suggests they would be hard to spot; Banks and Fischler have not yet worked out whether it is likely that they can be found.

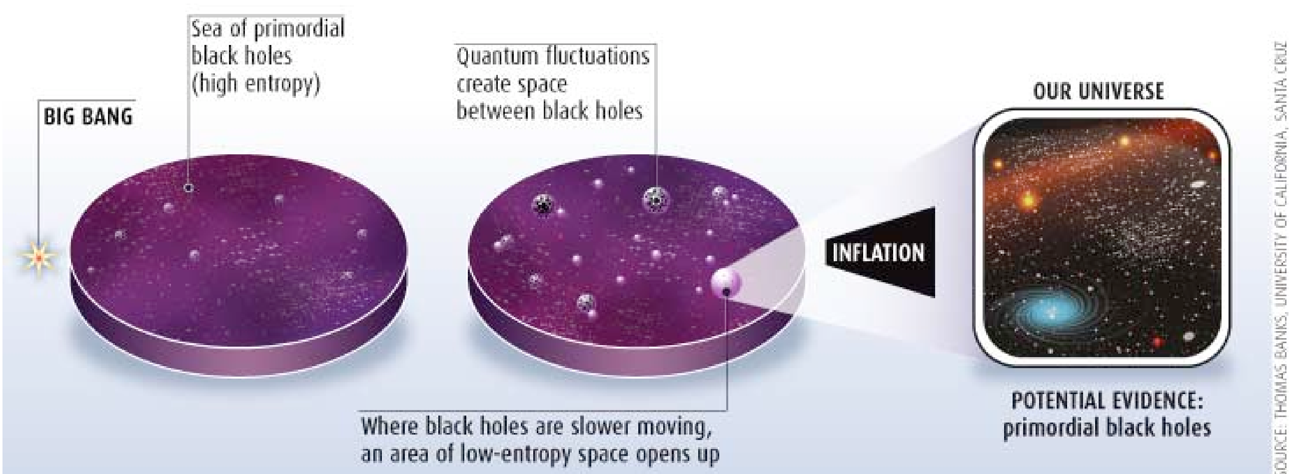
There may be another way to distinguish between the models. In the standard big bang picture, gravitational waves are generated during inflation from collisions of clumps of matter. Some of these waves might be observed, either by future gravitational wave detectors such as the ESA and NASA-sponsored Laser Interferometer Space Antenna, planned to launch in 2015, or by the imprint they would leave on the cosmic microwave background. In the colliding-brane model, however, inflation never happens, which means primordial gravitational waves would not be produced. Observing them would rule out that model, while leaving viable the black hole fluid and fragmenting universe scenarios.

The most likely outcome, however, is that none of the models will be proved correct any time soon. Indeed, the quest to understand the origin of the universe seems destined to continue until we can answer a deeper question: why is there anything at all instead of nothing?

There is another way to think about why our universe began in a highly ordered or "low entropy" state. In 2002, a group of physicists led by Leonard Susskind at Stanford University in California proposed that entities capable of observing the universe could arise via random thermal fluctuations, as opposed to the big bang, galaxy formation and evolution. This idea has been explored by others, including Don Page at the University of Alberta in Edmonton, Canada. Some researchers argue that under certain conditions, self-aware entities in the form of disembodied spikes in space-time - "Boltzmann brains" - are more likely to emerge than complex life forms. Because they depend on fluctuations of particles, Boltzmann brains would be more common in regions of high entropy than low entropy. If the universe had started out in a state of high entropy, it would be more likely to be populated by Boltzmann brains than life forms like us, which suggests that the entropy of our early universe had to be low. As a low-entropy initial state is unlikely, though, this also implies that there are a huge number of other universes out there that are unsuitable for us.

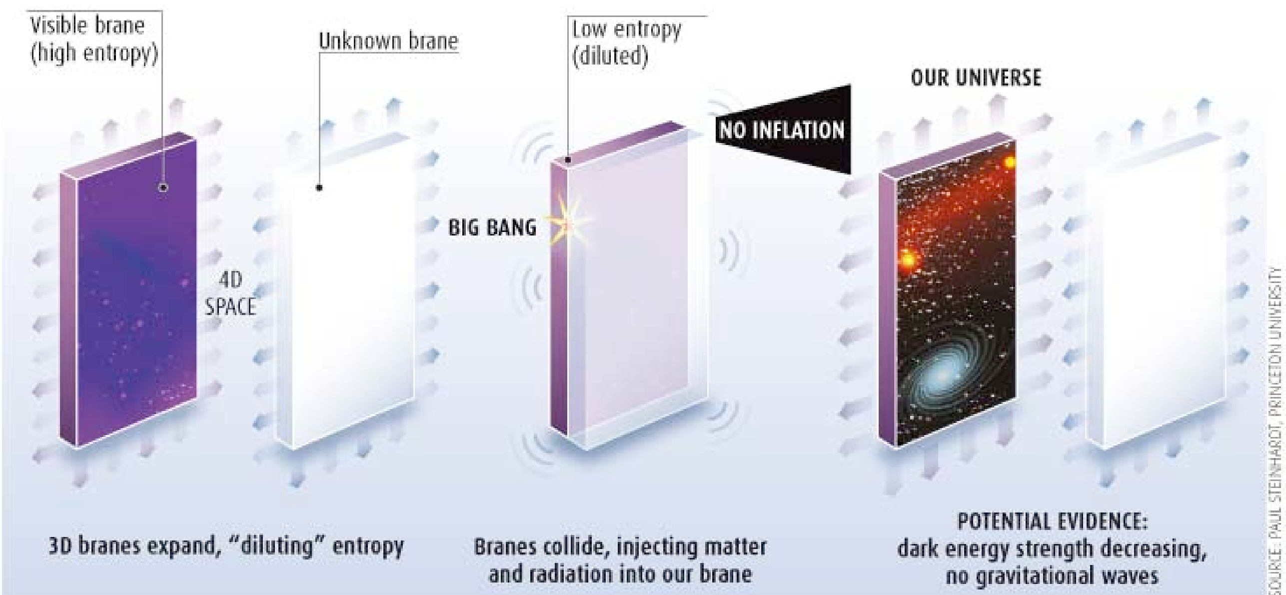
BLACK HOLE SEA

One theory says that the universe began as a black hole fluid, with maximum entropy or disorder. Our observable universe then emerged as a low-entropy bubble of ordinary space that underwent the conventional model of inflation



COLLIDING BRANES

Another scenario uses string theory to propose that our visible universe exists on a 3D membrane, or brane, floating in 4D space. Another brane collides with ours every trillion years, giving rise to what looks like a big bang and dark energy. In this model inflation does not occur



FRAGMENTING UNIVERSE

Was our universe originally ripped from a larger entity? Perhaps, as dark energy increased, it fragmented a mother universe into numerous baby universes, one of which underwent inflation to become ours. In turn, our universe is expanding and may give rise to babies of its own