Why Do Galaxies Exist?

Theorists can answer that question in remarkable detail if they postulate a universe filled with "cold" dark matter

There has always been a certain aura of mystery about galaxies. Why do they even exist? They seem to have formed in abundance only a few billion years after the big bang—quite fast, as these things go—yet there is no obvious reason why they should have formed at all. The observational astronomers have had little to say about the process, because newborn galaxies lie far beyond the limits of current generation telescopes. And the theorists can only guess.

Nonetheless, the problem of galaxy formation is entering a period of general ferment. With new instrumentation such as the ultrasensitive charged-coupled device detectors, astronomers are extending their observations far enough back in time to see some evolutionary effects in galaxies. And meanwhile, the theorists are producing models that are more realistic, more concrete, more detailed—and thereby more testable.

During the past year or so, in fact, one remarkable model has emerged that seems to explain most of the gross features of galaxies in terms of a single mechanism: the dynamics of "cold dark matter." As developed by George R. Blumenthal and Sandra M. Faber of Lick Observatory, Joel R. Primack of the University of California, Santa Cruz, and Martin J. Rees of Cambridge University (*1*), it is probably the most compelling picture of galaxy formation yet.

"It's unlikely to be the last word," says Jeremiah Ostriker of Princeton University, himself a prominent researcher in the field of galaxy formation and not a man to embrace new theoretical fashions lightly. "But it's certainly more promising than many of the other models. It's the first attempt I know where observers and theorists have tried to put everything together. And they are to be commended for that."

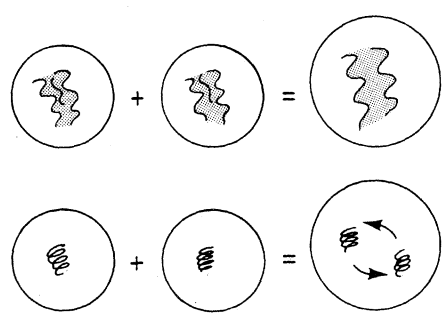
As its name suggests, the cold dark matter picture starts from the observation that in every galaxy and in every cluster of galaxies, roughly 90 percent of the mass is completely invisible. The argument is that spiral galaxies rotate too fast, and cluster galaxies likewise move too fast; the only way these structures could be gravitationally bound is if they were embedded like the organelles of a cell in some massive, invisible halo.

With 90 percent of the mass, this dark matter in the halo was clearly a dominant factor in galaxy formation; the luminous matter that looks so impressive to human eyes is little more than a trace element by comparison. But that just shifts the mystery from one place to another: What is the dark matter?

Astronomers have been debating that question for more than a decade. Blumenthal, Primack, and their colleagues arrive at their hypothesis of "cold" dark matter essentially by a process of elimination: there's not much else it could be.

The obvious assumption is that the dark matter is made of ordinary, "baryonic" stuff—that is, protons and neutrons. The problem is that it is very hard to make baryonic matter invisible. Gas or dust would screen out light from the central galaxy, and anything larger than 0.08 solar masses would ignite and become visible as a star. So baryonic dark matter would have to come in chunks no smaller than rocks and no bigger than planets. But no one has found a plausible reason why it should be restricted to just that size range.

Besides, the baryons came out of the big bang very homogeneously. Witness the fact that the 2.7 K microwave background radiation, emitted from the cosmic plasma some 100,000 years after the big bang, is uniform to a few parts in 10^4. So it is very hard to see how baryonic matter could have clumped into galaxies as quickly as it did without something else to help it along.



To cool or to coalesce?

In the cold dark matter model, low-mass clumps of dark matter form early in the life of the universe; they therefore tend to merge quickly, before the baryonic gas trapped inside them can cool (top). The result is just a larger clump of dark matter and a larger cloud of hot gas. But larger clumps form later and collisions take longer. The baryonic gas may then have time to cool and condense; if subsequent collisions produce anything, it will be a protocluster instead of one huge galaxy (bottom). Estimates of cooling rates and collision rates give galaxy sizes in the range of 10^8 to 10^{12} solar masses, about what is observed.

Thus, pure baryonic dark matter is considered highly unlikely. The presumption these days is that the Something Else is *non*-baryonic dark matter, some form of massive, but weakly interacting, particle that came swarming out of the big bang along with the photons.

This idea first gained strength about 1980, when it appeared that certain grand unified field theories predicted a small mass for the neutrino—and better still, when two independent experiments seemed to find evidence for such a mass (*2, 3*). Massive neutrinos were taken up and championed most vigorously by the Soviet physicist Ya. B. Zeldovich and his colleagues. Neutrinos would certainly be invisible, they pointed out, since the particles hardly interact with anything. Moreover, neutrinos are certainly abundant, since they must have been produced by nuclear reactions during the big bang. So even a tiny mass would allow them to dominate the universe gravitationally.

In fact, calculations by Zeldovich and his colleagues suggested that gravitational clumping of the massive neutrinos would have created "traps" for the baryonic matter with a natural mass scale of 10^{15} solar masses—which happens to be the typical mass of a supercluster of galaxies. Thus, the neutrinos offered a way around the constraints imposed by the 2.7 K background: the baryons could have started out as homogeneous soup, yet they could have still clumped up rapidly as they fell into the neutrino traps. Moreover, calculations showed that eventually, because of the random distribution of the clumps, these protosuperclusters would have evolved into strings of galaxies separated by great voids, which is more or less the large scale structure observed today (*4*).

The calculations of the Zeldovich group were confirmed by a number of independent groups at about the same time and, for a few years, massive neutrinos were very fashionable indeed. However, the predictions that looked so right in broad outline soon began to look very wrong in detail.

The most glaring problem is that the massive neutrino picture has the superclusters forming first, as 10^{15} solar-mass gas clouds. It was only afterward that the clouds gave rise to individual galaxies. But from spectroscopic evidence the galaxies are known to be quite old; if the age of the universe today is taken to be 18 billion years, then the galaxies must have formed only 1 or 2 billion years after the big bang. Yet numerical simulations of the massive neutrino model show that the supercluster collapse would not have occurred until quite late, roughly 3.5 billion years after the big bang. (More precisely, collapse is predicted at redshift $z < 2$, while galaxy formation is known to take place at redshift $z > 3$.)

This is embarrassing enough. But the massive neutrino model also has problems explaining the dark halos around individual galaxies. The particles would be much more prone to collect in the big clusters, meaning that observations should show much more dark matter there. In fact, observations show that the ratio of dark mass to luminous mass (that is, baryonic matter) is remarkably constant at about 10, from the tiny dwarf spheroidal galaxies through the biggest superclusters.

The upshot of all this is that massive neutrinos have lost a lot of their luster. They may be up there—although the 1980 mass detections are still controversial—but neutrinos probably had little to do with galaxy formation.

The problems with massive neutrinos can be traced to the fact that they come out of the big bang fireball still moving at relativistic speeds. In a word, they were "hot." This is what sets the scale of those 10^{15} solar mass traps: fluctuations on a smaller scale are simply washed out. Forming galaxies early requires that the dark matter has to form smaller, galaxy-sized traps, and that means starting with a particle that comes out of the big bang more slowly: "warm" dark matter, or even "cold" dark matter.

These alternative forms of dark matter were first studied by Princeton's P. James E. Peebles in 1982, and shortly thereafter by Blumenthal and Primack in 1983. It turned out that nothing much mattered about the cold dark matter particle, except that its nongravitational interactions had to be much feebler than the weak interactions experienced by the neutrino, and that it had to have some appropriate, nonzero mass. Everything else followed from the thermodynamics of the big bang. Specifically, these very weakly interacting particles would have decoupled from the cosmic plasma sooner than the neutrinos, they would have cooled and slowed down sooner, and they would have begun forming gravitational traps sooner.

Furthermore, if the big bang started out with primordial density fluctuations that were distributed independently of scale—as predicted in the popular "inflation" models of the early universe (*5*)—then the first clumps formed by these new particles would indeed be much less massive than the first clumps formed by neutrinos.

"Warm" dark matter, in this scheme, is the kind of stuff that has just the right properties to produce galaxy-sized perturbations. Unfortunately, as Blumenthal, Primack, and other workers quickly realized, not even the fertile minds of the particle theorists had come up with a good candidate particle.

However, for cold dark matter, the kind of stuff that would produce subgalactic perturbations, the particle physicists had provided numerous candidates: axions, which arise in theories that try to suppress any large CP violations in the strong interactions; photinos, the spin-½ partner of the photon in theories of supersymmetry; primordial black holes produced in the big bang; and even the massive "nuggets" of quark matter proposed by Princeton's Edward Witten.

Admittedly, there is no direct evidence for any of these objects—the neutrino, at least, is known to exist—but the state of particle physics is such that they all have to be taken seriously (*6*). Thus, Blumenthal and his colleagues, as well as several other groups, began to take a close look at how well cold dark matter would work as a mechanism for galaxy formation. It turned out to work very well.

In the picture outlined by Blumenthal, Faber, Primack, and Rees, the galaxy formation process begins shortly after the emission of the microwave radiation at 100,000 years, as baryonic matter begins to fall in with the cold dark matter fluctuations. In effect, the universe is littered with gravitational potholes, and each pothole is filling up with hot gas. (Actually, the smallest clumps tend to form first. The minimum size for the baryon clouds is the so-called Jeans mass, which under these conditions is a few hundred thousand solar masses; the gas is so hot that anything smaller simply evaporates again.) There follows a competition. On the one hand, the hot baryonic gas in each pothole is trying to cool off, radiate its energy, and contract into a dense core at the center. On the other hand, these are dynamic potholes, and they are simultaneously trying to move around and merge into larger structures. The outcome depends on the time scales: low-mass clumps of dark matter form early and tend to merge quickly, while their baryonic gas clouds are still energetic and diffuse. The result in each case is simply a bigger halo containing a bigger cloud. But the high-mass clumps form later, so that collisions are less frequent. The baryonic gas thus may have time to cool into compact protogalaxies, which would tend to swing right past one another without merging; at most, a collision might produce two protogalaxies orbiting separately—a kind of protocluster.

The Virgo cluster

A preference for elliptical galaxies over spirals—and an abundance of invisible dark matter.

Calculating the time scales for these two processes is straightforward. The result, according to Blumenthal and his colleagues, is that the protogalaxies only have time to cool and form dense cores if their masses lie in the range of 10^8 to 10^{12} solar masses—"[which] encompasses virtually all the mass that is observed to comprise galaxies," they write. Larger mass systems, which are the last to form in this picture, would naturally consist of swarms of galaxies. (The gas clouds would have cooled and contracted long before.) At the other extreme, the tiny dwarf spheroidal galaxies are only marginally able to hold themselves together and only marginally able to cool.

Thus, the cold dark matter model explains the existence of galaxies in just the mass range observed. This is already striking enough. But equally striking is that when the dark matter dynamics is folded in with the statistics of random density fluctuations, there arises a relation between the mass (or luminosity) of a galaxy and the internal velocities of the stars that comprise it: mass is proportional to the fourth power of the velocity. As it happens, this relation has already been established empirically. Known variously as the Tully-Fisher or Faber-Jackson law, it is perhaps the most reliable indicator of a galaxy's true luminosity. (The average orbital velocity of the stars can be determined directly through spectroscopy.)

The cold dark matter model has a number of other suggestive features:

- Within the hot protogalactic gas, which would have a temperature of some 10^5 to 10^6 K, regions of "cold" gas would form at a temperature of roughly 10,000 K. (This is the temperature at which hydrogen and helium become ionized by collisions, and is well known as one of the stable states of the interstellar medium.) The cold gas would then tend to form gravitationally bound structures of roughly 10^5 to 10^6 solar masses—about the size observed for globular clusters, which indeed seem to have some of the oldest stars in the universe.
- Within admittedly large uncertainties, the spiral galaxies seem to have arisen from 1σ fluctuations of the primordial mass distribution, and ellipticals from 2σ fluctuations. This agrees with the fact that more than half of all galaxies are spirals, while only 15 percent are ellipticals.
- Blumenthal, Faber, Primack, and Rees argue that high σ fluctuations will tend to develop smaller angular momentum: "You expect them to be *rounder*," explains Blumenthal. These fluctuations will thus be able to collapse quickly to form the big, featureless elliptical galaxies—which seem to have very little angular momentum—and the large central bulges of the spiral galaxies. (The spiral bulges seem to have all the characteristics of small elliptical galaxies, except that they happen to be embedded in a dusty disk.) Moreover, for statistical reasons, high σ fluctuations will tend to be strongly correlated; therefore, one would expect more ellipticals to be found in rich clusters—as is observed.

Also, there may be an environmental effect: the low σ fluctuations, which presumably yield the disk galaxies, tend to be relatively large and diffuse; such galaxies would thus form slowly, with the baryons falling in over large distances. But in rich clusters, the large halos would tend to collide and merge before the galaxy finishes forming. So one expects that in rich clusters, much of the gas that might have gone into disk galaxies instead gets scattered through the cluster as a diffuse haze—exactly what is seen in x-ray observations of rich clusters.

- Recent calculations by Blumenthal, Faber, Primack, and Brandeis University's Riccardo Flores suggest that as a galaxy condenses into the middle of a halo it significantly modifies the distribution of dark matter in the halo. This distribution can be probed observationally by measuring the orbital velocities of stars in a given galaxy as a function of distance from the center. The theory predicts that the orbital velocity in a spiral galaxy will be roughly independent of radius—exactly as observed.

Cold dark matter explains the existence of galaxies in just the mass range observed.

- N-body simulations suggest that cold dark matter can produce a large-scale structure of superclusters and voids that is similar to observed structure.

In addition to these features, the cold dark matter picture also seems to mesh well with another popular idea known as "biased" galaxy formation. First developed last year by James Bardeen of the University of Washington and Nicholas Kaiser of the University of California, Berkeley, bias attempts to reconcile a key prediction of the inflation model with a recalcitrant fact of observation: the average density of the universe is predicted to be very nearly equal to the critical density required for closure ($\Omega=1$ in the jargon), whereas observations of the galaxies and clusters find less than 20 percent of that amount ($\Omega<0.2$), even counting the dark halos.

Suppose, however, that for some reason galaxies only formed at the extremes of the density fluctuations. In other words, says Bardeen, "The first galaxies somehow poisoned the universe for the rest." From purely statistical arguments, high σ fluctuations would have been more strongly correlated than the matter distributions as a whole. Thus, one would naturally expect to see galaxies forming the kind of clumpy large-scale structure that is actually observed. But in between the clusters one would expect to find enough dark matter and uncondensed baryonic gas to boost Ω up to 1.

There is some intriguing observational evidence for unformed matter out in intergalactic space. Wallace Sargent and his colleagues at the California Institute of Technology have observed dense clusters of absorption lines in the spectra of quasars, which they interpret as Lyman-α absorption by clouds of hydrogen gas along the line of sight. These clouds contain very little mass themselves. But if they are embedded in much larger expanses of ionized hydrogen, as seems plausible, then the Bardeen-Kaiser picture does seem consistent.

On the other hand, no one has yet come up with a truly compelling mechanism to bias galaxy formation sufficiently. Perhaps the first galaxies to form sent out ultraviolet radiation, or supernova shock waves, thereby heating the surrounding gas and making it impossible for more galaxies to form. But then, as Princeton's Ostriker asks, how is it that galaxies were able to form in rich clusters? Wouldn't a "poisoning" effect have caused the galaxies to be *less* clustered, not more? Clearly, more work needs to be done.

Given the paucity of observational data on galaxy formation, there is inevitably an element of fashion when it comes to theorizing about the subject. It is important to remember that the cold dark matter picture is only the latest of many such fashions, and is almost certainly not the last word. For example, there is widespread albeit indirect evidence that many galaxies harbor million-solar-mass black holes in their cores. Indeed, quasars are thought to be otherwise ordinary galaxies that have *billion*-solar-mass black holes. Yet the cold dark matter picture says nothing about when or how such black holes might form.

Nonetheless, it is by far the best model currently available. It offers a rich array of testable predictions. And of course, it underscores once again the cosmologists' challenge to the particle physicists: find the particle or particles responsible for the dark matter.

—M. Mitchell Waldrop

References and Notes

1. G. R. Blumenthal, S. M. Faber, J. R. Primack, M. J. Rees, *Nature (London)* **311**, 517 (1984).
2. A. L. Robinson, *Science* **208**, 697 (1980).
3. M. M. Waldrop, *ibid.* **211**, 470 (1981).
4. ——, *ibid.* **219**, 1050 (1983).
5. ——, *ibid.*, p. 375.
6. Certain anomalous events recently detected at the CERN proton-antiproton collider may be consistent with photino production [——, *ibid.* **227**, 505 (1985)].