

GALAXY FORMATION AND DARK MATTER

Martin Rees
Institute of Astronomy
Madingley Road
Cambridge, CB3 0HA.

INTRODUCTION

My aim in this general talk will be to air some questions, rather than offer firm answers, because the most basic questions about galaxies are indeed still unresolved. In particular:

1. We do not know why such things as galaxies should exist at all — why these assemblages of stars and gas with fairly standardised properties are the most conspicuous large-scale features of the cosmos.

2. About 90% of the mass associated with galaxies is hidden. The luminous stars and gas contribute only about a tenth of the gravitating material inferred from dynamical arguments. What the rest consists of is still a mystery.

3. It is unclear why the nuclei of some galaxies flare up, and release the colossal amount of non-stellar radiation emitted from quasars and radio galaxies, as discussed by Maarten Schmidt in his discourse.

We are perplexed about these issues, just as 50 years ago our predecessors were perplexed about the nature of stars. But some of us are hopeful that the physical processes underlying galaxies are coming into focus, and can at least be seriously addressed. I must apologise in advance to specialists on this topic for the 'broad brush' and inevitably distorted exposition I shall be giving.

In their already-classic book on galactic dynamics, Binney and Tremaine (1987) make the point that galaxies are to astronomy what ecosystems are to biology. They are not only dynamical units, but chemical units as well. The atoms we are made of come from all over our Milky Way galaxy, but few come from other galaxies. The ecological analogy reflects other features of galaxies: their complexity, ongoing evolution, and relative isolation.

Single stars, the individual organisms in the galactic ecosystem, can be traced from their birth in gas clouds through their lifecycle. And we have come to understand why *stars* exist with the general properties we see. The question why *galaxies* exist is less straightforward than the equivalent question for stars. Galaxies formed at an earlier and remote cosmic epoch. We don't know how much *can* be explained in terms of ordinary processes accessible to study now, and how much has its causes in the earliest universe.

There is an elaborate taxonomy for galaxies, but the most obvious categories are disks and spheroids or ellipticals. There is a well-known cartoon model, dating back about 30 years, to account for this basic morphological distinction. Suppose

that a galaxy started life as an irregularly-shaped gas cloud contracting under gravity, and that the collapse of such a gas cloud were highly dissipative, in the sense that any two globules of gas that collided would radiate their relative kinetic energy and merge (Figure 1). The end result of the collapse of such a cloud would be a rotating disk. This is the lowest energy state that the cloud can reach if it does not lose or redistribute its angular momentum. On the other hand, stars do not collide with each other, and are unable to dissipate energy in the same fashion as gas clouds. So the *rate of conversion of gas into stars* could be the crucial feature determining the type of galaxy that results. Elliptical galaxies would be those in which the conversion is fast, so that most stars have already formed before the gas has had time to settle down in a disk. Disk galaxies result when the star formation is delayed until the gas has already settled into a disk. According to this traditional picture, disk galaxies are those with slower metabolism, which have not yet got so close to the final state in which essentially all the gas is tied up in low mass stars or dead remnants.

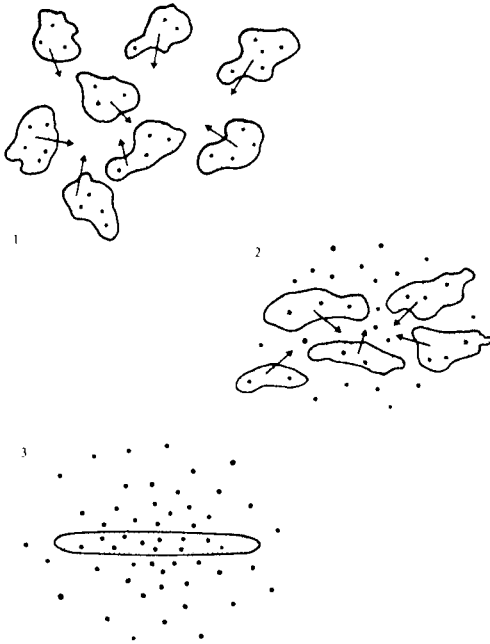


Figure 1. ‘Cartoon’ showing three stages in the traditional picture of protogalactic collapse

WHAT IS SPECIAL ABOUT GALACTIC DIMENSIONS?

This story has many inadequacies, and I’ll return to some of them later. In particular, there is no scale in the picture. Is there any physics that singles out clouds of galactic dimensions, just as, since Eddington and Chandrasekhar, we have known the natural scale of stars? All we have for galaxies is a simple but suggestive physical argument. Two timescales are important in determining how a self-gravitating

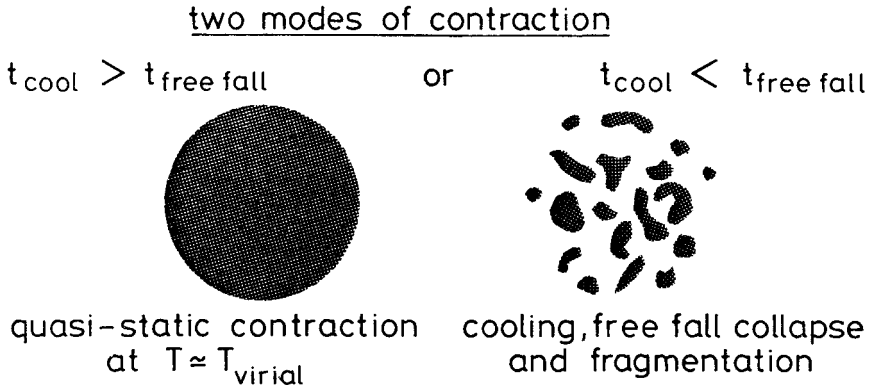


Figure 2. *Cooling and contraction of self-gravitating gas clouds.*

gas cloud evolves. The first of these is the dynamical or freefall time, which is of order $(G\rho)^{-\frac{1}{2}}$, its precise value depending on the geometry of the collapse. The second is the radiative cooling timescale. This depends on the gas temperature T_g , and can be written $T_g/\rho\Lambda(T_g)$ where Λ can be calculated from atomic physics.

If t_{cool} exceeds $t_{\text{dynamical}}$, a cloud of mass M and radius r can be in quasi-static equilibrium, with the gas at the virial temperature. But if $t_{\text{cool}} < t_{\text{dynamical}}$ such equilibrium is impossible (Figure 2). The cloud cools below the virial temperature and undergoes freefall collapse or fragmentation. We would expect clouds to collapse and fragment in the fashion depicted in Figure 1 only if they enter the part of $M - r$ plane where cooling is faster than freefall. A simple calculation shows that this criterion involves a characteristic mass-independent radius of order 75 kpc and a characteristic mass M_{crit} of order $10^{12} M_{\odot}$. Clouds less massive than M_{crit} will readily fragment, but above M_{crit} fragmentation is impossible unless the cloud contracts until its radius is below r_{crit} . This characteristic mass and radius, consequences of straightforward physics (Figure 3 *overleaf*), feature in many cosmogonic schemes as at least setting an upper limit to the scale of galaxies.

Eddington claimed that a physicist on a cloud-bound planet could have predicted the properties of the gravitationally-bound fusion reactors that we call stars. But these simple considerations don't suffice to predict galaxies, even with hindsight. This is because any true explanation of galaxies must involve setting them in a cosmological context.

THE COSMOLOGICAL CONTEXT

In a memorable invited discourse at the Patras IAU General Assembly, Zel'dovich (1982) discussed the hot big bang model, which he opined was as sure as that the Earth goes round the Sun. We may not all quite share his exuberant certitude. But most of us regard the hot big bang as the 'best buy' cosmology, more than 50% likely to be essentially correct. According to this picture everything emerged from a universal thermal soup which was initially smooth, and almost featureless,

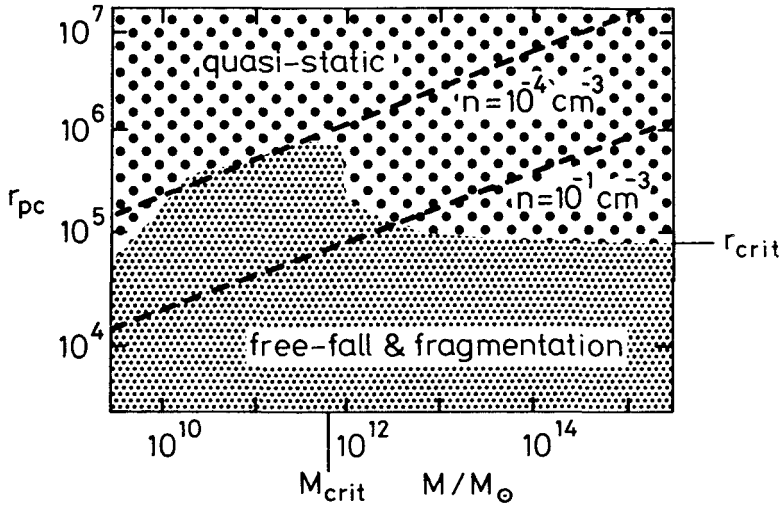


Figure 3. The quasi-static and free-fall regions (cf. Fig. 2) are here presented in a mass-radius plot. M_{crit} and r_{crit} should set characteristic upper limits to the dimensions of galaxies. Clouds with $M \gg M_{\text{crit}}$ would be quasi-static unless at very high densities (cf. gas in clusters of galaxies).

but not quite. There were, we don't really know why, small fluctuations from place to place in the expansion rate. Structures emerged via gravitational instability as over-dense regions lagged more and more behind the universal expansion, and eventually condensed out as embryo galaxies and clusters.

Theorists trace back the history of the hot big bang over 60 decades of logarithmic time. The events and stages in the cosmic expansion are summarised in Figure 4, which goes back to the earliest era, the intellectual habitat of the 'gee whiz' fringe of particle physicists. For our present purposes the uncertain details are irrelevant. It may, though, be conceptually useful to divide cosmic history into 3 parts. For the first 40 decades the microphysics is uncertain. When the universe cools below 10 MeV and the density falls below nuclear density, the microphysics become straightforward. Initial irregularities, owing their origin to the first era, amplify via gravitational instability, and things become less straightforward when the first of these condense out. Then we confront a set of new difficulties. The physics is just Newtonian gravity and gas dynamics, the easy bits of Landau and Lifshitz, but the complications are those of non-linearity. The 'recent' universe is hard to understand for the same reason that weather prediction is difficult.

The types of difficulty that one faces in studying the early and late phases were amusingly contrasted in a recent article by the distinguished Canadian relativist Werner Israel. The early universe was like the challenge of chess, he said, but the later non-linear stages were like mud wrestling. Maybe, but it's glorious mud and some of us have wallowed contentedly in it for years.

A key question is how much can be explained by processes occurring at the

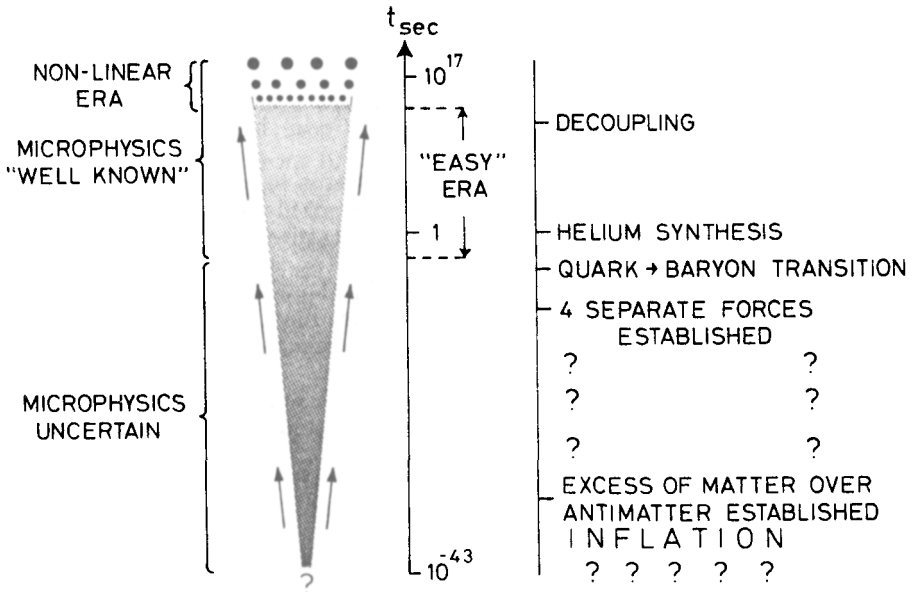


Figure 4. Stages in the evolution of the 'standard' big bang model universe

range of epochs accessible to astronomers, and how much has to be attributed to the 'chess playing' era.

The main types of relevant data are morphological classifications (dating back to Hubble); correlations between luminosity, velocity dispersion and size; and the statistics of galaxy clustering. Any quantitatively satisfactory theory must explain these things. We have no generally agreed theory yet. Indeed, as Saslaw has put it, 'if galaxies didn't exist, we'd have no problems explaining the fact'. Moreover, the seekers for any such theory must first face a most embarrassing circumstance: this is the *dark matter problem*, evidence that 90% of the mass of galaxies is unaccounted for, and takes some unknown form.

DARK MATTER

The evidence for dark matter dates back more than 50 years, but has firmed up since the classic papers of Einasto, Cowsik and Saar and Ostriker, Peebles and Yahil, both published in 1974. Vera Rubin (1985) reviewed this (now compelling) evidence in her discourse at the last IAU.

The masses inferred from relative motions of galaxies in apparently bound groups and clusters exceed by a factor 10 those inferred from the internal dynamics of the luminous parts of galaxies. This apparent discrepancy could be resolved if galaxies were embedded in extensive dark haloes. The halo hypothesis can be checked in some edge-on disk galaxies, where emission from gas can be observed out at radii far exceeding the extent of the conspicuous stellar disk. The mass of

this gas is itself negligible, but rotation velocities derived from its spectral lines do not fall off as $R^{-\frac{1}{2}}$, as would be expected if the gas were orbiting a mass distribution concentrated at much smaller radii. Instead the velocity remains almost constant, implying that the mass within radius R is proportional to R out to 80 kpc in some cases. Direct lower limits on the mass-to-light ratio in the outlying parts of some galaxies exceed 300 solar units.

In some elliptical galaxies also, the mass seems to increase proportional to R out to large R . In M87, such evidence comes from globular cluster orbits, and, still further out, is inferred from the X-ray temperature and profile of the diffuse gas. On a larger scale, we have evidence from clusters of galaxies, along the lines first discussed by Zwicky and Sinclair Smith in the 1930s. Many independent lines of evidence point towards the existence of dark matter (these are summarised in Figure 5). This has as good a claim to be termed a paradigm shift as any development one can think of in modern astronomy.

The dynamically-inferred dark matter, though ten times the luminous matter, still amounts to only 10 or 20 per cent of what is required for a closed universe: the corresponding value of the density parameter Ω , the ratio of the actual density to the cosmological critical density, is 0.1 or 0.2. I shall return later to the question of whether there could be enough dark matter to make $\Omega = 1$.

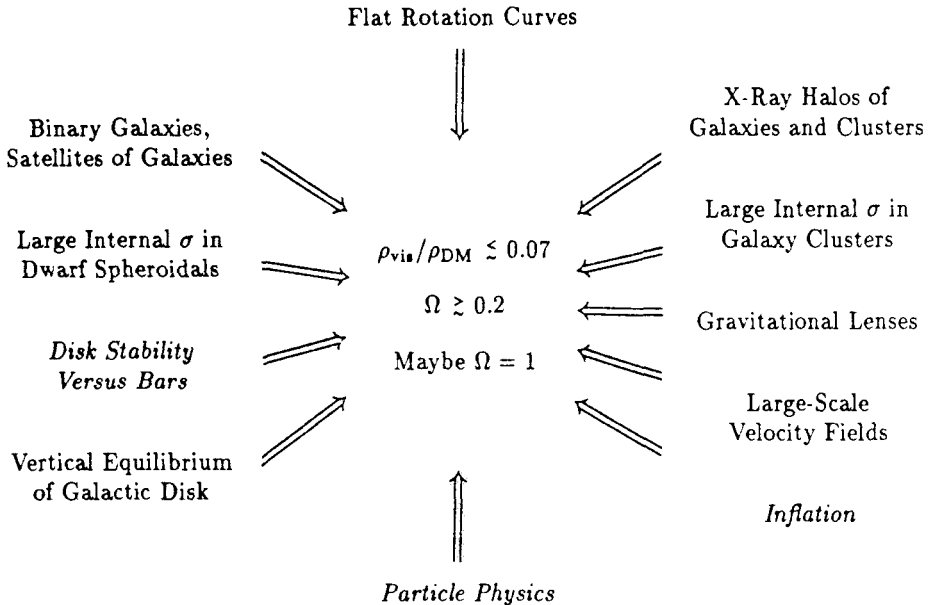


Figure 5. *The various lines of evidence for dark matter (diagram due to J. Kormendy)*

THE NATURE OF THE DARK MATTER: BARYONIC OR NOT?

What is the halo dark matter? The first possibility that comes to mind is faint stars or stellar remnants. Figure 6, due to Carr, Bond and Arnett (1984), quantifies the maximum hidden contribution to Ω that could be made by stars or their remnants in the mass range from 10^{-2} to $10^8 M_{\odot}$. There are two tenable dark matter candidates: very low mass stars, below $0.1 M_{\odot}$, or the remnants of very massive stars. Ordinary stars above $0.1 M_{\odot}$ would contribute too much background light unless they had all evolved and died, leaving dark remnants. But the remnants of ordinary massive stars of $10 - 100 M_{\odot}$ would produce too much material in the form of heavy elements. Stars with core masses above $200 M_{\odot}$ end their lives, via the pair production instability, by collapsing rather than exploding. These very massive objects, (VMOs for short), do not eject heavy elements, and leave black hole remnants. Such objects, if they constitute our own galactic halo, can't however exceed $10^6 M_{\odot}$ each, because otherwise dynamical friction, whereby a hole transfers energy to lighter stars close to its path, would have led to excessive thickening of the galactic disk.

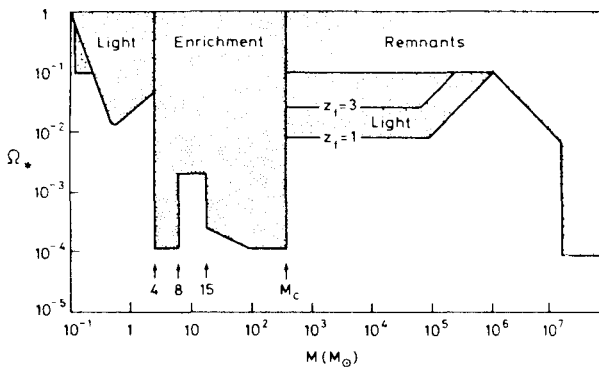


Figure 6. Constraints on the fraction of the critical density that could be present in stars or stellar remnants of various masses. The stars are presumed to have formed at some redshift z_f . Possible candidates for the dark matter are low mass stars (brown dwarfs or 'Jupiters') or very massive objects (VMOs).

Is it likely or unlikely that a forming galaxy should convert most of its mass into either ultra-low mass stars or objects heavier than a few hundred suns? We don't understand enough about star formation, even close at hand in for instance the Orion nebula, to be confident in saying how the initial mass function might be affected by intense background radiation, absence of heavy elements, lack of magnetic fields, and the rest. Theory therefore cannot arbitrate reliably between low mass and VMO options (Figure 7 *overleaf*).

Can we learn from observations about what the dark matter is? Low mass objects would be perhaps detectable in the infrared: the nearest would be less than a parsec away, with high proper motions. There are two handles on VMO remnants. They might reveal their presence by accretion on passage through interstellar clouds. Also, they imply that galaxies would be bright when young -

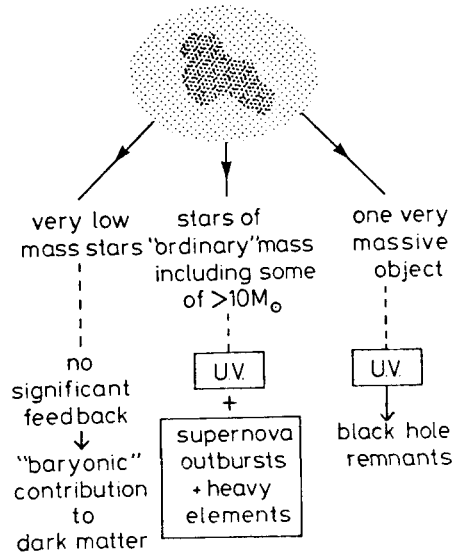


Figure 7. In many cosmogonic schemes, star formation would be initiated in baryonic clouds of around $10^6 M_{\odot}$, but we have no firm theoretical basis for deciding the characteristic mass, or the IMF, of these first stars.

there are constraints from the sky brightness, and from the faint galaxy counts, but the quantitative interpretation of these limits depends on the uncertain redshift of galaxy formation.

One way of detecting compact dark objects, and discriminating between the Jupiter and VMO options, is by searching for evidence of gravitational lensing. The probability of seeing lensing due to an object in our own halo is only about 10^{-6} . But the cross-section for effective lensing is proportional to distance, so there is, perhaps surprisingly, much more chance of detecting objects in the haloes of galaxies half way out to the Hubble radius (Figure 8). The probability that a compact source at redshift $z > 1$ is significantly microlensed by objects along the line of sight is of order Ω , independent of the individual lens mass involved (Refsdal 1970, Press and Gunn 1973). The *angular separation* of the images, proportional to $(\text{lens mass})^{1/2}$ is however a diagnostic of the masses. For masses above $10^5 M_{\odot}$, very long baseline radio interferometers provide adequate resolution. We could probably already exclude $\Omega = 1$ in such objects.

For brown dwarfs of below $0.1 M_{\odot}$, the angular scale is less than a micro arcsecond. This cannot be directly resolved by any technique, until optical interferometers are deployed in space. There is nevertheless a genuine prospect of detecting lensing of this kind because of the variability that would ensue if the lens were to move transversely (*e.g.* Gott 1981). An object at the Hubble distance moving at 100 km s^{-1} takes only a few years to traverse a micro arcsecond. The image structure and time variation are more complicated if the line of sight passes through, for example, a galactic halo, thereby encountering an above-average column density of dark matter. Several objects may then contribute to the imaging, yielding a frosted glass effect, whose pattern, though too small to be seen directly, would vary on a timescale of months or years.

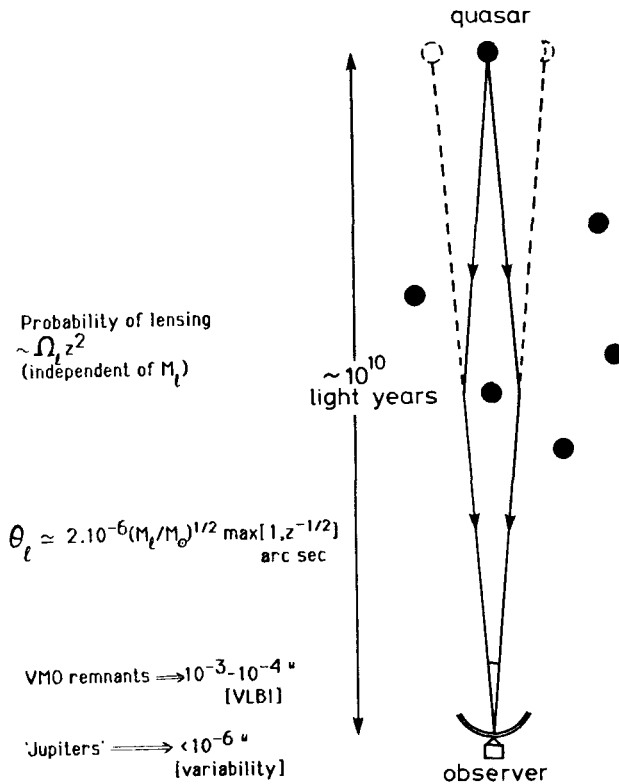


Figure 8. Properties of gravitational microlensing by a population of compact objects along the line of sight to a cosmologically-distant quasar.

Those I've just discussed are the 'dull man's' options for dark matter. The big bang may have left not just baryons and radiation, but other species as well, which may contribute to Ω . In the standard big bang model, neutrinos are almost as abundant as microwave background photons, outnumbering baryons by around 10^9 . Their mass would only need to be a few eV to make them dynamically important. More than 15 years ago Cowsik and McClelland (1973) and Marx and Szalay (1972) conjectured that neutrinos could provide the dark mass in galactic haloes and clusters. At that time the suggestion was not followed up very extensively. But by the 1980s physicists had become more open-minded about non-zero neutrino masses. A change in theoretical attitude, coupled with experimental claims that the electron neutrino had a mass around 36 eV (Lyubimov *et al.* 1980), stimulated astrophysicists to explore scenarios for galaxy formation in which neutrino clustering and diffusion played a key rôle. More recently, other kinds of non-baryonic matter have also been considered.

Provided that we know the mass and annihilation cross-section for any species of elementary particle, we can in principle calculate how many survive from the big bang, and the resultant contribution each species makes to Ω . Progress in

experimental particle physics may therefore reveal a particle which must contribute significantly to Ω , unless we abandon the hot big bang theory entirely.

Neutrinos have the virtue of being known to exist, but particle physicists are inventive, and have come up with a long shopping list of relics that *might* exist. The most theoretically-favoured option is some kind of electrically neutral weakly interacting massive particles, WIMPs for short. These have attractive cosmogenic consequences which I'll come back to in a moment. What is perhaps more remarkable is that such particles may be looked for in the lab.

If our Galactic Halo were composed of WIMPs with individual masses of a few GeV, they would be swarming through this room, with a density of 10^5m^{-3} and speeds of around 300 km s^{-1} . Collision cross-sections are small, but whenever a WIMP collided with an atomic nucleus, the nucleus would recoil with a similar velocity, and an energy around a keV. The collision rates depend on the physical details and the target nucleus, but are in the range 1 - 1000 events per day per kilogram of detector.

These collision events may be detectable by a variety of cryogenic techniques in a low background environment, at quite modest cost. Such experiments are being planned in various countries (see Primack *et al.*, 1988, for a review). Ingenious schemes for detecting a halo background of exotic particles are surely among the most worthwhile and exciting high risk experiments in physics or astronomy today – potentially as important as those that led to the discovery of the microwave background in the 1960s. A null result, with just upper limits, would surprise nobody. On the other hand, such experiments could reveal new particles, as well as determining what 90% of our universe consists of. Because the detection is sensitive to velocity, they would even reveal the halo's velocity dispersion and rotation. The mean velocity of halo particles relative to the detector would change during the year, owing to the Earth's motion round the Sun. The resultant annual modulation, with an amplitude of a few per cent and a peak in June, would be an unambiguous signature discriminating against spurious background.

DARK MATTER AND GALAXY FORMATION

A less direct line of attack on pinning down the dark matter entails exploring the consequences of each option for galaxy formation. If it is dynamically dominant, then non-baryonic matter plays a key rôle in the process whereby small primordial perturbations evolve into protogalaxies and clusters.

The key parameter is the spectrum of density fluctuations, the rms amplitude as a function of mass scale, at the recombination epoch, $z = 1000$. Density contrasts on all relevant scales amplify at the same rate thereafter, so the first bound systems to arise via gravitational instability will have mass scales for which this amplitude peaks. The spectrum depends on what is imprinted initially, possibly modified by preferential damping of smaller scales before recombination.

The left-hand panel in Figure 9 shows the spectrum expected if the universe is dominated by neutrinos with masses 10 or 20 eV. These are moving sufficiently fast that everything is homogenised on scales at least up to $10^{14} M_{\odot}$. The first bound systems would then be superclusters, and galaxies would result from some kind of secondary fragmentation process.

On the right is shown a 'white noise' spectrum, with amplitude larger for smaller scales. Here we have a hierarchical 'bottom-up' cosmogony, with the emergence first of subgalactic scales, then galaxies, and then clusters. (There may then

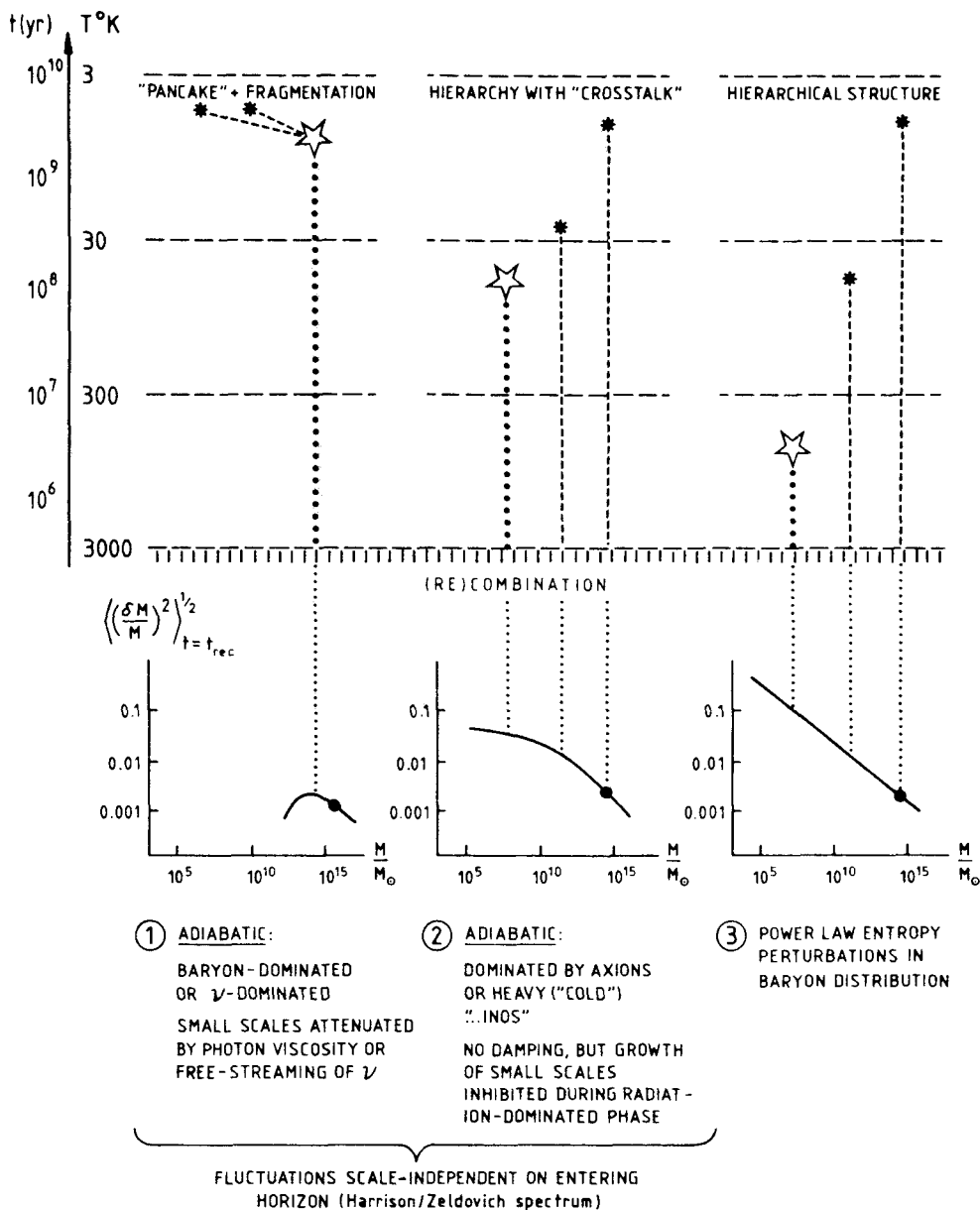


Figure 9. Cosmogenic scenarios corresponding to three different spectra for the post-recombination density perturbations. See text for further explanation.

be an interesting complication: radiative or explosive output from the first small bound objects could create secondary large-scale inhomogeneities that swamp those already present. We then have neogony as well as palaeogony.)

At recombination, when the universe was 10^6 years old, the microwave background shifted redward of the visible band, and the universe entered a literal dark age. The universe remained a simple place until the first bound systems condensed. We don't know when 'first light' occurred. The dark age may have been brief, as in the right-hand panel of Figure 9, or it could have lasted a billion years if the left-hand panel is closer to the truth. We remain more confused and ignorant about this phase of cosmic history than many seem to be about the first 10^{-35} seconds.

Let us focus now on the middle panel in Figure 9. The fluctuation spectrum here has the shape unambiguously calculable for WIMPs, or for any non-baryonic dark matter that is 'cold', in the sense that the individual particles move too slowly for damping due to free-streaming to occur, as it does for neutrinos. This 'cold dark matter' spectrum is nearly flat for small masses, so the typical fluctuation of $10^6 M_\odot$ would collapse no earlier than the epoch corresponding to $z = 10$. The build-up of structure is hierarchical, in the sense that smaller scales tend to form earlier. However, because of the flat spectrum, there would be complicated 'cross talk' between many different scales. The 3σ peaks in the density distribution on galactic scales, $10^{11} M_\odot$, would have the same amplitude as more typical peaks of mass $10^6 M_\odot$, and would therefore collapse at the same time. It is consequently hard to analyse, either analytically or numerically, even the purely dynamical and non-dissipative aspects of the clustering. However, those studies that have been done are encouraging, in that when the amplitude of the fluctuations is normalised so as to match the data on galaxy clustering, the finer scale disposition of the dark matter closely reproduces the sizes and profiles of individual galactic haloes. An example of how such clustering develops, based on simulations of Frenk *et al.* (1985), is shown in Figure 10. Figure 11 shows the final spatial disposition of the dark matter for a slightly different model.

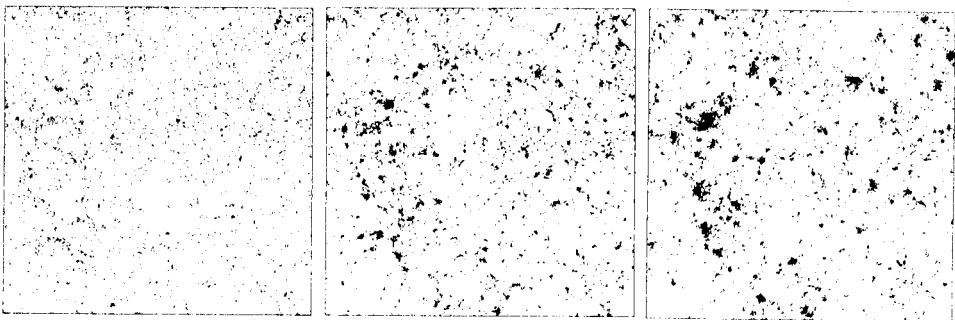


Figure 10. Three stages in the evolution of non-dissipative gravitational clustering within a comoving cubical volume, for an initial spectrum with $((\delta M/M)^2)^{1/2}$ proportional to $M^{-1/3}$. If the right-hand panel is taken to represent the present epoch, then the middle panel is $z = 0.9$ and the left-hand panel $z = 3.5$.

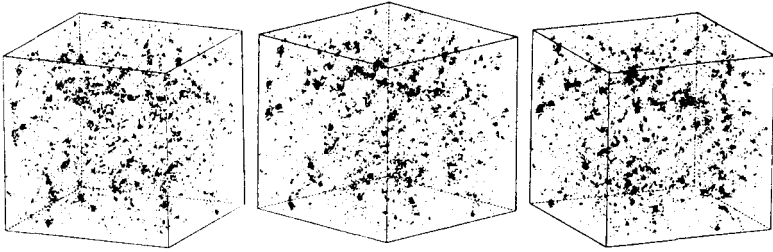


Figure 11. Three views to illustrate the spatial structure within a simulated cubical volume of the expanding universe. In this model the initial fluctuations were Poissonian, with amplitude proportional to $M^{-\frac{1}{2}}$.

Even though the dark matter may be dynamically dominant, it manifests itself only gravitationally. To predict what the universe would actually look like in this model — the luminosity function of galaxies and how they are clustered — we need to develop more understanding of several physical processes. Baryons are presumed to condense in virialised haloes of dark matter in the mass range $10^8 - 10^{12} M_{\odot}$. For larger masses, dissipative cooling may be inefficient for the reason mentioned earlier (*cf.* Figure 3). Below $10^8 M_{\odot}$ the potential wells may be too shallow to capture primordial gas. The mass distribution of isolated virialised systems can in principle be learned from N-body simulations. But even if the dissipationless clustering of the dark matter is accurately known, the fate of the baryonic component, how much gas falls into each potential well and how much is retained, involves complex gas dynamics. We need also to understand how the baryonic component behaves during mergers. If we trace back the history of the large haloes in Figure 10, half have experienced a merger since $z = 2$.

Theoretical fashions are often transient. But the cold dark matter model (Peebles 1982, Bond and Szalay 1983, Blumenthal *et al.* 1984 and references cited therein) has survived for more than 5 years (though there are two possible problems with it that I'll mention later). Insofar as it can account for galaxies and their haloes it offers circumstantial support for the idea that the dark matter is in WIMPs or axions. But this evidence is only circumstantial. The nature of the dark matter is still an open question. I am personally agnostic and would bet 25% on Jupiters, 25% on black holes, 25% on WIMPs or other cold dark matter, leaving the remaining 25% for things not yet thought of.

The late Professor Redman of Cambridge, warmly remembered by senior IAU members as a no-nonsense observer with little taste for speculation, once claimed that any competent astrophysicist can reconcile any theory with any set of facts. An even more cynical colleague extended this claim, asserting that the astrophysicist often need not even be competent. Dark matter theorists are perhaps exemplifying Redman's theorem and its extension. All things considered, the existence of dark matter is unsurprising. There are all too many forms it could take, and the aim of theorists and observers alike must be to narrow down the range of options. What is encouraging is that various lines of observations, experiments, and theoretical modelling should over the next few years do just this.

It would be specially interesting if we could, by astronomical methods, discover some fundamental particle which has been predicted by theorists. If such particles turned out to account for the dark matter, we would however have to view the galaxies, the stars, and ourselves, in a downgraded perspective. Copernicus dethroned the Earth from any central position. Early this century, Shapley and Hubble demoted us from any privileged location in space. But now even baryon chauvinism might have to be abandoned: the protons, neutrons, and electrons of which we and the entire astronomical world are made could be a kind of afterthought in a cosmos where photinos or neutrinos control the overall dynamics. Great galaxies could be just a puddle of sediment in a cloud of invisible matter ten times more massive and extensive.

IS THE UNIVERSE FLAT ($\Omega = 1$)?

Dark matter is relevant to the future of the entire universe, to eschatology. Will the universe eventually recollapse to a big crunch, or will it expand forever? This depends on whether or not the density parameter Ω exceeds unity.

In an influential review published back in 1974, Gott, Gunn, Schramm, and Tinsley summarised the arguments bearing on Ω . They concluded that the dynamical evidence favoured a value 0.1 or 0.2, and noted that if this matter were all baryonic, the lower end of the range was compatible with the value favoured by big bang nucleosynthesis, for a Hubble time of 2×10^{10} years, (a value consistent with the ages of the globular clusters, *etc.*.) Much new evidence has accumulated since 1974, especially on cluster dynamics and element abundances, and some relevant theoretical issues have been refined and elaborated. But if one were to update Gott *et al.*'s discussion, their net conclusion would not change much.

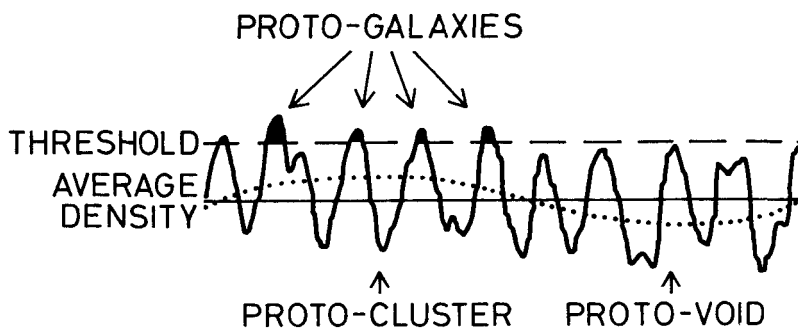


Figure 12. If galaxies form from exceptionally high-amplitude peaks in a gaussian density field, they display enhanced clustering (or biasing) because the probability of a high peak is sensitive to whether or not there is a positive contribution to the amplitude from longer-wavelength modes.

The attitudes of theorists, however, seem to have changed markedly. This is partly because non-baryonic matter is now taken much more seriously, seeming in some ways almost a natural expectation. But the other new element in the discussion is the concept of inflation. This resolves some stubborn paradoxes in

a rather natural way. In particular, it suggests why the expansion rate is so fine tuned that our universe has neither collapsed long ago, nor is expanding too fast for galaxies to have condensed. Inflation indeed suggests that the fine tuning should be so precise that Ω has almost exactly the value unity. If Ω is indeed 1, the balance of argument tilts in favour of non-baryonic dark matter, because standard big bang nucleosynthesis favours a value for Ω_{baryon} of around 0.1.

The fact that direct dynamical evidence suggests a lower Ω implies that, if Ω is actually unity, the galaxies must be more clustered than the overall mass distribution. Voids must not be as empty as they look. The efficiency of bright galaxy formation could be a sensitive function of the density or the depth of the potential well. Biasing might at first sight seem just an ad hoc contrivance introduced by theorists to save the philosophically attractive $\Omega = 1$ model when confronted with apparently conflicting evidence. But there are physical reasons for expecting it in the cold dark matter cosmogony. Bright galaxies would be more clustered than the mass for the same reason that in an ocean swell the highest waves come in groups: peaks are more likely to be exceptionally high if they are superimposed on a very large scale positive fluctuation (an incipient cluster) rather than in an incipient void (Figure 12). Nonetheless, the case for $\Omega = 1$ comes primarily from theoretical prejudice and has no really direct observational support.

LARGE-SCALE STRUCTURE

A word now about large-scale structure. Initial fluctuations imprinted in the early universe cannot 'know' what is special about galactic mass. They will spread up to larger scales – though the amplitude must fall off with increasing scale, because of the Universe's overall observed homogeneity. We may get cleaner evidence on these fluctuations from the bigger scales, because these have not yet been confused by non-linear and dissipative effects. Hence the interest in clustering and superclusters. The well-known Lick counts whose clustering properties have been analysed extensively by Peebles and his co-workers, are now complemented by data from the UK southern sky survey. Maddox and Efstathiou, using the APM machine in Cambridge, have now studied the galaxy correlation function and clustering data for this survey.

Objective statistical tests for large-scale clustering are sorely needed. Astronomers react to the data rather as to ink-blot psychological tests. Some see filamentary features, bubbles, or sheets. Others see only gaussian fluctuations, the contrast enhancement of the bright galaxies being perhaps enhanced by biasing.

Redshifts are now available for larger samples of galaxies, and apparent sheet-like structures are seen when the data are plotted with redshift as a radial coordinate. However, one cannot tell whether such a feature is a physically thin sheet, or a thick over-dense region expanding slower than the Hubble flow with a small velocity difference between front and back. What is really wanted for probing the dynamics are the velocities *relative to* the Hubble flow: the motions induced by the gravitational fields of clusters, or perhaps by giant explosive events. For this one needs not just redshift, but an independent measure of distance. This is just becoming possible for some samples. We shall soon know more about the reality of large-scale motions ('cosmic plate tectonics') and whether clusters really have large velocities relative to the Hubble flow. Also, are the edges of voids expanding faster than the Hubble flow? If not, the total density in voids cannot be much lower than outside.

Another line of evidence comes from the microwave background isotropy. This is amazingly precise, upper limits now being a few times 10^{-5} . The universe was therefore quite smooth at $z = 1000$, the redshift where, in most models, photons were last scattered. But perturbations grow by only a factor a 1000 at most since that epoch. Can we then reconcile the existence of conspicuous large-amplitude inhomogeneities in the present universe, on the scales of superclusters and voids, with the smoothness at the recombination epoch which these microwave limits imply – with the lack of gravitational or Doppler perturbations due to incipient clusters? We still have only upper limits, but these are now stringent enough to exclude some options. Theorists still tell the microwave observers, as they have been telling them for the last fifteen years or more, that another factor of two improvement will be crucial.

THE EPOCH OF GALAXY FORMATION AND THE FIRST QUASARS

The microwave background is our most direct link with very high redshifts, but Maarten Schmidt reminded us in his discourse that quasars are now seen with redshifts exceeding 4. Their light set out when the universe was less than a fifth of its present scale. The corresponding age is model-dependent, but for $H = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $\Omega = 1$, it is only around a billion years. Maarten Schmidt also displayed the dramatic peak in quasar activity at redshifts of 2 or 2.5. [It is an anti-anthropocentric irony that the most interesting time to be an astronomer was before the Earth formed.]

At least a few galaxies must therefore have formed, and evolved to the stage when a runaway catastrophe occurs in their nuclei, at a redshift exceeding 4. This is a severe problem for top-down models involving neutrinos. It is also a constraint on hierarchical models, especially the CDM model, in which galactic masses aggregate rather late. Only very exceptional CDM peaks would be on galactic scales at these redshifts. It is hard to quantify how much of a problem the data already pose for the CDM model. We need to know how big a galaxy has to be to ‘host’ a quasar, and how many generations of quasars there are. But the embarrassment threshold of CDM advocates will certainly need to rise if the redshift barrier gets pushed much beyond 5.

One *could* take the apparent quasar cut-off setting in beyond redshifts 2.5 as corroboration of the CDM model. Or one could attribute the peak to something else. And there is an argument, robust and insensitive to details, that an important stage of galaxy formation, the formation of disks, must have happened late, at a redshift less than 3, even if spheroids formed much earlier.

Angular momentum cannot be stored in the big bang. Protogalaxies would have had random shapes, and non-zero quadrupole moments. They would then have acquired angular momentum near the time of turnaround via tidal torques (Figure 13). This process imparts, however, only 10% of what is needed for rotational support. That is a problem unless there is a long ‘lever arm’, whereby the angular momentum could have been acquired when the protogalaxy had a ten times larger radius than now. The gas that makes a 10 kpc disk would then have to have fallen in from beyond 100 kpc. That infall would have taken more than a billion years, implying recent disk formation even if spheroids formed earlier (Figure 14).

Galaxies are generally isolated, but mergers occur today. We see many interacting galaxies. Moreover, mergers would have been much more frequent at larger redshifts. The next series of slides (not, unfortunately, reproducible in the

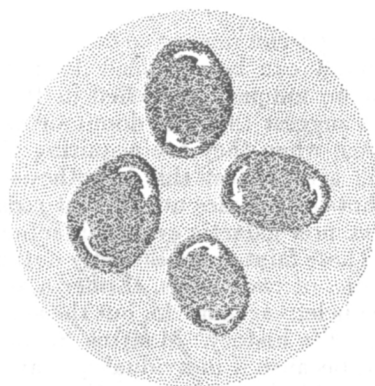


Figure 13. *Protogalaxies would have non-zero quadrupole moments, and their mutual gravitational interactions near the time of turnaround would have imparted angular momentum to each.*

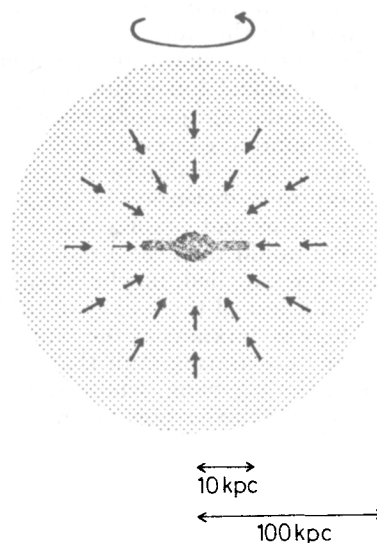


Figure 14 *Tidal torques typically impart less than 10 per cent of the rotational velocity needed for centrifugal support. The material that ends up in a disc of radius 10 kpc must therefore have fallen in from $\gtrsim 100$ kpc. This infall timescale is $\sim 10^9$ yrs, so the formation of discs cannot have been completed before the epoch corresponding a redshift of 2 or 3.*

written version) shows a merger of a disk and small elliptical, from simulations by Hernquist at Princeton. Many galaxies must have experienced such events.

'NORMAL' GALAXIES

Quasars may involve an atypical subset of galaxies. So we must be cautious about inferring anything about typical galaxies from the quasar redshift distribution. The same is true of radio galaxies, because they are exceptional too. [We heard just today, incidentally, that the galactic redshift record had been broken, with a newly discovered object at $z = 3.8$.] Until recently, hardly anything was known about *ordinary* galaxies sufficiently far back in time for evolutionary changes to really show up. But large telescopes and more sensitive detectors are changing

this. Images can now go so faint that there are 100,000 galaxies per square degree, and counts can be compiled down to fainter than 26 magnitude. These counts cannot be uniquely modelled in the absence of knowledge of the redshift. But models suggest that the dominant faint galaxy population may be being seen at the stage when they are acquiring disks. We must await the sharper images the ST will give to test this hypothesis. Next-generation telescopes should give us snapshots of galaxies at different redshifts, (different epochs), thereby allowing us to check how galactic evolution actually occurred.

These faint galaxies vastly outnumber quasars and radio sources, which could mean one of two things. *Either* a very small fraction of galaxies have long-lived active nuclei; *or* more do, but the activity represents a relatively brief phase in each galaxy's life history. Maarten Schmidt favoured something closer to this second option, because if individual quasars were too long-lived, they would build up to unacceptably large masses. If there were many generations of quasars, we would expect that dead quasars, massive black holes now starved of fuel, should lurk in the nuclei of many nearby galaxies. Recently, just such evidence has emerged, primarily from the work of Dressler, Kormendy, Richstone and Tonry. The stars near the centre of the Andromeda galaxy have a rotation and velocity dispersion revealing a central spike, and seem to be orbiting a dark compact object which at least fits the description of a black hole of $10^7 M_{\odot}$. Similar effects are seen in M32 and the Sombrero galaxy as well.

These holes could be reactivated, perhaps as radio galaxies or Seyferts, if the galaxy were disturbed by a merger. Otherwise they would be quiescent, but not quite. Now and again a star would wander so close that tidal forces ripped it apart. We would then see a flare persisting for as long as it took the debris to be swallowed or expelled, maybe a year or so. Searches for such a phenomenon would be a crucial test of the reality of these quiescent black holes.

There is darkness at the centre of even the most familiar galaxies. Moreover, 90% of the gravitating stuff that binds them may be a dark relic of the hot early phases of the big bang, whose elucidation transcends the physics we understand. Even normal galaxies point to new links between the cosmos and the microworld, depicted here in my final slide (Figure 15).

I argued earlier that the mundane physics of gas cooling and Newtonian collapse singles out a galactic mass and lengthscale, so that a favoured mass need not be imprinted *ab initio*. But there must have been some initial fluctuations. Otherwise the universe would still be amorphously uniform, with no galaxies, no stars, and no astronomers. There is still no agreed understanding of why the universe combines the small-scale roughness needed to initiate galaxies with the large-scale uniformity that has allowed it to expand smoothly for 10 billion years. This must await the ultimate synthesis depicted 'gastronomically' in Figure 15.

The problems of large-scale cosmogony are so intermeshed that we will not really solve any until the whole picture comes into sharper focus. For instance, we cannot test theories of galaxy formation and evolution until we understand the gas dynamics of star formation, and the possible rôle of active nuclei, as well as the exotic physics of the initial fluctuations.

The empirical data – observations in all wavebands, and laboratory experiments as well – are burgeoning and all advancing the subject. And theorists are injecting a range of not necessarily compatible ideas whose vector sum at least pushes the subject forward. Hubble's great book, 'The Realm of the Nebulae', concludes with these words. 'With increasing distance our knowledge fades and fades rapidly. Eventually we reach the dim boundary, the utmost limits of our telescope.

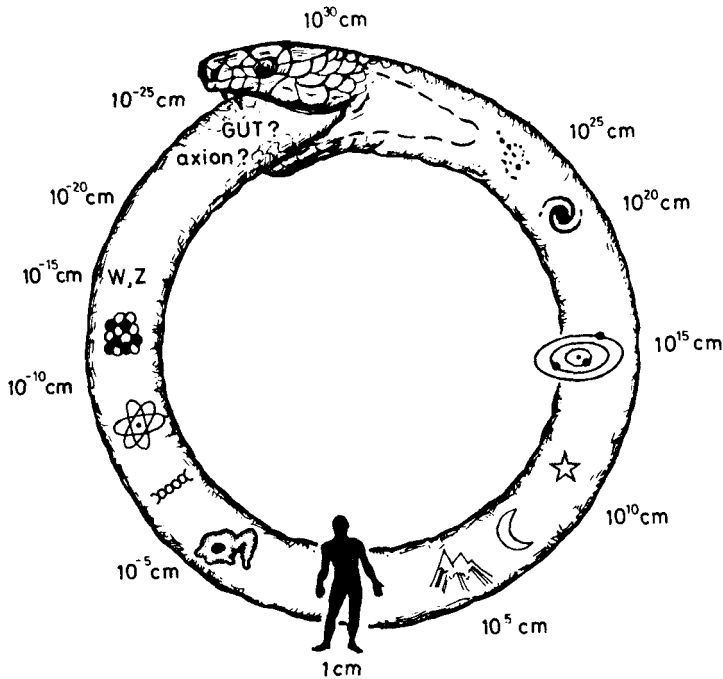


Figure 15. *The everyday world is determined by atomic structure; stellar evolution depends on nuclear physics. Recent ideas, accordingly to which galaxies and clusters may be held together by particles that are relics of the ultra-early universe, suggest further links between the ‘micro’- and the ‘macro’-world – the left and right segments of this picture.*

There we measure shadows, and we search among ghostly errors of measurement for landmarks that are scarcely more substantial. The search will continue. Not until the empirical resources are exhausted need we pass on to the dreamy realm of speculation.’

This search *has* continued as more powerful telescopes and detectors have been deployed. Observers have colonised the speculators’ former territory, and theory itself now has a speculative range undreamt of by Hubble’s contemporaries. The origin of the nebulae, and the emergence of cosmic structure, are still mysterious but the key questions are at least in clearer focus.

I am grateful to many colleagues for discussions and collaboration on topics mentioned in this talk, and to Judith Moss for her careful preparation of the typescript.

REFERENCES *

- Binney, J. and Tremaine, S. *Galactic Dynamics* (Princeton U.P.).
- Blumenthal, G.R., Faber, S.M., Primack, J.R. and Rees, M.J. 1984. *Nature* **311**, 517.
- Bond, J.R. and Szalay, A.S. 1983. *Astrophys. J.* **274**, 443.
- Carr, B.J., Bond, J.R. and Arnett, W.D. 1984. *Astrophys. J.* **277**, 445.
- Cowsik, R. and McClelland, J. 1973. *Astrophys. J.* **180**, 7.
- Einasto, J., Kaasik, A. and Saar, E. 1974. *Nature* **250**, 309.
- Frenk, C., Davis, M., Efstathiou, G.P. and White, S.D.M. 1985. *Nature* **317**, 595.
- Gott, J.R. 1981. *Astrophys. J.* **243**, 140.
- Gott, J.R., Gunn, J.E., Schramm, D.N. and Tinsley, B.M. 1974. *Astrophys. J.* **194**, 543.
- Lyubimov, V.A., Novikov, E.G., Nozik, V.Z., Tretyakov, E.F., and Kozek, V.S. 1980. *Phys. Lett.*, **394**, 266.
- Marx, G. and Szalay, A.S., 1972. *Proc. Neutrino 72, Technoinform, Budapest*, p.191.
- Ostriker, J.P., Peebles, P.J.E. and Yahil, A., 1974. *Astrophys. J. Letters* **193**, L.1.
- Peebles, P.J.E., 1982. *Astrophys. J. Letters* **263**, L.1.
- Press, W.H. and Gunn, J.E., 1973. *Astrophys. J.* **185**, 397.
- Primack, J.R., Seckel, D. and Sandoulet, B., 1988. *Ann. Rev. Nucl. and Particle Sci.* (in press).
- Refsdal, S., 1970. *Astrophys. J.* **159**, 357.
- Rubin, V., 1985. *Highlights of Astronomy* **7**, 27.
- Zel'dovich, Y.B., 1982. *Highlights of Astronomy* **6**, 29.

* It is not feasible to list original references to *all* topics touched on in this 'discourse'. This list primarily includes just the literature specifically mentioned in the text.