Galaxies, Galactic Nuclei and Dark Matter

By Martin J. Rees
Institute of Astronomy
Madingley Road, Cambridge, CB3 0HA, U.K.

I. *Introduction*

The properties of our own Milky Way Galaxy have been expertly reviewed by Professor Ken Freeman. My aim will be to describe some features of galaxies in general. Here we are still groping for answers to the most basic questions:

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.

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 exosystems 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.

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

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 of dead remnants. The main challenge is to elucidate this process, and to determine how and when the protogalactic clouds emerged during the overall expansion of the Universe.

II. *What is Special about Galactic Dimensions?*

This story depicted in Figure 1 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 gas cloud evolves. The first of these is the dynamical or freefall time, which is of order $(G\varrho)^{-1/2}$, 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 $T_g/\varrho\Lambda(T_g)$ where Λ can be calculated from atomic physics.

If $t_{\rm cool}$ exceeds $t_{\rm dynamical}$, a cloud of mass M and radius r can be in quasi-static equilibrium, with the gas at the virial temperature. But if $t_{\rm cool} < t_{\rm dynamical}$ 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 $M_{\rm crit}$ of order 10^{12} $M_\odot$. Clouds less massive than $M_{\rm crit}$ will readily fragment, but above $M_{\rm crit}$ fragmentation is impossible unless the cloud contracts until its radius is below $r_{\rm crit}$. This characteristic mass and radius, consequences of straight-

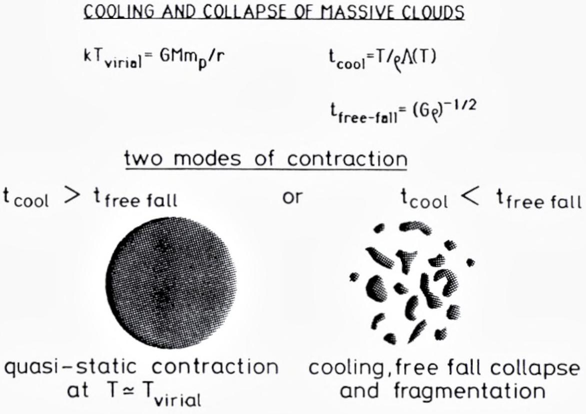

Figure 2. Cooling and contraction of self-gravitating gas clouds.

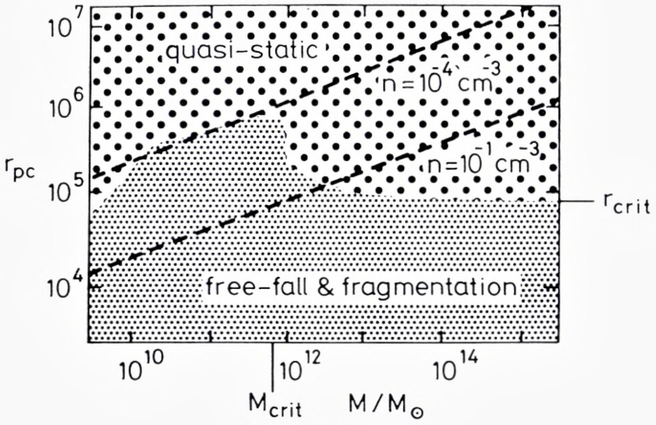

Figure 3. The quasi-static and free-fall regions (cf. Fig. 2) are here presented in a mass-radius plot (from Rees and Ostriker 1977). M_{crit} and r_{crit} should set characteristic upper limits to the dimensions of galaxies. Clouds with $M >> M_{crit}$ would be quasi-static unless at very high densities (cf. gas in clusters of galaxies).

forward physics (Figure 3), 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.

III. *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, 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

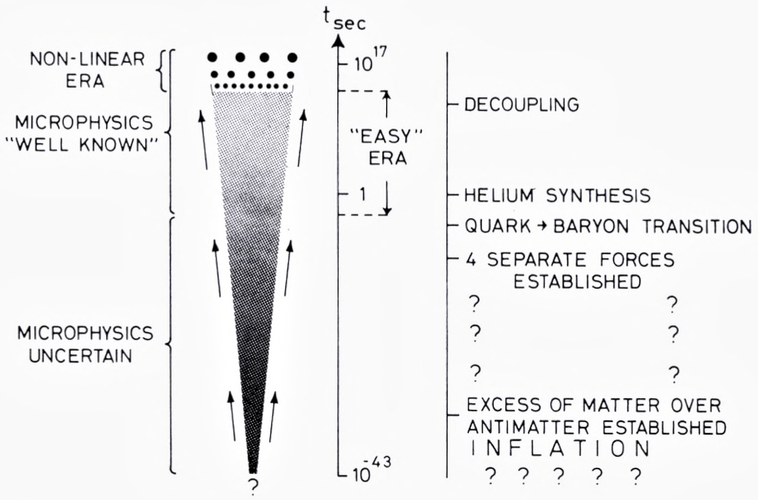

Figure 4. Stages in the evolution of the 'standard' big bang model universe.

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 *micro*physics 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, but the complications are those of non-linearity. The 'recent' universe is hard to understand for the same reason that weather prediction is difficult.

A key question is how much can be explained by processes occurring at the range of epochs accessible to observational astronomers, and how much has to be attributed to the uncertain physics at ultra-early eras.

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.

IV. *Dark Matter*

The evidence for dark matter dates back more than 50 years, but has firmed up since the classic papers of Einasto, Kaasik and Saar, and Ostriker, Peebles and Yahil, both published in 1974. 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 edgeon 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 $R^{-1/2}$, 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.

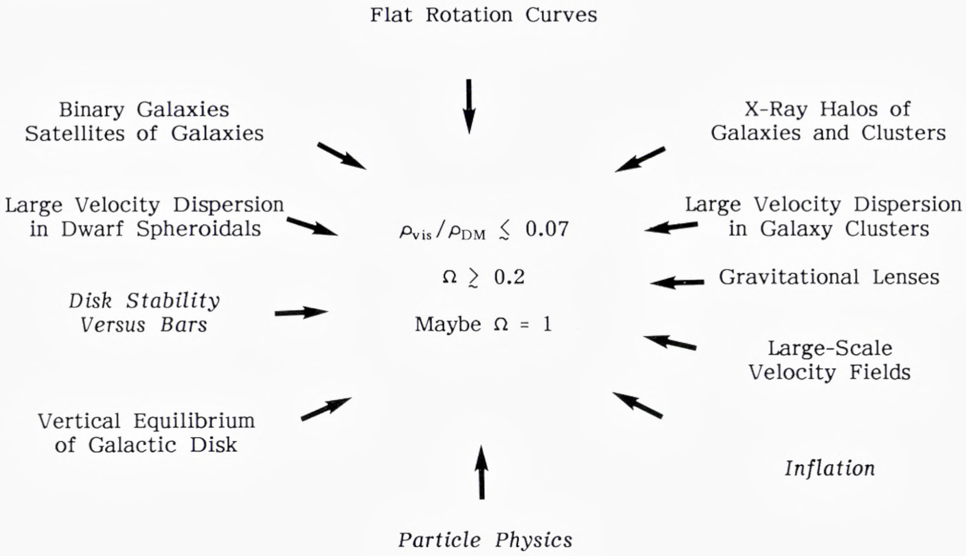

Figure 5. The various lines of evidence for dark matter. Items in italics refer to theoretical arguments (diagram due to J. Kormendy).

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.

V. *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 consti-

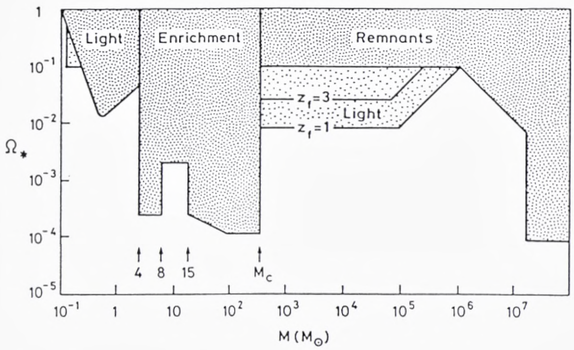

Figure 6. Constraints on the fraction of the critical density that would be present in stars or stellar remnants of various masses. The stars are presumed to have formed at some redshift z_f. Possible candidates for the dark matter are low mass stars (brown dwarfs or 'Jupiters') or very massive objects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

tute 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.

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).

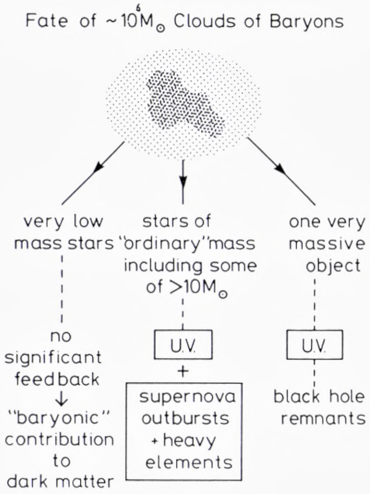

Figure 7. In many cosmogonic schemes, star formation would be initiated in baryonic clouds of around 10^6 $M_\odot$, but we have no firm theoretical basis for deciding the characteristic mass, or the IMF, of these first stars.

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 – 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

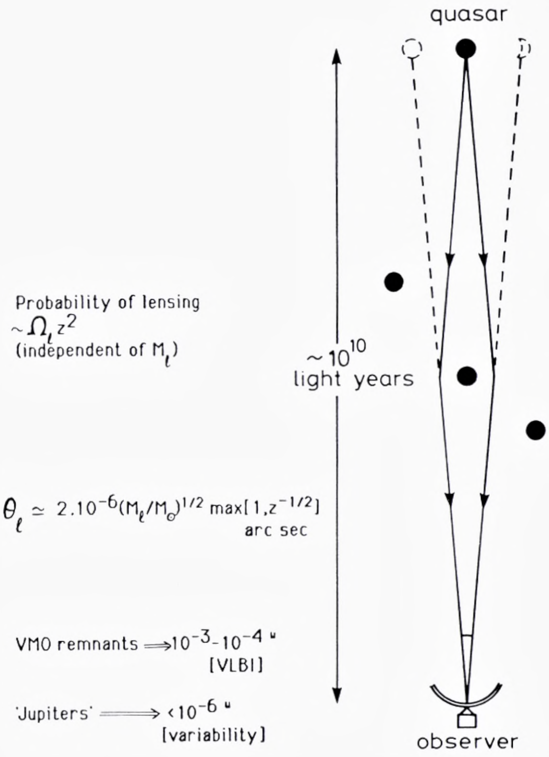

Figure 8. Properties of gravitational microlensing by a population of compact objects along the line of sight to a cosmologically-distant quasar.

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 (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 arc-second. 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 tranverse 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.

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 openminded 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 cosmogonic 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 $10^5 m^{-3}$ and speeds of around 300 kms^{-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 teckniques 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.

VI. *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 pertubations 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 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.)

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

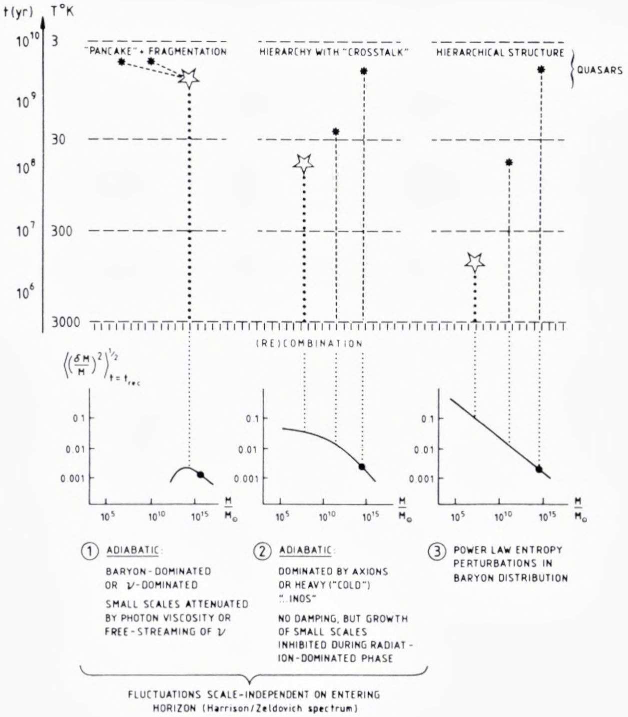

Figure 9. Cosmogonic scenarios corresponding to three different spectra for the post-recombination density perturbations. See text for further explanation.

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 $z \approx 10$. 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.

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 $z = 2$.

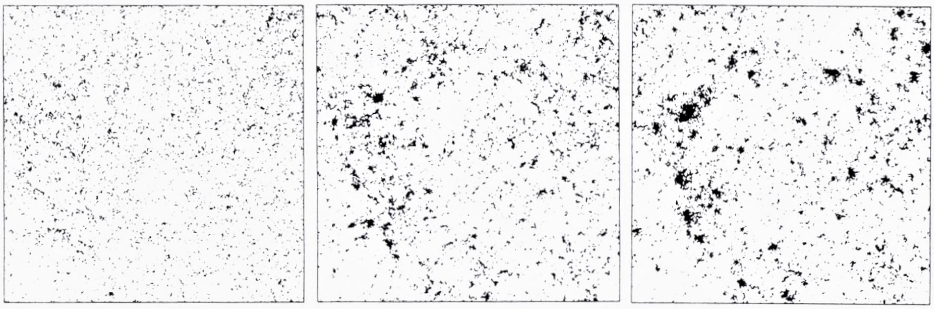

Figure 10. Three stages in the evolution of non-dissipative gravitational clustering within a comoving cubical volume, for an initial spectrum with $[(\delta M/M))^2]^{1/2}$ proportional to $M^{-1/3}$. If the right-hand panel is taken to represent the present epoch, then the middle panel is $z = 0.9$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

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. 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.

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.

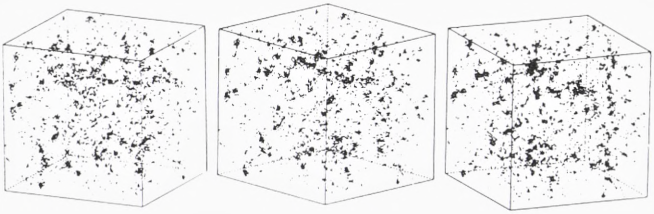

Figure 11. Three views to illustrate the spatial structure within a simulated cubical volume of the expanding universe. In this model the initial fluctuations were Poissonian, with amplitude proportional to $M^{-1/2}$.

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 after-thought 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.

VII. *Active Galactic Nuclei*

It has now been realised for 30 years that not all galaxies are 'mere' assemblages of stars and gas, but that in some there is a central power source – manifested as a quasar, radio galaxy, or Seyfert galaxy – which involves release of gravitational energy in and around a supermassive compact object (see *e.g.* Rees 1984 for a review). Close to such an object, a Newtonian '$1/r$' potential no longer applies. Indeed it is here, in the deepest part of the potential well, that the main power output is generated. We must therefore take the inherently relativistic features of the gravitational field into account.

The physics of dense star clusters and of supermassive objects are complex and poorly understood. In contrast, the final state of such a system, when gravitational collapse occurs, is comparatively simple, at least if we accept general relativity. According to the so-called 'no-hair' theorems, the endpoint of a gravitational collapse, however messy and asymmetric it may have been, is a standardised black hole characterised by just two parameters, mass and spin, and described exactly by the Kerr metric. If the collapse occurred in a violent or sudden way, it would take a few dynamical timescales for the black hole to settle down, and during that period gravitational waves would be emitted. But the final stationary state would be described by the Kerr solution, provided only that the material left behind outside the hole did not provide a strong gravitational pertubation.

Such models can, in broad terms, provide acceptable models for quasars. However, one cannot reliably predict the spectrum, nor whether the radiation is thermal or non-thermal; the hardest thing to estimate is what fraction of the power dissipated by viscous friction would go into relativistic particles (via shocks, magnetic reconnection, etc.) rather than being shared among all the particles. Nor do we know how steady or stable the inflow pattern might be. This is a topic where detailed numerical simulations would be worthwhile, particularly if these allowed us to treat unsteady accretion, non-axisymmetric instabilities, and realistic radiative emission and transfer processes.

Despite the lack of quantitative understanding of AGNs in general, the strong radio galaxies (such as Cygnus A) have a distinctive property that offers a clue to their central mechanism. The remarkable feature of these particular AGNs is that the 'kinetic' power required to energise the extended radio lobes (transmitted by the jets in the form of relativistic particles or Poynting flux) exceeds the radiative luminosity of the nucleus itself. Is there a mechanism that could generate an intense plasma outflow, even if the accretion rate and nuclear luminosity were low?

There is indeed another possible source of power over and above the gravitational energy released by infalling matter. The part of a spinning black hole's rest mass that is associated with its spin can in principle be extracted, as was first emphasised by Penrose (1969). By exploiting the analogy between a black hole's horizon and an electrical conductor, Blandford & Znajek (1977) suggested a realistic astrophysical context whereby electromagnetic torques can extract this energy, rather as the unipolar inductor mechanism brakes an ordinary spinning conductor. Three conditions are necessary, all of which can be fulfilled if the hole is surrounded by a small amount of plasma (as could result from low-level accretion).

(i) Magnetic fields threading the hole must be maintained by an external current system. The requisite flux could have been advected in by slow accretion; even if the field within the inflowing matter were tangled, around the hole it would nevertheless be well ordered. The surrounding plasma would be a good enough conductor to maintain surface currents that could confine such a field within the hole's magnetosphere. The only obvious upper limit to the field is set by the requirement that its total energy should not exceed the gravitational binding energy of the infalling gas.

(ii) There must be a current flowing into the hole. Although the relativistic plasma expected around the hole when $\dot{m}$ is low radiates very little, it emits some bremsstrahlung γ-rays. Some of these will interact in the funnel to produce a cascade of electron-positron pairs, yielding more than enough charge density to 'complete the circuit' and carry the necessary current; enough, indeed, to make the magnetosphere essentially charge-neutral, in the sense that $|(n_+ - n_-)| \ll (n_+ + n_-)$, so that relativistic m.h.d. can be applied.

(iii) The proper 'impedance match' must be achieved between the hole and the external resistance. Phinney (1983) has explored the physics of the relativistic wind whose source is the pair plasma created by $\gamma + \gamma \rightarrow e^+ + e^-$ in the hole's magnetosphere, which flows both outward along the funnel and into the hole. He finds consistent wind solutions in which ~ ½ of the hole's spin energy is transformed into Poynting flux and a relativistic electron-position outflow. [The ubiquity of pair-dominated plasmas in strong compact objects is, incidentally, something which has only been properly appreciated in recent years; this subject owes much to the contributions of NORDITA scientists].

The general scheme is depicted in Figure 12. Even a low-level and inefficient

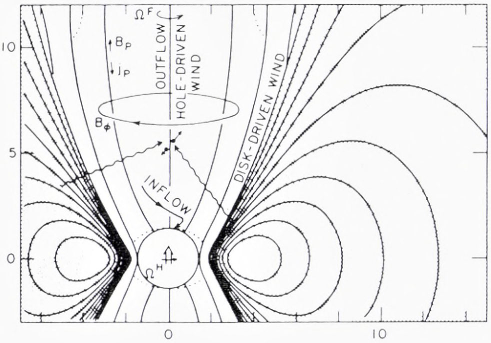

Figure 12. Schematic illustration of the hypothetical 'central engine' in radio sources (Rees *et al.* 1982; Phinney 1983). Interaction between a Kerr hole and the magnetic field generates a hydromagnetic wind. External plasma (stippled) confines a poloidal magnetic field B, of strength 10^3-$10^4 G$, to the hole. The precise geometry is unimportant; that shown is appropriate for a pressure-supported torus with constant specific angular momentum. γ-rays (wavy lines) radiated by the external plasma create pairs in the otherwise empty volume near the symmetry axis from which accreting material is excluded. On field lines which cross the event horizon, these pairs carry a current which extracts rotational energy from the hole in the form of a direct-current Poynting flux.

accretion flow can 'anchor' a magnetic field that threads the hole, and thereby tap the hole's spin energy; in these conditions the extracted power naturally goes predominantly into a relativistic bifurcated outflow. This mechanism seems specially appropriate for strong radio galaxies such as Cygnus A (Rees *et al.* 1982) – objects where the energy flowing along the jets dominates the radiative output of the AGN itself. Electron-positron pairs moving with Lorentz factors $\sim$ 100 would transport some kinetic energy, but most of the power outflow would initially be in the form of Poynting flux associated with the magnetic field coiled round the jet axis, and 'frozen in' to the pair plasma. This Poynting flux may be converted into fast particles where the jet encounters ambient material (perhaps on the scale of the VLBI radio components).

According to this general idea, radio galaxies harbour massive black holes formed long ago via catastrophic collapse (maybe during a quasar phase of activity). The holes lurked quiescent, the galaxy being swept clean of gas, for billions of years. Then some event (perhaps interaction with a companion) triggered renewed infall; maybe at a low rate, but sufficient to reactivate the nucleus by applying a magnetic field. This 'engages the clutch', tapping the hole's latent spin energy, and converting it into non-thermal directed outflow (Poynting flux and $e^+ - e^-$ plasma), which ploughs its way out to scales $\sim 10^{10}$ times larger. If this is indeed what happens in Cygnus A and M87, then these very large-scale manifestations of AGN activity could offer the most direct evidence for inherently relativistic effects.

How much energy is available? Up to 29 % of a hole's mass can be extracted in principle; models for relativistic winds with realistic efficiency factors allow conversion of rest mass with at least a few per cent efficiency – more than enough to power the largest radio galaxy, provided that the hole mass is greater than 10^8 $M_\odot$.

So there are two quite distinct ways in which massive black holes can generate a high luminosity: straightforwardly by accretion; or via the electromagnetic process just described, where the energy comes from the hole itself. The latter process tends to give purely non-thermal phenomena, whereas accretion yields an uncertain mixture of thermal and non-thermal power. The properties of an AGN must depend, among other things, on the relative contributions of these two mechanisms, which are functions of $\dot{M}$ and the spin of the hole. Other parameters may also be relevant, such as the nuclear mass M, the orientation and properties of the host galaxy, etc.

VIII. *What can we realistically expect from Theories and Models?*

How will we ever really know whether there are black holes in AGNs? The evidence can never be more than circumstantial. But we should not be too downcast by that. After all, the evidence that the Sun is powered by nuclear fusion, a cherished dogma never seriously contested, is also really just circumstantial. However, the confrontation of theory with observations, indirect even for stars, is more ambiguous still for quasars: in stars, energy percolates to the observable surface in a relatively steady, symmetric, and well-understood fashion; but in galactic nuclei it is reprocessed into all parts of the electromagnetic spectrum on scales spanning many powers of 10, in a way that depends on poorly-known environmental and geometrical effects in the host galaxy.

The generic black hole model is not infinitely flexible, and not invulnerable. It could be refuted in at least three ways:

1. by finding very regular periodicities, particularly on timescales below 1 hour;
2. by showing that the central masses were $\ll 10^6$ $M_\odot$ in Seyferts or $\ll 10^8$ $M_\odot$ in radio galaxies, or;
3. by developing a theory of gravity more convincing than general relativity which prohibits black holes.

But a clean-cut refutation, leading to abandonment of some theory, happens only rarely in astrophysics; many models have persisted unrefuted for a long time. A cynic might argue that they have survived only because they don't go beyond generalities, or else because their proponents have been adept at replacing or modifying faulty parts to keep shaky old models 'roadworthy'. Such a cynical attitude is not necessarily justified, and to explain why I must digress briefly into methodology. The way we are told science is done is like this: the data suggest a model, which suggests further tests, whereby the original model is either refuted or refined. Such a simple scheme is realistic in, for instance, particle physics, where the fundamental entities may be exactly reducable to a few basic constants and equations. But other sciences deal with inherently complex phenomena and no theoretical scheme can be expected to account for every detail. In geophysics, for instance, the concepts of continental drift and plate tectonics have undoubtedly led to key advances; but they cannot be expected to explain the shape – the precise topography – of the continents. What we should aim to do, in our attempts to understand quasars, is focus on those features of the data which genuinely test crucial ideas, and not to be diverted into measuring or modelling something which is accidental or secondary.

IX. *The AGN Population*

Ideally, one would like a unified model that explains the multifarious types of AGN in the same way that our theories for the Hertzsprung-Russell diagram do this for stars. Still more ambitiously, one would like to place quasars in the general context of galactic evolution. Do quasars die and get resurrected as radio galaxies? How and by what route does a condensed mass first accumulate in the centres of galaxies? What are the masses involved? How do the pyrotechnics in the nucleus react back on the rest of the galaxy? How common is it for massive holes, remnants of past activity, to lurk quiescent in the centres of normal galaxies? Already we have some clues from quasar demography and from the study of nearby galaxies.

Even if AGNs are precursors on the route towards black hole formation rather than involving black holes that have already formed, massive black holes should exist in profusion as remnants of past activity; they would be inconspicuous unless infall onto them recommenced, and generated a renewed phase of accretion-powered output or catalysed the extraction of latent spin energy.

Let us consider further the collective properties of quasars. The total energy output from all quasars in known to within a factor of order 2: it is $\sim$ 3000 $M_\odot$ c^2 per Mpc^3, about half coming from quasars with apparent magnitudes in the range 19-21. Galaxies contribute $\sim 10^5$ in the same units, and the microwave background $\sim 7.5 \times 10^6$. So, even though quasars may influence their host galaxies, they are collectively rather modest contributors to the cosmic energy budget, because of their low space density. (They may nevertheless have a crucial cumulative effect on the entire intergalactic medium, because their energy emerges largely in forms such as an ionizing continuum and high velocity jets and winds.)

The prime era of quasar activity is at $z = 2$ or 3. It is from these redshifts that most of the quasar background light originates. The population thereafter decays on a timescale of order $t_{\mathrm{Evo}} = 2 \times 10^9$ years. This is, however, merely an *upper limit* to the lifetime t_0 of each object: many generations of individual quasars could be born, and could die, in the period over which the population declines. Several important numbers depend on what t_0 actually is: the mass and the number of quasar remnants,

Table 1

(i)	(ii)
$t_0 \simeq t_{Evo}$	$t_0 \simeq 4 \times 10^7$ yr $\simeq 0.02\, t_{\mathrm{Evo}}$
$M = 2.5 \times 10^9\ L_{46}\ \varepsilon_{0.1}^{-1}\ M_\odot$	$M = 5 \times 10^7\ L_{46}\ \varepsilon_{0.1}^{-1}\, M_\odot$
$L << L_{\mathrm{Ed}}$	$L \simeq L_{\mathrm{Ed}}\ \varepsilon_{0.1}^{-1}$
Broad–line regions gravitationally bound	Broad–line region *not* gravitationally bound
Very massive remnants in ~2% of galaxies	$\sim 10^8$ $M_\odot$ remnants in most bright galaxies

the ratio of their luminosity to the Eddington limit, and the issue of whether the broadline region can be gravitationally bound.

Table 1 brings out these points. It contrasts two hypotheses: (i) that there was, in effect, only one generation of quasars, which were long-lived and massive; *versus* (ii) that there were $\sim$ 50 generations of quasars, so that their individual masses (for a given efficiency) need not have built up to such high values, and quasar remnants would be more numerous. In reality t_0 would not be a single number, but there would be a spread of ages, perhaps correlated with luminosity L.

X. *Evidence of Physical Conditions at $Z \gtrsim 2$: the first Quasars*

There are four lines of evidence which imply that conditions at redshifts of 2-4 were very different from those at the present epoch. These are:

(a) *Quasars*. These are now observed at redshifts exceeding 4. They offer evidence on the time-dependence of galactic activity. It has long been known that at $z \simeq 2$ the universe was more violently active, in the sense that the comoving density of powerful AGNs was then much higher. Enough high-z objects have now been found by well-defined search procedures to allow provisional attempts to infer how the luminosity function has evolved. These results, which should still be interpreted cautiously, suggest that there may be a decrease beyond $z \simeq 2.5$. There is, however, no incontrovertible evidence for a really sharp cut-off; and even if there was, it would be hard to decide whether this indicated a cut-off in the sources themselves, or if they became undetectable beyond a certain redshift because of stifling effects within the host galaxy (Rees 1986) or absorption in the intervening intergalactic medium by neutral gas (Rees 1969) or by dust (Ostriker and Heisler 1984). The first bound systems may have formed even earlier than the first quasars: different cosmogonic models (see Figure 9) place the first non-linear behaviour anywhere from $z \simeq 4$ to $z \simeq 1000$.

(b) *Quasars as probes of intervening gas*. The true nature of the gas that gives rise to multiple absorption redshifts in quasar spectra remains controversial. There can be no doubt, however, that the inferred gas clouds have great relevance to galaxy formation. Progress in interpreting these absorption features is, fortunately, not stymied by our poor understanding of the intrinsic properties of quasars: the quasar itself merely serves as a probe of material along the entire line of sight. Some of the absorption lines are caused by HI with a column density as high as 10^{21}cm^{-2} (Wolfe *et al.* 1986), and could involve a protogalaxy or a protogalactic disc. The much larger number of 'clouds' responsible for the Lyman α forest involve much lower HI columns; the relationship of these to galaxies is less clear. Although the details are still controversial, quasar spectra certainly reveal that the amount of HI increases with z over the range 1.8-3.8 where Lyman α can be studied from the ground using known quasars. It is unclear to what extent this implies that the amount of gas is increasing with z, rather than that the ionization level is decreasing with z.

(c) *Direct imaging of (proto)galaxies at high z*. Several extended emission-line objects have been recently observed, which are probably gas-rich galaxies at an early stage in their evolution (see Djorgovski 1988 and Cowie 1988 for reviews).

(d) *Radio structures*. Recent evidence from the radio – the band that can offer the most detailed structural information on high redshift objects – suggests that AGNs were not merely more numerous at high redshifts, but that their radio morphologies were qualitatively different. Barthel and Miley (1988) find that the high redshift radio sources, whether categorised as quasars or as radio galaxies, are more distorted than local sources of similarly high power. (At small redshifts, the powerful sources tend to be symmetric doubles; only those of lower power display strong bending or asymmetry). This suggests that at high z the medium around the sources is more disturbed; it might also suggest that these sources involve interacting galaxies (see, however, Gopal-Krishna and Wiita 1987). Studies of objects at lower redshift suggest that the rarer renewed activity in nuclei at recent epochs is triggered by close encounters or mergers.

XI. *When was Galaxy Formation completed?*

Redshifts $z \simeq 3 - 4$ correspond to a cosmic time $t \simeq 10^9$ yrs. The first bound systems *may* (though they need not) have formed *much* earlier than this. However, a general argument allows us to infer that galaxy formation would not have been completed much before this time: it should therefore not surprise us to find that conditions at $z \simeq 3$ were indeed very different from those prevailing now.

The luminous parts of many galaxies are now known to be embedded in dark halos extending out to $R \sim 100$ kpc. The free-fall time for protogalactic material is of order $(2GM/R^3{}_{\rm turn})^{-1/2}$, where $R_{\rm turn}$ is its radius at turnaround. The collapse factor before a non-dissipative system virialises is ~ 2 (though it can, of course, be larger when radiative cooling permits dissipation). If galaxies were only 10 kpc in size, they could therefore have formed on a timescale of 10^8 years from material that 'turned around' at redshift $z \gtrsim 20$. But the collapse phase of galaxies whose diffuse halos extend out beyond 100 kpc must have taken much longer.

The mean density of the Universe at the turnaround time of the halo material must be lower than the density of the proto-halo itself (by a factor 5.5 in the simplest model for a spherical protogalaxy), so a system whose mean density at turnaround was very low cannot have collapsed until a correspondingly late epoch. The low densities and long dynamical timescales in the outer parts of halos therefore have the crucial implication – irrespective of what these halos actually consist of – that galaxy formation (and, more specifically, the formation of the outer parts of disc galaxies) was not completed until the Universe was $\gtrsim 2.10^9$ years old: *i.e.* not until redshifts $z \simeq 2$, even if it started much earlier. If the angular momentum of discs were acquired via tidal torques (Figure 13) between neighbouring protogalaxies, the 'spin up' must have occurred at a radius $\gtrsim$ 10 times exceeding the present radius. This is a separate argument in favour of 'recent' galaxy formation, at least for disc galaxies (Fall and Efstathiou, 1980; Gunn, 1982).

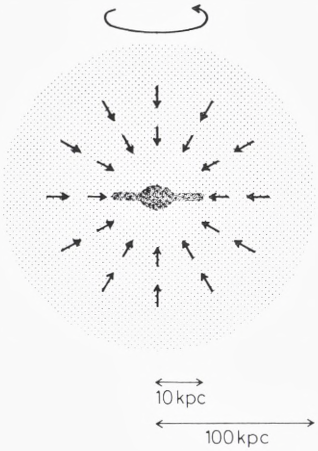

Figure 13. Tidal torques arise between neighbouring protogalaxies, because these typically have non-zero quadrupole moments, but these typically impart less than 10 per cent of the rotational velocity needed for centrifugal support. The material that ends up in a disc of radius 10 kpc must therefore 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 to a redshift of 2 or 3.

Quasar activity peaks at $z \simeq 2$; thereafter the comoving density of luminous quasars declines rapidly as the universe expands, on a timescale much shorter than the expansion timescale. The observed number density at $z \lesssim 2$ is so much smaller than the number density of bright galaxies that the formation of an ultra-luminous quasar within a bright galaxy must be a transient event and therefore sensitive to some (as yet not well understood) feature of the system.

XII. *Black Hole Remnants in Quiescent Galaxies*

Since quasars may involve an atypical subset of galaxies, 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. 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 (see Table 1). *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. Most authors favour 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.

Evidence on the masses, and frequency of occurrence, of black holes in present-day galaxies therefore offers important clues to quasar lifetimes, and to the relationships between different classes of AGNs. The dynamics of stars in the inner regions of nearby galaxies such as M31 and M32 indicate the presence of central concentrated dark masses (Tonry 1984, Kormendy 1988, Dressler and Richstone 1988), of $\sim 10^7\ M_\odot$, and one would like to know whether these might indeed be massive black holes. Since the number of surrounding stars, and their motions, are roughly known, the capture rate by the putative central hole can be estimated; the question then arises of whether the apparent quiescence of the nuclei of these galaxies is compatible with a massive hole's presence. The answer depends on what happens to the debris from each disrupted star. How much is accreted or expelled and with what associated luminosity or radiative efficiency? And, more specifically, *how long* does it take to digest or expel the debris from one star, in comparison with the interval between successive captures? If the stellar density in the central few parsecs is known, it is in principle straightforward to estimate how often a star gets close enough to the central hole to be tidally disrupted (or at least captured). The debris from one solar-mass star per ten thousand years, swallowed steadily with 10 per cent radiative efficiency, would yield a luminosity of 6.10^{41} erg s^{-1} – higher than is observed from M31 or M32.

Several features of the stellar capture process have as yet not been fully analysed:

(i) What fraction of the debris goes down the hole, rather than being expelled?

(ii) What is the radiative efficiency for the accretion process? In other words, how many ergs are radiated for each gram that is swallowed?

(iii) How long do viscous processes, etc., take to 'process' the debris?

The fate of the debris is discussed more fully elsewhere (Rees 1988). The bulk of the bound debris would be swallowed or expelled *rapidly* compared with the interval between successive stellar captures – the only conspicuous luminosity being a flare (predominantly of thermal UV X-ray emission) with $L = L_E$, the hole's Eddington luminosity, fading within a few years. The integrated output from this flare could in principle amount to a few per cent of the star's rest mass, but would probably be far less, because most debris would be 'fed' to the hole far more rapidly than it could be accepted if the radiative efficiency were high; much of the bound debris would then escape in a radiatively-driven directed outflow or be swallowed with low radiative efficiency.

Black holes formed at the era of peak quasar activity ($z \sim 2$) could be reactivated, perhaps as radio galaxies or Seyferts, if the host 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.

XIII. *Conclusions*

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, and perhaps points to new links between the cosmos and the microworld.

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.

The various 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. 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. Some of this written material is adapted from an invited discourse presented at the IAU General Assembly in August 1988.

References

Barthel, P. D. and Miley, G. K. 1988, *Nature* **333**, 319.

Binney, J. and Tremaine, S. 1987, *Galactic Dynamics,* (Princeton U.P.).

Blandford, R. D. and Znajek, R. L. 1977, *Mon. Not. R. astr. Soc.* **169**, 395.

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.

Cowie, L. L. 1988. In *'The Post-Recombination Universe',* ed. N. Kaiser and A. Lasenby (Reidel).

Cowsik, R. and McClelland, J. 1973, Astrophys. J. **180**, 7.

Djorgovski, G. 1988, in *'Towards Understanding Galaxies at Large Redshift',* eds. A. Renzini and R. Kron, (Reidel).

Dressler, A. and Richstone, D. O. 1988, *Astrophys. J.* **324**, 701.

Einasto, J., Kaasik, A. and Saar, E. 1974, *Nature* **250**, 309.

Fall, S. M. and Efstathiou, G. P. 1980, *Mon. Not. R. astr. Soc.* **193**, 189.

Frenk, C. S., White, S. D. M., Davis, M. and Efstathiou, G. P. 1985, *Nature* **317**, 595.

Gopal-Krishna and Wiita, P. J. 1987, *Mon. Not. R. astr. Soc.* **226**, 531.

Gott, J. R. 1981, *Astrophys. J.* **243**, 140.

Gunn, J. E. 1982, in *'Astrophysical Cosmology',* ed. H. A. Bruck *et al.* (Vatican Publications).

Kormendy, J. 1988, *Astrophys. J.* **325**, 128.

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. and Heisler, J. 1984, *Astrophys. J.* **278**, 1.
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.
Penrose, R. 1969, *Revista Nuovo Cim* **1**, 252.
Phinney, E. S. 1983, Cambridge Ph.D. thesis.
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.* **38**, 751.
Rees, M. J. 1969, *Astrophys. Lett.* **4**, 61.
Rees, M. J., Begelman, M. C., Blandford, R. D. and Phinney, E. S. 1982, *Nature* **295**, 17.
Rees, M. J. 1984, *Ann. Rev. Astr. Astrophys.* **22**, 471.
Rees, M. J. 1986. in *'Structure and Evolution of AGNs'*, eds. G. Giuricin *et al.* (Reidel) p. 447.
Rees, M. J. and Ostriker, J. P. 1977, *Mon. Not. R. astr. Soc.* **179**, 541.
Rees, M. J. 1988, *Nature* **333**, 523.
Refsdal, S. 1970, *Astrophys. J.* **159**, 357.
Tonry, J. 1984, *Astrophys. J. Letters* **283**, L27.
Wolfe, A. M., Turnshek, D. A., Smith, H. E., and Cohen, R. D. 1986. *Astrophys. J. Suppl.* **249**, 304.
Zel'dovich, Y. B. 1982, *Highlights of Astronomy* **6**, 29.

Submitted to the Academy August 1989
Published May 1990