What is the Dark Matter in Galactic Halos and Clusters?

M. J. Rees

Phil. Trans. R. Soc. Lond. A 1986 **320**, 573-583
doi: 10.1098/rsta.1986.0138

Phil. Trans. R. Soc. Lond. A **320**, 573–583 (1986)

Printed in Great Britain

What is the dark matter in galactic halos and clusters?

By M. J. Rees, F.R.S.

*Institute of Astronomy, Madingley Road, Cambridge CB*3 0*HA, U.K.*

The dark matter in galactic halos and clusters could be very faint low-mass stars, or in black holes (the remnants of an early generation of very massive stars); alternatively, it could be some species of particle surviving from the early Universe. Although none of these three very different possibilities can yet be excluded, there are real prospects that observations and experiments may soon allow us to discriminate among them. The dynamically inferred dark matter contributes a fraction $\Omega = 0.1$–0.2 of the critical cosmological density. The problem of reconciling the data with the theoretically appealing hypothesis that $\Omega = 1$ is briefly addressed.

1. Introduction

The inferred dark matter in the halos of individual galaxies and in clusters of galaxies apparently contributes a fraction $\Omega = 0.1$–0.2 of the critical cosmological density. Its smoother and less clumped distribution suggests that it underwent less dissipation during the processes of galaxy formation than the luminous stars and gas. This paper is primarily concerned with the nature of the dynamically inferred dark matter; in a final section I shall address the separate question of whether the data are compatible with there being still more dark matter; enough, in particular, to provide the entire critical density ($\Omega = 1$).

Three strands of evidence could eventually pin down what the halo material is.

(i) *Particle physics.* When our theories of high-energy physics become less speculative, and we can calculate how many particles of each species (with known mass) should have survived as relics of the Big Bang, it may turn out that one specific species is predicted, on the basis of standard cosmology, to contribute significantly to Ω.

(ii) *Cosmogonic models.* Evidence on the clustering scales of galaxies, the density profiles in halos, etc, can be compared with the outcome of simulations of gravitational aggregation of 'cold' and 'hot' dark matter (see Frenk 1986, this symposium).

(iii) *Direct detection.* The individual entities that our Galactic Halo is composed of may reveal themselves by astronomical observations, or by direct Earth-based experiments.

2. Baryonic candidates

Most of the initial baryons might have condensed into a population of stars that were either pregalactic, or else formed during the initial collapse phase of protogalaxies: these stars, or their remnants, could perhaps now have a high M/L and contribute to the unseen mass. Ideally, one would like to be able to calculate what happened when the first gravitationally bound clouds condensed from primordial material: did they form one (or a few) supermassive objects, or did fragmentation proceed efficiently down to low-mass stars? Our poor understanding of what determines the initial mass function (IMF) of stars forming now (in, for instance, the Orion nebula) gives us little confidence that we can calculate the nature of stars born in an environment very different from our (present-day) Galaxy. Arguments recently discussed elsewhere (see, for example, Carr *et al.* 1984; Rees 1986) imply that the individual masses of halo objects are *either* less than 0.1 $M_\odot$ *or else* in the range 10^2–10^6 $M_\odot$.

10^6 $M_\odot$ and 10^{-2} $M_\odot$ are two mass scales that emerge naturally in cosmogony. The Jeans mass in the baryonic component of the Universe after recombination is of order 10^6 $M_\odot$; a similar mass appears in a different, but equally relevant, context, as the Jeans mass in *ca.* 10^4 K gas that is in pressure balance with hotter gas in a collapsing protogalaxy (Fall & Rees 1985). If a 10^6 $M_\odot$ cloud collapses isothermally the Jeans mass falls, however, to *ca.* 10^{-2} $M_\odot$ before the onset of opacity effects; this is the case even in a primordial cloud containing no heavy elements. The IMF therefore depends crucially on the complexities of the fragmentation process. Does fragmentation of the first clouds proceed right down to the 'opacity limited' Jeans mass? Or, is fragmentation impeded by collisions (and coalescence) or protostars, or by tidal effects? I do not believe we can yet answer these questions with any confidence; it is therefore worth considering both the 'Jupiter' and 'VMO' options seriously, in the hope that observations can offer some firmer clues than theory.

2.1. *Low-mass stars*

The main constraint on low mass stars in galactic halos comes from limits to the observed optical and infrared emission. A very faint optical halo has been traced out to 300 kpc† in the giant elliptical galaxy M87 (Arp & Bertola 1969). In the edge-on spiral NGC 4565, there are limits on the near infrared emission, corresponding to 76 solar units in the I band (Hegyi & Gerber 1977), and 38 solar units in the K band (Boughn *et al.* 1981). The constraints thereby imposed on the slope of the IMF have been discussed by Peebles (1985) and by Hegyi & Olive (1986). If the entire halo mass were contributed by stars with an IMF such that $dn \propto M^{-(1+x)}\,dM$ down to some minimum mass $M_{\min}$, then if $M_{\min}$ exceeds 0.007 $M_\odot$, $x \geqslant 1.9$ (Salpeter's classic (1955) study derived $x = 1.35$ for our Galactic Disc.) This power-law IMF is only an illustrative example; the same amounts of mass could equally well be contributed by a population with a logarithmic Gaussian distribution peaking below 0.1 $M_\odot$.

One thing is clear: one cannot invoke a smooth extension of the IMF that is observed below 1 $M_\odot$, (which is, however, poorly known below 0.3 $M_\odot$). The objects constituting the dark matter must in some sense be a 'special creation'. However, this is not really a cogent objection to the idea. After all, the IMF derived by Miller & Scalo (1979) and Scalo (1986) for the solar neighbourhood does not look like a single power law, but rather resembles two superposed logarithmic Gaussian distributions. Larson (1986) interprets these as the products of two distinct modes of star formation, whose relative importance may have varied over galactic history. Conceivably the star formation process relevant to halos involved a third mode with a different characteristic mass. If the halo objects formed at an early pregalactic epoch, and subsequently clustered non-dissipatively into galaxies and clusters, then there would be even less reason to suspect that their IMF should resemble that of stars forming here and now. Indeed, we should remain open-minded even about the IMF of stars forming within other galaxies. Cooling flows in clusters of galaxies (reviewed by Fabian *et al.* 1984), where the gas pressure is *ca.* 100 times higher than in our Galactic Disc, imply star formation with a steep IMF. If stars in the

---

† 1 pc ≈ 30857×10^{12} m.

halo formed from a gaseous protogalaxy conditions would have been more similar to those in cooling flows than to those in our Galaxy now. Perhaps the Salpeter–Miller–Scalo function pertains only in environments of atypically low pressure.

Improved infrared limits to the brightness of halos (for example, NGC 4565) can tighten constraints on these hypothetical 'Jupiters', as of course can IRAS-type searches for high-proper-motion objects in our own Galaxy. However, luminosity is such a steep function of mass below 0.1 $M_\odot$ that even a substantial improvement of such tests only tightens the IMF constraints slightly.

2.2. *VMO remnants*

Massive stars of population III could have formed and died early in galactic history. If this happened at a substantial redshift, background light limits offer no constraints. A more conspicuous vestige of such objects would, however, be their contribution to chemical enrichment. Heavy elements are expelled from massive stars in their terminal phases unless they are so massive that they end their lives by collapsing to black holes after the pair-production instability, (Truran & Cameron 1971; Woosley & Weaver 1982; Carr *et al.* 1984 and references therein). Collapse rather than explosion is thought to occur for (non-rotating) core masses above 200 $M_\odot$. If the hidden mass were in VMOs, then the requirement that heavy elements be not overproduced requires an IMF for 'population III' such that only 10^{-3} or less of the mass goes into stars lighter than *ca.* 200 $M_\odot$. Moreover, there must be a cutoff above *ca.* 10^6 $M_\odot$, at least for objects within individual halos, because dynamical friction would have increased the velocity dispersion of disc stars to an excessive degree (Carr 1978; Lacey 1984; Ipser & Semenzato 1985; Lacey & Ostriker 1985). Within the context of VMO theories, we have little evidence on whether the preferred mass is closer to 10^2 or to 10^6 $M_\odot$.

The upper mass limit could be pushed downward if we had a firmer understanding of what luminosity would result from accretion on to black holes passing through the Galactic Disc (Ipser & Price 1977, 1982; McDowell 1985; Lacey & Ostriker 1985). The accretion rate for supersonic motion at speed V through gas of density n is proportional to $nV^{-3}M^2$. The luminosity depends on the accretion rate, and also on the uncertain efficiency ϵ. The angular momentum of infalling matter depends on the poorly known small-scale velocity field in interstellar matter; it is unclear whether there would be enough angular momentum to form a disc, or whether the inflow would be quasi-spherical. For spherical accretion, where the efficiency is low because the radiative cooling time is long compared to the free-fall time, ϵ should scale with M, making the luminosity proportional to M^3. For disc-like accretion, ϵ may be as much as 0.1, independent of M. The spectrum of the emergent radiation is also uncertain. The case of spherical inflow has been considered by Ipser & Price (1977, 1982), who argue that the radiation emerges mainly in cyclotron harmonics from marginally relativistic thermal electrons. This emission would peak in the infrared. Disc-type accretion could yield a higher luminosity, predominantly thermal radiation in the ultraviolet.

The most conspicuous holes would be those that were passing through dense clouds, and that had V much less than the mean velocity. Although, for a Maxwellian distribution, the fraction with speed less than V scales with V^3, the resultant higher $\dot{M}$ ($\propto V^{-3}$) makes these specially slow-moving holes more readily detectable.

The detectability of massive holes in our galaxy has been discussed by McDowell (1985). He shows, following the assumptions of Ipser & Price, that if the typical mass were 10^3 $M_\odot$

or more, the nearest objects passing through a dense interstellar cloud would be at 1 kpc distance, and would contribute 400 Jy flux at 10 μm, well above the IRAS detection limit; the same object would have an optical magnitude $V = 10$ (plus some correction for absorption). Lacey & Ostriker (1985), assuming disc-mode accretion, predict even higher luminosities, but suggest that this would be ultraviolet radiation giving rise to an H II region. The Ipser–Price estimates of luminosity are indeed rather conservative even on the basis of their assumed spherical infall, because a possible non-thermal tail of electrons is neglected. Unfortunately, the distinctive signature of an accreting black hole is hard to estimate, and so one cannot at the moment place firm limits on the number (or mass) of putative halo objects of this kind. Nevertheless, it already seems unlikely that the bulk of the mass could be in collapsed objects individually as heavy as $10^6\ M_{\odot}$.

There is, however an independent reason for conjecturing that our Galactic Halo may primarily consist of objects with masses near the upper limit permitted by dynamical friction constraints. Interaction with halo objects could then, in principle, account for the way the scaleheight and velocity dispersion of the disc stars depend on age. Other possible explanations for this trend, for instance, interactions with molecular clouds or spiral arms, do not account so naturally for the observed ratio of the velocity components in and perpendicular to the disc plane. Carr & Lacey (1986) propose a hybrid model for the halo objects which retains this feature but evades (or at least eases) the accretion constraints: they suggest that halos consist of *ca.* $2 \times 10^6\ M_{\odot}$ clusters of dark objects whose individual masses are stellar or substellar.

2.3. *Gravitational 'minilensing'*

One way of 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 of order 10^{-6} (Refsdal 1964; Paczynski 1986)' but there is, ironically, much more chance of detecting objects in the halos of galaxies (or in clusters) half way out to the Hubble radius. The probability that a compact source at redshift $z_s \gtrsim 1$ is significantly lensed by objects along the line-of-sight is order Ω_l, independent of the individual lens mass involved (Refsdal 1970; Press & Gunn 1974; for a source at redshift $z_s < 1$ the probability is *ca.* $z_s^2\, \Omega_l$. The angular separation, θ_l, of the lens images is, however, a diagnostic of the masses:

$$\theta_l = 2 \times 10^{-6} (M_l / M_{\odot})^{\frac{1}{2}} \max\,[z_s^{-\frac{1}{2}}, 1]''. \qquad (1)$$

For $M \geqslant 10^5\ M_{\odot}$, very long baseline radio interferometers provide adequate resolution. Characteristic image shapes were discussed by Press & Gunn (1973) and Blandford & Jaroszynski (1981). We can probably exclude $\Omega = 1$ in $10^6\ M_{\odot}$ objects, but not $\Omega = 0.1$.

For $M \lesssim 0.1\ M_{\odot}$ ('Jupiters') the angular scale is less than $(10^{-6})''$. 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 (Gott 1981; Young 1981). An object at the Hubble radius moving at *ca.* 10^2 km s^{-1} takes only a few years to traverse an angle $(10^{-6})''$. 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 (Young 1981; Paczynski

1986). 'Minilensing' could affect the optical continuum of quasars but not the spectral lines, because the latter come from a more extended region. If there were a firm observational limit to the scatter in the equivalent widths of the lines from quasar to quasar (i.e. in the line–continuum ratio) this would constrain the value of Ω contributed by small compact objects. Canizares (1982) uses this argument to rule out $\Omega = 1$ in any objects with mass greater $10^{-2}\ M_{\odot}$, but $\Omega \approx 0.1$ cannot yet be excluded.

To detect very small compact objects via lensing requires bright background sources whose intrinsic angular size is well below the value of θ_1 given by (1). The optical continuum of quasars probably comes from a region small enough to be lensed by Jupiters; its typical size, is, however, uncertain, and could be anywhere in the range 10^{14}–10^{17} cm.

No conventional astrophysical process could predominantly produce microscopic discrete masses much less than $10^{-2}\ M_{\odot}$. Such objects could however be the outcome of, for instance, phase transitions at early epochs; primordial black holes are another possibility. Is there any class of source, detectable out to large z, that could be even more compact than quasars, and thereby able to lens such masses? One such candidate would be *supernovae*, whose effective radius at peak light is a few times 10^{14} cm.

Type I supernovae can be detected out to cosmological distances, and have been suggested as 'standard candles' from which Hubble's constant and the deceleration parameter might be determined. A significant contribution to Ω in any compact objects of not less than about $10^{-4}\ M_{\odot}$ would prevent supernovae from behaving as standard candles (Efstathiou & Wagoner 1986). Highly lensed objects would be disproportionately represented in a magnitude-limited sample. If the compact masses were no larger than *ca.* $10^{-4}\ M_{\odot}$ (and θ_1, in (1) were comparable with the intrinsic angular size) the light curve would be distorted because, as the envelope expanded, the magnification along a typical line-of-sight would change. Small compact objects could also induce an apparent polarization in the light from supernovae. Each segment of the photospheric limb of a supernova envelope would be linearly polarized, because of the dominance of electron scattering opacity; but no net polarization would normally be seen unless the envelope were non-spherical (Shapiro & Sutherland 1982). However, compact lensing objects of *ca.* $10^{4}\ M_{\odot}$ would magnify different parts of the envelope by different amounts, thereby causing an observable net polarization even for spherical supernovae.

3. NON-BARYONIC CANDIDATES

The evidence on non-baryonic dark matter is so far restricted to the first two of the three categories mentioned in the introduction. We obviously will not detect gravitational lensing from individual microscopic particles. However, there are now genuine prospects of detecting such candidate particles by other methods.

3.1. *Local effects of 'inos'*

Provided that we know the mass and annihilation cross section for any species of elementary particle, we can in principle calculate how many of them survive from the Big Bang, and the resultant contribution 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 (see Ellis 1986, this symposium).

The prospects of direct detection are not so hopeless for heavy (*ca.* 1 GeV) supersymmetric 'inos' as for light (*ca.* 10 eV) neutrinos. Many proposals for detecting, or at least constraining, candidate particles have indeed been advanced. If the 'inos' were unstable, and photons were among the decay products, there may be observational traces even for a decay timescale as long as 10^{24} s. This is because limits to the hard-radiation background amount to only *ca.* 10^{-8} of the critical density. Antiprotons observed in the cosmic radiation may even be decay products of 'inos' (Silk & Srednicki 1984).

There has recently been a spate of interesting suggestions about how 'inos' might reveal their presence relatively close at hand. Weakly interacting massive (*ca.* 1 GeV) particles would have cross sections of order 10^{-36} cm^{-2} for interactions with nucleons. The 'optical depth' of the Sun is of the order of $(\sigma/(10^{-36}\ \mathrm{cm}^{-2}))$. Such particles, scattering elastically off nucleons in the Sun, would lose energy via the recoil, and could thereby become trapped (Steigman *et al.* 1978; Spergel & Press 1985). Despite being vastly outnumbered by ordinary nuclei, they could then contribute to energy transport in the solar core, because their mean free path is so long; the central temperature would then be slightly lower, with a consequent reduction in the B neutrino flux. Over the lifetime of the Sun, an isothermal core of 'inos' could build up a mass of $10^{-12}\ M_{\odot}$ if annihilations did not occur. Annihilations would restrict this buildup, unless one adopts a rather artificial model in which the cross section for annihilation is far below that for scattering (Krauss *et al.* 1985). However, even though annihilations may prevent a dense enough core building up to affect the standard solar-neutrino problem, high-energy neutrinos from these annihilations may reveal their presence in the underwater detectors developed to search for proton decay. Already, scalar or Dirac neutrinos with mass exceeding 6 GeV can be excluded. Analogous limits come from considering annihilations in the Earth rather than the Sun, as discussed by Silk *et al.* (1985) and Krauss *et al.* (1985).

Ingenious schemes for detecting a halo population of 'inos' or axions in the laboratory are discussed in the contributions by Jelley (1986, this symposium, and the discussion by Smith). These seem to me among the most worthwhile and exciting high-risk experiments in physics or astrophysics: a null result would surprise nobody; on the other hand such experiments could reveal new supersymmetric particles, as well as determining what 90% of our Universe consists of (and even permitting measurements of the halo's velocity dispersion and rotation).

3.2. *Macroscopic 'cold dark-matter' candidates: nuggets of strange matter or primordial black holes.*

Witten (1984) conjectured that grains or nuggets of 'strange matter', containing up, down, and strange quarks, may survive stably from the quark–hadron transition at $t = 10^{-4}$. Such objects, in some sense intermediate between elementary particles and lumps of astrophysical size, would count as non-baryonic matter in the context of nucleosynthesis. Recent work (Applegate & Hogan 1985; Alcock & Farhi 1985) suggests that neutrino heating when $kT > 1$ MeV would destroy nuggets unless they had a mass of planetary order and it is unclear that any larger ones could even form, because this would involve coordination over a scale larger than the particle horizon at the relevant epoch. If they indeed existed, these nuggets might be detectable (De Rujula & Glashow 1984); interesting constraints could be set from the results of monopole searches, and proton decay experiments, from the number of meteor showers, and from limits on the frequency of small-scale seismic events.

So three very different candidates for the dynamically inferred dark matter in halos and in clusters of galaxies are still in the running: low mass stars ('Jupiters'); black holes of

10^2–10^6 $M_\odot$, the remnants of an early generation of very massive objects (VMOS); and non-baryonic matter, most probably in the form of supersymmetric particles. I would lay roughly equal odds among these three options at the moment. However, it is gratifying that the odds could change quite rapidly, owing either to (i) improved observational and experimental searches for candidate objects, (ii) progress in particle physics or (iii) clearer evidence on how the dark matter is distributed (is it really present, for instance, in dwarf galaxies?).

4. A FLAT ($\Omega = 1$) UNIVERSE?

In an influential review paper published more than a decade ago, Gott *et al.* (1974) summarized the evidence bearing on Ω. They concluded that the dynamical evidence favoured a value 0.1–0.2, and noted that if the matter were all baryonic, the lower end of this range was compatible with the value favoured by standard Big-Bang nucleosynthesis (for a Hubble constant $H_0 \approx 50$ km s^{-1} Mpc^{-1}, a value consistent with the ages of globular clusters, etc.). Much new evidence has accumulated over the last 10 years, especially on cluster dynamics and element abundances; and some relevant theoretical issues have been refined and elaborated. But, if one were to update the Gott *et al.* discussion, their net conclusion would not change much.

There has, however, been a marked change in theorists' attitudes. This is partly because non-baryonic matter is now taken much more seriously, and seems in some ways almost a natural expectation. But the main element in the discussion is the concept of 'inflation': this is so appealing, and resolves some well-known and stubborn cosmological paradoxes in such a natural way, that it instils a strong prejudice in favour of $\Omega = 1$. It is perhaps worth spelling out the basis for this prejudice.

For all the present observable Universe to have evolved from a region that was in causal contact at the earliest times, inflation by a factor of at least 10^{30} is required. In most versions of inflation the exponential growth, once started, rapidly continues for many expansion timescales: it is likely to overshoot, stretching any small part of an initial chaotic hypersurface so that it becomes essentially flat over our present horizon scale. This would yield $\Omega = 1$, with a precision of order 1 part in 10^5 (the expected fluctuation amplitude). For inflation to yield the dynamically preferred value $\Omega \approx 0.1$ or 0.2 the inflation factor would have to be 'just' *ca.* 10^{30}, making the present Robertson–Walker curvature radius of the order of the Hubble radius. This would demand some coincidence. But there would then be an additional requirement that appears still more contrived: our presently observable part of the Universe would have to arise from a segment of the initial hypersurface with the seemingly very special property that its curvature was uniform to a few parts in 10^5; otherwise the curvature fluctuations that could induce quadrupole effects in the microwave background would be at least 10^4 times smaller than the overall Robertson–Walker curvature (see Wilkinson 1986, this symposium). Our Universe could thus not have inflated from a typical element of an initial chaotic hypersurface: the required region would have to be special, rather as a sphere would seem specially smooth if its surface irregularities amounted to 10^{-5} of the uniform mean curvature.

The only real dynamical evidence concerning M/L on scales exceeding clusters of galaxies comes from the Local Supercluster. Our infall towards Virgo is *ca.* 250 km s^{-1}. If the galaxies trace the overall mass distribution, this implies $\Omega \approx 0.2$ (Davis & Peebles 1983). If Ω were 1, the infall velocity, on this same assumption, would be *ca.* 800 km s^{-1}. To reconcile the data with the hypothesis that $\Omega = 1$, the supercluster must be presumed to be only a *ca.* 50% enhancement of the mean density, even though there are *ca.* 3 times as many galaxies within a sphere centred on Virgo, and extending out as far as our galaxy, as in a typical region of the same volume.

If $\Omega = 1$, the dynamical evidence requires that the dominant component of dark matter be more smoothly distributed than the galaxies (or, at least, the conspicuous galaxies included in surveys). Moreover, if the conventional Big-Bang model were correct, the nucleosynthesis constraints would imply that this dark matter was non-baryonic. Possibilities include the following:

4.1. *A homogeneous form of mass-energy contributing $\Omega \approx 0.8$*

(*a*) The Universe may be dynamically dominated by weakly interacting particles which are ultrahot, having random velocities much greater than 10^3 km s^{-1}, and so do not cluster. If such particles had survived from an early epoch they would have always been dynamically dominant over the baryons, and would have inhibited gravitational clustering altogether, as well as yielding an unacceptably fast expansion timescale at the era of nucleosynthesis; a somewhat contrived way round this difficulty involves supposing that these particles arise from decay of heavy particles with lifetimes *ca.* 1 Ga; the clustering could then have developed *before* the decays occurred.

(*b*) The Universe could be flat, even though the ordinary mass-energy density yielded $\Omega < 1$, if a non-zero cosmological constant contributed to the curvature. This hypothesis could be checked by classical cosmological tests: measurements of the deceleration parameter, etc. In some respects this idea resembles alternative (*a*) above; however, the Λ-term is dynamically unimportant at early epochs, and so would not have inhibited galaxy formation. The supposition that Λ should be comparable to ordinary matter in dynamical importance at the present epoch is somewhat unappealing, since it demands the sort of fine tuning that the $\Omega = 1$ hypothesis was intended to avoid, but does not conflict with any observations.

4.2. *Segregation of 'luminous' matter, biasing, etc.*

Alternative (*a*) in 4.1 entails postulating two quite distinct forms of dark mass energy: one completely unclustered, the other gravitationally bound in clusters and halos. This is not inconceivable: the Universe could in principle contain non-baryonic dark matter hot enough to be in category 4.1, the clumped dark matter all being baryonic (as discussed in §1). There could even be two types of non-baryonic matter (Bonometto & Valdarnini 1985). If, contrariwise, there is only one important kind of dark matter, then $\Omega = 1$ implies that it is less clumped than the galaxies: voids need not then be as empty as they look, and the Virgo Supercluster need not be a threefold enhancement in the *total* density. Could there, for instance, be a threshold value of $(\delta\rho/\rho)$, above which galaxy formation were efficient, but below which it was impossible? If so, there could be huge contrasts in the galaxy density from place to place even if the overall density were only slightly non-uniform.

Conceivably the baryons are less smoothly distributed than a dynamically dominant non-baryonic component, even on scales up to *ca.* 20 Mpc. This could, in principle, result from large spatial fluctuations in the photon:baryon ratio in the early Universe. In a neutrino-dominated cosmogony, where the first bound systems would be of supercluster scale, the fact that neutrinos behave as a collisionless gas, whereas the baryons are collisional, would inevitably segregate the two components to some degree, but the deviations from Hubble flow in such models would still be unacceptably large unless galaxies formed in regions of specially low density. Large-scale segregation could also arise if energetic events (explosions or strong winds) from first-generation bound systems had been able to push primordial gas over large distances.

An alternative, less extravagant of energy, is that the large-scale baryon distribution traces that of the dark matter, but the efficiency with which baryons transform into luminous galaxies is patchy. Baryonic matter need not, then, be segregated from non-baryonic matter on scales larger than 1–2 Mpc. (Dissipative infall into galactic halos and cooling flows in clusters would certainly have caused baryons to accumulate preferentially towards the centres of gravitationally bound systems on smaller scales than this.)

A widely discussed possibility is that galaxies formed only from exceptionally high peaks of the density distribution. This prescription reconciles the clustering properties of galaxies with an $\Omega = 1$ cold dark matter model (Davis *et al.* 1985), the reconciliation being achieved because the occurrence of high-σ peaks is specially sensitive to small-amplitude large-scale density fluctuations. Such peaks occur with enhanced probability in the crests rather than the troughs of a large-scale mode, so they display enhanced clustering (Bardeen *et al.* 1986 and references therein). But what astrophysical mechanism prevents lower-amplitude regions from also turning into galaxies, thereby neutralizing the effect?

The high-σ peaks would collapse earlier, and have higher density at turnaround, than more 'typical' fluctuations on a given mass scale. This could, in principle, in itself account for the biasing if star formation were highly sensitive to (for instance) Compton cooling on the microwave background, an effect that depends on $t^{-\frac{8}{3}}$. However, there are several ways whereby the first galaxies could have influenced their environment so as to modify the formation of later galaxies. Galaxy formation could, for instance, have switched off after some epoch (corresponding, say, to the collapse of $\nu_{\text{crit}}\sigma$ peaks where $\nu_{\text{crit}} \approx 3$) if some influence from the early-forming galaxies (with $\nu > 3$) inhibited all later galaxy formation: various physical processes have been suggested for bringing this about (Rees 1985; Silk 1985), but none would seem to do the job very convincingly.

To be relevant to the large-scale galactic distribution, any feedback influence must propagate sufficiently fast from a protocluster to a protovoid. To be more specific, suppose that the galaxies evolving from $\nu\sigma$ fluctuations, where $\nu > \nu_{\text{crit}}$, quench the formation of later galaxies (which would have developed from density peaks with $\nu < \nu_{\text{crit}}$); suppose also that the turnaround time for a critical ($\nu = \nu_{\text{crit}}$) perturbation is t_{crit}. Then the propagation time must be less than $(t_{\text{crit}}/\nu_{\text{crit}})$.

Any inhibiting effect due to energy output from a young galaxy would be more efficient in quenching an incipient galaxy if it was close by, rather than 10–20 Mpc away. Such an effect could lead to 'antibiasing'; the galaxies that formed would be anticorrelated, and less clustered than the overall mass distribution. The relative importance of 'antibiasing' and 'biasing' depends on a further timescale Δt, defined as the timelag between the last instant when an incipient galaxy can be quenched, and the time when (if not quenched) its energy would cause feedback. This timescale must *exceed* $t_{\text{crit}}/\nu_{\text{crit}}$ in order that negative feedback leads to enhanced clustering.

Note that biasing also results if galaxies enhance the formation of others close to them. (Some authors, e.g. Ostriker & Cowie (1981) and Ikeuchi (1981), have conjectured that such effects dominate the cosmogonic process.)

The most stringent demands on such mechanisms comes from the Local Supercluster, as Peebles (1986) has specially emphasized. If $\Omega = 1$ simple dynamical fits to the Virgo infall velocity imply that the Supercluster is now only a 50% density enhancement; when the first galaxies formed (at $z > 3$) it would have been only a 10% enhancement. Could such a small-amplitude effect have made galaxy formation substantially more efficient throughout a volume large enough to encompass several hundred galaxies? (The required enhancement factor is normally claimed to be about three.) To reconcile the large-scale voids and superclusters with $\Omega = 1$ may require some mechanism over and above the type of biasing process that could plausibly account for the galaxy–galaxy correlation function (cf. Davis *et al.* 1985).

4.3. *Non-random phases*

Other effects, aside from feedback processes and the special properties of high-ν Gaussian peaks, could generate a patchy distribution of galaxies, such that light failed to trace the overall large-scale mass distribution.

If the initial fluctuations had non-random phases, then one possibility might be that the Universe was inherently smoother in some places than in others: galactic-scale fluctuations with amplitudes large enough to form bound systems could abound in some regions, but be entirely absent in large 'calm patches'. The latter would then correspond to voids. The only specific model that incorporates this feature is string-induced galaxy formation. The correlation function of the stings can be calculated (Brandenberger & Turok 1985; Kibble & Turok 1986, this symposium); however, the precise way in which the dynamically significant constituents of the Universe (baryons, plus (maybe) a cold non-baryonic component) are influenced by strings has not yet been worked out.

Answers to the following questions would help to distinguish between various options outlined in this section.

(i) How much diffuse gas is there in the voids?

(ii) Are voids empty of all types of galaxy, or only those types that are most conspicuous? Any evidence that galaxies of different morphological types display unequal degrees of clustering is relevant here.

(iii) Are there any galactic-mass dark halos with no luminous galaxy within them? Such objects would be expected in, for instance, the cold dark matter model (assuming fluctuations with random phases) and could, if their core radius were small enough, account for gravitational lensing of quasars even when no lens is visible.

(iv) Do some galaxies (or even all galaxies of some morphological type) lack dark halos? Were this so – were it possible, for instance, for galaxies to form (even in a universe dominated by non-baryonic matter) from regions where the baryons had been compressed to (say) 10 times the average density – a complex multistage cosmogonic scheme would be indicated.

REFERENCES

Alcock, C. & Farhi, H. 1985 *Phys. Rev.* D **32**, 1273.
Applegate, J. & Hogan, C. J. 1985 *Phys. Rev.* D **31**, 3037.
Arp, H. & Bertola, F. 1969 *Astrophys. Lett.* **4**, 23.
Bardeen, J. M., Bond, J. R., Kaiser, N. & Szalay, A. S. 1986 *Astrophys. J.* **304**, 15.
Blandford, R. D. & Jaroszynski, R. 1981 *Astrophys. J.* **246**, 1.
Bonometto, S. & Valdarnini, R. 1985 *Astron. Astrophys.* (In the press.)
Bougn, S. P., Saulson, P. R. & Seldner, M. 1981 *Astrophys. J.* **250**, L15.
Brandenburger, R. & Turok, N. 1986 *Phys. Rev.* D **33**, 2175.
Canizares, C. R. 1982 *Astrophys. J.* **263**, 508.
Carr, B. J. 1978 *Comments Astrophys.* **8**, 161.
Carr, B. J., Bond, J. R. & Arnett, W. D. 1984 *Astrophys. J.* **277**, 445.
Carr, B. J. & Lacey C. G. 1986 *Astrophys. J.* (In the Press.)
Davis, M., Efstathiou, G., Frenk, C. S. & White, S. D. M. 1985 *Astrophys. J.* **292**, 371.
Davis, M. & Peebles, P. J. E. 1983 *A. Rev. Astr. Astrophys.* **21**, 109.
De Rujula, A. & Glashow, S. L. 1984 *Nature, Lond.* **312**, 734.
Efstathiou, G. & Wagoner, R. V. 1986 *Astrophys J.* In the press.)
Fabian, A. C., Nulsen, P. E. J. & Canizares, C. R. 1984 *Nature, Lond.* **310**, 733.
Fall, S. M. & Rees, M. J. 1985 *Astrophys. J.* **298**, 18.
Gott, J. R. 1981 *Astrophys. J.* **243**, 140.
Gott, J. R., Gunn, J. E., Schramm, D. N. & Tinsley, B. M. 1974 *Astrophys. J.* **194**, 543.
Hegyi, D. J. & Gerber, G. L. 1977 *Astrophys. J.* **218**, L7.
Hegyi, D. J. & Olive, K. A. 1986 *Astrophys. J.* **303**, 56.
Ikeuchi, S. 1981 *Publs astr. Soc. Japan* **33**, 211.
Ipser, J. R. & Price, R. H. 1977 *Astrophys. J.* **216**, 578.
Ipser, J. R. & Price, R. H. 1982 *Astrophys. J.* **255**, 651.
Ipser, J. R. & Semenzato, R. 1985 *Astron. Astrophys.* **149**, 408.
Krauss, L. M., Freese, K., Spergel, D. N. & Press, W. H. 1985 *Astrophys. J.* **299**, 1001.
Krauss, L. H., Srednicki, M. & Wilczek, F. 1985 Preprint.
Lacey, C. G. 1984 In *Formation and evolution of galaxies and large scale structures in the Universe* (ed. J. Audouze & J. Tran Thanh Van), p. 351. Dordrecht: Reidel.
Lacey, C. G. & Ostriker, J. P. 1985 *Astrophys. J.* **299**, 633.
Larson, R. B. 1986 *Mon. Not. R. astr. Soc.* **218**, 409.
McDowell, J. 1985 *Mon. Not. R. astr. Soc.* **217**, 77.
Miller, G. E. & Scalo, J. M. 1979 *Astrophys. J. Suppl.* **41**, 513.
Ostriker, J. P. & Cowie, L. 1981 *Astrophys. J.* Lett. **243**, L127.
Paczynski, B. 1986 *Astrophys. J.* **301**, 503.
Peebles, P. J. E. 1985 in *Theoretical aspects of astrophysics and cosmology* (ed. J. Sanz & L. Goicoechea), p. 253. Singapore: World Science Publications.
Peebles, P. J. E. 1986 *Nature, Lond.* (In the press.)
Press, W. H. & Gunn, J. E. 1973 *Astrophys. J.* **185**, 397.
Rees, M. J. 1985 *Mon. Not. R. astr. Soc.* **213**, 75.
Rees, M. J. 1986 In *Dark matter in the Universe* (*IAU Symposium no. 117*) (ed. G. Knapp & J. Kormendy). (In the press.) Dordrecht: Reidel.
Refsdal, S. 1964 *Mon. Not. R. astr. Soc.* **128**, 295.
Refsdal, S. 1970 *Astrophys. J.* **159**, 357.
Salpeter, E. E. 1955 *Astrophys. J.* **121**, 161,
Scalo, J. M. 1986 *Fundam. cosmic Phys.* (In the press.)
Shapiro, P. & Sutherland, P. 1982 *Astrophys. J.* **263**, 902.
Silk, J. I. 1985 *Astrophys. J.* **297**, 1.
Silk, J. I., Olive, K. A. & Srednicki, M. 1985 *Nucl. Phys.* (In the press.)
Silk, J. I. & Srednicki, M. 1984 *Phys. Rev. Lett.* **53**, 624.
Spergel, D. & Press, W. H. 1985 *Astrophys. J.* **294**, 663.
Steigman, G., Sarazin, C. L., Quintana, H. & Faulkner, J. 1978 *Astr. J.* **83**, 1050.
Truran, J. W. & Cameron, A. G. W. 1971 *Astrophys. Space. Sci.* **14**, 179.
Witten, E. 1984 *Phys. Rev.* D **30**, 272.
Woosley, S. E. & Weaver, T. A. 1982 In *Supernovae: a survey of current research* (ed. M. J. Rees & R. Stoneham), p. 79. Dordrecht: Reidel.
Young, P. J. 1981 *Astrophys. J.* **244**, 756.