PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Large Numbers and Ratios in Astrophysics and Cosmology [and Discussion]

M. J. Rees and W. H. McCrea

Phil. Trans. R. Soc. Lond. A 1983 **310**, 311-322
doi: 10.1098/rsta.1983.0093

Phil. Trans. R. Soc. Lond. A **310**, 311–322 (1983)
Printed in Great Britain

Large numbers and ratios in astrophysics and cosmology

By M. J. Rees, F.R.S.
Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, U.K.

The masses and lifetimes of stars can be expressed in terms of fundamental constants. Such expressions always involve powers of the number $\hbar c/Gm_p^2$, whose huge magnitude stems from the weakness of gravity on microphysical scales. Our physical understanding of what determines *galactic* dimensions is not yet, however, on the same firm footing. Observational cosmology gives us three basic numbers that characterize our Universe; (i) the Robertson–Walker curvature radius (whose present-day value is $\gtrsim 10^{60}$ Planck lengths); (ii) the baryon-to-photon ratio (of order 10^{-9}); (iii) the amplitude of the initial metric fluctuations which triggered galaxy formation (of order 10^{-4}). We are unsure how (or, indeed, whether) these cosmological numbers can be derived from known physics.

1. Introduction: the significance of α_G

The feebleness of the gravitational force is expressed quantitatively by the value of the 'gravitational fine structure constant' $\alpha_G = Gm_p^2/\hbar c \approx 10^{-38}$ (or by other numbers related to α_G by factors like $\alpha_f = e^2/\hbar c$ or m_e/m_p, where m_p is the proton mass). Familiar arguments, summarized in the accompanying paper by Press & Lightman (1983), tell us that stars - whether they are main sequence stars (gravitationally bound fusion reactors) or white dwarfs - have masses of order the Chandrasekhar mass, $\alpha_G^{-\frac{3}{2}} m_p$ (or $\alpha_G^{-1} M_P$, where M_P is the Planck mass). Stars contain as many as 10^{57} baryons, because this is the number needed for gravitational binding energy to compete with thermal or degeneracy pressure.

Straightforward arguments show, furthermore, that stars are *long lived* (as well as very massive) because gravity is weak. An upper limit to stellar luminosities is the so-called 'Eddington luminosity' $L_E = 4\pi Gm_p M/c\sigma_T$, where σ_T is the Thomson cross section. This is the luminosity for which radiation pressure on free electrons would balance gravity. The lifetime of a star can then be written as

$$t_* = Mc^2/L_E \times (\text{efficiency of rest-mass conversion}) \times (L_E/L_*), \tag{1}$$

which can be expressed as

$$t_* = \tfrac{2}{3}(\alpha_f m_p/m_e)\,(e^2/m_e c^2)\alpha_G^{-1} \times \begin{pmatrix}\text{efficiency of rest}\\ \text{mass conversion}\end{pmatrix} \times L_E/L_*. \tag{2}$$

The first term in brackets is of order 10; the efficiency of nuclear burning is $\lesssim 0.01$, and the stellar luminosity L_* is less than L_E because of other opacity additional to Thomson scattering, and because radiation provides only part of the total pressure. The important feature of (2) is that α_G^{-1} enters explicitly: the fiducial timescale determining the lifetime of stars – the time any object would take to lose its entire rest mass if it radiated with a luminosity L_E – is α_G^{-1} times the light travel time across a classical electron (or $\alpha_G^{-\frac{3}{2}} t_P$, where the Planck timescale, $t_P = (G\hbar/c^5)^{\frac{1}{2}} \approx 5 \times 10^{-44}$ s).

Figure 1 summarizes the physics of stars, planets, etc., in a mass–radius plot (see Press & Lightman 1983 for fuller discussion and references). The most striking feature is that significant phenomena occur for masses related to m_p by simple powers of α_G. The Planck mass, for which Compton and Schwarzschild radii are equal, is $\alpha_G^{-\frac{1}{2}} m_p$. A mass $\alpha_G^{-1} m_p$ corresponds to a black hole whose radius is the size of a proton: such a hole has Hawking (1975) temperature $kT \approx m_p c^2$, and radiates in a time of order $\alpha_G^{-\frac{3}{2}} t_P$ (i.e. a stellar lifetime, t_*). Stellar masses are of order $\alpha_G^{-\frac{3}{2}} m_p$. The mass scale $\alpha_G^{-2} m_p$ is also of significance as being the mass within a Hubble volume for a flat

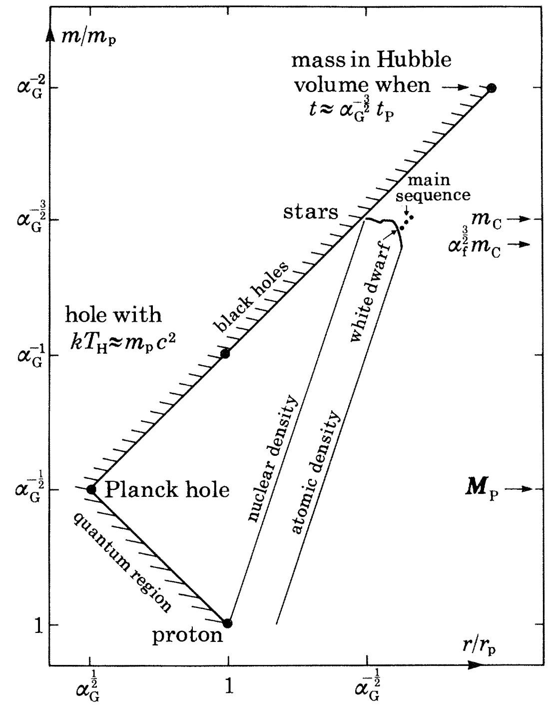

FIGURE 1. This diagram shows schematically, on a (logarithmic) mass–radius plot, how various characteristic scales of structure where gravitation is important involve simple powers of the 'gravitational fine structure constant' α_G. It is because α_G^{-1} is such a vast number, *ca.* 1.7×10^{38}, that so many powers of ten separate the astronomical and cosmological scales from the microphysical scale. Note, however, that the general shape of this diagram is insensitive to the value of α_G: if α_G^{-1} were somewhat smaller, the diagram would be altered only insofar as the separation between the atomic density and nuclear density lines ($(\alpha_f^{-1} m_p/m_e)$ horizontally) would look relatively larger.

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

Friedmann universe whose age is of order t_*. (Galactic masses can perhaps be included on this type of diagram, but as discussed in § 4.1 the physics is less clearcut.) This diagram emphasizes that it is only because α_G^{-1} is so huge that so many powers of ten separate the macrophysical from the astrophysical scales. Nothing in the diagram, however, depends sensitively on the actual value of α_G^{-1} (which is 1.7×10^{38}). If it were (say) of order 10^{32} rather than of order 10^{38}, one could envisage a small-scale speeded up universe where H-burning stars still existed, but were less massive by *ca.* 10^9 and had lifetimes shorter by *ca.* 10^6.

A prime aim of cosmologists is to understand the scale and homogeneity of our Universe. There are around 10^{80} baryons and around 10^{89} thermal photons within the 'Hubble volume'. A universe must plainly have a large space–time volume for stars (and other entities of astrophysical interest) to form and evolve within it, but one would hope to account for this as an outcome of natural initial conditions.

A universe in which stars can form and evolve must persist for a time $t \gtrsim \alpha_G^{-\frac{3}{2}} t_P$. Unless the density is orders of magnitude below 'critical', or unless the mass–energy is overwhelmingly in some non-baryonic form (and it would be hard to form stars at all if either of these were the case) then the total number of baryons must be greater than about α_G^{-2}.

Independently of the properties of stars, we can note that a hot Friedmann cosmological model will not permit small-scale departures from thermodynamic equilibrium until it cools to the stage when the free electrons combine to form atomic hydrogen. This happens at a decoupling temperature T such that $kT \approx (0.02)\,\alpha_f^2 m_e c^2$, where the numerical factor 0.02 arises because the much smaller phase-space volume for bound electrons delays the recombination until well below the temperature of 1 rydberg (equivalent to $0.5\alpha_f^2 m_e c^2$). The time t_{rec} taken to cool to this temperature in a radiation-dominated universe exceeds t_P by a factor of around $10^3 \alpha_G^{-1}(m_p/m_e)^2 \alpha_f^{-4}$: another very large number that explicitly involves α_G^{-1}.

The above considerations show – for what it is worth – that it is because α_G^{-1} is so large that a vast universe is a prerequisite for the existence of stars, cosmologists, and other manifestation of thermodynamic disequilibrium (see Carr & Rees 1979 for further discussion of these points).

2. Robertson–Walker curvature and the density parameter Ω

2.1. *Quantifying the curvature scale*

If the Universe had resembled a Friedmann model since t_P, a comoving scale initially equal to the Planck length would by now have grown by *ca.* 10^{30}, because the scale factor varies as $t^{\frac{1}{2}}$ for a radiation-dominated expression. But the present Hubble radius is *ca.* 10^{60} Planck lengths. The Robertson–Walker curvature radius, which is certainly not much smaller than the present Hubble radius, is thus $\gtrsim 10^{30}$ larger than the 'natural' scale. This problem – the so called 'flatness problem' – introduces another large number, and is perhaps solved by the concept of an 'inflationary' universe, as discussed in this volume by Kibble (1983).

We can write

$$\left\{\frac{\text{Robertson–Walker curvature radius}}{\text{Planck scale expanded to present epoch } t_0}\right\} = \left(\frac{t_0}{t_P}\right)^{\frac{1}{2}} \left(\frac{t_{eq}}{t_0}\right)^{\frac{1}{6}} \frac{1}{|\Omega - 1|^{\frac{1}{2}}}, \tag{3}$$

where Ω is the density parameter, defined as ρ/ρ_c, the critical density being $\rho_c = (\frac{8}{3}\pi G t^2)^{-\frac{1}{2}}$. The extra factor $(t_{eq}/t_0)^{\frac{1}{6}}$, where t_{eq} is the time at which the expansion switches from being radiation dominated to being matter-dominated, arises because for $t > t_{eq}$ the scale factor grows as $t^{\frac{2}{3}}$ rather than as $t^{\frac{1}{2}}$.

The actual density of the Universe – and, in particular, the question of whether $\Omega = 1$ – is of crucial importance. We must consider all forms of matter: dark as well as luminous; non-baryonic as well as baryonic.

2.2. *Contributions to Ω*

2.2.1. *Luminous matter: galaxies and gas*

The present Hubble constant (not in any sense, of course, a 'fundamental' constant) is still uncertain, so in quoting numerical values I will introduce a quantity $h = (3 \times 10^{17} s/t_H)$; the experts advocate values of h in the range 0.5–1. See Sandage & Tammann (1982) and Hanes (1982) for recent assessments from differing viewpoints. The baryon density is then $n_b = 3 \times 10^{-6}\,(\Omega_b h^2)$ cm^{-3}, where Ω_b is defined as the fraction of the critical density in baryonic form. An important number which *is* perhaps of fundamental significance is the ratio of the baryon density to the density of photons in the microwave background, apparently a black body with $T \approx 2.7$ K. This is then

$$\mathscr{S}^{-1} = n_b/n_\gamma \approx 3 \times 10^{-8} (T/2.7\,\mathrm{K})^{-3} (\Omega_b h^2), \tag{4}$$

$\mathscr{S}$ being a measure of the entropy per baryon.

A direct lower limit on Ω_b of order 10^{-2} can be set from the observed mean density of baryons in conspicuously 'luminous' form (visible galaxies, and intergalactic gas revealed by its X-ray emission), implying a baryon-to-photon ratio of not less than about $3 \times 10^{-10} h^2$. This is the number which GUT models must attempt to explain (see § 3).

There is no firm evidence for any anti-matter in the Universe (apart from a small fraction of anti-particles in cosmic rays, which could have been produced in high-energy collisions). Strong constraints on the presence of anti-matter in and around our Galaxy are set by the measured limits to the γ-ray background. Nevertheless, if one is strictly agnostic and free from theoretical preconceptions, one can certainly envisage that the Universe might possess matter–anti-matter symmetry (i.e. that the overall net baryon number, and $\langle n_b/n_\gamma \rangle$, are zero) provided that the scale of the regions of each 'sign' is at least as large as a cluster of galaxies.

2.2.2. *Evidence for unseen mass*

There are dynamical indications of some unseen mass that may or may not be baryonic. The most convincing evidence comes from applications of the virial theorem to rich clusters of galaxies (e.g. the Coma cluster), from the dynamics of our local group of galaxies, and from the statistical technique known as the 'cosmic virial theorem', whereby one analyses the deviations from Hubble-law motions induced in galaxies by their neighbours. These studies are still bedevilled by observational problems, but they broadly suggest that Ω is in the range $0.1 - 0.2$: in other words, there is perhaps ten times as much non-luminous matter as there is in stars and detectable gas. Ten times as much gravitating stuff is implicated in the relative motions of galaxies as in the internal dynamics of individual galaxies. The 'unseen' mass must be in diffuse halos around galaxies, or must pervade clusters or groups of galaxies. The evidence suggesting $\Omega = (0.1 - 0.2)$ comes from studying the dynamics of systems on scales $(1-2)\,h^{-1}$ Mpc.†

It is interesting to ask whether the data *permit* $\Omega = 1$, the value favoured by advocates of 'inflationary' cosmology. The short answer is that this is compatible with the data only if M/L continues to increase with length scale out to $\gtrsim 10\,\mathrm{h}^{-1}$ Mpc, so that on scales where the virial theorem can be reliably applied the unseen mass is less 'clumped' than the luminous mass. Much attention has been given to the velocity field in the local super-cluster, whose scale is 10–20 Mpc. Our infall velocity towards the Virgo cluster, relative to the mean Hubble flow, is apparently too small to permit $\Omega = 1$ *if* the total mass throughout the supercluster is distributed like the galaxies. However, if one drops this assumption, the results become quite inconclusive (Hoffmann & Salpeter 1982).

Only 10 % (and maybe as little as 1 %) of the mass–energy of the Universe is thus in 'known' form. All that can confidently be said about the unseen mass is that it is 'dark': it has a much higher mass-to-light ratio than the ordinary luminous content of galaxies. Whereas the inner parts of typical galaxies have M/L which is $\lesssim 10$ times the solar mass-to-light ratio, if the unseen mass contributes a density parameter Ω it must have M/L exceeding $2300\,\Omega h$ solar units. Possible

† 1 pc $\approx 3 \times 10^{16}$ m.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

forms of unseen mass are discussed elsewhere (see, for example, Einasto & Rees 1983) so I will give just a brief summary here.

2.2.3. *Baryonic forms for the unseen mass*

If galaxies and clusters were 'assembled' from sub-units that condensed earlier, most of the initial baryons might have been incorporated in a pre-galactic population of stars: 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 happens when a cloud of 10^6–$10^8 M_\odot$ condenses out soon after recombination: does it form one (or a few) supermassive objects, or does fragmentation proceed efficiently down to low-mass stars? Our poor understanding of what determines the masss spectrum of stars forming now (in, for instance, the Orion nebula), gives us little confidence that we can calculate the nature of pregalactic stars, born in an environment very different from our (present-day) Galaxy.

Although we cannot confidently predict what these pregalactic stars would be like (Kashlinsky & Rees 1983), there are several constraints which, in combination, imply that if there are enough of them to provide the unseen mass, the individual masses must either be less than $0.1 M_\odot$ or else in the range 10^3–$10^6 M_\odot$. Masses above *ca.* $0.1 M_\odot$ 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. Limits on the range 100–1000$M_\odot$ are uncertain because only ^{4}He may be ejected, the 'heavies' in the core collapsing into a black hole remnant. An uncertainty in the evolution of massive or supermassive stars is the amount of loss during H-burning; however the hypothesis that most mass goes into very massive objects (v.m.os) of greater than about $10^3 M_\odot$ is compatible with the nucleosynthesis constraints. A further consideration favouring these high masses is that v.m.os are likely to terminate their evolution by a collapse which swallows most of the mass: if most of the material were ejected, 'recycling' through several generations would be necessary in order to end up with most of the material in black holes rather than gas. Detailed discussions of pregalactic stars are given by Carr *et al.* (1983) and by Tarbet & Rowan-Robinson (1982).

2.2.4. *Non-baryonic unseen mass?*

If neutrinos have negligible rest mass, the present density expected for relic neutrinos from the big bang is $n_\nu = 110\,(T_\gamma/2.7\,\mathrm{K})^3\,\mathrm{cm^{-3}}$ for each two-component species. This conclusion holds for non-zero masses, provided that $m_\nu c^2$ is far below the thermal energy (*ca.* 5 MeV) at which neutrinos decoupled from other species and that the neutrinos are stable for the Hubble time. Comparison with the baryon density shows that neutrinos outnumber baryons by such a big factor (*ca.* $\mathscr{S}$) that they can be dynamically dominant over baryons even if their masses are only a few electron volts. In fact, a single species of neutrino would yield a contribution to Ω of $\Omega_\nu = 0.01\,h^{-2}(m_\nu)_{\mathrm{eV}}$, so if $h = 0.5$, only 25 eV is sufficient to provide the critical density.

The entire range $100\,h^2$ eV–3 GeV is incompatible with the hot big bang model (Gunn *et al.* 1978). (For $m_\nu > 3$ GeV, the rest mass term in the Boltzmann factor would kill off most of the neutrinos before they decoupled; the number surviving would be less than *ca.* n_b.) If any species of neutrino *were* discovered to have a mass in this excluded range, it would show that one cannot extrapolate the hot big bang back to $kT \gtrsim 5$ MeV, and that most of the photons must have been generated at later times.

(Such arguments are familiarly expressed by saying that the hypothetical particles of non-

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

zero rest mass would 'close the Universe by a large factor'. This loose phrase is in fact rather misleading. The geometry of the Universe is likely to have been laid down at very early stages by mechanisms that do not 'know' what the dominant constituents will be after 10^{10} years. If the Universe were indeed 'flat' (e.g. for 'inflationary' reasons), it would expand with $\Omega = 1$ (i.e. with a density of $(\frac{8}{3}\pi G t^2)^{-1}$) for all t. If the neutrinos were in the excluded mass range, or if there were, for instance, too many primordial monopoles, the observational incompatibility would be that the baryonic fraction of the total density would, for $t \approx 10^{10}$ years, be much less than 10^{-2}.)

Physicists have other particles 'in reserve' – right handed neutrinos, photinos or gravitinos, for instance – which (if they existed) could have been in thermal equilibrium with other species at very early times, and therefore contribute to Ω in an analogous way. The only difference would be that (n/n_γ) could be less than for neutrinos because the other '...inos' may have decoupled before muons (or even hadron pairs) annihilated: the later annihilations would then boost the neutrinos but not the still more weakly coupled '...inos'.

Any such particles would be dynamically important not only for the expanding Universe as a whole but also for large bound systems such as clusters of galaxies. This is because they would now be moving slowly: if the Universe had cooled homogeneously, primordial neutrinos would now be moving at around 200 $(m_\nu)_{\text{eV}}^{-1}\,\text{km}\,\text{s}^{-1}$. They would be influenced even by the weak (*ca.* $10^{-5}\,c^2$) gravitational potential fluctuations of galaxies and clusters. If the three (or more) types of neutrinos have different masses, then the heaviest will obviously be gravitationally dominant, since the numbers of each species should be the same.

It was conjectured more than a decade ago (Cowsik & McLelland 1972; Marx & Szalay 1972) that neutrinos could provide the 'unseen' mass in galactic halos and clusters. In recent years, astrophysicists have explored this possibility in some detail, and considered scenarios for galaxy formation in which neutrino clustering and diffusion play a key role. These scenarios have several appealing features, even though they lead to some new problems.

We may inhabit a universe where, on the largest scales, the baryons are merely a tracer for the distribution of a gravitationally-dominant neutrino sea. Neutrinos of mass *ca.* 10 eV, gravitinos, monopoles or axions are just some of many candidates for the unseen mass in the Universe: as far as the astronomical evidence goes, the unseen mass could equally well be low mass stars, or black holes of up to at least 10^6 solar masses (which could be either primordial, or the remnants of a generation of very massive pregalactic stars). There is no lack of candidates for unseen mass, either baryonic or non-baryonic.

2.3. *Primordial nucleosynthesis: need for non-baryonic matter if $\Omega = 1$?*

Some considerations based on primordial nucleosynthesis seem to favour the non-baryonic option, especially if the total density corresponds to $\Omega = 1$. Primordial nucleosynthesis depends on two things: the expansion timescale at 0.1–1 MeV and the baryon density (which is proportional to $\mathscr{S}^{-1} \propto \Omega_b h^2$). The predicted ^{4}He abundance is rather insensitive to the matter density: for $\Omega_b h^2 \gtrsim 10^{-2}$ (corresponding to $\mathscr{S} \lesssim 3 \times 10^9$) the density of baryons is high enough to ensure that most of the neutrons that survive when the neutron–proton ratio 'freezes out' at $kT \approx 1$ MeV get incorporated in ^{4}He.

The cosmic helium abundance can however be measured with sufficient precision to suggest that the primordial ^{4}He is less than 26 % at the 3σ level (Pagel 1982). This is compatible with $\Omega_b h^2 \lesssim 0.1$ but probably not with $\Omega_b h^2 = 1$ (for $\geqslant 3$ species of neutrinos). The strongest constraint on Ω_b from primordial nucleosynthesis comes, however, not from ^{4}He but from deuterium. This

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

is an intermediate product in helium formation, the amount emerging from the big bang being a steeply decreasing function of Ω_b. Only if $\Omega_b h^2 < 0.025$ can the observed deuterium abundance be produced in a standard hot big bang. The strength of this constraint stems from the failure of astrophysicists in the last decade to suggest any other plausible way of making deuterium.

The combined arguments from primordial nucleosynthesis suggest that $\Omega_b h^2$ is in the range 0.01–0.025 (permitting a value of Ω_b no higher than 0.1, even for $h = 0.5$). See Schramm (1983) for a recent summary. If the lepton number for ν_e and $\bar{\nu}_e$ were non-zero, then the neutron–proton equilibrium ratio would be shifted, affecting ^{4}He production. It is thereby possible in principle to accommodate a higher Ω_b (David & Reeves 1980). However, in order to make much difference, the neutrino lepton number must be of order the photon number; that is, $\mathscr{S}$ times larger than the baryon number. There are other possible complications and 'escape clauses' (involving large-amplitude inhomogeneities in the baryon distribution, etc.). But these considerations suggest that, unless one is to abandon the standard hot big bang model completely, the idea of *non*-baryonic unseen mass is very appealing, especially if (for other reasons) one favours an overall cosmological density as high as the critical value ($\Omega = 1$). Because the relevant parameter in primordial nucleosynthesis is $n_b/n_\gamma \propto \Omega_b h^2$, more precise comparison of models with observation must await a firmer value of the Hubble constant. If $h = 1$ (corresponding to Hubble time of 10^{10} years) then the simplest inference would be that most of the unseen mass – both in the halos of individual galaxies and in clusters and groups – was non-baryonic; but if $h = \frac{1}{2}$ (corresponding to a Hubble time of 2×10^{10} years) the *lower* limit to $\Omega_b h^2$ set by the requirement not to overproduce D + ^{3}He implies that some unseen matter – maybe that in halos, if not in intergalactic space – is baryonic, though only enough to contribute $\Omega_b \approx 0.1$. It is remarkable that the simplest and least arbitrary from of hot big bang model can, for a suitable choice of Ω_b, account for D, ^{4}He and ^{7}Li (Yang *et al.* 1983).

The production of primordial helium is among the few cosmic phenomena sensitive to the actual value of the weak interaction coupling constant. The resultant ^{4}He abundance is *ca.* 25%, rather than *ca.* 0% or *ca.* 100%, because the reactions controlling the neutron–proton ratio ($p + e^- \rightarrow n + \nu$, $p + \bar{\nu} \rightarrow n + e^+$) freeze out when kT has dropped to a value of order $(\Delta m)\, c^2$, where Δm is the proton–neutron mass difference, so the Boltzmann factor is neither close to unity nor ultra-small. (Another process sensitive to neutrino cross sections is the explosion of a supernova; were these cross sections smaller, neutrinos would not be sufficiently well trapped in the stellar core, bouncing at neutron densities, to drive the shock wave which blows off the stellar envelope. If the weak interactions were a factor of 10 weaker, we would have a universe composed primarily of ^{4}He, where supernovae could not explode.)

3. The baryon–photon ratio

Recent ideas on baryon synthesis – if the grand unified theories (GUTs) on which they are based are borne out by future developments – might allow us to test whether the Friedmann models apply back at temperatures of around 10^{15} GeV, corresponding to times *ca.* 10^{-36} s. The observed baryon-to-photon ratio (equation (4)), a measure of the fractional excess of baryons over their antiparticles at early times when $kT > m_p c^2$, is *ca.* 10^{-9}. If this number were much smaller than $\alpha_G^{\frac{1}{4}}$, the Universe would not be baryon-dominated when its age was of order a characteristic stellar lifetime. The value of the net baryon excess arising from out-of-equilibrium decay of X and Y particles can be computed, given a specific GUT (Kolb & Wolfram 1980); it involves a small parameter related to the CP-violation parameter in weak interactions. This work

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

is not yet on the same footing as the calculations of primordial helium and deuterium: it is perhaps at the same level as nucleosynthesis was in the poineering days of Gamow and Lemaitre. But if it could be firmed up it would represent an extraordinary triumph. The mixture of radiation and matter characterizing our Universe would not be *ad hoc* but would be a consequence of the simplest initial conditions. Also, as well as vindicating a GUT, it would reassure us about extrapolating in one bound, based on a Friedmann model, right back to the threshold of classic cosmology, almost back to the Planck time. On a logarithmic scale, this is a bigger extrapolation from the nucleosynthesis era than is involved in going to that era from the present time. It would also place constraints on dissipative processes arising from viscosity, phase transitions, black hole evaporation, etc., which might occur as the Universe cooled through the 'desert' between 10^{15} and 100 GeV. Although these ideas are still speculative, the 'prediction' of the photon–baryon ratio may turn out to offer one of the few empirical tests of GUTs.

If the baryon–photon ratio could be calculated, this would determine Ω_b. If $\Omega_b < 1$, then a strictly flat universe would require some non-baryonic contribution. Of course, one may eventually have theoretical knowledge of the rest masses of all other relevant particles; such information, in conjunction with knowledge of n_b/n_γ, would determine their contribution to Ω also. Looked at from this point of view, it perhaps seems coincidental that non-baryonic matter should dominate, but only by an order of magnitude rather than a vastly larger factor.

4. The origin of galaxies: primordial fluctuations

4.1. *Galaxies: the basic units*

The basic units delineating the Universe's large scale structure are, of course the *galaxies*. Much is now understood about their morphology and internal dynamics. But we still do not know *why* galaxies exist, why the most conspicuous large-scale features of the cosmic scene should be these gravitating aggregates of 10^{10}–10^{12} stars, with dimensions 10^4–10^5 light years. Galaxies have a broad luminosity function, but are no less standardized than stars. Whereas we understand stellar masses, we do not, however, have a convincing commensurate expression for galactic masses. Even worse, we do not know whether the explanation we seek lies within the province of the astrophysicist or the cosmologist. Conceivably the right characteristic mass is somehow 'imprinted' in the early Universe; alternatively, the galactic mass may be singled out by physical processes, just as stars in the stable mass range $\alpha_G^{-1} M_P$ are the end-product of condensation from a broad mass-spectrum of inhomogeneities (in, for example, the Orion nebula) without having to be favoured in the interstellar medium.

An indication that galactic dimensions may be determined by astrophysical processes comes from the following considerations (cf. Rees & Ostriker, 1977; Silk 1977; Binney 1977). A protogalaxy may have started its life as a massive gas cloud, not yet fragmented into stars. Two timescales are important in determining how a self-gravitating gas cloud evolves. The first of these is the dynamical, or free-fall, time t_{dyn}: this is of order $(G\rho)^{-\frac{1}{2}}$, its precise value depending on the geometry of the collapse. The second is the radiative cooling timescale: this depends on the gas temperature T_g, and can be written $t_{cool} = T_g/\rho f(T_g)$, where $f(T_g)$ depends on the composition and ionization of the gas and can be calculated from atomic physics (the cooling rate per unit volume is proportional to $\rho^2 f(T)$, which is why t_{cool} has the quoted dependence on T_g and ρ).

If $t_{cool} > t_{dyn}$, a cloud of mass M and radius R can be in quasi-static equilibrium, with T_g equalling the virial temperature $kT_g = GMm_p/R$. But if $t_{cool} < t_{dyn}$ such equilibrium is impossible:

the cloud cools below the virial temperature, and undergoes free-fall collapse or fragmentation. The criterion $t_{cool} \lesssim t_{dyn}$ also determines whether a shock developing during infall is isothermal (permitting a density enhancement of order the Mach number squared) or adiabatic (with a density enhancement not exceeding 4). We would expect clouds to fragment only if they enter the part of the M–R plane where $t_{cool} < t_{dyn}$. A simple calculation shows that this happens below a mass-independent radius

$$R_c = (e^2/m_e c^2)\,[\alpha_G^{-1}\alpha_f^2(m_p/m_e)^{\frac{1}{2}}] \approx 75\,\text{kpc}, \tag{5}$$

for all masses such that the virial temperature at this radius lies in the range $\alpha_f^2 m_e c^2 \lesssim kT_g \lesssim m_e c^2$ (i.e. for values of kT between a rydberg and the electron rest mass, when non-relativistic bremsstrahlung is the main radiative process). The mass for which the virial temperature at R_c is $kT_g = \alpha_f^2 m_e c^2$ is

$$M_c \approx [\alpha_G^{-2}\alpha_f^5(m_p/m_e)^{\frac{1}{2}}]\,m_p \approx 10^{12} M_\odot. \tag{6}$$

Masses below M_c cool efficiently by H and He recombination and line emission, even at larger radii than R_c. Clouds with mass below M_c will readily fragment; but above M_c fragmentation is impossible unless the cloud contracts to $R < R_c$.

These masses and radii are of the general order relevant to large galaxies, and play a role in many schemes for galaxy formation. However some cosmological processes in the early Universe must have given rise to gas clouds spanning the range around M_c: only then can mass-dependent cooling processes single out a preferred scale for galaxies. Two points are nevertheless uncontroversial:

(1) the early Universe must have contained *some* inhomogeneities (despite its overall Friedmannian character); otherwise its baryon content would now, after *ca.* 10^{10} years, still just comprise uniform neutral H and He.

(2) Whereas gas dynamical and dissipative effects must have been important in the formation of individual galaxies (so that the characteristics of galaxies are related only indirectly to the form of the primordial fluctuation spectrum), any inhomogeneities on much larger scales must be induced primarily by gravitation. The clustering properties of galaxies are the best evidence we have on this. (Note, however, that we can use these to infer the total density fluctuations only insofar as galaxies are a good tracer for the overall mass distribution.)

4.2. *Initial metric fluctuations*

If metric fluctuations exist in the early Universe, then we can constrain their amplitude ϵ on various length scales l as indicated in figure 2. The formation of galaxies by the present epoch from initial curvature fluctuations requires a certain minimum amplitude, which depends on the type of fluctuations (adiabatic or isothermal?) and on the nature of the hidden mass. These *lower* limits representative of classes of models are shown on the diagram. When these models are analysed in detail, other characteristic masses emerge which may be relevant to galaxies, so that the considerations leading to (5) and (6) may not be the complete story. In particular, a significant role is played by the largest mass scale of which free-streaming motions can homogenise the non-baryonic particles (neutrinos?). This is essentially the mass within the particle horizon when kT drops to $m_\nu c^2$: its value is $M_P(M_P/m_\nu)^2$, the analogue of the Chandrasekhar mass for a gravitationally-bound neutrino cloud (with m_ν replacing m_p).

The upper limits in figure 2 are not stringent except for the scales probed by anisotropy measurements of the microwave background, but the following conclusions can be drawn:

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

(i) if ϵ follows a power law $\epsilon(l) \propto l^x$, then $x > -0.15$;

(ii) if ϵ is *independent of scale* (the Harrison (1970)–Zel'dovich (1972) spectrum) as expected in 'inflationary' models, then its value is pinned down in the range 10^{-4}–10^{-5};

but,

(iii) if ϵ does not follow a smooth power law, we have significant constraints only for 4 (out of around 30) powers of ten in l.

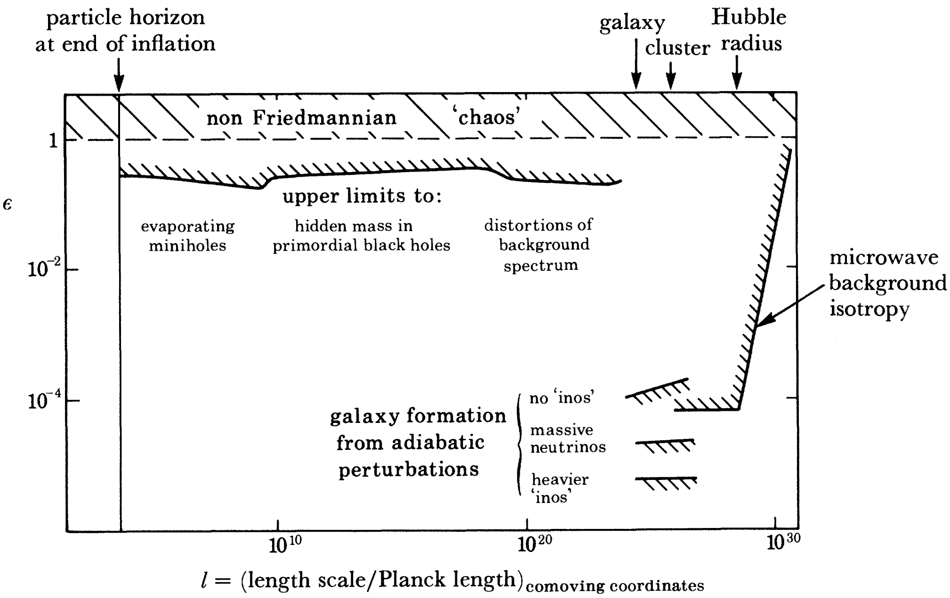

FIGURE 2. This diagram depicts the empirical limits on the amplitude ϵ of adiabatic metric perturbations on various scales. On large scales, the microwave background isotropy offers stringent upper limits. On smaller scales the limits are much less good. If bound systems (galaxies and clusters) in the present Universe evolved from adiabatic initial fluctuations, then *lower* limits are implied to the amplitude on the relevant scales. These limits depend somewhat on the detailed model, and on the nature of the unseen mass: three options are plotted. Because the various limits span a factor *ca.* 10^{30} in comoving length scale, these limits (rough though they are) constrain the slope of any power law $\epsilon \propto l^x$. If the Universe contains scale-independent metric fluctuations ($x = 0$), then the amplitude is pinned down to lie in the range 10^{-4}–10^{-5}. The diagram is drawn assuming a 'flat' background Universe with $\Omega = 1$, but only the large-scale limits are sensitive to this assumption.

5. CONCLUSIONS

The basic properties of stars can all be straightforwardly calculated: their large masses and timescales are a consequence of the vastness of the number $\alpha_G^{-1} \approx 1.7 \times 10^{38}$.

On the still larger scale of galaxies and the Universe, theories are much more provisional. Observational cosmology reveals three important constants.

(i) *The Robertson–Walker curvature radius.* The fact that the Universe is still expanding, after *ca.* $10^{60} t_P$, with a density within an order of magnitude of the 'critical' density, implies that the initial curvature radius at t_P (or at the end of an 'inflationary' phase) is more than *ca.* 10^{30} times larger than the horizon scale at that epoch. The precise value of this curvature radius depends on the density parameter Ω.

(ii) *The baryon-to-photon ratio.* This ratio $\mathscr{S}^{-1}$, given by (3), is a measure of the entropy per baryon. Grand unified theories suggest that it can be explained in terms of baryon non-conservation processes at $t \approx 10^{-36}$ s ($kT \approx 10^{15}$ GeV).

(iii) *The fluctuation amplitude*. The prime mystery is perhaps why the large scale Universe is so homogenous. However *some* fluctuations are essential in order to trigger galaxy formation. If these metric fluctuations have a scale-independent random-phase character, then the amplitude is pinned down to be $\epsilon \approx 10^{-4}$–10^{-5}, a number which theorists may hope to calculate; however if the fluctuations have a more general character, few constraints can be set.

The status of the galactic mass is still uncertain. It may be explicable in terms of physical processes at recent epochs; on the other hand, the scale of galaxies we see may be a consequence of fluctuations imprinted at early times.

Insofar as the aim of physics is to erode the number of independent underivable constants, it is gratifying that there is a serious chance of calculating the quantities listed above in terms of microphysical parameters.

References

Binney, J. J. 1977 *Astrophys. J.* **215**, 483.
Carr, B. J., Bond, J. R. & Arnett, W. D. 1983 *Astrophys. J.* (In the press.)
Carr, B. J. & Rees, M. J. 1979 *Nature, Lond.* **278**, 605.
Cowsik, R. & McLelland, J. 1972 *Phys. Rev. Lett.* **29**, 669.
David, Y. & Reeves, H. 1980 In *Physical cosmology* (ed. R. Balian, J. Audouze & D. N. Schramm), p. 443. Amsterdam: North-Holland.
Einasto, J. & Rees, M. J. 1983 *Nature, Lond.* (Submitted.)
Gunn, J. E., Lee, B. W., Lerche, I., Schramm, D. N. & Steigman, G. 1978 *Astrophys. J.* **223**, 1015.
Hanes, D. A. 1982 *Mon. Not. R. astr. Soc.* **201**, 145.
Harrison, E. R. 1970 *Phys. Rev.* D **1**, 2726.
Hawking, S. W. 1975 *Communs math. Phys.* **43**, 189.
Hoffmann, G. L. & Salpeter, E. E. 1982 *Astrophys. J.* **263**, 485.
Kashlinsky, A. & Rees, M. J. 1983 *Mon. Not. R. astr. Soc.* (In the press.)
Kibble, T. W. B. 1983 *Phil. Trans. R. Soc. Lond.* A **310**, 293.
Kolb, E. & Wolfram, S. 1980 *Nucl. Phys.* B **172**, 224.
Marx, G. & Szalay, A. S. 1972 *Proceedings of the Neutrino 72 Conference*, vol. I, p. 191. Budapest: Technoinform.
Pagel, B. E. J. 1982 *Phil. Trans. R. Soc. Lond.* A **307**, 19.
Press, W. H. & Lightman, A. P. 1983 *Phil. Trans. R. Soc. Lond.* A **310**, 323.
Rees, M. J. & Ostriker, J. P. 1977 *Mon. Not. R. astr. Soc.* **179**, 541.
Sandage, A. R. & Tammann, G. A. 1982 In *Astrophysical cosmology* (ed. H. A. Brück, G. V. Coyne & M. S. Longair), p. 23. Vatican Publications.
Schramm, D. N. 1983 *Phil. Trans. R. Soc. Lond.* A **307**, 43.
Silk, J. I. 1977 *Astrophys. J.* **211**, 638.
Tarbet, P. W. & Rowan-Robinson, M. 1982 *Nature, Lond.* **298**, 7.
Yang, J., Turner, M. S., Steigman, G., Schramm, D. N. & Olive, K. A. 1983 *Astrophys. J.* (Submitted.)
Zel'dovich, Y. B. 1972 *Mon. Not. R. astr. Soc.* **160**, 1P.

Discussion

W. H. McCrea, F.R.S. (*University of Sussex, Brighton BN*1 9*QH, U.K.*). If the amount of dark matter in the Universe greatly exceeds the amount of luminous (baryonic) matter, then, on the hypothesis that the dark matter is baryonic, the luminous matter would be only a small fraction of all the baryonic matter. The abundances of helium and deuterium quoted by Professor Rees are necessarily derived from observation of only that small fraction. Does Professor Rees consider that these abundances would have to be regarded as significant for all the rest? Otherwise, they would not be a compelling reason for rejecting the hypothesis that most of the mass of the Universe is in the form of baryonic matter.

M. J. Rees. The observed baryon content of the Universe could be atypical of all baryonic matter in its He and D abundance only if large-amplitude inhomogeneities already existed at the time of primordial nucleosynthesis ($t = 1-100$ s). This would not be expected according to most theoretical ideas, but is indeed possible in principle, as Professor McCrea proposes. Let me illustrate this by an (admittedly artificial) example. Suppose that the Universe were divided into 'cells', in half of which the baryon-to-photon ratio was 10 times higher than in the other half. If these were *isothermal* perturbations, they would involve only small-amplitude *metric* fluctuations even if the cell size corresponded to 10^3–$10^4\,M_\odot$, because the mass–energy of the uniformly distributed radiation would overwhelmingly dominate that of the baryons at the stage when the cells were larger than the horizon. If all the high-density cells developed into supermassive stars which collapsed into black holes, then all the baryons we now observe would have acquired a chemical composition characteristic of a low density universe even though the actual mean baryonic density was *ca.* 10 times higher.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES