ASTROPHYSICS, COSMOLOGY AND HIGH ENERGY PHYSICS

Martin J. Rees
Institute of Astronomy, Madingley Road, Cambridge CB3 OHA, England

ABSTRACT

A brief survey is given of some topics in astrophysics and cosmology, with special emphasis on the inter-relation between the properties of the early Universe and recent ideas in high energy physics, and on simple order-of-magnitude arguments showing how the scales and dimensions of cosmic phenomena are related to basic physical constants.

1. INTRODUCTION

Astrophysicists spend most of their time applying well-established physics to uncertain and ill-understood cosmic phenomena - they are, basically, users of physics who contribute nothing fundamental to that subject in return. But there are some areas of astronomy when the relevant physics is no better understood than the astronomy. The difficulty and uncertainty are then compounded; the credibility of any conclusions is thereby diminished. But astrophysicists then have the psychological compensation that their relationship with fundamental physicists is symbiotic rather than parasitic. In this lecture I shall try to give a general overview of our knowledge of the Universe and its contents, highlighting some ways in which cosmological considerations may permit some (albeit tentative) inferences about particle physics issues which cannot be settled by ordinary experimental methods.

2. STARS AND GALAXIES

Since I'm addressing physicists rather than astronomers, it may be best to begin right at the beginning, by defining a star. A good definition is this: a star is a self-gravitating aggregate of $N_o \simeq \alpha_G^{-3/2}$ baryons, where $\alpha_G^{-1} = \hbar c/Gm_p^2 \simeq 1.7 \times 10^{38}$ is the inverse of the "gravitational fine structure constant". The vastness of this number reflects the weakness of gravity on the microphysical scale.

Here is a justification of the above "definition". The virial theorem implies that the gravitational binding energy of a star must be of the order of its internal energy. Its internal energy comprises the kinetic energy per particle (radiation pressure being assumed negligible for the moment) and the degeneracy energy per particle. The degeneracy energy will be associated primarily with the Fermi-momentum of the free electrons, $p \sim \hbar/d$, where d is their average separation. Provided the electrons are non-relativistic, the degeneracy energy is $p^2/2m_e$, so the virial theorem implies

$$kT + \frac{\hbar^2}{2m_e d^2} \simeq \frac{GMm_p}{R} \simeq \left(\frac{N}{N_o}\right)^{2/3} \frac{\hbar c}{d} \qquad (1)$$

Here N is the number of protons in the star, and $R \simeq N^{1/3}d$ is its radius. As a cloud collapses under gravity, equation (1) implies that, for large d, T increases as d^{-1}. For small d, however, T will reach a maximum

$$kT_{max} \simeq \left(\frac{N}{N_o}\right)^{4/3} m_e c^2 \tag{2}$$

and then decrease, reaching zero when d is

$$d_{min} \simeq \left(\frac{N}{N_o}\right)^{-2/3} \frac{\hbar}{m_e c} \tag{3}$$

A collapsing cloud becomes a star only if T_{max} is high enough for nuclear reactions to occur, that is $kT_{max} > qm_e c^2$ where q depends on the strong and electromagnetic interaction constants and is $\sim 10^{-2}$ for H → He fusion. From equation (2), one therefore needs $N > 0.1\ N_o$. Once a star has ignited, further collapse will be postponed until it has burnt all its nuclear fuel. An upper limit to the mass of a star derives from the requirement that it should not be radiation-pressure dominated. Such a star would be unstable to pulsations which would probably result in its disruption. Using the virial theorem (that is, equation (1) with the degeneracy term assumed negligible) to relate a star's temperature T to its mass $M \sim Nm_p$ and radius R, the ratio of radiation pressure to matter pressure can be shown to be

$$\frac{P_{rad}}{P_{mat}} \sim \frac{aT^4R^3}{NkT} \simeq \left(\frac{N}{N_o}\right)^2 \tag{4}$$

so the upper limit to the mass of a star is also $\sim N_o m_p$. A more careful calculation shows that there is an extra numerical factor of the order of 10, so one expects stars to lie in the general range $0.1 \lesssim N/N_o \lesssim 10$ which is observed.

In other words, "gravitationally bound fusion reactors" are massive *because* gravity is weak (and α_G^{-1} so large): only assemblages of 10^{56} - 10^{58} particles can turn into stable stars with H-burning cores. Less massive bodies held together by their own gravity can be supported by electron 'exclusion principle' forces at lower temperatures (they would be planets or black dwarfs, not hot enough to undergo nuclear fusion unless squeezed by an external pressure); heavier bodies are fragile and unstable owing to radiation pressure effects.

When we look at the cosmos on a large scale, we find that the basic "building blocks" are entire *galaxies* - either disc (or spiral) galaxies or the amorphous and less photogenic ellipticals. 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 are about as standardised as stars. But whereas we understand physically why stellar masses are $\sim \alpha_G^{-3/2} m_p$ we have no very convincing commensurate expression for galactic masses (though there have been tentative suggestions on this point). 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 $\sim \alpha_G^{-3/2} m_p$ are the end-product of condensation from a broad mass-spectrum of inhomogeneities (in, e.g. the Orion nebula) without having to be favoured in the interstellar medium.

3. "GLOBAL FEATURES OF THE UNIVERSE

3.1 Scale and homogeneity

To the cosmologist, entire galaxies are just 'markers' or test particles scattered through space which indicate how the material content of the Universe is distributed. When a cosmologist makes a statement about how homogeneous the Universe is, he is referring to the distribution of galaxies. When he speaks of the expanding Universe he is basically talking about the motions of galaxies.

Galaxies are clustered; some in small groups like our local group, some in big clusters with hundreds of members. The Universe is manifestly inhomogeneous on all scales up to $\sim 10^8$ light years. But this is still small compared to the Hubble radius ℓ_H, and compared to the depth that can be probed with large telescopes; studies of remote faint galaxies, and of quasars and radio sources, show that these luminous objects are genuinely more isotropic around us on the largest scales we can observe. Unless we adopt anti-Copernican attitudes, this of course implies a high level of *homogeneity* on scales comparable with the Hubble radius. Moreover, the kinematics of the Universe - the expansion in accordance with Hubble's law where recession velocity (or redshift) is a measure of distance - seem to be isotropic.

It is only because the Universe, in its large-scale structure is *specially simple* and *symmetrical* that cosmology is feasible at all. Figure 1 shows a schematic space-time diagram to illustrate this: the regions of space-time about which we have direct evidence lie either close to our own world line (where we have inferences on the chemical and dynamical history of our Galaxy) or along our past light cone (where we have astronomical evidence). It is only because of the overall homogeneity that we can assume any resemblance between the distant galaxies whose light is now reaching us and the early history of our own Galaxy. This simplicity is something we must try to understand and explain.

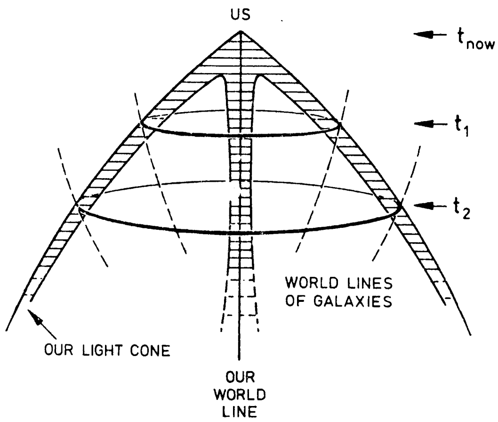

Figure 1. Schematic space-time diagram showing world line of our Galaxy and our past light cone. The only regions of space-time concerning which we have direct evidence are those shaded in the diagram, which lie either close to our own world line (inferences on the chemical and dynamical history of our Galaxy, 'geological' evidence, etc.) or along our past light cone (astronomical evidence). It is *only* because of the overall homogeneity that we can assume any resemblance between the distant galaxies whose light is now reaching us and the early history of our own Galaxy. In homogeneous universes we can define a natural time coordinate, such that all parts of the Universe are similar on hypersurfaces corresponding to a given value of t.

3.2 The isotropic microwave background

The only subsequent cosmological discovery that has matched Hubble's work in importance was the detection by Penzias and Wilson in 1965 of the cosmic microwave background radiation[1]. In the same issue of *Astrophysical Journal* in which this discovery was announced, Dicke and his Princeton colleagues[2] interpreted the microwave background as a relic of a "primordial fireball" phase when the Universe was hot, dense, and opaque. This concept had been developed many years earlier by Gamow, Alpher, and Herman.

The characteristics of this radiation - its isotropy, its apparently thermal spectrum, and its high energy density ($\sim$ 0.25 ev/cm^3 or $\gtrsim$ 10 times the energy density of intergalactic starlight) - were quickly recognized to be irreconcilable with most conventional astrophysical interpretations: a consensus thus emerged very quickly, even among the "conservative element" in astronomy, that this radiation was a relic of very early epochs indeed. Observations at various wavelengths have confirmed a broadly thermal spectrum. Searches for anisotropies have revealed no effect exceeding 10^{-4} on any angular scale down to a few arc min. (Being relative measurements these are much more precise than the intensity measurements.)

According to the "hot big bang" model the material in the early universe would constitute a hot plasma, strongly coupled to a thermal radiation field. The radiation temperature would decrease as $\rho_m^{1/3}$, where ρ_m is the particle density. When adiabatic expansion had cooled the material to approximately 4,000 ^oK, the plasma would recombine and the radiation would no longer undergo substantial scattering or absorption. The microwave photons that are now observed have not been scattered since an epoch when the mean density of matter was $\sim 10^9$ times higher than it is now - long before any galaxies came into existence. The effective source of the background is thus a "cosmic photosphere" at a temperature $\sim 4000\ ^{o}K$. The measured isotropy then implies that the observable part of the Universe has expanded isotropically, with a precision of $\lesssim 1$ part in 10^4, ever since it was compressed by a factor 10^3 in linear scale; moreover, the kinematics can be described by a single scale-factor R(t). The time-dependence of R(t) can be calculated on the basis of some dynamical field equations.

3.3 The Friedmann "hot big bang" model

Cosmological models governed by Einstein's equations evolve in a manner which is controlled by the mean density ρ. It is convenient to parameterise the present density of matter ρ_o as $\Omega_o \rho_{crit}$, where

$$\rho_{crit} = \left(\tfrac{8}{3}\pi G\ \tau_H^2\right)^{-\frac{1}{2}} = 4 \times 10^{-30} (\tau_H / 2 \times 10^{10} yr)^{-2}\ gm\ cm^{-3} \qquad (5)$$

τ_H is the Hubble time, whose value is still uncertain by a factor ~ 2. It is in the range $(1\text{-}2) \times 10^{10}$ yr. If the Universe is describable by a simple Friedmann model, with the cosmical constant Λ equal to zero, then it will expand for ever, or eventually recollapse, according as $\Omega \lesseqgtr 1$. Dynamical arguments seem to favour a value of Ω_o somewhat less than unity, but the issue is still unsettled. The main contribution to Ω_o may not necessarily come from baryons. There is direct evidence for a baryon contribution $\Omega_b \gtrsim 0.01$: this much comes from the stellar content of galaxies, and an intergalactic gas. But the bulk of the mass-energy in the Universe could be in some other form, such as black holes (either primordial, or the end-points of stars or supermassive objects), or neutrinos of non-zero mass. This last possibility, which I shall discuss later, is of course of particular interest to particle physicists.

All cosmological inferences relevant to particle physics are based on the "hot big bang" model of the early Universe. The apparently thermal spectrum of the microwave background suggests that this radiation is a relic of a phase when the Universe was highly opaque. The ratio of photons to baryons - a measure of the "entropy per baryon" in the Universe - is

$$\mathcal{S} = 1.4 \times 10^8 \left(\frac{T_o}{2.7K}\right)^3 \Omega_b^{-1} (\tau_H / 2 \times 10^{10} yr)^2 \qquad (6)$$

Note that the main uncertainty in this quantity stems from our ignorance of Ω_b. All we can say is that lies in the general range $10^8 - 10^{10}$; but somewhat model-dependent considerations based on the physics of the early Universe suggest that this range can be narrowed.

The photons make a negligible contribution to Ω at the present era despite the enormous factor by which they outnumber the baryons: this is because at 2.7K, the mean photon energy is only a few times 10^{-4}eV.

However the radiation temperature (or the energy per photon) varies as R^{-1} as the Universe expands, and the radiation energy density varies as R^{-4}. But the matter density varies only as R^{-3}. At early times, therefore, the expansion was dominated dynamically by radiation. The expansion would be radiation-dominated for $T \gtrsim 3 \times 10^4 \Omega_b$, and when kT is a few MeV the baryon contribution is quite negligible, basically because photons outnumber baryons by a factor $\mathcal{S}$. The Friedmann equations, according to which the expansion rate $\dot{R}/R \propto \rho^{\frac{1}{2}}$, then yield the temperature-time relation

$$t = 2(\mathcal{f}(T))^{-\frac{1}{2}} \left(\frac{kT}{1\ \mathrm{MeV}}\right)^{-2} \ \mathrm{sec.} \tag{7}$$

The quantity $\mathcal{f}(T)$ denotes the factor by which the energy density at temperature T exceeds that due to black-body photons alone. At a temperature of a few MeV, when one has e^+-e^- pairs and neutrinos,

$$\mathcal{f}(T) = {}^{11}/_4 + {}^{7}/_8\, N_\nu + \text{(small possible extra contributions from "exotic" species)} \tag{8}$$

N_ν being the number of two-component neutrino species. At higher temperatures still, other heavier species contribute more heavily to $\mathcal{f}(T)$; but the processes occurring when kT is of order an MeV turn out to be crucial for nucleosynthesis.

Helium is much more abundant, and much more uniformly distributed, than the heavier elements. The latter could all be the products of stellar nucleosynthesis. The helium, on the other hand, is commonly attributed to the hot dense early phase of the big bang : indeed, the most compelling reason for taking seriously the earlier phases (t = 1-100 sec) of a big bang is that the simplest assumptions (i.e. homogeneity, isotropy, no "new physics", Einstein's general relativity, etc.) yield a He abundance in gratifying accordance with observations[3),4)]. The crucial process that determines the amount of He is the neutron/proton "freeze-out" which occurs when the reactions $p + e^- \to n + \nu$, $p + \bar{\nu} \to n + e^+$ become slower than the expansion timescale. In a standard radiation-dominated Friedmann model, the reaction rate goes as T^5 (since the particle density goes as T^3 and the cross section as T^2) and the expansion rate $\sim (G\rho)^{\frac{1}{2}} \propto T^2$. The respective timescales are equal at $kT \simeq 1$ MeV. The neutron/proton ratio is then approximately $e^{-1.5}$, most of these neutrons being subsequently incorporated into D and ^{3}He, and then into ^{4}He, before they have time to decay freely. The predicted ^{4}He abundance depends only slightly on the matter density - for $\Omega_b > 10^{-2}$ (corresponding to $\mathcal{S} \lesssim 10^{10}$) the density of baryons is high enough to ensure that most of the neutrons which survive at "freeze-out" get incorporated in ^{4}He. However, the resulting ^{4}He abundance *is* sensitive to the *expansion rate* when $kT \simeq 1$ MeV : if the Universe expanded somewhat faster, then the neutron/proton ratio would "freeze out" at a higher temperature, when neutrons were less disfavoured by the Boltzmann factor; the ^{4}He abundance would then be *higher*.

Primordial helium synthesis is one of the few astrophysical contexts (supernovae being another) where the precise values of neutrino cross-sections, and the weak interaction coupling constant, enter at all sensitively. If these were made somewhat weaker (stronger), the big bang would produce $\sim$ 100% ($\sim$ 0%) helium. It is only because the weak interaction rates are such that the neutron/proton ratio "freezes out" when kT is of order the proton-neutron mass difference that an "interesting" fraction of helium is synthesised.

3.4 Limits on number of neutrino species

The observed fractional abundance of ^{4}He can be used to place interesting constraints on the number of types of neutrino[5]. This is because the amount of ^{4}He produced depends on the expansion rate (at a given T) and therefore on the number of independent species. One of the main uncertainties in quantifying this line of argument is that we do not know the primordial ^{4}He abundance - all that we can say is that it must be *below* the *lowest reliably-determined* abundance in any astronomical object (since stars can make extra helium during the course of galactic evolution). Most astronomers[6] would assess that the primordial ^{4}He must be $\lesssim$ 25%. This is consistent[4] with $N_\nu = 3$ for $\mathcal{S} > 2 \times 10^9$ (which corresponds to $\Omega_b < 0.08$ $(\tau_H/2 \times 10^{10}\ \text{yr})^2$ and permits $N_\nu = 4$ for $\mathcal{S} > 6 \times 10^9 (\Omega_b < 0.03(\tau_H/2\text{x}10^{10}\text{yr})^2)$. Lower densities are needed, for a given N_ν and expansion rate, if the primordial fraction of ^{4}He is less than (say) 23% rather than 25%. However, *very* low values of Ω_b lead to another inconsistency: the abundance of ^{3}He and D, intermediate products in ^{4}He synthesis, exceed what is observed[3]. These isotopes are both produced in the big bang; D can be burnt into ^{3}He in stars, but the primordial abundance of (^{3}He + D) is unlikely to have exceeded the presently-observed value of 8×10^{-5}. This constrains Ω_b to be $\gtrsim 0.05\ (\tau_H/2 \times 10^{10}\ \text{yr})^2$ (i.e. $\mathcal{S} < 3 \times 10^9$). On the other hand, if the observed D is a relic of the big bang (as is generally believed, owing to the failure of astrophysicists to come up with a plausible alternative way of making it), then Ω_b must be *less* than about $0.09\ (\tau_H/2 \times 10^{10}\ \text{yr})^2$.

These arguments, based on primordial nucleosynthesis, suggest that $N_\nu \lesssim 4$, *and* that Ω_b is of order $0.1\ (\tau_H/2 \times 10^{10}\ \text{yr})^2$. Higher values of N_ν could be reconciled only with more contrived and inhomogeneous models. If the lepton number for ν_e and $\bar{\nu}_e$ were non-zero, then the neutron-proton equilibrium ratio would be shifted, thereby affecting ^{4}He production. It is possible in principle to compensate for *any* speed-up factor in this way and thereby permit a higher value of N_ν or a higher Ω_b. However, in order to make any difference, neutrino lepton number must be of order the photon number - i.e. $\sim \mathcal{S}$ times larger than the baryon number.

3.5 Neutrino masses

A straightforward calculation shows that, if neutrinos have no rest mass, the present density, for each two-component species, is $n_\nu = 110(T_\gamma/2.7\text{K})^3\text{cm}^{-3}$.

This conclusion still holds for non-zero masses, provided that $m_\nu c^2$ is far below the thermal energy ($\sim$ 10 MeV) at which neutrinos decoupled from other species and that the neutrinos are stable for a time τ_H. Comparison with the baryon density of $\sim 3 \times 10^{-6} \Omega_b \ (\tau_H/2 \times 10^{10}\ \mathrm{yr})^{-2}$ shows that neutrinos outnumber baryons by such a big factor ($\sim \mathcal{S}$) that they can be dynamically dominant over baryons even if their masses are only a few eV. In fact a single species of neutrino would yield a contribution to Ω of

$$\Omega_\nu = 0.04 (m_\nu)_{eV} \ \ (\tau_H/2 \times 10^{10}\ \mathrm{yr})^2. \tag{9}$$

The entire range 100 eV - 3GeV is incompatible with the hot big bang model[7]. (for $m_\nu \gtrsim 3$ GeV, the rest mass term in the Boltzmann factor would kill off most of the neutrinos before they decouple; the number surviving would be $\lesssim n_b$). If a neutrino in this mass range *were* discovered, it would show that one cannot extrapolate the hot big bang back to kT $\gtrsim$ 10 MeV, and that most of the photons must have been generated at later times.Analogous limits can be set on the rest masses of possible *right*-handed neutrinos[8]. If these neutrinos interact even more weakly than left-handed neutrinos, they would have decoupled at an earlier stage in the hot big bang. Other species that annihilate *after* the right-handed neutrinos decouple would enhance the number of left-handed neutrinos but not of right-handed. The mass limits on the latter are correspondingly less stringent.

Neutrinos with a mass of a few eV 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 remained homogeneous, their velocities would now be $\sim 200\ (m_\nu)^{-1}_{eV}$ km s^{-1}. They would be influenced even by the weak ($\sim 10^{-5} c^2$) gravitational potential fluctuations of galaxies and clusters. I shall return to discuss neutrino masses later, in the context of "unseen" mass.

(I might note parenthetically that neutrino rest masses of $\lesssim$ 1 eV would have no important consequences for cosmology or large-scale astronomy. However, if neutrinos have very small rest masses and, in consequence, "oscillate", this may have detailed consequences in (e.g.) supernova explosions; even a mass of 10^{-6}eV would permit oscillations of MeV-neutrinos over a path length $\lesssim 10^{13}$ cm, and therefore affect the results of solar neutrino experiments.)

3.6 Unstable neutrinos

A new set of considerations apply if neutrinos are unstable. If the lifetimes are $\ll \tau_H$, all primordial neutrinos would have decayed long ago. For lifetimes $< 10^4$ sec, the decays would occur so early that the resultant energy would have been thermalised, leaving no trace except for an increased $\mathcal{S}$, compared to its value before the decays. (There are, however, other astrophysical constraints on lifetimes $\lesssim 10^4$ sec, from stellar evolution and supernova theory[9].) If the decay time is longer than 10^4 sec, there would be residual distortions in the microwave background.

If the lifetimes are $\gtrsim \tau_H$, the decay rate per comoving volume will have been essentially constant, and the most conspicuous effect would be a photon background due to decays at recent epochs.[10)12)] For masses 10-100 eV, the photons would be in the ultraviolet. The contribution of the ultraviolet background to Ω is $\lesssim 10^{-8}$; this means that lifetimes between τ_H and 10^{24} sec can be excluded (for $m_\nu > 10$ eV). An even more sensitive limit can be set by considering indirect effects of ultraviolet photons on the intergalactic medium,[12)] or (if the resultant photons have energies < 13.6 eV) by sensitive observations of clusters of galaxies in which the neutrino density may be enhanced above its mean value.

3.7 Inferences if we extrapolate back to $t \simeq 1$ sec

The "hot big bang" is not yet a firm "dogma" , in that there exist alternative (though somewhat contrived) models that are compatible with the sparse relevant data; but it is much more plausible than any specific alternative. If the "hot big bang" *is* accepted, the foregoing discussion tells us we can infer the following:

(i) There are unlikely to be more than 4 species of 2-component neutrinos.

(ii) No (left-handed) neutrinos have masses in the range 100 eV-3 GeV.

(iii) Some astrophysical considerations actually *favour* a neutrino mass $\sim$ 10 eV. Neutrinos, and not baryons, could then be gravitationally dominant in large systems of galaxies in the expanding Universe.

(iv) There are astrophysical constraints on neutrino lifetimes if they are unstable.

While there are "escape clauses" which render these inferences less than compelling, they can validly be cited as relevant evidence by particle physicists. These inferences depend on extrapolating back to $\lesssim 10^{-2}$ sec ($kT \gtrsim 10$ MeV), but the agreement between the predicted primordial nucleosynthesis and the observed helium and deuterium abundances give us some confidence in the validity of this.

4. BARYON PRODUCTION

The success of the hot big bang model in relating the origin of the light elements (^{4}He, ^{3}He, D) to processes occurring at $t \simeq 1$ sec has emboldened some physicists and astrophysicists to extrapolate back to still earlier times, when the physics is more uncertain and more exotic. Such extrapolation must stop at the $t_{Planck} = (G\hbar/c^5)^{\frac{1}{2}} \simeq 5 \times 10^{-44}$ sec, when $kT \simeq 10^{19}$ GeV and quantum gravity effects are crucial. However, if the expansion had indeed followed a Friedmann model (with $p = \frac{1}{3}\rho c^2$) ever since the threshold of "classical" cosmology, the temperature-time relation (7), extrapolated to the earliest stages, implies

$$kT \simeq 10^{19}(t/10^{-44}\,\text{sec})^{-\frac{1}{2}}\ \text{GeV} \tag{10}$$

For the first microsecond, kT exceeds a GeV; and during these initial stages the particle energies sweep down through the entire range of interest to theoretical high energy physicists - including, of course, the ultra-high energies unattainable by any feasible terrestrial accelerator (see Figure 2).

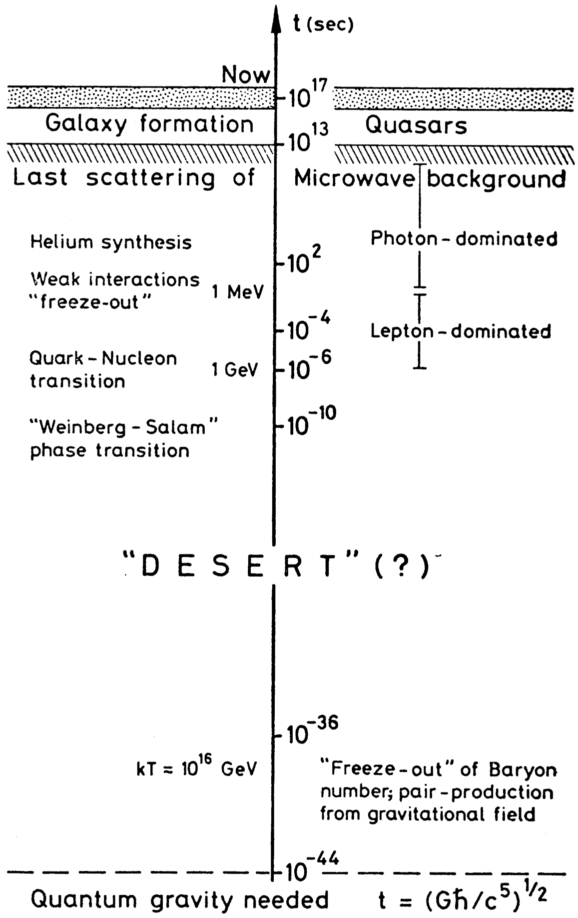

Figure 2. This diagram illustrates, in terms of logarithmic time, various key stages in the expansion of a standard Big Bang model. 60 "decades" separate us from the Planck time. Observations of individual galaxies and quasars permit us to probe only the last decade (stippled region of diagram); the last scattering of the microwave background may have occurred when the Universe had only $\sim 10^{-4}$ of its present age: primordial nucleosynthesis yields evidence on physical conditions when $t \simeq 1$ sec. Ideas that relate baryon production to the consequences of GUT schemes involve extrapolating back in time by a further 36 orders of magnitude!

When t $\lesssim 10^{-36}$ sec the particle energies exceed 10^{15} GeV, the characteristic mass of the X-boson hypothesised in grand unified theories (GUTs). The consequences of baryon non-conservation may then be crucial : indeed many authors have raised the exciting possibility that the baryon content of the Universe - i.e. the value of the parameter $\mathcal{S}$ - may have been imprinted at this era (for reviews see refs 13 or 14). Provided that C and CP violation occurs, and provided also that the relevant reactions are slow enough relative to the expansion rate to allow non-equilibrium effects to build up, the Universe can, as it cools below 10^{15} GeV, acquire an excess of baryons over antibaryons which is related to the CP-violation parameter. (The *non-equilibrium* requirement is essential; just as at the much later nucleosynthesis epoch the rapid expansion prevents everything from being transmuted into iron.) Detailed computations show that several GUT schemes lead to an asymmetry of 10^{-9}; this is the value which would yield $\mathcal{S} \simeq 10^{9}$, after the baryon-anti-baryon pairs annihilate when kT falls below 1 GeV. It would be remarkable if GUT theories did indeed account for the baryon content of the Universe without needing to impose it as an initial condition. This work is not yet on the same footing as the nucleosynthesis of helium and deuterium. It is perhaps at the same level as nucleosynthesis was in the pioneering 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 vindicate an extrapolation in one bound, based on a Friedmann model, right back to the threshold of classical 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 (see Figure 2). It would also place constraints on dissipative processes arising from viscosity, phase transitions, black hole evaporation etc., which might occur as the Universe cools through the "desert" between 10^{15} and 100 GeV. Although these ideas are still speculative, one could perhaps be more assertive and claim the "prediction" of the photon/baryon ratio as one of the few likely observational tests of GUTs.

5. OTHER POSSIBLE CONSEQUENCES OF GUTS, ETC.

A remarkable but quite essential feature of our Universe is its *scale* - the fact that it is still expanding after $\sim 10^{60}$ t_{Planck}. This timespan, and the corresponding vast cosmological lengthscales, are a prerequisite for the formation and evolution of stars and galaxies. Perhaps equally surprising is the *isotropy* - the fact that the Universe has not availed itself of the other degrees of freedom open to a more "chaotic" geometry.

At first sight it might seem odd to seek an explanation for the global homogeneity in terms of the physics at very early times. The mass encompassed within the Friedmann horizon (proportional to t) was then very small, so as one extrapolates back one seems further and further away from being able to offer a causal explanation. But there are two reasons why this may be wrong. The first stems from the possibly drastic consequences of phase transitions at the grand unified era; the second, still more speculative, involves quantum processes at Planck time.

In grand unified theories, there would be a phase transition at a temperature of order 10^{15} GeV; there may also be a phase transition at the electro-weak unification energy of $\sim 10^2$ GeV.[15)16)] One of the phases will in general have a lower energy density than the other. The differing energy densities of the two vacua would be equivalent to different values of the cosmical constant (or Λ-term).

The Λ-term is now very small (corresponding to a mass-energy density of $\lesssim 10^{-30}$ gm cm^{-3}; or, in other units, to $\lesssim 10^{-44}$ GeV^4). One would expect the *change* in the effective Λ-term associated with symmetry-breaking to be of order the energy density at the "grand unified temperature, i.e. $\sim 10^{60}$ GeV^4. Why things should be "fine tuned" to a precision of $\lesssim 1$ part in 10^{100}, so that the *post*-transition Λ-term is so small, is still a mystery.

If the dynamics of the very early Universe are effectively dominated by a Λ-term, then R(t) inflates exponentially (the "de Sitter cosmology"). The idea of an exponential growth phase has some appealing consequences.[15)16)] In particular, it suggests an answer to the problem of why the Universe is so large - why (in more technical language) the curvature radius of the hypersurfaces of homogeneity in the Robertson-Walker metric is $\gtrsim 10^{30}$ times the comoving scale of the horizon at the Planck time. If R grows by (say) e^{100} during this phase, the large scale of the Universe, or its "flatness", could be accounted for.

The main difficulty of this scheme centres on whether the Universe can expand by a gigantic factor and still, afterwards, achieve a "graceful exit" from its exponential growth - whether there can be a transition to a Friedmann phase where $R \propto t^{\frac{1}{2}}$. (For a discussion of this, see papers in ref 17). Note that, for this scheme to work, the heat released during the phase transition must raise kT above $m_X c^2$ so that baryon synthesis can occur after the exponential phase. Optimists might hope that this model could not only account for the Universe's overall homogeneity, but also generate fluctuations (needed to give galaxies and clusters) from microscopic effects. Much effort is being expended on the calculation of fluctuations, the net result being that the predicted fluctuations are generally too large, though their amplitude depends sensitively on the form of the effective potential of the Higgs scalar field responsible for the symmetry breaking.[17)] Despite the uncertainties, it is a measure of the progress made in recent years that the primordial fluctuation amplitude - the amplitude of the "ripples" superposed on the hypersurfaces of homogeneity in the Robertson-Walker metric - is regarded as calculable rather than "ad hoc".

It would indeed by remarkable if the entire observable universe could have sprung from a microscopic fluctuation. Since no non-zero conserved quantities are involved (if baryons are not conserved), this almost *ex nihilo* origin for the Universe is at least physically consistent

The production of *magnetic monopoles* is expected whenever the symmetry of an initial non-Abelian group is spontaneously broken, and the unbroken symmetry group contains the Abelian U(1) in its decomposition.[18][19] If this leads to even one monopole per horizon volume, and the monopole mass is $\sim 10^{15}$ GeV, then the resultant monopoles would exceed by 10^{10} the number needed to yield[20] $\Omega = 1$; the discrepancy is even greater when one takes into consideration the stringent constraints on the monopole density implied by the existence and persistence of large-scale cosmic magnetic fields. Some mechanism for suppressing or diluting the monopole density far below this naive estimate is obviously needed - the "inflationary" phase is one such mechanism.

Symmetry-breaking may lead to the formation of "domain walls" or "strings". The former can probably be ruled out, on the grounds that the effective mass-energy of the walls would be incompatible with present-day cosmological constraints. One-dimensional singularities - "strings" - cannot however be excluded.[21] It is not clear whether the topology of the strings would allow them to contract and disappear, or whether on the other hand they are stretched as the Universe expands. In the latter case, they could contribute $\Omega \approx 10^{-4}$ even at recent epochs, and might be the "seed" fluctuations that trigger galaxy formation.[22][23]

If phase transitions do not offer an explanation for the homogeneity and/or fluctuations, we must go further back still to the Planck time, when the classical concept of the particle horizon is transcended. Processes at the Planck time may allow us also to understand the isotropy. During the fifty years which have passed since the pioneering researches of De Sitter, Friedmann and Lemaitre, much theoretical attention has been focussed on devising other more complex and more general cosmological models - models with rotation, shear, or gross inhomogeneity. The present indications allow us to trace the history of our Universe back through 54 "decades" of logarithmic time, and over this entire period its evolution has proceeded in amazingly precise accordance with a Friedmann model - the most straightforward and mathematically calculable of all. But we have no idea why this should be so.

6. THE DENSITY OF THE UNIVERSE : "UNSEEN" MASS

Having inferred something about how our expanding Universe began, the next question concerns its future and eventual fate. Will the Universe expand for ever and the galaxies fade and disperse? Or, on the other hand, will it recollapse? The traditional way to try to answer this question has involved extending the work of Hubble to the study of galaxies at greater distances, or (equivalently) to greater look-back times. The redshift of a distant galaxy tells us its speed as the light set out on its journey towards us. So in principle one can compare the expansion of the Universe at early times with the present rate, and thereby infer how much it is decelerating. In practice this type of work is bedevilled by various observational difficulties and uncertain corrections, which have as yet prevented it from yielding a reliable answer.

But if we believe that the dynamics of the Universe are governed by Einstein's equations (and if the *present* Λ-term is zero), this question can be attacked in another way. It is then equivalent to the issue of the *mean density* : is this above or below the critical density ρ_{crit} defined in (5)? It is convenient to define a "density parameter" $\Omega = \bar{\rho}/\rho_{crit}$. We then would like to know whether $\Omega > 1$ (implying eventual collapse to a "big crunch") or whether $\Omega \lesssim 1$ (implying perpetual expansion).

The luminous stars in galaxies, together with the interstellar and intergalactic gas that can be directly observed, contribute $\Omega \approx 10^{-2}$, mainly in the form of ordinary baryons. But there are dynamical indications of some "unseen mass", which 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) 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 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 galaxy groups on scales (1 - 2) Mpc. It is interesting to ask whether the data *permit* $\Omega \gtrsim 1$. The short answer is that this *may* be consistent with the data provided that, on scales where the virial theorem can be reliably applied, the unseen mass is less "clumped" than the luminous mass.

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 $\frac{M}{L}$ which is $\lesssim 10$ times the solar mass-to-light ratio, if the unseen mass contributes a density parameter Ω it must have $\langle M/L \rangle$ exceeding $1100\ (\tau_H/2\times10^{10}\text{yr})\,\Omega$ solar units.

It needs no great ingenuity to invent possible forms for "unseen" mass in the Universe, baryonic or non-baryonic, among which are the following :

6.1 Baryonic forms

(a) Low-mass stars ("Jupiter") of $\lesssim 0.1\ M_\odot$. These would be very faint because they would be below the mass needed for hydrogen-burning (cf discussion of equations (1) - (4)).

(b) Remnants of very massive stars. 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 ^4He 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 (VMOs) of $10^3\ M_\odot$ is compatible with the nucleosynthesis constraints. A further consideration favouring these high masses is that VMOs 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[24].

The upper limit of $10^6\ M_\odot$ comes from noting that objects above this mass would have various detectable effects - e.g. dynamical friction or accretion - and can be ruled out as major constituents of galactic halos[25]. Higher masses could contribute to Ω if they were in intergalactic space, though systematic searches for gravitational lensing effects could soon reveal them, or allow us to exclude masses above $10^6\ M_\odot$. The probability of observing lensing along a line of sight to $z \simeq 1$ due to a population of compact objects is of order their contribution to Ω. The characteristic angular size of the image is, however, $\sim (M/10^6\ M_\odot)^{\frac{1}{2}}$ milliarc seconds.

6.2 Non-baryonic "unseen" mass

I have already mentioned (equation (9)) that primordial neutrinos would be dynamically important if their masses were even as little as a few e.v. If the three (or more) types of neutrinos have different masses, then the heaviest will be gravitationally dominant.

Physicists have other particles "in reserve" - right handed neutrinos, photinos or gravitinos, for instance - which could (if they existed) have been in thermal equilibrium with other species at very early times, and therefore contribute to Ω in an analogous way (see Dolgov & Zeldovich [26] for a review). 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 latter annihilations would then boost the neutrinos but not the still more weakly coupled "...inos". For gravitinos, we may have $(n/n_\gamma) \lesssim 10^{-2}$.

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 $\sim$ 10 ev, gravitinos, monopoles or axions are just some of many candidates for the "unseen" mass in the Universe : as far as

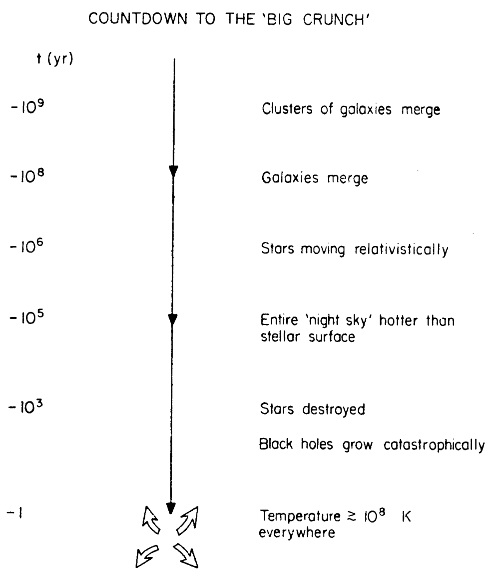

Figure 3. Stages in the contraction of a recollapsing Universe (from ref. 27). Times are measured backwards from the 'crunch'.

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 pre-galactic stars). There is no lack of candidates for "unseen" mass, either baryonic or non-baryonic.

Primordial nucleosynthesis of an adequate amount of D occurs in the simplest big bang models only if the baryonic contribution to Ω is < 0.1. This is therefore a reason for perhaps favouring a major non-baryonic contribution to Ω if the Universe has close to (or above) the critical density.

7. THE FAR FUTURE

The present evidence is that $\Omega \gtrsim 0.1$; we do not know whether $\Omega > 1$, though this cannot be ruled out. It is therefore interesting to speculate about the two very different "eschatologies" that could befall the Universe.

What will happen if $\Omega > 1$ and the Universe recollapses? The redshifts of distant galaxies would be replaced by blueshifts, and galaxies would crowd together again. Space is already becoming more and more 'punctured' as isolated regions - dead stars and galactic nuclei - collapse to black holes; but this would then be just a precursor of a Universal collapse, a 'big crunch' that engulfs everything (Figure 3). Some key stages in the 'count-down' are shown in the diagram. The final state would be a fireball like that which initiated the Universal expansion, though it would be somewhat more lumpy and unsynchronized.

THE FAR FUTURE OF AN EVER-EXPANDING UNIVERSE

10^{14} yr	Ordinary stellar activity completed
10^{17} yr	Significant dynamical relaxation in galaxies
10^{20} yr	Gravitational radiation effects in galaxies
[10^{31} - 10^{36} yr	Proton decay]
$10^{64}(m/m_\odot)^3$ yr	Quantum evaporation of black holes
10^{1600} yr	White dwarfs → neutron stars*
$10^{10^{26}}$ - $10^{10^{76}}$	Neutron stars undergo quantum* tunnelling to black holes, which then 'quickly' evaporate

*If proton decay does not occur

Figure 4. Timescales for an ever-expanding Universe (as discussed fully in reference 28).

But if, on the other hand, the Universe expands for ever to a "heat death", there is time for even the slowest processes to attain a terminal equilibrium. Some key stages here are shown in Figure 4.

These two outcomes are very different. But the initial conditions which could have led to anything like our present Universe are actually *very restrictive*, compared to the range of possibilities that could have been set up : the fact that the Universe has gone on for 10^{10} years without recollapsing, but is expanding sufficiently slowly to permit galaxies to form, in itself requires 'fine tuning' (see Figure 5 and caption). Moreover, our Universe looks even more special if we recall that the whole family illustrated in Figure 5 are just the Friedman models which are isotropic, their dynamics being described by the time-dependence of a single scale-factor R. Further free parameters can be introduced by admitting anisotropy, or large amplitude inhomogeneity : but would generally make the Universe less conducive to galaxy formation. The mystery is why the Universe is so homogeneous - why it has not availed itself of the other macroscopic degrees of freedom - and why the initial kinetic energy and gravitational energy were so closely balanced as its present properties indicate. This is another way of expressing the problems of the scale and flatness of the Universe - a problem which the idea of exponential inflation at the GUT era may be able to solve.

8. SCALES OF ASTROPHYSICAL STRUCTURE

As a final "unifying" comment, I would like to present Figure 6, which shows in order-of-magnitude how the scales of microphysics and astronomy

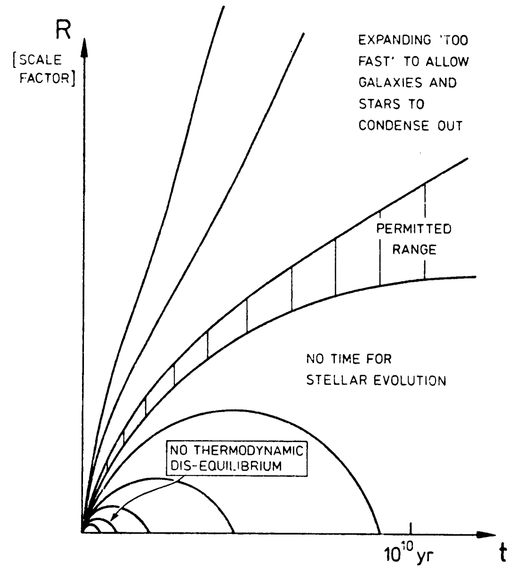

Figure 5. In the isotropic Friedmann models, the scale factor R(t) evolves essentially in the same manner as in 'Newtonian cosmology': the Universe either expands for ever (positive total energy) or recollapses. We know that our Universe is still expanding after 10^{10} yr; if it had recollapsed sooner, there would have been no time for stars to evolve; if it collapsed after less than $\sim 10^6$ yr, it would have remained opaque and close to thermal equilibrium throughout its life. The expansion rate cannot, however, be *too* much faster than 'parabolic': otherwise gravitational instability would have been ineffective and bound systems would not have condensed out. (This is equivalent to the statement that the present density is not orders-of-magnitude below the critical density.) There is therefore a sense in which the dynamics of the early Universe must have been 'finely tuned': in Newtonian terms, the fractional difference between the initial potential and kinetic energies of any spherical region must have been very small: in terms of the Robertson-Walker metric, the curvature of the hypersurfaces of homogeneity must have been very small compared to the particle horizon at early times: they must have been almost "flat". Our Universe seems ever more 'special' when we realize that the introduction of anisotropy, large-scale inhomogeneity, etc., offers many more degrees of freedom than the Friedmann models. Recent ideas on an exponential "inflationary" phase before the symmetry breaking transition at the GUT era suggest a possible line of explanation for the scale and "flatness".

are related by various powers of the "large number" α_G^{-1}, and the ordinary fine structure constant . (This topic is more fully discussed in ref. 29.) In this mass-radius diagram we see the "black hole line" $R = 2GM/c^2$ and the "quantum line" when $R = h/Mc$. These lines meet at the "Planck scale" $R_{Planck} = (G/hc^3)^{\frac{1}{2}} \simeq 10^{-33}$ cm; $M_{Planck} = (G/hc)^{-\frac{1}{2}} \simeq 10^{-5}$ gm. (It is because these dimensions lie so far from either laboratory or astrophysical relevance that our progress is not primarily impeded by the absence of a theory of quantum gravity!) The region of stellar masses ($M \simeq \alpha_G^{-3/2}\, m_p$) is shown (cf. equations (1)-(4)). A mass $\alpha_G^{-1}\, m_p$ is of interest, being the mass of a black hole which is about the size of a proton. Such a hole will evaporate in a time $\alpha_G^{-1}(h/m_p c^2)$ which turns out to be the rough lifetime of a typical star. A Friedmann cosmology with a density close to the critical density, as seems to be the case for our Universe (see Figure 5), contains $\sim \alpha_G^{-2}$ particles within a "Hubble sphere" when its age is $\sim \alpha_G^{-1}\, (h/m_p c^2)$. This

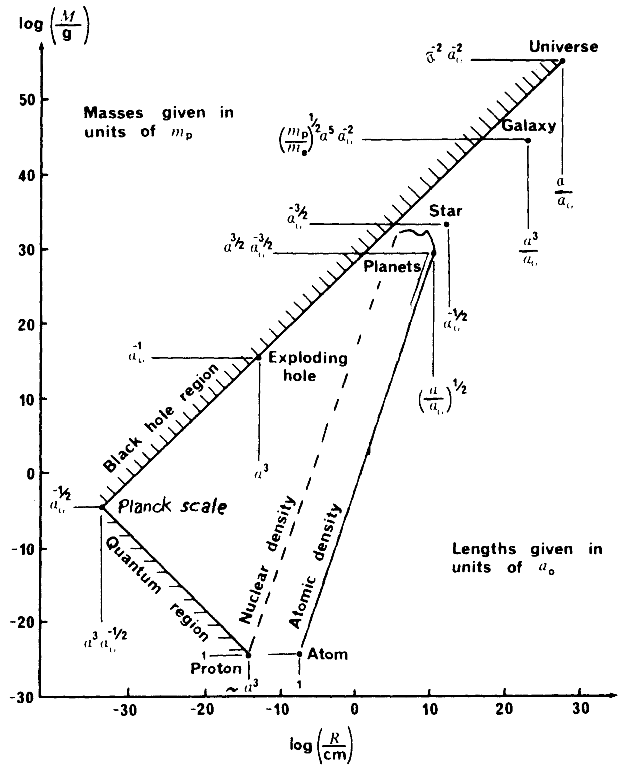

Figure 6. The mass and length scales of various natural structures expressed in terms of the electromagnetic and gravitational fine structure constants, α and α_G. Some of these scales also depend on the electron-to-proton mass ratio, but this we have eliminated by using $m_e/m_p \sim 10\alpha^2$. All these scales can be deduced directly from known physics except for the mass and length scale of the Universe, which depends on the age of the Universe being α_G^{-1} times the electron timescale $\hbar/m_e c^2$. Also shown are the atomic density line, the black hole density line and the 'quantum line' corresponding to the Compton wavelength. Most characteristic scales depend on simple powers of α_G. This diagram is reproduced from ref. 29.

also is plotted in Figure 6. The scales plotted for galaxies agree with observations, but the theoretical basis for them in terms of basic physics is still tentative. The wide span of so many orders of magnitude in Figure 6 - the fact that complex structures can become huge compared to atomic dimensions before their self-gravity becomes significant - is a consequence of α_G^{-1} being as large as $\sim 10^{38}$, a fact which theorists will perhaps eventually explain.

REFERENCES

1) A.A. Penzias and R.W. Wilson, Astrophys. J., 142, 419 (1965).

2) R.H. Dicke, P.J.E. Peebles, P.G. Roll and D.T. Wilkinson, Astrophys. J., 142, 414 (1965).

3) R.V. Wagoner, W.A. Fowler and F. Hoyle, Astrophys. J., 143, 3 (1967).

4) K.A. Olive, D.N. Schramm, G. Steigman, M.S. Turner and J. Yang, Astrophys. J., 246, 557 (1981).

5) V.F. Schwartzman, J.E.T.P. Lett., 9, 184 (1969).

6) B.E.J. Pagel, Phil. Trans. R. Soc., A307, 19 (1982).

7) J.E. Gunn, B.W. Lee, I. Lerche, D.N. Schramm and G. Steigman, Astrophys. J., 223, 1015 (1978).

8) K.A. Olive and M.S. Turner, Phys. Rev., D25, 213 (1982).

9) S.W. Falk and D.N. Schramm, Phys. Lett., 79B, 511 (1978).

10) R. Kimble, S. Bowyer and P. Jacobsen, Phys. Rev. Lett., 46, 80 (1981).

11) A. De Rujula and S.I. Glashow, Phys. Rev. Lett., 45, 942 (1980).

12) A.L. Melott and D.W. Sciama, Phys. Rev. Lett., 46, 1369 (1981).

13) E. Kolb and S. Wolfram, Nucl. Phys., B172, 224 (1980).

14) S. Weinberg in "Astrophysical Cosmology" ed. H.A. Brück, G.V. Coyne and M.S Longair (Vatican Publications) p. 503 (1982).

15) A.D. Linde, Rep. Prog. Phys., 42, 389 (1979).

16) A. Guth, Phys. Rev. D., 23, 347 (1981).

17) Papers by many authors in "The Very Early Universe" ed. G. Gibbons, S. Hawking and S. Siklos (CUP) (1983).

18) A. Polyakov, J.E.T.P. Letters, 20, 194 (1974).

19) G. t'Hooft, Nucl. Phys. B79, 276 (1974).

20) J.R. Prestki'l, Phys. Rev. Lett., 43, 1365 (1979).

21) T.W.B. Kibble, Physics Reports, 67, 183 (1980)

22) Y.B. Zeldovich, Mon. Not. R. astr. Soc., 192, 246 (1980).

23) A. Vilenkin, Phys. Rev. D23, 852 (1981).

24) B.J. Carr, J.R. Bond and W.D. Arnett, Astrophys. J. (in press)

25) B.J. Carr, Comments on Astrophys., 7, 161 (1978).

26) A.D. Dolgov and Y.B. Zeldovich, Rev. Mod. Phys., 43, 1 (1981).

27) M.J. Rees, Observatory, 89, 193 (1969).

28) F.J. Dyson, Rev. Mod. Phys., 51, 447 (1979).

29) B.J. Carr and M.J. Rees, Nature, 278, 605 (1979).