Cosmology and elementary particles

A. D. Dolgov

Institute of Theoretical and Experimental Physics, Moscow-259, USSR

Ya. B. Zeldovich

Institute of Applied Mathematics of Academy of Sciences of the USSR, Moscow, USSR

The restrictions on elementary particle properties which can be derived from cosmological and astrophysical data are considered. The inverse relations between micro- and macrophysics are also discussed, in particular the origin of the baryon asymmetry of the universe.

CONTENTS

I. Introduction 1
A. A brief cosmological review 2
B. Organization of the paper 6
II. Particle Concentrations in the Hot Universe 7
III. Bounds on the Number of Light Particle Species 9
IV. Light Particle Masses 11
V. Quenching and Contemporary Concentrations of Heavy Stable Particles 12
VI. Heavy Lepton Masses 13
VII. Unstable Leptons. Limits on Their Mass and/or Lifetime 14
A. The total energy density in the universe 14
B. The spectrum of cosmic electromagnetic radiation 15
C. Primordial nucleosynthesis 16
D. Stellar evolution 17
VIII. Astrophysical Bounds on Electromagnetic Parameters of Neutrinos 18
IX. Neutrino Degeneracy 19
X. Astrophysical Consequences of the Existence of a Heavy Neutral Lepton 20
XI. Some Other Particles (Higgs Mesons, Axions, Magnetic Monopoles, etc.) 21
A. Et cetera 21
B. Higgs mesons, axions 22
C. Magnetic monopoles 23
D. θ particles 25
XII. The Photon Mass 25
XIII. Nonconservation of Electric Charge. Unstable Electrons 26
XIV. Cosmological Consequences of Spontaneous Symmetry Breaking in Elementary Particle Physics 27
XV. Baryon Asymmetry of the Universe 28
A. The cold universe and entropy generation 29
B. Spatial separation of matter and antimatter 29
C. Some general theorems on particle–antiparticle asymmetry 29
D. Concealment of baryon number in black holes 33
E. Baryon asymmetry of the universe in a $C(CP)$-violating, B-nonconserving theory of elementary particles 35
F. Baryon asymmetry due to CPT violation 37
XVI. Vacuum Polarization by Curved Space-Time 37
Acknowledgments 38
References 38

I. INTRODUCTION

Astronomy nowadays tends to have closer and closer connections with physics. Astronomy by itself is a branch of science studying things which really have happened, are happening, or will happen in Nature.

Physics deals with the fundamental laws which govern different phenomena in Nature, and in particular the phenomena which indeed happen under natural conditions according to these laws. Apart from this, physics describes and predicts what will happen under special conditions in a controlled physical experiment. Astronomy adopts possible scenarios from physics and fits them to meager and often ambiguous observational data.

It is said that the paintings by the famous English artist Turner changed the weather on the coast of England; they made the sunsets more magnificent. We have witnessed how cosmic masers were discovered only after the development of laser technology in the laboratory. The increasing belief in general relativity and the detailed investigation of nuclear reactions as the power supply of the stars led theoreticians to the conclusion that neutron stars and black holes must exist. The discoveries of radio and x-ray pulsars confirmed the existence of neutron stars. Some compact x-ray sources are possible candidates for the high title of black holes. These are only some examples of physics helping astronomy to comprehend the observed phenomena, and the list could be enlarged.

But just the opposite cases are of special value for astronomers. These are the cases when astronomy teaches physicists and puts at their disposal specific data which are difficult or even impossible to obtain under laboratory conditions. One well known instance of this is the measurement of the speed of light. Roemer determined it in 1675 from the delay of the Jupiter satellite eclipse. It could not be measured in laboratories until centuries later. Another example is the discovery of the law of gravity by Newton. Celestial mechanics permitted it to be established in 1686. Only in 1798 was Cavendish able to measure the gravitational constant in the laboratory.

Does physics continue to be enriched by astronomy? Can astronomy, by virtue of its gigantic scale, get information inaccessible in a modern superequipped laboratory? Is it possible, in particular, to obtain, by means of astronomy, new data on elementary particles? This is the question which is considered in this paper, and the answer is positive.

We begin with a review of the basic facts known about the universe, which will be used in what follows. The consequences of these facts for elementary particles are touched here only briefly, as they will be discussed in detail in later sections.

The celestial mechanics of the solar system confirms with magnificient precision the theory of general relativity. The data leave almost no room for modifications of the theory. Therefore, the existence of gravitational waves (which is predicted by general relativity) is indirectly confirmed. Going a step further, quantum mechanics tells us that gravitons (spin-2 massless bosons) must exist. We may conclude that up to energies of the order of $m_p \simeq 10^{19}$ GeV the theory is established, with the possible uncertainty with regards to the cosmological term.

Our failure to date to observe gravitational waves directly is somewhat discouraging, but the theory is by no means compromised if one compares the sensitivity of modern measuring equipment with the expected intensity of the sources. Recent observations of a binary pulsar have given evidence in favor of gravitational radiation (Taylor *et al.*, 1979. See also MacCallum, 1979).

Another confirmation of general relativity follows from checking the equivalence principle, i.e., the equality of gravitational and inertial masses. The impressive accuracy of 10^{-12} is obtained here. This means that the equation[1] $E = \Delta m c^2$ is valid not only for strong and electromagnetic interactions but also for weak interactions. The experiments testing the equivalence principle were performed with electrically neutral bodies. They proved that there are no long-range forces (like the electromagnetic one) associated with the particles forming the bodies (Lee and Yang, 1955).

On the other hand, measurements of the Earth's magnetic field and those of Jupiter and of galaxies show no deviations from Maxwellian electrodynamics. This means that the photon mass is zero or extremely small, much smaller than the bounds obtained in laboratory experiments.

A new era in the interrelationship of physics and astronomy began about 20 years ago, due to the development of modern cosmology and, especially, due to experimental confirmations of the hot universe model. We shall now briefly present the basic facts of big bang cosmology.

A. A brief cosmological review

Many physicists owe their aquaintance with modern cosmology to the book *The First Three Minutes* by S. Weinberg (1977). They know that the universe is expanding in accordance with Hubble's law, i.e., $u = Hr$. The contemporary value of the Hubble constant[2] H (constant means no variation in space; that is, H does not depend on the value and direction of r but depends on time) is 55 km sec^{-1} Mpc.$^{-1}$ Here Mpc is the abbreviation for megaparsec, 1 Mpc $= 10^6$ $pc = 3.085 \times 10^{24}$ cm, so that $H = (1.8 \times 10^{10}\ \mathrm{yr})^{-1}$.

On large scales the universe is with high precision uniform, and its expansion is isotropic (i.e., equivalent in all directions in space). The largest distance we have information about is of order $c/H \simeq 5000$ Mpc $\simeq 1.5 \times 10^{28}$ cm. On such a scale deviations from the uniform and isotropic picture are less than 10^{-3}–10^{-4} in relative units. At an earlier time, however, there could have been larger perturbations.

The theory of the expanding universe connects the age of the universe t_0 with the present expansion rate H and the average matter density ρ. The lower the density of matter, the smaller the velocity variation during the expansion. For $\rho \to 0$ the universe's age $t_0 = R/u = H^{-1}$. With increasing density gravitational deceleration becomes important. This means that in the past the expansion was faster, so that the age t_0 is less:

$$t_0 = H^{-1} f(\Omega) = f(\Omega) 18 \times 10^9\ \mathrm{yr}\,,$$

where $\Omega = \rho/\rho_c$ and the function f is smaller than unity. The critical value of the matter density is defined by the equation $\rho_c = 3H^2/8\pi G = 6 \times 10^{-30}$ g cm^{-3}. If $\Omega = 1$ (i.e., $\rho = \rho_c$), $f(1) = 2/3$. For $\Omega \gg 1$ the function f tends to $1/\sqrt{\Omega}$ so that the H-independent upper bound on the age of the universe can be obtained, $t_0 < (8\pi G\rho/3)^{1/2}$. These formulas are valid in nonrelativistic cases, i.e., for $p \ll \varepsilon = \rho$. In the extreme relativistic limit where $p = \varepsilon/3$, the simple expression $f = (1 + \sqrt{\Omega})^{-1}$ can be found. The lower bound on the age of the universe is given by the age of the solar system and the Earth ($t_0 \geqslant 4.5 \times 10^9$ yr). Nuclear dating methods and in particular the abundance ratio of uranium 235 to uranium 238 require a time interval $t = 8 \times 10^9$ yr between today and the epoch of nucleosynthesis (nowadays $U^{235}:U^{238} = 7 \times 10^{-3}$ and, in accordance with the element formation theory, $U^{235} : U^{238} \gtrsim 1$ at the moment of production). The use of the Re-Os chronometer gives $t_0 = (11-18) \times 10^9$ yr (Hainebach and Schramm, 1976). Astronomical observations, together with the stellar evolution theory, give evidence that the age of some stars is as great as $(14-16) \times 10^9$ yr (Demarque and McClure, 1977). The situation was summarized by Tammann *et al.* (1980), who claim that the age of the universe is definitely greater than 12×10^9 yr and probably even greater than 15×10^9 yr. An upper bound for the density of matter compatible with the modern values of H and t_0 is approximately $\rho_{\max} = 10^{-29}$ g cm^{-3}. A more probable value is $\rho \simeq 2 \times 10^{-30}$ g cm^{-3}. There are some indirect data in favor of this result. The average density of matter contained in stars relative to the whole volume among galaxies and clusters of galaxies is about 5×10^{-31} g cm^{-3}.

The upper bound on ρ seems indeed to be reliable. It can be used to restrict the total energy density of forms of matter in the universe which are not observable directly. This idea was probably first applied to elementary particles by Zeldovich and Smorodinsky (1961). They found a bound on the density of particles such as neutrinos and gravitons which cannot be directly detected.

The next well established fact is the existence of the three-degree microwave background radiation. This radiation is highly isotropic, deviating from exact isotropy by less than 3×10^{-4} (in relative units $\Delta T/T$).

[1] In what follows we shall use the natural system of units such that $\hbar = c = k = 1$.

[2] Some authors, even today, claim that $H = (75-100)$ km sec^{-1} Mpc^{-1}. A mistrust of the existence and cosmological origin of the redshift in spectra of remote objects, however, is becoming less and less popular. On the contrary, it seems surprising now that the Hubble expansion was discovered primarily at small scales.

The degree of linear polarization is less than 10^{-3}, and no deviations from the Planck spectrum larger than 20–30% are found. (See, however, the paper by Woody and Richards, 1979). The radiation temperature now is $2.8 \pm .1$ K. The number density of the photons is about 500 cm^{-3}, and their energy density is 6×10^{-13} erg cm^{-3} Since the universe is transparent to the cosmic background radiation now, and in the recent past, the isotropy (relative to the solar system) tells us that this radiation is uniform in space. Because of the negligible interaction of the background photons with matter in the universe, their spectrum and spatial distribution provide information on the remote past of the universe.

The expansion of the universe, together with the blackbody cosmic background, leads to the conclusion that in the past the temperature was higher. The spectrum of the cosmic background radiation agrees with the prediction of the hot universe (big bang) model first formulated by Gamow. Extrapolating the temperature to earlier stages we obtain $T \simeq 10^{10}$ K $\simeq 1$ MeV at a time of the order of 1 sec from "the beginning"; T is about 10 MeV for $t = 0.01$ sec, and so on; approximately, $T(\mathrm{MeV}) \simeq (t\ \mathrm{sec})^{-1/2}$. For a temperature above 0.1 MeV there should be electrons and positrons in thermal equilibrium with the radiation. With rising temperature, thermal equilibrium establishes itself, generally speaking, faster than the rate of temperature variation. (There can be a violation of the equilibrium, however, if the interaction rate falls off with the rise in temperature.) So for large T there is thermal equilibrium, and all particles with $m \lesssim T$ contribute to the plasma in about equal amounts, with gravitons a possible exception. (Thermalization of gravitons was discussed by Kobzarev and Peshkov, 1974.) During the expansion and corresponding cooling of the universe, the number of light particles is conserved, and in first approximation their number density ratio to that of photons is constant. Therefore in big bang cosmology the present number density of a given type neutrino (together with antineutrino) with one helicity state is about 150 cm^{-3}. This result is valid for any sufficiently light neutrino. Today at least two types of neutrinos are known to exist: electronic neutrinos, ν_e (with corresponding antiparticle $\bar{\nu}_e$), and muonic neutrinos, ν_μ (with corresponding $\bar{\nu}_\mu$). Probably there is a third neutrino type, ν_τ and $\bar{\nu}_\tau$, associated with the charged τ lepton, the latter having a mass of about 1800 MeV. With the known number density of neutrinos (N_ν) and the upper bound for the matter density ($\rho_{\max}$) one can obtain an upper limit for the neutrino rest mass, the condition $\sum N_\nu m_\nu < \rho_{\max}$ being used. This argument was first presented in a paper by Gerstein and Zeldovich in 1966 and later by Cowsik and McClelland (1972).

In the framework of big bang cosmology the abundances of the relic quarks were considered. If quarks are not permanently confined there should be at least one stable quark type (quark flavor) which would survive until the present epoch as a free particle. When the temperature was sufficiently high the quark density would have been of the order of that of photons. In the course of cooling down, stable quarks could disappear only through mutual annihilation. The probability of quark-antiquark collisions would decrease rapidly as the universe expanded due to decreasing number density per unit volume. Therefore, a rather high abundance of relic quarks was predicted (Zeldovich, Okun, and Pikelner, 1965). Comparison of this result with the experimental bounds on quark concentration in matter was a strong argument in favor of quark confinement. There have been, however, recent attempts to get a smaller relic quark concentration (see Sec. V) generated by the claim (LaRue *et al.*, 1977, 1979) of free-quark observation. Along the same lines the relic abundances of stable or long-lived particles were later calculated. One example is a hypothetical heavy neutral lepton.

Let us return now to astronomical data and discuss the density of "ordinary matter," i.e., protons, neutrons, and the nuclei built of them. We shall consider an epoch when the temperature is less than 100 MeV so there are no antibaryons in the plasma.

First of all, we note that astronomical observations give no indication that antimatter occurs in an amount comparable to that of matter. The universe seems to be baryonic charge asymmetric. The asymmetry becomes significant when the temperature falls lower than 100 MeV ($t > 10^{-4}$ sec). At higher temperatures the asymmetry exists but it is relatively small and becomes important only after cooling down. If there were a baryonic excess in the primeval plasma the equilibrium density of antibaryons would be negligibly small for $T < 100$ MeV. Moreover, baryon–antibaryon annihilation would occur at an early time when the temperature was higher than 100 MeV. The thermodynamic equilibrium predicts early annihilation. There are no signs of later annihilation. Such signs might include contributions to the spectrum of cosmic rays, distortions in the spectrum of cosmic radio waves, or changes in the chemical content of primary matter (some details about this content are presented below). The combined observational data indicate large charge asymmetry. One has only to get rid of the prejudice that the symmetry of some particle properties (masses, absolute values of charges, and so on) demands an equal number of particles and antiparticles in the universe. Charge conjugation symmetry (C) is violated in nature, as is CP (P is parity). It is generally believed that CPT (T is time reversal, future $\longleftrightarrow$ past) holds exactly. But in a universe subject to evolution with a given "arrow of time," CPT does not prevent matter-antimatter asymmetry.

So, by the time $T < 100$ MeV, we have some protons and neutrons in a lepton-and-photon plasma. At first, owing to the weak interaction, the $p \rightleftarrows n$ transitions are kept in equilibrium (Hayashi, 1950), then at lower temperatures nucleosynthesis proceeds. The earliest nuclei to be produced are the deuterium nuclei, then H^3, He^3, and at last He^4. After that element production practically stops because of the low density of cosmological matter. Consequently, the reactions $\mathrm{He}^3 + \mathrm{He}^4 \to \mathrm{Be}^7$, $3\mathrm{He}^4 \to \mathrm{C}^{12}$, and so on almost do not proceed.

The standard big bang cosmology gives definite predictions about primordial element formation which

TABLE I. Element production in the standard big bang cosmology (Wagoner, 1974).

ρ_B ($T=2.7$ K) g cm^{-3}	$R(D)$	$R(\mathrm{He}^3)$	$R(\mathrm{He}^4)$	$R(\mathrm{Li}^6)$	$R(\mathrm{Li}^7)$	$R(\mathrm{B}^{11})$	$R(A \geq 12)$
7.15×10^{-33}	8.5×10^{-3}	3.6×10^{-4}	0.089	2.6×10^{-11}	2.0×10^{-9}		
1.27×10^{-32}	5.5×10^{-3}	2.8×10^{-4}	0.131	3.7×10^{-11}	3.0×10^{-9}		
2.26×10^{-32}	3.1×10^{-3}	1.9×10^{-4}	0.171	3.6×10^{-11}	2.8×10^{-9}		
4.02×10^{-32}	1.4×10^{-3}	1.1×10^{-4}	0.200	2.3×10^{-11}	1.5×10^{-9}		
7.15×10^{-32}	5.8×10^{-4}	6.7×10^{-5}	0.217	1.1×10^{-11}	5.0×10^{-10}		
1.27×10^{-31}	2.2×10^{-4}	4.3×10^{-5}	0.227	4.5×10^{-12}	2.2×10^{-10}		
2.26×10^{-31}	8.9×10^{-5}	2.8×10^{-5}	0.234	2.0×10^{-12}	3.4×10^{-10}		
4.02×10^{-31}	3.6×10^{-5}	1.8×10^{-5}	0.240		1.2×10^{-9}		
7.15×10^{-31}	1.3×10^{-5}	1.2×10^{-5}	0.246		3.5×10^{-9}		
1.27×10^{-30}	3.3×10^{-6}	8.5×10^{-6}	0.251		7.2×10^{-9}		
2.26×10^{-30}	3.9×10^{-7}	5.8×10^{-6}	0.255		1.2×10^{-8}		
4.02×10^{-30}	9.8×10^{-9}	4.1×10^{-6}	0.260		1.7×10^{-8}		
7.15×10^{-30}	1.2×10^{-11}	3.3×10^{-6}	0.265		2.5×10^{-8}		
1.27×10^{-29}		2.7×10^{-6}	0.270		3.8×10^{-8}	1.0×10^{-12}	2.4×10^{-12}
2.26×10^{-29}		2.4×10^{-6}	0.275		6.0×10^{-8}	1.7×10^{-12}	1.0×10^{-11}
4.02×10^{-29}		2.1×10^{-6}	0.280		9.4×10^{-8}	2.7×10^{-12}	5.0×10^{-11}
7.15×10^{-29}		1.8×10^{-6}	0.284		1.5×10^{-7}	4.0×10^{-12}	2.5×10^{-10}
1.27×10^{-28}		1.5×10^{-6}	0.289		2.2×10^{-7}	5.4×10^{-12}	1.2×10^{-9}
2.26×10^{-28}		1.1×10^{-6}	0.294		3.0×10^{-7}	6.4×10^{-12}	5.4×10^{-9}
4.02×10^{-28}		7.8×10^{-7}	0.299		3.7×10^{-7}	6.2×10^{-12}	2.1×10^{-8}
7.15×10^{-28}		4.3×10^{-7}	0.304		3.7×10^{-7}	4.6×10^{-12}	6.5×10^{-8}

are summarized in Table I for different values of the baryon density (ρ_B). For reasonable values of ρ_B the theory predicts the primordial mass fraction of He^4 in the range 25–30% almost independently of ρ_B. This prediction is a triumph for big bang cosmology. Investigations of the abundance of the elements in first-generation stars and interstellar gas agree well with the theory.[3] The accuracy of the astronomical observations is difficult to evaluate, but a helium abundance exceeding 35–40% is safely excluded. Shvartsman (1969) (and later Steigman *et al.*, 1977) noted that big bang nucleosynthesis was affected by the total number of different sorts of massless particles. The energy density at a given temperature depends on this number, and so does the relation between time and temperature, and as a consequence the rate of the reaction $p \rightleftarrows n$.

If there were 10 or 20 different neutrino types, the helium mass fraction would be about 40 or 50%. This is absolutely forbidden by the observations.

This result was not taken seriously 10 to 15 years ago, but now with new particles being discovered under more and more extreme conditions, the problem of the "demographic explosion" in the elementary particle world is of primary importance. Probably the bounds on the number of types and the properties of neutrinos, at least today, have proved to be the most fruitful area of cosmological applications to elementary particle physics. The bound on the number of different neutrino types could be destroyed if there were an excess of electronic neutrinos over antineutrinos. In our opinion this is hardly probable (see Sec. IX).

The deuterium mass fraction in interstellar gas is about 3×10^{-5}. This value is observed in ionized gas regions free of distortion caused by partial isotope separation in chemical reactions. The admixture of gas emitted by stars increases the helium fraction and decreases the deuterium one. Some deuterium could also be burned in the outer layers of stars. On the other hand, a secondary production of deuterium by cosmic rays or in shock waves is possible. In all cases the absence of a noticeable anomaly in the abundance of He^3 restricts possible fluctuations in the amount of primordial deuterium. The current abundance observations lead to the bound $R(D) > 2\times10^{-5}$ for the primordial deuterium mass fraction.

The idealized theory of deuterium synthesis in a universe with a homogeneous distribution of matter predicts the strong dependence of D abundance on matter density. The higher the ρ_B the deeper is the burning out of D in the nuclear reactions, so the less D survives the era of nucleosynthesis. For example, if $\rho_B = 5\times10^{-30}$ g cm^{-3} then $R(D)=10^{-8}$, and if $\rho_B = 5\times10^{-32}$ g cm^{-3} then $R(D)=10^{-3}$ (see Table I). Therefore, possible inhomogeneity in the distribution of matter strongly affects the deuterium abundance.

The relatively small amount of deuterium forbids some extreme hypotheses which predict energetic particle production after the era of nucleosynthesis. For example, a late baryon annihilation could produce deuterium through the reaction $\bar{p}+\mathrm{He}^4 \to D+n+\pi^+ +\pi^-$. The same is true also for intermediate-mass (10^{-5} g $< m < 10^{10}$ g) black hole evaporation.

Note that the observations of deuterium abundance demand that the density of baryons in the universe should be rather low ($\rho_B \simeq 5\times10^{-31}$ g cm^{-3}). This,

[3]Only stars and interstellar gas containing small percentages of heavy elements (such as carbon and iron) are investigated, in order to exclude the possibility of secondary helium production by stars. According to the theory of stellar evolution, the helium created out of hydrogen by nuclear burning inside a star is not released to enrich the interstellar gas until the star explodes—at which time heavy elements are also released. Therefore, a low abundance of heavy elements indicates a first-generation star with original helium abundance. The same is true for the interstellar gas.

however, does not restrict density of other forms of matter, e.g., massive neutrinos.

To conclude, the observed deuterium abundance, although posing some questions for big bang cosmology, on the whole confirms the theory.

Field theories with degenerate vacuums, e.g., with the Higgs phenomenon, lead to important conclusions about the extreme past of the universe. At sufficiently high temperature there would be no degeneracy and the ground state would be symmetric. As the temperature dropped the symmetric state would become energetically unfavored so spontaneous symmetry breaking and phase transitions could occur (Kirzhnits and Linde, 1972).

The universe's homogeneity contradicts the simple version of such a theory when only two discrete ground states at low temperature are possible (Lee, 1973). As was noted by Zeldovich, Kobzarev, and Okun (1974), two types of domain corresponding to two different ground states should appear in the course of cooling down. These domains are bounded by heavy walls which would violate the homogeneity of the universe.

In other versions of spontaneously broken symmetry theories vortex lines, magnetic monopoles, phase transitions, and other interesting phenomena take place. The observational data, however, permit no definite conclusion about this.

One of the most important observational facts is the ratio of baryon number density to the relic photon density, $\beta \equiv N_B/N_\gamma = 10^{-8}-10^{-10}$. This number can be easily obtained from the temperature of the microwave background (3 K) and the matter density ($5\times 10^{-30}-10^{-31}$ g cm^{-3}). The numerical value of β is poorly known, but it is hardly outside the above mentioned boundaries, so it is definitely much smaller than unity. In the framework of theories with strictly conserved baryonic charge and conserved entropy (i.e., for an adiabatic expansion) β is a constant defined by initial conditions. There have been, however, many attempts to determine β theoretically.

One of these attempts is connected with the existence of the human race, including the readers and authors of this review. If β differed significantly from the known value it would be highly improbable that life would have appeared and evolved up to the present level. This is far from physics, however.

More physical are the attempts to find conditions under which with a given $\beta \gtrsim 1$ and N_B conserved (in a comoving volume) N_γ arises. We mean an initial metric leading to oscillations which provide energy through subsequent damping. An extreme assumption is the generation of primordial black holes (Zeldovich and Novikov, 1967), which would evaporate through the Hawking mechanism (1974) and heat the initial baryon fluid with $\beta \gtrsim 1$. More realistic and tied to the structure of the universe is the concept of small inhomogeneities damping (Zeldovich, 1972).

In such a way the cold universe model may revive in a new guise. A version of this model was considered by one of the authors (Ya. Z., 1962) just before the relic photon discovery [Penzias and Wilson (1965); see also Dicke *et al.* (1965) and Doroshkevich and Novikov (1964)].

In some papers the possibility that baryonic charge is not absolutely conserved was considered.[4] The bound on proton lifetime, $\tau_p > 10^{30}$ yr, does not exclude essential baryon nonconservation at superhigh temperatures (Sakharov, 1967; Kuzmin, 1970). It is also possible that effective baryon nonconservation occurs during the formation and evaporation of black holes (Hawking, 1975; Zeldovich, 1976).

The latest development, and one of the most interesting, is connected with the grand unified theories in which baryonic charge nonconservation is natural and is not introduced *ad hoc*. The initial state in such a theory can be arbitrary. Because of thermodynamic equilibrium established at high temperature, initial conditions are effectively "forgotten" and charge symmetry is obtained. Then in the course of cooling down a definite value of β evolves as a result of nonequilibrium. This independence of the initial state is a very attractive feature of the theory, irrespective of concrete numerical results.

Let us turn now to the information which can be extracted from stellar and especially from solar investigations. It seems that, in their general features, the thermonuclear processes supplying stellar energy are satisfactorily described by contemporary theory. A discrepancy (Davis and Evans, 1976, and references therein) in the neutrino flux from the sun (the theoretical predictions are 2-4 times larger than the observations) may not be of crucial importance. It can probably be explained either by neutrino oscillations or (first proposed by Pontecorvo, 1958) or by an inhomogeneity in heavy element concentration inside the sun (a smaller amount in the core and a larger amount in the outer layer). It is, however, rather disturbing that the theory predicts a 1.6-fold increase of the sun's luminosity during the course of its 4.5×10^9 yr evolution. The geological data hardly accommodate such a variation, and absolutely forbid a greater luminosity change or a lesser age of the solar system. Hypothetical light neutral particles with a rather weak interaction have been postulated which, under certain conditions, could be formed within the sun's core and, having a long mean free path, could effectively transfer energy from the central region to the periphery or even the outside of the sun. This would speed up solar evolution. Going along the same lines one can restrict the properties of these hypothetical particles.

At the beginning of the century Eddington, an astronomer, suggested that hydrogen could be transformed into helium, and that this reaction could supply the energy of the stars. This prediction, made long before the birth of nuclear physics, is an example of a brilliant astronomer's intuition. Later Hoyle was able to make a comparable imaginative leap. He predicted the excited energy level of the carbon nucleus, which was close to the energy of three helium nuclei. This level is necessary for carbon production inside stars.

[4]There is no massless field connected with baryonic charge, as the electromagnetic field is connected with electric charge (Lee and Yang, 1955), so its conservation is not necessarily exact. The possibility of baryonic charge nonconservation was mentioned by Yamaguchi (1959) and Weinberg (1964). See footnote 20.

The recent contributions of astronomy to our understanding of particle physics are on a more modest scale than the achievements mentioned above. Now mostly negative results and some restrictions on particle properties (neutrino mass, number of neutrino types, photon mass, etc.) are obtained. But pride and optimism again revive when one sees that elementary particle physics may be able to explain the most important feature of the universe: the existence of matter, of baryons.

In a conversation with young Planck at the end of the nineteenth century, one of his elder colleagues spoke about the clear sky of theoretical physics on which only two small clouds could be noticed: the ultraviolet divergence in the theory of equilibrium radiation and the Michelson experiment. It is well known now that these clouds brought the thunderstorm of relativity and quantum theory.

On the blue sky of modern cosmology there is a cloud: the observed structure of the universe shows a considerable inhomogeneity at the scale of less than or about 100 Mpc. This could have developed because of gravitational instability. The observed structure by itself is compatible with the assumption of small but finite metric perturbations from the initial singularity until the decoupling of matter and radiation (Lifshitz, 1946). The isotropy of the microwave background radiation supports the assumption of small perturbations (about 10^{-2}–10^{-3} in relative units). These perturbations decrease for larger scales and become less than 10^{-4} for the largest scale, which is of the order of today's horizon (6000 Mpc).

We do not know, however, the amplitude of the short-wave perturbations. Could there have been strong short-wave perturbations smoothed down long before the present epoch? Could strong and very short-range perturbations produce primordial black holes in the extremely dense cosmological plasma? This question was first posed by Zeldovich and Novikov in 1967. The black hole evaporation discovered by Hawking in 1974 shows that primordial black holes with a mass of less than 10^9 g could indeed have appeared and then vanished without leaving a trace.

How will modern cosmology be modified after the whole spectrum of perturbations, which distinguishes our real world from the idealized Friedman model, is investigated and understood?!

The development of elementary particle physics shows that Nature is far from being as symmetric as was believed even a quarter of a century ago. The first among seemingly well established symmetries to collapse was parity conservation. Experiment showed that looking at some physical processes through a mirror one would see phenomena that did not exist in reality. Symmetry was reestablished by the assumption that particles should be exchanged with their antiparticles simultaneously with space reflection. In less than ten years this hypothesis also proved to be incorrect. Such a destruction of well believed principles was not without consequence. Now all earlier unassailable dogmas are called into question. Theoreticians demand that we reject the notions of baryonic charge conservation and of absolute proton stability (see Sec. XV). In some papers the authors have expressed doubts even about such sacred principles as CPT invariance and S-matrix unitarity. A related problem is the nonconservation of electric charge. Owing to the extremely small mass of photons, electric charge nonconservation leads to tremendous difficulties for a theory, and in this sense is very close to violation of CPT invariance or S-matrix unitarity.

In this review astronomical, but not especially cosmological, data are used to check the validity of the Maxwell equations. In particular, bounds on the photon rest mass are presented and the question of electron stability is discussed. On the other hand, we do not consider the problems of verifying general relativity, in spite of its close relation to the theory of one type of elementary particle—the graviton (and probably to its supersymmetric relatives such as gravitinos and others). There are excellent review papers devoted to the latter problem; many of them appeared in connection with the hundredth anniversary of Einstein's birth. All the experiments give excellent agreement with general relativity. We can only add that the so-called cosmological constant Λ, which is neither in contradiction with nor obligatory in general relativity, is bounded by the inequality $\Lambda < 2\times 10^{-56}$ cm^{-2}. This limit is fairly reliable. Probably it will be improved by an order of magnitude in the near future. The restriction for the vacuum energy density can be written

$$|\rho_{\text{vac}}| = |\Lambda/8\pi G| < 10^{-29}\ \text{g cm}^{-3} = 5\times 10^{-47}\, m_N^4$$

(Zeldovich, 1968). It would be reassuring in a field theory to have such a small value for ρ_{vac}. S. Weinberg (1980), in his Nobel Lecture, refers to the problem of a vanishing cosmological term, "the old mystery of why quantum corrections do not produce an enormous cosmological constant;... one is concerned with ... [a] term in the effective Lagrangian which has to be adjusted to be zero. ... The adjustment must be precise to some fifty decimal places."

B. Organization of the paper

The paper is organized as follows. In Sec. II we present general formulas for particle densities in the hot expanding universe and derive an analytic expression for the limiting concentrations of heavy stable particles as $t\to\infty$. These concentrations are determined by the particle masses and their annihilation cross sections. Then the plasma heating that results from the annihilation of particles with $m > T$ is briefly discussed.

In Sec. III a restriction on the number of light particle types is given, based upon the observed He^4 abundance in the universe. If the mass fraction of He^4 [R (He)] is less than 0.25 then in addition to the known ν_e, ν_μ, and ν_τ only one massless (or with $m < 1$ MeV) neutrino can exist. With the more cautious estimate R(He)<0.29 not more than four new neutrino types (in addition to those known) are permitted. All this is true if the baryon density ρ_B is at its lower observational limit. For larger values of ρ_B the restriction on the number of neutrino types is more stringent. Evaluating the influence of neutrinos on primordial nucleosynthesis one is able to get a bound on the right-handed weak current coupling constant which is much stronger than that obtained in laboratory measurements. Analogous considerations are applicable to any light particle with a weaker interaction than the neutrino.

In Sec. IV a limit is imposed on the masses of light weakly interacting particles, which follows from the condition $\rho_{\text{particles}} < \rho_{\max}$. Analogous limits for heavy stable particles are obtained in Sec. V. In particular the concentrations of relic unconfined quarks and protons are given for the case of charge symmetric initial state and conserved baryonic charge. The lower bound on the annihilation cross section of any stable particle is obtained, $\sigma v > 10^{-36}$ cm^2 [see Eq. (5.4)]. In Sec. VI stable neutral weakly interacting particles are considered. It is shown that no such particle can exist in the mass interval 40 eV $< m_L <$ 3 GeV in a cosmology with vanishing Λ. This means, in particular, that the mass of the new neutrino connected with $\tau^{\pm}$ is small, i.e., $m_{\nu_\tau} < 40$ eV (if ν_τ is stable). In Sec. VII astrophysical constraints on the mass versus lifetime of a neutral unstable lepton are presented. The constraints obtained are based upon analysis of the energy density in the universe, the spectrum of cosmic electromagnetic radiation, primordial nucleosynthesis, and the dynamics of supernovae. Various arguments forbid much of the area in the $(m-\tau)$ diagram, leaving very little room for a possible neutral lepton. In Sec. VIII electromagnetic parameters of neutrinos such as electric charge, magnetic moment, and charge radius are constrained. The results are based on consideration of the neutrino luminosity of the sun and white dwarfs.

In Sec. IX we discuss the possibility of a strong degeneracy of the neutrino background, i.e., of a considerable leptonic charge of the universe. In this case the He^4 abundance gives no reasonable restriction on the number of new neutrino types. However, the cosmological deuterium abundance proves to be sensitive to this number. Grand unified theories of elementary particle interactions argue against this degeneracy; $(\nu - \bar{\nu})$ excess should be of the same order as that of $(B - \bar{B})$.

In Sec. X the solution of the galactic hidden mass problem by a neutral lepton halo is briefly discussed, as are some other astrophysical implications of massive neutrinos.

Astrophysical constraints on the properties of different hypothetical particles are presented in Sec. XI. There is a serious contradiction between the cosmological prediction of the relic monopole concentration and the observations. Probably this indicates that the magnetic monopole does not exist. Some other possibilities are briefly discussed.

In Sec. XII astronomical observations which severely restrict the mass of the photon are considered. In Sec. XIII, the closely related possibility of electric charge nonconservation and electron nonstability are discussed. Some astrophysical and cosmological consequences of this assumption are briefly examined. Arguments are presented in favor of an extremely long electron lifetime or even of absolute stability because of the zero photon rest mass.

In Sec. XIV the implications for cosmology of spontaneously broken elementary particle models at times close to the initial singularity are considered.

Various ways of explaining the baryon excess in the universe are discussed in Sec. XV. The following models are considered:

1. a cold universe with the subsequent generation of entropy;
2. spatial separation of matter and antimatter;
3. concealment of antibaryons in primordial black holes;
4. baryonic charge nonconservation in elementary particle interactions and dominant baryon production in the primeval plasma because of C and CP violation;
5. an excess of B over $\bar{B}$ because of CPT violation.

Particular attention is paid to hypotheses 3 and 4. In our opinion model 4 is the most attractive. It would be very elegant to connect one of the fundamental numbers in cosmology, the ratio of the number of baryons to that of photons, with the laws of elementary particle physics. A short discussion of vacuum polarization in curved space-time is presented in Sec. XVI.

II. PARTICLE CONCENTRATIONS IN THE HOT UNIVERSE

As we have already noted, the observational data support the big bang scenario. For a universe filled with relativistic particles so that $p = \varepsilon/3$, the energy density depends on time for small t in the well-known way

$$\rho = \frac{3}{32\pi G t^2}. \tag{2.1}$$

Here, $G = 6.67 \times 10^{-8}$ cm^3 g^{-1} sec$^{-2} = 0.6 \times 10^{-38} m_N^{-2}$ is the gravitational constant, m_N is the proton mass, t is the time counted from the singularity, and $t_P = G^{1/2} = 10^{-43}$ sec is the Planck time. Hence the total number density of all sorts of particles is $N_{\text{tot}} = \text{const} \times (t_P t)^{-3/2}$.

A detailed discussion of the expansion law can be found, for example, in reviews by Zeldovich (1963, 1965c) and in books by Weinberg (1972, 1977) and by Zeldovich and Novikov (1975). We should like only to note that the dependence of ρ on t is universal. There are no additional parameters. In particular there is no dependence on the number of different particle species, so long as their masses are small. This number is essential, however, in the connection between the time t and the plasma temperature (see below).

Equation (2.1) is connected with the expansion law

$$\dot{a}/a = H(t),$$

where a is some scale and $H(t)$ is the Hubble constant, $H(t) = 1/2t$ for small t. From this it follows that the temperature of relativistic particles drops as $T \sim E = p \sim a^{-1} \sim t^{-1/2}$. The temperature of nonrelativistic particles admixed with relativistic gas falls faster:

$$T_{\text{nonrel}} \sim \frac{p^2}{m} \sim t^{-1} \tag{2.2}$$

if there is no energy exchange between nonrelativistic and relativistic particles. As for the energy density of massive particles, two cases should be distinguished. If massive particles are in chemical equilibrium with plasma and the average values of all conserved charges inherent in these particles are zero then their energy density falls rapidly with time, as $\exp(-m/T)$. If, however, the number of massive particles in a comoving volume is conserved, their energy density falls with time more slowly than that of massless ones and ultimately the energy density of the universe is dominated by nonrelativistic particles.

In the early stages it is assumed that the energy density of the universe was dominated by relativistic particles. This assumption could be wrong if the density of states rose exponentially with mass, i.e., if the Hagedorn picture with a limiting temperature were correct. According to the modern point of view, however, elementary particle interactions are governed by quark dynamics and are asymptotically free. There is no exponential increase of the number of elementary constituents in this picture. Moreover, the consideration of primordial He^4 formation (see Sec. III) provides a rather strict upper bound on the total number of neutrino types and, because of quark-lepton symmetry (if it exists), on the number of quark flavors too.

For what follows the problem of thermal equilibrium in the primeval plasma is of crucial importance. Usually equilibrium is established after a sufficiently long time, but in our case the situation is just the opposite. The older the universe, the farther away it is from equilibrium. When t is large, the expansion is slow but the reaction rates decrease faster than the rate of expansion, and consequently equilibrium is not established. By contrast, at an early stage reactions were fast due to the higher particle number and energy densities, so the gas of elementary particles was, generally speaking, in equilibrium. Let us discuss this in more detail. The universe's expansion rate is $\dot{a}/a \sim t^{-1}$ (here a is the radius of the universe). On the other hand, the rate of approach to equilibrium is

$$(Na^3)^{-1} d(Na^3)/dt \sim N\sigma v \ ,$$

where N is the particle number density, v is the particle velocity, and σ is their interaction cross section. The equilibrium destroyed by the expansion of the universe could be reestablished if

$$N\sigma v t \geq 1 \ . \qquad (2.3)$$

This condition will be made more precise below [see Eq. (2.9)]. For $T \gtrsim m$ particle number density is of the order $N(t) \simeq (tt_P)^{-3/2}$, where $t_P = G^{1/2} \simeq 10^{-43}$ sec is the Planck time; the inverse quantity $T_P = t_P^{-1} \simeq 10^{19}$ GeV is the Planck temperature (or mass). In gauge theories $\sigma \simeq \alpha^2/T^2$, if the temperature is higher than the intermediate boson masses (here $\alpha \simeq 10^{-2}$ is the gauge coupling constant). The temperature depends on the time according to the law $T \simeq (t_P t)^{-1/2}$; the condition for equilibrium (2.3) is fulfilled when $t > \alpha^{-4} t_P$ (the equilibrium for the gauge boson decays could be established earlier). As the universe grows older the temperature drops and equilibrium is no longer maintained, first because the equilibrium particle density falls steeply, as $\exp(-m/T)$ for $T < m$, and second because the rise in cross section ($\sim T^{-2}$) ends when the temperature gets lower than the gauge boson masses. As for extremely early times $t < t_P$ nothing is known about this. Some people believe that the gravitational interaction at this instant could have been strong, and it is possible that equilibrium existed. Then there would have been a nonequilibrium period $t_P < t < \alpha^{-4} t_P$ and the equilibrium would again be maintained for $\alpha^{-4} t_P < t < t_1$, t_1 being defined to an order of magnitude by the condition $T(t_1) = m$ (for a more precise discussion see below).

The equilibrium energy density of massive particles is given by

$$\rho_m = \frac{g_s}{2\pi^2} \int_0^\infty \frac{(p^2+m^2)^{1/2} p^2 dp}{\exp\left[\frac{(p^2+m^2)^{1/2}}{T}\right] \pm 1}$$

$$= \begin{cases} g_s \dfrac{\pi^2}{30} T^4 b_\pm , & T > m \\[2ex] g_s \dfrac{m}{\sqrt{2}\,\pi^{3/2}} (mT)^{3/2} \exp(-m/T) , \end{cases} \qquad (2.4)$$

where g_s is the number of particle spin states; for photons and electrons $g_s = 2$ and for massless left-handed neutrinos $g_s = 1$. The plus sign corresponds to fermions and the minus sign to bosons; $b_+ = \frac{7}{8}$ and $b_- = 1$. Thus taking into account only relativistic particles ($m < T$) we obtain for the total energy density

$$\rho = \frac{\pi^2}{30} T^4 N_{DF} \ , \qquad (2.5)$$

where N_{DF} is the number of effective degrees of freedom: the γ quanta contribution to N_{DF} is 2, electrons and positrons together give $\frac{7}{2}$, and each type of neutrino (with one helicity state) gives $\frac{7}{4}$. Thus if there are photons, $e^\pm$, ν_e, $\bar{\nu}_e$, ν_μ, and $\bar{\nu}_\mu$ in the thermal equilibrium, then $N_{DF} = 9$. In modern elementary particle models N_{DF} can be of the order of a hundred when the temperature is sufficiently high so that quark, gluon, and other degrees of freedom are excited.

Comparing Eqs. (2.1) and (2.5) we find the following time dependence of the temperature:

$$T = \left(\frac{45}{16\pi^3}\right)^{1/4} N_{DF}^{-1/4} (t_P t)^{-1/2}, \quad \text{i.e.,} \quad T_{\rm Mev} = 1.5 N_{DF}^{-1/4} (t_{\rm sec})^{-1/2} \qquad (2.6)$$

As the universe expands and the temperature drops, the real concentration of massive particles fails to follow the equilibrium value, which decreases as $\exp(-m/T)$. As we shall see below, the concentration of stable massive particles in a comoving volume tends to a constant when the rate of their annihilation becomes smaller than the expansion rate of the universe. Note that during the expansion, when the temperature is falling, the energy distribution of massless noninteracting particles conforms to a blackbody curve. For massive particles the situation is different. Expansion without interaction would lead to a large departure from equilibrium. The concentration of massive particles X is determined by the equation (see, for example Zeldovich and Novikov, 1975)

$$\dot{N}_X = -N_X^2 v\sigma(X\bar{X} \to \text{all}) - 3HN_X + \psi(t) \ , \qquad (2.7)$$

where the second term on the right-hand side comes from the expansion of the universe and $\psi(t)$ gives the

X-particle creation. It is convenient to introduce the relative concentration $r_X = N_X/N_\gamma$, where N_γ is the concentration of conserved massless particles (i.e., satisfying the equation $\dot{N}_\gamma = -3HN_\gamma$). If we neglect the processes of entropy generation, such as plasma heating owing to decay or the annihilation of massive particles when they depart from thermal equilibrium, then N_γ coincides with the photon concentration, $N_\gamma \simeq 0.24T^3$. So we obtain

$$\dot{r}_X = -\sigma v N_\gamma (r_X^2 - r_{X_{eq}}^2), \tag{2.8}$$

where $r_{X_{eq}}$ is the relative equilibrium concentration,

$$r_{X_{eq}} \approx \begin{cases} 1 & \text{if } \theta \equiv T/m > 1 \\ \frac{1}{4} g_s \theta^{-3/2} \exp(-\theta^{-1}) & \text{if } \theta < 1 . \end{cases}$$

It is noteworthy that the substitution of $r_X N_\gamma$ for N_X excludes the effect of the expansion. Indeed the term $3HN_X$ is present in Eq. (2.7) but is absent in Eq. (2.8). When deriving Eq. (2.8) from (2.7) the thermodynamic relation between direct and inverse reactions was used. This permits us to express $\psi(t)$ through σ, $\psi = N_{eq}^2 v\sigma$ ($X\bar{X} \to$ all). Equation (2.8) can be integrated numerically (Wolfram, 1978), but a sufficiently accurate result can be obtained in the following way. When t is small, r is close to its equilibrium value $r_{X_{eq}}$. The equilibrium is destroyed when the rate of change of the equilibrium concentration due to the temperature decrease turns out to be comparable to or larger than the reaction rates, i.e.,

$$\dot{r}_{eq}/r_{eq} = 2\sigma v N_\gamma r_{eq} . \tag{2.9}$$

As $\dot{r}_{eq}/r_{eq} \simeq (m/T)t^{-1}$, where t is as usual counted from the "beginning," the equilibrium holds if $2\sigma v N_\gamma r_{eq} tT/m > 1$. The instant of equilibrium destruction [the instant of quenching according to the pioneering papers by Zeldovich (1965a) and Zeldovich *et al.* (1965); see also Zeldovich (1975)], or the instant of freeze-out as it is termed in the English literature, is defined by

$$\frac{g_s}{4} \theta_f^{1/2} e^{-1/\theta_f} = \frac{7N_{DF}^{1/2}}{\sigma v m_P m_X}$$

Equation (2.6) and the expression for the equilibrium photon concentration were used here. Substituting the numerical values of the constants we obtain

$$\frac{1}{\theta_f} - \tfrac{1}{2}\ln\theta_f = \ln\frac{g_s \sigma v m_P m_X}{28N_{DF}^{1/2}} = 40 + \ln\frac{g_s \sigma v m_N m_X}{N_{DF}^{1/2}} , \tag{2.10}$$

where m_N is the nucleon mass.

With θ_f defined by this equation the relative concentration of X particles at the moment of quenching (freeze-out) can be found,

$$r_{fX} = (2\sigma v N_\gamma t_f \theta_f)^{-1} .$$

The asymptotic (as $t \to \infty$) value of the relative concentration is defined by Eq. (2.8) in which zero is substituted for r_{eq} and the boundary condition is $r = r_{fX}$ at $t = t_f$. So one finds,

$$r_{0X} = 2\theta_f r_{fX} = \frac{14N_{DF}^{1/2}\theta_f}{\sigma v m_P m_X} = 10^{-18} N_{DF}^{1/2} (\sigma v m_N^2)^{-1} \frac{m_N}{m_X} \left(40 + \ln\frac{g_s \sigma v m_N^2}{N_{DF}^{1/2}} + \ln\frac{m_X}{m_N}\right). \tag{2.11}$$

This is valid if $\theta_f < 1$.

Equation (2.8) was integrated numerically by Wolfram (1979). In terms of the parameter Z introduced in his paper our result (2.11) can be written as $r_{0X} = 2Z^{-1}\ln Z$. Comparison of this expression with the numerical results shows agreement to within an order of magnitude. We do not understand, however, the factor m_X^4 in the final expression for Z given by Wolfram.

If after the quenching of X particles the annihilation of some massive particles heats the plasma, the number of photons slightly increases and the ratio r decreases. For example, e^+e^- annihilation increases the photon gas temperature by the factor $(11/4)^{1/3} = 1.4$ as compared to the neutrino temperature. This was first shown by Peebles (1966). Indeed from entropy conservation in the course of the expansion it follows that

$$S_b(\gamma) + S_b(e^+e^-) = \frac{4\pi^2}{15} T_b^3 (1 + \tfrac{7}{4}) = S_a(\gamma) = \frac{4\pi^2}{15} T_a^3 , \tag{2.12}$$

where the indices b and a refer to the quantities before and after the annihilation, respectively. At the moment of annihilation the neutrinos have already decoupled from the plasma and so $T_\nu = T_b$ and, because of Eq. (2.12), $T_\gamma = T_a = 1.4\, T_\nu$. That is why the present photon number density $N_{0\gamma} \sim T_\gamma^3$ is three times that of the neutrino.

III. BOUNDS ON THE NUMBER OF LIGHT PARTICLE SPECIES

A very impressive result of modern cosmology is an upper bound on massless particle types. This restriction is based on the study of primordial element formation (mainly He^4). It can be said that not less than half of the total number of neutrino types is known at present (ν_e, ν_μ, and ν_τ).

The first person to discuss the influence of massless particles on primordial nucleosynthesis was Shvartsman (1969). Later writers were able to treat the subject more precisely (Steigman *et al.*, 1977; and Schramm, 1978) by the use of the latest observational data. The arguments are the following. The amount of He^4 produced during the big bang is dependent on the neutron–proton ratio at freeze-out. Neutron freeze-out (or quenching) takes place when the rate of the weak interaction processes $e^-p \leftrightarrow n\nu_e$ and $e^+n \leftrightarrow p\bar{\nu}_e$ drops below the expansion rate. Later at about 0.1 MeV almost all neutrons are bound into He^4 so the mass fraction of He^4 is $2(n/p)/[1+(n/p)]$, where n/p is the neutron-proton ratio. The neutron-proton ratio at freeze-out is

$$N_n/N_p = \exp(-\Delta m/T_f) , \tag{3.1}$$

where N_n and N_p are the neutron and proton concentrations, respectively; Δm is their mass difference, and T_f is the freeze-out (quenching) temperature. As was mentioned earlier, the instant of quenching is determined by the condition $\tau_{\text{reaction}} \simeq t$, where $\tau_{\text{reaction}} = (\sigma v N_\nu)^{-1}$ is the characteristic time of the reactions involved. The neutrino density is proportional to T^3, the cross section of the $(n-p)$ transition is $\sigma \sim T^2$, and the universe's age is related to the plasma temperature through $t = N_{DF}^{-1/2} T^{-2}$. Hence the change of the freeze-out temperature due to variation of N_{DF} is defined by

$$T_f' = T_f (N_{DF}'/N_{DF})^{1/6} \tag{3.2}$$

and, accordingly, the relative neutron concentration is

$$\frac{N_n}{N_p} \sim \exp\left[-\frac{\Delta m}{T_f}\left(\frac{N_{DF}}{N_{DF}'}\right)^{1/6}\right], \tag{3.3}$$

where T_f is the freeze-out temperature for $N_{DF} = N_{DF}'$.

The standard calculations (Schramm and Wagoner, 1977) predict a mass fraction of $R(\mathrm{He}^4) = 0.23 - 0.26$ (Table I). The uncertainty in this result is due to the dependence of $R(\mathrm{He}^4)$ on the baryon density, $\rho_B = (10^{-31} - 5\times 10^{-30})$ g cm^{-3}. The equilibrium ratio N_n/N_p which is determined by Eqs. (3.1)–(3.3) is independent of ρ_B, but the further reactions $n + p \to D$ and $D + D \to \mathrm{He}^4$ are two-body with respect to baryons and so the total result of the transition $2n + 2p \to \mathrm{He}^4$ depends on ρ_B(rather weakly for realistic values of ρ_B). The higher ρ_B, the greater the resulting He^4 abundance.

From Eq. (3.3) it follows that

$$\frac{N_{DF}'}{N_{DF}} = \frac{T_f}{\Delta m}\left(\ln \frac{2 - R(\mathrm{He}^4)}{R(\mathrm{He}^4)}\right)^{-6}.$$

Using the observational upper bound on $R(\mathrm{He}^4)$ and normalizing T_f by the standard $R(\mathrm{He}^4)$ calculation result, we can limit N_{DF}'/N_{DF} from above. Taking into account that

$$\frac{N_{DF}'}{N_{DF}} = 1 + \frac{7}{36}(k_\nu - 2), \tag{3.4}$$

where k_ν is the total number of neutrino species, we obtain an upper bound on k_ν. [It is $(k_\nu - 2)$ that enters Eq. (3.4) because ν_e and ν_μ have already been taken into account.] It has been argued (Peimbert, 1976) that $R(\mathrm{He}^4) < 0.29$. This gives $k_\nu < 8$. Recently a much more restrictive bound was reported (Schramm, 1978), limiting the total number of all neutrino types to less than four. The latter can be obtained if $R(\mathrm{He}^4) < 0.25$ (Thum *et al.*, 1979; Tayler, 1979). If this is true then only one new neutrino type in addition to the known ν_e, ν_μ, and ν_τ can exist. Correspondingly, there can be only one new charged lepton connected with this neutrino. Because of quark-lepton symmetry the total number of quark flavors should be less than or equal to eight. Up to now, five types of quarks have been observed in laboratories: u, d, s, c, and b. So no more than three new quark flavors can be discovered. It is possible, however, that this point of view is too optimistic (or pessimistic?). First of all, the reliability of the calculations, as well as of the observational data, may be overestimated. Then there is the inelegant but formally not excluded possibility of a considerable degeneracy of the electronic neutrino. This could well invalidate the bounds obtained (see Sec. IX).

Recently Stecker (1980) has argued that the observed abundance of He^4 is less than 0.23. This is incompatible with the standard scenario, even with three types of neutrons. So Stecker recommends revising the usual big bang theories. However one way out of this difficulty is to assume a lower baryonic matter density and a dominating non-nucleonic mass present in the universe, e.g., in the form of massive neutrinos (see Sec. X). In view of modern trends in neutrino physics, this seems to be most appealing.

The above limits on the value of k_ν refer to massless left-handed neutrinos (with only one possible helicity state). If there were right-handed neutrinos too, the number of degrees of freedom would be twice as large and k_ν would be effectively doubled. If this were the case k would be equal to six even for the known neutrinos, ν_e, ν_μ, and ν_τ. It has been shown, however, that $k_\nu < 4$ [for the extreme limit $R(\mathrm{He}^4) < 0.25$]. One might be tempted to conclude that neutrinos are strictly massless. (If neutrino mass does not vanish, both helicity states are permitted. The number of states is not doubled, however, for the Majurana neutrino.) The conclusion would be wrong, however. If the neutrino mass were small but nonvanishing and the right-handed neutrinos had no direct interactions (except for the gravitational one, of course) or interacted very weakly, then the contribution of the right-handed neutrinos to the energy density and to N_{DF} would be suppressed by the factor $(T_R/T_L)^4$. Here T_R (T_L) is the temperature of the right- (left)-handed neutrino gas. The inequality $T_R \neq T_L$ is based on the fact that the interaction of ν_R is weaker than that of ν_L. That is why ν_R is out of thermal contact with the rest of the universe at much higher temperature (T_{DR}) than is ν_L. The particles with masses in the interval $T_{DR} > m > T_{RL}$ heat the plasma by their annihilation (Peebles, 1966) so the temperature of left-handed neutrinos rises but the temperature of ν_R does not change because the latter is already decoupled. This leads to a relative suppression of the role of ν_R. A similar conclusion was reached by Shapiro *et al.* (1980). It is noteworthy that if neutrinos are massive, then the mass matrix should be either of the Dirac form (i.e., $m_D \bar{\nu}\nu$) or of the Majurana form (i.e., $m_M \nu C \nu$), but not both because in the latter case all four degrees of freedom of a neutrino would be excited in the primeval plasma. (Dolgov, 1980b).

The decoupling temperature for the case $T > m$ is defined by

$$m_N T_D \sigma(T_D) = 2\times 10^{-18} N_{DF}^{1/2}/\bar{N}_{DF}. \tag{3.5}$$

Here use was made of Eqs. (2.3) and (2.6) and of the expression $N \simeq 0.2 T^3 N_{DF}$ for the number density of relativistic particles. $\bar{N}_{DF}$ is the number of particle species interacting with neutrinos. Different species are weighted in proportion to the value of their cross section. For left-handed electronic neutrinos

$$\sigma_L(T) = 4G^2T^2/\pi \simeq 4\times 10^{-10} T^2 m_N^{-4} \pi^{-1}$$

and, as we shall see, the decoupling temperature is such that $N_{DF} = 5.5$ and $\bar{N}_{DF} = 1$ for ν_μ and ν_τ, and $\bar{N}_{DF} = 2$ for ν_e. So Eq. (3.5) gives $T_{DL} \approx 3$ MeV. If ν_R has no direct interactions then $\sigma_R = (m_\nu/T)^2 \sigma_L$ is to be substituted into Eq. (3.5) and $T_{DR} = 10^{-8} m_N (m_N/m_\nu)^2$ if T_{DR}

$<T_P$. If ν_R interacts in the same way as ν_L but with a suppressed cross section, i.e., $\sigma_R = \beta_R \sigma_L$, $\beta_R < 1$, then Eq. (3.5) gives

$$\beta_R (T_{DR}/m_N)^3 \simeq 10^{-8}. \tag{3.6}$$

The consideration of ν_R's effect on nucleosynthesis permits us to limit the value of β_R. To suppress the contribution of ν_R to the energy density, the decoupling temperature should be at least larger than the muon mass. This constraint is insufficient, however, because the plasma temperature rises only 1.1 times due to muon annihilation. Therefore ν_R contributes 1.4 times less to the energy density than does ν_L. However, even this rather weak requirement results in a bound on the coupling constant of the right-handed current,

$$\beta_R < 10^{-5}, \tag{3.7}$$

which is much stronger than those obtained in the laboratory.

Strictly speaking we cannot exclude the possibility that one of the three known neutrinos has a strong coupling with the right-handed current, but the other two must satisfy the coupling condition (3.7). One should keep in mind, however, that the conclusion depends upon the accuracy of our determination of He^4 abundance.

Analogous reasoning was applied (Olive *et al.*, 1978) to hypothetical massless particles with a weaker interaction than that possessed by neutrinos. The weaker the interaction, the lower the particle temperature at the time of neutron quenching (freeze-out), and the smaller the influence on nucleosynthesis. Accordingly the number of permitted particle types depends on their interaction strength. For the weakest of all known interactions, gravitation, which is inherent in any particle, the number of particle species which does not change the nucleosynthesis is, as stated by Olive *et al.* (1978), smaller than 20. This result, however, depends on the number of heavy particle types which heat the plasma by their annihilation. If one takes into account a large number of gauge bosons and Higgs particles omitted in that paper, the discussed bound can become 2–3 times larger. With an increase in interaction strength the limit on the number of massless particle types grows increasingly restrictive and reaches four for neutrinos, independent of heavy particles. It is noteworthy that with the same arguments one can restrict the energy density of relic gravitons (gravitational waves) which came from the earliest epoch ($t < t_P$). The graviton density is not restricted otherwise because gravitons perhaps never were in equilibrium with other particles (Kobzarev and Peshkov, 1974).

IV. LIGHT PARTICLE MASSES

The instant of particle quenching (freeze-out) depends on two competing factors (other than the expansion rate): their mass and their interaction strength. The larger the mass the earlier the annihilation stops; on the other hand, the larger the cross section, the longer the annihilation continues. If particle interactions are sufficiently weak their annihilation stops and the concentration freezes out at a high temperature $T > m$. So their number density now is about that of the relic photons. If the particle mass is larger then $3\,K \simeq 3\times10^{-4}$ eV their energy exceeds the energy of the electromagnetic background radiation. The upper bound for the total energy density is known to be $\rho_{\max} = 10^{-29}$ g cm^{-3} $= 5.6$ keV cm^{-3}. The energy density of massive stable particles should be smaller than this value. These arguments were presented by Gerstein and Zeldovich (1966) to limit the muonic neutrino mass. Later the numerical result of their paper was improved (Cowsik and McClelland, 1972; Szalay and Marx, 1976) by using newer and more accurate observational data and with the assumption of equal masses of all neutrino types.

As we saw in the preceding section the quenching (freeze-out) temperature for neutrinos is about 3 MeV (to be more precise, $T_f \simeq 3$ MeV for ν_e and $T_f \simeq 5$ MeV for ν_μ and ν_τ). This result is valid for any type of neutrino with $m < 3$ MeV. After the instant of quenching, ν_μ and ν_τ remain in kinetic equilibrium with the electron-photon plasma because of the neutral current interaction. For electronic neutrinos, charged currents are also operative. After e^+e^- annihilation ($T < 0.5$ MeV) the kinetic equilibrium is destroyed and ν's decouple from the plasma. If $m_\nu > 3$ MeV then at the instant of freeze-out the neutrino would be nonrelativistic. We shall return to this possibility in Sec. VI. Here let us consider limits on the light ($m < 3$ MeV) neutrino masses. At $T \simeq 3$ MeV the number density of neutrinos is $N_\nu = (\frac{3}{4}) N_\gamma$. The origin of the factor $\frac{3}{4}$ is the following: The neutrino has half the polarization states that the photon has; the equilibrium number density of relativistic fermions differs from that of bosons by a factor of $\frac{3}{4}$; and the factor 2 comes from taking into account ν and $\bar{\nu}$. At the temperature $T \simeq m_e = 0.5$ MeV, e^+e^- annihilations heat the plasma, and the photon gas temperature rises by a factor of 1.4 over that of the neutrino (see Sec. III). This increases the photon concentration, and the relation between N_ν and N_γ now is

$$N_{\nu 0} = \tfrac{3}{4}(1.4)^{-3} N_{\gamma 0}, \tag{4.1}$$

where $N_{\nu 0}$ and $N_{\gamma 0}$ are the contemporary values of the neutrino and photon number densities, respectively. As $N_{\gamma 0} \simeq 550$ cm^{-3} so $N_{\nu 0} \simeq 150$ cm^{-3}. The condition $150 \sum m_{\nu i}$ cm$^{-3} < \rho_{\max}$ leads to

$$\sum m_{\nu i} < 40 \text{ eV } [\text{if } m_{\nu_i} < 3 \text{ MeV}], \tag{4.2}$$

where the summation is over all neutrino types ν_e, ν_μ, ν_τ, etc.

In the same way bounds can be obtained on the masses of hypothetical light particles which interact more weakly than neutrinos. One has to take into account only that the weaker the interaction of a particle, the earlier it decouples from the plasma and the lower is its temperature now (see discussion in Sec. III). In this connection the statement that for massive neutrinos the factor 2 should be introduced into the right-hand side of Eq. (4.1) seems to be incorrect. In fact the contribution of right-handed neutrinos into ρ is suppressed by the factor $(T_R/T_L)^3$, and as we saw in Sec. III, $T_R < T_L$. The value of T depends on the ν_R interaction strength and the number of heavy particle species.

To appreciate the result (4.2) we recall the laboratory bounds (Particle Data Group, 1978) on the masses of different neutrino types: $m_{\nu_e} < 30$ eV, $m_{\nu_\mu} < 5\times 10^5$ eV, $m_{\nu_\tau} < 2\times 10^8$ eV. For ν_μ and ν_τ the laboratory limits cannot compete with the cosmological ones. (See, however, the discussion in Sec. X.)

V. QUENCHING AND CONTEMPORARY CONCENTRATIONS OF HEAVY STABLE PARTICLES

If particles X have a sufficiently strong interaction they become nonrelativistic when they are still in thermodynamic equilibrium. They go out of equilibrium for $T < T_{fX} \simeq m_X/\ln(\sigma m_P m_X)$. In this case their residual (as $t\to\infty$) concentration is defined by Eq. (2.11). Possible plasma heating by the annihilation of particles with $m < T_{fX}$ is not taken into account in this expression. It will not, however, markedly change the results presented below.

The substitution of $m_X = m_N$ and $\sigma_N v = 10^{-26}$ cm^2 into Eq. (2.11) yields the relative baryon concentration (Zeldovich, 1965a; Chiu, 1966; Zeldovich and Novikov, 1975)

$$r_B = 10^{-18}. \tag{5.1}$$

This number was obtained under the assumption of baryon–antibaryon symmetry. This is in conflict with observation, however. There are only baryons and no antibaryons in the visible part of the universe, and the number of baryons is much larger than that given by Eq. (5.1). The ratio of baryon number density to photon density is 10^{-8}–10^{-10}. So either one has to assume that the baryon excess always existed in the universe, or one has to seek a dynamical explanation of the observed baryon asymmetry (see Sec. XV). Note that if there is an excess of baryons over antibaryons, then the residual antibaryon concentration is much smaller than that given by Eq. (5.1).

Along the same lines, the number of free relic quarks has been estimated (Zeldovich, Okun, and Pikelner, 1965). For $m_q = 100 m_N$ and $\sigma_q v_q \simeq \sigma_N v_N \simeq 10^{-26}$ cm^2 one obtains $r_q \simeq 10^{-20}$, i.e., one quark for 10^{10}–10^{11} nucleons. In other words, quarks (if they are unconfined) are as abundant in our world as gold. The searches for fractionally charged quarks in different media surely exclude this possibility. The bounds on the quark-to-nucleon number ratio vary from 10^{-19} for iron to 10^{-27} (Ogorodnikov *et al.*, 1979) for recently studied ocean water. These results precent a strong argument in favor of quark confinement. Strictly speaking, however, an exotic quark behavior leading to a small abundance in the samples investigated cannot be excluded. Those who are interested in these problems can find a list of references in the review by Jones (1978). Recently the Stanford group (La Rue *et al.*, 1977, 1979) claimed the discovery of several fractional charges on niobium samples contaminated with tungsten. One needs a further independent experimental confirmation, however, and in this case a revolutionary change of our ideas would be necessary.

The baryonic asymmetry of the universe could yield an excess of quarks over antiquarks at the hot stage. This, however, would not lead to an increase in the relic quark abundance because extra quarks would disappear, for example through the reaction $q + q \to B + \bar{q}$.

In some papers attempts have been made to obtain a smaller relic quark abundance. Nakamura *et al.* (1977) assumed that at an early stage ($t < 10^{-5}$ sec) the thermal history of the universe differed from that of the standard model. The basic idea is that after quark quenching a strong heating of the primeval plasma took place. The temperature reached was, however, smaller than m_q. Decays of heavy unstable neutrinos ν_h were assumed to be the energy source, the ν_h gas being strongly degenerate. For the relative chemical potential of ν_h the following estimate was obtained: $\xi_{\nu_h} \equiv \mu_{\nu_h}/T \gtrsim 10^8$ (for $m_q = 100$ GeV). Decays of ν_h increased the photon density but did not change the density of quarks (because $T < m_q$). In this way the quark concentration could be suppressed to the level of today's limits, $r_q = N_q/N_\gamma \simeq 10^{-30}$. In this model some degeneracy ($\mu/T \simeq 1$) of ν_e and/or ν_μ is predicted. A huge leptonic charge in the universe at the beginning seems to be mysterious, but if free quarks were discovered a modification of the standard scenario would deserve attention.

In a paper by Wagoner and Steigman (1979) the arguments in favor of a small relic quark abundance are presented. A brief summary is as follows. It is assumed that the quark interaction potential increases with distance up to some radius $r_C \gg 1$ fermi. Then the potential reaches the limiting value, which equals twice the mass of a free quark $2M$. The effective quark mass is small, however, when the density is high, and, correspondingly, the average distance between neighboring quarks is small.[5] The quark quenching temperature is defined by this smaller mass, or more precisely, by the characteristic hadronic scales, i.e., $T_f \simeq m_\pi$. At $T = T_f$ almost all quarks condense into hadrons. Wagoner and Steigman assume that only those quarks remain free whose energy is greater than their mass, $E > M$. The equilibrium distribution of quarks being valid until $T = T_f$, the concentration of free quarks proves to be very small, $\sim \exp(-M/T_f)$ and for $M = 15$–30 GeV it is possible to obtain $n_q/n_B \simeq 10^{-20}$.

In our opinion this conclusion is not obligatory. Because of statistical fluctuations the color could be noncompensated in a volume V_0 which is on the one hand large as compared to the quark capture volume ($V_C = r_C^3$), and on the other hand sufficiently small so that the relative size of fluctuations is large enough that $\Delta N/N_{\rm tot} \gtrsim 10^{-20}$ (here $N_{\rm tot}$ is the total number of particles in the volume V_0 and ΔN is the color excess in this volume). It is noteworthy that due to the saturation of the quark force at $r > r_C$, fluctuations develop in the same way as in an ideal gas. However, even much smaller fluctuations are sufficient for our purposes. If the separation distance between quarks is smaller than r_C, they form hadrons when the temperature drops

[5] The picture of confinement is not yet understood, however, and quark behavior is probably quite different.

below some value $T = 300-500$ MeV. But there should be a rather large excess of color in volume $V_0 \gg r_C^3$ and the corresponding quarks should remain free because they can find a partner for forming a white hadron only far away at the distance $r \approx V_0^{1/3} \gg r_C$. The energy necessary for liberation could come from the energy released in the course of hadron production. The question can be raised, however, whether the statistical fluctuations in volume V_0 of the primeval plasma had enough time to develop. The answer is yes. Indeed, evaluating the quark diffusion distance in the time $t(T = 1 \text{ GeV}) \simeq 10^{-6}$ sec as $\lambda = \sqrt{t/T} \simeq 10^{-4}$ cm, we find that fluctuations can develop in the volume $V = \lambda^3$, which is much larger than V_C. In conclusion we should like to note that the whole problem is not as yet clear and deserves further investigation.

The above mentioned difficulties with the relic quark abundance are inherent also in possible new stable hadrons (Wolfram, 1979; Dover *et al.*, 1979). The ratio of the new stable particle density to that of baryons should be

$$r_X/r_B = (m_N/m_X)(\sigma_N v_N/\sigma_X v_X) \tag{5.2}$$

in the case of charge symmetry with respect to baryons and X particles. The baryon asymmetry of the universe leads to a drastic increase in r_B. Thus if there were no X asymmetry the ratio (5.2) would be 10^{10} times smaller. The latter result is close to the existing experimental bounds. It follows that the mechanisms producing baryon asymmetry should not work for the new stable hadrons if they existed.

Using Eq. (2.11) it is easy to calculate the contemporary energy density of X particles:

$$\rho_X = m_X r_X N_\gamma = 3\times 10^{-14} N_{DF}^{1/2} (\sigma_X v_X m_N^2)^{-1} (\tilde{T}_\gamma/T_\gamma)^3 (m_N \text{ cm}^{-3}), \tag{5.3}$$

where the factor $(\tilde{T}_\gamma/T_\gamma)^3$ takes into account plasma heating after the decoupling of X particles. From the condition $\rho_X < \rho_{\max}$ one obtains the lower bound on the $X\bar{X}$ annihilation cross section

$$v_X \sigma_X > C \times 10^{-36} \text{ cm}^2 , \tag{5.4}$$

where $C = 2N_{DF}^{1/2}(\tilde{T}_\gamma/T_\gamma)^3$. This factor depends upon m_X and upon the strength of X particle interactions which keep them in thermal equilibrium with the plasma. For reasonable values of the parameters, C is not far from unity.

It is noteworthy that accelerator experiments which search for new particles give upper bounds on their cross sections. Cosmology permits us to limit the cross sections from below.

VI. HEAVY LEPTON MASSES

A study of the properties of the new heavy lepton τ shows that as it decays a new type of neutrino ν_τ is produced. In other words τ and ν_τ possess a new conserved charge analogous to electronic or muonic charges. The existing laboratory data give a rather weak bound on the ν_τ mass $m_{\nu_\tau} < 250$ MeV. So in principle ν_τ could be rather heavy. Cosmological arguments lead to a much stronger limitation on m_{ν_τ}, that of the inequality (4.2). Limit (4.2) was obtained for rather light neutrinos, i.e., those with $m < 3$ MeV. With an increase in the neutrino mass the density of relic neutral leptons drops and the condition $\rho_\nu < \rho_{\max}$ holds if m_ν is high enough. This upper value of the mass is happily of the order of some GeV. So the only possibility left for ν_τ is (4.2). In what follows we determine an upper bound for the forbidden mass region (Vysotsky, Dolgov, and Zeldovich, 1977 a, b; Lee and Weinberg, 1977; Hut, 1977; Sato and Kobayashi, 1977). Our consideration is valid for a neutral stable lepton interacting in accordance with the Weinberg–Salam model.

The $L\bar{L}$ annihilation which determines the relic L concentration in accordance with Eq. (2.11) proceeds through the neutral current interaction. To evaluate the cross section value we take into account that L's become nonrelativistic at the moment thermodynamic equilibrium breaks down. The cross section of $L\bar{L}$ annihilation (at rest) into light particles is of course proportional to v^{-1} and in the Weinberg–Salam model is equal to

$$\sigma v = \frac{G_F^2 m_L^2}{6\pi} C , \tag{6.1}$$

where $G_F = 10^{-5} m_N^{-2}$ is the Fermi coupling constant and C is determined by the number of open annihilation channels. If one takes into consideration the processes $L\bar{L} \to \nu_e\bar{\nu}_e,\ \nu_\mu\bar{\nu}_\mu, e^+e^-,\ \mu^+\mu^-,\ u\bar{u},\ d\bar{d}$, and $s\bar{s}$ then $C = 9$. It was assumed that $\sin^2\theta_W = 0.25$ and that there were three quark colors. As we shall see below, the boundary value of m_L is about 3 GeV. The quenching freeze-out) temperature of the massive neutrino in this case is equal to 150 MeV. At such a temperature there are $\gamma,\ e^\pm,\ \mu^\pm,\ \nu_e,\ \nu_\mu,\ \pi^\pm,\ \pi^0$ in thermodynamic equilibrium and $N_{DF} = 31/2$. To evaluate the present energy density of L one needs to take into account photon gas heating by the annihilation of e^+e^-, $\mu^+\mu^-$, and $\pi^+\pi^-$, and by π^0 decays. Because of all of these the number density of photons became $\frac{31}{2}\,\frac{1}{9}\,\frac{11}{4} \simeq 5$ times higher, and correspondingly the present ratio of relic L's to photons became five times smaller than that given by Eq. (2.11). Taking this into account and substituting Eq. (6.1) into Eq. (2.11) one finds for the present energy density of L

$$\begin{aligned}\rho_{L0} + \rho_{\bar{L}0} &= 0.4 r_{0L} m_L N_{\gamma 0}\\ &\simeq 100\left(\frac{m_N}{m_L}\right)^2\left(1 + 0.2\,\frac{m_L}{m_N}\right) \text{keV cm}^{-3} .\end{aligned} \tag{6.2}$$

The energy density ρ_L is inversely proportional to the lepton mass squared owing to the increase of the annihilation cross section with increasing m_L. Demanding that $\rho_{L0} + \rho_{\bar{L}0}$ should be smaller than $\rho_{\max}$ we obtain $m_L > 4$ GeV. So no neutral stable lepton can exist in the mass interval 40 eV $< m_L <$ 4 GeV.

The freezing temperature is determined from Eq. (2.10) as

$$T_f/m_L = \tfrac{1}{20} . \tag{6.3}$$

From this it follows that leptons L are indeed nonrelativistic at the instant of freeze-out and that our estimates of C and of the plasma heating are correct.

Equation (6.1) for the annihilation cross section is valid in the local four-fermion limit when $m_L < m_Z$ where Z is the neutral intermediate boson mediating weak interactions. With increasing m_L the cross section should reach maximum value and then decrease. That is why there must be an upper bound on m_L. If only processes with Z exchange are taken into account the corresponding upper bound is $m_L < 3$ TeV. So in this case the following mass intervals are permitted: $0 < m_L < 40$ eV, and 4 GeV $< m_L < 3$ TeV. If, however, the mass of L is generated by a vacuum expectation value of the Higgs field χ, the interaction between L and χ should be strong for large m_L and no conclusion can be drawn in such a simple way.

If heavy neutral leptons indeed exist and if they could be captured by stellar systems, their energy density today would be considerably smaller than that predicted by Eq. (6.2). The increase in density of L because of this capture could lead to the appearance of secondary annihilation. So fewer L would survive to our time. At first glance this would seem to make the limit on L much worse. Actually the situation is just the opposite (Vysotsky *et al.*, 1977). The point is that about half of the entire energy released in annihilation ultimately is converted into electromagnetic radiation. The energy density of the latter is known to be much smaller than $\rho_{\max}$. The density of the relic radiation is $4\times 10^{-34}\,\mathrm{g\,cm^{-3}}$, that of optical radiation is $4\times10^{-36}\,\mathrm{g\,cm^{-3}}$, and that of x rays is $10^{-37}\,\mathrm{g\,cm^{-3}}$. Thus despite the energy of L decay products being $(Z+1)$ times less than ρ_{L0}, due to the redshift, the advantage of the stronger observational bounds on ρ_{em} compensates the loss. If a fraction β of the heavy leptons annihilates at a certain Z, the total energy density of L, $\bar{L}$, and their decay products at the present time will be

$$\rho_{\mathrm{tot}} = 2(1-\beta)\rho_{L0} + \frac{2\beta\rho_{L0}}{Z+1}\,, \tag{6.4}$$

where ρ_{L0} is defined by Eq. (6.2). Protostar formation had to take place during the epoch of matter dominance when $(1+Z_{cr}) \simeq 10^4\Omega(H/60)^2$. So in Eq. (6.4) $Z < 10^4$. Even for the maximum value of Z a stronger limit on m_L than that obtained above follows from the condition $2\beta\rho_{L0}(1+Z)^{-1} < \rho_{e.m.}$ for say, $\beta > 0.1$.

This result, however, is obtained under the assumption that secondary annihilation occurs sufficiently late so that the photons resulting from the annihilation are not thermalized in the primeval plasma. As the formation of stars takes place after the plasma becomes neutral, this assumption is indeed true. There is, however, another possibility. At an early stage in the development of the universe, the energy density could be dominated by nonrelativistic heavy leptons. In such a case leptonic star formation is possible. Increased annihilation in such localized concentrations of leptonic matter would not be visible in the spectrum of electromagnetic radiation today but would lead to a considerable decrease in the density of L. If this were the case, the upper limit on m_L would be much weaker. For low values of the mass of L the annihilation inside these stars should be complete, otherwise the condition $\rho_L < \rho_{\max}$ would not be satisfied. For example, if $m_L \simeq 100$ MeV, not more than 1% of the total leptons could survive the secondary annihilation.

Leptonic matter would predominate when the temperature dropped lower than $T_D = 10^{-7}(m_N/m_L)^2 m_N$ (for $m_L > 3$ MeV). So in the case of $m_L < 10 m_N$ leptons would dominate the energy density before ordinary celestial bodies appeared. At $T < 1$ MeV leptons would not interact with the plasma and their temperature would drop faster than the plasma temperature, $T_L \simeq T^2_{\mathrm{plasma}}/m_L$. So the leptonic gas would cool quickly, and as a result of gravitational instability heavy leptons could form primary inhomogeneities in the system. These inhomogeneities could serve as condensation centers for aggregates of ordinary matter. We have not yet considered this problem in detail, however. The behavior of leptonic matter in the case where leptons constitute the bulk of the matter in the universe deserves further investigation.

Leptonic condensation because of gravitational capture by stellar or galactic size objects was briefly discussed by Gunn *et al.* (1978). They concluded that this process was unlikely or even impossible. Their view was later reconsidered by Zeldovich *et al.* (1980), who showed that gravitational binding of massive leptons could take place because of time variation in the gravitational field of the collapsing gas. The corresponding increase in annihilation could disturb the spectrum of electromagnetic cosmic radiation. From the absence of noticeable anomalies in this spectrum the limit $m_L < 100$ GeV was obtained. This result refers, however, to a rather late period ($Z < 5$) and to the case of a low lepton density as compared to the total matter density. The case of high lepton density calls for special investigation. Nevertheless, as was noted above, the necessary annihilation of almost all leptons inside lepton stars seems improbable. So the bounds on m_L obtained in this section should not be considerably changed.

VII. UNSTABLE LEPTONS. LIMITS ON THEIR MASS AND/OR LIFETIME

The bounds on neutral lepton mass obtained in Sec. VI were based on the lower limit on the age of the universe and the corresponding upper limit on the energy density $\rho_{\max}$. These results are valid only for stable particles or, more precisely, for particles whose lifetime exceeds the age of the universe: $t_0 = 3\times10^{17}$ sec. With more refined considerations a wide range of lifetime values versus mass can be excluded. In short, the bounds obtained are based on a study of the influence of hypothetical unstable leptons and their decay products on

1. the total energy density in the universe;
2. the cosmic electromagnetic radiation spectrum,
3. primordial nucleosynthesis;
4. stellar evolution

In what follows we shall discuss these in more detail.

A. The total energy density in the universe[6]

A heavy neutral lepton could have the following decay channels: $L \to \nu\nu\bar{\nu}$, $\nu\gamma$, $e^+e^-\nu$, etc. The number density

[6]For further discussion see Vysotsky, Dolgov, and Zeldovich, 1977a,b; Sato and Kobayashi, 1977; Dicus, Kolb, and Teplitz, 1977, 1978; Goldman and Stephenson, 1977.

of electrons N_e resulting from the decay $L \to e^+e^-\nu$ is determined by the density of heavy leptons (2.11) and by the branching ratio of this decay, B_e. For any non-negligible B_e density, N_e proves to be so high that the e^+e^- annihilation rate $N_e v_e \sigma(e^+e^- \to 2\gamma)$ exceeds the expansion rate. So all the energy from L decay is transformed ultimately into massless particles. In contrast with the case of a stable L the energy density of the decay products decreases further because of the redshift.

The variation of the temperature with time in the course of the expansion is defined by $T(t_1)=T(t_2)(t_2/t_1)^\lambda$, where $\lambda=\frac{2}{3}$ for the matter-dominated universe and $\lambda=\frac{1}{2}$ for the radiation-dominated universe. In standard cosmology radiation dominance is succeeded by matter dominance at $t=t_m=3\times10^{11}\Omega^{-2}$ sec (for $H=60$ km sec^{-1} Mpc^{-1}). If a heavy neutral lepton L exists, the picture is more complicated. The contribution of a stable L to the energy density at a redshift Z in accordance with the previous calculations is

$$2\rho_L(Z)=(Z+1)^3\rho_{\max}f(m_L)\,, \tag{7.1}$$

where

$$f(m_L)=(4m_N/m_L)^2 \quad \text{if } m_L>3 \text{ MeV}$$

and

$$f(m_L)=(m_L/40\,\text{eV}) \quad \text{if } m_L<3 \text{ MeV}\,.$$

There is no exact agreement at $m_L=3$ MeV because neither expression for $f(m_L)$ is approximate. Matter dominance begins when $2\rho_L(Z'_m)=(Z'_m+1)^4\rho_\gamma(0)$, where $\rho_\gamma(0)$ is the energy density of the electromagnetic background radiation at the present time. Hence

$$Z'_m \approx \frac{\rho_{\max}}{\rho_\gamma(0)}f(m_L)=2.5\times10^4 f(m_L)\,. \tag{7.2}$$

If L is stable, $f(m_L)<1$ and the expansion regime is changed at about the same time as in the standard case, $t_m=3\times10^{11}\Omega^{-2}$ sec.

For unstable L, $f(m_L)$ can exceed 1 and leptonic matter could dominate the total energy density earlier than ordinary matter ($t'_m<t_m$) if the lepton lifetime was large enough ($\tau_L>t'_m$). After L decayed, i.e., when t was larger than τ_L, radiation could dominate again. Depending on m_L and τ_L this period could last up to the present. The contribution of heavy unstable leptons and their decay products to the energy density is

$$\tilde\rho_L(Z)=\rho_L(Z)e^{-t/\tau}+\int_0^t\frac{dt'}{\tau}e^{-t'/\tau}\rho_L(Z')\left(\frac{Z+1}{Z'+1}\right)^4, \tag{7.3}$$

where $\rho_L(Z)$ is the energy density of the stable leptons at the moment corresponding to a redshift Z as defined by Eq. (7.1); t is the time counted from the instant of the singularity. t and Z are related by $t=t_0(1+Z)^{-3/2}$ for $t_m<t<t_0$ (t_0 is the age of the universe) and $t=t_m(1+Z)^{-2}$ for $t<t_m$. Clearly the expression for t can be modified to include the possible existence of heavy leptons. In standard cosmology t_m is determined by the condition $\rho_B=\rho_\gamma$, where $\rho_B=N_Bm_B$ is the total energy density of baryons and $\rho_\gamma=N_\gamma T$ is the energy density of the relic photons. Substituting $\beta=N_B/N_\gamma=10^{-8}$–$10^{-10}$ into this condition, we find $T_m=T(t_m)=0.1$–10 eV. Hence $Z_m=T_m/3\text{ K}=3\times10^{3\pm1}$ and $t_m=t_0(1+Z_m)^{-3/2}=t_0(10^{-4}$–$10^{-7})$, where t_0 is the age of the universe. If a heavy unstable L exists, the expansion regime can be different from the standard one. However, this complication does not significantly affect the boundaries of the forbidden region in Fig. 1.

From the condition [see Eq. (7.3)] that $\tilde\rho_L(0)$ is smaller than $\rho_{\max}$, a restriction on τ_L for fixed m_L can be found. In Fig. 1 the forbidden region in the (τ_L, m_L) plane is shown (curve 1). Line *aa* indicates lifetime as a function of mass for the standard weak interaction: $\tau_L=2\times10^{-6}(m_L/m_\mu)^{-5}$ sec.

B. The spectrum of cosmic electromagnetic radiation[7]

If $\tau_L>10^{12}$ sec, the photons originating from L decay do not interact with the cosmic plasma because by this time, the density has decreased and the plasma become transparent to electromagnetic radiation. Thus a considerable distortion in the spectrum of cosmic radiation could appear. In this case cosmological arguments exclude for some values of τ_L an even larger mass range than for a stable lepton. This is so because the photons from L decays can be directly observed (in contrast to L itself) and that is why the observational bounds for them are more restrictive. The decay $L\to\nu\gamma$ directly changes the spectrum of cosmic γ quanta. The decay $L\to e^+e^-\nu$ distorts the spectrum because of the subsequent annihilation $e^+e^-\to2\gamma$ or because of Compton scattering of the electrons on cosmic photons. Only the decay $L\to\nu\nu\bar\nu$ creates unobservable particles in the final state and does not influence the cosmic electromagnetic radiation. If $m_L>2m_e$ the probability of decay into $e^+e^-\nu$ should be of the same order as that into $\nu\nu\bar\nu$. If $m_L<2m_e$, the decay mode $\nu\nu\bar\nu$ should be dominant. These statements, however, are model dependent.

Thus the results obtained in this subsection are valid in the case when photons are produced in a significant fraction of L decays. This assumption is natural if m_L is large, and probably is wrong if $m_L<2m_e$.

If decays of L produce photons with energy κm_L ($\kappa<1$) a limit on particle lifetime versus mass can be obtained from the condition

$$\frac{B\kappa\rho_L(Z)}{(Z+1)^4}<\rho_\gamma\left(\frac{\kappa m_L}{Z+1}\right). \tag{7.4}$$

[Compare with Eq. (7.3).] Here B is the branching ratio of the corresponding decay, the energy density $\rho_L(Z)$ is defined by Eq. (7.1), and $\rho_\gamma(\omega)$ is the energy density of the cosmic electromagnetic radiation with frequency ω as measured today. For energy density of, say, x rays, $\rho_\gamma<10^{-8}\rho_{\max}$, so better restrictions on the proproperties of L can be obtained than those obtained in the previous subsection where the condition $\tilde\rho<\rho_{\max}$ was used.

For $\tau_L<10^{12}$ sec, the photons and charged particles produced by decays of L are in thermal contact with the cosmic plasma and the "kinetic" equilibrium is rather quickly established. If, however, $\tau_L>10^4$ sec, there is insufficient time for establishment of the

[7]For further discussion see Sato and Kobayashi, 1977; Lee, Lerche, Schramm, and Steigman, 1978; Dicus, Kolb, and Teplitz, 1978; Cowsik, 1977.

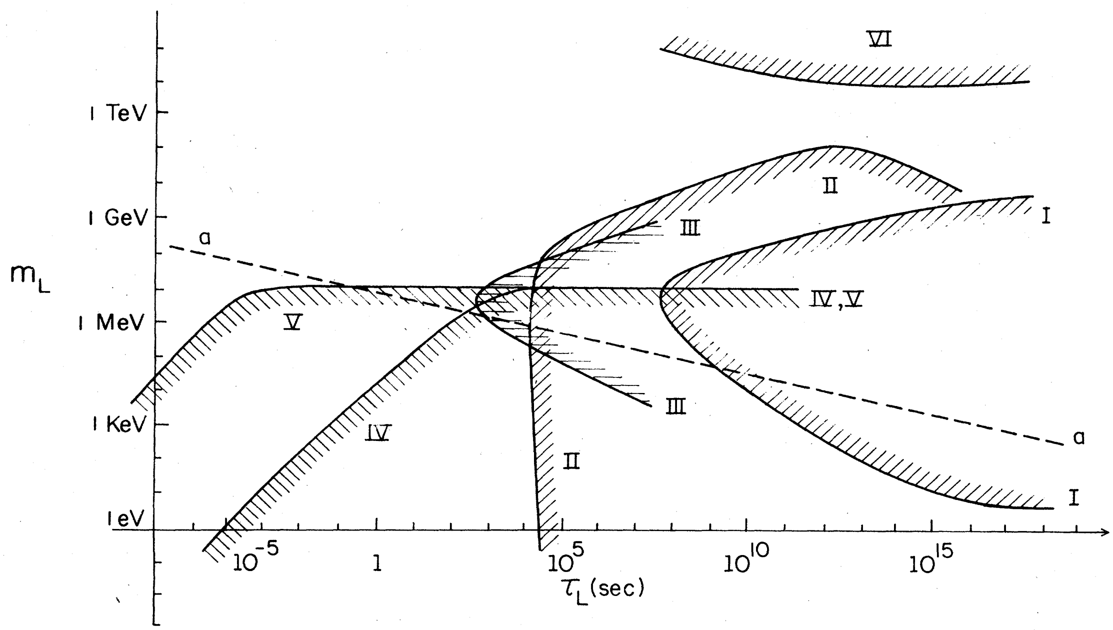

FIG. 1. Astrophysically forbidden region of lifetimes and masses of a neutral lepton (shaded). The curves are obtained by the analysis of (I) the lower bound on the age of the universe and the corresponding limit on the total energy density in the universe; (II) the spectrum of the cosmic electromagnetic radiation; (III) the primordial nucleosynthesis; (IV) γ-ray spectrum from supernovae; (V) energetics of supernovae. All the curves except (I) are based on the assumption that in a considerable part of the L decays photons are ultimately produced. Restriction (I) is valid also in the case when L dominantly decays into $\nu\nu\bar{\nu}$. Region (VI) is obtained in the same way as (I) and refers to neutral leptons if they interact through exchange of vector bosons of the Weinberg–Salam model. Using arguments similar to that of cases (II) and (III) the forbidden region for superheavy leptons can be enlarged. Line aa is the lifetime dependence on mass for the standard weak interaction.

"chemical" equilibrium. The point is that "kinetic" equilibrium is achieved through the fast processes of Compton scattering, and "chemical" equilibrium through the much slower processes of thermal photon emission and absorption. This question has been considered in more detail by Zeldovich and Novikov (1975) (see also Sunyaev and Zeldovich, 1970). They stated that a distortion in the blackbody spectrum of relic photons would be observable up to $Z = 10^8$ if the relative energy release were of the order of 10%. This leads to the limit $\tau_L < 3\times 10^3$ sec. The same result was obtained by Gunn *et al.* (1978).

Later R. Sunyaev (1980) showed that the process $\gamma e \to 2\gamma e$ smoothed over distortions in the spectrum until $Z \simeq 10^7$ or 3×10^4 sec. This process has a smaller cross section than bremsstrahlung in ep collisions. The smallness is, however, compensated by the greater number density of photons. If $\Omega < 0.1$, the process $\gamma e \to 2\gamma e$ dominates. For larger Ω, photon creation by $ep \to ep\gamma$ is also important. For $\Omega = 1$, the Planck spectrum would be restored by this reaction if $Z > 2\times 10^7$. The above mentioned shift of the critical value of Z reflects a change in the input parameters from those used in the original paper by Sunyaev and Zeldovich (1970); that is why the double Compton scattering comes into play. So when $t > 3\times 10^4$ sec the number of photons in the plasma changes rather slowly and the production of new photons by the decay $L \to \gamma + \cdots$ leads to the Bose–Einstein energy spectrum of radiation;

$$F(\omega) \sim \omega^3 \left[\exp\left(\frac{\mu+\omega}{T}\right) + 1\right]^{-1},$$

the chemical potential μ being nonzero.

The deviations of this expression from the blackbody spectrum are the most prominent at small frequencies $\omega < T$. In this region the spectrum of the electromagnetic background radiation has been very well studied and no deviations from the Planck formula (that is from the Bose–Einstein formula with $\mu = 0$) have been found. These arguments permit us to set the limit $\tau_L < 3\times 10^4$ sec (see curve II in Fig. 1). With rising m_L the bound becomes weaker because the number density of heavy leptons is a decreasing function of m_L.

C. Primordial nucleosyntheses [8]

The bound on the number of neutrino types obtained by analyzing He^4 production in the early universe (see Sec. III) suggests using analogous arguments for restrictions on heavy neutral lepton properties. Heavy leptons would affect helium formation in the following two ways:

1. by an increase of the expansion rate $H \sim \dot{a}/a$ and, because of this, by a larger ratio of neutron to photons; and

[8]For further discussion see Sato and Kobayashi, 1977; Dicus, Kolb, Teplitz, and Wagoner, 1977; Miyama and Sato, 1978.

2. by an increase in total nucleon number density at the time of nucleosynthesis (or a decrease in the entropy per baryon) as compared to that accepted in the standard model. The entropy decrease is connected with heating in the plasma by the decays of L into $\nu\gamma$, νe^+e^-, $\nu\mu^+\mu^-$, etc., which occur after the nucleosynthesis is over. So the contemporary value of the ratio N_B/N_γ would be smaller than this ratio at the time of nucleosynthesis. In the standard approach it is assumed that the ratio N_B/N_γ does not change from the moment of nucleosynthesis to our time except for some change in the photon number by e^+e^- annihilation.

Both of these effects (increase of ξ and of N_B/N_γ) lead to an increase in He^4 formation. This increase, however, is rather small for the interesting values of τ_L and m_L [an assessment which is in agreement with that of Gunn *et al.* (1978) and Dicus *et al.* (1977)]. To evaluate the effect of a variation of $\xi=(\rho/\rho_{\text{standard}})^{1/2}$ we recall that modern data on He^4 abundance permit six new types of neutrinos to exist [for the conservative estimate of the He^4 mass fraction $R(He^4)<0.29$]. As was discussed in Sec. III the parameter ξ determines the value of the ratio N_n/N_p at quenching (freeze-out), the quenching temperature T_f being about 1 MeV. Hence the ratio of the contributions to ρ from a heavy lepton and from k types of new neutrinos is $\eta=2N_L m_L/kN_\nu T_f$. For $k=6$ the value of η is less than 1 for any m_L so that a change of ξ because of the contribution from L does not lead to a considerable variation in He^4 production.

One can easily calculate the change of the ratio $\beta=N_B/N_\gamma$ if one takes into account that the entropy in a comoving volume is conserved [see for example, the book by Weinberg (1972)].

If $\delta\beta$ is a contribution to β from the decay of L, the following relation is valid:

$$\frac{\delta\beta}{\beta_0}=\text{const}\times\frac{m_L}{T}r_L\ ,$$

where const$=0(1)$, r_L is determined by Eqs. (2.11) and (6.2), and T is the temperature at the instant of the decay. The uncertainty in the value of β_0 is as large as two orders of magnitude, giving rise to an uncertainty in the mass fraction of He^4, $\delta R(He^4)\simeq 0.04$ which is rather small. So in the value of the ratio $\delta\beta/\beta_0$ a variation by two orders of magnitude is allowable. The product $m_L r_L$ reaches its maximum value at $m_L\simeq 1$ MeV. Even in this case the plasma temperature at the instant of the decay should be greater than $0.01\times m_L\simeq 10^{-2}$ MeV and consequently $\tau_L<10^4$ sec. This is close to the result obtained in the preceding subsection.

Thus the study of He^4 production does not lead to a significant expansion of the forbidden region in the (τ_L, m_L) diagram (Fig. 1). The abundance of cosmological deuterium is a much more sensitive indicator of the existence of heavy leptons because in contrast to He^4, the H^2 abundance is a steep function of the present baryon density.

It is well known that the prediction of the He^4 abundance is one of the triumphs of big bang cosmology. As the calculations of other element abundances were made in the same framework, the results obtained seem to be reliable. From modern observations and estimates of the nuclear processes inside stars, it follows that the mass fraction of primordial deuterium should be larger than 10^{-5}.

An increase in the expansion rate due to L contributions would lead to a larger amount of H^2 because in this case deuterium nuclei would have less time to transform into other elements. Such a change in the expansion rate is the only effect which could change the deuterium abundance if L were stable or decayed into $\nu\nu\bar{\nu}$. This effect is rather small, however, and no restriction on the lifetime of L can be obtained. If, however, L decayed into $\nu\gamma$, νe^+e^-, etc., the ratio $\beta=N_B/N_\gamma$, as we noted above, would change from the moment of nucleosynthesis to the present time, and the deuterium fraction is very sensitive to this parameter. Calculations of the influence of L on H^2 abundance were made by Dicus *et al.* (1977). Their results are presented in Fig. 1 (curve III).

If $m_L>2m_N$ it makes possible the decay $L\to B\bar{B}+$ anything. The relative number density of antibaryons is $N_{\bar{B}}/N_B=(1-100)(2m_N/m_L)^3 BR(L\to B\bar{B}+\text{anything})$. For $\tau_L>100$ sec $\bar{B}$ produced by the decay could directly influence nucleosynthesis.

D. Stellar evolution [9]

If the neutral lepton mass is less than the temperature inside stars and it is unstable, L can strongly affect stellar evolution. The astronomical observations in this case give some restrictions on the properties of such neutral leptons. These restrictions are, of course, applicable to a rather light L but their advantage is that one need not travel so deeply into the past as in the preceding cases. In particular the limits presented below are valid in the hot model as well as in the cold one.

A rather strong bound on the lifetime of the electronic neutrino in the case of its decay into a photon plus anything can be obtained from an analysis of the x-ray radiation from the sun (Cowsik, 1977). The modern theory of nuclear reactions inside the sun predicts the solar neutrino flux to be about 10^{11} cm^{-2} sec^{-1} with the average neutrino energy 200 KeV.[10] The neutrino decay would produce an intense x-radiation. But modern observations find the x-ray flux to be less than 10^{-4} cm^{-2} sec^{-1}. This results in the limits $\tau_{\nu_e}>5\times10^{17}$ sec or $\tau_{\nu_e}<1$ sec. In the latter case neutrinos would mainly decay inside the sun; however, the laboratory measurements exclude this possibility.

A stronger limitation and, moreover, one valid for any type of neutrino with $m<10$ MeV, can be obtained

[9]For further discussion see Cowsik, 1977; Falk and Schramm, 1978.

[10]The experiment by the Davis group shows that the actual flux of solar neutrinos is about $\frac{1}{3}$ of the calculated value. The bound on τ_ν given here is not considerably changed if one uses the experimental result instead of the theoretical one. Recently the ITEP group presented evidence in favor of a nonvanishing neutrino mass (Lyubimov *et al.*, 1980 and Kosic *et al.*, 1980). If such is indeed the case, one could easily believe in neutrino oscillations and so the contradiction between the observations of solar neutrinos and the theory could be resolved.

from the study of supernovae. We recall that a supernova explosion releases about 10^{53} erg of energy in the form of neutrino radiation, the average neutrino energy being ~10 MeV. Decays of these neutrinos would yield a large γ-ray flux. For the latter the observational limit is, however, 10^{-3} cm^{-2} sec^{-1} sterad^{-1}. So a neutral lepton, if it exists, must be rather long-lived, $\tau_L > 3 \times 10^{23}$ sec, or must decay while still inside the star. In the latter case $\tau_L < 10^3$ sec$/\varepsilon$ where $\varepsilon = E_L/m_L \simeq 10$ MeV$/m_L$ and the coefficient 10^3 is connected with the size of a red giant which precedes the supernova; this size is larger than 3×10^{13} cm. It is noteworthy that laboratory limits on τ_L are valid only for ν_e and ν_μ.

The bound on τ_L from above can be further improved (Falk and Schramm, 1978) if one takes into account the fact that the visible energy released by a supernova explosion is about 10^{51} erg. This means that not more than 1% of the total number of leptons emitted by the explosion can decay outside the core. This in turn means that $\tau_L < R_C/\varepsilon \simeq 10$ sec$/\varepsilon$, where R_C is the radius of a supernova core. The bounds on τ_L, which are dependent on m_L, are presented in Fig. 1 (curves IV and V).

VIII. ASTROPHYSICAL BOUNDS ON ELECTROMAGNETIC PARAMETERS OF NEUTRINOS

If a right-handed neutrino existed, a nonvanishing magnetic moment for it would, in principle, be possible. Moreover, gauge theories with right-handed currents predict a nonzero magnetic moment. If the matrix element of the neutrino interaction with the electromagnetic current is of the form

$$\langle \nu(p_2) | j_\alpha^{em} | \nu(p_1) \rangle = \frac{e\kappa}{2m_e} \bar{u}_2 \sigma_{\alpha\beta}(p_2 - p_1)_\alpha u_1 \tag{8.1}$$

for the value of κ the following estimate is valid:

$$\kappa = \frac{G_F m_e m_M}{\pi^2} \sin\varphi = 2.5 \times 10^{-13} \frac{m_M}{m_e} \sin\varphi \,, \tag{8.2}$$

where m_e is the electron mass, m_M is the mass of the charged lepton connected with the given type of neutrino, and $\sin\varphi$ characterizes the admixture of right-handed currents in leptonic interactions. For $\mu\nu_\mu$ and $e\nu_e$ currents the limit is known to be $\sin\varphi < 0.1$.

The laboratory limits on the value of κ have been collected (Beg *et al.*, 1978) and are about 10^{-9} for ν_e and about 10^{-8} for ν_μ. Astrophysics helps us to improve upon these limits by 1–2 orders of magnitude. The arguments presented below were first stated by Bernstein *et al.* (1963) and later by Sutherland *et al.* (1976) and used in the paper by Beg *et al.* (1978). In the dense matter inside stars, photons acquire an effective mass of the order of the plasma frequency. Owing to this, interaction (8.1) would lead to photon decay into $\nu\bar{\nu}$ and energy losses from stars would be considerably enhanced. If one assumes that the neutrino luminosity of the sun is no greater than the photon luminosity, the limit $\kappa < 10^{-9}$ can be obtained. This limit is valid for any neutrino type with a mass smaller than 150 eV. The latter restriction arises because the plasma frequency in the central region of the sun is about 300 eV. As for the initial assumption about the neutrino luminosity of the sun, its violation would lead to a rate of hydrogen burning inside the solar core at least twice that assumed in the standard models. This can hardly be accepted. The point is that, according to modern estimates, about half of the total amount of hydrogen inside the sun should be burned by the present time. The doubling of the energy loss would mean that no hydrogen would now be left inside the solar core.

A stronger limit on the value of the neutrino magnetic moment can be obtained by analysis of white dwarf evolution. A detailed discussion of this question can be found in Chiu *et al.* (1966), Stothers (1970), and Bernstein *et al.* (1963). The achieved limit, valid for neutrinos with masses up to 20 KeV, is $\kappa < 10^{-10}$. It is claimed (Beg *et al.*, 1978) that a more accurate study of white dwarfs will permit us to improve this limit by an order of magnitude.

It should be noted that the astrophysical bounds on κ for ν_e and ν_μ are far from the expected value [Eq. (8.2)], but for ν_τ the bound on $\sin\varphi$ is better than the one obtained by experiments on τ decays. This is partly a consequence of less accurate experimental information about properties of τ and partly of a large value for m_τ.

Clearly the arguments presented above can also limit the exotic possibility of nonzero electric charge for the neutrino. As the process of transformation of photons into neutrinos is defined by the probability of photon decay into $\nu\bar{\nu}$, a limit on the neutrino charge follows from the substitution $\kappa \to e(\nu) \cdot 2\sqrt{2}\, m_e/\omega_p$, where $e(\nu)$ is a (hypothetical) neutrino charge, ω_p is the plasma frequency, and m_e is the electron mass. Hence, from the bound on κ, one easily obtains $e(\nu) < 10^{-13} e(e)$, where $e(e) \equiv e$ is the electron charge.

For electronic neutrinos a stronger limit $e(\nu_e) < 10^{-17} e$ (Zorn *et al.*, 1963) follows from electric charge conservation because during the β decay of neutrons, $e(n)$ and $e(p) + e(e)$ can be independently measured (see, however, Sec. XIII). For muonic neutrinos the laboratory bounds are much weaker, $e(\nu_\mu) < 10^{-5} e$ [see the discussion by Bernstein *et al.* (1963)].

To believe in the existence of neutrino charge is difficult even without the astrophysical bounds, but in any case the neutrino charge radius, that is the vertex $\gamma\nu\bar{\nu}$ for $q^2 \neq 0$, should be nonvanishing:

$$\langle \nu | j_\alpha^{em} | \nu \rangle = \tfrac{1}{6} \langle r^2 \rangle q^2 \bar{\nu} \gamma_\alpha \nu \,.$$

In the framework of gauge models of weak and electromagnetic interactions the following estimate for the value of the neutrino charge radius can be obtained: $\langle r^2 \rangle = G_F/2\pi \simeq 10^{-34}$ cm^2. Data on neutral current interactions give $\langle r^2 \rangle \approx 10^{-30}$–$10^{-31}$ cm^2. The astrophysical bound on $\langle r^2 \rangle$ is easy to find by recalculating the bound on κ with help of the substitution $\langle r^2 \rangle \longleftrightarrow 2\kappa/\omega_p m_e$. As $\omega_p \simeq 300$ eV for the sun, it follows from the restriction on solar neutrino luminosity that $\langle r^2 \rangle < 10^{-28}$ cm^2 and, from the consideration of white dwarfs for which $\omega_p = 50$ KeV, that $\langle r^2 \rangle < 10^{-30}$ cm^2. The results discussed in this section are collected in Table II.

TABLE II. Limits on electromagnetic parameters of neutrinos.

		κ	$e(\nu)/e$	$\langle r^2\rangle$ cm^2
Laboratory limits	ν_e	10^{-9}	4×10^{-17}	10^{-30}
	ν_μ	10^{-8}	3×10^{-5}	10^{-31}
Limits from the neutrino luminosity of the Sun	$\nu_{e,\mu}$	5×10^{-10}	10^{-13}	10^{-28}
Limits from the neutrino luminosity of white dwarfs	$\nu_{e,\mu}$	10^{-10}	10^{-13}	10^{-30}

IX. NEUTRINO DEGENERACY

In the preceding section it was implicitly assumed that the universe is neutral with respect to any conserved quantum number such as electric charge Q, electronic charge E, muonic charge M, and a possible charge connected with the new lepton T. An exception is baryonic charge, which is known to be nonvanishing but very small $(N_B - N_{\bar{B}})/N_\gamma \simeq 10^{-8}$–$10^{-10}$. It is also known that the electric charge of the universe as a whole should be equal to zero with a very high precision because noncompensated charge leads to a long-range interaction which is 10^{18} times stronger than the gravitational interaction. More precisely, for the gravitational attraction to balance the electrostatic repulsion of two protons one has to add to each proton 10^{18} neutrons. The presence of a nonvanishing electric charge in the universe would contradict its homogeneity. For a closed universe the total electric charge should be strictly zero. (See also Sec. XII.) As for leptonic charges, no long-range forces are connected with them (see, however, Sec. X) and so observational data do not contradict their considerable values. In particular the possibility cannot be excluded that $|(N_{\nu_e} - N_{\bar{\nu}_e})|/N_\gamma \gtrsim 1$. The assumption that E, M, and T are vanishingly small seems, however, natural when we consider that the charges which can be measured, i.e., Q and B, prove to be very small.

Recently attempts have been made to calculate the baryon asymmetry of the universe in the framework of grand unified theories (see Sec. XV). The value obtained for baryonic charge in these models seems to be somewhat smaller than is needed, but it is possible that further attempts in this direction may give better results. It seems, however, absolutely improbable in such models, where neither baryonic nor leptonic charge is conserved, that the leptonic charge of the universe would be much larger than the baryonic charge.

Thus in big bang cosmology it is natural to expect that the excess of leptons over antileptons (or vice versa) should be of the same order of magnitude as the excess of baryons over antibaryons. Of course such a small leptonic charge would not influence the results of the preceding sections. In the rest of this section we reject the assumption that leptonic charges are small and discuss the observational limits on neutrino chemical potentials; we also consider how a considerable degeneracy of the neutrino gas would affect the conclusions reached above.

The best limit on the chemical potentials of all types of neutrinos follows from the bound on the total energy density of the universe (see Sec. II):

$$\sum_i \rho_{\nu_i} \simeq \frac{1}{8\pi^2} \sum_i \mu_{\nu_i}^4 < \rho_{\max} .$$

This expression is obtained under the assumption that $\xi_i \equiv \mu_{\nu_i}/T > 1$. From this it follows that $\mu < 0.01$ eV for any neutrino type. The boundary value of μ corresponds to the number density of neutrinos, about 10^6 cm^3. This is three orders of magnitude higher than the present number density of relic photons. If one takes into account that the temperature of relic neutrinos today is about 2 K (if $m_\nu = 0$) the following limit for ξ can be found (Weinberg, 1972):

$$\left(\sum \xi_i^4\right)^{1/4} < 45 . \tag{9.1}$$

We note here that the value of ξ should be a constant during the free expansion of the universe because the chemical potential of the neutrino is proportional to its temperature.

A stronger bound valid only for the chemical potential of the electronic neutrino can be obtained if one considers light element production in the early universe. A study of primordial formation of He^4, H^2, and Li^7 with the neutrino chemical potentials restricted by condition (9.1) and otherwise arbitrary has been made by Yahil and Beaudet (1976; Beaudet and Yahil, 1977) [see also the earlier paper by Wagoner *et al.* (1967)]. The degeneracy of ν_μ, ν_τ, etc. influences nucleosynthesis only because it causes an increase in the total energy density of the universe and a corresponding increase in the expansion rate. In this sense it is analogous to the existence of new neutrino types (see Sec. III). As for the nonvanishing chemical potential of the electronic neutrino, it not only contributes to ρ, but also directly and strongly affects the concentration of neutrons frozen out. Indeed, in equilibrium $\mu_n - \mu_p = \mu_e - \mu_{\nu_e}$, μ_e being much smaller than T because of electric charge neutrality, so that

$$\frac{N_n}{N_p} \simeq \exp\left[-\left(\frac{\Delta m}{T} + \xi_{\nu_e}\right)\right] . \tag{9.2}$$

Based on an analysis of He^4 production the conclusion was reached that

$$-0.25 < \xi_{\nu_e} < 1.8 . \tag{9.3}$$

This restriction was found under the assumptions (a) that the baryon density in the universe is confined within the limits 10^{-31} g cm$^{-3} < \rho_B < 10^{-28}$ g cm^{-3} and (b) that $|\xi_{\nu_\mu}| < 45$. If there are other neutrino species than ν_e and ν_μ the latter condition should be changed by $(\sum \xi_i^4)^{1/4} < 45$.

We should like to note that if $\xi_{\nu_e} = 0$ then from the consideration of He^4 formation it follows that $(\sum \xi_i^4)^{1/4} < 4$. If neutrino oscillations existed, the number of independent chemical potentials would be smaller and the bound on ξ_i would be changed. Oscillating neutrinos in the early universe were recently considered by Dolgov (1981).

The possibility discussed here, that the universe can possess nonvanishing leptonic charges, was the starting point for criticism (Linde, 1979b) of the bounds on the number of new neutrino types (see Sec. III). Indeed it is well known that by a small increase of ξ_{ν_e} it is possible to compensate for the higher neutron concentration connected with new types of neutrinos. Thus the limit $k_\nu < 6$ would not necessarily be satisfied if there were a relative excess of relic ν_e over $\bar{\nu}_e$ of the order of unity (i.e., $\xi_{\nu_e} > 0$ and $|\xi_{\nu_e}| \simeq 1$). Of course some bound, though extremely weak, can be obtained from the condition $\sum \rho_\nu < \rho_{\max}$:

$$k_\nu < 10^5 . \qquad (9.4)$$

This limit can be improved if one considers the deuterium production. The result, however, depends very much upon the density of baryons in the universe, which is known to within about two orders of magnitude. In the standard model the following estimate is found (Reeves, 1974):

$$R(\mathrm{H}^2)/R_0(\mathrm{H}^2) = (10^{-30}\ \mathrm{g\,cm^{-3}})/\rho_B .$$

Here $R_0(\mathrm{H}^2)$ is the mass fraction of deuterium observed today and $R(\mathrm{H}^2)$ is the mass fraction of primordial deuterium. With rising ρ_B the rate of transition of H^2 into He^4 increases. That is why $R(\mathrm{H}^2)$ drops as ρ_B rises. An increase in the total energy density due to the contribution of new particles or neutrino degeneracy results in an increase in $R(\mathrm{H}^2)$. This is because the expansion rate rises with increasing ρ and correspondingly less time is left for H^2 to transform into H^4. So a limit on k_ν should be stronger for smaller ρ_B. According to Linde (1979 b) $k_\nu < 10^4$ for $\rho_B = 10^{-29}\ \mathrm{g\,cm^{-3}}$ and $k_\nu < 30$ for $\rho_B = 10^{-30}\ \mathrm{g\,cm^{-3}}$. It is rather difficult, however, to reach a conclusion about the mass fraction of primordial deuterium based upon the abundance observed today because the formation or destruction of H^2 in later stages is not well known.

Note that the assumption of strong degeneracy of the neutrino sea permits us to improve the limit on neutrino mass. If the contemporary value of the neutrino chemical potential is larger than m_ν, the energy density of relic neutrinos is equal to $(\mu^3 m_\nu)/6\pi^2$. If it is to be smaller than $\rho_{\max}$, the bound $m_\nu < 4$ eV should be valid (for $\xi = 50$ and a vanishing cosmological constant).

In conclusion we should like to stress once more that cosmological data do not contradict considerable neutrino degeneracy (i.e., $|N_\nu - N_{\bar{\nu}}|/N_\gamma \gtrsim 1$) and as a result do not contradict the existence of a large number of different massless particle species. However this seems to be very unlikely.

X. ASTROPHYSICAL CONSEQUENCES OF THE EXISTENCE OF A HEAVY NEUTRAL LEPTON

In the preceding section restrictions on the properties of leptons obtained with the help of astrophysics were considered. Now we shall look at the subject from another point of view and discuss how the existence of a massive neutral lepton helps us to resolve some astrophysical problems.

A heavy neutrino is a remarkable candidate for bearing the galactic missing mass. As is well known the mass-to-light ratio for galaxies as a whole is considerably higher than for average stars in the middle part of a galaxy. The galactic mass can be determined by the gravitational attraction between galaxies in a cluster of galaxies. Another way to find the mass of a galaxy is to measure the variation of the velocities of matter with an increase of the distance from the galactic center. Such measurements also show that the bulk of the matter in galaxies is outside the region where luminous objects are seen. So the dynamics of galaxies are determined by some dark, invisible matter. Marx and Szalay (1972; Szalay and Marx, 1976) and Cowsik and McClelland (1972) propose that neutrinos with a mass of about 10 eV could solve the problem of the missing mass. The idea was later considered by Marx (1977), Gunn *et al.* (1978), and Tremaine and Gunn (1979). They argued that neutrinos mainly cluster with galaxies forming galactic halos.

Using the condition

$$\Omega^* = (\text{the matter density in galaxies})/\rho_{\max} < 0.05 \qquad (10.1)$$

Tremaine and Gunn conclude that $\sum m_\nu < 0.6$ eV. From the Pauli exlusion principle it follows that the maximum neutrino density $n_{\max}$ in phase space cannot exceed $2g_\nu$, where g_ν is the number of helicity states. With a more accurate treatment of neutrino dynamics Tremaine and Gunn conclude that $n_{\max} < g_\nu$. On the other hand, if neutrinos in a galactic halo form a Maxwellian thermal gas then it follows from the known value of $n_{\max}$ that $m_\nu > 10$ eV. The contradiction between the upper and lower limits on m_ν leads one to conclude that light neutrinos cannot solve the problem of the missing mass. Heavy leptons, however, with a mass around some GeV, can do this. In our opinion the accuracy of bound (10.1) is overestimated and it could be an order of magnitude weaker. In this case the result of Tremaine and Gunn could be invalid.

Recently Lyubimov *et al.* (see also Kosic *et al.*, 1980) have claimed that a thorough investigation of the electron spectrum in tritium decay proves that the neutrino (electronic antineutrino) has a rest mass of the order of 30 eV (14 eV $< m_\nu <$ 46 eV) on a 99% confidence level. Using the data of the Reines experiment, Barger *et al.* (1980) find oscillations in the antineutrino beam inducing the reaction $\bar{\nu}_e + d \to 2n + e^+$. The difference in the squares of the neutrino masses $(m_1^2 - m_2^2)$ was estimated to be of the order of 1 eV^2 but other solutions with a larger difference are also possible. Reines *et al.* (1980) claim to have detected oscillations by comparison of the charge current reaction $\bar{\nu}_e + d \to 2n + e^+$ and the neutral current reaction $\bar{\nu}_e + d \to p + n + \bar{\nu}_e$ induced by the beam of reactor antineutrinos. Oscillations $\nu_e \leftrightarrow \nu_\tau$ seem also to have been observed by the CHARM collaboration [see Winter (1979).].

All this taken together is changing the prevailing attitude of the astrophysical establishment towards neutrinos with nonvanishing rest mass of the order of some eV or tens of eV. Earlier it was commonplace to put upper limits on the neutrino mass assuming that it would converge to zero. Now one is tempted to explain real phenomena in conjunction with the rest mass. Marx and Szalay, who adopted this point of view ten years ago, thus find themselves in the position of forerunners.

The first success of such an approach is the explanation of the deficit of solar neutrinos in the Davis experiment by oscillations $\nu_e \leftrightarrow \nu_\tau \leftrightarrow \nu_\mu$ as proposed by Pontecorvo. The prediction depends on mixing angles, but a ratio of $\frac{1}{3}$ of ν_e to the total amount of ν's in the Earth's orbit is reasonable.

One obvious astrophysical implication is that if the sum of neutrino masses $\Sigma = m_1 + m_2 + m_3$ exceeds 25 eV the universe is closed and its expansion must be changed to contraction in the remote future. The age of the universe is also shortened. To a good approximation it is

$$t = 21 \times 10^9 \text{ yr}/[(H/50) + (\Sigma/40)^{1/2}]. \tag{10.2}$$

Here H is the Hubble constant in kilometers per second per megaparsec (the best value is $H = 55$) and Σ is taken in eV.

If for all neutrinos $m_1 = m_2 = m_3$, then when $\Sigma = 90$ one obtains $t = 8.5 \times 10^9$ yr, which is less than the values found by Os/Re chronometry ($11-18 \times 10^9$ yr) and by stellar evolution theory applied to old globular clusters ($14-16 \times 10^9$ yr).

A possible outcome has been suggested by Zeldovich and Sunyaev (1980), who note that the age of the universe can be lengthened by the introduction of the cosmological constant. One gets thereby a closed but ever-expanding universe. It is only for $\Lambda = 0$ that the two features coincide, i.e., from the fact that the universe is closed in space it follows that it oscillates in time. If $\Lambda \neq 0$ the curvature of space can be positive, so that the universe is closed but it expands forever. The upper limit on m_ν which is valid for models with a nonvanishing cosmological constant is about 200 eV. Accidently we return to the first rough estimate of Gerstein and Zeldovich (1966).

Another way of violating bound (4.2) without revising any fundamental idea is to assume that an unknown source of background photons existed, so that the ratio N_ν/N_γ is smaller than in the standard model. An example of such a source would be a new long-lived particle decaying into photons.

As has already been pointed out by Marx and Szalay, if $m_\nu \neq 0$ the growth of perturbations leading to galaxy formation and clustering is also changed: It begins now with the growth of neutrino density perturbations. In a recent investigation Doroshkevich *et al.* (1980) conclude that smaller initial perturbations are needed. So it is easier to understand the negative result of all searches for background temperature fluctuations.

Neutrinos more massive than those considered by Tremaine and Gunn (1979) are good candidates for the explanation of the "hidden mass" of galaxies and clusters of galaxies. Despite the fact that massive neutrinos could solve some cosmological problems we should like to end this discussion with a warning: Astronomy will use gratefully the neutrino rest mass found in laboratories, but astronomy can not be used to prove or even to support a definite value or even the very existence of a nonzero mass for neutrinos or some other unknown particles.

XI. SOME OTHER PARTICLES (HIGGS MESONS, AXIONS, MAGNETIC MONOPOLES, ETC.)

The arguments of the preceding section can be directly applied to other elementary particles. In what follows we briefly discuss how cosmology helps to restrict the region of permitted values for parameters of various hypothetical particles and interactions. These particles are not yet seen in the laboratory so astronomical observations are the only sources of information about them. Out of respect for chronology we shall start from the earliest considerations of massless vector particles.

A. Et cetera

In 1955 Lee and Yang posed the question, whether baryonic photons existed, i.e., massless vector particles interacting with baryonic charge in the same manner as photons interact with electric charge (but with a weaker coupling). Later Okun (1969) considered the hypothesis that each conserved charge is a source of its own type of photons, i.e., of baryonic, electronic, muonic, etc. A very strong limit on the coupling constants of baryonic and electronic photons follows from the equality of the gravitational and inertial masses of lead and of copper found by Dicke (1962) (see also Braginsky and Panov, 1971). The restrictions on the couplings obtained by Okun (1969) are, respectively, for baryonic and leptonic photons

$$\alpha(B) < 10^{-43}\alpha(Q), \quad \alpha(E) < 10^{-45}\alpha(Q), \tag{11.1}$$

where $\alpha(Q) = \alpha = \frac{1}{137}$ is the electromagnetic coupling constant.

For muonic charge a comparable restriction cannot be applied because there are no macroscopic samples of matter with a nonzero muonic charge. The laboratory data on $(g-2)$ give the rather weak limit

$$\alpha(M) < (10^{-3} - 10^{-4})\alpha(Q).$$

Stellar luminosity data permit us to put greater constraints on the muonic coupling constant. The point is that the interaction of a muonic neutrino with a muonic photon would result in the electromagnetic interaction of a muonic neutrino due to the transition of a muonic photon through a virtual $\mu^+\mu^-$ pair into a usual photon. The induced charge is of the order of

$$e_{\text{ind}} \simeq e\alpha(M)\ln(\Lambda/m_\mu).$$

This estimate is based on a divergent Feynman integral and depends on additional conditions imposed on the theory. It should be valid, however, if there were no special requirement that the induced charge vanish. Now using the results of Sec. VI (see Table II) and neglecting the logarithmic factor we obtain

$$\alpha(M) < 10^{-13} \quad if\ m_{\nu_\mu} < 100 \text{ eV (sun)}$$

$$\alpha(M) < 3 \times 10^{-12} \quad if\ m_{\nu_\mu} < 20 \text{ KeV (white dwarfs)}.$$

By analyzing the relic neutrino background (Sec. V) one may determine that $m_\nu < 40$ eV. If, however, the neutrino mass exceeded this limit, this would mean for instance that the neutrino possessed a new stronger-than-usual interaction. From this the lower limit on $\alpha(M)$ for heavy ν_{II} is as follows:

$$\alpha(M) > 3.3\times 10^{-7}\frac{m_\nu}{m_N}\left[1+\frac{1}{42}\;\frac{\pi\alpha^2(M)m_N}{m_\nu}\right]^{-1/2}. \qquad (11.2)$$

This inequality is valid for $m_\nu > 40$ eV. If, however, $m_\nu < 40$ eV, $\alpha(M)$ can vanish, and this is probably the case.

B. Higgs mesons, axions

The scalar Higgs field χ is probably a necessary ingredient in the modern theory of the weak and electromagnetic interactions. A nonzero vacuum expectation value for this field results in a spontaneous generation of masses of fermions and gauge bosons. The mass of χ is a free parameter of the theory. From the absence of the decay $K^+\to\pi^+\chi^0$ one can conclude that $m_\chi > 300$ MeV. There are theoretical arguments (Linde, 1976a; Weinberg, 1976) in favor of a lower bound on m_χ an order of magnitude larger.[11] The derivation of such a bound is based, however, on the assumption that fermion masses are much smaller than those of intermediate bosons, $m_f < m_w$, and that there is a single χ field. If $m_f \gtrsim m_w$ the Higgs meson mass is, generally speaking, arbitrary. Spontaneous symmetry breakdown, however, proceeds in such a way that we get a bound on fermion masses $m_f < 76$ GeV; otherwise the condition of vacuum stability is not fulfilled (Anselm, 1979; Linde, 1976b; Linde and Krive, 1976).

In any case the Higgs boson mass should be larger than several hundred MeV independently of the theory. Such heavy bosons must have a very short lifetime. For example, if $m_\chi = 500$ MeV, $\tau_\chi = 10^{-16}$ sec. In this case cosmology provides no restriction on the properties of the Higgs bosons. The existing bounds (Sato and Sato, 1975 a, b) refer to the unrealistic case of very light scalar particles which decay into photons with the lifetime $\tau_\chi \simeq 7\times 10^7$ (KeV/m_χ)^3 sec. Such particles definitely have nothing to do with the standard Higgs meson. Nevertheless the restrictions on their properties are of interest because no laboratory limits in this region of mass and interaction strength are known.

By studying the spectrum of the cosmic electromagnetic background Sato and Sato (1975a) determined that the mass region 0.1 eV$< m_\chi <$400 eV is forbidden. This forbidden region can be enlarged with the help of the arguments presented in Sec. VII.B, namely, that the Planck spectrum of relic photons, destroyed by the decay $\chi\to 2\gamma$, would not be restored to the spectrum we observe at the present time if $\tau_\chi > 3\times 10^4$ sec. This would permit us to exclude the region 0.1 eV$< m_\chi <$10 KeV. However, the lifetime value used by Sato and Sato (1975a) is about two orders of magnitude higher than the standard estimate. So the upper boundary of the forbidden region shifts to 1 KeV. By the study of stellar evolution we obtain in what follows stronger bounds on masses of light scalar and pseudoscalar mesons having semiweak interactions with fermions (Sato and Sato, 1975b; Vystosky *et al.*, 1978; Sato, 1978a; Falk and Schramm, 1978; Dicus, Kolb, Teplitz, and Wagoner, 1978).

The light pseudoscalar particle known as the axion (Weinberg, 1978; Wilczek, 1978) was born theoretically as the result of attempts to explain parity conservation in the framework of quantum chromodynamics. The axion emerges from the spontaneous breakdown of U(1) symmetry as the Goldstone boson, and acquires a small mass through interaction with instantons. The mass of the axion was estimated to lie within the limits 10 KeV $< m_a < 1$ MeV.

The search for the decay $K^\pm\to\pi^\pm a$ and for axion production by reactor neutrinos has given negative results, so probably the axion does not exist, but it may be early to pronounce the final sentence.

The interaction of axions and of Higgs-type scalar bosons with fundamental fermions is of the form

$$\mathcal{L}_{\text{int}} = \sqrt{G_F}\, m_f(\bar{u}\gamma_5 u a + \bar{u} u \chi), \qquad (11.3)$$

where m_f is the fermion mass, $G_F = 10^{-5} m_N^{-2}$, a is the operator of the axion field, and χ is the operator of the scalar meson field. If $m_{a,\chi} < 2m_e$ the dominant decay mode for both a and χ is 2γ. The lifetime for this decay is defined by the famous triangle diagram and equals

$$\tau = 10^{-3} k^{-2}(1\ \text{MeV}/m)^3\ \text{sec}. \qquad (11.4)$$

Here $k = \sum Q_i^2$; the summation over all quark and lepton species is performed (for the model with three lepton generations and six quark flavors, $k=8$), m is the mass of a or χ. The factor k^{-2} is due to the equal contribution of all fermions in the intermediate state to the decay probability. If $m < m_f$ the diagram alone gives a result proportional to m_f^{-1} and the coupling constant of a or χ with fermions is $\sim m_f$.

If light particles with interaction (11.3) existed, they should be produced inside stars owing to the reactions $\gamma e\to ea(e\chi)$ or $\gamma\gamma\to a(\chi)$. The cross sections of χ and a photoproduction on electrons near threshold are

$$\sigma(\gamma e\to e\chi) = \frac{\alpha G_F}{4}\left(1-\frac{m_\chi^2}{\omega^2}\right)^{1/2}\left[\left(\frac{m_\chi}{m_e}\right)^2 + 2\left(1-\frac{m_\chi^2}{\omega^2}\right)\right],$$
$$\sigma(\gamma e\to ea) = \frac{\alpha G_F}{4}\left(\frac{m_a}{m_e}\right)^2\left(1-\frac{m_a^2}{\omega^2}\right)^{1/2}. \qquad (11.5)$$

Here $\alpha = \frac{1}{137}$ and ω is the photon energy. After thermal averaging, each factor $(1-m^2/\omega^2)$ should be changed by $2T/m$, where T is the temperature inside the star (if $T < m$).

The density of matter inside the solar core is $\rho_c = 10^2$ g cm^{-3} and the temperature is $T \simeq 1$ KeV. The average density of solar matter, however, is $\rho_s = 1$ g cm^{-3}. Thus the mean free paths of a and χ connected with the reactions of photoproduction and the inverse one of radiative capture are

[11]Recently Linde (1980a) presented arguments based on cosmology that the mass of the lightest Higgs particle should be smaller than 9 GeV if a single multiplet of the Higgs meson exists.

$$\lambda(a) = \frac{1}{\sigma n_e} = 10^{11}\left(\frac{m_e}{m_a}\right)^2\left(\frac{m_a}{T}\right)^{1/2}\ \text{cm},$$
$$\lambda(\chi) = 10^{11}\left(\frac{m_\chi}{T}\right)^{1/2}\left(\frac{4T}{m_\chi} + \frac{m_\chi^2}{m_e^2}\right)^{-1}\ \text{cm}. \tag{11.6}$$

These are much larger than the solar radius $R_\odot = 7\times 10^{10}$ cm. The decay length of axions (or χ particles) in accordance with Eq. (11.4) is

$$\lambda_{\text{decay}} = 10^{14}k^{-2}(10\ \text{KeV}/m)^3(T/m)^{1/2}\ \text{cm}. \tag{11.7}$$

Thus for $m < 100$ KeV axions could freely escape from the center of the sun and carry a considerable amount of energy. The specific energy losses due to the reactions $\gamma e \to ae$ and $\gamma\gamma \to a$ have been calculated (Vysotsky *et al.*, 1978). They are dominated by the last reaction and equal

$$q = 4\times 10^8 m^{11/2} e^{-m/T} T^{3/2}(k^2/64)\ \text{erg cm}^{-3}\,\text{sec}^{-1}. \tag{11.8}$$

Here $m = m_a$ or m_χ and T is expressed in KeV. The total energy loss by the sun due to light a or χ particles is $L = qM_c/\rho_c$, where $M_c \simeq 10^{-2}M_\odot$ is the mass of the solar core and $\rho_c = 10^2$ g cm^{-3} is its matter density. Requiring energy loss due to axion (or χ particle) radiation to be smaller than the photon luminosity of the sun we obtain

$$m_{a,\chi} > 25\ \text{KeV}. \tag{11.9}$$

As the considered mechanism provides energy mainly in the form of x rays, the data (Womack and Overbeck, 1970) on the solar x-ray flux on the earth permit us to get the better limit $m > 50$ KeV. In this way the bound $m > 100$ KeV was obtained, but in our opinion such an estimate is a bit high. Indeed even for $m = 50$ KeV the decay length of axions is $\lambda_{\text{decay}} = 4\times 10^9$ cm [see Eq. (11.4)] so the x-ray quanta from the decay mainly "stick" inside the sun, and for $m = 100$ KeV the x-ray flux at the Earth's orbit should be negligible. We will not dwell on it because stronger though more model-dependent limits on m can be obtained from the study of red supergiants (Vysotsky *et al.*, 1978; Sato, 1978; Dicus, Kolb, Teplitz, and Wagoner, 1978). The core temperature of a red supergiant is $T \simeq 15$ KeV (Warshawsky and Tutukov, 1972 and 1973) which is much hotter than that of the sun, and the radius of the core is $R_c \simeq 10^{10}$ cm $\gg \lambda_{\text{decay}}$ for $m > 100$ KeV. So the axion luminosity is determined by axion emission from the outer layer, whose thickness is equal to λ_{decay}. The energy transmission by axions from the central region to the periphery in the hydrodynamic time should be smaller than the envelope binding energy. For characteristic values of the mass of the core $M_c = 4M_\odot$, that of the envelope $M_{en} = 12M_\odot$, and the envelope radius $R_{en} = 10^{12}$ cm, the binding energy is $E_b = 10^{49}$ erg. Thus from the condition

$$E = L_a t_H = 2\pi R_c l_a q t_H < E_b, \tag{11.10}$$

where $t_H \simeq 10^4$ sec is the hydrodynamic time and q is determined by Eq. (11.9), we obtain (Vysotsky *et al.*, 1978)

$$m_{a,\chi} > 200\ \text{KeV}.$$

A similar limit follows also from data on the lifetime of red supergiants. Cosmological implications of axions and light Higgs bosons were also considered by Dicus, Kolb, Teplitz, and Wagoner (1978). They discussed the influence of these particles on primordial deuterium production, the spectrum of the electromagnetic background, and stellar evolution. Their best bound on particle mass is obtained from the analysis of red giants and coincides with that given above.

C. Magnetic monopoles

The requirement that there be symmetry between electricity and magnetism led Dirac (1934; see also Dirac, 1948) to hypothesize the existence of an elementary magnetic charge, i.e., a magnetic monopole. The value of a magnetic charge is related to the electron charge by the quantization condition $g/e = \alpha^{-1}k/2$ where $\alpha = \frac{1}{137}$ and $k = 0, \pm 1, \pm 2, \ldots$. In the Schwinger (1966) model only even values of k are permitted. The huge magnetic charge of the monopole is thus quantized and well defined. The mass of the monopole was, however, unknown. Broken symmetry theories have given new life to the monopole idea (Polyakov, 1974; t'Hooft, 1974). The existence of monopoles and their properties depends crucially on the type of theory (see discussion at the end of this section). There are monopoles which are sources of the massive gauge fields and genuine monopoles which are sources of a Maxwellian magnetic field. In what follows only these genuine monopoles are considered. The monopole mass in these theories is expressed through the gauge boson mass. If the gauge boson coincides with the intermediate boson of the weak interaction then

$$m_M = \alpha^{-1}m_w \simeq (5-10)\ \text{TeV}.$$

See, however, the discussion at the end of this subsection.

Despite intensive searches for magnetic monopoles in accelerator experiments as well as in cosmic rays they have still not been discovered. The only candidate described in the literature (Price *et al.*, 1975, 1978) for the role of the monopole seems to be doubtful. Observational bounds on the number density of magnetic monopoles in the universe are presented by Bludman and Ruderman (1976). The best limit follows from the existence of galactic magnetic fields (Bludman and Ruderman, 1976; Domogatsky and Zheleznykh, 1969; Parker, 1970). If the monopole density were sufficiently high, monopoles would destroy galactic magnetic fields by accelerating in these fields and so taking away the field energy. From this it follows that

$$n_{M0} < 10^{-26}\ \text{cm}^{-3}. \tag{11.11}$$

Because of the large coupling constant of monopoles with a magnetic field they are effectively accelerated in galactic magnetic fields and their average energy is about 10^{11} GeV. The scattering of energetic monopoles by the relic electromagnetic background was considered by Osborne (1970) where, using the data on the flux of cosmic γ rays, it was shown that the monopole number density is smaller than 10^{-24} cm^{-3} for $m_M = 10$ TeV and smaller than 10^{-26} cm^{-3} for $m_M = 2.5$ TeV.

In what follows we shall evaluate the number density of the relic monopoles (Domogatsky and Zheleznykh, 1969; Adams *et al.*, 1976; Zeldovich and Khlopov, 1978) along the same lines as in Sec. III, where we calculated the number density of relic leptons and nucleons. The number density of monopoles at the moment of thermodynamic equilibrium violation is defined by Eq. (2.11). We make a crude estimate of the monopole–antimonopole annihilation cross section by assuming that annihilation must take place if the distance between monopoles is such that their kinetic energy is equal to the potential energy. So the annihilation radius is

$$r_0 = \frac{g^2}{4\pi T} \tag{11.12}$$

where $g = e/2\alpha$ and $e^2 = 4\pi\alpha$. Hence for the annihilation cross section one finds $\sigma_0 = \pi r_0^2 = \pi/16\alpha^2 T^2$. This estimate will be made more precise in what follows. The realistic cross section does not rise so fast when temperature drops. The temperature at the instant of violation of thermodynamic equilibrium is determined by Eq. (2.10) and is equal to $T_1 \simeq 0.02\, m_M$. In the now fashionable models of elementary particles the number of excited degrees of freedom is $N_{DF} \simeq 100$ at this temperature. Taking this into account we find for the relative concentration of monopoles at the moment of equilibrium violation

$$r_1(M) \simeq 10^{-19}\ \mathrm{cm}^{-3}. \tag{11.13}$$

If after the violation of thermodynamic equilibrium the burning out of the monopoles stopped, their concentration today could be as great as 10^{-16} cm^{-3}. However, considerable annihilation goes on even when $t > t_1$. At this time the equilibrium monopole concentration becomes smaller than the real concentration, so we neglect r_{eq}^2 in Eq. (2.8) when evaluating the residual monopole concentration. As an initial condition for r_M the value given in (11.13) is used. This expression is somewhat underestimated because too high a value for the cross section was assumed in the calculation of r_{1M}. As we shall see, however, the final result is independent of the initial condition. Equation (2.8) can be easily solved yielding the following expression for the relative concentration of relic monopoles today: $r_{0M} = r_1(1 + r_1 I)^{-1}$, where $I = \int_{t_1}^{t_0} dt\,(\sigma v T^3)$. The value of I is so large that $r_1 I \gg 1$ and $r_{0M} \simeq I^{-1}$ independently of r_1. If one substitutes for the cross section our first crude expression σ_0, the integral diverges when $t_0 \to \infty$ as $t_0^{1/4}$. As an upper bound, the time until hydrogen recombination in the primeval plasma should probably be substituted, but not the age of the universe. Until recombination monopoles are in kinetic equilibrium with the plasma and their temperature drops together with the plasma temperature. Later, when matter develops into galaxies, the surviving monopoles are accelerated in the (possible) magnetic fields. Substituting $t = 10^{12}$–10^{13} sec as the upper bound into the expression for I we find $I = 2\times 10^{24}$ and $n_{0M} = 10^{-22}$ cm^{-3}, which is four orders of magnitude greater than the bound (11.11). Even if the burning out of monopoles kept on until the present time, their concentration would be two orders of magnitude greater then the existing upper bound (11.11). In fact the discrepancy is even more serious. The point is that the annihilation cross section used above is considerably overestimated. The radius $r_0^2 = g^2/4\pi T$ is so large that in the volume $V = 4\pi r_0^3/3$, a large number of particles is contained, $N = VT^3 \simeq (16\alpha^3)^{-1} \gg 1$. That is why monopole annihilation should be treated in the diffusion approximation (Zeldovich and Khlopov, 1978). The mean free path of monopoles in the plasma is $\lambda = (\sigma_{90} N_\gamma)^{-1}$ where σ_{90} is the cross section of monopoles multiply scattering on 90°, $\sigma_{90} \simeq 2\times 10^{-33} \times (m_M/T)$ cm^2. The value of λ becomes larger than r_0 rather late, when $t > t_2 \simeq 10^{-5}$ sec. Until this time the quantity $4\pi D r_0 = 4\pi\lambda v r_0/3$ should be substituted into the expression for I instead of $\sigma_0 = \pi r_0^2$. Then we find for the relative monopole concentration $r_{2M} = 10^{-21}$ at $t = t_2$. When $t > t_2$, one can neglect the effect of other particles on monopole–antimonopole collisions. We evaluate the annihilation cross section assuming that capture takes place if the monopole energy loss due to bremsstrahlung is of the order of the kinetic energy of the monopoles. In the classical approximation we obtain $\sigma = (T/M)^{0.6}\sigma_0$. As a result the integral I becomes convergent on the upper limit and the relative concentration of monopoles practically does not change for $t > t_2$. Finally we have[12] $n_{0M} = 5\times 10^{-19}$ cm^{-3}.

This estimate was made under the assumption that the monopole mass is fixed independently of temperature. It has been observed (Kirzhnits and Linde, 1972), however, that when $T > m_w$ there is no symmetry breakdown in the theory and the gauge boson masses vanish. The particles obtain nonzero masses only when T drops below m_w. In accordance with this view we assume that $M_M = \alpha^{-1}(m_w^2 - T^2)^{1/2}$. The thermodynamic equilibrium would be violated now at a temperature three times lower than in the case $m_M = \mathrm{const}$: $T_1' = 0.9\, m_w$ and the relative monopole concentration would be $r_1' = r_1/3$ [see Eq. (11.13)]. The final result r_{0M}, however, would not be noticeably changed.

Thus the relic monopole concentration proves to be unacceptably huge. Probably our extrapolation into the temperature region of about 100 GeV or even higher is not correct. It would surely be incorrect if the Hagedorn limiting temperature existed. In this connection an interesting possibility has been discussed by Polyakov (1974). There is some support for the idea of a phase transition in quantum chromodynamics with rising temperature, during which quark confinement at low temperature gives way to liberation at high temperature. For colored monopoles the picture is just the opposite, and at high temperature monopoles are confined. Such a phase transition is connected with the gluon string condensation at a high temperature. The temperature of the phase transition is estimated to be about 1 GeV. The number of relic monopoles in this case should be extremely small, $N_M = \exp(-m_M/1\ \mathrm{GeV}) \simeq 10^{-4000}$, and no contradiction with the observations would exist. One can reverse the arguments and regard the absence of the relic monopoles as a hint regarding the phase transition in quantum chromodynamics (QCD). Monopole confinement has also been

[12] Analogous arguments applied to charged particles predict their concentration $n_{ch} = 2\times 10^{-15}$ cm^{-3}. This means that no stable charged particle with a mass larger than 3×10^9 GeV can exist.

mentioned by Daniel *et al.* (1979) and Linde (1980b).

The value of the monopole mass used above is rather optimistic. In the Weinberg–Salam model based on the group $SU(2)\times U(1)$ monopoles are not obligatory. Monopoles should exist in a gauge theory based on a semisimple group with a single coupling constant. Such groups are considered now as a basis of grand unified theories. The masses of the intermediate bosons in the majority of the known models are extremely large, $m_X \simeq (10^{15}-10^{16})$ GeV. This leads to a monopole mass of the order of the Planck mass or one or two orders of magnitude smaller. The calculation made above gives in this case an even larger monopole concentration. It is possible, however, that the singularity in the solution of the equations of general relativity would be smoothed down because of gravitational quantum corrections when temperature tended to the Planck mass. If this were the case the monopole concentration would be suppressed by the factor $\exp(-m_M/T_{\max})$, which could help. If, however, the original symmetry group G were broken in such a way that electromagnetic interactions were described by a non-Abelian semisimple subgroup G' of G and the masses of the gauge bosons appearing through a breaking of G were small, i.e., $m_v \ll m_X$ and $m_M = m_v/\alpha \ll m_P$, then the scenario described above would not be possible. If monopole confinement at high temperature does not occur this can be considered as an indication of the character of the symmetry breaking.

Heavy monopoles in grand unified theories have been discussed by Preskill (1979). He claims that the bound given in Eq. (11.11) is valid for monopoles which are lighter than 10^{16} GeV. (One criticism of this view, however, is that, as noted by Khlopov and Linde, monopoles with practically any mass would destroy galactic magnetic fields at the period of galactic formation.) If $m_M \simeq 10^{16}$ GeV, a rather strict upper bound on their abundance can be obtained from the limit on the total mass density of the universe and from consideration of He^4 formation.

Preskill considered different scenarios of heavy monopole production in the course of phase transitions connected with spontaneous symmetry breaking and concluded that the initial monopole density was not necessarily close to the thermal one but could be much smaller, depending on the type of phase transition. He argued that monopole production could be suppressed if the phase transition were of the first order. For very small initial monopole densities (r_{1M}) the final result does depend on r_{1M} and probably could be in agreement with observations. Analogous reasoning has been applied to the problem by Einhorn *et al.* (1980).

Further investigations of the type of phase transition in connection with the monopole problem have been made by Guth and Tye (1980), who argue that in the SU(5) group the phase transition is indeed of the first order if some strict constraints on coupling constants are imposed (see also Einhorn *et al.*, 1980). It is not clear whether these constraints can be derived in a natural way. A detailed study of this point is now in progress (Linde, private communication). To conclude, the problem of relic monopoles remains unsettled; surely many new efforts to resolve it are underway.

D. θ particles

Recently Okun (1980a) has discussed the existence of new long-range forces and especially of a new kind of particle with large, probably microscopic, confinement radius—the so-called θ particle (Okun, 1980b). The most restrictive bounds on its properties can be derived from cosmology. For instance, the rank of the gauge group generating interactions between θ particles can be limited from above by consideration of primordial He^4 formation. If the latter is SU (N) then $N<3$. Relic θ matter has been considered by Okun (1980a), Dolgov (1980a), and Khlopov (1980). They reached the conclusion that the existence of θ quarks (i.e., particles possessing color charge and θ charge) in a simple version of the model contradicts big bang cosmology. Either the model should be modified or quarks do not exist. If q_θ existed their mass would have to exceed 15 GeV (because they were not seen at PETRA), so bound states of gluons (g), being light, should be practically stable. They could burn out in three-body collisions like $3g\to 2g$. Their abundance has been calculated by Dolgov (1980a), and it has been shown that the confinement radius of such particles should be larger than 100 fm.

XII. THE PHOTON MASS

The photon is the only particle (except for the graviton?) for which the theory demands that the mass be vanishing. This demand is connected with the principle of gauge invariance, basic to electromagnetic interactions (and probably to weak and strong ones too), and with strict current conservation. Thus in contrast to the neutrino, whose zero mass is weird, the photon has a vanishing mass as a natural consequence of the theory. History teaches us, however, that a symmetric picture often proves to be crooked at a closer look. So a skeptic would expect that gauge symmetry is only approximate. In this case the photon could have a small but nonvanishing mass. It is noteworthy that a finite photon mass does not lead to serious difficulties in pure quantum electrodynamics, but a massive photon can spoil the renormalizability of the unified theory of weak and electromagnetic interactions. A renormalizable model of massive photon interactions with a spontaneous breakdown of electromagnetic current conservation has been considered by Okun and Zeldovich (1978) in connection with a possible electron instability. In this model, however, a very light charged scalar particle emerges in contradiction with experiment. The problem of a possible electron instability is discussed in the next section. Here we consider the Maxwell equations at the classical level.

Putting aside the theoretical difficulties, we note that the last word, as always, is left to experiment. In obtaining limits on the photon mass, terrestrial experiments cannot compete with the results obtained with astronomical data. The laboratory limits on m_γ are summarized in a review by Goldhaber and Nieto (1971). The best limit was obtained by measuring the electric field in a closed cavity with conducting walls (Williams *et al.*, 1971). As is known, for massless photons $E=0$. From the upper bound on E it follows that

$$m_\gamma < 10^{-14}\ \text{eV} \quad \text{or} \quad \lambda_\gamma \equiv m_\gamma^{-1} > 3\times 10^9\ \text{cm}\,. \tag{12.1}$$

In connection with experiments of this type but using static fields (in the cited paper an electric field with a frequency of 4×10^6 Hz was used) we mention the paper by Dolgov and Zakharov (1971) where the use of the Earth's electrostatic potential was proposed to obtain a limit on m_γ. The point is that for $m_\gamma \neq 0$ the absolute value of the potential φ becomes observable and directly affects the electric field inside the conducting cavity. It is known that the potential difference between the Earth's surface and the ionosphere is about 5×10^5 V and the electric charge of the Earth is 6×10^5 C. Were it not for the ionospheric charge, the potential of the earth would be 10^9 V, but because of ionospheric charge compensation it is about 5×10^5 V. This huge potential can in principle be used for restricting the photon mass.

The next step in the constraining of m_γ is the study of planetary magnetic fields at a large distance. In the case of massive photons the magnetic field has an extra decreasing factor $\exp(-m_\gamma r)$ and the existence of magnetic fields at a great distance from the source permits us to exclude too large m_γ. The best limit obtained in this way was established (Davis *et al.*, 1975) by the study of the magnetic field of Jupiter as measured by *Pioneer 10*,

$$m_\gamma < 6\times 10^{-16}\ \text{eV},\quad \lambda_\gamma > 5\times 10^{10}\ \text{cm}\,. \tag{12.2}$$

This is the best bound known today based on direct measurements. Further improvement can be achieved by the analysis of galactic magnetic fields. The results found in this way are discussed by Chibisov (1976). The most restrictive limit on m_γ was obtained through consideration of the equilibrium of the interstellar gas in the Magellanic clouds. The term proportional to m_γ^2 contributes to the magnetic pressure $p = m_\gamma^2 A^2/8\pi$, where A is the vector potential of the magnetic field. If the magnetic field B differs from zero at a scale l then $A \simeq Bl$ and the pressure is $p = (l m_\gamma)^2 B^2$. It is noted that this pressure leads to compression of the interstellar gas. Since the energy density of the gas in the Magellanic clouds is dominated by the energy density of the magnetic field, the equilibrium condition demands $B^2 > m_\gamma^2 A^2 \simeq l^2 m_\gamma^2 B^2$, i.e., $m_\gamma l < 1$. As $l \simeq 10^{22}$ cm, one obtains

$$m_\gamma < 3\times 10^{-27}\ \text{eV},\quad \lambda_\gamma > 10^{22}\ \text{cm}\,. \tag{12.3}$$

The cosmological horizon of the universe, in comparison, is $t_0 c \simeq 10^{28}$ cm.

In conclusion we should like to mention two recent papers in which cosmological consequences of the assumption of a finite photon mass were considered. In the paper by Kuzmin and Shaposhnikov (1978) the hypothesis of photon condensation at an early stage in the history of the universe was discussed. By the transmission of the condensate energy into longitudinal photons the relic longitudinal radiation could be produced. The authors consider the possibility that the hidden mass of the universe is concentrated in this radiation. If this were the case then there should now be about 10^{22} photons cm^{-3}. This hypothesis can be experimentally excluded, for example, by measuring slowly varied electric fields inside conducting cavities. Such experiments have been performed, as we noted above, to check the Coulomb law. The electric field strength of the longitudinal photons should be

$$E = (m_\gamma/\omega)\rho_{\max}^{1/2} \simeq 10^{-10}\ \text{V}/m\,.$$

Barnes (1979) considered a cosmological model with a nonzero charge density. It is believed usually that noticeable charge density contradicts the observed cosmological picture. Indeed the electric field connected with charge density through the equation div $E = 4\pi\sigma$ would destroy the isotropy of the universe and lead to a considerable increase in the expansion rate. From the assumption that the electrostatic interaction of matter in the universe is weaker than the gravitational interaction, it follows that the charge density is smaller than 10^{-24} cm^{-3}. A similar bound can be obtained from the observed homogeneity and isotropy of the universe. A naive order of magnitude estimate can be made as follows. If the electric charge $Q\neq 0$, the universe should be open and its metric should be quasihyperbolic, spherically symmetric, but locally anisotropic (except the "center"). The electric field should tend to a constant because the volume and the surface of the universe expand in the same way, $\exp(2r)$. From the condition that large-scale energy density variations be smaller than 10^{-3} one finds $E^2/8\pi < 10^{-12}$ erg cm^{-3}. Evaluating E as $E = \sigma e \times 10^{28}$ cm, where 10^{28} cm is the horizon radius, e is the electron charge, and σ is the number density of (noncompensated) charged particles, we obtain $\sigma < 10^{-24}$ cm^{-3}.

If, however, $m_\gamma \neq 0$ the relation between E and σ is changed to

$$\text{div}E + m_\gamma^2\varphi = 4\pi\sigma \tag{12.4}$$

where φ is the electrostatic potential, $E = -\nabla\varphi$. For a uniform charge distribution $\varphi = 4\pi\sigma m_\gamma^{-2}$ and $E = 0$. Nonzero charge density influences the dynamics of the universe only through contribution to the total energy density:

$$\rho = \frac{1}{8\pi}(E^2 + m_\gamma^2\varphi^2)\,. \tag{12.5}$$

In our case $E=0$ and $\varphi = 4\pi\sigma m_\gamma^{-2}$. From the condition $\rho<\rho_{\max}$ and inequality (12.3) we find

$$\sigma < m_\gamma(\rho_{\max}/2\pi)^{1/2} < 10^{-18}\ \text{cm}^{-3}\,. \tag{12.6}$$

Because the value of $\rho_{\max}$ corresponds to about 10^{-5} protons cm^{-3}, limit (12.6) means that the charge of only one in 10^{13} protons can be noncompensated by electrons.

Even for such a charge density, this model runs into difficulty in accounting for the primordial nucleosynthesis. With conserved electric charge this model is described by the "stiffest equation of state" at large density as giving the contribution $\rho\sim\sigma^2\sim N^2$ which will dominate over all other terms ($\rho\sim Nm$ for nonrelativistic particles and $\rho\sim N^{4/3}$ for degenerate ultrarelativistic particles).

XIII. NONCONSERVATION OF ELECTRIC CHARGE. UNSTABLE ELECTRONS

The electron is known to be the lightest charged particle. Its instability automatically means electric charge (and of course electromagnetic current) nonconservation. In classical Maxwellian electrodynamics, current conservation immediately follows from the gauge-invariant equation $\partial_\mu F_{\mu\nu} = j_\nu$. The simplest way to get a theory with a nonconserved current is to introduce the "massive term" into this equation, $m^2 A_\nu$. We know, however, that photon mass (if any) is extremely small (see the preceding section) and it proves to be difficult to find a consistent theory with nonconserved current.

One is free, however, to check the above dogmas experimentally. Moreover these dogmas ought to be checked in this world where physical principles undergo continuous revision. We note here that a field theory with electric charge nonconservation has been considered by Ignatier *et al.* (1979).

Searches have been made for the decay $e \to \nu\gamma$. In setups well shielded from cosmic rays, photons with energy 250 KeV were looked for. No such decay was observed and so the upper limit for this specific channel was reached:

$$\tau(e \to \nu\gamma) > 10^{22} \text{ yr}.$$

As noted by A. A. Pomansky (1976), a similar result can be obtained without these expensive experiments, and what is more this result is valid for any decay channel (e.g., $e \to 3\nu$). The point is that the decay of electrons inside terrestrial matter would lead to the accumulation of an electric charge inside the Earth at the rate

$$i = \frac{M_\oplus}{2m_N} \frac{Q_e}{\tau_e} = 3 \times 10^{22} \text{ C}/\tau_e. \tag{13.1}$$

The electric charge of the Earth is approximately constant and equal to 5.7×10^5 C. The electric field produced by this charge generates a current of $\sim 2 \times 10^3$ A from the upper layers of the atmosphere to the Earth's surface. It should be noted that the direction of this current is opposite to that necessary for compensation of possible electron decays. The stability of the terrestrial charge is maintained by a current from the Earth to the atmosphere which flows during thunderstorms, and which is estimated to be several thousand amperes. Hence stability of the charge of the Earth leads to the limit $\tau_e > 10^{29}$ sec $\simeq 3 \times 10^{21}$ yr. A compensating current should ultimately be created by cosmic electrons. It is known that the electronic component in energetic cosmic rays is smaller than 10^{-3} A. The necessary current of 10^3 A corresponding to the existing limit on τ_e could be picked up from low-energy electrons. The cosmic electron flux in this case should be about 10^3 cm^{-2} sec^{-1}.

It is possible to use astrophysical arguments which can yield a stronger bound on τ_e. The following reasoning (Okun and Zeldovich, 1978) shows, however, that there is no need for this. Electron energy depends on the value of the electrostatic potential at the point where the electron is found. However, the energy of electrically neutral decay products does not depend on the potential φ. Hence the potential affects the energy and the probability of the decay. The fact that the value of φ is observable is not surprising because the theory is not gauge invariant. An analogous effect was discussed in the preceding section for $m_\gamma \neq 0$.

One cannot formulate a Lorentz-invariant field theory of massless vector particles with positive definite energy if there is no current conservation. So in consideration of unstable electrons one has either to prescribe a finite mass for the photon or to abandon the basic principles of the theory. In what follows we accept the former possibility. Returning to electron decay in an external field φ we note that for a uniform charge distribution the decays stop as soon as the potential becomes equal to m_e. In accordance with Eq. (12.4) this takes place when the charge density reaches the value $\sigma = m_\gamma^2 m_e / 4\pi e$. Taking into account restriction (12.3) this corresponds to an excess of protons over electrons in the universe,

$$\Delta = N_p - N_e \simeq 3 \times 10^{-30} \text{ cm}^{-3}.$$

In the course of the expansion the decay would proceed in such a way that the difference $\Delta = N_p - N_e$ should be constant. This means that the decay rate should be $3H = 10^{48}$ cm^{-3} sec^{-1}. Taking into account that the electron number density today is $N_e \simeq 10^{-6}$ cm^{-3}, we find that the lifetime of electrons under cosmic conditions is longer than 10^{42} sec. An electron in a region of large negative potential could decay much faster. The decay should proceed into a large number $[\sim(m_e/m_\gamma)^{2/3}]$ of soft longitudinal photons (Okun and Zeldovich, 1978; Voloshin and Okun, 1978) because of the huge probability of the longitudinal photon bremsstrahlung, which is proportional to m_γ^{-2}. It is interesting that in the case of the conserved current, on the other hand, the probability of longitudinal photon emission is negligibly small, $\sim m_\gamma^2$.

We noted above that a vector particle interacting with a nonconserved current should be massive. The arguments can be reversed and it can be shown that, due to the smallness of m_γ, electromagnetic current should be conserved and electrons be stable (Voloshin and Okun, 1978). This result seems to be highly probable physically, despite the calculation's being based on divergent Feynman graphs. So from the limit on m_γ it follows that the lifetime of the electron is, if not infinite, extremely large: $\tau_e > \exp(10^{20})$ (arbitrary units).

XIV. COSMOLOGICAL CONSEQUENCES OF SPONTANEOUS SYMMETRY BREAKING IN ELEMENTARY PARTICLE PHYSICS

The principle of spontaneous symmetry breaking has proven very fruitful in elementary particle physics, and its validity has been supported by the recent success of the gauge models. The starting points of the theory are the invariance of the Lagrangian with respect to a symmetry group and a degeneracy of the ground states. An observable ground state is realized by chance and leads to a symmetry violation for physical states [see, for example, the review by Coleman (1975)]. This picture has important consequences for cosmology. The form of the realized ground state, corresponding to a minimum of the free energy, depends on the temperature. At low temperature the symmetry is violated and at high temperature it is restored (Kirzhnits, 1972). In the case of a discrete symmetry the symmetric ground state, which existed in the vicinity of the cosmological singularity, goes over into one of the possible asymmetric states. In causally unconnected regions these vacuum states, generally speaking, should be different. As a result, a domain structure of the universe arises. Cosmological applications of this picture have been discussed by Zeldovich, Kobzarev, and Okun (1974). The domain structure proves to be energetically unfavored and each domain tends to expand and absorb the others. This results in the movement of the domain walls almost with the speed of light. The surface energy of the domain walls, however, proves to be so large (for reasonable values of the parameters) that this structure would destroy the observed isotropy of the universe. This is a rather strong argument against the model of a spontaneous violation of CP invariance proposed by Lee (1973).

Cosmological consequences of theories of spontaneously broken particle symmetries are considered in detail by Linde (1979a). Here we describe them only briefly. First of all we mention the vortex lines which should appear when a U(1) symmetry is spontaneously broken, and magnetic monopoles which are generated by non-Abelian gauge symmetry breaking (Nilsen and Olesson, 1973; Kibble, 1976). As is claimed by Linde (1979a), the vortex lines of the Salam–Weinberg model do not produce significant cosmological effects.[13] As for magnetic monopoles, their cosmological concentration was discussed in Sec. XI, where it was shown that a large discrepancy exists between the observations and the theoretical estimates. It should be noted, however, that the restoration of symmetry depends not only on temperature but also on the value of the fermion number density F in the universe (F is the difference between the number densities of fermions and antifermions). For some models no phase transition from the asymmetric to the symmetric phase occurs if $F \neq 0$ independently of temperature (Linde, 1979a). Such models can explain the absence of monopoles and of the domain structure. The only possible source of the fermion numbers are neutrinos. According to the estimates by Linde, the relative excess $|N_\nu - N_{\bar{\nu}}|/N_\gamma \simeq 1$ is enough for destruction of the phase transition (see, however, Sec XI).

Depending on the value of F, two scenarios of the development of the universe at small t are possible. If $F=0$, then for T exceeding some critical value $T_{\rm crit}$, all the masses except for the mass of the Higgs particles vanish and the universe is hot. Another possibility is that the universe possesses a large leptonic charge and as $t\to 0$ and $T\to\infty$ the minimum of the effective potential of the Higgs field tends to $-\infty$. In this case all the masses tend to infinity and the energy density in the universe is dominated by nonrelativistic particles.

An interesting possibility arises when the weak charges of the leptons and of the baryons are mutually compensated so there are no neutral current interactions. In this case the universe at an initial stage is a dense cold medium consisting of baryons and leptons. In the course of expansion, the fermion density decreases and a phase transition from the symmetric state into the asymmetric state is possible.[14] At the phase transition a large amount of energy is released because the new vacuum state is more favored energetically, and the universe becomes hot. In such a model one can naturally obtain $N_B/N_\gamma = 10^{-8}$–10^{-10} with a cold symmetric initial state. There are models discussed in the literature where the effective potential of the Higgs field has two minima, the relative position of which depends upon temperature and the Higgs particle mass. A very interesting situation is possible when the universe begins at the absolute mimimum of the effective potential, but with dropping temperature this minimum gets higher and becomes a local, but not an absolute, minimum. In this case a tunnel transition into an energetically more favored vacuum is possible. Just this example of the cold universe heating was considered above. The choice of the vacuum depends on the Higgs meson mass. It is possible in principle but difficult to believe that we live in a metastable world. The discovery of the Higgs meson and the measurement of its mass could help to assuage doubts about the stability of the universe. The theory of tunneling can be found in Voloshin *et al.* (1974), Frampton (1977), and Coleman (1978).

In conclusion, we should like to note that the results of this section depend on the mechanism of symmetry breaking. The examples discussed were based on the Higgs mechanism. The picture can be considerably different if, for example, the scheme of Dimopoulos and Susskind (1979a) is valid, in which symmetry violation of the weak and electromagnetic interaction is connected with a Goldstone meson in a world of superheavy particles ($m \simeq 10^4$ GeV). The cosmological consequences of such a model have not yet been considered and present a rich field for investigation.

XV. BARYON ASYMMETRY OF THE UNIVERSE

The world we see around us is (fortunately) extremely charge asymmetric—there are protons, neutrons, and electrons forming all around, but there are no antibaryons and positrons. The question of how this could happen is probably the most important one in fundamental physics and cosmology. There are in principle three possibilities. First, the baryonic excess is assumed to be the "initial condition." It is rather natural in the cold universe model and artificial in the big bang scenario. In the cold universe case one has to invent mechanisms which would give the modern large value of the entropy per baryon $s = 10^8$–10^{10}. Second, the universe is assumed to be charge symmetric but spatially nonuniform, in some regions having an excess of baryons and in others an excess of antibaryons. Third, and most interesting, a dynamical mechanism for the asymmetry is proposed. Much work on this subject is now in progress, but it seems to us that all the sources of the baryon asymmetry which are in principle possible can be enumerated at this time.

---

[13]Recently Zeldovich (1980) investigated the vortex lines of this theory with parameters of the order of those used in grand unified theories ($m \simeq 0.01\ m_P$). Perhaps these vortex lines would explain the perturbations built from an initially Friedmanian universe during cooling.

[14]It is a first-order phase transition with hysteresis and nucleation. Therefore there is an entropy gain by the transition.

This section is organized as follows. In Sec. XV.A we consider entropy generation in an initially cold universe. In Sec. XV.B the spatial separation of matter and antimatter is briefly touched upon. In Sec. XV.C some general statements about the origin of charge asymmetry are formulated, and in Secs. XV.D, XV.E, and XV.F dynamical models of baryonic charge generation are considered.

A. The cold universe and entropy generation

The common feature of papers on this subject is the assumption that the initial state is that of a cold baryon fluid with a small entropy per baryon $s=0$. Later a large amount of energy is somehow released, the universe becomes hot, and the entropy rises. As the energy source, an excitation and subsequent dissipation of sound has been proposed (Zeldovich, 1972). The sound waves could be generated by a nonuniformity in the metric and the matter density. For the extremely rigid equation of state (Zeldovich, 1961) of the baryon fluid, i.e., $\varepsilon=p$, the relative size of the fluctuations should be about 10^{-4}. In this case the value of the entropy per baryon observed today can be determined to be $s_0 \simeq 10^8-10^{10}$. For uniform and isotropic cosmological models the increase in entropy due to bulk viscosity (Weinberg, 1971) or to quark condensation into hadrons (Kazanas, 1978) proves to be negligible in comparison with the necessary value s_0. If this condensation is going smoothly, without hysteresis, supersaturation, or nucleation, the process is an adiabatic. The expansion is governed by gravitation, and is slow compared with elementary particle processes, and entropy generation is small.

Just the opposite point of view was advocated recently by Lasher (1979). He assumed that the quark–nucleon phase transition was of the first order and the quark phase survived rather late after the transition point. The parameters of this model can be chosen so that the observed value of entropy per nucleon can be obtained. Lasher's assertion, however, that the model predicts density fluctuations sufficient for the formation of galaxies is open to question. The density fluctuations are of an unstable type and so they tend to decrease (Zeldovich, 1980).

A rather powerful source of entropy could be primary black hole evaporation (Novikov *et al.*, 1979; Hawking, 1975). By the estimates of Novikov *et al.* the entropy s_0 observed today could arise from this mechanism if about one half of the matter in an initially cold universe collapsed into black holes with a mass of $M=10^{-4}\Omega^{-1}$ g.

An interesting possibility for entropy generation was considered by Linde (1979a). In this model it was assumed that at the initial moment the energy density was dominated by the cosmological term, and that this was connected with a local (not absolute) minimum of the effective potential of the Higgs field around $\chi=0$ (see Sec. XIV). In the course of the universe's expansion a condensate of the χ field is formed (spontaneous symmetry breaking). The new vacuum state could have a smaller energy than the initial vacuum state. The energy of the initial vacuum (or in other words, of the cosmological term) would go over into matter energy. Thus by choosing values of the parameters one could obtain the needed values of s_0.

B. Spatial separation of matter and antimatter

In a paper by Omnes (1969) it is assumed that due to repulsion of B and $\bar{B}$ at a temperature of about $0.3m_N$, thermodynamic equilibrium leads to separation into two phases with B excess and $\bar{B}$ excess. Owing to surface tension and annihilation, these domains grow to galaxy size. The idea of the repulsion between nucleons and antinucleons has been seriously criticized [see, for example, Zeldovich and Novikov (1975)]. We should only like to mention here the paper by Bogdanova and Shapiro (1974), in which arguments are presented in favor of an attraction between nucleons and antinucleons. We should also like to note that for a temperature larger than 300 MeV, where according to Omnes the separation between matter and antimatter takes place, there are probably no nucleons but free quarks in the plasma. The quark interaction is now believed to be described by QCD, which does not lead to matter–antimatter separation. So this mechanism probably is not realized.

C. Some general theorems on particle-antiparticle asymmetry

Some of the statements presented below are well known and are included here for completeness. First we formulate the conditions for charge asymmetry at the elementary particle level. A necessary condition is of course a violation of C invariance. If, however, CP is conserved, no excess of particles over antiparticles emerges in developing a symmetric initial state. A violation of C invariance with simultaneous CP conservation leads to correlations of spins or angular momenta with particle momenta, or to T-odd triple products of momenta. The number of particles and antiparticles can be different in different regions, of phase space, but after summing up the spin projections and integrating over the phase space, the difference disappears and the total probabilities of the processes $i\to f$ and $\bar{i}\to\bar{f}$ (where $\bar{i}$ is the charge-conjugated state with respect to i) should be equal: $\Gamma_{if}=\Gamma_{\bar{i}\bar{f}}$. So, to obtain the inequality $\Gamma_{if}\neq\Gamma_{\bar{i}\bar{f}}$ a joint violation of C and CP is necessary. Instances of C and CP violation have been established experimentally. The branching ratios of the neutral K_L meson decays—into the charge-conjugated final states $K_L^0\to\pi^+e^-\nu$ and $K_L^0\to\pi^-e^+\bar{\nu}$—are proved to be different. Hence one can expect an asymmetric result in the development of a symmetric initial state.

Further restrictions on the dynamics of charge asymmetry generation can be obtained by using the CPT theorem. This theorem states that the amplitude of a charge-conjugated, time-reversed and parity-reversed process is equal to the amplitude of the original process. In particular it leads to equal masses and lifetimes for particles and antiparticles and to equality of the total cross section, $\sum_f\Gamma_{if}=\sum_f\Gamma_{\bar{i}\bar{f}}$, but the decay branching ratios as well as the probabilities of specific reaction channels can be different.

Differences between the probabilities of charge-conjugated elementary processes can arise if several (two or more) *inelastic* reaction channels are open. This can be easily proved with the help of the unitarily relation

$$i(T_{if}-T^{\dagger}_{if})=-\sum_n T_{in}T^{\dagger}_{nf}=-\sum T^{\dagger}_{in}T_{nf}\,, \qquad (15.1)$$

where T_{if} is the amplitude of the transition from an initial state i into a final state f. The T matrix is related to the S matrix by $S=1+iT$. Equation (15.1) follows from the S-matrix unitarity condition $SS^{\dagger}=S^{\dagger}S=1$. The summation over intermediate states includes integration over phase space. Owing to the CPT theorem, $T_{fi}=\tilde{T}_{\bar{i}\bar{f}}$, where the tilde means PT transformation over the variables, namely, change of spin signs. Taking into account the identity $T^{\dagger}_{if}=T^{*}_{fi}$ one finds $T_{if}=\tilde{T}_{fi}$ if the right-hand side in Eq. (15.1) can be neglected. Thus in the Born approximation probabilities of charge-conjugated processes summed over spin projections are equal to each other, $\Gamma=\bar{\Gamma}$. A stronger result can also be proved, namely, if only two reaction channels i and f, are open, then independently of rescattering, $\Gamma=\bar{\Gamma}$. Indeed in this case it follows from Eq. (15.1) that

$$2\,\mathrm{Im}\,T_{ii}\{\lambda\}=\int|T_{ii}\{\lambda\}|^2 d\tau_i+\int|T_{if}\{\lambda\}|^2 d\tau_f\,,$$

where $\{\lambda\}$ is the longitudinal component of the spin variables (e.g., helicities) and $d\tau$ is the phase-space element. The CPT theorem ensures that $T_{ii}\{\lambda\}=T_{\bar{i}\bar{i}}\{-\lambda\}$. So after summing over polarizations we find $\Gamma_{if}=\Gamma_{\bar{i}\bar{f}}$. Hence to destroy the equality $\Gamma=\bar{\Gamma}$ at least three open channels are needed, $i\to f$, $i\to k$, and $k\to f$.

Consider now the decays of a neutral particle into two charge-conjugated sets of states f and $\bar{f}$. If there is no direct transition $f\to\bar{f}$ then

$$\sum_f\Gamma(\kappa\to f)=\sum_f\Gamma(\kappa\to\bar{f})\,. \qquad (15.2)$$

In fact the probability of the decay into f is

$$\Gamma(\kappa\to f)=\langle\kappa|H|f\rangle_{\rm out\;out}\langle f|H|\kappa\rangle$$

where H is the decay Hamiltonian, $|f\rangle_{\rm out}$ is the state of outgoing waves when $t\to\infty$, and $|f\rangle_{\rm in}$ is that of ingoing waves. For the one-particle states $|\kappa\rangle_{\rm in}=|\kappa\rangle_{\rm out}$. Because of CPT invariance

$$\Gamma(\kappa\to f)=\langle\kappa|H|\bar{f}\rangle_{\rm in\;in}\langle\bar{f}|H|\kappa\rangle\,. \qquad (15.3)$$

Substituting into (15.3) the complete set of states and summing over f, we find

$$\sum_f\Gamma(\kappa\to f)=\sum_{\kappa,l,f}\langle\kappa|H|k\rangle_{\rm out\;out}\langle k|\bar{f}\rangle_{\rm in\;in}\langle\bar{f}|l\rangle_{\rm out\;out}\langle l|H|\kappa\rangle\,.$$

Here use was made of the condition $\sum_k|k\rangle\langle k|=1$. If there is no transition $f\to\bar{f}$ then, first, the sets k and l should coincide with $\bar{f}$ and, second, the S matrix, $S={}_{\rm out}\langle k|f\rangle_{\rm in}$, is unitary in subspaces f and $\bar{f}$ separately. So Eq. (15.2) is valid. Along similar lines it can be proved (Nanopoulos and Weinberg, 1979) that no baryon asymmetry arises in the first order of B-nonconserving interactions.

The arguments presented above cannot be directly applied to the case of two degenerate levels, analogous to the system $(K^0-\bar{K}^0)$, between which mutual transitions are possible. If CP is violated, diagonalization of such a system with respect to a total Hamiltonian results in two eigenstates κ_L and κ_S which are not eigenstates of the CPT operator (Lee and Wu, 1966). This is a cause of charge asymmetry even in the case of negligible rescattering. A well known example is the charge asymmetry in the decays $K^0_L\to\pi^+e^-\bar{\nu}$ and $K^0_L\to\pi^-e^+\nu$.

Thus at the level of elementary particles there is an asymmetry which leads to different numbers of particles and antiparticles. This mechanism could be responsible for the baryonic excess in the universe. For this to be true, it would be necessary of course that baryonic charge not be a strictly conserved quantum number, even though the observed proton stability convinces us of the opposite. The large proton lifetime $\tau_p>10^{30}$ yr and B nonconservation could be compatible if a B-nonconserved interaction switches on at a very high energy. At high temperatures, however, an important factor smoothing out-charge asymmetry comes into play. The point is that at high temperature the primeval plasma is in thermodynamic equilibrium, a state in which the contribution of every reaction must be considered. Let, for example, baryonic charge be generated by the decays of a neutral meson $\kappa\to B\bar{L}$ and $\kappa\to\bar{B}L$, where B is a baryon and L is a lepton and $(B\bar{L})\leftrightarrow(\bar{B}L)$. In the equilibrium state not only decays of κ but also the inverse reactions $B\bar{L}\to\kappa$ and $\bar{B}L\to\kappa$ as well as direct scattering $B\bar{L}\leftrightarrow\bar{B}L$ should be taken into consideration. Taking into account all possible processes leads to the vanishing of particle–antiparticle number differences. The absence of charge asymmetry in an equilibrium state was noted by Okun and Zeldovich (1976) and later in a series of papers on the generation of baryon asymmetry by means of the mechanism discussed here.[15]

If T invariance is violated, however, the validity of standard equilibrium thermodynamics can be questioned. There is no detailed balance condition in this case but nevertheless S-matrix unitarity proves to be sufficient to provide the classical equilibrium distributions. So if there is no conserved quantity, the chemical potential vanishes in equilibrium and there is an equal number of particles and antiparticles. At the last step the equality between masses, owing to CPT, is essential. We return to this later but for now we note that because of the expansion of the universe the thermodynamic equilibrium is destroyed, and so is the balance of the reactions which maintain the condition $B=0$. Thus during the nonequilibrium stage, which comes about because of the expansion of the universe, a nonzero baryonic charge can arise.

A very attractive feature of this model is that the result is independent of initial conditions. At the beginning one can have anything, $B=0$ or $B\neq0$, but during the equilibrium stage everything is smoothed out and, necessarily, $B=0$. Subsequent baryon generation in the nonequilibrium stage is completely determined by the dy-

[15]For a formal proof see Dimopoulos and Susskind (1979b).

namics. This feature is inherent in all models [Sakharov (1967); Kuzmin (1970); and new ones] but it has been stated especially clearly by Ellis, Gaillard, and Nanopoulos (1979). Okun and Zeldovich (1976) use it as an argument against the Pati–Salam model with unstable quarks.

In the literature there is the incorrect statement that the baryon excess generated by the decays $\kappa \to B\bar{L}$ and $\kappa \to \bar{B}L$ is compensated in thermal equilibrium by the inverse processes $B\bar{L} \to \kappa$ and $L\bar{B} \to \kappa$. However, owing to the CPT theorem,

$$T(\kappa \to B\bar{L}) = T(\bar{B}L \to \kappa) \quad \text{and} \quad T(\kappa \to \bar{B}L) = T(\bar{L}B \to \kappa),$$

so the contribution of inverse decay processes can only increase the baryon excess. The compensation is achieved by the direct processes $B\bar{L} \longrightarrow \bar{B}L$ which, as was noted above, should exist to provide the inequality $\Gamma(B\bar{L}) \neq \Gamma(\bar{B}L)$. An analogous paradox can be found by examining decays of degenerate states κ_L and κ_S. It can be shown that charge asymmetries in decays of κ_L and κ_S are of the same sign. The rescattering in the final state can be negligible. It seems thus that even in thermal equilibrium, decays of κ_L and κ_S could produce a baryon excess. As in the preceeding example, the inverse processes can only make it worse. The contradiction can be resolved if one takes into account the interference effects in decays of a coherent mixture of κ_L and κ_S. One can check this statement by considering the CPT constraints on the scattering processes $BX \to \kappa_{L,S} \to BY$ in the case when direct transitions $BX \longrightarrow BY$ are negligible.

We shall now show that S-matrix unitarity (in fact a weaker condition) provides the standard statistical equilibrium distributions (Dolgov, 1979; Weinberg, 1979; Toussaint *et al.*, 1979). Let there be the processes

$$a_i + b_i + c_i + \cdots \to a_k + b_k + c_k + \cdots .$$

We consider the time variation of the number density of a_i:

$$\dot{N}(a_i) = \frac{d}{dt}\int n(a_i)\,d^3p = \text{const}\int d^4P\,d\tau_i \sum_k d\tau_k\,(\Pi n_k |A_{ki}|^2 - \Pi n_i |A_{ik}|^2), \tag{15.4}$$

where $n(a_i)$ is the distribution function in momentum space for a_i, $\Pi n_k = n(a_k)n(b_k)n(c_k)\cdots$; $P = p_a + p_b + \cdots$ is the total momentum

$$d\tau_i = \delta^4(P - \Sigma p_i)\Pi d^3p_i/[(2\pi)^3\,2E_i]$$

is the phase space element, and A_{ik} is the amplitude of the transition from i into k. A_{ik} is connected with the T matrix introduced above by the relation $T_{ik} = (2\pi)^4\delta^4(P - \Sigma p_i)A_{ik}$.

To simplify the notations we assume that the occupation numbers are small and the corresponding quantum statistical corrections are not essential. Taking these corrections into account leads only to unessential technical complications, as will be sketched below. We have to check that the Boltzman distribution function

$$n(a_i) = \exp\{[\mu(a_i) - E(a_i)]/T\}$$

indeed represents an equilibrium, i.e., that it provides the equality $\dot{N} = 0$. Because of the conservation of chemical potential (in equilibrium) and of energy, the following relation holds:

$$\Pi n_i = \Pi n_k . \tag{15.5}$$

Hence Eq. (15.4) can be rewritten as

$$\dot{N}(a_i) = \text{const}\int d^4P\,d\tau_i\,\Pi n_i \sum_k d\tau_k\,(|A_{ki}|^2 - |A_{ik}|^2). \tag{15.6}$$

In a T-invariant theory, $|A_{ik}| = |A_{ki}|$ (the detailed balance condition) and $\dot{N}(a_i) = 0$. If T invariance is violated, the equality $\dot{N}(a_i) = 0$ is valid as before, but because of the weaker condition

$$\sum_k \int d\tau_k\,(|A_{ki}|^2 - |A_{ik}|^2). \tag{15.7}$$

This expression can be named the cyclic balance condition. Note the resemblance to the cyclic equilibrium of electrons in a magnetic field.

Equation (15.7) follows from S-matrix unitarity $SS^\dagger = S^\dagger S = 1$. However, not all the unitarity restrictions are needed; we need only to normalize the diagonal elements of S, i.e., the probability $\sum_f w_{if} = 1$ (the probability of anything happening is unity) and the inverse relation $\sum_f w_{fi} = 1$. Apart from this, S-matrix unitarity demands the vanishing of the off-diagonal elements as a result of the completeness of the basis of physical states.

One may derive Eq. (15.7) by using only the physically trivial condition $\sum_f w_{if} = 1$ and CPT invariance and saying nothing of unitarity.

The conclusion that the form of equilibrium distributions is independent of T invariance is of course valid when the occupation numbers are not small. One has to check that the condition $\dot{N}(a_i) = 0$ is fulfilled when

$$n(a_i) = \frac{\exp\{-[E(a_i) - \mu(a_i)]/T\}}{1 \pm \exp\{-[E(a_i) - \mu(a_i)]/T\}}. \tag{15.8}$$

If the occupation numbers are not small, we substitute $\Pi n_i \to \Pi n_i \Pi(1 \pm n_k)$ in Eq. (15.3) and $d\tau_k \to d\tau_k \Pi(1 \pm n_k)$ in Eq. (15.7). They are connected with the well known factor $\langle 1+n|\,a^\dagger|\,n\rangle \sim \sqrt{n+1}$ for Bose fields and the exclusion principle for Fermi fields ($a^\dagger$ is a particle creation operator). Note that the dependence of the density on the initial particles is of the same form as before when n was considered to be small. One more remark about the substitution of n dependence into the unitarity condition should be made. Because of unitarity the following relation (say for Bose fields) is valid:

$$\sum_k |\langle\{n_i\},\{n_k\}|T|\{n_i - 1\},\{n_k + 1\}\rangle|^2 = \sum_k |\langle\{n_k + 1\},\{n_i - 1\}|T|\{n_i\},\{n_k\}\rangle|^2 .$$

Each term in this equation is proportional to $\Pi(n_k + 1)\Pi n_i$. The dependence on Πn_i is factored out so Eq. (15.7), with the substitution mentioned above, is obtained. It is noteworthy that there is some dependence of the transition amplitudes on occupation numbers which is not factorizable and is unknown if the theory is not specified, but this is unimportant for our results.[16]

For functions (15.8) the relation

$$\Pi n_i \Pi(1 \pm n_k) = \Pi\, n_k \Pi(1 \pm n_i) \tag{15.9}$$

provides the equilibrium condition $\dot{N} = 0$.

Let us turn now to the nonequilibrium case. Deviations from equilibrium appear because of the expansion of the universe. This is taken into account by the addition of the term $-3HN(a_i)$ into the right-hand side of Eq. (15.6), where $N(a_i) = \int d^3p\, n_i(p)/(2\pi)^3$ is the total number of particles of type i in a unit volume. If H is much smaller than the characteristic reaction rates this term can be neglected and the equilibrium has time to follow the expansion. Following the paper by Dolgov (1979) we consider the case when baryonic charge nonconservation is provided by the reaction

$$qq \to \bar{q}l\ , \tag{15.10}$$

where q is a quark and l is a lepton.[17] Of course for charge symmetry violation other reaction channels should be open. The particular form of the reactions is, however, unimportant for what follows. The arguments presented are of general validity and are not based on the specific form of B-violating reactions.

As the universe expands the slower processes are the first to be frozen out. So when reaction (15.10) becomes too slow to keep pace with the expansion, processes with q and l which proceed through the usual strong or weak interaction are still in equilibrium. Because of this the quark density in phase space should be

$$n_q = \exp\left[\frac{\mu_q(t) - E}{T}\right]\ . \tag{15.11}$$

The density for $\bar{q}$, l, and $\bar{l}$ should be analogous. Quantum corrections [see Eq. (15.8)] will not change the results but will lengthen the formulas. In expression (15.11) we neglected terms of the order $H\tau_{\text{tot}}$, where $\tau_{\text{tot}} = (\sigma_{\text{tot}} v_n)^{-1}$ is the characteristic time (inverse rate) of reactions in which q and/or l take part. These terms will be taken into account later. We should like to note that in this approximation no baryonic excess is generated even if reactions with B nonconservation are out of equilibrium ($H\tau_{\Delta B} \gtrsim 1$) while other reactions with q and l are still in equilibrium.

The nonequilibrium character of reaction (15.10) suggests a possible violation of the equality $2\mu_q = \mu_{\bar{q}} + \mu_l$. If, however, at the initial instant t_0 the condition $\mu_q(t_0) = \mu_{\bar{q}}(t_0)$ were fulfilled, giving an equal amount of quarks and antiquarks, it would be valid also after reaction (15.10) and other processes with $\Delta B \neq 0$ went out of equilibrium. The equality $\mu_q = \mu_{\bar{q}} = 0$ is fulfilled until processes which are faster than (15.10) provide the canonical form of the phase-space distribution function (15.11) and in particular until these processes equalize temperatures of different sorts of particles. Indeed, let $\mu_q(0) = \mu_{\bar{q}}(0) = 0$. When t is small [and correspondingly $\mu_q(t)$ and $\mu_{\bar{q}}(t)$ are small] the proportionality $\Delta N = N_q - N_{\bar{q}} \simeq \mu(t)$ holds. Here use was made of the condition $\mu_q(t) = -\mu_{\bar{q}}(t) \equiv \mu(t)$ which should be valid provided that fast processes, e.g., $q\bar{q}$ annihilation into photons and elastic redistribution over phase space, are in equilibrium. It follows from Eq. (15.4) that for small t,

$$N(q) = F(t)\mu(t)\ , \tag{15.12}$$

$$N(\bar{q}) = \bar{F}(t)\mu(t)\ . \tag{15.13}$$

Subtracting these equations from each other we obtain $\dot{\mu} = a\mu$ ($a < 0$). The only solution of this equation such that $\mu(0) = 0$ is $\mu(t) = 0$.[18] This result is stable with respect to fluctuations.

To get a quantitative estimate of the baryon asymmetry resulting from a departure from chemical equilibrium which in turn is caused by the expansion of the universe, we proceed as follows. Let $n_i = n_{eqi} + n_i'$ where n_{eq} is the equilibrium value (for a given temperature) of the particle density in phase space. Substituting it into Eq. (15.4) with the expansion term taken into acount, we obtain

$$\dot{N}_i' = -3HN_i' + S_i(n_{eqj} + n_j') - (3HN_{eqi} + \dot{N}_{eqi})\ . \tag{15.14}$$

Here $N = N_{eq} + N_{eq}'$ is the total density of particles of type i and S is the collision integral; note that $S(n_{eq}) = 0$. The value of N_{eq} is given by

$$N_{eq} = \frac{g_S T^3}{2\pi^2} \int_0^\infty \frac{x^2\, dx}{\exp\{[(x^2 + m^2)/T^2]^{1/2}\} \pm 1}\ ,$$

where T satisfies the equation $\dot{T} = -HT$. Hence for the function $F_i = 3HN_{eqi} + \dot{N}_{eqi}$ we obtain

$$F_i = 2\left(\frac{m_i}{T}\right)^2 HT^3 \frac{dI}{dz}\ , \tag{15.15}$$

where

$$I(z) = \int_0^\infty dx\, x^2[\exp(x^2 + z)^{1/2} \pm 1]^{-1}$$

and $z = (m/T)^2$. For $m < T$, $F(m/T) \sim (m/T)^2 HT^3$; for $m > T$, $F(m/T) \sim H(mT)^{3/2}\exp(-m/T)$; and for $m \simeq T$, $F(m/T) \sim HT^3$.

Note that F vanishes if $m = 0$. This means that no deviation from equilibrium occurs for massless particles (Toussaint *et al.*, 1979).

To proceed further one needs to specify the form of the collision integral S. Strictly speaking no self-contained equation can be obtained for total number den-

[16]All this is done in the ideal gas approximation. This is justified for asymptotically free field theories. It seems, however, that the conclusion is valid in the general case.

[17]Note that this reaction respects color symmetry and fermionic charge conservation $F = B - L$. In some models discussed in the literature the fermionic charge F is strictly conserved, so to get the observed baryon excess, one has to assume that the "initial value" of F is small. In this case the lepton asymmetry of the universe is of the same order as the baryon one. If F were not small initially, the fermionic charge would now be lost in the neutrino sea. In the framework of the models considered this could happen only if the masses of fundamental baryons were larger than the freezing temperature of B-nonconserved reactions. This is not the case, however.

[18]Note the resemblance of this to the statement (Zeldovich, 1938) that the equilibrium in a mixture of ideal gases is unique and charge symmetric.

sities N_i because S depends on phase-space number densities $n_i(p)$. With plausible approximations, however, Kolb and Wolfram (1980) have found a system of coupled kinetic equations governing the behavior of total number densities N_i. They considered the simplified color-nonconserving model in which baryon asymmetry arises in reactions $X \leftrightarrow qq$ and $X \leftrightarrow q\bar{q}$ where X is a heavy meson. The numerical solution of these equations was found, showing that in the case of the light X meson (or strong coupling of X to quarks) baryon asymmetry was suppressed. A more realistic color-conserving model has been considered by Dolgov (1980b), who discussed the decays $X \leftrightarrow qq$ and $X \leftrightarrow \bar{q}\bar{l}$. Kinetic equations for phase-space number densities $n(p)$ were solved analytically in the limit of strong coupling of X mesons to fermions. The solution was in qualitative agreement with that of Kolb and Wolfram. Fry, Olive, and Turner (1980a, b, c) have solved the kinetic equations numerically for the SU(5) theory and their results agree quantitatively with those of Kolb and Wolfram (1980).

The result obtained for strong coupling by Dolgov (1980) matches very well with the weak coupling solution found by Weinberg (1979) (see Sec. XV.E). It seems to be accurate for all values of the coupling constant. The algebra is rather complicated, however, and in what follows a simplified treatment of Eq. (15.14) is given. The result obtained in this way is correct to within an order of magnitude.

Consider the equation for the quark number density. The last term in equation (15.14) in this case vanishes because $m_q = 0$. There are deviations from equilibrium, however, because some other particles interacting with quarks are massive (in our example the X boson has a nonvanishing mass). In the equation governing the time dependence of N_X [Eq. (15.14)] we make the substitution $S_X = -N_X'\Gamma$, where Γ is the decay width of the X meson. Now the solution of Eq. (15.14) is easily found:

$$N_X'(t) = e^{-\Gamma t}\int_0^t dt' e^{\Gamma t'} F_X(t') , \tag{15.16}$$

where $F_X(t)$ is defined by Eq. (15.15) with $T = T(t)$. Note that Γ is a constant and $N' \to 0$ as $t \to \infty$.

In the equation for N_q we approximate S_q as $S_q = \tilde{\Gamma}(t)(aN_q' + bN_X')$, where a and b are some numerical factors of the order of unity. The essential point is that $\tilde{\Gamma}(t)$ is not a constant but a decreasing function of time. In the example considered quarks are produced and annihilate during the inverse and direct decays of the X meson so

$$\tilde{\Gamma}(t) = \Gamma \exp(-m_X/T) .$$

Now for the baryon number per comoving volume $\Delta = (N_q - N_{\bar{q}})/T^3$ the following approximate equation holds:

$$\dot{\Delta} \simeq -\Gamma \exp(-m_X/T)\Delta + N_X' T^{-3}\Delta\Gamma , \tag{15.17}$$

where $\Delta\Gamma$ is the difference between the partial decay rates of $X \to qq$ and $\bar{X} \to \bar{q}\bar{q}$. We have neglected here the term $N_q'\Delta\Gamma$ and some others which could change the result by a numerical factor of the order of unity.

The solution of Eq. (15.17) is

$$\Delta = \Delta\Gamma \exp\left\{-\int_0^t dt'\tilde{\Gamma}(t')\right\} \times \int_0^t dt' \exp\left\{\int_0^{t'} dt'' \tilde{\Gamma}(t'')\right\} N_X'(t')T^{-3}(t') . \tag{15.18}$$

For large decay width Γ such that $\Gamma t_m \gg 1$ [where t_m is defined by the condition $T(t_m) = m_X$] the integral can be easily evaluated and we finally get

$$\beta = \frac{N_B}{N_\gamma} = (10^{-1}\text{–}10^{-2})\,\frac{\Delta\Gamma}{\Gamma}\,\frac{1}{\Gamma t_m} . \tag{15.19}$$

The first numerical factor in Eq. (15.19) is connected to the increase of the photon number with time because of annihilation of different massive particles which were present in the primeval plasma.

This result shows that if a particle interacts strongly when $T \simeq m$ (i.e., $\Gamma t_m > 1$), its contribution to baryon asymmetry is suppressed by the factor $(\Gamma t_m)^{-1}$. As we shall see below, this is just the case of superheavy gauge bosons X with the currently accepted mass value of 3×10^{14} GeV (see Sec. XV.E). As for Higgs particles H with about the same mass, no such suppression appears because the coupling strength of H to fermions is approximately two orders of magnitude smaller than that of X.

The results of this section are used below (Sec. XV.D and XV.E) where some more specific models are discussed.

D. Concealment of baryon number in black holes

The quantum radiation of black holes discovered by Hawking (1975) makes possible the generation of baryon asymmetry from a symmetric initial state with (on the field theory level) strictly conserved baryonic charge. In the independent particle approximation a black hole evaporates baryons and antibaryons symmetrically. It was conjectured by Hawking (1975) that a CP-noninvariant interaction could destroy the B-$\bar{B}$ symmetry. Zeldovich (1976) has given an example of how CP-noninvariant decay of a hypothetical particle, emitted by a black hole, could lead to asymmetry. In the recent paper by Toussaint *et al.* (1979) it is stated, however, that if baryonic charge is microscopically conserved the net flux of baryons from a black hole is equal to that of antibaryons, so no charge asymmetry occurs in the space outside the black hole. We disagree with this statement and prove the opposite with an explicit example (Zeldovich, 1976). Let there exist a neutral meson A with decay modes $A \to q\bar{h}$ and $A \to \bar{q}h$. Owing to C (and CP) invariance violation,

$$\Gamma(q\bar{h}) > \Gamma(\bar{q}h) . \tag{15.20}$$

Here q is a light quark (or a baryon) and h is a heavy one. As was noted in Sec. XV.C the total set of the reactions $A \leftrightarrow X_i$ and $X_i \leftrightarrow X_j$ in equilibrium provides an equal number of particles and antiparticles. The process of quantum radiation of a black hole into empty space is, however, essentially a nonequilibrium one. In particular, particle collisions practically do not happen because of the low density outside the black hole. The outstreaming particles occupy a cone in phase space, in every point in space outside the black hole horizon.

The net number of baryons generated by the decay is equal to that of antibaryons. If, however, the masses are such that light baryons q are extremely relativistic and heavy ones h are slow, there will be a baryon asymmetry in the space outside the black hole. The point is that owing to Eq. (15.20) the number of heavy baryons is larger than that of light baryons and the probability of the inverse gravitational capture, that is, the reabsorption by a black hole of particles which had evaporated from it, is larger for a slow particle than for a relativistic one. That is why a black hole could accumulate antibaryonic charge and generate a baryonic excess in the space outside the black hole.

For numerical estimates we assume that the temperature of the black hole is $T_{BH} \simeq m_A$ [for a lower temperature $T_{BH} < m_A$ the A-meson emission would be suppressed by the factor $\exp(-m_A/T_{BH})$]. Hence the black hole mass is

$$m_{BH} = \frac{1}{8\pi} (GT)^{-1} = \frac{m_P^2}{25 m_A} , \tag{15.21}$$

where $m_P = G^{-1/2} = 10^{19}\ m_N$. The evaporation time for such a black hole is

$$t_{\text{vap}} = (G^2 m_{BH} T^4)^{-1} (8\pi N_{DF})^{-1} = m_P^2/(N_{DF}\, m_A^3) , \tag{15.22}$$

where N_{DF} is the number of particle species with masses smaller than m_A, $N_{DF} \simeq 10^2$. The mechanism described above would be operative if $t_{\text{vap}} > \tau_A$, where τ_A is the lifetime of the A meson. In grand unified theories for a sufficiently large m_A the estimate $\tau_A = (\alpha m_A N_{DF})^{-1}$ holds where $\alpha \simeq 10^{-2}$ is the unification coupling constant. Hence

$$m_A < \sqrt{\alpha}\ m_P . \tag{15.23}$$

The estimate used here for τ_A is valid for the superheavy gauge boson which exists in some versions of the grand unified theories. If the A meson has different properties, its lifetime can be different but the condition $\tau_A < t_{\text{vap}}$ should be satisfied.

At the instant $t = t_{\text{vap}}$ the energy in the universe is

$$\rho = C(G t_{\text{vap}}^2)^{-1} = C N_{DF}^2\, m_A^6\, m_P^{-2} , \tag{15.24}$$

where the constant C depends on the equation of state. For relativistic gas $C = 3/(32\pi)$.

Black hole radiation produces particles with energy of the order of $T_{BH} = m_A$ and number density of a certain type $N_i = \rho/(N_{DF} T_{BH})$. In particular the baryon charge density is

$$B = C\, \delta (N_{DF}^B / N_{DF}) N_{DF}^2\, m_A^5\, m_P^{-2}\, \eta , \tag{15.25}$$

where N_{DF}^B is the number of baryon species. We assume that $N_{DF}^B/N_{DF} = 0.5–0.25$, η is determined by the difference in probabilities of the inverse captures of heavy and light baryons, δ is the value of the charge asymmetry in an elementary process, $\delta = [\Gamma(q\bar{h}) - \Gamma(\bar{q}h)]/\Gamma_{\text{tot}}$. No asymmetry is known to exist in the Born approximation, so δ is small. Its value is of course model dependent and, strictly speaking, unknown. In the existing models of CP violation (see some discussion at the end of Sec. IV.E) δ is probably about 10^{-6}–10^{-8}.

After the release of energy due to black hole evaporation a new thermal equilibrium is established with the temperature T determined by

$$\rho = \frac{\pi^2}{15} N_{DF} T_e^4 , \tag{15.26}$$

where ρ is defined by Eq. (15.24). The inverse specific entropy on a unit of baryonic charge is

$$s^{-1} = \frac{B T_e}{\rho} = C^{1/4} N_{DF}^{1/4} \frac{N_{DF}^B}{N_{DF}} \left(\frac{15}{\pi^2}\right)^{1/4} \left(\frac{m_A}{m_P}\right)^{1/2} \eta\delta$$

$$\simeq 0.5 \left(\frac{m_A}{m_P}\right)^{1/2} \eta\delta . \tag{15.27}$$

The observed value of the baryonic asymmetry can be obtained if $m_A > 10^{-4} m_P = 10^{15}$ GeV (if $\delta = 10^{-8}$). Some years ago such masses seemed ridiculously huge but now the attitude towards them is changed thanks to grand unified theories in which gauge mesons with about the same masses are postulated.[19] With the parameters chosen the mass of the black holes should be ~0.1 g. Cosmological data do not contradict the existence of such black holes (Novikov *et al.*, 1979) but no reliable estimates of their number are available.

A somewhat different mechanism has been considered by Dolgov (1980c). Particle behavior in the vicinity of a black hole is known to be governed by the equation

$$[\partial_\xi^2 - \epsilon^2 - V(\xi)] R(\xi) = 0 ,$$

where ϵ is the particle energy, $\xi = \rho + \ln(\rho - 1)$, ρ is the usual radius, and V is a potential (everything is in units of the inverse gravitational radius of a black hole). The potential V vanishes as $\xi \to -\infty$ and V tends to the particle mass squared as $\xi \to +\infty$; a potential barrier is in between. It was assumed that an A meson could decay before penetrating the barrier. In this case the condition $T_{BH} \gtrsim m_A$ need not be fullfilled; only the weaker condition $T_{BH} \gtrsim m_q$ should be respected. In fact if $T_{BH} \simeq m_q$, only light baryons could penetrate to infinity and owing to C and CP violation [see Eq. (15.20)] the flux of q could exceed that of $\bar{q}$.

Returning to the paper by Toussaint *et al.* (1979), we should like to note that the contradiction between our results and theirs is probably due to an oversimplification of their model. They describe particle transmission through the gravitational field outside the event horizon by the Lagrangian $\varphi_i V_{ij}(x) \varphi_j$. This leads to a linear equation of motion. However, the processes of particle production (e.g., decays) are not taken into account in their model Lagrangian. If particle number can be changed in a reaction (e.g., one particle becom-

[19] For a noticeable difference in the inverse capture probabilities of h and q, m_h should be of the order of m_A. Fermions that heavy cannot be squeezed into the framework of the standard Weinberg–Salam model (see discussion in Sec. XI). So the described mechanism could operate only if superheavy fermions with new interactions existed.

ing two or more) the corresponding equation becomes nonlinear and the arguments of that paper, as they are presently stated, become inapplicable.

Recently Turner (1979) discussed baryon production by primordial black holes. With plausible assumptions about the mass spectrum of black holes he obtained a result for β close to the observed value. The author, however, erroneously claimed that no baryonic excess could be generated by this mechanism if baryonic charge were conserved in microscopic processes.

Baryon asymmetry in a charge-invariant theory was considered by Beletsky *et al.* (1979), who assumed that antimatter disappeared in black holes because of statistical fluctuations. The probability of this happening is very small, but a single event could happen with any probability.

E. Baryon asymmetry of the universe in a $C(CP)$-violating B-nonconserving theory of elementary particles

This possibility, first prophetically suggested by Sakharov (1967), has become very popular recently in connection with the ideas of the grand unified theories.[20] In most such models leptons and quarks are in the same multiplet of a symmetry group and so baryonic charge nonconservation naturally results. Ways in which baryonic charge conservation might be retained are reviewed by Slansky (1978). The observed stability of the proton despite B conservation is explained by the extremely large mass of the intermediate mesons which interact with B-nonconserved currents. A dimensional estimate of proton lifetime gives $\tau_p = \alpha^{-2} m_X^4/m_N^5$. With $m_X = (10^{15}-10^{16})$ GeV one obtains $\tau_p = (10^{33}-10^{37})$ yr. More recent estimates of the proton lifetime however, give smaller values. Experimentally $\tau_p > 10^{30}$ yr.

The superweak B-nonconserving interaction becomes stronger with energy and for $E \simeq m_X$ it becomes of the same order as all other (except for gravitational) interactions. In particular for X-meson decays $\Delta\Gamma/\Gamma \simeq 1$ ($\Delta\Gamma$ is the width of decays with $\Delta B \neq 0$). Meson decays as a source of the baryon asymmetry in the universe were first considered by Sakharov (1979), Toussaint *et al.* (1979) and Weinberg (1979). Weinberg assumed that $m_X < m_P = 10^{19}$ GeV and that for $t < t_P = m_P^{-1}$, gravitational interactions were strong and capable of establishing thermal equilibrium at the instant $t = t_P$ and enforcing the condition $\Delta B(t = t_P) = 0$. Note that for a model in which processes with $\Delta B \neq 0$ are in equilibrium after $t = t_P$, there is no need to assume charge symmetry at $t \leq t_P$ because at the equilibrium stage baryonic charge disappears independently of initial conditions. Subsequent generation of baryons in our best of all possible worlds is dependent only upon elementary particle interactions and should be determined by microscopic parameters. In the lowest order in $\alpha \simeq 10^{-2}$ baryonic charge is produced by X-meson decays. The total decay rate is

$$\Gamma_X = \alpha m_X^2 N_D (m_X^2 + T^2)^{-1/2} . \tag{15.28}$$

Then there are scattering processes with virtual X-boson exchange which also violate B conservation. Their rate is

$$\Gamma_C = \alpha^2 T^5 N_C (T^2 + m_X^2)^{-2} . \tag{15.29}$$

The number density of X mesons is governed by their decays and by annihilation (which is also B nonconserving) and of course by the inverse processes. The annihilation rate is

$$\Gamma_a = \alpha^2 T N_a r_X , \tag{15.30}$$

where $r_X = n_X/T^3$ is the relative X-meson concentration, and N_D, N_a, and N_C are some numerical factors depending on the model; usually they are of the order of N_{DF} or an order of magnitude smaller. For $T < m_X$, $r_X \sim \exp(-m_X/T)$ and the annihilation rate is small in comparison with the scattering rate Γ_C; for $T > m_X$, $\Gamma_a \simeq \Gamma_C$. Note that for $T > m_X$ all the interactions (including those with $\Delta B \neq 0$) are of the same strength, and for $T < m_X$ the interactions associated with light particle exchange have the annihilation rate described in Eq. (15.30) with $r_X = 1$, i.e., they are $(m_X/T)^4$ times stronger than the processes with X exchange.

Consider first the case in which temperature falls below m_X in a time shorter than the lifetime of X, i.e.,

$$\Gamma_X t(m_X) < 1 . \tag{15.31}$$

In accordance with Eq. (26), $t(T) = 0.3 N_{DF}^{-1/2} T_P T^{-2}$, so, for condition (15.31) to be fulfilled, it is necessary that $m_X > 0.3\alpha N_D N_{DF}^{-1/2} T_P$. For such a large-$m_X$ baryonic charge, variation due to fermion scattering (15.29) and to $X\bar{X}$ annihilation (15.30) is always small. Immediately after $t = t_P$, X bosons decouple from the plasma and their number density in a comoving volume is (almost) constant until $t \simeq \tau_X$. Then for $t > \tau_X$ the decays of X become significant but the inverse X-meson production is negligible because of the Boltzmann factor $\exp(-m_X/T)$ [for the case of $T(\tau_X) < m_X$]. Owing to C and CP violation, the baryonic charge released in decays of X is not equal to that of $\bar{X}$. This baryonic excess is conserved in the course of the expansion and survives to our time in the form of the observed matter. The relative baryon concentration is

$$\Delta r_B = \frac{\Delta N_B}{N_\gamma} = \frac{\Delta\Gamma_X - \Delta\Gamma_{\bar{X}}}{\Gamma_X} \frac{N_X}{N_{tot}} . \tag{15.32}$$

Here N_{tot} is the total number of particle species and N_X is that of X bosons; we assume $N_X/N_{tot} \simeq 0.1$, $\Delta\Gamma_X/\Gamma_X$ is the relative baryonic charge generation in X-

[20]Earlier, Weinberg (1964) noted that there were no massless fields connected with baryonic charge, so the latter is not necessarily conserved. He also noted that baryon asymmetry of the universe might result from B nonconservation, but in the framework of a steady-state universe with spontaneous production of matter from vacuum. Later Weinberg rejected the idea of baryonic charge nonconservation and returned to it only in 1979 (Weinberg, 1979). The difference between the old and new models is that in the old ones the quark was short-lived, but the proton decay proceeded in a third-order, B-nonconserving interaction, single quark decay inside the proton being energetically forbidden. In the new models proton decay is of the same order as the process $qq \to \bar{q}\bar{l}$.

meson decays, and $\Delta\Gamma_{\bar{X}}/\Gamma_X$ is the same for $\bar{X}$. The difference $(\Delta\Gamma - \Delta\bar{\Gamma})$ depends upon the model of CP violation. If CP violation proceeds spontaneously due to a relative phase of the vacuum expectation values of the Higgs fields, symmetry restoration at high temperature (if it takes place at $T < m_X$) leads to the vanishing of Δr_B. No CP violation can be introduced into the interaction of fermions with gauge bosons because CP odd phases in the coupling can be canceled by a redefinition of the wave functions. Some CP violation appears in this case in the fermion mass matrix and in the Higgs sector. This gives, however, a small value of $(\Delta\Gamma - \Delta\bar{\Gamma})$. It is claimed (Weinberg, 1979) that with a natural choice of the parameters one can obtain $(\Delta\Gamma - \Delta\bar{\Gamma})/\Gamma \simeq 10^{-9}$ which gives the necessary value of Δr_B.

A too high value of $m_X = 10^{18}$–10^{17} GeV is an evident shortcoming of the case considered. First, in the existing models a smaller unification mass is needed. Second, one needs to make extra assumption that an initial state is symmetric because in this model the universe is never in equilibrium with respect to B-violating processes, except probably in an unknown stage at $t < t_P$.

Now let condition (15.31) be unsatisfied, i.e., $m_X < 0.3\,\alpha N_D N_{DF}^{-1/2} m_P$. There is a time interval in this case when X-meson decays and the inverse processes are faster than the expansion rate, so X-meson concentration is close to equilibrium. The equilibrium condition is $\Gamma_X r_X t (m_X/T) > 1$ for $m_X > T$, and $\Gamma_X t > 1$ for $m_X < T$, which leads to the following equilibrium temperature interval:

$$\eta^{1/3} > (T/m_X) > (\ln\eta + \tfrac{5}{2}\ln\ln\eta)^{-1} , \tag{15.33}$$

where $\eta = 0.3\alpha N_{DF}^{-1/2} N_D m_X^{-1} m_P$. The temperature interval for which reactions (15.29) are fast, i.e., $\Gamma_C > H \sim t^{-1}$, is

$$\alpha\eta > (T/m_X) > (\alpha\eta)^{-1/3} . \tag{15.34}$$

For simplicity we set $N_D = N_C$. This condition is not self-contradicting if $\alpha\eta > 1$. From the comparison of Eqs. (15.33) and (15.34) it follows that for $\eta < 10^3$ the process of direct and inverse X-meson decays remain in equilibrium longer than the fermion scattering. Hence for

$$m_X > 10^{-3}\ 0.3\alpha N_{DF}^{-1/2} N_D m_P \simeq 10^{-6} m_P,$$

the B-nonconserving scattering can be neglected and the relative baryon concentration is determined by Eq. (15.32) with

$$\frac{N_X}{N_{\text{tot}}} \simeq 0.1\left(\frac{m_X}{T_f}\right)^{3/2} e^{-m_X/T_f}$$
$$\simeq 0.1(\eta\ln^{5/2}\eta)^{-1}[\ln(\eta\ln^{5/2}\eta)]^{3/2}. \tag{15.35}$$

For the boundary value $\eta = 10^3$, $T_f = m_X/15$, and $N_X/N_{\text{tot}} \simeq 10^{-6}$, making necessary the too large ratio $(\Delta\Gamma_X - \Delta\Gamma_{\bar{X}})/\Gamma_x \simeq 10^{-2}$. If $\eta > 10^3$ or $m_X < 10^{-6} m_P$ the burning out of the baryon charge due to reactions between quarks and leptons (15.29) should be taken into account. This leads to the vanishing of r_B with the rate (Dimopoulos and Susskind, 1979)[21]

$$\dot{r}_B = -\Gamma_C r_B . \tag{15.36}$$

As the initial condition, the value of r_B should be used which arises from X-meson decays at the instant of their freezing out [see Eqs. (15.32) and (15.33)]. As a result we obtain unacceptably small r_B. Thus we can conclude that if the baryon excess is generated by X decays, the mass of the latter should be larger than 10^{13} GeV and closer to 10^{16}–10^{18} GeV. The mass of the X meson necessary to give the observed r_B depends on the model of CP violation; the weaker the CP violation, the larger this mass should be. If, however, there are some other B-nonconserving reactions which go on longer than the processes with X mesons, then, the universe effectively "forgets" about the baryonic charge which existed during the X-meson era. In this case the bounds obtained above should be modified to take into account these other processes. This picture could be realized if, in the Higgs meson interactions, baryonic charge were not conserved (Ellis *et al.*, 1979; Weinberg, 1979). The estimates obtained are easily reformulated for this case with the substitution $\alpha \to \alpha_H = (10^{-2}$–$10^{-4})\,\alpha$. The dominant processes are now decays and inverse production of H mesons.

Some other mechanisms of baryonic charge generation have been considered in the literature. Ignatiev *et al.* (1978) discuss a universal superweak interaction which violates CP and B conservation. In this model baryon charge is generated by decays of a scalar particle χ which is an isodoublet with respect to the weak Weinberg–Salam group. In our opinion the result of this paper is overestimated, because the authors do not take into account the burning out of χ in the reaction $\chi\bar{\chi} \to (Z^0) \to l\bar{l}$. A paper by Yoshimura (1978) which stimulated the recent boom on this problem considers baryonic charge generation by the reaction $qq \to \bar{q}\bar{l}$. The resulting baryon asymmetry should be small, however, for the reasons mentioned in Sec. XV.C. Quite recently, after we finished our work on this review, some other papers on baryon asymmetry have appeared. First, we should like to mention the papers by Nanopoulos and Weinberg (1979) and by Barr *et al.* (1979). They conduct a detailed investigation of C (CP) violation at superhigh energy. The authors conclude that Higgs particles, H, are more likely than gauge bosons, X, to generate large baryon asymmetry. In addition to a larger $C(CP)$-violating parameter in H decays, there is no kinetic suppression of baryonic asymmetry because of a weaker coupling of H to fermions (Kolb and Wolfram, 1980; Fry, Olive, and Turner, 1980b; Dolgov, 1980b). In these papers kinetic equations governing baryonic excess are analyzed.

[21]However, Treiman and Wilczek (1980) argue that in the framework of the SU(5) model the quantity $\Delta N_{dR} - \Delta N_{lL} - \Delta N_{\nu L}$ is effectively conserved in the course of this burning out of r_B (here ΔN is the difference between the number of particles and antiparticles). This quantity is not conserved in the Higgs meson decays and so, if baryon asymmetry were caused by H decays, a nonzero value could be generated and survive later.

In papers by Ignatiev *et al.* (1979b), and Kuzmin and Shaposhnikov (1980) a detailed analysis of baryon asymmetry generation is made in the framework of the SO(10) symmetry group. In the second of these papers the interesting conclusion is reached that right-handed weak currents should be strongly suppressed as compared to left-handed currents, otherwise the baryon asymmetry would be too small. The flow of papers on this subject continues, although the supply of new ideas seems to be exhausted. Now is the time for hard calculational work.

In conclusion, we should like to note that the possibility of explaining the existence of our world by microphysical phenomena seems exciting. It is interesting that cosmological data can serve as a criterion for the choice of a grand unified group and of the mechanics of CP invariance violation. The experimental discovery of proton instability would be of extreme importance for cosmology.

Last but not least, most work in this field is done on the assumption of an exact Friedman expansion. This is not proved yet!

F. Baryon asymmetry due to *CPT* violation

Of course if CPT is violated we cannot speak of any self-consistent model of elementary particle interactions. The standard equilibrium statistical mechanics should be valid, however, because, as we see in Sec. XV.C, the equilibrium statistical distributions are provided only by the diagonal elements of the unitarity condition. If the CPT theorem were violated, baryon asymmetry could arise even during the equilibrium stage because of a different energy spectrum of particles and antiparticles. Of course for the generation of a baryonic excess, B nonconservation is, as before, necessary. Note that in the case of conserved B the vanishing of the average baryonic charge would lead to the nonzero chemical potentials of baryons and antibaryons $\mu_B = -\mu_{\bar{B}} = m_B - m_{\bar{B}}$. If B is not conserved then in thermal equilibrium $\mu_B = \mu_{\bar{B}} = 0$ and

$$(N_B - N_{\bar{B}})/N_B \simeq \delta m/T$$

for $m > T$ and

$$(N_B - N_{\bar{B}})/N_B \simeq (\delta m/T)(m/T)$$

for $m < T$. For the mass difference of p and $\bar{p}$ the limit is known to be $\delta m_p/m_p < 10^{-4}$. So in principle a baryon asymmetry could be connected with $(p - \bar{p})$ mass difference,[22] but this should be small.

Cline *et al.* (1977) assumed a noticeable antiproton instability. They claimed that if $\bar{p}$ lifetime lay within the limits 10^{-8} sec $< \tau_{\bar{p}} < 10^{17}$ sec the B-nonconserving decays of $\bar{p}$ could result in the observed baryon excess. Later cosmological arguments were used to diminish this interval. Their results, however, do not coincide with each other. Sato (1978b) obtains the limits 3.4×10^{-4} sec $< \tau_{\bar{p}} < 7.4 \times 10^{-4}$ sec, and Demaret and Vandermeulen (1978) obtain 10^{-8} sec $< \tau_{\bar{p}} < 10^{-7}$ sec. In our opinion, however, one cannot obtain the right value of Δr_B with any value of antiproton lifetime. The point is that the proton is a strongly interacting particle: at $T = m_N$ the annihilation rate is large, $\Gamma_a \simeq 10^{22}$ sec^{-1}, whereas the expansion rate of the universe is 10^6 sec^{-1}. Because of the arguments presented in Sec. XV.C, deviations from equilibrium are small for p and $\bar{p}$ when $\Gamma \geq m_N$ and so the average baryonic charge is small. On the other hand, when later Γ_a and H are of the same order and deviations from equilibrium become noticeable, the baryon (antibaryon) concentration due to the factor $\exp(-m/T)$ becomes small, and decay of all the antiprotons cannot give a large excess of protons. Using Eq. (15.19) the following estimate for Δr_B can be obtained:

$$\Delta r_B = (\sigma_a v T T_P)^{-1} , \tag{15.37}$$

where σ_a is the annihilation cross section. This formula is valid when the temperature is greater than the proton (antiproton) quenching temperature $T_f = m_N/40$. Expression (15.37) predicts the proton concentration to be about 10^{-19} of the photon concentration.

In concluding this section, we want to stress once more that we are probably very close to the solution of the grand problem of how our beautiful and stable (with respect to annihilation) world was created from an explosive initial state. Now we need to understand the origin of the Friedman expansion after the singularity and all cosmological problems will be solved.

XVI. VACUUM POLARIZATION BY CURVED SPACE-TIME

The above mentioned topic may seem to be somewhat outside the scope of our paper, but actually the distinction between particles and fields, and vacuum polarization is now artificial. In the ninteenth century charged particles and electromagnetic fields were considered separately but quantum mechanics has bridged the gap between them. Still we shall cover the topic very briefly, reviewing mostly the Soviet papers less known to a Western reader.

A field theory quantized straightforwardly leads to the prediction of zero-point oscillations of the boson field with nonzero expectation values for all field strengths squared (E^2, H^2, for example) and divergent energy density and pressure

$$\epsilon(m) = \tfrac{1}{2}\int d^3p(p^2+m^2)^{1/2}/(2\pi)^3$$

$$P(m) = \tfrac{1}{6}\int d^3p\, p^2(p^2+m^2)^{-1/2}/(2\pi)^3 .$$

Free fermionic fields yield negative divergent integrals of the same type. It is customary to renormalize the contribution of every single field, for example by taking normal ordered products of creation and annihilation operators instead of $(aa^\dagger + a^\dagger a)/2$ in the Hamiltonian.

What is actually known from observations? The sum total of all contributions to the vacuum energy is zero or at least very small:

$$\epsilon_v < 10^{-8}\ \mathrm{erg\,cm^{-3}}.$$

[22]The equality of masses of K and $\bar{K}$ is an argument against CPT violation and p-$\bar{p}$ mass difference.

Written as $\epsilon_v = m_v^4$ this means $m_v < 10^{-3}$ eV. No arguments based on fundamental principles have as yet explained this vanishingly small value.

If $\epsilon_v \neq 0$ then relativistic invariance ensures that $P = -\epsilon_v$ or $T_{ik} = \Lambda g_{ik}$ in an arbitrary frame (but still in Minkowski space) where Λ is the cosmological constant. A direct proof was given by Zeldovich (1967) that the sums $\epsilon = \sum C_i \epsilon(m_i)$ and $P = \sum C_i p(m_i)$ have the property $p = -\epsilon$.

As is mentioned above, there are no indications that $\Lambda \neq 0$ if neutrinos are massless, but if they are massive, perhaps this will necessitate the nonvanishing of Λ, ϵ_v, and p_v.

The next question is the nature of the change of ϵ_v and p_v (or, taken together, T_{ikv}, subscript v is for vacuum) in a curved space. The difference $T_{ik} = T_{ikv}$ (curved) $- T_{ikv}$ (flat) is of the same type as the Casimir effect in quantum electrodynamics. Even those who assume that T_{ikv} (flat) vanishes for every field agree that this difference is real. This quantity is of the same sign for fermions and bosons; therefore it is not canceled even if T_{ikv} (flat) is zero.

Thus T_{ik} appears on the right-hand side of the Einstein equations. As far as it is local it depends on R_{jklm} and its derivatives:

$$T_{ik} = F(R_{jklm}) \ ,$$

where F is a functional. In this approximation we can put it into the left-hand side and speak of modified Einstein equations. The modification can be achieved formally by adding more complicated scalars to the action of general relativity:

$$\int R\, dv \to \int (R + C_1 R^2 + C_2 R^2 \ln R + \cdots) dv \ .$$

This ensures all general properties of the Einstein equations.

We begin with zero-point oscillations partly for pedagogical reasons, but also because we believe in the "minimal interaction theory," which means that one should investigate only the modifications which are inescapable consequences of the existence of various fields (gravitons in the weak field approximation included), but not the most general modifications of the Lagrangian.

The modification was considered by Ginzburg *et al.* (1971). They point out that by general covariance T_{ikv} can be written as

$$T_{ikv} = aG_{ik} + T'_{ik} \ .$$

The first term can be incorporated in the left-hand side of the Einstein equation, it is equivalent to a change (renormalization) of the gravitational constant

$$G_{ik} = G'(T'_{ik} + T_{ikm}) \ ,$$

where $G' = 8\pi\, G/(1 - 8\pi Ga)$ is the renormalized constant and T_{ikm} is the energy-momentum tensor of matter. So the term aG_{ik} is unobservable. In contrast to this T'_{ik} is a genuine vacuum polarization term. It is observable, at least in principle.

The most important point is that T'_{ik} is not subject to the so-called "energy dominance principle," which means that $T_{00} > T_{ik}$, i, $k = 1, 2, 3$. This condition is fulfilled for free particles and fields and it is very probable that it is also true for interacting fields.

The vacuum polarization term T'_{ik} is built from the contributions of various fields. One might think naively that T'_{ik} is also subject to the laws governing these fields. That is true for the energy-momentum conservation law $T'_{ik;k} = 0$. But it does not hold for the energy dominance principle simply because it is an inequality!

The energy dominance condition is of utmost importance, as pointed out by Hawking and his colleagues. The theorems about the inescapable occurrence of singularities are based on energy dominance. Hawking has shown that if energy dominance is always fulfilled, no particle creation from a vacuum is possible by space-time distortion, i.e., by gravitational fields. *Apropos* of this problem, Zeldovich and Pitayevsky (1971) explicitly pointed to a violation of the energy dominance condition by vacuum polarization. This violation was shown to permit construction of a singularity-free cosmological model with a DeSitter-type initial metric,

$$ds^2 = dt^2 - e^{2Ht}(dx^2 + dy^2 + dz^2),$$

consistent with vacuum polarization energy density (Gurovich and Starobinsky, 1979; Starobinsky, 1980; Zeldovich, 1981).

We should like here merely to call attention to the unusual properties and very important consequences of vacuum polarization. It is a task for future investigations to obtain all the properties of the universe from the laws of fundamental quantum field physics.

ACKNOWLEDGMENTS

We thank S. S. Gerstein, M. Yu. Khlopov, A. D. Linde, L. B. Okun, and R. A. Sunyaev for valuable discussions.

REFERENCES

Adams, P. J., V. Canuto, and H.-Y. Chiu, 1976, Phys. Lett. B **61**, 397.
Anselm, A. A., 1979, Zh. Eksp. Teor. Fiz. Pis'ma Red. **29**, 645.
Barger, V., K. Whisnant, D. Cline, and R. J. N. Phillips, 1980, Preprints COO-881-135 and COO-881-146.
Barnes, A., 1979, Astrophys. J. **227**, 1.
Barr, S., G. Segrè, and H. A. Weldon, 1979, Phys. Rev. D **20**, 2494.
Beaudet, G., and A. Yahil, 1977, Astrophys. J. **218**, 253.
Beg, M. A. B., W. J. Marciano, and M. Ruderman, 1978, Phys. Rev. D **17**, 1935.
Beletsky, Yu. A., A. I. Bugrij, E. S. Martynov, and A. A. Trushevsky, 1979, Nature (London) **282**, 194.
Bernstein, J., M. Ruderman, and G. Feinberg, 1963, Phys. Rev. **132**, 1227.
Bludman, S. A., and M. A. Ruderman, 1976, Phys. Rev. Lett. **36**, 840.
Bogdanova, L. N., and I. S. Shapiro, 1974, Zh. Eksp. Teor. Fiz. Pis'ma Red. **20**, 217.
Braginsky, V. B., and V. I. Panov, 1971, Zh. Eksp. Teor. Fiz. **61**, 873.
Chibisov, G. V., 1976, Usp. Fiz. Nauk **119**, 551.
Chiu, H.-Y., 1966, Phys. Rev. Lett. **17**, 712.
Chiu, C. W., H.-Y. Chiu, and R. B. Stothers, 1966, Ann. Phys. (N.Y.) **39**, 280.

Cline, D., P. McIntyre, and C. Rubbia, 1977, Phys. Lett. B **66**, 429.
Coleman, S., 1975, in *Laws of Hadronic Matter: Proceedings of the 11th Course of the "Ettore Majorana" International School of Subnuclear Physics*, edited by A. Zichichi (Academic, New York), p. 139.
Coleman, S., 1978, Phys. Rev. Lett. **40**, 665 and references therein.
Cowsik, R., 1977, Phys. Rev. Lett. **39**, 784.
Cowsik, R., and J. McClelland, 1972, Phys. Rev. Lett. **29**, 669.
Daniel, M., G. Lazaridis, and Q. Shaffi, 1979, TH-2800 CERN.
Davis, L., Jr., A. S. Goldhaber, and M. M. Nieto, 1975, Phys. Rev. Lett. **35**, 1402.
Davis, R., Jr., and J. C. Evans, Jr., 1976, paper delivered at Leningrad Conference on Active Processes in the Sun and the Solar Neutrino Problem, BNL Report No. 21837.
Demaret, J., and J. Vandermeulen, 1978, Phys. Lett. B **73**, 471.
Demarque, P., and R. D. McClure, 1977, in *The Evolution of Galaxies and Stellar Population*, edited by B. M. Tinsley and R. B. Larson (Yale University Observatory, New Haven), p. 199.
Dicke, R. H., 1962, Phys. Rev. **126**, 1580.
Dicke, R. H., P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson, 1965, Astrophys. **142**, 414.
Dicus, D. A., E. W. Kolb, and V. L. Teplitz, 1977, Phys. Rev. Lett. **39**, 168.
Dicus, D. A., E. W. Kolb, and V. L. Teplitz, 1978, Astrophys. J. **221**, 327.
Dicus, D. A., E. W. Kolb, V. L. Teplitz, and R. V. Wagoner, 1977, Phys. Rev. D **17**, 1529.
Dicus, D. A., E. W. Kolb, V. L. Teplitz, and R. V. Wagoner, 1978, Phys. Rev. D **18**, 1829.
Dimopoulos, S., and L. Susskind, 1979a, Phys. Lett. **81B**, 416.
Dimopoulos, S., and L. Susskind, 1979b, Phys. Rev. D **18**, 4500.
Dirac, P. A. M., 1934, Proc. R. Soc. London Ser. A **133**, 60.
Dirac, P. A. M., 1948, Phys. Rev. **74**, 817.
Dolgov, A. D., 1979, Zh. Eksp. Teor. Fiz. Pis'ma Red. **29**, 254.
Dolgov, A. D., 1980a, Yad. Fiz. **31**, 1522.
Dolgov, A. D., 1980b, Yad. Fiz. **32**, in press.
Dolgov, A. D., 1980c, preprint ITEP-9 and Zh. Eksp. Teor. Fiz. **79**, 337.
Dolgov, A. D., and V. I. Zakharov, 1971, Zh. Eksp. Teor. Fiz. **60**, 468.
Domogatsky, G. V., and I. M. Zheleznykh, 1969, Yad. Fiz. **10**, 1238.
Doroshkevich, A. G., and I. D. Novikov, 1964, Dokl. Akad. Nauk SSSR **154**, 809.
Doroshkevich, A. G., Ya. B. Zeldovich, R. A. Sunyaev, and M. Yu. Khlopov, 1980, Astron. Zh. Pis'ma **6**, 457 and 465.
Dover, C. B., T. K. Gaisser, and G. Steigman, 1979, Phys. Rev. Lett. **42**, 1117.
Einhorn, M. B., D. L. Stein, and D. Toussaint, 1980, Phys. Rev. D **21**, 3295.
Ellis, J., M. K. Gaillard, and D. V. Nanopoulos, 1979, Phys. Lett. B **80**, 360; **82**, 464 (E).
Falk, S. W., and D. N. Schramm, 1978, Phys. Lett. B **79**, 511.
Frampton, P. H., 1977, Phys. Rev. D **15**, 2922.
Fritzsch, H., and P. Minkowski, 1975, Ann. Phys. (N.Y.) **93**, 193.
Fry, J. N., K. A. Olive, and M. S. Turner, 1980a, Phys. Rev. D **22**, 2953.
Fry, J. N., K. A. Olive, and M. S. Turner, 1980b, Phys. Rev. D **22**, 2977.
Fry, J. N., K. A. Olive, and M. S. Turner, 1980c, Phys. Rev. Lett. **45**, 2074.
Georgi, H., and S. L. Glashow, 1974, Phys. Rev. Lett. **32**, 438.
Gerstein, S. S., and Ya. B. Zeldovich, 1966, Zh. Eksp. Teor. Fiz. Pis'ma Red. **4**, 174.
Ginzburg, V. L., D. A. Kirzhnits, and A. A. Lyubushin, 1971, Zh. Eksp. Teor. Fiz. **60**, 451.
Goddard, P., and D. I. Olive, 1978, Rep. Prog. Phys. **41**, 1357.
Goldhaber, A. S., and M. N. Nieto, 1971, Rev. Mod. Phys. **43**, 277.
Goldman, T., and G. J. Stephenson, Jr., 1977, Phys. Rev. D **16**, 2256.
Gunn, J. E., B. W. Lee, I. Lerche, D. N. Schramm, and G. Steigman, 1978, Astrophys. J. **223**, 1015.
Gurovich, V. T., and A. A. Starobinsky, 1979, Zh. Eksp. Teor. Fiz. **77**, 1683.
Guth, A. H., and S.-H. Tye, 1980, Phys. Rev. Lett. **44**, 631.
Hayashi, C., 1950, Prog. Theor. Phys. **5**, 224.
Hawking, S. W., 1974, Nature **248**, 30.
Hawking, S. W., 1975, Commun. Math. Phys. **43**, 199.
Hainebach, K. L., and D. N. Schramm, 1976, Astrophys. J. Lett. **207**, L79.
Hut, P., 1977, Phys. Lett. B **69**, 85.
Ignatiev, A. Yu., N. V. Krasnikov, V. A. Kuzmin, and A. N. Tavkhelidze, 1978a, in *Neutrino-77,* conference proceedings (Nauka, Moscow), Vol. II, p. 293.
Ignatiev, A. Yu., N. V. Krasnikov, V. A. Kuzmin, and A. N. Tavkhelidze, 1978b, Phys. Lett. B **76**, 436.
Ignatiev, A. Yu., V. A. Kuzmin, and M. E. Shaposhnikov, 1979, preprint of the Institute for Nuclear Research H-0123.
Jones, L. W., 1978, Phys. Rev. D **17**, 1462.
Kazanas, D., 1978, in *Proceedings of the Seminar on Proton Stability*, edited by D. Cline (University of Wisconsin, Madison).
Khlopov, M. Yu., 1980, Yad. Fiz. **32**, (in press).
Kibble, T. W. B., 1976, J. Phys. A **9**, 1387.
Kirzhnits, D. A., 1972, Zh. Eksp. Teor. Fiz. Pis'ma Red. **15**, 745.
Kirzhnits, D. A., and A. D. Linde, 1972, Phys. Lett. B **42**, 471.
Kobzarev, I. Yu., and P. I. Peshkov, 1974, Zh. Eksp. Teor. Fiz. **67**, 428.
Kolb, E. W., and S. Wolfram, 1980a, Phys. Lett. B **91**, 217.
Kolb, E. W., and S. Wolfram, 1980b, Nucl. Phys. B **172**, 224.
Kosik, V. S., V. A. Lyubimov, E. G. Novikov, V. S. Nosik, and E. F. Tretyakov, 1980, Yad. Fiz. **32**, 301.
Kuzmin, V. A., 1970, Zh. Eksp. Teor. Fiz. Pis'ma Red. **5**, 32.
Kuzmin, V. A., and M. V. Shaposhnikov, 1978, Zh. Eksp. Teor. Fiz. Pis'ma Red. **27**, 665.
Kuzmin, V. A., and M. E. Shaposhnikov, 1980, preprint of the Institute of Nuclear Research, P-0138.
La Rue, G. S., W. M. Fairbank, and A. F. Hebard, 1977, Phys. Rev. Lett. **38**, 1011.
La Rue, G. S., W. M. Fairbank, and J. D. Phillips, 1979, Phys. Rev. Lett. **42**, 142 and 1019.
Lasher, G., 1979, Phys. Rev. Lett. **42**, 1646.
Lee, B. W., and S. Weinberg, 1977, Phys. Rev. Lett. **39**, 165.
Lee, T. D., 1973, Phys. Rev. D **8**, 1226.
Lee, T. D., and C. S. Wu, 1966, Annu. Rev. Nucl. Sci. **16**, 471.
Lee, T. D., and C. N. Yang, 1955, Phys. Rev. **98**, 1501.
Lifshitz, E. M., 1946, Zh. Eksp. Teor. Fiz. 16, 587.
Linde, A. D., 1976a, Zh. Eksp. Teor. Fiz. Pis'ma Red. **23**, 73.
Linde, A. D., 1976b, Trieste preprint TC-76-26/76.
Linde, A. D., 1979a, Rep. Prog. Phys. **42**, 389.
Linde, A. D., 1979b, Phys. Lett. B **83**, 311.
Linde, A. D., 1980a, Phys. Lett. B **92**, 119.
Linde, A. D., 1980b, Zh. Eksp. Teor. Fiz. Pis'ma Red. (in press).
Linde A. D., and I. V. Krive, 1976, Nucl. Phys. B **117**, 265.

MacCallum, M., 1979, Nature **280**, 449.
Marx, G., 1977, ITP-376 preprint, Budapest.
Marx, G., and A. S. Szalay, 1972, in *Neutrino-72* conference proceedings (Edit. Technoinform, Budapest), Vol. I, p. 191.
Miyama, S., and K. Sato, 1978, Prog. Theor. Phys. **60**, 1703.
Nakamura, T., S. Miyama, and K. Sato, 1977, Prog. Theor. Phys. **58**, 1052.
Nanopoulos, D. V., and S. Weinberg, 1979, Phys. Rev. D **20**, 2484.
Nilson, N., and P. Olesson, 1973, Nucl. Phys. B **61**, 45.
Novikov, I. D., A. G. Polnarev, A. A. Starobinsky, and Ya. B. Zeldovich, 1979, Astron. Astrophys. **80**, 104.
Ogorodnikov, D. D., I. M. Samoylov, and A. M. Solntsev, 1979, Zh. Eksp. Teor. Fiz. **76**, 1881.
Okun, L. B., 1969, Yad. Fiz. **10**, 358.
Okun, L. B., 1980a, preprint, ITEP-41.
Okun, L. B., 1980b, preprint ITEP-6.
Okun, L. B., and Ya. B. Zeldovich, 1976, Comments Nucl. Part. Phys. **6**, 69.
Okun, L. B., and Ya. B. Zeldovich, 1978, Phys. Lett. B **78**, 597.
Olive, K. A., D. N. Schramm, and G. Steigman, 1978, in *Proceedings of the Seminar on Proton Stability*, edited by D. Cline (University of Wisconsin, Madison).
Omnes, R., 1969, Phys. Rev. Lett. **23**, 38.
Omnes, R., 1972, Phys. Rep. **36**, 1.
Osborne, W. Z., 1970, Phys. Rev. Lett. **24**, 1441.
Parker, E. N., 1970, Astrophys. J. **160**, 383.
Particle Data Group, 1978, Phys. Lett. B **75**, i.
Peccei, R. D., and H. R. Quinn, 1977, Phys. Rev. Lett. **38**, 1440.
Peebles, P. J. E., 1966, Phys. Rev. Lett. **16**, 410.
Peimbert, M., 1975, Annu. Rev. Astron. Astrophys. **13**, 113.
Penzias, A. A., and R. W. Wilson, 1965, Astrophys. **142**, 419.
Polyakov, A. M., 1974, Zh. Eksp. Teor. Fiz. Pis'ma Red. **20**, 430.
Pomansky, A. A., 1976, in *Proceedings of International Neutrino Conference, Aachen*, edited by H. Faissner, H. Reithler, and P. Zerwas.
Preskill, J. R., 1979, Phys. Rev. Lett. **43**, 1365.
Price, P. B., E. K. Shirk, W. Z. Osborne, and L. S. Pinsky, 1975, Phys. Rev. Lett. **35**, 487.
Price, P. B., E. K. Shirk, W. Z. Osborne, and L. S. Pinsky, 1978, Phys. Rev. D **18**, 1382.
Reeves, H., 1974, Annu. Rev. Astrons. Astrophys. **12**, 437.
Reines, E., H. W. Sobel, and E. Pasierb, 1980, Phys. Rev. Lett. **45**, 1307.
Sakharov, A. D., 1967, Zh. Eksp. Teor. Fiz. Pis'ma Red. **5**, 32.
Sakharov, A. D., 1979, Zh. Eksp. Teor. Fiz. **76**, 1172.
Sato, K., 1978a, Prog. Theor. Phys. **60**, 1942.
Sato, K., 1978b, Kyoto University preprint, KUNS 396.
Sato, K., and H. Kobayashi, 1977, Prog. Theor. Phys. **58**, 1775.
Sato, K., and H. Sato, 1975a, Prog. Theor. Phys. **54**, 912.
Sato, K., and H. Sato, 1975b, Prog. Theor. Phys. **54**, 1564.
Schramm, D. V., 1978, in *Neutrino-78*, conference proceedings (Purdue University, Lafayette), p. 87.
Schramm, D. V., and R. V. Wagoner, 1977, Ann. Rev. Nucl. Sci. **27**, 37.
Schwinger, J., 1966, Phys. Rev. **144**, 1087.
Shvartsman, V. F., 1969, Zh. Eksp. Teor. Fiz. Pis'ma Red. **9**, 315.
Shapiro, S. L., S. A. Teukolsky, and I. Wasserman, 1980, Phys. Rev. Lett. **45**, 669.
Slansky, R., in *Proceedings of the Seminar on Proton Stability*, edited by D. Cline (University of Wisconsin, Madison).
Starobinsky, A. A., 1980, Phys. Lett. B **91**, 99.
Stecker, F. W., 1978, Astrophys. J. **223**, 1032.
Stecker, F. W., 1980, Phys. Rev. Lett. **44**, 1237; 1437(E).
Steigman, G., D. N. Schramm, and J. E. Gunn, 1977, Phys. Lett. B **66**, 202.
Stothers, R. B., 1970, Phys. Rev. Lett. **24**, 538.
Sunyaev, R. A., 1980, Astron. Zh. Pis'ma (unpublished).
Sunyaev, R. A., and Ya. B. Zeldovich, 1970, Astrophys. Space Sci. **7**, 20.
Sutherland, P., J. N. Ng, E. Flowers, M. Ruderman, and C. Inman, 1976, Phys. Rev. D **13**, 2700.
Szalay, A. S., and G. Marx, 1976, Astron. Astrophys. **49**, 437.
't Hooft, G., Nucl. Phys. B **79**, 276.
Tammann, G. A., A. Sandage, and A. Yahil, 1980, preprint of Astronomy Institute, University of Basel, No. 1.
Tayler, R. J., 1979, Nature 282, 559.
Taylor, J. H., L. A. Fowler, and P. M. McCulloch, 1979, Nature **277**, 437.
Thum, C., P. Mezger, and V. Pankonin, 1979, Max Planck Institute preprint No. 64.
Toussaint, D., S. B. Treiman, F. Wilczek, and A. Zee, 1979, Phys. Rev. D **19**, 1036.
Treiman, S. B., and F. Wilczek, 1980, Phys. Lett. **95B**, 222.
Tremaine, S., and J. E. Gunn, 1979, Phys. Rev. Lett. **42**, 407.
Turner, M. S., 1979, Phys. Lett. B **89**, 155.
Voloshin, M. B., I. Yu. Kobsarev, and L. B. Okun, 1974, Yad. Fiz. **20**, 1229.
Voloshin, M. B., and L. B. Okun, 1978, Zh. Eksp. Teor. Fiz. Pis'ma Red. **28**, 156.
Vysotsky, M. I., A. D. Dolgov, and Ya. B. Zeldovich, 1977a, Zh. Eksp. Teor. Fiz. Pis'ma Red. **26**, 200 and, 1977b, in *Neutrino-77*, conference proceedings (Moscow, Nauka), p. 42.
Vysotsky, M. I., Ya. B. Zeldovich, M. Yu. Khlopov, and V. M. Chechetkin, 1978, Zh. Eksp. Teor. Fiz. Pis'ma **27**, 533.
Wagoner, R. V., W. A. Fowler and F. Hoyle, 1967, Astrophys. J. **148**, 3.
Wagoner, R. V., 1974, in *Confrontation of Cosmological Theories with Observational Data*, edited by M. Longair (Reidel, Boston).
Wagoner, R. V., and G. Steigman, 1979, Phys. Rev. D **20**, 825.
Warshawsky, V. I., and A. V. Tutukov, 1972, Nauchn. Informatsia Astromsoveta AN SSSR **23**, 47; **26**, 35(E).
Weinberg, S., 1964, in *Lectures on Particle and Field Theory*, edited by S. Deser and K. Ford (Prentice Hall, Englewood Cliffs), Vol. II, p. 482.
Weinberg, S., 1971, Astrophys. J. **168**, 175.
Weinberg, S., 1972, *Gravitation and Cosmology* (Wiley, New York).
Weinberg, S., 1976, Phys. Rev. Lett. **36**, 294.
Weinberg, S., 1977, *The First Three Minutes* (Basic, New York).
Weinberg, S., 1978, Phys. Rev. Lett. **40**, 223.
Weinberg, S., 1979, Phys. Rev. Lett. **42**, 850.
Weinberg, S., 1980, Rev. Mod. Phys. **52**, 515.
Wilczek, F., 1978, Phys. Rev. Lett. **40**, 279.
Williams, E. R., J. E. Faller, and H. Hill, 1971, Phys. Rev. Lett. **26**, 721.
Winter, K., 1979, in *Proceedings of the International Symposium on Lepton and Photon Interactions*, edited by T. B. W. Kirk and H. D. Abarbanel (FNAL, Batavia, Illinois), p. 258.
Womack, E. A., and J. W. Overbeck, 1970, J. Geophys. Res. **75**, 1811.
Woody, D. P., and P. L. Richards, 1979, Phys. Rev. Lett. **42**, 925.
Yahil, A., and G. Beaudet, 1976, Astrophys. J. **206**, 26.
Yamaguchi, Y., 1959, Prog. Theor. Phys. Suppl. No. 11, 1.
Yoshimura, N., 1978, Phys. Rev. Lett. **41**, 281; **42**, 746(E).
Zeldovich, Ya. B., 1938, Zh. Fiz. Khim. **11**, 697.
Zeldovich, Ya. B., 1961, Zh. Eksp. Teor. Fiz. **41**, 1609.
Zeldovich, Ya. B., 1962, Zh. Eksp. Teor. Fiz. **43**, 1561.
Zeldovich, Ya. B., 1963, Usp. Fiz. Nauk **80**, 357.
Zeldovich, Ya. B., 1965a, Zh. Eksp. Teor. Fiz. **48**, 986.
Zeldovich, Ya.B., 1965b, Advances Astron. Astrophys. **3**,

242.
Zeldovich, Ya. B., 1967, Zh. Eksp. Teor. Fiz. Pis'ma Red. **6**, 883.
Zeldovich, Ya. B., 1968, Usp. Fiz. Nauk **95**, 209.
Zeldovich, Ya. B., 1972, Mon. Not. R. Astron. Soc. **160**, p. 1.
Zeldovich, Ya. B., 1976, Zh. Eksp. Teor. Fiz. Pis'ma Red. **24**, 29.
Zeldovich, Ya. B., 1980, Mon. Not. R. Astron. Soc. **192**, 246.
Zeldovich, Ya. B., 1981, Usp. Fiz. Nauk **133**, No. 2 (in press).
Zeldovich, Ya. B., and M. Yu. Khlopov, 1978, Phys. Lett. B **79**, 239.
Zeldovich, Ya. B., A. A. Klypin, M. Yu. Khlopov, and V. M. Chechetkin, 1980, Yad. Fiz. **31**, 1286.
Zeldovich, Ya. B., I. Yu. Kobzarev, and L. B. Okun, 1974, Zh. Eksp. Teor. Fiz. **67**, 3.
Zeldovich, Ya. B., and I. D. Novikov, 1967, Zh. Eksp. Teor. Fiz. Pis'ma Red. **6**, 772.
Zeldovich, Ya. B., and I. D. Novikov, 1975, *Structure and Evolution of the Universe* (Nauka, Moscow).
Zeldovich, Ya. B., L. B. Okun, and S. B. Pikelner, 1965, Usp. Fiz. Nauk **84**, 113.
Zeldovich, Ya. B., and L. P. Pitajevsky, 1971, Commun. Math. Phys. **23**, 185.
Zeldovich, Ya. B., and Ya. A. Smorodinsky, 1961, Zh. Eksp. Teor. Fiz. **41**, 907.
Zeldovich, Ya. B., and R. A. Sunyaev, 1980, Astron. Zh. Pis'ma **6**, 451.
Zeldovich, Ya. B., R. A. Sunyaev, A. G. Doroshkevich, and M. Yu. Khlopov, 1980,
Zorn, J. C., G. E. Chamberlin, and V. W. Hughes, 1963, Phys. Rev. **129**, 2566.