COSMOLOGY and PARTICLES

XVIth Rencontre de Moriond
Moriond Astrophysics meeting
Les Arcs-Savoie-France, March 15-21, 1981

COSMOLOGY AND PARTICLES
(Cosmologie et Particules)

ISBN 2-86332-009-2

Printed and bound in Singapore by Singapore National Printers (Pte) Ltd.

Proceedings of the
SIXTEENTH RENCONTRE DE MORIOND
ASTROPHYSICS MEETING, 1, 1981, Les Arcs.
Les Arcs-Savoie-France, March 15-21, 1981

COSMOLOGY and PARTICLES

edited by
J. Audouze - P. Crane
T. Gaisser - D. Hegyi
and
J. Tran Thanh Van

SECRETARIAT :
MORIOND ASTROPHYSICS
Bâtiment 211 - Université Paris-Sud
91405 ORSAY CEDEX (FRANCE)

This book is dedicated to the memory of our colleague and friend Beatrice M. Tinsley, professor of astrophysics at Yale University, who contributed so much to the progress of most of the topics covered here.

The recollection of her enthusiasm and her exceptional human qualities will continue to inspire those who had the chance to know and esteem her.

The Editors

L'ensemble de ce livre est dédié à la mémoire de notre collègue et amie Béatrice M. Tinsley, professeur d'astrophysique à l'Université de Yale, qui a contribué à l'essor et aux progrès de la plupart des sujets traités ici.

Le souvenir de son enthousiasme et de ses qualités humaines exceptionnelles continuera à inspirer ceux qui ont eu la chance de la connaître et de l'estimer.

Les Organisateurs

PREFACE

The astrophysicists are first realizing with most interest that elementary particle physics influence directly the primordial evolution of the Universe. The Grand Unification theories determine the ways by which the Universe has been born and also in some sense its remote future by predicting a finite lifetime for the proton.

In the framework of the Rencontres de Moriond who gather since many years the elementary particle physicists and the biologists in a ski resort, about two dozens of astrophysicists have met to discuss very informally but also very deeply these problems which excite both astrophysicists and elementary particle physisists communities. These two communities had the chance to interact effectively during these rencontres since a whole day was devoted to joint discussions. There are obvious proofs of this interaction in these proceedings with the contributions of R. Cowsik, J. Ellis and D.N. Schramm for instance. The astrophysicists present at Les Arcs have also considered other aspects of high enrgy astrophysics : then in a very friendly and cheerful atmosphere entertained by the very effective organization of J. Tran Thanh Van who promoted these rencontres, we have discussed topics as fundamental and various as the observation and the analysis of the very high energy cosmic rays (Section I), the implications of the Grand Unification and the phase transition theories on the primordial evolution of the Universe (Section II), the astrophysical consequences of the finite mass (if any) of the neutrinos (Section III), the primordial nucleosynthesis (Section IV), the anisotropies of the 2.7 K blackbody radiation (Section V) and the search of "unvisible" mass in the Universe (Section VI).

Thanks to the efforts of the authors of these excellent contributions whom I thank very cheerfully, we have been able to publish very quickly this book providing an account of the recent progresses performed in very "burning" topics. I hope that this book will be useful to the two scientific communities who met at Les Arcs.

This informal gathering would not have been as fruitful and pleasant without the organization and the hospitality of J. Tran Thanh Van and his wife, without the charming collaboration of Nicole Mathieu and the effective and friendly collaboration with my three co-organizers Philippe Crane, Tom Geisser and Dennis Hegyi. I want to emphasize the fundamental role of Dennis in this enterprise.

During the meeting which is the first of a series that we wish to be as fruitful and successful than those of the biologists and the elementary particle physicists our community underwent the immense loss of Beatrice M. Tinsley who was one of the most brillant, cheerful and productive members. I had personnaly the chance to know her very well, to be one of her friends and also to work very closely with her on different problems related to the evolution of galaxies especially in the writing of a review paper on this topic. This collaboration was indeed for me one of the most pleasant and inspiring among those I undertook. At the end of this preface I want to say all the good things that we all think about her not only professionally but also and especially in our friendly relationship. I feel very sadly the decease of Beatrice who worked so hard in astrophysics and cosmology and that I consider as one of the most courageous and noble colleague.

By her works and her memory, she will continue to inspire this field who should allow us to pursue our meetings in the frame of these rencontres.

Jean Audouze

PREFACE

Les astrophysiciens viennent de se rendre compte avec interêt que la physique des particles élémentaires exerce une influence directe sur l'évolution de l'Univers et en particulier sur sa naissance. Les théories tentant d'unifier les interactions fondamentales (l'interaction forte, l'interaction faible et l'interaction nucléaire) déterminent la façon dont l'Univers a pu prendre naissance et fixent également d'une certaine manière son lointain futur en prédisant une durée de vie finie pour le proton.

Dans le cadre des rencontres de Moriond qui rassemblent depuis de nombreuses années, dans une station de sports d'hiver, les physiciens des particules élémentaires et les biologistes, une vingtaine d'astrophysiciens se sont réunis pour discuter de façon très approfondie ces problèmes qui intéressent à la fois la communauté des astrophysiciens et celle des physiciens des particules. Les deux communautés ont eu la chance d'interagir de façon effective aux Arcs puisqu'une journée entière a été consacrée à des discussions communes. On en trouve des preuves manifestes dans ces comptes-rendus avec les contributions de R. Cowsik, J. Ellis et D.N. Schramm. Les astrophysiciens rassemblés aux Arcs se sont également intéressés à d'autres aspects de l'astrophysique des hautes énergies : c'est ainsi que dans une atmosphère extrêmement agréable et amicale entretenue par l'organisation attentive de J. Tran Thanh Van, fondateur de ces rencontres, nous avons discuté de sujets aussi fondamentaux et variés que l'observation et l'analyse du rayonnement cosmique de très haute energie (Section I), l'implication des théories d'unification des interactions fondamentales et de celles faisant appel à des transitions de phases sur l'évolution primordiale de l'Univers (Section II), les conséquences astrophysiques de la découverte d'une masse finie pour les neutrinos (Section III), la nucléosynthèse primordiale (Section IV), l'anisotropie du rayonnement à 2.7 K (Section V) et la recherche de la masse "invisible" dans l'Univers (Section VI).

Grâce aux efforts des auteurs de ces excellentes contributions, que je tiens à remercier, nous sommes en mesure de publier très rapidement ce livre rendant compte des progrès effectués dans des sujets très "brûlants" et qui, je l'espère, sera utile aux deux communautés qui se sont rencontrées aux Arcs.

Cette rencontre informelle n'aurait pas été aussi fructueuse et agréable sans l'organisation et l'esprit d'hospitalité de J. Tran Thanh Van et de son

épouse, sans la souriante collaboration de Nicole Mathieu et sans la collaboration efficace et amicale de mes trois co-organisateurs, Philippe Crane, Tom Gaisser et Dennis Hegyi en soulignant le rôle fondamental de ce dernier.

Au moment où s'est tenue cette réunion qui inaugure une série que nous souhaitons aussi fructueuse que celles qui concernent les biologistes et les physiciens des particules, notre communauté a subi la perte immense de l'un de ses membres les plus brillants , chaleureux et productifs en la personne de Béatrice M. Tinsley. J'ai eu personnellement la chance non seulement de la connaître, mais surtout de travailler de façon étroite avec elle sur plusieurs problèmes d'évolution des galaxies et de collaborer à la rédaction d'un article de revue sur ce sujet. Cette collaboration a été certainement l'une des plus agréables et des plus fécondes (tout au moins pour moi) parmi celles que j'ai entreprises. Qu'il me soit permis de dire à la fin de cette préface tout le bien que nous pensions d'elle, non seulement sur le plan professionnel, mais aussi et surtout sur le plan de l'amitié. Je ressens avec une profonde tristesse le décès de Béatrice qui a tant travaillé pour l'astrophysique et la cosmologie et qui restera pour moi l'undes plus beaux exemples de courage et de noblesse.

Elle continuera donc à inspirer par ses travaux et sa mémoire ce domaine qui devrait nous permettre de nous réunir encore souvent dans ces rencontres.

Jean Audouze

CONTENTS

PROSPECTS FOR COSMIC RAY PHYSICS AROUND 10^{15} eV†

T. K. Gaisser
Bartol Research Foundation of The Franklin Institute
University of Delaware, Newark, Delaware 19711 U.S.A.

ABSTRACT

Knowledge of the chemical composition of primary cosmic ray nuclei is a prerequisite to understanding the origin, acceleration and propagation of the cosmic rays. Details of composition up to 100 GeV/nucleon have been studied extensively for many years, and the subject is a mature one in this low energy region. At much higher energies the flux becomes so low that the primaries cannot be studied directly, but can be observed only through their secondary cascades in the atmosphere. In the air shower energy region from 10^{14} to 10^{20} eV the subject of cosmic ray composition is therefore still in its infancy. New air shower experiments now operating promise to give qualitatively new data on longitudinal development of individual, large showers (10^{17}-10^{20} eV). New colliding beam experiments will soon give direct information on hadron-hadron collisions at energies equivalent to 10^{14}-10^{15} eV in the lab, which will enable significant progress in interpreting existing data on small showers and other atmospheric cascade experiments. We review the situation, with emphasis on energies around 10^{15} eV.

I. Introduction

Because the primary cosmic ray beam extends at least five orders of magnitude beyond 10^{15} eV, cosmic ray experiments can in principle provide information both about composition of the primary beam and about high energy physics well beyond accelerator energies. At energies above 10^{14} eV the primary flux is, however, so low that only indirect measurements of cascades in the atmosphere have been possible. Thus in practice ambiguities have so far made it difficult to draw definitive conclusions from experiment: there are too many parameters to be determined from data that necessarily has significant systematic uncertainties (see Table 1).

This situation should change soon. On the one hand, new colliding beam experiments will provide direct information about hadronic interactions around 10^{14}-10^{15} eV (see Table 2). This should make it possible to determine at least the general features of primary composition in this energy range. Such a result will be of considerable interest for astrophysics because it is already known[1] that there is structure in the energy spectrum just above 10^{15} eV and there are hints[2] of changes in composition also. Table 3 illustrates the importance of knowledge of the energy dependence of the primary composition around 10^{15} eV for understanding cosmic ray origin and propagation. Keeping in mind that the thickness of the galactic disc is about 300 pc and that the scale of large magnetic irregularities is $\sim$ 1 pc, it is clear, for example, that the interpretation of the change in the energy spectrum around 10^{15} eV[1] as well as the increasing anisotropy beyond 10^{15} eV[3] depend crucially on the composition. (Recall that cascade measurements determine total energy per nucleus rather than rigidity.)

At higher energies, several new cosmic ray air shower experiments now measure longitudinal development of individual air showers, including early portions of the shower before maximum. These include the atmospheric Cerenkov experiments of the Moscow State University[4a] and of the University of Durham [4b] groups, which operate around 10^{17} eV ($\sqrt{s} \sim 20$ TeV) and the Fly's Eye experiment at Utah[5] which now operates in the range 10^{18}-10^{19} eV ($\sqrt{s} \sim 100$ TeV). Because these techniques reflect early portions of shower development they are in a better position than conventional extensive air shower (EAS) experiments[6] to determine simultaneously properties of particle interactions and of the primary beam. In particular it should be possible to measure both the relative fraction of heavy and light primary nuclei and the nucleon-air cross section in an energy range that is significantly higher than will be reached by machines in the foreseeable future. The technique for achieving this has been described by Cassiday et al.[5] and is discussed elsewhere by

Table 1. Unknowns for high energy cosmic ray experiments

Interaction properties		Primary beam
σ	cross section—how does it depend on energy?	Chemical composition of primary nuclei
κ	inelasticity—what fraction of energy goes into particle production?	Energy spectra of components of primary beam.
f(x)	inclusive cross sections—how do shapes change with interaction energy	Possibility of new objects among high energy primaries.
<n>	multiplicity	
correlations		
P_T	Transverse momentum—jet structure?	

Table 2. Planned hadron colliding beam machines

Machine	Completion Date	C of m Energy	Equivalent Lab Energy
CERN $\bar{p}p$	1981	540 GeV	1.5×10^{14} eV
FNAL $\bar{p}p$	1984	2000 GeV	2.1×10^{15} eV
Isabelle	1986	800 GeV	3.4×10^{14} eV

Table 3. Larmor radius (pc) in 3μG field.

E_{total}(eV)	10^{15}	10^{17}	10^{19}
protons	0.3	30	3000
iron nuclei	0.01	1	100

Gaisser et al.[7], taking into account detailed effects of nuclear fragmentation on shower development. In the remainder of this paper I concentrate on the situation around 10^{15} eV.

II. Present Situation Around 10^{15} eV

Information about primary composition and properties of hadronic interactions in this energy range come at present from indirect observations of hadronic cascades in the atmosphere initiated by primary cosmic rays.[8] Three types of experiments can be distinguished: a) energy and depth dependence of uncorrelated fluxes of hadrons, of muons and of γ-rays;[9]b) fluxes and internal properties of γ-ray families[10,11] (jets of particles associated with a single primary but involving a cascade of hadronic interactions); and c) properties of small EAS.[12,13] There are two contrasting points of view about the implications of this data:

1) That hadronic scaling (limiting fragmentation in the form of radial scaling) is valid and that most observations can be understood by a suitable adjustment of the primary composition. In this case it is necessary to assume that the cross section rises with energy at the rate of about 15% per decade for $\sigma^{inel}_{p\text{-}air}$.[14,8]
2) That scaling cannot under any circumstances fit all the data, and that drastic changes in hadronic interactions are required.[9,12,13] Such changes should become significant already at beam energies of 100 TeV in the lab and would have the property that produced particles are relatively softer at high energies than at low; i.e. that interactions become more inelastic.

Hillas[14] has shown that all data on air showers (except the ratio of neutral/charged hadrons, but including N_μ/N_e data) can be fit by the picture with radial scaling and increasing cross section. Data on neutral/charge ratio conflict with each other[15] and in any case the non-scaling picture will probably also fail in this respect.[16] On the other hand, Stamenov concludes that no combination of scaling, increasing cross section and composition can give agreement with the N_μ/N_e data.

As for data on families, some properties of showers appear in agreement with scaling, while others, especially the calculated intensities, appear to be in strong disagreement.[10,11] Sensitivity to composition has not, however, been completely studied here.

Erlykin and Kuzina[9] conclude that a scaling model, even with increasing cross section, cannot simultaneously account for fluxes of hadrons and of the electromagnetic component at various atmospheric depths. In contrast Astafiev and Mukchamedshin[17] do find agreement with both. They also, however, find

fluxes of families in agreement with experiment, unlike the discrepancy found by others.[18,19]

In view of these conflicting claims and results, it is clear that further progress requires removing some of the ambiguity in the first column of Table 1. In the remainder of this paper I will therefore discuss to what extent proposed experiments at the $\bar{p}p$ collider[20] will be able to determine properties of hadronic interactions necessary to make definitive conclusions about composition around 10^{15} eV.[21]

III. Measuring Energetics of Hadronic Interactions with High Energy Colliding Beams

The diagram in Fig. 1 illustrates the physical basis in the quark picture for hadronic scaling. An incident hadron (nucleon in this example) considered as a bag of quarks strikes a target. The way in which the bag fragments into quarks is independent of the nature of the target and of the incident energy, but reflects only the momentum distribution of the quarks inside the incident hadron. (At this level, hadron fragmentation is like nuclear fragmentation, where fractional momenta of fragments reflect their Fermi momentum inside the incident nucleus.) Radiation processes during and after the collision, in which color is adjusted in accordance with confinement, are also assumed to scale as in the electromagnetic analog, leading to some form of Feynman scaling[22] or limiting fragmentation[23] for the overall process.[24] (The symmetric fragmentation of the target hadron and radiation in the pionization region are not shown in Fig. 1.)

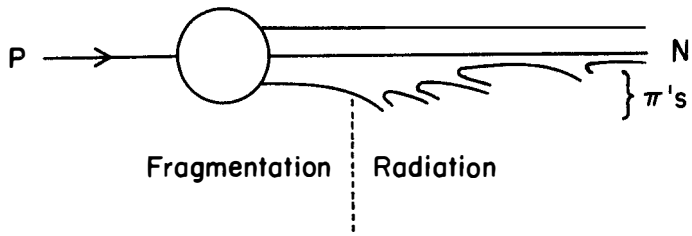


Fig. 1 Fragmentation of an incident nucleon in the quark picture.

Detailed phenomenology of hadronic processes within this type of picture has been carried out by various authors.[25,26,27] It is possible to account in detail for such features as quantum number ratios in various processes in

the fragmentation region. Although such analyses are not of the same theoretical status as those of hard scattering processes, where QCD can be used, the successful phenomenology gives strong confirmation of the underlying picture. We cannot, therefore, believe that hadronic scaling will suddenly disappear at cosmic ray energies. Moreover, if nucleon initiated collisions have a diffractive or leading component, then meson initiated processes must also.

If hadronic scaling is broken, the cosmic ray data suggest that processes become more inelastic with increasing energy. One can imagine two ways this could happen within the quark picture described above: (1) all processes gradually become more inelastic (steeper Feynman x distributions for inclusive cross sections), perhaps with a logarithmic dependence on energy as in breaking of Bjorken scaling; and/or (2) a new very inelastic component becomes increasingly important at high energies. Such a component could be associated with the rising cross section[10] and could conceivably even be associated with some completely new process (quark liberation, Centauros[11], Long-Flying Component[28], etc.). The main point I wish to make is that if either of these possibilities is realized presently planned experiments at the CERN $\bar{p}p$ collider will be able to measure and distinguish between them even if there are no associated exotic processes.

Many features of cosmic ray cascades are determined by the overall energetics of hadronic collisions. Fast secondaries and fragments with $|x| > 0.1$ thus play a very important role. This region of phase space may not be fully accessible to the earliest experiments at $\bar{p}p$ colliders, and this has led to some reservations about the extent to which these experiments will be able to remove the ambiguities in present interpretations of cosmic ray data. If, however, less energy is carried by particles with $|x| > 0.1$ then necessarily more must appear in the region $|x| < 0.1$, and this can be detected.

The detector of Ref. 20 will be able to measure the total hadronic energy emitted within the angular region $1° < \theta < 179°$ of the beam axis in interactions at $\sqrt{s}$ = 540 GeV. This corresponds to $x = 0.1$ at p_T = 500 MeV/c. The hadronic component and the electromagnetic component (from π° - decay) will be separately measured. As an example, I have evaluated the quantity $\langle F_h \rangle = \int_0^{0.1} x \frac{dn}{dx} dx$ in several models now in use for cosmic ray calculations. Here F_h is approximately the fraction of energy within the visible angular range carried by hadrons, excluding neutral pions. Table 4 shows $\langle F_h \rangle$ in three representative models: scaling, modified scaling[10], and scaling violation[9]. Reference 10 is an example of a two component model in which the importance of the non-scaling component increases with energy, whereas Ref. 9 is an example of a model in which scaling is broken in all interactions,

the degree of breaking increasing rapidly with energy starting at $E_{Lab} \sim 1$ TeV.

Table 4. Average values of F_h in various models

Interaction Energy (Lab in TeV)	Scaling[24]	Modified Scaling[10]	Scaling Violation[9]
0.1	.12	.12	.12
100	.12	.24	.23

It will also be possible and instructive to study the distribution of F for a sample of minimum bias events at the $\bar{p}p$ collider. Figure 2 shows schematically the results expected in the three models. If there is substantial violation of scaling between $\sqrt{s}$ = 60 and $\sqrt{s}$ = 540 GeV, this distribution might distinguish between a two component and a single component violation.

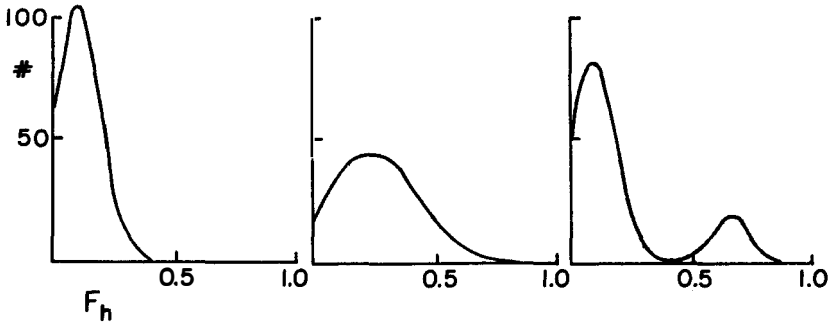


Fig. 2 Schematic illustrations of distribution of F_h in three types of model.

IV. Conclusion

The essential point that I wish to make is that forthcoming experiments at $\bar{p}p$ colliders will be able to distinguish among various models of hadronic interactions in the energy range $\sqrt{s} \gtrsim 500$ GeV (corresponding to lab energies $\gtrsim 10^{14}$ eV). This is true even if the experiments only detect secondaries with $1° \leq \theta_{cm} \leq 179°$, provided experiments are of calorimetric type as planned.[20,21] We can therefore expect substantial progress toward determining composition of primary cosmic rays in the energy range 10^{14}-10^{16} eV total

total energy per nucleus. This region is of particular interest for astrophysics because of the known structure in the spectrum in this region.[1] Furthermore, because of the low flux, it is accessible at present only to indirect experiments which require a knowledge of hadronic interactions in order to unfold the atmospheric cascading.

References

[1] A. M. Hillas, Physics Report 20, 59 (1975).

[2] See G. B. Yodh, this conference, for a review of this subject and its possible implications. See also J. Linsley and A. A. Watson, Phys. Rev. Letters.

[3] A. A. Watson, Proceedings of this conference.

[4] N. N. Kalmykov, et al., Proc. 16th ICRC (Kyoto) 9, 73 (1979) and G. B. Khristiansen, Ibid., 14, 360 (1979). K. Orford and K. E. Turver, Nature 264, 727 (1976) and R. T. Hammond et al., Nuovo Cim. C1, 315 (1978).

[5] G. L. Cassiday et al., Proc. 15th ICRC (Plovdiv) 8, 270 (1977) and Proc. Bartol Conference, A.I.P. Conf. Proc. #49, p. 419 (1979). Preliminary results for showers in the energy range 10^{18}-10^{19} eV have been presented at this conference by J. W. Elbert.

[6] Recall that a conventional EAS experiment samples each shower at one depth only.

[7] T. K. Gaisser, T. Stanev, Phyllis Freier and C. J. Waddington, Proc. IAU/IUPAP Symposium No. 94 (Bologna) 1980.

[8] For a general review and references see T. K. Gaisser and G. B. Yodh, Ann. Reviews of Nuclear and Particle Science 30, 475 (1980).

[9] A. D. Erlykin and N. P. Kuzina, Proc. Sofia Seminar on Cosmic Ray Cascades p. 99 (1980). See also P. N. Lebedev, Cosmic Ray Laboratory preprint #95 (1980) and Proc. 16th ICRC (Kyoto).

[10] A. M. Dunaevskii et al., Proc. Sofia Seminar p. 1 (1980). See also P. N. Lebedev preprints #206 and 207 (1978).

[11] C. M. G. Lattes, Y. Fujimoto and S. Hasegawa, Physics Reports 65, 151 (1980).

[12] J. N. Stamenov, et al., Proc. Sofia Seminar, p. 61 (1980).

[13] J. Wdowczyk, Proc. Sofia Seminar, p. 185 (1980).

[14] A. M. Hillas, Proc. 16th ICRC (Kyoto) 9, 13 (1979). See also M. Ouldridge and A. M. Hillas, J. Phys. G4, L35 (1978). See also T. K. Gaisser, R. J. Protheroe, K. E. Turver and T. J. L. McComb, Revs. Mod. Phys. 50, 858 (1978).

[15] T. K. Gaisser and P. Rudolf, J. Phys. G2, 781 (1976).

[16] Erlykin and Kuzina (Ref. 9) find a value of neutral/charge ratio lower than the experimental value for the uncorrelated flux.

[17] V. A. Astafiev and R. A. Mukchamedshin, Proc. 16th ICRC (Kyoto) 7, 204 (1979). See also K. Kasahara, et al., Proc. 16th ICRC (Kyoto 13, 70, 76 (1979).

[18] R. W. Ellsworth et al., these Proceedings.

[19] Yu. A. Fromin et al., J. Phys. G4, 1911 (1978).

[20] A. Astbury et al., (C. Rubbia, spokesman) Proposal CERN/SPSC/78-06, SPSC/P92 (1978). I consider the CERN $\bar{p}p$ collider here; the results, however, should also apply in a general way to the FNAL $\bar{p}p$ collider and to ISABELLE.

[21] Quantitative results are described more fully in T. K. Gaisser, Physics Letters 100B, 425 (1981).

[22] R. P. Feynman, Phys. Rev. Letters 23, 1415 (1969).

[23] J. Benecke, T. T. Chou, C. N. Yang and E. Yen, Phys. Rev. 188, 2159 (1969).

[24] The version of scaling that appears best to fit all present accelerator data on inclusive cross sections is radial scaling proposed by E. Yen, Phys. Rev. D10, 836 (1974). See Hillas, Proc. 16th ICRC (Kyoto) 6, 13 (1979) for a recent review of this subject. In all cases, for fast secondaries, the variable $x \to E/E_0$, the ratio of the energy of the produced particle to that of the incident particle. Scaling means the cross section depends only on the ratio x but not on the two energies separately.

[25] R. Hwa, Phys. Rev. D22, 1593 (1980). This is the version of the quark picture shown in Fig. 1 in which the nucleon fragments into dressed valence quarks ("valons").

[26] D. W. Duke and F. E. Taylor, Phys. Rev. D17, 1788 (1978).

[27] See J. F. Gunion, Proc. XI Int. Symposium on Multiparticle Dynamics, Bruges (1978) for a review. See also S. Pokorski and S. Wolfram, CALT 68-795.

[28] V. I. Yakovlev et al., these proceedings and Proc. 16th ICRC (Kyoto) 6, 59 (1979).

COMPOSITION OF COSMIC RAYS AT HIGH ENERGIES

G. B. Yodh
Department of Physics and Astronomy
University of Maryland
College Park, Maryland, USA 20742

Abstract: A critical analysis of experiments pertaining to the composition of primary cosmic rays from 10^2 to 10^{11} GeV is presented. Experimental observations on time structure of hadrons near air shower cores, on multiple muons with energy in the several TeV range, on variation of muon number with electron number, on fluctuation of muon number in showers of fixed size, on energy variation of longitudinal development of showers and on isotropy of cosmic rays are examined assuming no drastic change in the nature of high energy interactions. One finds that the composition of cosmic rays is continually varying from being predominantly light at 100 GeV to mostly heavy at about 10^6 GeV and reversing back to predominantly light above 10^8 GeV. The "nested leaky box" model of origin, propagation and acceleration of cosmic rays, with a galactic cut off at a rigidity of 10^5 GV/nucleon and a dominant extra-galactic component above 10^8 GeV per nucleus naturally predicts this behavior.

1. Introduction

The cosmic ray all particle spectrum extends from energies of a few GeV up to 10^{11} GeV and spans about nineteen decades in intensity. Apart from a steepening at 10^{7} GeV and a flattening at 10^{9} GeV the spectrum is remarkably smooth (Sreekantan 1979). As of now, there exists no satisfactory theory of origin, propagation and acceleration of cosmic rays over the whole energy span in spite of our thirty years of study. (Lingenfelter 1979).

At energies below a hundred GeV per nucleon, modulation in the heliosphere and energy variation of path lengths modify the spectra of cosmic rays. It is generally believed that at still higher energies, where modulation effects become negligible and where the effective path length becomes a constant one may be able to measure the source spectra. There may exist a regime of energies, where galactic confinement and scattering allows one to study these source spectra by measuring high energy cosmic rays and differentiate between various models of cosmic rays. (Cesarsky 1980).

Current models for the origin, acceleration and propagation of cosmic rays generally address relatively low energy region; below hundred GeV per nucleon. Sources of cosmic rays are assumed to be located within galaxies. The low energy cosmic rays are supposed to arise from sources distributed in the disc of our own galaxy. Acceleration of cosmic rays to hundreds of GeV per nucleon is considered to be due to processes such as the Fermi mechanism, betatron acceleration and acceleration in pulsars. Isotropy of cosmic rays is considered to be a consequence of diffusive scattering by magnetic irregularities in the interstellar medium. Energy variation of ratio of secondary to primary elements in the cosmic ray beam implies an energy dependent path length for cosmic rays. There are at least two distinct ways to obtain the required path length variation. One assumes that all variation takes place outside the sources and is due to an energy dependent leakage from a galactic leaky box. The other assumes that path length variation occurs in the sources themselves, the residence time in the neighborhood of the sources being longer for lower energy cosmic rays which can be more easily contained in the source region. In the first approach, augmented leakage from the galaxy must start at a relatively low rigidity ($\sim$ 10 GV/nucleon), while for the second approach (Nested Leaky box) leakage from the galaxy need not occur till much higher rigidity, of about 10^{5} GV/nucleon where galactic magnetic fields can no longer contain these energetic cosmic rays effectively. These models predict energy variation of spectra of cosmic ray nuclei which are in reasonable agreement with experimental data at low energies. However, they lead to quite different behaviour at high energies as regards the spectral composition of cosmic rays and isotropy of cosmic rays (details of these

ideas can be found in the Proceedings of IAU-IUPAP Symposium on "The Origin of Cosmic Rays" held in Bologna, June 1980). The galactic leaky box model predicts either that all components will exhibit flat source spectra or that all spectra will become steep depending on whether path lengths become energy independent or continue their decrease indefinately, respectively. Such a scenario could lead to large anisotropies above 1000 GV/nucleon. The nested leaky box model, in contradistinction, would require all spectral indices to exhibit source properties (an index of about -2.6) above 100 GV/nucleon which will continue up to a rididity where augmented galactic leakage sets in, all spectra steepening by the same power. This model predicts isotropy below the rigidity cut off.

Any model must account for several features of cosmic rays observations at high energies. They must give (i) high degree of isotropy up to 10^5 GeV, (ii) they knee bend at 10^7 GeV, (iii) the increasing anisotropy with energy and (iv) the "ankle" bend at 10^9 GeV.

This paper examines the available data on cosmic rays at high energies and attempts to narrow down the allowed variation of elemental composition with energy. It is shown that there is no "Normal" composition of cosmic rays. The composition varies from predominantly light to mainly heavy as one goes from 10 to 10^6 GeV, after which it appears to again become predominantly light by 10^8 GeV.

In order to set the stage for extension to higher energies, a short summary of direct measurements of cosmic rays is given in this introduction.

Direct measurements of cosmic ray particles above the atmosphere have been done using a variety charged particle telescope carried aloft by balloons and satellites. Even after several decades of effort, limitations of weight and exposure times has prevented experimenters from obtaining direct information about elemental composition above energies a few hundred GeV per nucleon.

At energies above 10 GeV/nucleon, particle energies have been measured by Cherenkov counters (Juliusson 1974; Lezniak and Webber, 1980; Caldwell, 1977; Balasubrahmanyan et. al. 1980; Arens et. al. 1979), by magnetic spectrometers (Smith et. al. 1973, Orth et. al., 1978), by ionization calorimeters (Ryan et. al. 1972, Ormes and Balasubrahmanyan 1973, Schmidt et. al. 1976, Grigorov et. al. 1971, Simon et. al. 1980) and in some experiments by suitable combinations of these elements. Recently, emulsion chambers of large area are being flown (JACEE collaboration 1979) to extend the knowledge of higher energies.

In a set of pioneering experiments, Grigorov and his colleagues (Grigorov et. al. 1971) in series of PROTON satellites measured the spectrum of all particles up to an energy per nucleus of almost 10^{15} eV or 10^6 GeV. The most important result obtained was that the all particle spectrum has a single effective spectral index of 1.64 ± .01 from 100 to 10^5 GeV per nucleus, (see figure 1).

Their measurement of proton spectrum, however, has been shown to be susceptible to systematic effects of backscatter (Ellsworth et. al. 1977). Their all particle spectrum is consistent with the sum of individual elemental spectra measured in other experiments up to about 2000 GeV per nucleus.

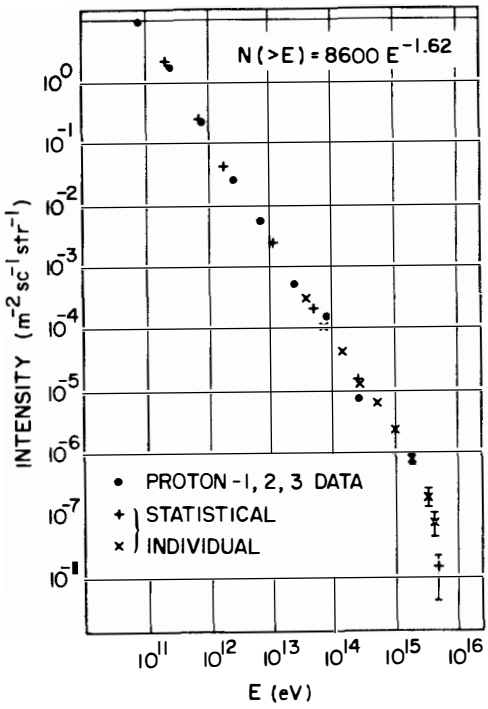


Figure 1: All particle spectrum as measured by the series of "PROTON" Satellites (Grigorov et. al. 1971). Data from different satellites was normalized to fit world data below 100 GeV. The curve can be fitted at a spectral index of 1.64 ± .01 from 100 to 10^6 GeV per nucleus.

At energies where elemental spectra have been directly measured one salient feature is that the observed spectral indices of different elements are different. Between 10 and 200 GeV per nucleon the spectral indices vary from being 2.75 ± .02 for protons to 2.20 ± 0.2 for the iron component (Julisson 1974, Ryan et. al. 1972, Ormes and Balasubrahmanyan 1973). These variations have been interpreted as due to variation of effective path lengths. A summary of world data on elemental spectra at high energies are shown in figure 2. (Ormes and Freier, 1978 and Simon et. al. 1980). An examination of these graphs show that better data

are needed to clarify the situation and to determine whether all spectral indices become the same at high energies or they remain different. In particular, if all indices were to become the same then it is important to determine whether they become about -2.3 or about -2.7.

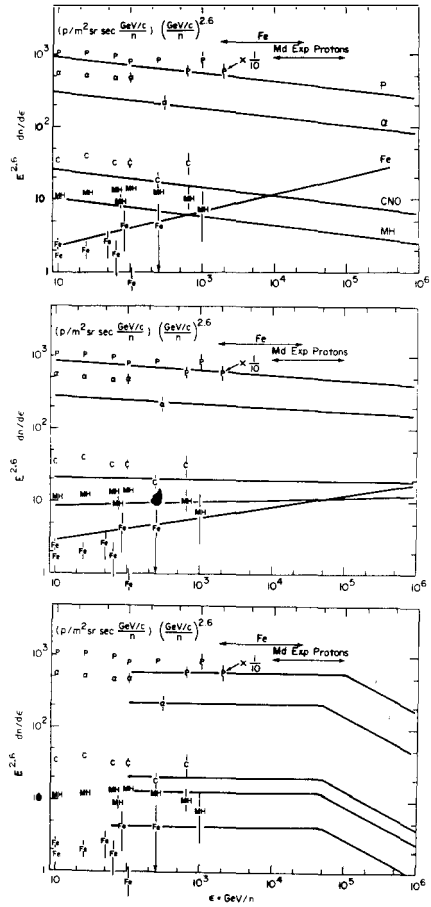


Figure 2: Differential spectra of major elemental components of cosmic rays above 10 GeV per nucleon. The ordinate is multiplied by $\varepsilon^{2.6}$ and proton points have been scaled down by a factor of ten. The fits used in Models are superposed.

The data show a continuously varying composition between 1 and 200 GeV per nucleon, however, due to the dominance of light species the change in average atomic weight, <A>, is rather small. If the flat iron spectrum were to continue up to 10^6 GeV per nucleus than it is obvious that at such energies cosmic ray mix would become predominantly heavy and have a much larger <A> (see figure 15).

Section 2 outlines the essential procedure used in this analysis. Section 3 discusses the experimental observations and draws conclusions from them regarding models proposed in Section 2. Conclusions are summarized in Section 4.

2. Outline of the Method

The method of analysis is to relate an assumed set of primary spectra with experiments listed below (3(a) and 3(e)) using a well defined particle interaction model and keeping the number of free parameters of primary spectra to a minimum. The analysis is based upon the following requirements:

(1) The cosmic ray spectra used in extrapolations to higher energies are anchored to direct measurements discussed in Section 1 at an energy where the quoted errors in elemental fluxes are less than or equal to twenty-five percent. (The flex values and energies used for normalization are given in Table I).

(2) The extrapolated spectra must add up to give an all particle spectrum which agrees with PROTON satellite measurements up to 10^6 GeV and above that with the best estimate of the flux from air shower measurements. (Gaisser and Yodh 1980, Gaisser et. al. 1978, Hillas 1979a,b). The entire composite spectrum is shown in figure 3.

(3) Compatibility with air shower experiments listed below:

- (a) Recent observations on time structure of hadrons near air shower cores (Goodman et. al. 1979).
- (b) Measurements of rates of high energy (>1000 GeV) multiple muons (Elbert, J. W., 1978).
- (c) Variation and fluctuations of total number of muons, N_μ, with total number of electrons, N_e, in a shower (Kalmykov et. al. 1975, and Elbert et. al. 1976).
- (d) Lateral structure of high energy muons in showers (Acharya et. al. 1979).
- (e) Variation of shower maximum with energy (Thornton, Clay, 1980, Linsley 1977 a,b, 1980 a,b 1981, Aguirre' et. al. 1973, Gaisser et. al. 1978).

(4) Extrapolation of total interaction cross sections and particle production distribution functions, according to current ideas of strong interactions (Ellsworth et. al. 1981 a,b).

TABLE I

Normalization Energy ε_o		$< A >$	$(dn/d\varepsilon)^*\varepsilon_o$
GeV/n	GeV/nucleus		
2000	2000	1	1.5×10^{-5}
300	1200	4	8×10^{-5}
250	3750	15	1.2×10^{-5}
75	1950	26	1.73×10^{-4}
63	3528	56	9×10^{-5}

*$(dn/d\varepsilon)\varepsilon_o$ is given in units of particles per (m^2, sr, s, GeV/n). The spectra are assumed to be of the form $(dn/d\varepsilon)_A = K_A^\gamma \varepsilon^{-\gamma A}$ and k_A^γ is adjusted for each value of γ_A assumed, at energy ε_o.

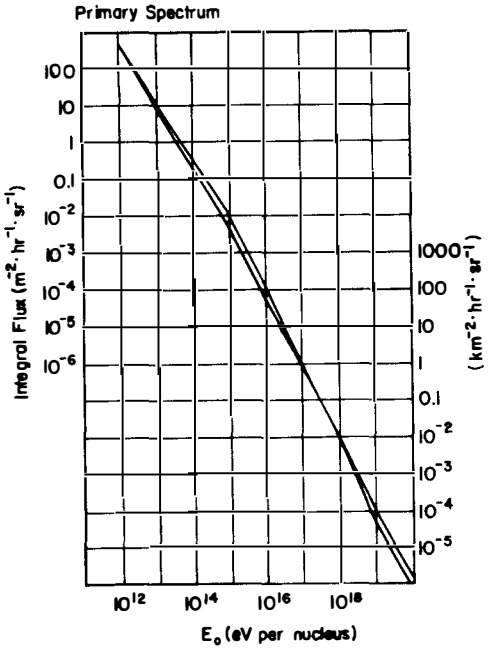


Figure 3: A composite integral energy spectrum of all particles from 10^{11} to 10^{20} eV. Shaded areas give a measure of uncertainties in the knowledge of the absolute flux (taken from Gaisser and Yodh 1980).

(5) Requirement that the above extrapolations of particle interactions and the assumed cosmic ray spectra one accounts for:

(a) Uncorrelated hadron and muon fluxes in the atmosphere (Ellsworth et. al. 1981c) and

(b) Detailed features of C-jets and A-jets from Japan-Brazil emulsion chamber experiment at Mt. Chacaltaya (Lattes et. al. 1973, Fujimoto and Hasegawa 1978, Lattes et al. 1980).

We therefore, use a high energy interaction model satisfying points (4) and (5) stated above, which incorporates; (a) rising cross sections (Gaisser and Yodh 1980), (b) an independent particle emission scaling model for particle production which uses x-radial as the variable to extrapolate to high energies ($x_R = \frac{2E^*}{\sqrt{s}}$ where E^* is total energy of paritcle in C. M. and $\sqrt{s}$ is the available energy in C. M.) according to work of Ellsworth et. al. (Ellsworth et. al. 1981a and b), (c) a transverse momentum distribution of a two component form (Halzen and Luthe 1978) for which the non-exponential contribution become about 30 percent at 10^5 GeV (Ellsworth et. al. 1981a), (d) leading particle effects are contained in the model separately for nucleon and pion collisions and (e) which agrees with highest energy accelerator data.

Hadron-nucleus collisions are treated as hadron-nucleon collisions. This simplification does not affect the results in any significant way because the steep cosmic ray spectra emphasize only large x processes for which this approximation has been shown to be good (Busza et. al. 1975). Collisions of primary nuclei are treated usually as superposition of A independent cascades of energy E/A, E being the energy per nucleus. However, where it is important we have used fragmentation models based upon actual data to break up nuclei in the atmosphere (Freier and Waddington 1975).

To summarize, we fix the interaction model, then simulate (either by a full Monte-Carlo or by analytical calculations) the experimental observations, varying two to three parameters which quantify primary cosmic ray spectra until an adequate representation of the data is obtained. This restriction to a minimum set of parameters is imposed by the fact that experiments generally determine, independently, only two or three parameters of air showers (such as, N_μ and N_e or N_e and shower maximum (X_m) or trigger rate and rate of delayed hadrons or number of events with 1, 2 and 3 energetic parallel muons), and because the resolution of the techniques is not sufficient to separate events from individual elements on an event by event basis. We also utilize the general belief that primary spectral indices should vary relatively slowly with A at least up to 10^7 GeV where the famous bend in all particle spectral index by 0.5 takes place. (See figure 3).

In the simulations to be discussed in the next section, three different models for primary cosmic ray spectra are considered below 10^7 GeV.

Model I: A two component model wherein it is assumed that P,α, CNO, and MH nuclei have the same spectral index γ_1 and H component has a different spectral index γ_2. For this model one must find mechanisms for origin, acceleration and propagation which preferentially treat iron (see for example the model for a "pulsar" bump by Wdowczyk and Wolfendale 1973, see also, Elbert et. al. 1975).

Model II: Energy spectra determined by an acceleration mechanism which depends on Z, A or Z^2/A. For such a model the parameters are γ_o and α where the spectral index for the species A is given by $\gamma_A = \gamma_o - \alpha\ (A-1)$.

Model III: All spectra above 2000 GeV become flatter with a slope of $\gamma_o \sim 2.6$ and have a cut-off at a rigidity R_c volts/nucleon, after which the slope increases by 0.5. The two parameters are γ_o and R_c (Cowsik 1968).

We again emphasize here that all particle spectrum has a slope of 2.64 ± .01 below 10^6 GeV. However, slopes of the predominant components, P and α, at low energies, between 10 and 2000 GeV, are about 2.75, hence to obtain the flatter all particle spectral index one should have some components with indices smaller than those for P and α or P and α index must decrease at high energies. The models considered have this flexibility and furthermore can represent most current theoretical ideas outlined in the introduction in a phenomenological way.

3. Experimental Observations:

Before giving the results of the analysis on composition of cosmic rays, the experimental data to be explained are presented and their sensitivity to various components of the analysis discussed. Five different experiments are considered:

a) Time structure of hadrons near air shower cores:

The recent experiment of the University of Maryland group (Goodman et. al. 1979 (a)(b)) measured the arrival time distribution of energetic hadrons (E > 3 GeV) with respect to the shower front. The trigger required was three fold: (i) a minimum shower particle density near the detector (> 18 p/m^2), (ii) a minimum hadronic energy over an area of $4m^2$ (> 35 GeV in a calorimeter) and (iii) a minimum pulse height in the 0.5 m^2 counter which measured the arrival time of hadrons (pions, nucleons or kaons) with respect to the shower counters (> 3 particles). The basic data is shown in figure 4, which shows a scatter plot of pulse height of T3 (which is a measure of hadron energy) and its time structure. A total of 21,700 triggers were obtained and the fraction of hadrons with delay > 15 ns was = 120 ± 10/21,700 $\sim$ 0.55 ± 0.05%. The philosophy of the analysis is to reproduce these two basic measurements; the trigger rate and the fraction de-

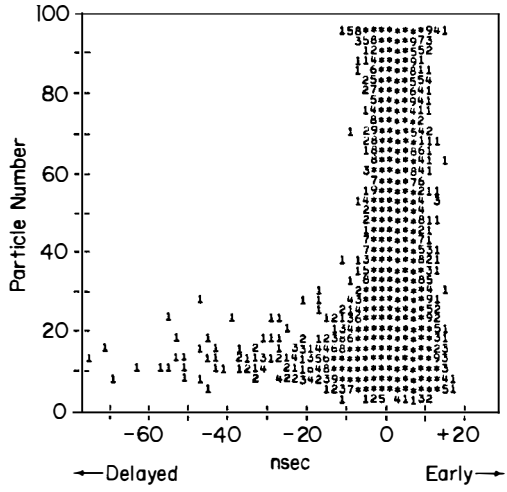


Figure 4: A scatter plot of time delay versus pulse height of hadrons in the University of Maryland delayed hadron experiment (Goodman et. al. 1979). Energy of the hadron in GeV is approximately equal to the numerical value of pulse height. The plot contains a total of 21,700 events.

layed. One can also try and reproduce the actual density distribution of points in the scatter plot.

The experiment detects the debris of an air shower, started by a cosmic ray particle of high energy, after about 8 interaction lengths. Trigger rates and shower Monte-Carlo calculations indicate that the events are being generated by primaries with energies above 10^4 GeV. Why should the experiment have any sensitivity to the nature of the primary? The qualitative explanation is that a particle density of 18 p/m^2 and energy of 35 GeV in the calorimeter places the detector within few meters of the core. To acquire a significant delay the hadron must be a low energy nucleon. Low energy nucleons tend to be distributed at a sizable distance from the core. The lateral distribution of shower particles initiated by an iron primary of the same total energy per nucleus as a proton initiated event will be older and also flatter and hence trigger conditions can be satisfied further from the core for Fe events than for proton events. Therefore the delayed fraction is an A dependent quantity. Thus for a given trigger rate iron induced events have a much greater efficiency for producing a delayed event than protons; Monte Carlo calculations give a ratio for efficiencies greater than 50! Simultaneously fitting trigger rate and delayed fraction allows one to determine two parameters. The results require substantial increase of the heavy component by 10^6 GeV for all three models.

The actual procedure is to use a four dimensional Monte-Carlo for the cascade to calculate the trigger effectiveness as a function of energy for each species, $N_A(\varepsilon)$, where ε = Energy per nucleon. As most of the low energy nucleons come not from the highest energy collisions but lower energy ones, the sensitivity to details of very high energy collisions is not large. Physics of lower energy is fairly well known hence the calculation of $N_A(\varepsilon)$ is not very sensitive to extrapolations of the interaction model. The delayed particles contain leading nucleons, recoil nucleons and produced nucleons. In figure 5 are shown calculated curves for $N_A(\varepsilon)$ for protons and iron primaries. The steep rise is due to meeting the shower density criterion and the bend occurs when shower requirement is met every time. The function continues to increase slowly because multiplicity increases only slowly with energy. Trigger rate R_A for each species is found by convoluting $N_A(\varepsilon)$ with $(dn/d\varepsilon)_A$ and is a function of γ_A, the spectral index. The total trigger rate is then found by summing over A; $R = \Sigma R_A(\gamma_A)$. The same Monte-Carlo is used to find the delayed fraction for each species, f_A. This quantity is a slowly increasing function of energy and relatively independent of γ_A, but rapidly varying with A. Thus we simulate the two experimental numbers: the trigger rate, R, and number delayed, $D = \Sigma f_A R_A$.

For each model, contours for fixed values of input parameters are drawn in R, D plane and compared with data points. The results are displayed in figures 6 (a), (b) and (c) for the three models respectively. In generating these figures, values of $N_A(\varepsilon)$ and f_A have been fixed at the best values given by the Monte-Carlo calculations.

It is reasonable to ask how much can these values be changed by changing the model of high energy interactions within the constraints discussed in the previous section. We have found that it difficult to change $N_A(\varepsilon)$ by more than ~ 30% at a given energy. The delayed fraction is even less sensitive to change in the model. Propagating these systematics into the final determination of spectral parameters the results given in Table II are obtained for the three models. In order to indicate how good these fits are the predictions of these models at low energies, they are compared to data shown in figure 2. Alternatively, the predictions of the ratio H/CNO are compared with data in figure 7.

Below 100 GeV/nucleon, all three model predictions are changed by the variation of effective path lengths. All of them can be made to agree with the observed decrease in iron to CNO ratio with decreasing energy.

To better illustrate the significance of this analysis, the variation of percentage of iron in cosmic ray flux as a function of energy is shown in figure 8. Most striking result is that these data on delayed particles demand a dominance of H component at 10^6 GeV.

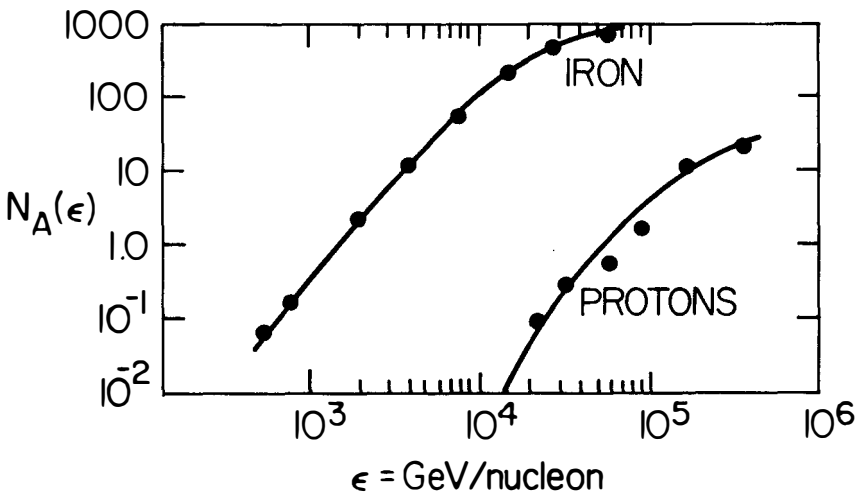


Figure 5: The trigger effectiveness, $N_A(\epsilon)$, for elemental groups as a function of energy per nucleon. The points were calculated by a 4-dimensional Monte-Carlo program using a superposition model.

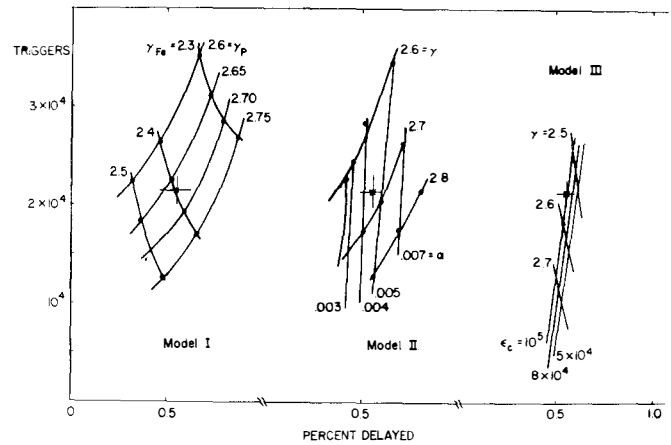


Figure 6: Fixed parameter contours for trigger rate versus fraction delayed for the three models: (a) Model I - shows that $\gamma_{FE} \sim 2.4$ while $\gamma_p \sim 2.7$ is required; (b) Model II - the best fit is $\gamma = 2.67 - .0045(A-1)$ and (c) Model III - $\gamma = 2.6$ and $\varepsilon_c \sim 10^5$ GeV/nucleon gives reasonable agreement. These curves show that delayed fraction is affected much less than trigger rate when spectral indices are varied. Curves should be considered an illustration of what the procedure is, the fits used are given in Table II.

TABLE II

Model Parameters

Model I: Bimodal, $dn/d\varepsilon = K_a \varepsilon^{-\gamma_A}$ P(m^2, sr, s, GeV/n), and ε in GeV/n.

Model II: $\gamma_A = \gamma_o - \alpha(A-1)$

Model III: COMMON SPECTRAL INDEX AND RIGIDITY CUT OFF

Model	Parameters	Elemental Group				
		p	α	CNO	$10 \leq z \leq 16$	$26 \leq z \leq 30$
I	γ	2.71	2.71	2.71	2.71	2.36
	K	12,271	390	36	14	1.3
II	γ	2.67	2.65	2.61	2.57	2.45
	K	10,135	305	22	8	1.89
III	Rigidity CUT OFF at 10^5 GV for P					
	γ	2.5	2.6	2.6	2.6	2.6
	K	5,738	221	21	13	4.4

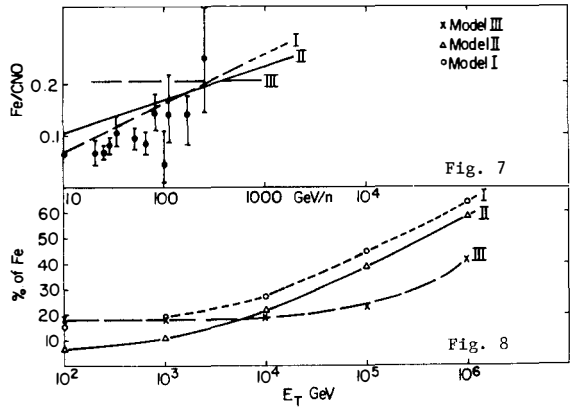


Figure 7: Energy variation of ratio of rion group to the CNO group fluxes. The normalization of the models used with data at about 100 GeV/n is indicated.

Figure 8: Variation with energy of percentage of 'iron' in the cosmic ray mix for the three models - below 10^6 GeV per nucleus.

We note further that the trigger rate would be greatly exceeded if the proton spectrum were to become flatter beyond 2,000 GeV with a spectral index of 2.3 (source spectrum?). It will also be difficult to account for the slope and flux of Grigorov's all particle spectrum.

If all spectra were to acquire the same slope of 2.7 or greater beyond 500 GeV/nucleon then again one obtains inconsistency with these data. The delayed fraction would be too small, the total trigger rate too low and all particle spectral index in disagreement with Grigorov.

b) High Energy Multiple Muons

Muon and neutrino detectors placed deep underground can study the frequency distribution of arrival of multiple muons with energy above a threshold set by the overburden of rock. This threshold is high generally around 10^3 GeV. The high energy muons come from decay of energetic pions and kaons high in the atmosphere and their flux is sensitive to the nature of the first few interactions of primary cosmic rays in the atmosphere. In general one is dealing with hadronic interactions in 10^4 to 10^7 GeV range. Fixing the high energy physics model, one can study the sensitivity of multiple muon frequency to the nature of primary responsible for these events. As the probability for decay of pions and kaons is greater at higher altitudes, those species which can interact earlier will give rise to more multiple muons.

The data from experiments by the Utah group (Mason et. al 1975), by the Homestake Mine experiment (Deakyne et. al. 1978) and by the Baksan experiment in

caucus mountains (Chudakov et. al. 1979 and Alexlyev, 1979) can be analyzed to determine the primary composition (Elbert 1978, Mason et. al. 1975, Lowe et. al. 1976, Elbert et. al. 1975 a,b).

Careful analysis by Elbert (Elbert et. al. 1981) using a model similar to I show that the data require the composition given in Table II below 10^6 GeV which implies dominance of heavies at 10^5 to 10^6 GeV. The contribution to the multiple muon flux, with multiplicity n, can be shown to be proportional to about the second power of atomic weight A (Elbert 1978) and so high multiplicities can be very sensitive to the fraction of heavies. For n = 10 and E_μ > 2.6 TeV the Homestake group (Deakyne et. al. 1978) obtain a ratio of $\sim$ 50 for the relative rates due to iron and proton components for Model I. The average primary proton energy which give rise to doubles of E > 3 TeV is about 200 TeV, while average energy for tenfolds would be above a 1,000 TeV. In figure 9 the sensitivities to atomic weight of the primary of the multiple muon method and delayed hadron methods are compared. The curve for muons is for the case of obtaining ten muons through detector. It is clear that with a large enough detector placed deep underground one obtains high sensitivity to A.

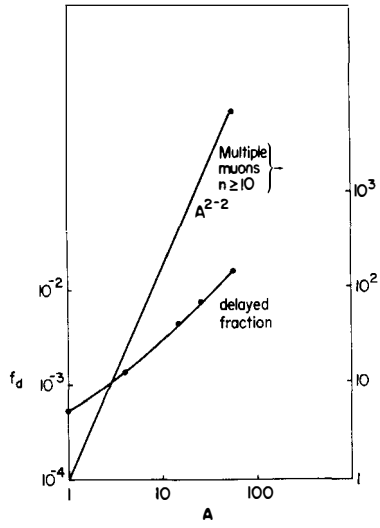


Figure 9: Sensitivity to atomic weight of the primary initiating (i) a tenfold or greater than tenfold multiple muon events with E_μ > 3 TeV and (2) a delayed hadron near air shower core. The primary energies involved are greater than 10^4 GeV per nucleus.

Mason et. al. have shown that using an approach quite similar to the one used here the muon-pair rate can only be accounted for if the heavy component has a flatter slope up to the break in the primary spectrum, which they take to be about 3 x 10^6 GeV total energy. At 3 x 10^6 GeV they require 80% iron in the cosmic ray mix.

Recently, data on multiple muons arriving from the vertical with E > 2.6 TeV has been obtained at Homestake. The rates of double and triple muons of the same energy but from vertical and inclined directions can be compared with Monte-Carlo simulations. Again, the comparison shows the need of increasing fraction of heavies consistent with models of this paper (Elbert, et. al. 1981).

c) Relations between total number of muons, N_μ and total number of electrons, N_e, in showers:

The experimental data on average value of N_μ as a fucntion of average value N_e is shown in figure 10. (Kalmykov et. al. 1975, Olejniczak et. al. 1977).

Muons arise primarily from charged pion decay. This decay probability depends on pion energy. Number of electrons, on the other hand, are related to the energy going into the electromagnetic component via π^os. It is reasonable to expect that the relation of N_μ to N_e should depend on energy per nucleon of the cosmic ray primary for any high energy model. It is found from Monte-Carlo simulations (see for instance a review paper by Gaisser, Protheroe and Turver 1978) that A nuclei give more muons for a given electron size than protons. If scaling is assumed to be valid, the calculation requires a preponderance of A=56 nuclei to account for N_μ vs. N_e data in the energy range 10^5 to 10^7 GeV. (Shower sizes of 10^4 to 10^7 particles). If primary cosmic rays were mainly protons, then one would require gross violation of scaling and very high multiplicities. The energy range covered by these experiments corresponds to 10^5 to 10^7 GeV.

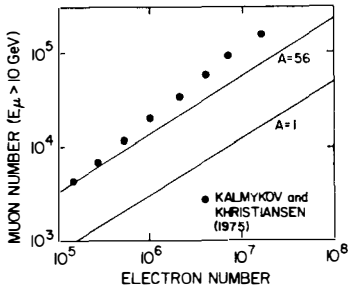


Figure 10: The variation of number of muons of energy > 10 GeV as a function of average electron shower size.

Through a study of lateral distribution of muon above some minimum energy (say 2 GeV) one can determine N_μ for each shower of size N_e. If σ_μ is the variance in N for fixed N_e then one studies variation σ_μ/N_μ as a function of N_e. A summary of each measurements is shown in figure 11.

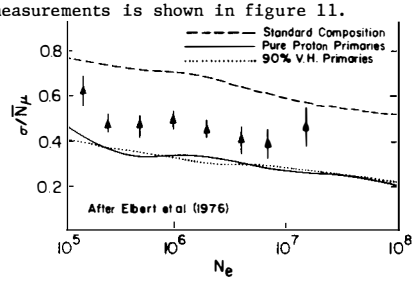


Figure 11: Relative dispersion of muon number in showers if fixed sea level electron size. Note that standard composition corresponds to mainly proton initiated events. A mix containing 20 or 30% protons and 70 or 80% heavies and medium heavies would give agreement with data.

Shower size varies from 10^5 to 10^7 (corresponding to an E_o range: 2×10^5 GeV to 2×10^7 GeV approximately) σ_μ/N is seen to lie between 0.4 and 0.6. Monte-Carlo simulations by the Utah group (Elbert et. al. 1975) point out that it is not possible to account for these observations (they assume a scaling model) with either pure protons or pure iron. They conclude, however, that it is possible to explain the data with a preponderance of iron with a small but non-negligible fraction of protons.

Both of these results are consistent with a changing composition which becomes iron rich by 10^6 GeV.

d) Lateral structure of high energy muons in showers generated by primaries of 10^5 to 10^6 GeV.

Using underground muon detectors in correlation with air shower array, the Tata Institute group has studied showers initiated by primaries of energy around 10^5 GeV. (Acharya et. al. 1979). The lateral distributions of muons with energy greater than 220 GeV are compared with that for muons with energy greater than 1 GeV.

Detailed Monte Carlo simulations using models similar to the one considered here show that the data require the primary composition to be a mixed one. A composition like the one outlined here is needed, some 30% of the showers must be due to heavy component at 10^5 GeV.

e) Study of energy dependence of shower maximum and shape of shower development:

(1) Rapid longitudinal development of showers:

From observations of zenith angle variation of shower size for fixed intensity of showers, it is possible to study the longitudinal development of

the nuclear-electromagnetic cascade in the atmosphere above 10^{14} eV. Constant intensity restriction fixes the primary energy of cosmic rays initiating the shower, and changing the zenith angle varies the depth at which air shower is sampled. To illustrate the method we show, in figure 12, observations by the BASJE experiment done at Mt. Chacaltaya, Bolivia, 5200 meters above sea level (Aguirre et. al. 1973). The position of the shower maximum y_{max} can be easily shown to be related to the energy per nucleon by the relation

$$y_{max} = c + B\ln(E_o/A$$

One solves this equation to calculate the effective A of the primaries of energy Eo giving

$$A_{eff} = E_o \exp[(C - y_m(obs))/B]$$

with B being the cascade length in air, 37 g/cm^2. Two sets of curves are shown, one for A=1 and the other for A=56, based upon detailed calculations based upon scaling models. The rapid rise in the cascade curves is again indicative of preponderance of iron nuclei. (Gaisser, Protheroe and Turver 1978).

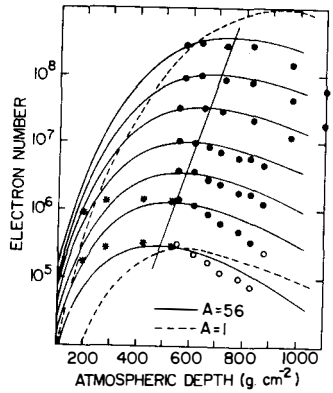


Figure 12: Longitudinal development of electron cascades initiated by primaries of the same energy (seven different groups of equal intensity curves are shown) compared with computations by Gaisser et. al. (1978). The slanted line shows estimated position of shower maximum for the different energy bins.

(2) Variation of depth of shower maximum with energy:

The depth of shower maximum for a pure EM cascade, initiated by a primary of energy E_o is given by $X_M^{EM} = (t_R/\ell n\ 2)\ell n(E_o/\varepsilon)$ where radiation length, t_R, is 37 g/cm^2 and the critical energy, ε, is 80 MeV in air. For a nuclear-electromagnetic cascade, initiated by a primary of mass A one finds that dependence of X_M on E_o, A and interaction lengths can be described by $X_M \simeq X_1 + t_R(1-\alpha)\ell n(E_o/A)$

+ $\beta(\lambda_\pi + \lambda_N)$ (Linsley 1977, Hillas 1979, also see Gaisser et. al. 1978) where α is a parameter that reflects the energy dependence of multiplicity and β is related to inelasticity in meson and nucleon interactions. If scaling holds in the fragmentation region then $\alpha = 0$.

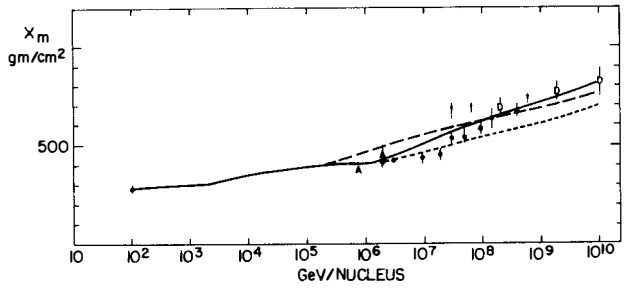


Figure 13: Variation of X_m with energy compared with calculations based upon different assumptions discussed in the text.

Several points are to be noted about this relation:

(a) If the primary composition is changing with energy, the rate of change of X_m with E_o will be affected. An increase in H component would slow down the growth of X_m and vice versa.

(b) If scaling is violated in the fragmentation region, $\alpha \neq 0$, and as α is a positive number less than one, rate of growth of X_m is again reduced.

(c) It is known that inelastic cross sections increase with energy. This implies that λ_π and λ_N decrease with energy. This again slows down the rate of growth of X_m with E_o.

(d) If scaling is valid, composition is constant and interaction lengths are energy independent one obtains the largest rate of growth of X_m with energy.

Measurements of time structure of Cherenkov light from air showers have been used to find X_m as a function of energy (shown size) (Orfond and Turver 1976, Hammond et. al. 1978, Protheroe and Turver 1979, Thornton and Clay 1980, Efimov et. al. 1973 and Kalymkov et. al. 1979). Shower maxima can also be deduced from equal intensity curves (Aguirre et. al. 1973) and by making shower studies as a function of altitude (Antonov et. al. 1979). The variation of X_m with energy is shown in figure 13 in which recent corrected data of Thornton and Clay have been shown along with other data. (Linsley and Watson 1981).

Let us examine these data according to the method used by Linsley and Watson to derive information about primary composition at still higher energies, energies above 10^6 GeV. First we "nail" down X_m by Monte-Carlo calculation at 10^2 GeV where particle physics is experimentally known. A value of 285 ± 15 g/cm^2 down from the top of the atmosphere is obtained (W. V. Jones 1978). One notices that no single straight line on this semi-logarithmic plot fits the data. Such a

straight line would arise if α is a constant, λs' are energy independent and effective A of primary beam is constant.

We have pointed out in section 2 that up to 10^5 GeV, scaling holds in the fragmentation region and that cross sections are increasing. We calculate an energy variation of X_m by (i) requiring the curve to pass through 285 g/cm^2 at 100 GeV and (ii) incrementing X_m by energy variation of λ_π and λ_N, using β = 1.5, while keeping the primary composition above 10^5 GeV, the same as that at low energies (dashed line) and for a composition changing according to Model III (solid line up to 10^6 GeV and its dotted continuation) and examine the resulting curve in figure 13 up to 10^{10} GeV.

It is immediately clear that a constant composition won't work! Model III fits the data up to 10^7 GeV quite well.

Now between 5 x 10^6 and 5 x 10^8 GeV X_m is increasing from about 450 g/cm^2 to 700 g/cm^2 an increase of 250 g/cm^2 in two decades of energy. Clearly something is changing rapidly.

One plausible explanation for this change is that in two decades the composition changes back to predominantly light by 5 x 10^8 GeV.

Violations of scaling in the fragmentation region would make $\alpha \neq 0$ and cause a reduction in the rate of rise of X_m. Cross sections would have to start decreasing rapidly to mimic the increase of X_m, in fact, $\lambda_\pi + \lambda_N$ must change by one hundred percent in two decades! Both of these methods to explain the change seem implausible.

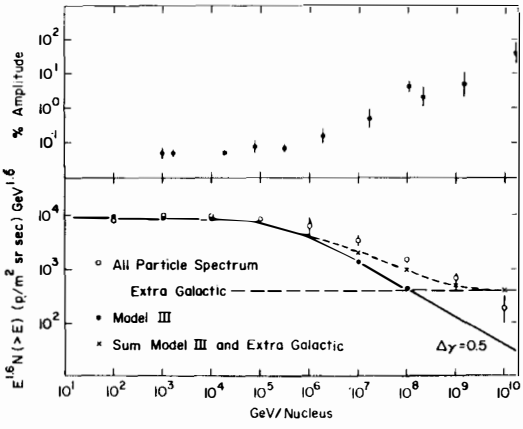


Figure 14: (a) Predictions of Model III compared with all particle composite spectrum of figure 3. The integral intensity has been multiplied by $E^{1.6}$ where E is energy per nucleus. Above 10^7 GeV Model III falls below the data indicating the need for a new component. An estimated extragalactic component is added to account for the highest energy cosmic rays. (b) Energy variation of anisotropy measurements. The increase in anisotropy from less than 0.1 percent below 10^5 GeV to several tens of percent by 10^{10} GeV is to be noted. The corrèlation is suggestive of the existence of an extra galactic component.

It is therefore reasonable to conclude, in a manner similar to that of Linsley and Watson, that chemical composition of cosmic rays changes from iron dominant at 10^6 GeV to proton dominant at 10^9 GeV. It is harder to infer particle interaction features because a combination of decreasing cross sections and scale violations in fragmentation region could occur simultaneously to mimic the rate of increase of X_m with energy above 5×10^8 GeV.

How does one combine a new component, mostly protons with Model III. The procedure used is shown in figure 14. Above energies where cutoff allows particles to leak out of the confinement region, the extragalactic component will become progressively more abundant with energy. Two decades above 10^5 GV this component must become comparable to the remaining galactic flux. This component is required to have a slope of 2.6, like all other components, and is able to account for the flux of air showers above 10^{10} GeV. The resultant spectrum is seen to agree with cosmic ray spectrum from 100 to 10^{10} GeV.

With this addition, energy variation of X_m is accurately reproduced as shown in figure 13 by the solid curve.

4. Discussion and Conclusions:

Detailed analysis of available data from air shower studies above 10^4 GeV strongly suggests that the elemental composition of cosmic rays is continuously changing from 10 GeV to 10^{10} GeV. The analysis relies on anchoring extrapolations of spectra at about 100 GeV/nucleon, requires consistency with all particle spectra, uses particle physics model which is a reasonable conservative extrapolation of accelerator data according to direct observations of interactions of cosmic ray hadrons around 10^4 GeV and with incorrelated hadron and muon fluxes at high energies. The analysis examines data on time structure of hadrons, multiple muons at high energies, comparisons of total muon and electron components in showers and variation of depth of shower maximum with shower size or energy, in terms of several models of primary cosmic ray spectra.

The following conclusions are obtained:

A predominantly light composition, similar to that at lower energies cannot account for the data. It is necessary to have at least 40% H component at 10^6 GeV. This can be achieved by several models, we consider three. (I) A bimodal model, wherein the 'light' and heavy components have two different spectral indices; (II) a continuously changing spectral index model with H having the flattest spectra; and (III) a model with constant spectral indices but with a rigidity dependent cut off.

The first two models must be modified above 10^6 GeV to describe the bend in all particle spectrum. The third model has the change in slope built in at the outset. In this model, however, the composition must remain predominantly

heavy above the bend. In Models I and II one could preferentially cut off the heavies with a mechanism that could leave protons and helium nuclei undisturbed.

The variation of X_m with energy below 10^6 GeV is consistent with all of the models. The striking increase in X_m between 10^7 and 10^9 GeV could be accommodated in models I and II by choosing the cut off mechanism to eliminate Hs' rapidly in two decades and recover proton dominance. If this is done for model I, where one must eliminate iron by changing its spectral index by one unit, the resulting composition at 10^{10} GeV is proton dominant and the flux is in reasonable agreement with air shower data.

If in model II we cut off the spectra in an A dependent way, such that the change in spectral index is greatest for iron and least for protons it could be possible to fit the flux and X_m curves above 10^8 GeV. Such a complex pattern of variation would require a rather specific astrophysical model which may be difficult to construct.

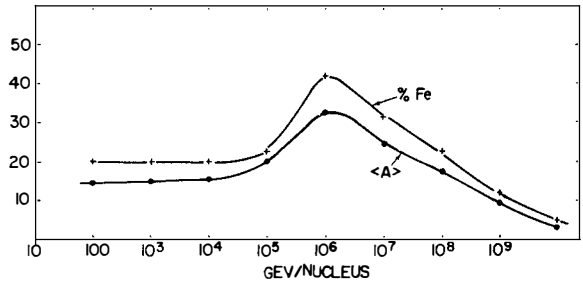


Figure 15: Energy variation effective atomic weight and percentage of iron component for primary cosmic rays.

In Model III, however, there is a natural mechanism to derive the required variation by assuming the existence of an extra-galactic (or an extra-confinement region) component of very high energy cosmic rays. (Ginzburg and Syrovatsky 1964, Peters 1960). This new component will become fully important at 10^7 GeV, about two decades above the cut off of low energy galactic component. The variation of effective atomic weight, A, with energy in this model is shown in figure 15. The agreement with energy spectra is shown in figure 14 and variation with X_m in figure 13. Model III, in fact is the "nested leaky box" model (Cowsik and Wilson 1973, 1975) in which cosmic ray sources are distributed within galaxies. The low energy variations are explained by energy dependent residence time within the sources. The cut off at 10^5 GV/nucleon rigidity is due to leakage out of our galaxy. The extra-galactic component is always present but becomes dominant at high energy only. Only one average source composition is needed to cover the full energy range.

We conclude, therefore, that composition of cosmic rays at high energies is varying with energy; from being light dominated below 10^4 GeV, to becoming heavy between 10^5 to 10^7 GeV and then again becoming light dominated at extremely high energies.

I want to thank Jordan Goodman, R. W. Ellsworth, J. Ormes, V. K. Balasubrahmanyan, R. Protheroe, T. K. Gaisser, R. Cowsik and John Linsley for many discussions and suggestions. This research was supported in part by the National Science Foundation.

References

Acharya, B. S., et. al. 1979. Proc. 16th ICRC, Kyoto, Japan, 13:272

Aguirre, C., et. al. 1973. Proc. 13th ICRC, Denver, U.S.A., 4:2598

Alexylev, E. N., et. al. 1979. Proc. 16th ICRC, Kyoto, Japan, 10:276

Antonov, R. A., et. al. 1979. Proc. 16th ICRC, Kyoto, Japan, 9:258

Arens, J. F., et. al. 1979. Space Science Instr., 4:303

Balasubrahmanyan, V. K. and Ormes, J. F. 1973. Ap. J., 186:109

Balasubrahmanyan, V. K., et. al. 1980. Astr. Sp. Sci., 71:135

Busza, W., et. al. 1975. Phys. Rev. Letters 34:383

Caldwell, J. H., 1977. Ap. J. 218:269

Cesarsky, C. J., 1980. Ann. Rev. of Astronomy and Astrophysics. 18, 289

Chudakov, A. E., et. al. 1979a. Proc. 16th ICRC, Kyoto, Japan, 10:192

Chudakov, A. E., et. al. 1979b. Proc. 16th ICRC, Kyoto, Japan 10:188

Cowsik, R., 1968. Canadian Journal of Physics, 46:S142

Cowsik, R., and Wilson, L., 1973. Proc. 13th ICRC, 1, 500

Cowsik, R., and Wilson, L., 1975. Proc. 14th ICRC, 2, 659

Deakyne, M., et. al. 1978. Proc. of Bartol Conf. on Cosmic Rays and Particle Physics, Ed. T. K. Gaisser, AIP Conf. Vol. 49:48

Efimov, N. N., et. al. 1973. Proc. 13th ICRC, Denver, U.S.A., 4:2378

Elbert, J. W., et. al. 1975a. Phys. Rev. D12, 660

Elbert, J. W., et. al. 1975b. J. Phys. G8, L 13

Elbert, J. W., et. al. 1976. J. Phys. G2, 971

Elbert, J. W. 1978. Proc. Dumand Summer Workshop, La Jolla Ed. A. Roberts, Vol. 2, 101

Elbert, J. W., et. al. 1979. Proc. 16th ICRC, Kyoto, Japan, 10:405

Elbert, J. W., Gaisser, T. K. and Stanev, T., 1981, Private Communication

Ellsworth, R. W., et. al. 1977. Astr. Sp. Sci. 52:415

Ellsworth, R. W., et. al. 1978. Proc. of Bartol Conf. in Cosmic Rays and Particle Physics, Ed. T. K. Gaisser, Am. Inst. of Phys. 49:111

Ellsworth, R. W., et. al. 1981a. Phys. Rev. D23:764

Ellsworth, R. W., et. al. 1981b. Phys. Rev. D23:771

Ellsworth, R. W., et. al. 1981c. Nucl. Phys. B., to be published

Freier, P. and Waddington, C. J., 1975. Astr. and Sp. Sci. 38:419

Fujimoto, Y. and Hasegawa, S., 1978. Cosmic Rays and Particle Physics, 1978. Proc. of Bartol Conf., Ed. by T. K. Gaisser (AIP, New York, 1978) p. 137

Gaisser, T. K., et. al. 1978. Rev. Mod. Phys. 50:859

Gaisser, T. K., et. al. 1979. Proc. 16th ICRC, Kyoto, Japan, 9:275

Gaisser, T. K. and Yodh, G. B., 1980. Ann. Rev. Nucl. Part. Sci. 30:475

Ginzburg, V. L. and Syrovatski, S. I., 1964. The Origin of Cosmic Rays, Oxford: Pergamon

Grigorov, N. L., et. al. 1971. Proc. 12th ICRC, Hobart, Tasmania, 5:1946

Goodman, J. A., et. al. 1979. Phys. Rev. Letters 42:854

Halzen, F. and Luthe, J., 1978. Phys. Lett. 48B:440

Hammond, R. T., et. al. 1978. Nuovo Cim. 1C.N.4:315

Hillas, A. M., 1975. Phys. Rep. 20 C:59

Hillas, A. M., 1978. Cosmic Rays and Particle Physics-1978, Proc. of Bartol Conf., Ed. T. K. Gaisser. (AIP, New York, 1978):373

Hillas, A. M., 1979. Proc. 16th ICRC, Kyoto, Japan, 8:7

JACEE Exp. Capdevielle, J. N., et. al. 1979. Proc. 16th ICRC, Kyoto, Japan, 6:324

Jones, W. V., 1978. Proc. of 1978 Dumand Summer Workshop, Ed. A. Roberts, (Fermilab, Batavia, Ill. 1979)1:313

Julisson, E. 1974. Ap. J. 191:331

Kalmykov, N. N., et. al. 1975. Proc. 14th ICRC, Munich, W. Germany, 8:2861

Kalmykov, N. N., et. al. 1979. Proc. 16th ICRC, Kyoto, Japan, 9:73

Lattes, C. M. G., et. al. 1973. Proc. 13th ICRC, Denver, U.S.A., 3:2227 and 4:2671

Lattes, C. M. G., et. al. 1980. Phys. Rep. 65, 3

Lezniak, J. A. and Webber, W. A. 1980. Ap. J. 233:676

Lingenfelter, R. E. 1979. Proc. 16th ICRC, Kyoto, Japan, 14:135

Linsley, J. 1977a. Proc. 15th ICRC, Plovdiv, Bulgaria (Bulgarian Academy of Sciences, Plovdiv, Bulgaria, 1977) 12:89

Linsley, J. and Watson, A. A., 1977, ibid. 12:203

Linsley, J. 1980. Invited talk of Very High Energy Cosmic Rays, to be published in Proc. of IUPAP/IAU Symposium No. 94 on "Origin of Cosmic Rays," Bologna, Italy, June 1980.

Linsley, J. and Watson, A. A. 1981, Phys. Rev. Letters 46:459

Lowe, G. H., et. al. 1976. Phys. Rev. D 13:2925

Mason, G. W., et. al. 1975. Proc. 14th ICRC, Munich, W. Germany. 8:2943

Olejniczak, J., et. al. 1977. J. Phys. G3,847

Ormes, J. F. and Freier, P. S., 1978. Ap. J. 222:471

Orth, C. D. et. al. 1978. Ap. J. 226:1147

Oxford, K. J. and Turver, K. E., 1976. Nature 264:727

Peters, B. 1961. Nuovo Cimento 22:800

Protheroe, R. J. and Turver, K. E. 1979. Nuovo Cimento 51A:277

Ryan, M. J. et. al. 1972. Phys. Rev. Letters 28:985

Schmidt, W. K. H. et. al. 1976. Astro. Ap. 46:49

Simon, M. et. al. 1980. Ap. J. 239:712

Sreekantan, B. V. 1979. Proc. 16th ICRC, Kyoto, Japan. 14:345

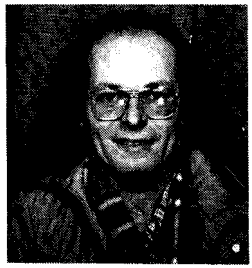
Smith, L. H., et. al. 1973. Ap. J. 180:987

Thornton, G. and Clay, R. 1980. Phys. Rev. Letters 44:959 and Phys. Rev. Letters 45, 1463 (E), 1980

Wdowczyk, J. and Wolfendale, A. W. 1973. J. Phys. A 6:1594

COSMIC RAY ANISOTROPY : 10^{12} - 10^{20}eV

A.A. Watson
Department of Physics, University of Leeds,
Leeds LS2 9JT, England.



The results of experiments designed to study the arrival direction distribution of cosmic rays of energy 10^{12} - 10^{20}eV are reviewed. It is shown that at all energies there is evidence for anisotropy, the amplitude of which ranges from 0.075% at the lowest energies to 90 ± 20% above 4.10^{19}eV. The increase of anisotropy with energy is not smooth, showing features which occur at energies similar to those at which features are observed in the cosmic ray energy spectrum. At least up to 2.10^{17}eV it seems probable that the acceleration sites lie within our Galaxy, and it is hard to escape the conclusion that particles of energy $>10^{19}$eV are extragalactic. Sources of the highest energy particles ($\sim 10^{20}$eV) must lie within 200Mpc, and considerably closer if, as seems likely, the intergalactic medium is such as to prevent rectilinear propagation. Between 2.10^{17} and 10^{19}eV the location of the sources is less certain. The aim of future arrival direction experiments should be to study anisotropy as a function of primary mass composition.

1. Introduction

Seventy years after the discovery of cosmic rays, the question of their origin remains unsolved. In the early days of investigation of the properties of the radiation, experimenters hoped to find that some regions of the sky were brighter than others when viewed in cosmic ray light. It is now known that below about 10^{11}eV cosmic ray intensities are strongly influenced by the spatial and temporal changes in the magnetic fields of interplanetary space so that the Galactic anisotropy cannot be determined from the earth. At higher energies, where solar effects can be accounted for or neglected, the amplitude of anisotropy is found to be small, being less than 0.1% up to 10^{14}eV. The high degree of isotropy of low energy cosmic rays was recognized early on, and this fact, coupled with the large density of cosmic rays (0.5eVcm^{-3} at 10^{9}eV and 0.1eVcm^{-3} at 10^{11}eV), was a significant clue in early thinking about the possibility that our Galaxy might contain a weak but extensive magnetic field. It is of course the Galactic magnetic field, at least at the lowest energies, that thwarts attempts to locate cosmic ray sources in a manner analgous to that used to find optical or X-ray sources: the 'seeing' is spoiled for cosmic ray telescopes because the Larmor radius of charged particles is very small by comparison with Galactic dimensions except at the highest energies. At 10^{14}eV a proton has a radius of gyration of only 0.04pc in a 3μG field; at 10^{18}eV the corresponding radius is 370pc, rather larger than the thickness of the Galactic disc.

In this paper I will outline the detection techniques used to measure cosmic ray anisotropies, describe the results available and discuss some tentative interpretations. I will show that features in the variation of the character of the anisotropy with energy are linked to features in the primary energy spectrum. A full understanding of the meaning of these linkages is not yet clear because information about cosmic ray mass composition, particularly above $\sim 10^{14}$eV, is still rudimentary.

2. Primary Energy Spectrum and Mass Composition

As a background to our discussion of cosmic ray anisotropy it will be helpful to keep in mind our present knowledge of the cosmic ray energy spectrum and mass composition. Figure 1 shows a schematic version of the energy spectrum. The integral intensity has been multiplied by $E^{1.5}$ (E is the primary energy in eV) and is plotted on the y-axis. This procedure serves to show clearly the two prominent features in the spectrum, namely the 'knee', near 10^{15}eV, and the 'ankle' above 10^{19}eV. Both these features are well established although the flattening observed above 10^{19}eV has been observed only from the Northern Hemisphere.[1] Integral spectrum slopes are -1.65,

between 10^{11} and 10^{15}eV, -2.1 between 10^{17} and 10^{19}eV and -1.4 ± 0.1 above 10^{19}eV. Between 10^{15} and 10^{17}eV there is evidence[2] that the spectral slope is steeper than at higher energies but the experimental situation is not yet clear. It should be borne in mind that, because the observed particle densities are greater and because the shower maximum is relatively lower in the atmosphere, the properties of primaries above 10^{17}eV (and at least up to 10^{19}eV) are somewhat better established than the properties of lower energy primaries which produce smaller showers.

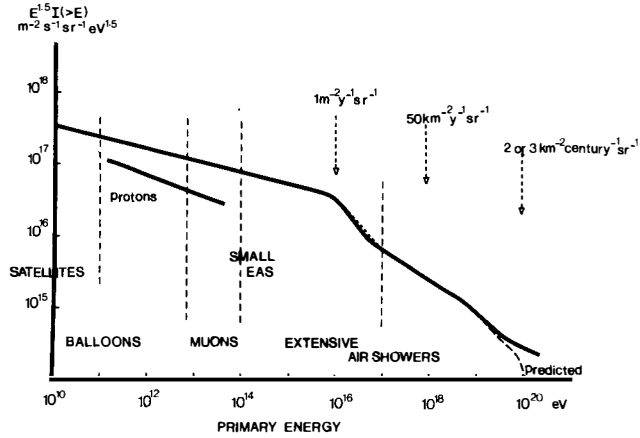


Figure 1 The cosmic ray energy spectrum above 10^{10}eV. The detection techniques used in various energy regions are indicated. The dashed curve near 10^{16}eV marks an energy range in which the spectral shape is uncertain. The curve above 10^{19}eV marked 'predicted' is the calculation[6] of the Greisen-Zatsepin effect[4,5].

Explanations for the two major features are not yet agreed. About 20 years ago Peters proposed that the 'knee' at 10^{15}eV might reflect the inability of our Galaxy to retain cosmic rays protons of this energy (Larmor radius $\sim$0.4pc). A prediction of this hypothesis was that heavy nuclei would begin to dominate the primary beam above 10^{15}eV. No convincing evidence has been obtained to support this prediction and counter evidence has been offered that Fe-nuclei are becoming dominant at energies less than 10^{15}eV.[3] Hillas [2] has reviewed the situation and concluded that composition and spectrum data cannot be reconciled with the Galactic leakage model. He proposes that the change is induced in the spectrum by photonuclear or pair production reactions occuring

near to the source. The spectral feature at 10^{15}eV would thus reflect the radiation field in the acceleration region.

The 'ankle' feature above 10^{19}eV is even more puzzling. Following the discovery of the 2.7K black body radiation in 1965 it was predicted by Greisen[4] and by Zatsepin and Kuzmin[5] that if ultra high energy cosmic rays were produced only in sources which were at cosmological distances (e.g. quasars) then the intensity of cosmic rays above $\sim 5 \times 10^{19}$eV would fall so dramatically that no events of 10^{20}eV would be observed with the giant air shower arrays then being brought into operation. Reality is different. The three Northern Hemisphere installations (Volcano Ranch (U.S.A.), Haverah Park (U.K.) and Yakutsk (U.S.S.R.)) all find that the spectrum is flatter above 10^{19}eV than at lower energies. The total exposure achieved at the highest energies in these experiments is $\sim 300\,\mathrm{km^2\,yr\,sr}$ and at least 7 events with energies above 10^{20}eV have been observed. Using the theoretical predictions of Strong et al[6] about 0.2 events would have been expected. There is a consensus that this result implies that the age of the cosmic rays above about 3×10^{19}eV is less than 10^8 yr (e.g. Puget et al[7]). The arrival direction distribution of these particles, which will be discussed in detail below, is not that expected if the most energetic particles were produced within our Galaxy.

For the purposes of interpreting arrival direction information details about the mass composition of the primary cosmic rays are crucial. At energies less than 10^{13}eV the arrival direction distributions observed refer mainly to protons as the experiments which have been possible are selective in their triggering and respond dominantly to muons produced by proton primaries. At higher energies the only method of getting information about primary cosmic rays is by studying features of the extensive air showers which they produce (electrons, muons, air-Čerenkov light, air-scintillation light etc). The problem of extracting the mass composition is particularly difficult because the particles cannot be observed directly and because the necessary nuclear physics must be extrapolated from lower energies. Recently progress has been made by studying the change of the depth in the atmosphere at which the number of electrons in the shower reaches its maximum. The variation of this depth with energy (the elongation rate) for a fixed primary composition depends principally upon the multiplicity of the pions produced in hadronic interactions.[8] If it is assumed that p-p cross-sections continue to rise at the rate observed at accelerator energies then it appears that the primary particles are very light ($\overline{\ln A} = 0^{+0.6}_{-0}$) above 3×10^{16}eV and rather heavy ($\overline{\ln A} = 4 \pm 2$) near 10^{15}eV.[9] The rapid change of the depth of maximum with energy between 10^{15}eV and 3.10^{16}eV expected on such a picture is indeed

observed.[10] A detailed summary of mass composition determinations is given elsewhere in these Proceedings by Yodh. A review of properties of the highest energy cosmic rays has been given by Linsley.[11]

3. Arrival Direction Data

Kirally et al [12] have reviewed arrival direction data at all energies while more recently Elliot [13] has discussed the measurements available at energies below 10^{14}eV. The present experimental situation is summarized in figure 2. What is plotted (figure 2(a)) is the amplitude of the 1st harmonic of the cosmic ray anisotropy as a function of energy. At energies less than 6.10^{16}eV the first harmonic is computed in sidereal time and is an average over the range of declinations (typically $\pm 30^0$) which lie within the reception cone of the detector system. At the higher energies, where information on the energy and direction of individual events is available, the first harmonic is measured in right ascension. The data shown refer to all declinations above -6^0; analyses in declination are discussed briefly below. The directions of maximum amplitude (the phase) are shown in figure 2(b).

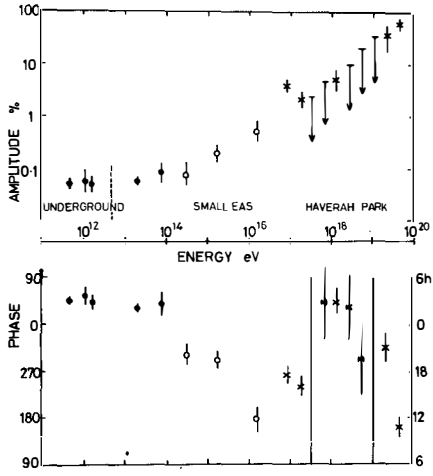


Figure 2 Summary of cosmic ray anisotropy data, after(11). The top half of the diagram (figure 2a) shows the amplitudes measured in various experiments. The filled circles are from references (14-18) as reviewed in (13). The open circles are from the compilation of Linsley and Watson[19] while the crosses are from the Haverah Park experiment[20,21]. The lower half of the diagram (figure 2b) shows the time of maximum amplitude (the phase). The error bars correspond to ± 1 standard deviations; the upper limits are shown at the 95% confidence level.

The three points below 10^{13}eV are from measurements made in underground laboratories[14,15,16] at depths between 40 and 507 m.w.e. where muons from the decay of pions produced high in the atmosphere by the incoming primaries are detected. The energy refers to the energy/nucleon of the primaries which are dominantly protons. The next two points (between 10^{13} and 10^{14}eV) are from the work of the Japanese[17] and Hungarian[18] groups who have measured the time variation of the rates of small air showers produced at mountain altitudes. Above 5.10^{11}eV the trajectories of the primaries are relatively little influenced by the interplanetary magnetic field (or the effect has been accurately allowed for). The constancy of the amplitudes and directions between 5.10^{11} and 10^{14}eV is striking, and must be considered as well established.

Kirally et al[12] in a clear discussion of these data, have linked the invariance of phase and amplitude to the persistence of the direction of the local interstellar magnetic field out to distances of a few parsecs, the field direction being inferred from the steady flow of interstellar gas in the region just outside the solar cavity. Kirally et al also deduce that under these circumstances features of cosmic ray propagation do not change very much between 10^{11} and 10^{14}eV.

Above 10^{14}eV all measurements are derived from studies of air showers. Between 10^{14} - 10^{17}eV (open circles) the data have been taken from a compilation[19] of all available evidence recorded between 1951 and 1965. The results are based on 23 experiments in which the counting rate of small air showers was recorded as a function of time. The directional accuracy possible in these early experiments was rather poor as it relied upon atmospheric collimation. Likewise the energy resolution was limited as analysis of individual events was not possible. Nonetheless there is evidence for an anisotropy which changes both in amplitude and phase as the energy increases The validity of this claim is strengthened by the fact that the amplitude of the first of the older air shower points, at 3.10^{14}eV, is $0.075 \pm 0.020\%$, in excellent agreement with the mean amplitude of the 5 independent measurements made with EAS and muon techniques in more recent experiments at lower energies. The difference between the phase of the first of the older points and that of the lower energy data is not particularly disturbing as the Larmor radius of a proton of this energy is $\sim$0.1pc in the canonical field, and variations in magnetic field and/or cosmic ray gradient on this scale may well occur in directions away from the Galactic plane. As the energy increases the amplitude of anisotropy starts to rise and there is a steady change in the phase of maximum.

Above about 6.10^{16}eV the data shown in figure 2 derive from the Haverah Park experiment[20,21]. At the present time these data are the most numerous available; between 10^{17} and 10^{18}eV the number of showers recorded exceeds those from other experiments by roughly an order of magnitude; between 10^{18} and 10^{19}eV the Haverah Park data set exceeds that of Volcano Ranch[22] by more than a factor of 5 while above 10^{19}eV the events available from Northern Hemisphere experiments are Yakutsk[23] (34), Volcano Ranch[24] (44), Haverah Park[24] (144). The discussion of data above 6×10^{16}eV which follows concentrates mainly on the Haverah Park results but where comparison has been possible the agreement of these results with the broad features of other experiments is found to be good (see reference (20) for more details). In the Haverah Park experiment the direction of each event is known to within 10^{-2} sr and the energy resolution is better than 40%. Partition of the events into energy bins a factor of 2 wide has been adopted.

The data from the Yakutsk, Volcano Ranch and Haverah Park cover mainly the sky region above $\delta = 0^0$. There are limited statistics from Haverah Park down $\delta = -6^0$ and from Volcano Ranch down to $\delta = -30^0$. A major experiment has been operated in the Southern Hemisphere by the University of Sydney[25]. Analysis is still in progress and the final results are keenly awaited as the exposure achieved in this single experiment was comparable to that from the three Northern Hemisphere arrays combined. Particularly at the highest energies the absence of data from the Southern Hemisphere severely hampers interpretation of the measurements which are available.

The variation of anisotropy with energy above 10^{16}eV which is summarised in figure 2 is extremely complex. Near 10^{17}eV there appear to be two significant amplitudes, in adjacent energy bins separated by a mean energy of a factor 2. The direction of maximum is similar in both bins when the data are examined over all declinations, but differences in detail are revealed when individual declination strips are examined. In the lower energy bin (E1, 6×10^{16} - 1.25×10^{17}eV) there is a slight excess near 240^0RA in each bin, whereas bin E2 (1.25 - 2.5×10^{17}eV) shows an excess dominantly in the strip with $20^0 < \delta < 30^0$ (figure 3). In the search for anisotropy[20] 100 such declination strips were examined; the chance of finding one such remarkably anisotropic strip is computed, taking into account the number of strips, as 8×10^{-3}. No data from other experiments are available to check the reality of this result but figure 4 shows the phases and amplitudes from all other experiments in which measurements were made near 10^{17}eV. The agreement with the Haverah Park results, averaged over all declinations, is impressive.

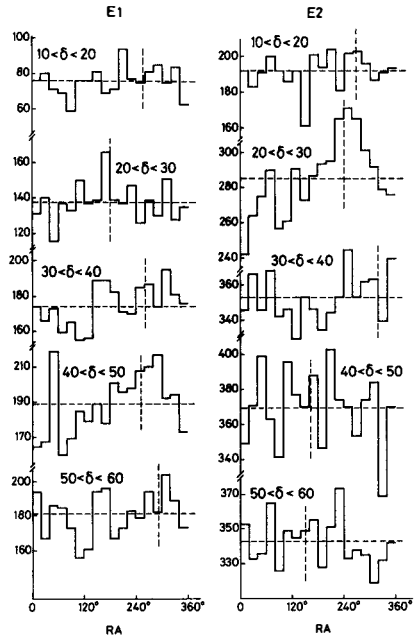


Figure 3 Right ascension distributions in 5 declination strips for energy bins E1 (6×10^{16} - 1.25×10^{17} eV) and E2 (1.25 - 2.5×10^{17}eV) of Haverah Park data[21]. The dashed horizontal line in each histogram is the mean number for that interval; the vertical dashed line shows the phase of the first harmonic in right ascension.

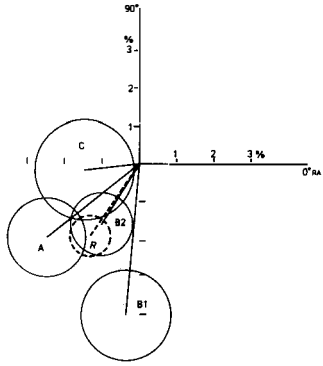


Figure 4 Comparison of data from a number of experiments near 10^{17}eV. The circles indicate rms amplitudes. A (Cornell [26]), 19000 events, 7.5×10^{16}eV); B (Haverah Park [20]) (1, 13825 events 6×10^{16} - 1.25×10^{17}eV; 2, 31266 events, 1.25 - 2.5×10^{17}eV)); C (other experiments, 11059 events, see [20] for details, 6×10^{16}-5×10^{17}eV). R is the resultant vector.

In the two energy bins immediately above those just discussed only upper limits can be set to an anisotropy. The limits shown (figure 2) are 95% confidence limits and it therefore appears that the amplitude of anisotropy may not increase as rapidly with energy beyond 2.10^{17}eV as it appears to do between 10^{15} and 10^{17}eV. Note that the phase of bin E3 ($2.5 - 5 \times 10^{17}$eV) is unknown. By contrast bin E4, again showing only an upper limit, has a phase which can be estimated and is similar to that in the 2×10^{18}eV region (bins E5 and E6) where there is limited evidence for an anisotropy in right ascension. This result is supported (see figure 5) by experimental evidence which is available from other work[22,26-29] carried out near 10^{18}eV.

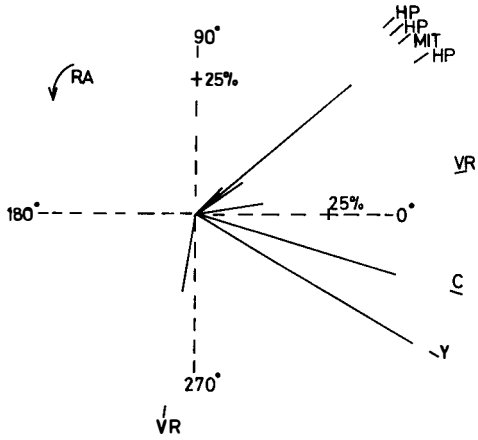


Figure 5 Comparison of data from a number of experiments near 10^{18}eV. C(Cornell[33]), MIT(MIT group[27]), HP(Haverah Park[28]), VR(Volcano Ranch[22]), Y(Yakutsk[29]). The variation in amplitude between experiments is large but the agreement in phase is striking.

At the very highest energies (>4.10^{19}eV) there is accumulating evidence for an excess of events from high Galactic latitudes. Such a possibility was first noted in 1973[30] and as further events have been recorded the tendency for energetic events to arrive preferentially at large angles to the Galactic plane has persisted. The current situation (for $\delta > 0^0$ and E>4.10^{19}eV) is summarised in figure 6; the Volcano Ranch and Haverah Park events are those listed in the World Data Centre Catalogue[24] while the Yakutsk points were reported by Krasilnikov[23]. The cut in energy has been taken at 4.10^{19}eV so as to correspond with the highest energy band discussed by Krasilnikov.

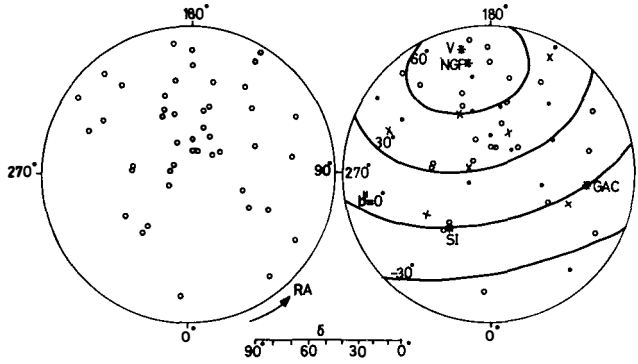


Figure 6 The arrival directions of the 45 most energetic events ($>4 \times 10^{19}$eV) observed at $\delta > 0^0$. The projection used is an equal area projection; $\delta = 90^0$ lies at the centre of the circle which has circumference, $\delta = 0^0$. The open circles in the left hand diagram are identifiable in the right hand diagram (13 filled circles, Yakutsk[23]; 7 crosses, Volcano Ranch[24] and 25 open circles, Haverah Park[24]). In the right hand diagram lines of Galactic latitude, $b'' = 60^0$, 30^0, 0^0, -30^0, are marked as are the directions of spiral-in, the Galactic anti-centre, the north Galactic pole and the centre of the Virgo cluster.

Individual energies have yet to be published for the Russian data. The first harmonic in right ascension of these 45 events has an amplitude of $90 \pm 20\%$ (probability of arising by chance from an isotropic distribution = 1.4×10^{-4}) and a phase of 180 ± 14^0RA. It is also clear that there is no enhancement towards the Galactic plane as predicted by Syrovatsky[31]. The probability of a 2:1 enhancement in the region of the plane with respect to the Pole ($|b''| < 30^0$ compare with $|b''| > 30^0$) can be strongly rejected, ($p \ll 10^{-3}$).

An alternative way of presenting this result is to determine the deviation from expectation of the mean Galactic latitude of these data. It has been pointed out before[1] that the Haverah Park data set reveal an increase of $\langle \sin b'' \rangle$, over the value expected for isotropy, above the energy at which the energy spectrum flattens. A similar analysis has been made with the available showers from Volcano Ranch and Yakutsk and is shown alongside the Haverah Park result in figure 7. There is some support from these experiments for a change in the Galactic latitude distribution at the energy in question. Above 4.10^{19}eV, $\langle \sin b'' \rangle$ for 48 events is 0.50 ± 0.07 compared with the isotropic expectation of 0.30. For $1\text{-}2 \times 10^{19}$eV (the other energy band for which Yakutsk directions are available) $\langle \sin b'' \rangle = 0.210 \pm 0.045$ (isotropic expectation = 0.22) for the showers recorded in the three

experiments.

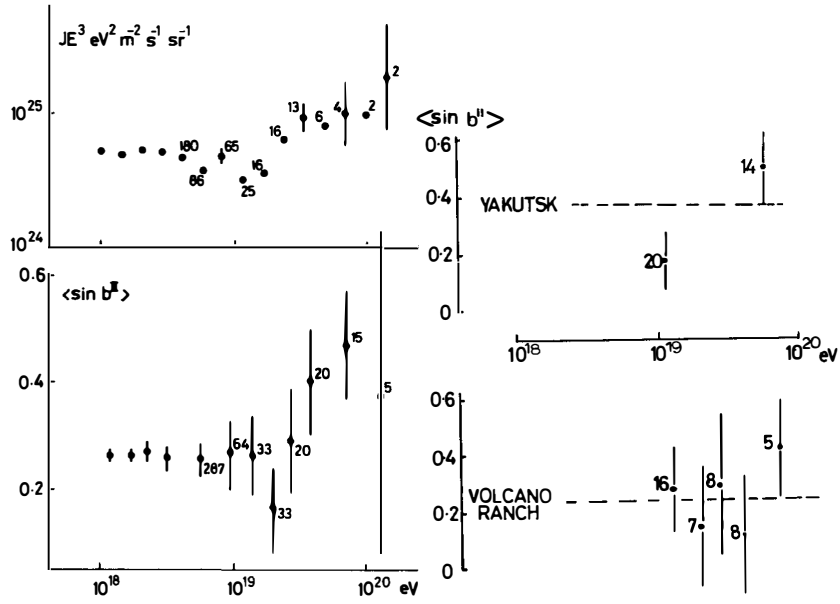


Figure 7 Variation of <sin b''> with energy. The left hand diagram shows the Haverah Park results[1] for <sin b''> and the energy spectrum. The right hand diagram shows the results derived for <sin b''> for the Volcano Ranch[24] and the Yakutsk[23] data sets.

The reason, 'a priori', for investigating the anisotropy above and below 4×10^{19}eV is that above this energy the spectrum is flatter than at lower energies whereas if the sources of these multi-Joule particles are at cosmological distances (quasars for example), the spectrum is predicted to steepen rapidly. I have been unable to think of any systematic error which, if it were present, would simultaneously flatten the spectrum and cause the showers to arrive with an anomalous intensity at high Galactic latitudes. The two effects combine to provide an important clue as to the origin of the most energetic particles.

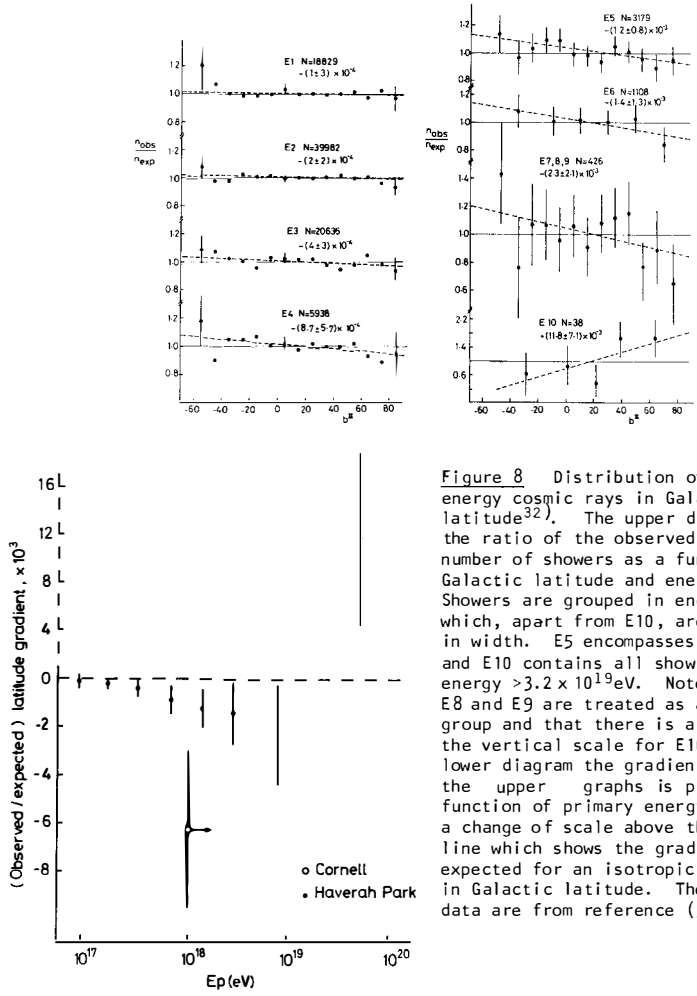


Figure 8 Distribution of high energy cosmic rays in Galactic latitude[32]. The upper diagram shows the ratio of the observed to expected number of showers as a function of Galactic latitude and energy. Showers are grouped in energy bins which, apart from E10, are a factor 2 in width. E5 encompasses $1\text{-}2 \times 10^{18}$eV and E10 contains all showers with energy $>3.2 \times 10^{19}$eV. Note that E7, E8 and E9 are treated as a single group and that there is a change in the vertical scale for E10. In the lower diagram the gradient of each of the upper graphs is plotted as a function of primary energy. There is a change of scale above the dotted line which shows the gradient expected for an isotropic distribution in Galactic latitude. The Cornell data are from reference (33).

A more detailed analysis in Galactic latitude has recently been made by the Haverah Park group[32] for showers with energies below 4.10^{19}eV. The 9×10^4 events with energies above 6×10^{16}eV have been grouped into factor of 2 bins in energy. For each energy bin the ratios of the observed to the expected number of events have been calculated as a function of Galactic latitude and are shown in figure 8. The statistics allow only a description of the systematic deviations from uniformity in Galactic latitude. The gradient of a linear least

squares fit to the data of each energy bin is shown in the figure. The result for bin E10 is a restatement of the result just discussed, that showers above 4×10^{19}eV arrive preferentially from high Galactic latitudes. The remaining energy bins show a systematic enhancement at low Galactic latitudes. Although no energy bin shows a very significant enhancement on its own, a trend of steeper gradients at higher energies is apparent when the fitted gradients are plotted as a function of energy. Direct comparison is possible only with data from the Cornell group[33] who have reported the co-ordinates of 166 events having $E>10^{18}$eV. The gradient is comparable in magnitude and of the same sign as those obtained with Haverah Park data. For the Haverah Park events alone the mean gradient at energies less than 3.2×10^{19}eV is $(-3.1 \pm 1.4) \times 10^{-3}$, significantly different from zero.

4. Discussion and interpretation

I have attempted to summarise the present state of knowledge about cosmic rays in figure 9 where the energy dependence of the mean mass, the integral slope of the energy spectrum and parameters associated with the anisotropy, are displayed. The breaks in the various parameters are not, of course, as sharp as indicated but are shown as discontinuities to highlight the possibility that changes in spectral features appear to coincide with changes in features of the

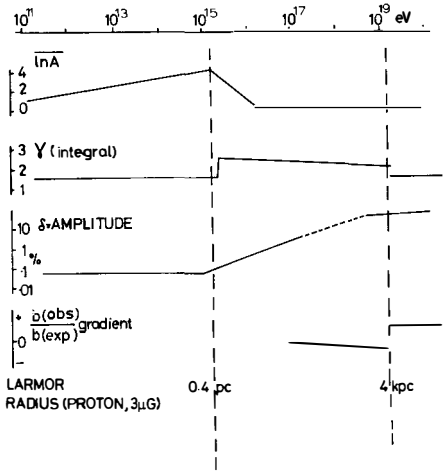


Figure 9 Summary of experimental data on cosmic rays. $\overline{\ln A}$ is the mean of the logarithm of the primary mass, A; γ is the slope of the integral energy spectrum; δ is the amplitude of the 1st harmonic in sidereal time or right ascension and the b'' gradient ratio is a summary of the data of figure 8. The Larmor radius of a proton in a 3μG magnetic field is shown at two energies.

anisotropy. The two most clearly established changes in the slope of the energy spectrum occur near 10^{15}eV and 3×10^{19}eV. Near the lower energy the

amplitude of the first harmonic in right ascension is observed to start to increase with energy; at the higher energy a tendency for particles to arrive preferentially from high Galactic latitudes is observed. There may be a less pronounced spectral feature near 10^{17}eV where the anisotropy amplitude changes and where the phase and declination distribution behaves in a complicated manner; more measurements of the spectrum are required in this region.

It is evident from figure 9 (and figures 2 and 8) that cosmic rays are anisotropic at all energies between 5.10^{11} and 10^{20}eV. There is suggestive evidence, derived from analysis of the distribution of 100MeV γ-rays, that cosmic rays of 10^{9} - 10^{10}eV have their origin in our Galaxy[34]. It is conservative and unexceptionable to suppose that cosmic rays up to at least 2.10^{17}eV are also of Galactic origin. The relatively large amplitudes, several percent, observed near 10^{17}eV and the complex distribution in declination (see figure 3) could hardly arise as a result of some anisotropic, extragalactic, distribution of sources because diffusion of the particles in intergalactic space is expected to render their distribution highly isotropic. Rather one might seek to interpret the observed patterns as arising from the convolution of the magnetic field structure within a few Larmor radii of the earth with an anisotropic distribution of Galactic objects which would accelerate such particles. In this context it is noted that the currently attractive shock acceleration model might operate up to energies of 10^{18}eV[35]. Suitable shocks might be, or might have been, associated with features such as the Heiles supershells[36]. Theoretical studies of methods of acceleration of light nuclei in such systems would be of enormous help in interpreting the data.

Extensive efforts have been made in the last 10 years to understand the fact that cosmic rays of 10^{20}eV are observed at a rate $(5(^{+2.4}_{-4.2})\times10^{-16}\mathrm{m}^{-2}\mathrm{s}^{-1}\mathrm{sr}^{-1})$ significantly greater than expected if the sources of such particles lie only at cosmological distance. The point is that protons of this energy interact very strongly with the 3K radiation (producing pions) and have an attenuation length of about 200Mpc at 10^{20}eV. Thus 10^{20}eV protons from the Coma cluster would be reduced in intensity by $\sim$2.7 even if diffusion in intergalactic space was negligible. If all of the primaries were iron nuclei (in contradiction to the limited experimental evidence) photodisintegration by the 3K photons would be effective in depleting the intensity by a similar amount[7]. The sources must thus be relatively local, certainly lying within 200Mpc.

Evidence for anisotropy in the arrival directions of the highest energy events (figure 7) appears at variance with any reasonable model of Galactic origin, and it has become popular[6,7,37-39] to consider the possibility that highest energy particles are produced by sources within the Virgo cluster. A

major drawback of this hypothesis is that the energy requirements for such sources are very large. To maintain the observed local energy density of $\sim 10^{-8}$eV cm^{-3} throughout the Local Supercluster when the 3K-limited lifetime is $\sim 10^{8}$y requires an energy input, into 10^{20}eV particles and above, of $\sim 5 \times 10^{41}$erg s^{-1}. Additionally the sources would have to pump vast quantities of energy into lower energy cosmic rays, as much as 10^{46}erg s^{-1} into cosmic rays >1GeV if the high energy spectral shape ($\gamma = 1.3 \pm 0.1$) persists to lower energies. Such an energy output is about 10^{6} times that of our own Galaxy in cosmic rays whereas the number of similar galaxies in the Virgo cluster may only be $\sim 10^{3}$.

A particularly interesting and detailed discussion of this problem has been given by Giler et al[40] in a development of some earlier ideas of Wdowczyk and Wolfendale[41]. They have made a detailed study of a model in which the sources of the particles lie in the Virgo cluster and in neighbouring clusters. Two diagrams from their paper are shown in figure 10. The spectral shape at the highest energies can be reproduced by the multi-cluster model with a large diffusion coefficient ($3 \times 10^{35}\ E_{19}^{1.5}$ cm^2 s^{-1}) though a better description of the anisotropy data is provided by the model in which the Virgo cluster contains the most significant sources. A test between the two models may be possible when

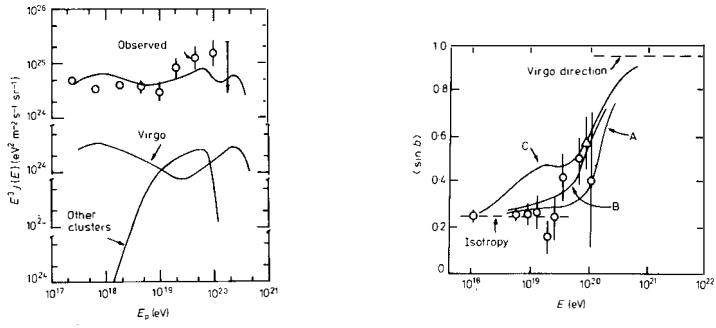


Figure 10 Results of calculations of the Virgo cluster and multi-cluster models of Giler et al[40]. The left hand diagram shows the fit to the spectral shape achieved with the multi-cluster model (clusters within 100Mpc) in which the diffusion coefficient is energy dependent, $D = 3 \times 10^{35}\ E_{19}^{1.5}$ cm^2 s^{-1} (E in units of 10^{19}eV). The right hand diagram shows the variation of <sin b''> predicted by models: A, Virgo cluster only, $D = 5 \times 10^{33}\ E_{19}^{\frac{1}{2}}$ cm^2 s^{-1}; B, Virgo cluster alone, $D = 1.58 \times 10^{34}\ E_{19}$ cm^2 s^{-1}; C, Virgo plus other clusters, $D = 3 \times 10^{35}\ E_{19}^{1.5}$ cm^2 s^{-1}. The open circles are from[1], the triangle is a smaller sample of Yakutsk data than that of[23] which is shown in figure 7. The diagrams are figures 7 and 9 of reference (40).

data on the spectra and anisotropy in the Southern Hemisphere ($\delta<0^0$) become available. On the Virgo hypothesis it seems clear that a smaller anisotropy than seen in the Northern Hemisphere would be expected together with a less pronounced flattening of the spectrum. This arises because of the well-known concentration of galaxies in the northern Galactic hemisphere[42]. Predictions based on the multi-cluster model require more detailed knowledge of the positions, sizes and distances of Galactic clusters than are presently available.

Although the model just described successfully explains features associated with the highest energy cosmic rays it predicts that the direction of anisotropy of 10^{18} - 10^{19}eV particles should lie above the Galactic plane (figure 10) rather than below it as is found experimentally (figure 8). The origin model can be reconciled with observations if the primaries which reach us from southern Galactic latitudes have undergone some preferential acceleration, perhaps as a result of having spent a substantial part of their lives within the Local Group of galaxies, the centre of which lies in a direction 22^0 below the Galactic plane. The possibility of intergalactic acceleration, suggested originally by Cocconi[43], has been revived more recently in the context of shock acceleration by Blandford and Ostriker[35]. It is not being suggested here that particles receive all of their energy in intergalactic space. Rather it is being proposed that the anisotropy arises because particles receive a boost of energy, ΔE, leading to an anisotropy of $(\gamma + 1)\frac{\Delta E}{E}$, where γ is the slope of the integral energy spectrum. Galactic winds and the motion of massive gas clouds, such as those of the Magellanic stream[44], may generate the necessary shocks in the medium of the Local Group.

If a heavy mass composition is eventually established as being more probable than the proton dominated composition supposed in the above discussion then the observations between 10^{18} and 10^{19}eV could be reconciled with a Galactic origin. A cosmic ray gradient caused by inefficient trapping within the Galactic magnetic field would result in an anisotropy in a direction orthogonal to the local magnetic field and the gradient. The direction of the observed excess is consistent with a field direction of $\ell'' \sim 90^0$, $b'' \sim 0^0$ provided the sun lies in the direction towards the inner boundary of the local spiral arm as it may do[45].

5. Summary and Future Experiments

The conclusions which have been reached above as to the origin of cosmic rays are:-

1. 10^{12} - 2.10^{17}eV : It is likely that the majority of these particles are

accelerated in sources which lie within our own Galaxy. The levels of anisotropy observed from 10^{15} to 2.10^{17}eV are much larger than expected on any extragalactic theory of origin and since there is evidence from studies of γ-ray data that 10^9 - 10^{10}eV cosmic rays are Galactic it would be surprising if particles of intermediate energy had a different origin. Furthermore a number of plausible Galactic models have been devised[46-49] which, though differing in their assumptions, lead to predictions for the amplitude of anisotropy in rough agreement with experiment.

It should be borne in mind, however, that a mixture of an isotropic extragalactic component and an anisotropic Galactic component could be present at any energy. For example the observed anisotropy near 10^{17}eV might arise from a Galactic component which had a 100% 1st harmonic amplitude (a point source produces a 200% amplitude) and an isotropic extragalactic component which comprised $\sim$98% of the total intensity.

2. $\underline{2.10^{17}\text{-}10^{19}\text{ eV}}$: It is not clear at the present time whether cosmic rays in this energy band originate dominantly from within our Galaxy or outside of it. If the major mass component is established to be hydrogen then it will be hard to sustain the view that the majority of the particles originate within our Galaxy.

3. $\underline{>10^{19}\text{eV}}$: The evidence for anisotropy in a direction nearly normal to the Galactic plane, coupled with the difficulty of accelerating particles to these energies in any known structure within our Galaxy, makes it probable that the particles observed are of Galactic origin. The continuation of the spectrum to just beyond 10^{20}eV, however, requires that the sources lie within 200Mpc and probably closer.

What experiments need to be undertaken to test these conclusions? At the highest energies the major needs are for data from the Southern Hemisphere, which should come shortly from the Sydney experiment[35] and for a three-to-tenfold increase in statistics which may be achieved by the University of Utah's Fly's Eye experiment[50] in the near future. If 10^3km^2 of collecting area could be monitored for 10 years only about 5 events with E>10^{21}eV would be expected from an extrapolation of the known rate at 10^{20}eV. Between 10^{17} and 10^{19}eV high statistics experiments in which it is possible to measure the arrival direction pattern of individual mass components are being discussed and preliminary data of this type may come from the recently commissioned large array in Japan[51]. At lower energies the Nagoya group has plans to extend their high counting rate, small air shower experiments to energies above 10^{14}eV[52] while between 10^{15} and 10^{17}eV experiments currently running at Haverah Park enable an interesting region to be surveyed with current

technology. Below 10^{17}eV however, the smallness of the showers makes it hard to see how mass assignments can be made efficiently for individual events.

Of immense benefit to the interpretation of existing data would be theoretical effort aimed at pinning down likely acceleration regions. In particular the enormous power required to be input to high energy cosmic rays and the acceleration of particles to 10^{20}eV pose challenging problems for astrophysicists.

Acknowledgements

I am grateful to the organizers of the 1st Moriond Astrophysics meeting for inviting me to attend and for the financial support which helped to make it possible. I would like to thank my co-authors of reference (32) for permission to quote our results prior to publication, and Jeremy Lloyd-Evans for lively and helpful discussions. The continuing support for air shower studies at Haverah Park given by the Science Research Council is gratefully acknowledged.

References

1. G. Cunningham et al., Ap.J.236, L71, 1980.
2. A.M. Hillas, Proc. 16th Int. Cosmic Ray Conference (Kyoto) 8, 7, 1979.
3. J.A. Goodman et al., Phys.Rev.Letters 42, 854, 1979.
4. K. Greisen, Phys.Rev.Letters 16, 748, 1966.
5. G.T. Zatsepin and V.A. Kuzmin, Soviet Phys - JETP Letters 4, 78, 1966.
6. A.W. Strong, J. Wdowczyk and A.W. Wolfendale, J.Phys A 14, 1767, 1974.
7. J. Puget, F.W. Stecker and J.H. Bredekamp, Ap.J. 205, 638, 1976.
8. J. Linsley, Proc. 15th Int. Cosmic Ray Conference (Plovdiv) 12, 89, 1977.
9. J. Linsley and A.A. Watson, Phys.Rev.Letters 46, 459, 1981.
10. G. Thornton and R. Clay, Phys.Rev.Letters 43, 1622, 1979, and Phys.Rev. Letters 45, 1463 (E), 1980.
11. J. Linsley, in Proceedings of IUPAP/IAU Symposium N° 94, "Origin of Cosmic Rays", Bologna, Italy, 1980 (in the press).
12. P. Kirally et al., Rivista del Nuovo Cimento, 2, 1, 1979.
13. H. Elliot, Proc. 16th Int. Cosmic Ray Conference (Kyoto) 14, 200, 1979.
14. S.T. Davies, H. Elliot and T. Thambyahpillai, Proc. 16th Int. Cosmic Ray Conference (Kyoto) 4, 210, 1979.
15. A.G. Fenton and K.B. Fenton, Proc. Int. Cosmic Ray Symposium on High Energy Cosmic Ray Modulation (Tokyo), p.313, 1976.
16. H.E. Bergeson et al., Proc. 16th Int. Cosmic Ray Conference (Kyoto) 4, 188, 1979.
17. S. Sakakibara et al., Proc. 16th Int. Cosmic Ray Conference (Kyoto) 4, 216, 1979.
18. G. Benko et al., Proc. 16th Int. Cosmic Ray Conference (Kyoto) 4, 205, 1979.
19. J. Linsley and A.A. Watson, Proc. 15th Int. Cosmic Ray Conference (Plovdiv) 12, 203, 1977.
20. D.M. Edge et al., J.Phys.G 4, 133, 1978.
21. J. Lapikens et al., Proc. 16th Int. Cosmic Ray Conference (Kyoto) 4, 19, 1979.
22. J. Linsley, Proc. 14th Int. Cosmic Ray Conference (Munich) 2, 598, 1975.

23. D.D. Krasilnikov, Contributed paper to 7th European Cosmic Ray Symposium, Leningrad, 1980 (unpublished).
24. Catalogue of Highest Energy Cosmic Rays, N^{O} 1, (World Data Center C2 for Cosmic Rays, Institute of Physical & Chemical Research, Itabashi, Tokyo) (Editor M. Wada) p.5 and p.61, 1980.
25. C.J. Bell et al., J.Phys.A 7, 990, 1974.
C.J. Bell, J.Phys.G 2, 857 and 867, 1976
A.D. Bray et al., Proc 16th Int. Cosmic Ray Conference (Kyoto) 12, 365, 1979.
26. J. Delvaille et al., J.Phys.Soc. Japan (Suppl A 3) 17, 76, 1962.
J. Delvaille Ph.D. Thesis, University of Cornell, 1962.
27. G.W. Clark et al., Phys.Rev. 122, 637, 1961.
28. A.M.T. Pollock, R.J.O. Reid and A.A. Watson, Proc. 15th Int. Cosmic Ray Conference (Plovdiv) 2, 292, 1977.
29. D.D. Krasilnikov et al., Proc. 15th Int. Cosmic Ray Conference (Plovdiv) 2, 189, 1977.
30. R.G. Brownlee et al., Proc. 13th Int. Cosmic Ray Conference (Denver) 4, 2530, 1973.
31. S.I. Syrovatskii, Preprint N151, P.N. Lebedev Physical Institute, Moscow, 1969. Comments on Astrophysics & Space Physics 3, 155, 1971.
32. S.M. Astley, G. Cunningham, J. Lloyd-Evans, R.J.O. Reid and A.A. Watson to appear in Proc. 17th Int. Cosmic Ray Conference (Paris), July, 1981.
33. J.D. Delvaille et al., Proc. 5th Inter-American Seminar on Cosmic Rays (La Paz) 2, XL-1, 1962.
34. D. Dodds, A.W. Strong and A.W. Wolfendale, Mon.Not.Roy.Astron.Soc. 171, 569, 1975.
35. R.D. Blandford and J.P. Ostriker, Ap.J. 237, 793, 1980.
36. C. Heiles, Ap.J. 229, 533, 1979.
37. A.M. Hillas, Can.J.Phys. 46, S623, 1968.
38. F.W. Stecker, Phys.Rev.Letters, 21, 1016, 1968.
39. V.S. Berezinsky and G.T. Zatsepin, Sov.J.Nucl.Phys. 13, 453, 1971.
40. M. Giler, J. Wdowczyk and A.W. Wolfendale, J.Phys.G 6, 1561, 1980.
41. J. Wdowczyk and A.W. Wolfendale, Nature 281, 356, 1979.
42. A. Yahil, A. Sandage and G.A. Tammann, Ap.J. 242, 448, 1980.
43. G. Cocconi, Nuovo Cimento 3, 1433, 1956.
44. D.S. Mathewson, M.P. Schwarz and J.D. Murray, Ap.J. 217, L5, 1977.
45. H.R. Dickel et al., IAU Symposium N^{O} 38 (D. Reidel), p.213, 1970.
46. I. McIvor, Mon.Not.Roy.Astron.Soc. 179, 13, 1977.
47. M.C. Bell, J. Kota and A.W. Wolfendale, J.Phys.A 7, 420, 1974.
48. A.J. Owens and J.R. Jokipii, Ap.J 215, 677, 1977.
49. B. Peters and N.J. Westergaard, Ap.Sp.Science 48, 21, 1977.
50. H.E. Bergeson et al., Phys.Rev.Letters 39, 847, 1977.
J.W. Elbert, paper in these Proceedings.
51. T. Hara et al., Proc. 16th Int. Cosmic Ray Conference (Kyoto) 8, 135, 1979.
52. K. Murakami and K. Nagashima (private communication).

THE STUDY OF AIR SHOWERS BY THE FLY's-EYE

Jerome W. Elbert
Physics Department, University of Utah

ABSTRACT

The Fly's-Eye is an optical-electronic system designed to detect scintillation and Cherenkov light from ultra-high-energy cosmic ray air showers. For about 9 months, 48 out of the total 67 mirrors of the system have been operated. Air showers with greater than 10^{17} eV are being detected and reconstructed by the University of Utah Cosmic Ray Group. The Fly's-Eye will allow the development of individual showers in the atmosphere to be studied and will be able to observe showers many kilometers away. Besides observing the spectrum and anisotropy of air showers from 10^{17} to 10^{20} eV cosmic ray nuclei, the system may also detect point sources of 10^{14} to 10^{16} eV γ-rays. Preliminary data from a search for γ-rays from the vicinity of the Crab Pulsar are presented.

I. INTRODUCTION

The Fly's-Eye is the first apparatus designed to provide truly remote detection of cosmic ray air showers. In this paper I would like to describe the Fly's-Eye to astrophysicists and physicists who are not cosmic ray specialists and to point out the goals of the air shower work to be done by the Fly's-Eye. Since it is likely, or at least possible, that cosmic rays are produced in supernova remnants, active galactic nuclei, and quasars,[1-3] some cosmic ray studies can have implications outside the field of cosmic ray phenomena. For example, models of cosmic ray production have been proposed [3,4] in which a major part of the emitted energy of astronomical objects consists of cosmic rays or very high energy γ-rays which are observable by cosmic ray detectors. For the Fly's-Eye there are two general kinds of observations which are of special astrophysical interest.

One kind of observation is the study of cosmic rays within the energy range 10^{17} to 10^{21} eV. Some astrophysical implications of these studies will be described in Section V of this paper. Somewhere above 10^{20} eV, cosmic rays travel relatively direct paths to the earth, and large anisotropies are expected in their arrival directions. In this extreme high energy limit, cosmic ray studies resemble other forms of astronomy, with nearly undeflected charged particles arriving at the earth and giving information about their sources. The study of the chemical composition of the ultra-high-energy cosmic ray nuclei also has astrophysical significance.

A second kind of observation of astrophysical interest is the detection of primary γ-rays at higher energies than have been studied so far. In this case the relevant energy range is from about 10^{14} eV to 10^{16} eV. Preliminary results of γ-ray observations from the vicinity of the Crab Nebula are presented in Section VI.

In Section II, the production of light by air showers and the basic idea of the Fly's-Eye is described. In Section III, the idea of the Fly's-Eye is described. Section IV describes the Fly's-Eye itself. The ways in which the Fly's-Eye may contribute to knowledge of extremely high energy cosmic rays, astrophysics, and particle physics are described in Section V. Section VI discusses ultra-high-energy γ-ray detection by the Fly's-Eye.

II. PRODUCTION OF LIGHT FROM EXTENSIVE AIR SHOWERS

Extremely high energy primary cosmic rays (mostly protons and other nuclei) enter the earth's atmosphere and collide with the nuclei of air atoms. Mesons are produced in these interactions. The nucleons and some of the mesons from the

collisions go on to collide with other air nuclei, producing more mesons. Since tens of mesons are produced in each such collision and the atmosphere is about ten interaction lengths thick, the resulting cascades consist of many thousands of mesons, most of which are π mesons. This is the hadronic part of the "air shower."

Unlike the charged π mesons, which live long enough (if they are very energetic) to collide with nuclei and produce other mesons, about a third of the mesons are neutral and decay almost immediately into two γ-rays. The γ-rays produce electron-positron pairs in collisions with air nuclei. The electrons and positrons produce more γ-rays by the bremsstrahlung process. The atmosphere is very "thick" compared to the typical amount of material needed for these processes. The γ-rays from the decay of a single energetic neutral π meson can result in an electromagnetic cascade consisting of thousands or even millions of electrons, positrons, and lower energy γ-rays.

Part of the shower energy is transferred to the electromagnetic cascade In each generation of the hadronic cascade. By the time such a shower has reached the surface of the earth most of the shower energy has been shifted from the hadronic to the electromagnetic cascade, and most of the charged particles are positrons and electrons. At first in the development of such an "extensive air shower" the numbers of electrons and positrons increase. The electromagnetic cascade procedes until electron and positron energies decrease to about 50 to 100 MeV. At these energies, the electrons and positrons lose about as much of their energy by ionization as by bremsstrahlung, and the multiplication process comes to an end.

Because of the processes described above, the number of charged particles in an extensive air shower gradually increases to a maximum value and then gradually decreases. The very highest energy showers, however, have not reached their maximum number of particles (or maximum "size") before the shower strikes the ground, if the shower is vertical. The cascade propagates in a straight line at essentially the speed of light. The width of the cascade, about 100 meters, is mainly the result of multiple Coulomb scattering of the low energy electrons and positrons.

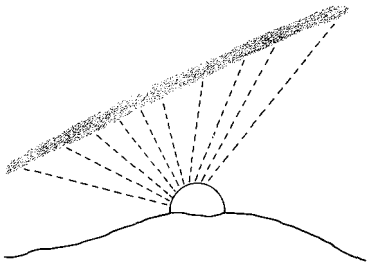
The Fly's-Eye detects light produced by energetic air showers. The most important part of the light detected by the Fly's-Eye is scintillation light. Ionizing shower particles produce excited nitrogen molecules and ions. Except at extremely high altitudes, only a small fraction of the excited molecules or ions emit photons. Most of the excited molecules and ions lose energy in collisions with other air molecules. The number of excited molecules or ions is proportional to the air pressure, while the probability of emitting a photon before a de-exciting collision takes place is inversely proportional to the pressure. As

a result of this, the number of emitted scintillation photons per particle per meter is roughly constant, independent of the altitude. The light is emitted isotropically. At fixed pressure, the light output is proportional to the ionization energy loss of a particle, and amounts to about 4 photons per particle per meter for minimum-ionizing particles. The scintillation light occurs in numerous bands in the near ultraviolet and violet end of the spectrum.

Cherenkov light is also produced by air showers. All charged particles traveling at greater than the speed of light in air produce this light. Near sea level, a highly relativistic particle produces about 25-30 visible photons per meter in air. These photons are emitted at an angle of about 1.4° with respect to the particle's direction. However, due to multiple Coulomb scattering and other effects, shower particles have average angles of about 8° with respect to the direction of propagation of the shower as a whole. Consequently, this angle dominates the angular distribution of Cherenkov light. At sea level, only about half the shower particles have velocities so close to c that they can produce Cherenkov light. At higher altitudes, a somewhat smaller fraction of particles can emit Cherenkov light, because the index of refraction of the air is closer to unity. Although the number of produced Cherenkov photons is greater than the number of scintillation photons, the extreme difference between the angular distributions of the two light components results in the dominance of scintillation light when showers are viewed at large angles from the shower direction.

III. THE IDEA OF THE FLY'S-EYE

The classical method of detecting air showers is by constructing a grid of of particle detectors and registering events in which a number of adjacent detectors count particles simultaneously. Such grids can cover areas of tens of square kilometers.[5] These grids or air shower arrays have given us most of our present knowledge about the highest energy air showers. Since the highest energy cosmic rays are extremely rare, it is desirable to extend the detection areas for cosmic ray air showers to the largest feasible values. Optical detection of remote air showers allows the detection area to be increased to the range of hundreds or perhaps thousands



FLY'S-EYE AIR SHOWER DETECTOR

of square kilometers. The basic idea of the Fly's-Eye (which we owe to Professor K. Greisen),[6] was to develop an optical-electronic system to exploit this possibility.

There is another major advantage of the use of remote detection of cosmic ray air showers. Unlike an array of particle detectors, which measures the air shower "sizes" (numbers of particles) only at the surface of the earth, the remote detection of air showers allows the longitudinal development of individual air showers to be studied. That is, the number of shower particles can be obtained at many positions along the air shower track. This information allows studies to be made of the chemical composition of the primary cosmic rays, the interaction length of the extremely high energy protons in air, and the "elongation rate" of showers, which is dependent on the number of produced particles in high energy collisions. These studies will be discussed in Section V.

The origin of the name, "Fly's-Eye," is that the apparatus is somewhat analogous to the eye of an insect in which a large number of optical elements are arranged in such a way that each element can detect the intensity of light in a particular narrow angular range. An air shower, as seen in the figure, produces a signal in optical elements arranged on a great circle drawn on the imaginary hemisphere representing the field of view of the Fly's-Eye. In the Fly's-Eye there are 870 photomultiplier tubes (PMTs) at the focal points of 67 spherical mirrors. Each of the PMTs views a hexagonal-shaped aperture of about 2.7° half-angle. The pattern of PMTs which see light from a given shower defines the directions of a plane (the shower-detector plane) containing the Fly's-Eye and the linear trajectory of the air shower. The fact that the shower crosses the the sky at the speed of light is used together with the measured times at which the PMTs observe the shower to obtain the distance of the shower and the angle between the shower path and a horizontal line in the shower-detector plane. This completely determines the position and direction of the shower trajectory. The intensity of the light observed by the different PMTs which view the shower gives the number of particles at numerous positions along the shower trajectory. Thus, the Fly's-Eye obtains a great deal of information about each shower which it detects.

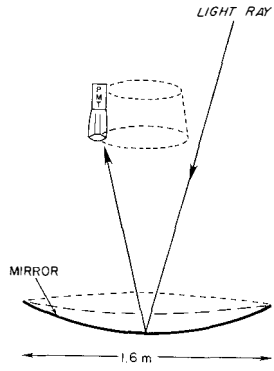
The particular uses of the Fly's-Eye described above are based on the detection of scintillation light from air showers at distances of about 1 km or more from the Fly's-Eye. There are also very many lower energy air showers which fall within about 200 meters of the Fly's-Eye. For most of these showers the observed light is predominantly Cherenkov light emitted within a few degrees of the direction of travel of the shower. The "image" of such a shower in the Fly's-Eye is a brief light flash in one or more PMTs. The idealized geometry for such an event, called a "blast" is an instantaneous flash from a specific direction. In this case the shower direction is nearly the direction from which the light is ob-

served, and the time integral of the light pulse gives a lower limit to (and an approximate value of) the shower energy. These "blasts" will be discussed further in Section VI, in connection with the search for sources of ultra-high-energy γ-rays.

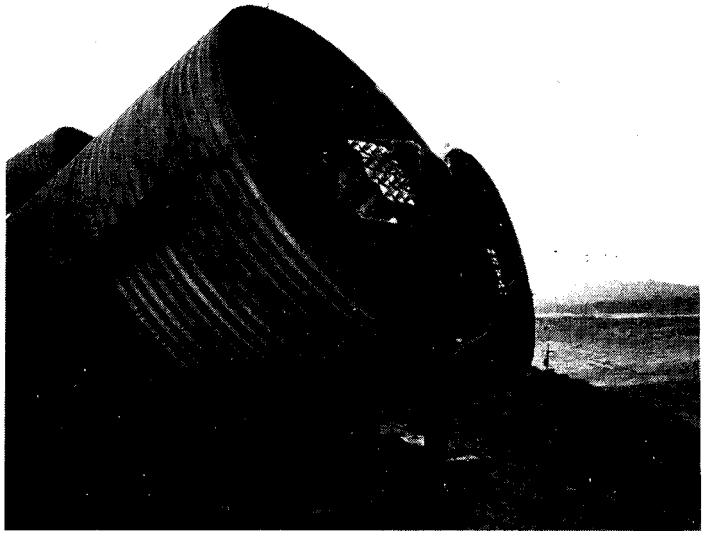
IV. THE FLY'S-EYE

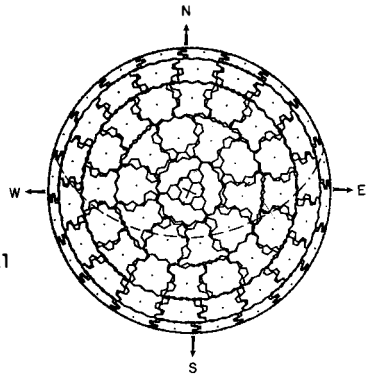
The Fly's-Eye is funded by the U.S. National Science Foundation. The principle investigator of the project is Professor George Cassiday, of the University of Utah Physics Department. The location of the Fly's-Eye is on the Dugway Army Proving Grounds, Utah, U.S.A. This location is about 140 km southwest of Salt Lake City. The site is on a small mountain which is about 150 m above the surrounding country. The sparsely inhabited location is away from the lights and air pollution of large cities. The location is also desirable because the sky is free of clouds a large part of the time.

The figure represents a mirror unit of the Fly's-Eye. The 1.6 m diameter mirror focuses light (focal length 1.5 m) onto a cluster of 12 or 14 hexagonal-faced light collectors, each of which reflects the light onto the face of a PMT. The mirrors are spherical and are made of aluminized glass. The mirrors are shaped by the University of Utah Cosmic Ray Group by gravity-slumping of a circular piece of ordinary plate glass onto a graphite mold in an electrically heated oven. Aluminum is evaporated onto the mirrors in a very large vacuum chamber. The light collectors are plastic funnel-shaped devices which are aluminized on the inside surface.



Cylindrical steel housings shelter the mirrors. The mirror is located on the lower left-hand side of the cylinder shown in the photograph. The housing is shown in position for viewing the sky. During the day the cylindrical unit is rotated so that the open end faces the ground. If the fields of view of all the mirror boundaries are projected onto a hemispherical dome, the resulting mirror boundaries are shown in the following figure. The mirror units have been aligned to agree with this pattern with an accuracy of 1/2 degree. The Fly's-Eye is operated at night when the moon is not up and the sky is cloud-free. Because of these constraints, the Fly's-Eye operates about 10% of the time. There will be 67 mirrors units in the complete Fly's-Eye. During the last 9 months 48 mirror units have been in operation. All 67 units will be operated later this year.





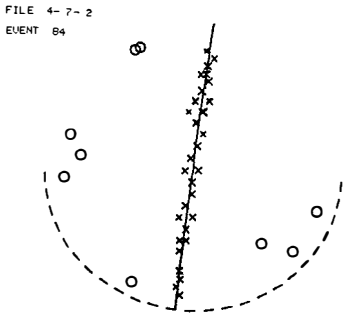
The sensitivity of the Fly's-Eye is limited by the noise produced by fluctuations in the background light, produced by stars and chemical processes in the upper atmosphere. In order to optimize the sensitivity of the Fly's-Eye in the presence of the varying background light intensities, the triggering rate of PMTs in each mirror is monitored by our Digital Equipment Corporation PDP-1134 computer and the thresholds are automatically adjusted to keep the counting rates within the desired range. A typical sensitivity is such that the scintillation light from a 10^{17} eV air shower is just detectable if the closest point of the shower trajectory is 1.7 km from the Fly's-Eye.

Showers which are, for example, 1 km from the Fly's-Eye pass through the field of view of a PMT in a much shorter time (~ 1/3 μs) than showers which are 20 km away (~ 7 μs). The signal currents are integrated over time periods fixed by the electronics. It is not optimal to integrate data for 10 μs for nearby showers since background fluctuations would be increased during the more than

9 μs of excess integration time. Thus, some showers need to be integrated for long periods, while other showers are best integrated for short time periods. This problem is treated by fanning out each PMT signal into 4 data channels with integration times spanning the range from 0.8 to 20 μs. In each shower, data are recorded in 4 channels for each tube which triggered. After the distance of a given shower is determined, the selection of data from the best data channel can be done for each PMT.

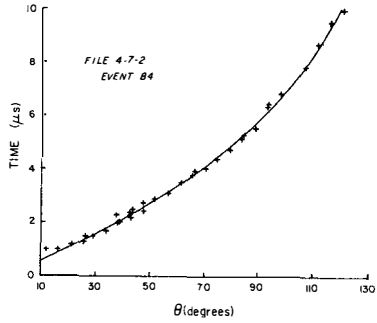
Triggering (deciding that a possible event is present and that data should be recorded) is accomplished in this system by a two-level hierarchy. First, a mirror trigger is formed when two PMTs in a given mirror pass over threshold at nearly the same time. A system trigger occurs when two mirrors trigger at nearly the same time. By requiring that these "coincidences" be separated by a minimum interval, the triggering accepts cases in which a shower crosses the field of view of the Fly's-Eye, but rejects events in which the light arrives nearly simultaneously in all PMTs. The latter events are the "blasts" mentioned previously, which are only of interest when looking for γ-ray sources.

During normal operation, about 10 track-like events are observed per hour. There is no difficulty in distinguishing showers from the background noise because real showers produce a very orderly sequence of pulses in a line across the sky. Data which are stored for each event are the time at which the event occurred, the identities of all PMTs which triggered in the event, the values of thresholds at the time of the event, the relative times at which different PMTs triggered (accurate to 50 ns), and the pulse integrals from the four data channels of each PMT which triggered.



The reconstruction of the geometry of the shower is done in two stages. The first is to fit the shower-detector plane by analysis of the directions of tubes in which signals were detected in a given event. The figure shows the pattern of tubes which fired in particular event. The view is of a dome representing the sky, with x's representing positions of tubes which were part of the event. Circles represent tubes in which triggers occurred but were identified as noise tubes by the pattern recognition program. The dashed line represents the horizon. The great circle representing the fit to the shower-detector plane is drawn by the computer through the representation of the data. There is little difficulty in obtaining the orientation of this plane with very good accuracy.

The next part of the fit is done to the times at which the PMTs triggered. The following figure shows a fit of the times, t_i, as a function of the angles θ_i between the direction of the shower and the direction the light traveled to the i-th PMT. The relation between the variables is given by



$$t_i = t_o + \frac{R_p}{c} \tan(\theta_i/2)$$

R_p is the "impact parameter" (measured from the Fly's-Eye to the closest point, P, on the shower trajectory) and t_o is the time at which the shower particles reach point P. Since the θ_i depend on a choice of an angle ψ between the shower direction and a horizontal line from the Fly's-Eye to the shower, the fit is to the 3 parameters t_o, R_p, and ψ.

If a shower is viewed only at small values of θ_i then $\theta_i/2 \approx \tan(\theta_i/2)$ and $t_i \approx t_o + R_p/c(\theta_i/2)$. This is an equation for a straight line relating the t_i to the θ_i, and it does not contain sufficient information to give the values of the 3 unknowns. However, the tangent function is non-linear and this curvature of the $t_i(\theta_i)$ plot (which is evident in the figure) allows the shower to be reconstructed. In the shower shown here, the impact parameter $R_p = 1.73 \pm 0.10$ km and the light emission angles θ_i range from 12 to 122 degrees, with an uncertainty of 2 degrees in these angles. The angular uncertainties of showers range from about 1 degree up to infinity, depending on the angular range in which the showers are visible.

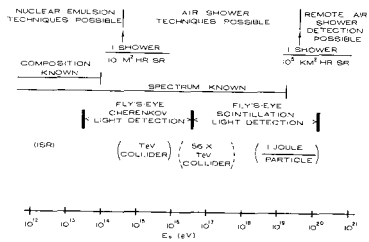
After the direction and position of a shower are determined, the PMT signals can be analyzed in order to infer the number of particles present in the different regions of the shower. The determination of the number of particles within the field of view of a given PMT requires the relative amount of light expected from scintillation and from Cherenkov emission to be estimated. The procedures for this analysis are being developed and the development of the cascades in the atmosphere should be known soon.

Artificial light sources have been built into the mirror units which allow the long term variation of the reflectivity of the mirrors and light collectors or changes of the PMT gains to be measured. Procedures have been developed to equalize the sensitivities of all PMTs in the system. An optical flasher has been developed to produce artificial "showers" to be detected by the Fly's-Eye. An intense pulse of collimated light was sent over the Fly's-Eye at "impact parameters" of 1-3 km. The flasher unit was accurately aimed to give a beam of

known direction. Part of the beam was scattered by the air and was received by the Fly's-Eye, as in a real shower. The beam direction was calculated from the Fly's-Eye data to check the reconstruction procedures. The detected light intensities were studied to check the atmospheric scattering model. Both checks have given us confidence that the real showers can be successfully analyzed using the Fly's-Eye.

V. THE GOALS OF THE FLY'S-EYE PROJECT

In order to introduce the Fly's-Eye goals, I would like to draw a rough picture of the current knowledge in this field. The figure illustrates the information available at various energies. At the top of the figure, the energy domain is shown in which nuclear emulsion techniques or other direct measurements are possible. These measurements are done in space or at very high altitudes and require the cosmic ray to penetrate the apparatus so that direct measurements can be made of the ionization rate and energy of the nucleus. Somewhere around 3×10^{14} to 10^{15} eV these direct methods become impractical as the intensity of primary cosmic rays above this energy becomes less than about 0.1 particle m^{-2} hr^{-1} sr^{-1}. This energy limit may be increased in the future using experiments in space, but the primary spectrum falls so rapidly with increasing energy that it is difficult to produce major changes in the limit.



Above the energies of the direct measurements of the primary cosmic rays the ground-level air shower arrays can be used to detect cascades resulting from the primary cosmic rays. Such techniques are indirect, in that the characteristics of the primary cosmic rays need to be deduced from the cascades which they produce. The gain in sensitivity by the air shower technique is enormous, however, since individual showers spread over areas of about 10^5-10^6 m^2, so that particle detectors can be separated from each other by more than 100 m with a total grid area covering tens of square kilometers. The air shower arrays can measure total shower energies as well as the arrival directions of the air showers. The air shower spectrum and anisotropies of arrival directions have been measured up to about 10^{20} eV.[5] Indirect methods have been used to obtain some information about the primary composition in the air shower energy region.[7,8] Although these investigations have given interesting results, it is too early to say that a consensus of opinion has been reached on the composition problem.

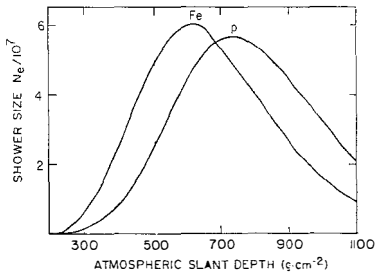
At energies above 10^{20} eV, the most promising measurement technique for studying the air shower spectrum and the cosmic ray chemical composition is remote optical detection, as developed in the Fly's-Eye. In this case, showers with $E_o > 10^{20}$ eV can probably be viewed at distances of 20 km, yielding a detection area of about 1200 km^2. Showers in about 3 steradians of solid angle may be detectable, and the system would run about 10% of the time. Since the flux of such showers is about 10^{-6} km^{-2} hr^{-1} $steradian^{-1}$, about 3 of them could be detected per calendar year. A few years of operation would greatly improve the world's statistics on such ultra-high-energy showers.

Above about 3×10^{16} eV, the Fly's-Eye can detect air showers by the scintillation light which they produce. It also has the capability, however, of detecting lower energy showers by means of the Cherenkov light which they emit. It was mentioned earlier that the Cherenkov light from showers is very intense when, and only when, the light has been emitted at small angles with respect to the direction of propagation of the shower. As a result, lower energy showers can be detected by Cherenkov light, but only when the shower axes fall within a few hundred meters of the Fly's-Eye. Although this implies a shower detection area of only about 10^5 m^2, the rate of such showers is very high, since they can be observed down to energies of about 3×10^{13} eV for which the flux is about 2 per m^2 per hour per steradian.

Historically, one of the important uses of cosmic ray studies has been the possibility of exploring nuclear interactions at higher energies than accelerators can produce. The figure also illustrates the relationshiip of the Fly's-Eye to present and future accelerator experiments. Currently, the highest energy accelerator studies are at the Intersecting Storage Rings (ISR) at the European Center of Nuclear Research (CERN) in Geneva. The peak ISR energy is equivalent to 200 GeV protons striking "stationary" targets such as the earth's atmosphere. This year, CERN may start operating colliding proton-antiproton beams at an equivalent stationary target energy of 1.5×10^{14} eV. In a few years, colliding beams of protons and antiprotons at Fermilab in the U.S.A. will reach the equivalent of 2×10^{15} eV primary cosmic rays. This is still outside the energy range for which the Fly's-Eye was designed to study nuclear interaction characteristics. However, the interaction characteristics observed at Fermilab at 2×10^{15} eV would allow us to predict the characteristics of 10^{17} eV ($56 \times 2 \times 10^{15}$ eV) air showers produced by iron nuclei quite accurately.

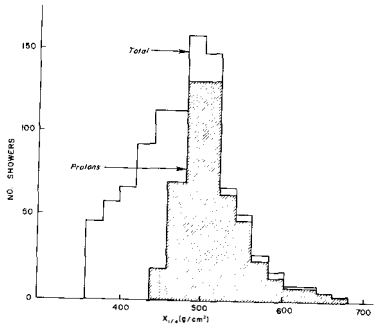
One of the most difficult problems in the interpretation of air shower data is the uncertainty about whether certain somewhat surprising characteristics of air showers are due to high energy interaction characteristics unlike those seen at accelerators or that the characteristics are due to a composition of the primary cosmic rays which is very different from that which is seen at lower en-

ergies. The problem is illustrated by the next figure. For some specific model of the nuclear interaction process one may obtain the longitudinal development curves chosen in the figure for 10^{17} eV air showers from primary cosmic rays which are protons and iron nuclei. If the experimental curve is like that labeled Fe, it may mean that the primary cosmic rays are mostly iron nuclei or that the correct interaction model gives development curves from primary protons which are identical in shape with the curve labeled Fe.



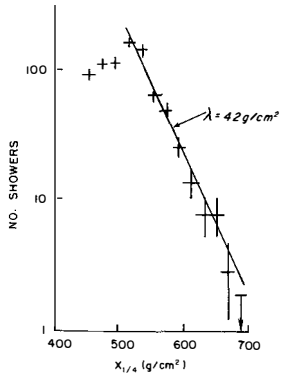
The solution to this problem may come from accelerator experiments, where the proton-antiproton collisions may be done at high enough energies so that the expected shape of the Fe development curve may be given with reasonable certainty at 10^{17} eV, as mentioned earlier.

However, it may be possible to solve this ambiguity of composition vs. interaction model using Fly's-Eye data. Suppose we look at the very early part of the shower development curves, where the shower size is rapidly increasing with depth and is only 1/4 the size it will be at maximum development. The depth at which this occurs is less dependent on the interaction model than is the depth of the shower's maximum size. But this depth, called $x_{1/4}$, is strongly affected by the depth at which the primary cosmic ray first interacted. For example, Professor T.K. Gaisser has calculated the distribution of $x_{1/4}$ for monoenergetic air showers produced by a composition in which protons and various other nuclei were present. The figure shows the results, with the contribution due to protons indicated by shading.



It can be seen that protons are very dominant at the deeper (higher $x_{1/4}$). part of the distribution. In fact, when the same data are plotted with a logarithmic scale, as in the following figure, the distribution gives the absorption curve for the proton primaries, with a slope equal to that used for protons in the calculation. If only Fe nuclei were present the distribution would fall off more rapidly at large depths and the fluctuations in $x_{1/4}$ would be much smaller. This kind of observation can give the proton-air inelastic

cross section at very high energies and it can give an estimate of the fraction of protons among the primary cosmic rays at high energies. The cross section measurement is of interest to high energy nuclear physicists and the fraction of protons in the primary composition is of astrophysical interest. The ability to identify certain showers as proton showers would make the implications of the depth of the maximum of the showers less ambiguous. Knowledge of the composition would make the nuclear physics implications of other cosmic ray experiments clearer, as well.



There may be surprising showers waiting to be discovered by the Fly's-Eye. Since the Fly's-Eye will be the first instrument to obtain reasonably detailed information about the longitudinal development of large numbers of showers above 10^{17} eV, certain rare interactions might be observed. For example, if interactions of the "Centauro" type occur in very high energy air showers, some showers might start very deep in the atmosphere and then develop very quickly.

Another example of a possible exotic event would be air showers produced by interactions of ultra-high-energy neutrinos. These would be identifiable because they would interact and produce showers at great depths in nearly horizontal or upward-going trajectories through the atmosphere. Other presently unknown particles might be produced far away in nearly horizontal showers and then interact or decay near the Fly's-Eye to produce showers that are clearly at an abnormally large atmospheric depth.

Two kinds of results which have astrophysical implications are the ultra-high-energy spectrum and the anisotropy of the arrival directions. There is a lack of isotropy in the highest energy cosmic rays which are observed.[5] The excess showers are not observed in the galactic plane, so they do not appear to be of galactic origin. Their interaction with the universal black-body radiation puts an upper limit on the distances which they are likely to have traveled. This interaction would produce a large attenuation of showers with $E_0 > 10^{20}$ eV traveling distances much more than 50 Mpc, with much less attenuation at significantly lower energies. The observed spectrum shows some flattening in this energy range which is the opposite of the effect expected from the attenuation. This implies that the likely origin of these particles is in relatively local clusters of galaxies. Since the degree of anisotropy appears to increase with energy, the observation of showers in the 10^{20} eV region is of great interest.

One of the goals of the Fly's-Eye work is to add to the world supply of air showers in this energy range.

An activity which has recently been started with the Fly's-Eye is the search for 10^{14} to 10^{16} eV γ-ray sources using the Fly's-Eye. Our interest in this search has been stimulated by the observations of Dzikowski et al.[9] and by our results which are described in the next section.

VI. POSSIBLE GAMMA-RAY DETECTION BY THE FLY'S-EYE

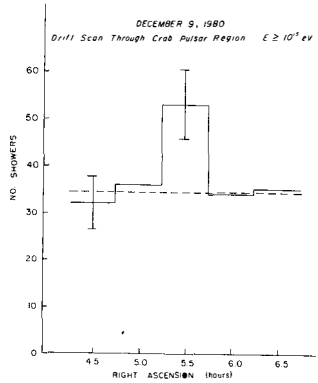
There are a number of reasons why the search for 10^{14} to 10^{16} eV γ-ray sources is appealing to our research group. The possibility of such searches represents a capability of the Fly's-Eye that has emerged without being specifically "designed into" the system. The production of such gamma rays is by very high energy particles. This implies that sources of energetic γ-rays are likely to be cosmic ray sources as well. However, the search for the γ-ray sources is more direct than attempts to determine the sources of the cosmic rays because the γ-rays point directly to the sources rather than being deflected by magnetic fields. The γ-rays from pulsars are especially interesting because the directions of likely sources are known and the radiation may be pulsed with the previously measured period of the pulsar. The search for sources with these particular characteristics can be quite sensitive.

Our interest was also stimulated by the work of a cosmic ray group at the University of Lodz. Using an air shower array with a muon detector, Ozikowski et al.[9] saw a 3.6σ excess of showers from the direction of the Crab Pulsar compared to directions differing by 90, 180 and 270 degrees of right ascension from the Crab direction. For these showers, the numbers of muons was smaller by a factor of 0.60 ± 0.12 than in the comparison directions. This is consistent with a fraction of the showers being produced by primary γ-rays, since the number of muons expected in showers from γ-rays is much lower than in ordinary showers. The angular resolution of the array was poor, and data were taken for showers within 15° of the direction of the Crab Pulsar. The data were taken between 1975 and 1979. The estimated γ-ray flux was $3 \pm 2 \times 10^{-13}$ cm^{-2} s^{-1} for $E_0 > 10^{16}$ eV.

The Fly's-Eye has a number of advantages for searching for γ-rays compared to the array described above. The collection area is roughly 10^5 m^2 rather than 10^3 m^2. The angular resolution is about 3° or 4°. Also, the Fly's-Eye is at least as sensitive to γ-rays as to cosmic rays of a given energy, while an air shower array taking data in this energy range at sea level and at zenith angles greater than about 30° would tend to discriminate against γ-rays. These circumstances allowed a preliminary comparison to be made to the Polish group's results in a few night's running time of the Fly's-Eye.

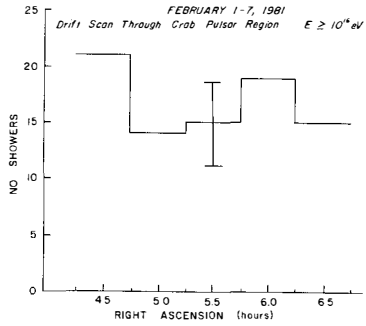
Special triggering conditions were set up for the Fly's-Eye to search for γ-rays. Events were accepted in which only one mirror detected light. For showers in the energy range 10^{15} to 10^{16} eV, the sensitivity of the system needs to be reduced to keep the large signals from Cherenkov light from saturating the electronics.

The dashed line in the second figure in Section IV represents the trajectory which the Crab Pulsar followed through the field-of-view of the Fly's-Eye. Data were collected from the different mirrors which viewed the entire region of right ascension between 4.5 and 6.5 hours. The distribution of showers from the first night of γ-ray search data is shown in the next figure. During this run very few showers above 10^{16} eV were observed, but the 10^{15} eV data show an excess of weak statistical significance near the right ascension of the Crab Pulsar.

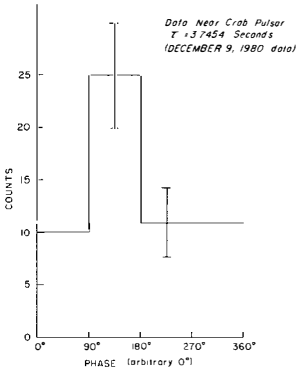


During five of the eight nights from February 1-7, data were taken again. With this longer running time a significant number of showers were obtained at $E_0 > 10^{16}$ eV. In this case, no excess was observed near the Crab Pulsar direction (see the figure). During this period the data at 10^{15} eV were mostly below threshold because the running conditions were set up differently than during the December run.

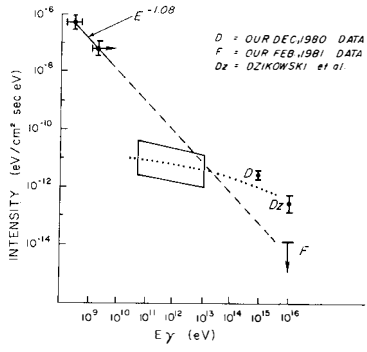
At present the Fly's-Eye clock is only accurate to about 1 second during a night's data run. There is a pulsar near the Crab Pulsar with a period of 3.7454 seconds. The December 9, 1980 data were checked to see if there was an indication of this period being associated with the excess of events seen on that night from near the Crab Pulsar. The data were analyzed at this period and the results are shown in the next figure. There is an excess of events occurring during about 1 second out of each period. The probability of this excess occurring by chance is ~ 3%. The data shown are for an energy of about 10^{15} eV. The excess of events is about 15, similar to the excess of events near the direction of the Crab Pulsar in the same data.



The statistical significance of the positive effects shown here is not overwhelming. However, more observations ought to be made in this energy region. The last figure compares results for γ-rays from the Crab Pulsar (or its vicinity in the case of the 3 highest energy observations). The dashed line is an extrapolation from measurements in the GeV region. The upper limits and observations represented by the parallelogram,[10] were made by detecting atmospheric Cherenkov light, as with the points D and F from the Fly's-Eye December and February data. The Polish group's data are identified by Dz. The dotted line is a possible trend of the high energy positive observations. Our February data disagree with the Dzikowski et al. observation by a large factor. However, the production of high energy γ-rays from these pulsars is thought to be variable, so there is not a contradiction between the two results.



Our weak, preliminary results do not prove that measurable fluxes of γ-rays are present in this energy range, but they suggest that more observations should be made. A clock that is accurate to about 1 ms over a 2 week monthly running time would be a worthwhile improvement to the Fly's-Eye because it would allow a more sensitive search to be made for showers occurring with the periodicity of various pulsars.



REFERENCES

1. Ostriker, J.P. and Gunn, J.E., Astrophys. J. 157 1395 (1969).
2. Harrison, E.R., Nature 264, 525 (1976).
3. Lovelace, R.V.E., Nature 262, 649 (1976).
4. Eichler, D. and Vestrand, W.T., Proc. 16th Int. Cosmic Ray Conf. 1, 147 (1979).
5. See the following and the references given therein: Cunningham, G. et al., Astrophys. J., 236, L71 (1980).
6. Greisen, K. Proc. 9th Int. Cosmic Ray Conf. 2, 609 (1965).
7. Goodman, J.A. et al., Phys. Rev. Lett. 42, 854 (1979)
8. Linsley, J. and Watson, A., Phys. Rev. Lett. 46, 459 (1981).
9. Dzikowski, T. et al., University of Lodz Preprint, 1980.
10. Weekes, T.C., Proc. of the 1978 DUMAND Summer Workshop, A. Roberts, ed. (Scripps Institution of Oceanography, LaJolla, 1978) p.313.

THE PROCESSING OF NUCLEI BY ELECTRONS IN ACTIVE GALAXIES

R. SCHAEFFER

Service de Physique Théorique
CEN Saclay
91191 Gif-sur-Yvette Cedex, France

The cross-section for spallation of nuclei by electrons and γ rays are calculated and some astrophysical consequences are presented.

Many active galaxies contain large fluxes of relativistic electrons and a rather accurate estimate of the energy they contain can be obtained from the observation of the radio emission[1]. These electrons process the galactic matter they encounter and may lead to abnormal abundances. In particular the more abundant elements such as Carbon may be transformed into Li, Be, B by a spallation process quite similar to the one that builds up[2] the latter elements in the galaxy, except that now the active agents are electrons, and not protons. The relevant cross-sections have been estimated recently and are large enough so as to lead to non-negligible astrophysical consequences[3]. The spallation of a nucleus by electrons, such as $e + C \rightarrow e + p + B$, is at first guess due to the knock-out of a proton by an electron. It is then quite similar to the proton spallation process $p + C \rightarrow p + p + B$, but due to the Coulomb force rather than to the nuclear force. The corresponding cross-section is thus reduced by roughly a factor $\alpha^2 \sim 10^{-4}$, α being the fine structure constant. The accurate calculation made in ref.[3] leads to $\sigma_e/\sigma_p \sim 3.3\ 10^{-4}$ for the $^{12}C \rightarrow {}^{11}B$ reaction, with a threshold for the electron energy near $E_e \sim 20$ MeV. This is the process currently measured[4] in electron scattering experiments, and the σ_e/σ_p ratio calculated[3] is consistent with these experiments. There are, however, besides the Coulomb interaction also the so-called "transverse" excitations that are negligible at the scattering angles where the measurements are currently made, but that become extremely large at small electron deflection angles. This process excites almost only the giant dipole resonance that in turn decays by particle emission. Its magnitude can be obtained by well-known sum rules. For the $C \rightarrow B$ reaction one gets $\sigma_e \sim 0.2$ mb, that is $\sigma_e/\sigma_p \sim 4\ 10^{-3}$, which is quite larger than the Coulomb process. There is however an important difference : proton spallation, as well as electron spallation by the Coulomb interaction produce B, Be and Li with nearly equal abundances. The giant dipole resonance, on the other hand, decays to the B ground state only, so little Be and Li is produced. Since this process dominates the spallation by electrons, we should retain that the spallation of C by electrons produces B, but no Be nor Li, in contrast to spallation by protons. The B production cross-section is about 200 times smaller for electrons than for protons. There is however still another factor of enhencement in dense media, due to the high energy γ rays produced from the electrons by Bremstrahlung. These γ rays also excite the giant dipole resonance, so one can simply add the two processes so as to obtain an effective spallation cross-section for B production

$$\sigma_{eff} = \left(0.2 + 8.6\ \frac{n(H)R}{\lambda}\right) \text{ mb} \qquad n(H)R < \lambda$$

$$\sigma_{eff} = 8.8 \text{ mb} \qquad n(H)R > \lambda$$

where n(H) is the hydrogen density, R the thickness of the gas and $\lambda \sim 3\ 10^{25}$ part/cm^2 the radiation length. Note the important increase of σ_{eff} as soon as

n(H)R is non negligible as compared to λ. In this case, too, there is no Be or Li production.

Formulas that can be used with a minimal knowledge of nuclear physics, and that are valid for any nucleus, are given ref.[3] for the electro and photo-spallation cross-sections as well as for the excitation of other nuclear states besides the giant dipole resonance.

Let us now consider the radio-sources where an upper limit of the ratio $n(B)/n(C) \leq 1/20$ for the relative Boron to Carbon abundance has been determined[5] from the search for some of the B atomic γ-ray lines. From this upper limit, these authors conclude that the observed gas cannot be emitted by the very central region since there would be too much B present. Also, assuming the observed gas is within a radius of 30 pc of the active center, they get a limit on the proton flux present there ($\varepsilon_p < 4\ 10^{53}$ erg). Using the same model, but considering now the spallation by electrons, we get a limit on the electron flux present : $\varepsilon_e < 10^{56}$ erg. Such a value can be directly compared to the value of ε_e deduced from the radio-data.

The observations of Baldwin et al were made in a medium where little γ's are produced ($n(H)R \ll \lambda$), but in more dense media where the 20 MeV photons induce the spallation an interesting comparison can also be made between the observed high energy γ emission and the Boron abundances. As well as the radio emission, the observation of the γ flux gives the instantaneous electron flux, whereas the B abundance would lead to the time-integrated flux. The γ emission, however, depends on the density of the galactic medium, whereas the radio emission does not. A detailed study of these emissions will thus provide an useful information on the density of the emitting matter as well as the time of irradiation of this matter by the electrons.

1) G.R. BURBRIDGE et al., Astrophys. J., 193, p.43 (1974).
2) H. REEVES, Ann. Rev. Astrophys. 12, p.437 (1974).
3) R. SCHAEFFER, H. REEVES and H. ORLAND, to be published.
4) J. MOUGEY et al., Nucl. Phys. A 262, p.461 (1976).
5) J. BALDWIN et al., Astron. Astrophys. 61, p.165 (1977).

A GUT-ED TOUR THROUGH THE EARLY UNIVERSE

D.V. Nanopoulos
CERN, Geneva

ABSTRACT :

A simplified view of the evolution of the Universe is presented. The implications of Grand Unified Theories at each characteristic period in the History of the Universe are pointed out. A new mechanism for Cosmological baryon production, through the decays of Superheavy fermions, is discussed in some detail.

Prologue

My talk is naturally divided in two parts. Part I contains a general survey of the evolution of the Universe, from $t = t_{Planck} \simeq 10^{-43}$ sec up to $t = t_o \sim 10^{10}$ years, as most of us imagine that it happened. Part II contains a new mechanism of Baryon creation, through the use of superheavy fermions, and some discussion of other competing mechanisms of Baryon production. I am going to assume, that the reader is familiar with the standard Big-Bang model and the basics of Grand Unified theories[1], at least at the level, discussed by John Ellis at the parallel meeting on Particle Physics.

I.EVOLUTION OF THE UNIVERSE : A PICTURE.SQUE APPROACH.

Picture 1 : THE BIG-BANG. ($T > M_{Planck} \sim 10^{19}$ GeV)

It seems that it all started at least ten to twenty billion years ago with a big explosion, very appropriately called the Big-Bang, under conditions which are beyond our current understanding but perhaps not hopelessly away from our grasp. Extrapolating back in time, our "classical" laws of physics, entail, that it is difficult, if not impossible, to avoid an initial singularity. Some of us feel unhappy with such a situation and hope that Quantum Gravity or some other yet unthinkable way will prevent such a disaster. Another "Collapse disaster" springs to mind, the Rutherford atom. It remains to be seen, if, Quantum Theory is going to save us again.

Anyway, we plead full ignorence at least for now, and start our scientific discussion from temperatures below the Planck temperature ($\sim 10^{19}$ GeV), where it is not inappropriate to ignore quantum gravitational effects.

Picture 2 : THE SMOOTHING ERA (10^{15} GeV $< T < 10^{19}$ GeV)

We assume now, that we are given an expanding baby-Universe with matter and radiation in equilibrium, and where the laws of Physics are known to us, are applicable. In the era under discussion, Grand Unified Interactions play an important role : they smooth things out[2]

At some stage during this era, Grand Unified Interactions come to equilibrium, and so through their Baryon number violating component wipe out any huge (local or global) excess of baryons or vice-versa.

That is very helpful because it makes subsequent creation of any Baryon asymetry free from initial conditions. At the same time, Grand Unified interactions,

through the so called Grand Unified viscosity, provide some help[2] in homogenizing and isotropizing a possibly initial inhomogeneous and anisotropic universe. It seems, that after a violent explosion and a few subsequent drastic strong gravitational effects, the Universe needs some time to relax and smooth things out. Grand Unified interactions seem to play an important role in this smoothing era.

Picture 3 : BARYOSYNTHESIS ($T \sim 10^{15}$ GeV (?))

Let's take for granted, that the Universe has been brought in a state, where it is almost homogeneous and isotropic and that any baryon asymmetry has been erased.

Around $T \sim 10^{15}$ GeV $\sim M_X$, i.e the mass of superheavy gauge bosons that mediate Grand Unified Interactions, according to the usual lore of thinking in gauge theories a kind of phase transition occurs. We pass, from a phase where the group G that unifies all interactions is unbroken to a phase where G is broken down to SU(3) x SU(2) x U(1), the "observable" well-known group(s), that describe the usual strong and electro-weak interactions.

The question, that naturally emerges is, what kind of phase transition takes place, namely of the first or of second order ? This is still, an unanswered question, and at the same time a very fundamental one. There are some interesting features related to a first order phase transition. Since, the Universe is going to stay for some time in the false vacuum, the standard phenomenon of supercooling occurs, which, if it is of the right size ($T_{supercool}/T_{final} \sim 10^{-28}$!) may solve[3] simultaneously, the horizon, asymptotic flatness and abundance of monopoles, problems (3). Some people think though, that the price to pay may be too high, because once trapped in the false vacuum, a graceful exit may be impossible. Anyway, let's assume that the phase transition, of whatever order ended successfully.

Next thing in the row, is the creation of the baryon asymmetry. Grand unified theories contain all the necessary ingredients[4] (B, C, CP violation), and the existence of an expanding Universe helps in getting things out of equilibrium, so that a net baryon creation is possible[5] . We should not be suprised with the smallness of the effect, because remember that it has to involve higher orders so that CP-violation be active.

A super-naive estimate indeed gives

$$\frac{1}{s_B} \equiv \frac{n_B}{n_\gamma} \sim \alpha_G^q . \varepsilon \simeq 10^{-8} \qquad (1)$$

where α_G is a typical "fine GUT structure constant" ($\sim 1/50$), and ε some CP-violation parameter (may be $\varepsilon\ 10^{-3}$ (?)). The experimental fact that

$$n_B/n_\gamma \sim 10^{-9\pm1} \qquad (2)$$

indicates that (1) is a rather succesfull prediction of GUTS. (For a more detailed discussion see part II).

It should not escape our notice the fact that, GUTS, in strictly Robertson-Walker-Friedman (R-W-F) Universe, predict only adiabatic (or iso entropic) fluctuations. This is readily seen from eq (1), where the specific entropy per baryon (S_B) is expressed in terms of microscopic physics parameters, i.e S_B is a space-time independent constant.

As is well-know, the type of fluctuations adiabatic or isothermal, is very crucial in the theory of galaxy formation. If recent observation[6] on cosmic background radiation and their theoretical interpretation are correct, then it seems that we need also isothermal fluctuation[7]. The easiest way to marry the GUT program with isothermal fluctuations is to allow the very early Universe to deviate from a strictly R-W-F form by adding say inhomogeneous shear[8]. Then things work out O.K. Let me not go further on this very important subject since other participants in this meeting will discuss it in great detail.

Finally, it should be mentioned the severe constraints, that one puts on the form and properties of interactions and particles that "wake up" after the baryosynthesis has occured. They better not wipe-out any baryon asymmetry, if at the same time are important in creating any new baryon asymmetry. Such situation, occur some times in theories where new interactions are introduced at intermediate scales between 100 GeV and 10^{15} GeV. These kinds of interactions usually violate baryon numbers, but while they manage to keep the proton stable enough, they come to equilibrium at some scale much lower than say 10^{15} GeV, so that they wipe out everything without being able to build in a new asymmetry.

Moral : watch out "funny intermediate scale" interactions, that intend to bloom the desert or to create oases, they may as well be mirages !

Ways out from this problem will be discussed is part II.

Picture 4 : THE DULL PERIOD (?) (10^3 GeV < T < 10^{14} GeV)

In the most conventional picture, during this long-long period **NOTHING HAPPENS** (the desert hypothesis). The Universe just expands easy and tranquile. In part II of this talk, we will try to provide some possible excitement occuring in this

era in an effort to make it less dull.

Picture 5 : THE ELECTRO-WEAK ERA ($1\ \text{GeV} < T < 10^3\ \text{GeV}$)

The most dramatic effect in this era is the phase transition, that occurs between the phase where, the electro-weak group SU(2) x U(1) is unbroken to the phase where, SU(2) x U(1) is broken down to U(1) electromagnetic. Here, we better watch out the kind of the phase transition. If it turns out to be of the first order, we should make it sure that not too much entropy is created, otherwise our painstaking created baryon asymmetry will be unacceptably diluted. Even worse, at these "low energies" there is no potential agent to employ for the creation of the baryon asymmetry. Things are not arbitrary as they look at first sight. For example, the aesthetically appealing and maybe technically necessary, for solving the gauge hierarchy problem, Coleman-E.Weinberg mechanism[9] , i.e a radiatively triggered spontaneous breakdown of the electroweak group SU(2) x U(1), may be in grave trouble. John Ellis in his talk[10] at this meeting, discussed the limits on the dilution factor, that a GUT-created n_B/n_γ may suffer, by using a recently discovered[11] amusing connection, between the electric dipole of the neutron and n_B/n_γ the dilution factor should be at least less than say O(1000).

Picture 6 : THE STRONG INTERACTIONS ERA ($T \sim 1$ GeV)

Around $T \sim 1$ GeV the strong interactions are really "strong", such that quarks glue together and create the so highly desired baryons. Three quarks for Muster Mark.

Picture 7 : NUCLEOSYNTHESIS ($T \sim 1$ MeV)

One of the most startling successes of the standard Big-Bang cosmology, is indeed the prediction of the ~ 25 % (by mass) ^{4}He abundance. There are three basic tacit assumptions involved in these calculations :

i) There are no more that 3 or 4 species of neutrinos. GUTS constrain the number of species of neutrinos, through the succesful prediction (19) of the b quark mass in terms of the τ-lepton mass (or vis-versa !)

$$\frac{m_b}{m_\tau} \simeq 2.5\text{-}3 \qquad (3)$$

More than 3-4 species of neutrinos, or equivallently more than 6-8 quark flavours will undo the succesful eq (3).

ii) The neutrinos are almost massless (m_ν <<1MeV). Grand Unified theories predict that[1] , if neutrinos have some mass, then it is of the order of

$$m_\nu \simeq m^2/M \sim (10^{-5} \text{ eV} \rightarrow \text{few eV}) \qquad (4)$$

where m and M are scales related to "low energy" and "super-high energy" physics respectively.

iii) There is no neutrino degeneracy $[(n_\nu - n_{\bar{\nu}})/n_\gamma \sim 1]$
But, one of the main characteristics of GUTs is the democratic handling of quarks and leptons, which in this case imply[13] (modulo bizarre occassious)

$$\frac{n_B}{n_\gamma} \sim \frac{n_\nu - n_{\bar{\nu}}}{n_\gamma} \sim 10^{-8} \qquad (5)$$

GUTs indeed help indirectly nucleosynthesis.

Picture 8 : THE MATTER-RADIATION SPLIT ($T \sim 1$eV)

There are two main characteristics of this epoch. First, there is a transition from a Radiation dominated era to a Matter dominated era (ρ radiation becomes smaller than ρ matter).. Second, recombination occurs. The photons have no more enough energy to ionize the continously forming atoms out of protons, neutrons and electrons. Matter and radiation cannot be anymore in equilibrium, they decouple.
Needless to stress the importance of this splitting. From now on, radiation cannot smooth out any preexisted density fluctuation, which may now GROW. Clearly enough, the presently observed homogeneous and isotropic cosmic background radiation, red-shifted from the decoupling time, provides very useful information on the history of the Universe at the decoupling time.

Picture 9 : GALAXY FORMATION (10^{-2} eV < T < 1 eV)

Here, the situation is obviously more complicated. Still, GUTS may help in providing the form of perturbation adiabatic and/or isothermal), plus some possible ways to create such perturbations during phase transitions and even calculate

their strength. Another important GUT ingredient that may help here, is the fact that, neutrinos may have masses (see eq. 4) of the right order of magnitude ($\sim$ few eV), so that, they may play a fundamental role in this epoch. Other talks in this meeting cover fully this topic.

Picture 10 : HEAVY ELEMENTS,... , DNA-RNA/PROTEINS, ... HOMO SAPIENS

$$(10^{-4}\text{eV} \sim T < 10^{-2}\text{eV})$$

The story from now on is rather well-known. Heavy elements are created in the cores of the stars and diffused around by supernovae explosions. Later on, the first molecules are formed, eventually macromolecules are created and after that we are left in the capable hands of Watson and Crick. AND HERE WE ARE !

If GUTS are right, and protons do indeed decays, then even Woody Allen figured out in his "Stardust Memories", that our future is rather dark, because we will dissolve to the primer parts, that we all have been constructed off.

II.A NEW MECHANISM FOR BARYON PRODUCTION : THE CASE FOR SUPERHEAVY FERMIONS

One of the biggest successes of Grand Unified theories, is the quantitative understanding of the observed baryon asymmetry in the Universe. Since by now, this is a rather well-known subject[5] , I will just review quikly here the basic ingredients, in order to make easier the connection with the new proposed mechanism. Grand Unified interactions contain normally and naturally baryon number, CP and C violating components. Also, the expansion of the Universe make it easy for the abovementioned interactions, to go out from equilibrium. But then, we have all the necessary and sufficient conditions[4] to create a net baryon number.

What we believe that happened is someting like the following[5] : some superheavy particles (the Myrtons (A) of ref. 5) have been kept in a large number (proportional to n_γ) at temperatures below their masses

$$\left.\begin{aligned} n_A = n_{\bar{A}} &\sim n_\gamma \\ T &< M_A \end{aligned}\right\} \qquad (6)$$

Clearly, at some temperature T_A their decay rate will become equal or bigger than the expansion rate of the Universe. At that moment, they will decay like mad, but, the decay products because of the Boltzmann supression factor $\exp(-M_A/T_A) \ll 1$, (remenber $T_A \ll M_A$), will be unable to form back their parents. So, we almost have free decay of equal number of myrtons (A) and anti-myrtons ($\bar{A}$). (and actually $\sim n_\gamma$) which as we all know from our basic particle physics course, may create a baryon asymmetry given all the needed "violating" interactions. Numerous calculations[5] have shown, that the most likely candidate to be identified with the myrtons (A-particles) are the superheavy Higgs bosons (H). In this case, it is easy to see that eq.(6) is satisfied, if

$$\left.\begin{aligned} M_H &\geq O(\alpha^2\ M_{Planck}) \\ T_A &\geq O(10^{14}\ \text{GeV}) \end{aligned}\right| \qquad (7)$$

and in the "standard" 2 $\underline{5}$ of Higgs SU(5) model on gets (14)

$$n_B/n_\gamma \sim 10^{-2}.(\Delta B)_H \simeq 10^{-2}.(G_F.m_b^2)\ \varepsilon \qquad (8)$$

where $(\Delta B)_H$ is the CP-violating asymmetry in the decays of H and $\bar{H}$, $m_b \simeq 2$ GeV (the mass of the b quark at 10^{15} GeV), $G_F.m_p^2 \simeq 10^{-5}$ the usual Fermi constant and ε is a CP-violation parameter. We immediately see that eq.(8) is compatible and pretty close to the observed experimental number, as given by eq(2).
All these are fine. Though, some people may have hard time to swallow the fact that, all our existence depends on (superheavy) Higgs bosons, while even the low energy Higgsons seem to be trouble-makers (gauge hierarchy problem), and they have indeed very elusive properties. Some other people, may have difficulties to believe, that such an extremely early created baryon asymmetry, has not been erased by one way or the other, before coming down say to 1 GeV. For example, as discussed in Part I, the existence of intermediate interactions which violate Baryon number (and any global symmetry involving baryon number), make difficult the conservation of the already created baryon asymmetry.
May be, all such thoughts are too pessimistic, and everything at the end will turn out to be O.K, i.e there are no problems to have Higgsons around, and simply there are no intermediate scale, B-violating interactions around, but, let's keep an open mind and look for alternatives.
The most obvious candidate is of course the superheavy gauge bosons. Though,

there are rather severe problems related to the creation of baryon asymmetry, through superheavy gauge boson decays to n_B/n_γ is

$$n_B/n_\gamma \simeq (1.5.10^{-1}).(\Delta B)_G . \left[160\ (10^{15}\text{GeV}/M_G)\right]^{-1\,3} \qquad (8')$$

The factor in square parentheses in eq(8') represents the suppression effects of $2 \leftrightarrow 2$ interactions which tend to dilute the asymmetry generated by gauge boson decays and inverse decays. The factor $(\Delta B)_G$ is the CP-violating asymmetry in the decays of gauge particles and antiparticles. The parenthesized suppression factor in eq(8') is quite fierce if $M_G \simeq 6.10^{14}$ GeV as expected in minimal SU(5). Also in simple models, $(\Delta B)_G$ of eq(8') turns out to be one order higher in α_G than $(\Delta B)_H$ of eq(8). Moreover eq(8) does not involve any suppression factor, at least for a physically interesting range of mass and coupling parameters. Simply, superheavy gauge bosons seem to be a disfavorite candidate for the creation of baryons.

What's left ? A very natural candidate : superheavy fermions. I will like now to discuss some work (15) that R. Barbieri, A. Masiero and myself have being doing at CERN, concerning a possible mechanism to create baryons through superheavy fermion decays. Grand Unified theories, except the minimal SU(5) model, contain naturally superheavy fermions. Even the minimal SU(5) model, may be trivially extended to include superheavy fermions. After all, if there are low and superheavy mass gauge bosons, low and superheavy mass Higgs bosons, why not low and superheavy mass fermions ?

Let us concentrate on a special category of superheavy fermions, that they decay mainly to three light fermions, through the exchange of a superheavy Higgs boson. The reader may envisage it, as a kind of $\mu \to e\bar{\nu}_e\nu_\mu$ decay with the only difference that the W-boson is replaced by a Higgs boson. In this case the width is given by.

$$\Gamma_F = \left(\frac{\lambda^2}{\mu_M}\right)^2 \times \frac{m_F^5}{12\pi\ M_H^4} \qquad (9)$$

in an obvious notation, where λ indicates an average universal Higgs coupling, m_F is the mass of the superheavy fermion and $M_H \simeq 10^{15}$ GeV is the mass of the exchanged superheavy Higgs boson.

Then we distinguish the following stages in the history of the superheavy fermions, during the cooling of the early Universe :

i) Superheavy fermions in equilibrium.

In general, the two body interactions mediated by superheavy bosons with rates $\simeq \alpha_G^2 \, T^5/M^4$, will remain in equilibrium up to a temperature

$$T_o' \simeq M\left(\frac{1.6.N^{1/2}}{\alpha_G^2}\,\frac{M}{M_P \ell}\right)^{1/3} \simeq 10^{15}\ \text{GeV} \tag{10}$$

where the expansion rate of the Universe

$$H = \frac{(1.6)\ T^2\ N^{1/2}}{M_P \ell} \tag{11}$$

becomes dominant. Here, N = 10-100 is the suitably weighted number of all particles with masses $< T_o'$, and $M \simeq 10^{15}$ GeV is a typical Grand Unified mass scale. Since some of these interactions will in general be B-violating, one expects no net baryon (or antibaryon) survival below T_o' *. If the superheavy fermions F are not SU(3) x SU(2) x U(1) singlets, will still be kept in equilibrium below T_o', or even, below their mass m_F, by annihilation and compton scattering with "light" gauge bosons. In this case eq.(10) should be replaced by (assuming $m_F < 10^{14}$ GeV)

$$\begin{aligned} T_o &\simeq m_F \cdot \left[\ell_n Q + 1/2\, \ell_n\ \ell_n Q\right]^{-1} \\ Q &\equiv \alpha_G^2 \cdot \left(\frac{M}{8N}\right)^{1/2} \cdot \frac{1}{M^2} \cdot \frac{M_{Pe}}{m_F} \end{aligned} \tag{12}$$

while the relative number density is given by

$$\left.\frac{n_F}{n_\gamma}\right|_{T^\circ} \simeq Q^{-1}\, \ell_n\ Q \tag{13}$$

---

*Footmote : It is a remarkable property of superheavy fermion interactions to violate, in general, any global symmetry involving the B-number. With superheavy fermions around, it is difficult to keep alive any primordial baryon asymmetry.

In the following we will use eqs(12) and (13), as being more realistic and general.

ii) Superheavy fermions out of equilibrium.

At later times ($T < T_o$), until they decay, the relative number density of the superheavy fermions (F) (eq. 13) remains constant, with their energy density being dominated by their rest mass, $\varrho_F \simeq n_F \cdot m_F$.
Actually, at $T \simeq m_F/3N$, this same energy density (ϱ_F) will start dominating the full energy density of the Universe, up to the temperature T_F of their decays.
So, if this scenario has something to do with reality, it introduces a new epoch in the early Universe for $T_F < T < m_F/3N$, where basically the Universe is Matter dominated !
This new era may have some consequences on the subsequent evolution of the Universe.

iii) Superheavy fermions decay

When the time comes, i.e when

$$H \simeq \left(\frac{8M}{3M_{Pl}^2}\ \varrho_F\right)^{1/2} = \left(\frac{8M}{3M_{Pl}^2} \cdot \frac{m_F \cdot T^3}{M^2}\ \left.\frac{n_F}{n_\gamma}\right|_{T_o}\right)^{1/2} \qquad (14)$$

falls below the decay rate given by eq.(9), superheavy fermions will decay. This happens at a temperature

$$T_F \simeq 1/10 \cdot \left(\frac{\lambda^a}{4\pi}\right)^{4/3} \left(\frac{m_F}{M}\right)^3 (M\, M_{Pl}^2)^{1/3} \left(\left.\frac{n_F}{n_\gamma}\right|_{T_o}\right)^{-1/3} \qquad (15)$$

In these decays a generation of baryon asymmetry is expected. In fact, the general scenario analyzed in the beginning is applicable. F's can be identified as myrtons and indeed one finds

$$\frac{n_B^{\sim}}{n_\gamma} \simeq (10^{-3} - 10^{-9}) \cdot \left.\frac{n_F}{n_\gamma}\right|_{T_o} \qquad (16)$$

depending on the parameters of the particular model. Though, one should keep in mind, that, since $\varrho(T_F)$ is dominated by the rest mass of the F's, their decay

will also reheat the Universe up to a temperature T_F', so that.

$$\rho(T_F) \simeq \frac{m_F T_F^3}{M^2} \left.\frac{n_F}{n_\gamma}\right|_{T_o} \simeq \frac{3N T_F'^4}{M^2} \qquad (17)$$

and in turn dilute the baryon asymmetry by a factor

$$d \simeq \left(\frac{T_F}{T_F'}\right)^3 = \left(\frac{3N T_F}{m_F}\right)^{3/4} \left(\left.\frac{n_F}{n_\gamma}\right|_{T_o}\right)^{3.4} \qquad (18)$$

The final observed asymmetry will then be given by

$$\frac{n_B}{n_\gamma} = \frac{\tilde{n_B}}{n_\gamma} \cdot d \simeq (10^{-3}\text{-}10^{-9}) \left(\frac{3N T_F}{m_F}\right)^{3/4} \left.\frac{n_F}{n_\gamma}\right|_{T_o}^{1/4} \equiv (10^{-3}\text{-}10^{-9})\, d' \qquad (19)$$

Demanding that, the factor d' should not be less then 10^{-5}, and using eqs.(13), (15) and (19), we put a lower limit on m_F, as well as, on T_F :

$$\begin{aligned} m_F &\gtrsim 10^{11}\text{-}10^{12} \text{ GeV} \\ T_F &\gtrsim 10^{3}\text{-}10^{4} \text{ GeV} \end{aligned} \qquad (20)$$

where a typical value $\lambda^2/4\pi \sim 10^{-4}\text{-}10^{-5}$ has been taken. We feel that our expectations have been fulfilled.

We notice that, we may create a baryon asymmetry through F's at temperatures as low as $T_F \simeq 1\text{TeV}$

Even if Higgsons do not exist, or even if there are dangerous "wiping out" intermediate scale interactions, still there is hope inside GUTS, to create a respectable baryon asymmetry at a very low energy scale ($\sim$ 1 TeV), through superheavy fermions.

I believe, we don't have to take the extremes. The following scenario looks to me highly probable. As the Universe cools down, F's and H's eventually get out from equilibrium (in the case of F's at least their B-violating interactions get out of equilibrium). Then, the H's decay and may (if we have a complicated enough Higgs structure) create a substantial baryon asymmetry. As the Universe

cools further down, intermediate scale B-violating interactions (if they exist), may partially or completely erase any preexisted baryon asymmetry. Eventually the Universe cools down to a temperature that the relevant F's decay and provide through their decays some baryon asymmetry. This baryon asymmetry should be added up (hopefully constructively) to any preexisted baryon asymmetry, to make up what we observe today. GUTs have enough complexity to cover any out-of hand case, and provide enough baryons to build up our Universe.
Finally, I would like to close my talk with a few comments on the specific new mechanism proposed here.

1) The key point is that we should have around very-long lived superheavy particles, which make the important condition imposed by eq.(6), trivial to satisfy, and at the same time creating baryons at somehow "low temperatures" ($\sim$ 1 TeV), so that, subsequent "accidents" are easier to avoid.

2) Very long-lived particles are not strangers. Remember the protons ! Actually, one may push the analogy further. May be, at some point there was a microscopic copy of our Universe. Galaxies and even "people" made up from "Superheavy protons" were around, etc, etc...
The very existence of a "Superheavy Matter" dominated era, as an intermentzo in a radiation dominated Universe, may have some important consequences on the subsequent evolution of the Universe.

3) I would also like to stress the fact that the occurence of the scale of 1TeV not for above the electo-weak phase transition, may not be accidental. Since, our estimates are rather naive, one may think what will happen, if, superheavy fermions may live up to the moment, where the electroweak phase transition takes place. F's, may even help a first order phase transition to occur faster... There are some amusing consequences, that they deserve some futher thinking.

4) If indeed, superheavy fermions are the main source of baryon production, then, the recent connection[11] between the electric dipole of the neutron and the observed baryon asymmetry, does not hold at least in its present form.
On the other hand, if this connection will be found to be experimentally unacceptable, and the other tacit assumptions in its derivation hold, then we will have A CASE FOR THE SUPERHEAVY FERMIONS.

ACKNOWLEDGEMENTS

I would like to thank Jean Audouze and Tran Thanh Van for organizing such a stimulating meeting.

REFERENCES

1) For a review of Grand Unified Theories see
D.V. Nanopoulos, "Grand Unified Models" in :
The Proceedings of the XV Rencontre de Moriond, Vol. II, ed. by J. Tran Thanh Van, p. 427 (1981).

2) J. Ellis, M.G. Gaillard and D.V. Nanopoulos, Phys. lett. 90B, 253 (1980).

3) A.H. Guth, Phys. Review D23, 347 (1981).

4) A.D. Sakharov, Pis'ma Zh. Eksp. Teor. Riz. 5, 32 (1967).

5) For a general review see
D.V. Nanopoulos,"Cosmological Implications of GUTS", in :
Progress in Particle and Nuclear Physics, Vol. 6, ed. by Sir D. Wilkinson (Pergamon Press, p. 23.

6) G.F. Smoot et al. Phys. Rev. Lett. 34, 898 (1977).
R. Fabri et al. Phys. Rev. Lett. 44, 1563 (1980).
S.P. Boughon et al, Ap. J. Lett. 243, 113 (1981).

7) J. Silk and M.L. Wilson, Ap. J. Lett. 244, 37 (1981).

8) J.R. Bond, E.W. Kolb and J. Silk, U.C. Berkeley preprint (1981).
J.D. Barrow and M.S. Turner, EFI-preprint -81-02 (1981).

9) S. Collman and E. Weinberg, Phys. Rev. D7, 1888 (1973).

10) J. Ellis, talk in these proceedings.

11) J. Ellis, M.K. Gaillard, D.V. Nanopoulos and S. Rudaz,
Phys. Lett. 99B, 101 (1981) ; CERN preprint TH-3056 (1981).

12) A.J. Buzas, J. Ellis, M.K. Gaillard and D.V. Nanopoulos,
Nucl. Phys. B135, 66 (1978)
D.V. Nanopoulos and D.A. Ross, Nucl. Phys. B157, 273 (1979)

13) D.V. Nanopoulos, D. Sutherland and A. Yildiz,
Lett. Nuovo Cim. 28, 205 (1980).

14) D.V. Nanopoulos and S. Weinberg, Phys. Rev. D20, 2484 (1979).

15) R. Barbieri, A. Masiero and D.V. Nanopoulos, Phys. Lett. 98B, 191 (1981).

COSMOLOGY AND THE NEUTRON ELECTRIC DIPOLE MOMENT

John Ellis
CERN - Geneva

ABSTRACT

There is a contribution to the neutron electric dipole moment d_n from the CP violating θ vacuum parameter of QCD. Diagrams analogous to those responsible for the baryon number of the universe also contribute to θ, providing an order of magnitude lower bound on d_n in terms of the baryon-to-photon ratio n_B/n_γ. GUTs sufficiently complicated to explain the observed n_B/n_γ predict that d_n should be close to the present experimental upper limit. The comparison between d_n and n_B/n_γ gives us information about entropy generation after the epoch of baryon generation in the very early Universe.

1. - INTRODUCTION

Much experimental effort [1),2)] has been devoted to the search for an electric dipole moment of the neutron, d_n, which violates CP. Detection of a non-zero d_n would add much to our knowledge of CP violation, confined hitherto to the K^o-$\bar{K}^o$ system [3)], and the present experimental limit [2)]

$$d_n < 6 \times 10^{-25} \text{ e-cm} \tag{1}$$

is already a strong constraint on different models of this phenomenon [1)]. Theoretical analysis of d_n has proceeded along two lines, the first being calculation of perturbation theory diagrams in CP-violating weak interaction models. More recently it has been realized [4)] that non-perturbative effects in QCD will violate CP if a hitherto unremarked parameter θ_{QCD} is non-zero. While it is not renormalized by the strong interactions alone, θ_{QCD} is in general renormalized by the non-strong interactions, even if its bare value is zero.

Mary Gaillard, Demetre Nanopoulos, Serge Rudaz and I have recently emphasized [5)] that there are contributions to θ_{QCD} from Higgs exchange diagrams in GUTs, which are closely analogous to those responsible for the baryon number of the Universe produced in the decays of very heavy Higgs particles [6)]. These contributions to θ_{QCD} give us an order of magnitude lower bound on d_n of

$$d_n \gtrsim 2.5 \times 10^{-18} \left(\frac{n_B}{n_\gamma}\right) \text{e-cm} \tag{2}$$

If we take from astrophysics [7)] a lower bound on n_B/n_γ of 2×10^{-10} then we find that

$$d_n \gtrsim 5 \times 10^{-28} \text{ e-cm} \tag{3}$$

to be compared with the upper bound (1).

In this talk I will review our work [5),8)], starting with a reminder of the meaning of θ_{QCD}, its relation to Higgs couplings and its renormalization [9)]. Then we will see how n_B/n_γ is believed to be generated in CP-violating decays of Higgs bosons [6)] via diagrams similar to some contributing to θ_{QCD}. In fact, in many models θ_{QCD} and hence d_n is much larger [5),8)] than the lower bound (2) or (3). Finally we will see how the experimental upper limit [2)] on d_n

already gives us useful information on the amount of entropy that could have been generated subsequent to the creation of a baryon-antibaryon asymmetry. The bound (1) already poses problems [10] for the scenario of symmetry breaking by radiative corrections in the Weinberg-Salam model, and constrains [8] deviations from the usual Robertson-Walker-Friedmann (RWF) big bang cosmology with some implications for models of galaxy formation [11].

2. - THE QCD θ PARAMETER

In perturbation theory, QCD automatically conserves CP [12]. However, there is a possible term [4] in the QCD Lagrangian

$$\mathcal{L}_{QCD} \ni \theta_{QCD} \frac{g^2}{32\pi^2} G^a_{\mu\nu} \tilde{G}^{a\mu\nu} \tag{4}$$

where $G^a_{\mu\nu}$ is the gluon field strength tensor and $\tilde{G}^{a\mu\nu}$ is its dual. The θ_{QCD} term (4) is a total derivative and hence does not show up in perturbation theory, but it is non-zero for non-perturbative field configurations such as instantons [4]. It conserves charge conjugation C because it is the product of two gauge field strengths, but it violates parity P because the dual tensor $\tilde{G}^{a\mu\nu}$ contains an antisymmetric $\varepsilon_{\mu\nu\rho\sigma}$ symbol. Therefore the θ_{QCD} term (4) violates CP.

However, it is not directly related to the CP violation seen [3] in the K^o-$\bar{K}^o$ system since there it occurs in an effective interaction with $|\Delta S| = 2$, whereas the interaction (4) clearly conserves strangeness. It will, however, contribute to the neutron electric dipole moment d_n, in an amount estimated [13] as

$$d_n \lesssim 4 \times 10^{-16} \theta_{QCD} \text{ e-cm} \tag{5}$$

from both MIT baggery and chiral perturbation theory techniques. Using the most recent [2] upper limit (1) on d_n we infer that

$$\theta_{QCD} \lesssim 1.5 \times 10^{-9} \tag{6}$$

[This upper bound on θ_{QCD} means that the $O(10^{-3})$ CP violation in the K^o-$\bar{K}^o$ system cannot be due to a combination of θ_{QCD} and a CP conserving $|\Delta S| = 2$ interaction.]

For comparison, the best upper limit on d_n by an American group [1] is $d_n \lesssim 3\times10^{-24}$ e-cm, and they are now starting an extensive series of experiments to improve this by as many as three orders of magnitude. A major problem of theoretical physics is to determine why θ_{QCD} is so small (6).

In pure QCD, θ_{QCD} is an arbitrary parameter which is not renormalized by the strong interactions but has no natural reason to be small or zero. When one takes into account the non-strong interactions θ_{QCD} is in general renormalized [9]. The most obvious contribution to renormalization of θ_{QCD}, which seems to give the correct order of magnitude in general, arises from the renormalization of the quark matrix which is necessary at each order of perturbation theory in the non-strong interactions. If we define an effective quark mass parameter $m(q)$ from the inverse quark propagator :

$$S_F^{-1}(q) = \not{q} - m(q) \tag{7}$$

then it gets renormalized in each order of perturbation theory by diagrams shown generically in Fig. 1a. The renormalized mass matrix m^{Ren} should then be restored to real and diagonal form at each order in perturbation theory. This entails phase rotations on the left- and right-handed quark fields which feed through via the QCD U(1) anomaly into a renormalization of θ_{QCD} by an amount

$$\delta\theta_{QCD} = \arg \det m^{Ren} \tag{8}$$

The leading contribution to $\delta\theta_{QCD}$ comes from one particle irreducible diagrams F_1 making a transition between left- and right-handed quarks :

$$\delta\theta_{QCD} = \arg \det m^{Ren} \approx \mathrm{Im}\ \mathrm{Tr}(m_0^{-1} F_1) \tag{9}$$

where m_0 is the zeroth order quark mass matrix. Since the effective mass parameter m^{Ren} (7) is a function of the momentum scale q, so also will be $\delta\theta_{QCD}$. The observation (or non-observation) of d_n constrains the effective value of θ_{QCD} renormalized at a momentum scale of order 1 GeV or so. Presumably there is an eventual theory of everything (TOE) which fixes θ_{QCD} at some enormous scale $q \gtrsim O(10^{19})$ GeV. Between this enormous scale and the scale of the neutron θ_{QCD} is renormalized by all the non-strong interactions, and we can distinguish [5] the following contributions :

$$\theta_{QCD}(1\ \text{GeV}) = \theta_{TOE} + \delta\theta_{GUT} + \delta\theta_{KM} \tag{10}$$

We denote by $\delta\theta_{GUT}$ the renormalization due to GUTs at energy scales $\approx 10^{15}$ GeV, and by $\delta\theta_{KM}$ the renormalization due to conventional six-quark Kobayashi-Maskawa (KM) weak interactions at momentum scales $< 10^{15}$ GeV. These latter effects have been estimated [9] as

$$\delta\theta_{KM} = O(10^{-16}) \tag{11}$$

resulting in a contribution to d_n of order 10^{-31} to 10^{-32} e-cm. This contribution to d_n is in fact dwarfed by the direct contribution to d_n from perturbation theory in the strong and weak interactions which has been estimated [9),14)] as

$$\delta d_n = m_{u,d} \left(\frac{\alpha_s}{\pi}\right)^2 \left(\frac{\alpha}{\pi}\right)^2 \frac{m_c^2}{m_W^4} S_1^2 S_2 S_3 \sin\delta \tag{12}$$

$$\approx 10^{-30\pm1}\ \text{e-cm}$$

in the standard KM model. We see that the standard KM weak interaction contributions to d_n are very small.

As yet we do not know how to estimate θ_{TOE}, but we can calculate $\delta\theta_{GUT}$. This is derived from the quark mass matrix, which is in turn given by a Higgs vacuum expectation value multiplied by a Higgs coupling matrix H_1 as illustrated in Fig. 1 : $F_1 = v H_1$. For convenience we will work directly with the coupling matrix of this so-called m-Higgs. We will see that in general a GUT sufficiently complicated to explain the observed n_B/n_γ will yield a $\delta\theta_{GUT} >> \delta\theta_{KM}$. In the absence of a TOE we will assume that there is no conspiratorial cancellation between θ_{TOE} and $\delta\theta_{GUT}$ so that in order of magnitude

$$\theta_{QCD}\ (1\ \text{GeV}) \gtrsim \delta\theta_{GUT} \tag{13}$$

and we will now go on to consider the relation of $\delta\theta_{GUT}$ to n_B/n_γ in some detail.

3. - THE COSMIC CONNECTION

In his talk at this meeting, Demetre Nanopoulos [6] has already described to you the mechanisms offered by GUTs for generating a baryon-antibaryon asymmetry in the very early Universe. It seems most likely that the net baryon number originates from the C- and CP- violating decays of some very heavy particles, typically gauge bosons, Higgs bosons or fermions with masses $O(10^{14}$ to $10^{15})$ GeV. Figure 2 shows how [15] a net baryon asymmetry of order 10^{-9} can be built up at high temperatures if one postulates heavy particles X with a suitable B-, C- and CP-violating asymmetry in their decays. Superheavy fermions are not present in the simplest GUTs, though they are commonplace in larger ones [6]. Most quantitative analyses of baryon number generation have focused on the decays of gauge and/or Higgs bosons, and we shall concentrate on them here. Their contributions to n_B/n_γ in simple models have been calculated [16] numerically to be

$$(n_B/n_\gamma) \approx \begin{cases} 1.5 \times 10^{-1.1}\, \epsilon_G \left[160 \left(\frac{10^{15}\text{GeV}}{m_G}\right)\right]^{-1.3} & \text{for gauge bosons} \\ 0.5 \times 10^{-1.1}\, \epsilon_H & \text{for Higgs bosons} \end{cases} \qquad (14)$$

The factor in square parentheses for gauge bosons in (14) represents the suppression effects of 2-2 interactions which tend to dilute the asymmetry generated by gauge boson decays and inverse decays. There are no such factors for Higgs bosons in (14) if $m_H > 2\times10^{14}$ GeV and the $g_H^2/4\pi \equiv \alpha_H \approx 6\ 10^{-4}$ as seems plausible. The factors ϵ_G and ϵ_H in (14) are the CP-violating asymmetries in decays of gauge and Higgs particles and antiparticles. In simple models [6] ϵ_G is one order higher in α_{GUT} than is ϕ_H. Furthermore, the parenthesized suppression factor in (14) is quite fierce if $m_G \approx 6\times10^{14}$ GeV as expected [17] in minimal SU(5). We are therefore inclined [6] to expect that Higgs boson decays would dominate the production of a net baryon asymmetry, and estimate

$$(n_B/n_\gamma) \lesssim (10^{-1} \text{ to } 10^{-2})\ \epsilon_H \qquad (15)$$

Let us call the Higgs boson with the CP-violating decay asymmetry ϵ_H the d Higgs. The lowest order in which ϵ_H can become non-zero is fourth-order, via diagrams like those in Fig. 3a, which appear in models with ≥ 2 Higgs multiplets coupling to fermions. An individual contribution to the decay asymmetry ϵ_H can be written as the product of a Higgs emission vertex a_d and an absorption vertex b_d^+, at least one of which contains a final state interaction. These are represented in Fig. 3b alongside the typical lowest order example of Fig. 3a. The

fermion lines in Fig. 3 must be summed over all fermion generation indices yielding us the trace of the product I, of coupling matrices indicated in Fig. 4a. This gives us a decay asymmetry

$$\varepsilon_H \gtrsim \frac{\mathrm{Im\ Tr}(I_1)}{T_r(h_d h_d^+)} = \frac{\mathrm{Im\ Tr}(a_d b_d^+)}{T_r(h_d h_d^+)} \tag{16}$$

where h_d is the lowest order coupling matrix of the d Higgs. In general the d Higgs is a subset of a GUT multiplet ϕ_d of Higgs fields, e.g., a colour triplet $\underline{3}$ in a five-dimensional representation of SU(5) Higgses. The m Higgs is also in general a subset of a GUT multiplet ϕ_m, e.g., an SU(2) doublet $\underline{2}$ in a $\underline{5}$ of SU(5). In the simplest minimal SU(5) with just one $\underline{5}$ of Higgs fields ϕ, the m and d Higgses must be partners in the same GUT multiplet ϕ. However, this is not in general true but there will usually be a non-trivial overlap between the SU(5) multiplet ϕ_m containing the m Higgs and that (ϕ_d) containing the d Higgs. In a basis where the d Higgs is pure :

$$\phi_m = \sum_j \phi_j U_{jm} \tag{17}$$

where we will suppose that U_{dm} is non-zero and of order one. We recall from Eq. (9) that $\delta\theta_{GUT} \gtrsim \mathrm{Im\ Tr}(m_o^{-1} F_1)$, which we can rewrite in terms of the Higgs coupling matrix H_1 as

$$\delta\theta_{GUT} \gtrsim \mathrm{Im\ Tr}\ (h_m^{-1} H_1) \tag{18}$$

where h_m is the lowest order coupling matrix of the m Higgs. Figure 4b demonstrates that there is a contribution to H_1 of the form

$$(h_m a_d b_d^+) \times |U_{dm}|^2 \tag{19}$$

which we can insert into Eq. (18) finding that

$$\delta\theta_{GUT} \gtrsim |U_{dm}|^2 \mathrm{Im\ Tr}\ (h_m^{-1} h_m a_d b_d^+) \tag{20}$$
$$\gtrsim \mathrm{Im\ Tr}(a_d b_d^+)$$

A specific example of this lower bound to the renormalization of θ_{QCD} by GUTs is shown in Fig. 5 : part (a) shows the fourth order contribution to (n_B/n_γ) while part (b) shows a corresponding contribution to the renormalization of θ_{QCD}.

Combining the expectations (13) and (20) we have

$$\theta_{QCD}(1GeV) \gtrsim \text{Im Tr}(a_d b_d^+) \tag{21}$$

and the expression (16) for ε_H therefore means that

$$\theta_{QCD}(1GeV) \gtrsim \varepsilon_H \times \text{Tr}\ (h_d h_d^+) \tag{22}$$

It seems reasonable to suppose that an order of magnitude estimate of $\text{Tr}(h_d h_d^+)$ is given by the corresponding quantity $\text{Tr}(h_m h_m^+)$:

$$\text{Tr}(h_d h_d^+) \simeq \text{Tr}(h_m h_m^+) \gtrsim \sqrt{2}\ G_F\ m_t^2 \gtrsim 6 \times 10^{-4} \tag{23}$$

Using the results (15) of numerical calculations we therefore deduce from (22) and (23) that

$$\theta_{QCD}(1GeV) \gtrsim 10\left(\frac{n_B}{n_\gamma}\right) \text{Tr}\ (h_d h_d^+) \tag{24}$$

$$\gtrsim 6 \times 10^{-3}(n_B/n_\gamma)$$

Putting in the connection (5) between θ_{QCD} and d_n we finally reach [5] the promised bound (2) on d_n, which leads in turn to the cosmological lower bound (3) of

$$d_n \gtrsim 5 \times 10^{-28}\ \text{e-cm}$$

if we accept the astrophysical estimate [7] that $(n_B/n_\gamma) \geq 2\times10^{-10}$.

It is worth mentioning [5),8)] that this bound is not saturated in many of the models we have investigated. For example, in minimal SU(5) with a single $\underline{5}$ of Higgs the renormalization of θ_{QCD} takes place in order $(\alpha/\pi)^2$, whereas $\text{Im Tr } I_1 \neq 0$ only in eighth order. As for SU(5) with two or more $\underline{5}$ of Higgs, $\delta\theta_{GUT}$ is of order (α_H/π) while $\text{Im Tr } I_1$ is of order $(\alpha_H/\pi)^2$. In SU(5)

with a $\underline{5}$ and a $\underline{45}$ of Higgs, the lowest order contribution to $\delta\theta_{GUT}$ is $O((\alpha/\pi)^{-1})$ larger than our bound. It may well be, therefore, that d_n is $O((\alpha/\pi)^{-1})$ larger than the lower bound (3), and hence within an order of magnitude of the present experimental upper limit, in any GUT which is sufficiently complicated to explain the observed baryon-to-photon ratio.

4. - A COSMIC SEISMOMETER

The neutron electric dipole moment is unique among observable low-energy CP-violating parameters in being sensitive to aspects of physics at short distance scales and high energy scales up to 10^{15} GeV and beyond. From the point of view of a cosmologist, high energies translate into high temperatures and hence very early times. The neutron electric dipole moment is therefore unique in its potential for probing CP-violating processes at these very early times, and whatever may have happened subsequently during the expansion of the Universe. In particular, since the bound on θ_{QCD} relates it to the primordial generation of baryon number, any subsequent generation of an extra factor E of entropy would dilute the observed n_B/n_γ by $1/E$ relative to the GUT calculation. The bound on d_n would therefore be a factor E larger :

$$d_n \gtrsim 2.5 \times 10^{-18}\, E\left(\frac{n_B}{n_\gamma}\right) \tag{25}$$

The experimental upper bound [2] on d_n already constrains E rather severely :

$$d_n \lesssim 6 \times 10^{-25}\ \text{e-cm} \Rightarrow \left(\frac{n_B}{n_\gamma}\right) \lesssim 2.4 \times 10^{-7}/E \tag{26}$$

and if we accept that $(n_B/n_\gamma) \geq 2\times10^{-10}$ we deduce [8]

$$E \lesssim 1.2 \times 10^{3} \tag{27}$$

Thus there is not much scope after the creation of the baryon asymmetry for entropy generation, e.g., during one of the subsequent phase transitions in the Universe. For example, it has been pointed out [10] that if the $SU(2)\times U(1)$ Weinberg-Salam symmetry is spontaneously broken by radiative corrections, then one should expect a strongly first order phase transition with much supercooling. It has been

estimated [10] that during the reheating of the Universe subsequent to the phase transition a factor $E \geq 10^4$ of entropy would be generated. Our bound (27) may therefore mean trouble for the scenario of Weinberg-Salam symmetry breaking by radiative corrections. This problem would be exacerbated if the present experimental upper bound on d_n were improved, or if the favoured GUT turns out to be one in which θ_{QCD} is renormalized by a graph of lower order than those related to the baryon number in Section 3.

Another example of an irregularity in the early Universe that is probed by the neutron electric dipole moment is the inhomogeneous shear proposed by some authors [11] as a way of generating isothermal density fluctuations which could give rise to galaxies. If the Universe were shear-dominated at the time of baryon generation, grand unified viscosity [18] would help to damp it down shortly afterwards, and some entropy would be generated during this dissipation. People have pointed out previously that the value of (n_B/n_γ) now should enable one to bound the amount of primordial shear. Our connection with d_n enables us to establish a quantitative bound, stating that the energy density at the epoch of baryon generation cannot have been shear-dominated by a factor Σ larger than $O(10^{18})$, about twenty orders of magnitude better than the previous best limit from cosmological nucleosynthesis [19]. This does not, however, rule out the scenario of galaxy formation from isothermal fluctuations due to shear inhomogeneities, which can work [11] if $\delta\Sigma$ (and hence Σ) is as small as 10^{-3}.

These examples may serve to demonstrate the utility of our cosmic connection [5),8] between baryon generation and the neutron electric dipole moment. This low energy observable is unique in being related to physics at ultra-high energies. As we gradually refine our knowledge of d_n, (n_B/n_γ) and narrow down the field of possible GUTs this connection will become ever more quantitative and restrictive.

FLOREAT THE INTERPLAY BETWEEN PARTICLE PHYSICS AND ASTROPHYSICS !

ACKNOWLEDGEMENTS

This talk is based on work done in collaboration with Mary K. Gaillard, Demetre Nanopoulos and Serge Rudaz. It is a true pleasure to thank them for the enjoyable times we have had working together. Also I thank Jean Audouze, Phil Crane, Tom Gaisser and Dan Hegyi for arranging such a stimulating opportunity to discuss with astrophysicists, and J. Silk for interesting conversations. Finally thanks are due to Tran Thanh Van for making such meetings possible.

REFERENCES

1) W.B. Dress, P.D. Miller, J.M. Pendlebury, P. Perrin and N.F. Ramsey - Phys. Rev. D15 (1977) 9 ;
See also : N.F. Ramsey - Physics Reports 43C (1978) 409.

2) I.S. Altarev et al. - Pis'ma Zh.Eksp.Teor.Fiz. 29 (1979) 794 and Leningrad Nuclear Physics Institute Preprint 636 (1981).

3) J.H. Christensen, J.W. Cronin, V.L. Fitch and R. Turlay - Phys.Rev.Letters 13 (1964) 138.

4) G. 't Hooft - Phys.Rev.Letters 37 (1976) 8 and Phys.Rev. D14 (1976) 3432 ;
R. Jackiw and C. Rebbi - Phys.Rev.Letters 37 (1976) 172 ;
C.G. Callan, R.F. Dashen and D.J. Gross - Phys.Letters 63B (1976) 334.

5) J. Ellis, M.K. Gaillard, D.V. Nanopoulos and S. Rudaz - Phys.Letters 99B (1981) 101.

6) See the talk of D.V. Nanopoulos at this meeting for a review, and also :
J. Ellis, M.K. Gaillard and D.V. Nanopoulos - "Unification of the Fundamental Particle Interactions", ed., S. Ferrara, J. Ellis and P. van Nieuwenhuizen (Plenum Press, N.Y., 1980), p. 461.

7) D.N. Schramm - Enrico Fermi Institute Preprint EFI-81-03 (1981), and talk at this meeting based on work to be published with J. Yang, K.A. Olive and G. Steigman.

8) J. Ellis, M.K. Gaillard, D.V. Nanopoulos and S. Rudaz - CERN Preprint TH.3056/LAPP TH-34 (1981).

9) J. Ellis and M.K. Gaillard - Nuclear Phys. B150 (1979) 141.

10) E. Witten - Nuclear Phys. B177 (1981) 477 ;
A.H. Guth and E.J. Weinberg - Phys.Rev.Letters 45 (1980) 1131.

11) J.D. Barrow and M.S. Turner - Enrico Fermi Institute Preprint EFI-81-02 (1981) ;
J.R. Bond, E.W. Kolb and J. Silk - "The generation of isothermal perturbations in the early Universe", U.C. Berkeley Preprint (1981).

12) S. Weinberg - Phys.Rev.Letters 31 (1973) 494 ;
D.V. Nanopoulos - Nuovo Cimento 8 (1973) 873.

13) V. Baluni - Phys.Rev. D19 (1979) 2227 ;
R.J. Crewther, P. Di Vecchia, G. Veneziano and E. Witten - Phys.Letters 88B (1979) 123 ; E 91B (1980) 487.

14) See also ;
J. Ellis, M.K. Gaillard and D.V. Nanopoulos - Nuclear Phys. B109 (1976) 213 ;
E.P. Shabalin - Yad.Fiz. 28 (1978) 151 ;
B.F. Morel - Nuclear Phys. B157 (1979) 23 ;
D.V. Nanopoulos, A. Yildiz and P.H. Cox - Phys.Letters 87B (1979) 53 and Ann.Phys. (N.Y.) 127 (1980) 1.

15) E.W. Kolb and S. Wolfram - Phys.Letters 92B (1980) 217 and Nuclear Phys. B127 (1980) 224.

16) J.N. Fry, K.A. Olive and M.S. Turner - Phys.Rev.Letters 45 (1980) ; Phys.Rev. D22 (1980) 2953, 2977.

17) See my talk at the Particle Physics Meeting held in parallel with this one.

18) J. Ellis, M.K. Gaillard and D.V. Nanopoulos - Phys.Letters 90B (1980) 253.

19) J.D. Barrow - Mon.Nat.Roy.Ast.Soc. 175 (1976) 359 ;
D.W. Olson and J. Silk - Ap.J. 226 (1978) 50.

FIGURE CAPTIONS

Figure 1 (a) Diagrams contributing to the renormalization of the quark mass matrix and thence to θ_{QCD}, which are related
(b) to the renormalization of the coupling of the m Higgs.

Figure 2 Illustration [15] how the baryon-antibaryon asymmetry can be built up in the decays of superheavy particles $X(Y_+ \equiv (n_X+n_{\bar{X}})/n_\gamma)$, $Y_B = (n_q-n_{\bar{q}})/n_\gamma$, $Y_- = (n_X-n_{\bar{X}})/n_\gamma$.

Figure 3 (a) Typical fourth order contributions to ϵ_H, and
(b) the generic vertices a_d and $b_d^\dagger$.

Figure 4 (a) The C- and CP-violating imaginary part of the trace of I_1 which gives ϵ_H, and
(b) an analogous contribution to the m Higgs coupling matrix.

Figure 5 (a) Fourth order contribution to Im $Tr(I_1)$ in an SU(5) model with two $\underline{5}$'s of Higgs (the solid lines are $\underline{10}$'s of fermions, the zigzags are $\underline{5}$'s of fermions, the dashed lines are d Higgs and the dot-dashed lines are other Higgses), with
(b) an analogous contribution to the m Higgs coupling matrix.

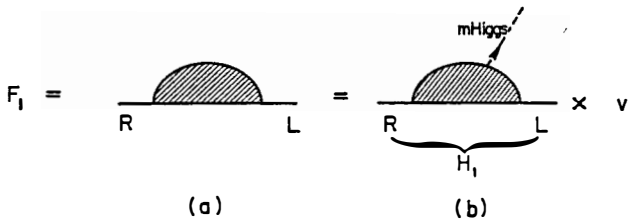

mHiggs
F1 =
R
L
=
R
L
H1
× v
(a)
(b)

Fig. 1

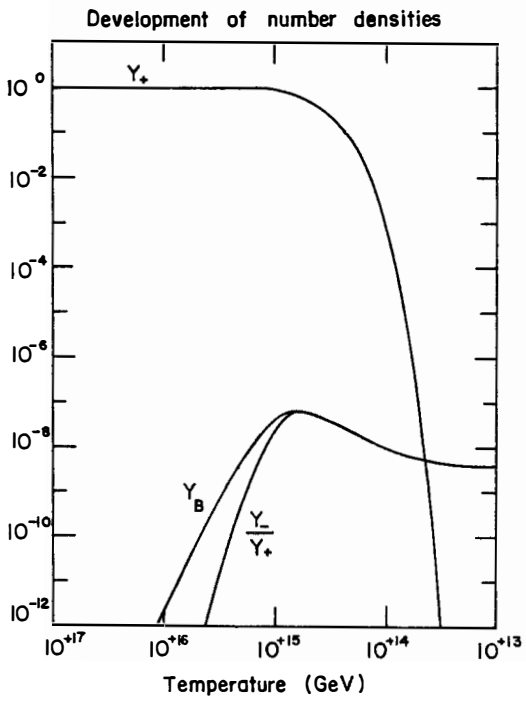

Development of number densities
Y+
10 0
10-2
10-4
10-6
10-8
10-10
10-12
YB
Y-
Y+
10+17
10+16
10+15
10+14
10+13
Temperature (GeV)

Fig. 2

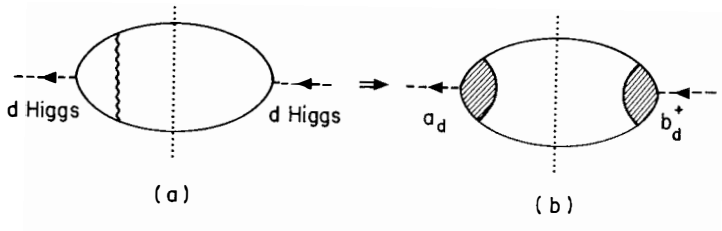


Fig. 3

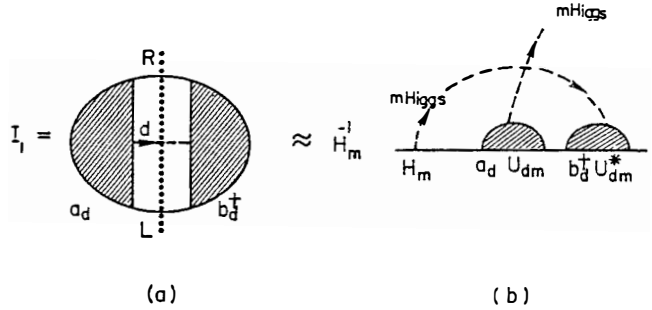


Fig. 4

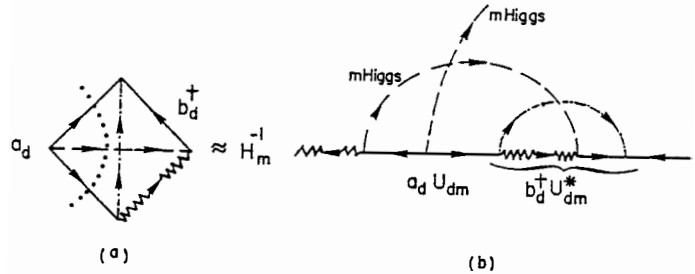


Fig. 5

ELEMENTARY PARTICLE PHASE TRANSITIONS IN THE VERY EARLY UNIVERSE

S. A. Bludman†
Department of Physics, University of Pennsylvania
Philadelphia, Pennsylvania 19104, U.S.A.

ABSTRACT: The restoration of elementary particle symmetry at high temperatures induces a huge cosmological constant (vacuum energy density) which exceeds the thermal energy density if the Higgs meson mass is small enough. If the Universe began at low entropy (Tepid Universe), this prevents any initial cosmological singularity and prevents massive monopole production in the initial GUTS phase transition.

> "The tentative conclusion is reached that if general relativity is to be treated as a self-contained theory, then the 'cosmical terms' that contain the cosmical constant should be omitted. But if general relativity is only part of what is needed to construct a theoretical model of physical reality, then the cosmical terms ought to be retained as affording additional freedom in linking up with other parts of physical theory."
>
> W. H. McCrea, "The Cosmical Constant", Q. Jour. Royal Astron. Soc. 12 (1971) 140

I. INTRODUCTION

The spontaneous broken gauge theory (SBGT) of elementary particles has already unified weak and electromagnetic interactions and promises to lead to a grand unification of strong and electroweak interactions (GUTS). The Higgs mechanism, elementary or dynamical, for each symmetry breaking of the elementary particle vacuum predicts that above some temperature T_c the symmetry will be unbroken, i.e. that, if the temperature was high enough, the present cold universe of elementary particles must have been preceded by an earlier phase in which elementary particle symmetry was restored. The gap in energy density between the hot (disordered) phase and the present cold (ordered) phase, goes to materialize meson and fermion masses, and is calculable, positive and huge.[1]

Gravitational observations show that the vacuum energy density (cosmological constant) of the present universe is negligible. Why this is so remains a deep mystery we cannot answer. (Perhaps $\varepsilon_{vac} = 0$ is necessary as a condition of asymptotic flatness in an expanding universe in order that the total energy be definable.) Precisely because the vacuum energy density vanishes in the present universe, SBGT requires that it was huge and positive in the very early hot universe. In fact, if the Higgs meson mass M_H associated with the earliest phase transition is close to its minimum allowable value M_{CW}, this vacuum energy density dominates the radiation-matter density over a temperature range in the early phase universe, so that the positivity condition $\varepsilon + 3p > 0$ was not then satisfied. In this paper, we show that if the initial entropy is chosen small enough, this fixes a maximum temperature and a minimum radius for the early universe. In this way, what is now standard particle physics can avoid any initial singularity which, if it occurred, would have signalled the breakdown of classical gravitational theory. After symmetry breaking, the universe is in the $\varepsilon_{vac} = 0$ phase with which we are familiar, so that the standard scenario for baryosynthesis, nucleosynthesis, plasma recombination and galaxy formation holds.

Especially in cosmology, where so few observations are available, it is desirable to assume a theory with the minimum of free parameters. For this reason, we choose for cosmology the Robertson-Walker theory (whose one adjustable parameter is the initial adiabat or RT value) and, for GUTS, the minimal SU(5) theory (whose one adjustable parameter is, as in the Weinberg-Salam theory, the Higgs

mass M_H).

II. EFFECTIVE POTENTIAL FOR MINIMAL SU(5) GUTS THEORY

A. Zero Temperature

The effective potential $V(\phi)$ of the Higgs field ϕ for the minimal SU(5) GUTS theory in one-loop approximation is, at zero temperature,[2,3]

$$V_o(\phi) = \tfrac{1}{2}\mu^2\phi^2 - (\frac{\mu^2}{4\sigma^2} + \frac{1}{2}B)\,\phi^4 + B\phi^4 \ln(\phi^2/\sigma^2) + K \qquad (2.1)$$

where $B = (75/8)^2(g^2/4\pi)^2$ and K is a constant depending on the absolute zero of energy. We consider the weak Higgs sector in which the vector mesons are much heavier than the scalar, so that the $B \ln(\phi^2/\sigma^2) \sim B \ln (T^2/M_V^2)$ radiative correction receives contributions from vector meson loops but scalar or fermion loops are neglected. The gauge vector meson coupling is g and $V(\phi)$ has been written in a form making explicit the minimum at $\phi=\sigma$ and the additional extremum at $\phi = 0$. The values determined for these minimal GUTS parameters[4] are given in Table 1, together with those for the familiar $SU(3)_c \times SU(2)_w \times U(1)_{em}$ theory.

Since

$$\frac{d^2V_o}{d\phi^2} = \begin{cases} \mu^2 , & \phi = 0 \\ M_H^2 , & \phi = \sigma \end{cases} , \qquad (2.2)$$

where $M_H^2 \equiv M_{CW}^2 - 2\mu^2 \equiv 2(M_{LW}^2 - \mu^2)$, the one still undetermined parameter μ^2 fixes the Higgs mass and the nature of the V_o extrema. Here $M_{CW}^2 \equiv 8B\sigma^2 \equiv 2M_{LW}^2$ is a natural mass unit. The $\phi = 0$ extremum is a minimum for $M_H < M_{CW}$ ($\mu^2 > 0$) and a maximum for $M_{CW} < M_H$ ($\mu^2 < 0$).

Because the gap in energy density

$$\Delta\varepsilon \equiv V_o(0) - V_o(\sigma) = \tfrac{1}{8}(M_H^2 - M_{LW}^2)\sigma^2, \qquad (2.3)$$

we consider only $M_{LW} < M_H$, so that the broken symmetry phase $\phi = \sigma$ is absolutely stable with respect to any local minimum at $\phi = 0$. We will thus have two cases: (i) $M_{LW} < M_H < M_{CW}$, a false vacuum at $\phi = 0$ which is (at all temperatures) metastable with respect to transitions to the true vacuum at $\phi = 0$; (ii) $M_{CW} < M_H$, instability at $\phi = 0$ (at T = 0 and low enough temperatures).

B. Finite Temperatures

At finite temperature[2,3], $V_o(\phi)$ must be replaced by $V(\phi(T),T) = V_o(\phi) + V_T(\phi,T)$ where

$$V_T(\phi(T),T = \frac{3NT^4}{2\pi^2}\int_0^\infty \ln\{1-\exp[-(x^2 + \frac{m_V^2}{T^2})]\}\; x^2\,dx \qquad (2.4)$$

is the free energy density of an ideal Bose gas of N = 12 vector mesons of temperature-dependent mass,

$$M_v(\phi(T)) = (25/8)^{\frac{1}{2}}\, g\phi(T). \tag{2.5}$$

Since for $M_v/T < 1$,

$$V_T(\phi,T) = 3NT^4 \left\{-\frac{\pi^2}{90} + \frac{(M_v/T)^2}{24} + (M_v^3/T^3)\right\}, \tag{2.6}$$

the coefficient of ϕ^2 in V_o is replaced at finite temperature by $\frac{1}{2}(\mu^2 + g''^2T^2)$, where $g''^2 \equiv 75\, g^2/8$. Finite temperature therefore tends to stabilize the symmetric state $\phi = 0$ and to destabilize the broken-symmetry state $\phi = \sigma$. Indeed, above a temperature T_2, defined by

$$(d^2V/d\phi^2)_{\sigma(T_2)} = 0 \quad , \tag{2.7}$$

there is no broken symmetry. As shown in Fig. 1a), for $M_{LW} < M_H < M_{CW}$ there was already at $T = 0$ a false vacuum ($\phi=0$) as well as the true vacuum ($\phi=\sigma$); for $0<T<T_2$, the barrier between these two vacua is reduced. As shown in Fig. 1b), for $M_W < M_H$, $\phi = 0$ was an unstable equilibrium at $T = 0$, this remains so for $0<T<T_1$ where T_1 is defined by $(d^2V/d\phi^2)_{\phi=0} = 0$ so that

$$T_1^2 = (-\mu^2)g''^2 = 4(M_H^2 - M_{CW}^2)/75g^2.$$

For $T_1<T$, however, $\phi = 0$ becomes a minimum.

At $T = 0$, the broken symmetry phase lay lower than the symmetric phase by (2.3). We define the symmetry temperature T_c at which the two phases have equal minima, $V(0,T_c) = V(\sigma(T_c),T_c)$. Thus

$$\Delta N\, \frac{\pi^2}{10}\, T_c^4 \approx \varepsilon_{vac}\,, \tag{2.8}$$

where $\Delta N=12$ is the number of originally massless vector mesons which become massive on symmetry breaking.

C. First Order Phase Transitions

In Fig. 2) the heavy lines plot the location of the minima of $V(\phi(T),T)$. Above T_2 only the symmetric phase $\phi = 0$ is stable. Below T_1, only the broken symmetry phase $\phi = \sigma(T)$ is stable. But, at intermediate temperatures, there are two locally stable minima separated by a potential barrier. At $T = T_c$, these two vacua are degenerate in energy: this would be the transition temperature in a quasi-static expansion or contraction. For $T < T_c$, the broken-symmetry vacuum is the true vacuum and the symmetric vacuum is a metastable "false vacuum". For $T > T_c$ the reverse is the case.

In fact, because tunneling through a barrier is exponentially

slow, a contracting universe superheats up to $T_{sh} \approx T_2 > T_c$, and a expanding universe supercools down to $T_{sc} < T_c$. In case $M_{CW} < M_H$, $T_{sc} \approx T_1$, the temperature at which the barrier disappears. In case $M_{LW} < M_H < M_{CW}$, $T_{sc} \approx 0$ unless another phase transition intervenes.

During this drawn-out first-order transition, the universe is far from thermal equilibrium. When the transition to the broken-symmetry phase finally takes place, the temperature rebounds almost up to T_c, with entropy multiplication $E \approx (T_c/T_{sc})^3$. As shown in Fig. 3) for $M_{CW} < M_H$, the supercooling temperature $T_1 \sim (M_H^2 - M_{CW}^2)^{\frac{1}{2}}$ and the rebound temperature $\approx T_c \sim (M_H^2 - M_{LW}^2)^{\frac{1}{2}}$. While E, shown by the dashed curve in Fig. 3, becomes very large for $M_{LW} < M_H \lesssim M_{CW+}$, it approaches unity quite rapidly for $M_H \gtrsim M_{CW}$. Because $M_{CW} = \sqrt{8B}\,\sigma$ is the measure of radiative corrections that make the transition first-order, the strength of the first-order transition (measured say by the latent heat released) decreases continuously as M_H increases from its minimum value M_{LW}. Defining $\beta^2 \equiv (M_H/M_{LW})^2 - 1$, the transition becomes practically second order for $M_H/M_{CW} = [\frac{1}{2}(1+\beta^2)]^{\frac{1}{2}} > 1$. While the radiative corrections always make the phase transition technically first-order, we will see in Sec. IIIB that empirical limits on entropy multiplication show that the transition can only be moderately first-order, i.e., we must consider only $M_H \gtrsim M_{CW}$ or $\beta \gtrsim 1$.

D. Definition of Vacuum Energy

Evaluated at its minima $\phi = 0$, $\sigma(T)$, the Higgs potential defines the free energy density of the vacuum sector

$$f \equiv \frac{E - TS}{\text{Volume}} = \varepsilon - T s = -p. \tag{2.9}$$

The last equality follows from the fundamental thermodynamic relation $E + pV = TS + \Sigma N_i \mu_i$, since in the charge-symmetric vacuum all net conserved charges are zero. f, ε and $s = -df/dT$ are functions of T only. Thus the energy density divides naturally into a constant zero-temperature part $\varepsilon_o = f_o = V_o(\phi(0),0)$ and a thermal part $\varepsilon_T(T) \equiv \varepsilon(T) - \varepsilon_o$. The zero temperature part obeys the equation of state $p_o = -\varepsilon_o$, while in the radiation-dominated universe we consider, $f_T \sim T^4$ and $p_T = -\varepsilon_T + T s = \frac{1}{3}\varepsilon_T$.

The stress-energy tensor for an ideal fluid $T^{\mu\nu} = (\varepsilon + p)u^\mu u^\nu + p g^{\mu\nu}$ is independent of four-velocity u^μ, if and only if, it obeys the equation of state $\varepsilon + p = 0$. Such a Lorentz-invariant vacuum, endowed with dynamical degrees of freedom demanded by quantum field theory, deserves to be called the Lorentz-

invariant ether. While the vacuum $\varepsilon_{vac} = \varepsilon_o$ is Lorentz-invariant, like the temperature itself the thermal energy is not. In a non-empty universe, the matter-radiation flow defines a preferred co-moving frame.

The gravitational properties of the vacuum are just those of a dynamically induced cosmological constant, since Einstein's equation $G^{\mu\nu} + \Lambda g^{\mu\nu} = -\kappa\, T^{\mu\nu}$, where $\kappa \equiv 8\pi G = 8\pi(\hbar c/M_P^2)$ and $T^{\mu\nu}$ is the radiation-material stress energy tensor, can be rewritten

$$G^{\mu\nu} = -\kappa\ (T^{\mu\nu} + T^{\mu\nu}_{vac}),\ \kappa\varepsilon_{vac} \equiv \Lambda \tag{2.10}$$

Empirically, the vacuum energy vanishes in the present cold universe. Therefore choosing in Eq. (2.1)

$$\kappa = (\tfrac{1}{2}B\sigma^2 - \tfrac{1}{4}\mu^2)\sigma^2 = \frac{1}{8}(M_H^2 - M_{LW}^2)\sigma^2, \tag{2.11}$$

we finally have

$$\varepsilon = \frac{1}{2}\mu^2\phi^2 - (\frac{\mu^2}{4\sigma^2} + \frac{1}{2}B)\phi^2 + B\phi^4 \ln(\phi^2/\sigma^2) + \varepsilon_{vac}$$
$$+ 3N\ \frac{\pi^2}{30}\,T^4 - \frac{1}{2}g''^2\,T^2\,\phi^2 + \ldots \tag{2.12}$$

which is sketched in Fig. 4a). Defining T_x so that $\varepsilon_T(T_x) = \varepsilon_{vac}$, we see that there is a temperature interval $T_x > T > T_1$, in the supercooled symmetric phase in which the vacuum energy density dominates the radiation-material density. Fig. 4b) shows $\varepsilon + 3p = 2(\varepsilon_T - \varepsilon_{vac})$, which is negative in the vacuum-dominated plane.

III. THE EXPANDING UNIVERSE

The success of the standard cosmological model for nucleosynthesis shows that when the universe had cooled down to ~1 Mev, it was already quite isotropic and homogeneous. We assume that this was already the case at the much earlier GUTS epoch (although allowing appreciable anisotropy at baryosynthesis would permit initial isothermal density perturbations, from which it is easier to later evolve galaxies than would be the case with the initial adiabatic baryon density perturbations permitted in an isotropic universe). The expansion is then governed by

$$H^2(t) \equiv (\frac{\dot{R}}{R})^2 = \frac{8\pi G\varepsilon}{3} - \frac{k}{R^2}\ ,\ k = 0, \pm 1\ , \tag{3.1}$$

which together with the First Law of Thermodynamics $d(\varepsilon V) + pdV = 0$

or

$$d\varepsilon = -(\varepsilon + p)\frac{dV}{V} \tag{3.2}$$

implies

$$\ddot{R} = -\frac{4\pi G}{3}(\varepsilon + 3p)R \quad .$$

Where the positivity condition $\varepsilon + 3p > 0$ is obeyed, $\ddot{R} < 0$, so that a plot of R(t) vs. t is concave downwards; if this were the case, there would be a physical singularity T, $\varepsilon \to \infty$ as $R \to 0$. For the vacuum-dominated radiation universe, however,

$$\varepsilon + 3p = 2\,\varepsilon_T - 2\,\varepsilon_{vac} < 0 \tag{3.3}$$

so that R(t) is convex upwards. Particularly as the universe supercools towards T_1 so that $\varepsilon_T \ll \varepsilon_{vac}$, the scale grows exponentially $R(t) \sim \exp(ct/\sqrt{2}\,R_E)$, where

$$R_E \equiv M_p\,(3/16\pi\varepsilon_{vac})^{\frac{1}{2}} \quad . \tag{3.4}$$

A. Advantages of Exponential Growth

Guth[5] has pointed out three advantages to exponential growth and supercooling in the very early universe:

1. The particle horizon

$$d_H(t) = R(t)\int_0^t \frac{cdt'}{R(t')} \tag{3.5}$$

grows rapidly, as $R_E \exp(ct/\sqrt{2}\,R_E)$ in the vacuum-dominated phase, instead of as 2ct in a radiation-dominated universe. This allows early chaos to homogenize through causal processes.

2. Flatness problem: The spatial curvature term in Eq.(3.1), $|k/R^2| \ll H^2 \sim 8\pi G(\varepsilon/3)$ in the present universe. Why is our present universe as large ($R > 10^{28}$ cm) as it is? Since, at present, T = 2.7 K, why is $RT > 10^{28} \sim e^{64}$ or the dimensionless entropy presently in photons and neutrinos

$$S/k_B = \frac{4}{3}\,\frac{\pi^2}{15}\left[1 + 3\left(\frac{7}{22}\right)\right] T^3 R^3 > 10^{86} \; ? \tag{3.6}$$

3. Exponential growth and extreme supercooling would also suppress[6] the cosmological production of magnetic monopoles whose mass $M_m \simeq 10^{16}$ Gev would otherwise dominate the present expansion. Since $\rho_m \lesssim \rho_{CR}$ and $\rho_B \sim .01\,\rho_{CR}$, the present numbers of monopoles

must be in the ratio

$$n_m/n_B \leq 100\ (M_B/M_m) = 10^{-14} \quad . \tag{3.7}$$

At present, the observed baryon-photon ratio and specific entropy are

$$\eta \equiv n_B/n_\gamma = (2\text{-}8) \times 10^{-10}$$
$$s/k_B n_B = 10^{11.4 \pm 0.3} \quad . \tag{3.8}$$

The limits for monopoles are therefore[7]

$$n_m/n_\gamma < 10^{-23}$$
$$s/k_B n_m > 10^{25} \quad . \tag{3.9}$$

GUTS predicts $n_m \gtrsim n_B$ or

$$s/k_B n_m \sim 10^{11},\ n_m/n_\gamma \sim 10^{-9} \quad , \tag{3.10}$$

in strong disagreement with (3.9).

B. Inescapable Problems with Extended Exponential Growth

Guth preferred to attribute the hugh size or entropy needed to solve the horizon and flatness problems to at least 64 e-folds of exponential growth in a drawn-out first-order phase transition. He himself realized[5], however, that such extensive exponential growth would have fatal consequences.

1. Limits on Entropy Multiplication after Baryosynthesis.

GUTS can in the first place generate only a baryon/photon ratio $\eta_o \leq 10^{-5}$ or specific entropy $(s/k_B n_B)_o > 10^6$. The presently observed value (3.8) therefore permits only entropy multiplication $E < 10^6$ through dissipative processes of all kinds since baryosynthesis. In particular, the electroweak phase transition at $T_c \sim$ 23 Gev can allow moderate supercooling down to only $T_{sc} > 0.01\ T_c =$ 230 Mev. (If the recent theoretical bound[8] on the neutron's electric dipole moment, $d_N > 2.5 \times 10^{-8}\ \eta_o$ e-cm is combined with the present experimental upper bound 6×10^{-25} e-cm $> d_N$, then stronger limits $\eta_o < 2 \times 10^{-7}$, $E < 6000$, $T_{sc} > .06\ T_c = 1$ Gev are obtained. Witten's idea[9] of arresting electroweak supercooling by a chiral-symmetry breaking transition at 100-300 Mev is thus incompatible with the theoretical bound on d_N and just compatible with the original baryosynthesis bound.)

Two conclusions emerge concerning the entropy multiplication or baryon dilution admissable after baryosynthesis: (1) Supercooling from the electroweak phase transition must stop at $T_{sc} > 230$ Mev, because of the intervening chiral-symmetry phase transition[9] or because the Weinberg-Salam Higgs meson has mass $M_H > 1.002\ M_{CW} = 9.2$ Gev.[10,11] (2) Any entropy multiplication $E > 10^{14}$ which would reconcile the monopole ratios (3.9) and (3.10) would be incompatible with the baryon dilution $E < 10^6$ permitted.

2. Inescapable Exponential Growth?

There are no such limits on entropy generation before baryosynthesis fixes the baryon-photon ratio η_o and magnetic monopoles are synthesized. To explain the present uniformity, flatness, and absence of monopoles by an extremely drawn-out first-order phase transition, Guth had wanted exponential growth and supercooling by a factor $10^{28} = e^{64}$. Nevertheless, Guth himself realized[5] that out of such exponentially rapid expansion, the universe as a whole would never undergo the transition into the broken-symmetry phase we now inhabit.

Extreme supercooling takes place when $M_H \approx M_{CW}$. In this Coleman-Weinberg approximation, $\mu^2 = 0$ and scale invariance is broken only by the logarithmic radiative corrections $B \ln(\phi^2/\sigma^2) \sim B \ln(T^2/M_V^2)$: the growth rate of bubbles of the broken-symmetry phase proceeds at a rate per unit volume per unit time $\sim T^4 \exp[-A(T)]$, where $A(T)$ is the minimum 0(3)-symmetric Euclidean action.[9] Meanwhile, the universe expands at a rate per unit volume per unit time $\chi^4 = (8\pi G\, \varepsilon_{vac}/3)^2$. The nucleation rate never catches up with the expansion rate because $A(T) \sim (g^3 \ln \frac{M_V}{T})^{-1}$ decreases so slowly with T that $T^4 \exp[-A(T)] < \chi^4$ for all $T > 0$. Consequently, although individual bubbles expand almost at the speed of light, they never coalesce (percolate) to include an appreciable fraction of the universe. This catastrophic conclusion is not quite right, even for the Coleman-Weinberg theory[11], but does show that for $M_H \approx M_{CW}$, T_{sc} is extremely small. Indeed, in the electroweak phase transition, in order to have $T_{sc} > 230$ Mev so that entropy multiplication $E < 10^6$, the Weinberg-Salam Higgs mass $M_H > 1.002\ M_{CW} = 9.2$ Gev.[10,11] When $M_H > M_{CW}$ we saw that the metastable false vacuum becomes unstable for $T < T_1$.

C. Extended Supercooling Is Not Needed

Excessive entropy multiplication after baryosynthesis is ruled out by the specific entropy presently observed. Before baryosynthesis, we cannot endure supercooling extensive enough to make the present RT > 10^{28}. These conclusions rule out any drawn-out first-order phase transition as the source of the present entropy $S/k_B > 10^{86}$ and of the reduced monopole production.

To solve the horizon and flatness problems, however, requires only some huge entropy before baryosynthesis. This huge entropy could be an initial condition, a consequence of an anthoropic principle, or a consequence of <u>any</u> very strong dissipation. What strongly dissipative processes are there before GUTS? The finger points at gravitational processes which are known to be strongly dissipative near the Planck temperature and density. Processes coming immediately to mind are the isotropization of initial chaos, repeated bounces of a closed universe, particle production and Hawking radiation from black holes (event horizons) or horizon radiation from deSitter space (particle horizons).[12]

The magnetic monopole problem will be solved if the universe was never hotter than 10^{16} Gev or if the breaking of the originally compact symmetry down to a U(1) subgroup is delayed to a low temperature.

IV. A TEPID UNIVERSE

After seeing why first-order transitions cannot be excessively drawn-out, we now turn our attention to gravidynamic processes in the very early symmetric universe. We will see how the cosmological constant induced at symmetry restoration avoids the initial cosmological singularity and leads to an initial temperature $T_x \sim 9 \times 10^{13}$ Gev, low enough that magnetic monopoles will not be initially produced.

A. Unstable Static Einstein Universe: Non-Singular Initial Conditions

We assume the universe began at $t = -\infty$ in a minimal SU(5) symmetric false vacuum defined by some Higgs mass $M_H > M_{CW}$. (Since we will allow only a mildly first-order phase transition $\beta^2 \equiv 2(M_H/M_{CW})^2 - 1$ cannot be too close to unity.) We assume an Einstein universe, closed (k = +1) and static with scale factor

$$R_E = M_P(3/16\pi\varepsilon_{vac})^{\frac{1}{2}} = 8.9 \times 10^{-13}\ \beta^{-1}\ \text{Gev}^{-1}$$

$$= 1.8 \times 10^{-26}\ \beta^{-1}\ \text{cm} \qquad (4.1)$$

and temperature

$$T_X \approx (30\ \varepsilon_{vac}/N\pi^2)^{\frac{1}{4}} = 0.92 \times 10^{14}\ \beta^{\frac{1}{2}}\ \text{Gev}, \qquad (4.2)$$

chosen so that

$$\frac{3}{16\pi G R_E^2} = \varepsilon_T(T_X) = \varepsilon_{vac} = (1.7 \times 10^{14}\ \beta^{\frac{1}{2}}\ \text{Gev})^4 \qquad (4.3)$$

will make $\dot{R} = 0 = \ddot{R}$ initially. Since $R_E T_X = 82\ \beta^{-\frac{1}{2}}$, the initial entropy

$$S/k_B = (2N\pi^2/15)\ (R_E T_X)^3 = 8 \times 10^6\ \beta^{-3/2}$$

is relatively small. We call this non-singular alternative to the Hot Big Bang, the Tepid Universe. It should not be confused with Linde's Cold Universe[13] in which $S = 0 = \varepsilon_T$ initially, and almost all of the radiation-matter is supposed to be created out of the initial vacuum instability.

As is well known, Einstein's static universe is very unstable gravitationally. If allowed to expand slowly from rest at $t = -\infty$, it expands and supercools until at

$$T_1 = M_{CW}(\beta^2-1)^{\frac{1}{2}}/2g'' = 0.83 \times 10^{14}\ (\beta^2-1)^{\frac{1}{2}}\ \text{Gev}, \qquad (4.4)$$

the false vacuum becomes an unstable phase, permitting the delayed transition into the broken-symmetry phase. During the supercooling, initially mild exponentially accelerated expansion takes place from rest as $\varepsilon + 3p = 2(\varepsilon_T - \varepsilon_{vac})$ decreases from zero. Guth' exponential growth, on the contrary, begins from a universe that is already exploding out of the Big Bang.

Because in the initial symmetric phase the temperature never exceeded $T_X = 9 \times 10^{13}\ \beta 1/2$ Gev, magnetic monopoles (whose mass $\gtrsim 10^{16}$ GeV) were never made in the original tepid beginning. If we are in the first expansion, this explains the absence of magnetic monopoles in the present universe.

B. The Initial Expansion:
Suppression of Monopole Production

During the elementary particle phase transition, the temperature rebounds to almost

$$T_c \approx (90\ \varepsilon_{vac}/\Delta N\pi^2)^{\frac{1}{4}} = 1.2 \times 10^{14}\ \beta^{\frac{1}{2}}\ \text{Gev}$$

with entropy multiplication

$$E \approx (T_c/T_1)^3 = 3[\beta/(\beta^2-1)]^{3/2} \qquad (4.5)$$

from $S/k_B = 8 \times 10^6\ \beta^{-3/2}$ Gev to $2 \times 10^7\ (\beta^2-1)^{-3/2}$ Gev. This entropy could be made much larger, by finely tuning β to be very, very close to one. Besides being unnatural, however, this would lead to extreme supercooling to very low T_1.

If our model is correct, we therefore expect the SU(5) Higgs mass $M_H > M_{CW} = 2.7 \times 10^{14}$ Gev, and the universe to be closed. Since the hugh present entropy cannot be generated in a long, drawn-out first-order transition, we require other strongly dissipative processes to explain the present entropy or size of the universe. These other strongly dissipative processes in the very early universe must almost certainly be gravitational, such as dissipation of initial anisotropy, repeated bounces of the closed universe, or radiation from event or particle horizons.

C. Conventional Evolution of a Closed Universe

Once in the broken-symmetry phase at about 10^{14} Gev, the universe continues with the conventional evolution through baryosynthesis, further symmetry-breaking transitions, nucleosynthesis, recombination and galaxy formation. Being closed, the universe expands to a maximum size and minimum temperature determined by

$$c^2 = \frac{8\pi G}{3}\,\varepsilon(T_{min}) \cdot R_{max}^2 \qquad . \qquad (4.6)$$

We, with $H^2 = (\dot{R}/R)^2 \sim \frac{8\pi G}{3}\,\varepsilon$ are nowhere near maximum size now. In the subsequent recontraction and reheating, the universe will become radiation-dominated, but because $\ddot{R}$ remains negative, it will never again become vacuum-dominated.[14]

Although its origin was non-singular in the tepid symmetric phase, the universe must subsequently go through repeated non-exponential crunches. In those subsequent infinite contractions and reheatings, entropy and magnetic monopoles will be created.

V. CONCLUSIONS

Without appealing to quantum gravity or supergravity, by using only classical general relativity and by-now conventional particle physics, we have avoided an initial singularity with a choice of original scale size at $t = -\infty$, $\hbar/M_P c \ll R_E \ll H_o^{-1}$. Because before symmetry-breaking the temperature never exceeded $T_X \sim 10^{14}$ Gev, magnetic monopoles are not made in the first symmetry-breaking, but will be made in later crunches.

This scenario requires that the universe be closed and that the Higgs mass exceed $M_{CW} = 2.7 \times 10^{14}$ Gev. Because, just like Guth, we cannot derive the present entropy from extended exponential supercooling, we require other, probably gravitational, processes to explain the present entropy or size of the universe.

"I would as soon think of reverting to Newtonian theory as of dropping the cosmical constant."
A. S. Eddington, The Expanding Universe
(Cambridge University Press, Cambridge, 1933)

REFERENCES

1. S. A. Bludman and M. A. Ruderman, Phys. Rev. Lett. 38 (1977) 255; S. A. Bludman, UPR-0143T (1979), unpublished.
2. D. A. Kirzhnits and A. D. Linde, Phys. Lett. 42B (1972) 471; D. A. Kirzhnits, Zh. Eksp. Teor. Fiz. Piz. Red. 15 (1972) 529 [JETP Lett. 15 (1972) 529].
3. S. Weinberg, Phys. Rev. D9 (1974) 3357; L. Dolan and R. Jackiw, Phys. Rev. D9 (1974) 3320.
4. T. J. Goldman and D. A. Ross, Phys. Lett. 84B (1979) 208; Cal-Tech preprint 68-759 (1980); W. J. Marciano, Phys. Rev. D20, (1979) 274; J. Ellis, M. K. Gaillard, D. V. Nanopoulos and S. Rudaz, Annecy preprint LAPP-TH-14 (1980).
5. A. H. Guth, Phys. Rev. D23 (1981) 347; A. H. Guth, SLAC-PUB 2576 (1980).
6. A. H. Guth and S.-H. H. Tye, Phys. Rev. Lett. 44 (1980) 631.
7. Y. B. Zel'dovich and M. Y. Khlopov, Phys. Lett. 79B (1978) 239; J. Preskill, Phys. Rev. Lett. 43 (1979) 1365.
8. J. Ellis, M. K. Gaillard, D. V. Nanopoulos and S. Rudasz, CERN TH. 2967 (1980).
9. E. Witten, Nucl. Phys. B177 (1981) 477.
10. A. H. Guth and E. J. Weinberg, Phys. Rev. Lett. 45 (1980) 1131.

11. P. J. Steinhardt, Nucl. Phys. B179 (1981) 492.

12. D. Eardley and W. H. Press, private communication.

13. A. D. Linde, Rep. Prog. Phys. 42 (1979) 390.

14. M. A. Sher, Phys. Rev. D22 (1980) 2989.

	$SU(3)_c \times SU(2)_w \times U(1)_{em}$	Minimal SU(5)
Statistical weight	106.75	160.75
σ(Gev)	246	4.5×10^{14}
	$g^2/4\pi = 1/31.5$, $\frac{1}{4\pi}(g^2+g'^2)=1/24.3$	1/45
g"2	$\frac{3g^2+g'^2}{16} = (0.287)^2$	$\frac{75}{8} g^2 = (1.6)^2$
Massive Vector Mesons	$W^{\pm}$ (2) , Z^o(1)	X (12)
M_V (Gev)	$\frac{g\sigma}{2} = 77.7 \sqrt{\frac{g^2+g'^2}{2}}$ σ=88.5	$\frac{25}{8}^{\frac{1}{2}} g\sigma = 4.2 \times 10^{14}$
One-loop Gauge Meson Radiative Correction		
B	$\frac{3}{8^2}\left[2\left(\frac{g^2}{4\pi}\right)^2 + \left(\frac{g'^2}{4\pi}\right)\right]^2 = 1.74 \times 10^{-5}$	$\frac{75}{8}^2 \left(\frac{g^2}{4\pi}\right)^2 = 4.3 \times 10^{-2}$
Higgs Meson Mass Scale		
M_{CW} (Gev)	9.2	2.7×10^{14}
M_{LW} (Gev)	6.5	1.9×10^{14}
Phase Transition		
ϵ_{vac} $(Gev)^4$	$(24\ \beta^{\frac{1}{2}})^4$	$(1.7 \times 10^{14} \beta^{\frac{1}{2}})^4$
T_c (Gev)	$24\ \beta^{\frac{1}{2}}$	$1.2 \times 10^{14} \beta^{\frac{1}{2}}$
T_X (Gev)	$9.7\ \beta^{\frac{1}{2}}$	$0.92 \times 10^{14} \beta^{\frac{1}{2}}$
T_1 (Gev)	$16\ (\beta^2-1)^{\frac{1}{2}}$	$0.83 \times 10^{14} (\beta^2-1)^{\frac{1}{2}}$

Table 1. Parameters of Minimal SU(5) and of SU(3)xSU(2)xU(1) Spontaneously Broken Gauge Theories. Aside from the order parameter and coupling constants, the effective Higgs potential in both cases depends only on the Higgs mass $M_H^2 = M_{LW}^2(1+\beta^2) = M_{CW}^2\ (1 + \beta^2)/2$.

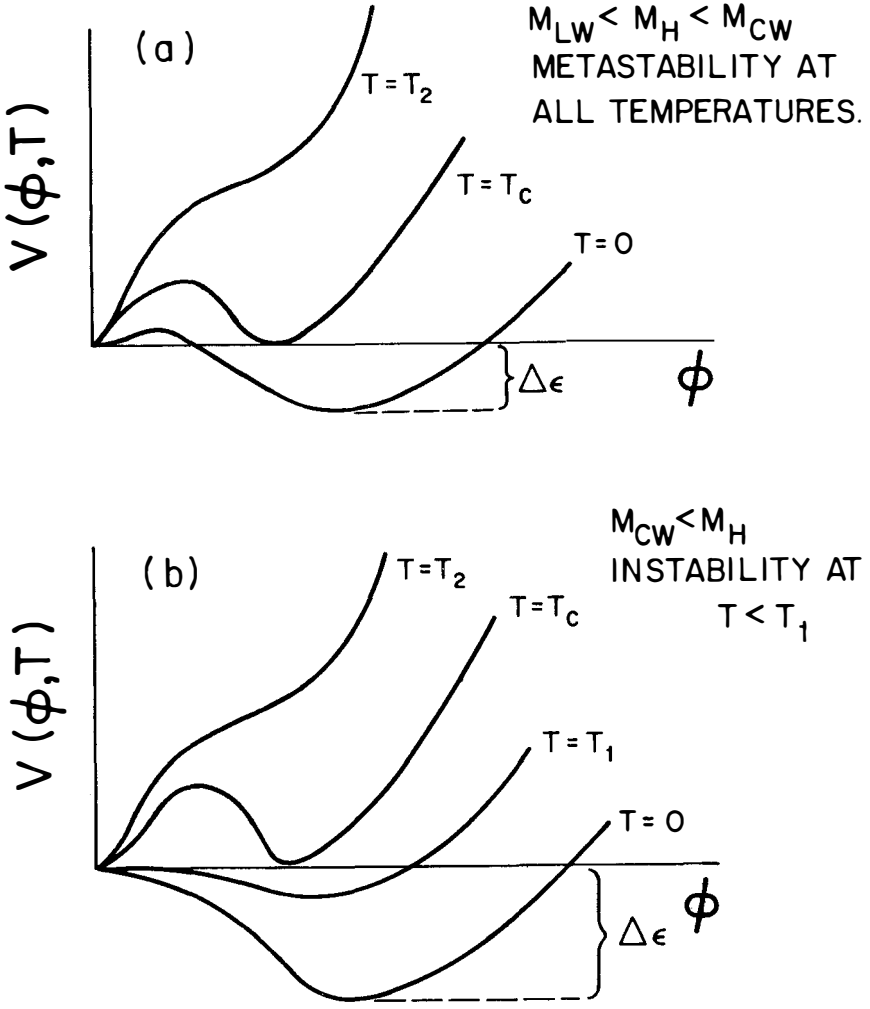


Figure 1. Temperature dependent effective Higgs potential for a) $M_{LW} < M_H < M_{CW}$, in which case the symmetric state $=0$ is metastable at all temperatures $< T_c$ and b) $M_{CW} < M_H$, in which case the symmetric state is unstable at $T < T_1$. The symmetric and asymmetric states are degenerate at T_c. The broken symmetry state is metastable at $T > T_c$ and unstable at $T > T_2$.

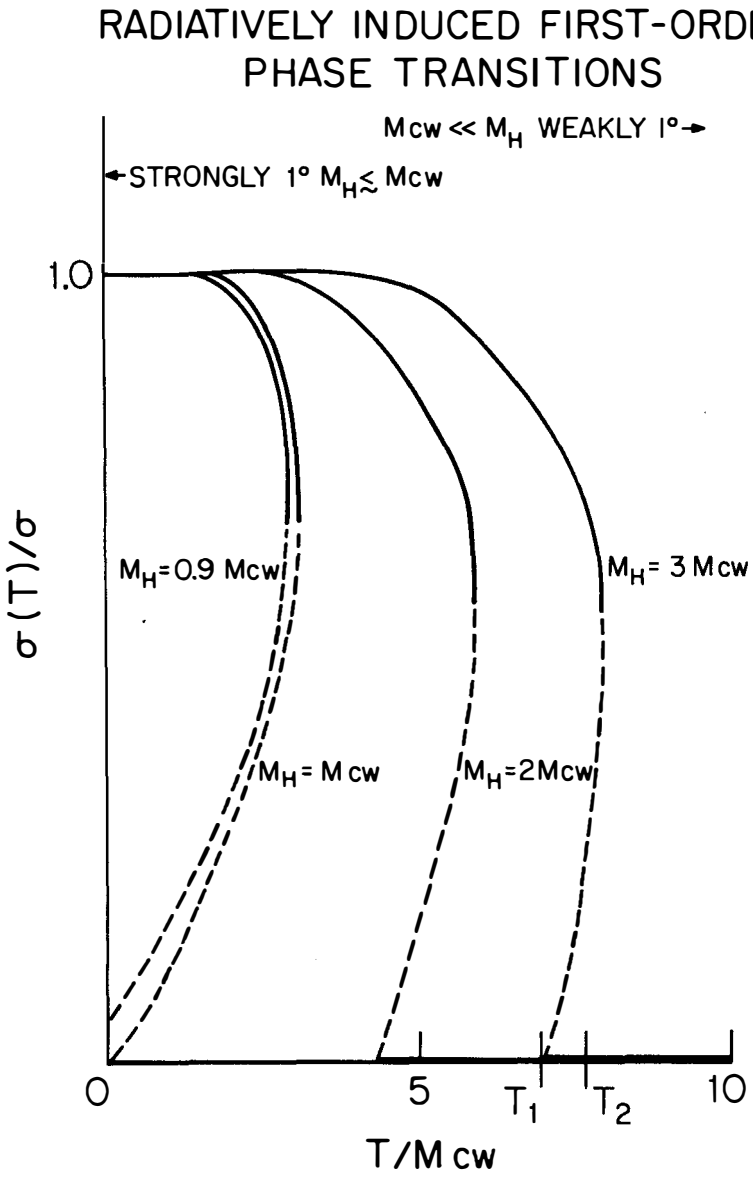


Figure 2. The temperature-dependent order parameter $\sigma(T)$ in units of σ as function of T/M_{CW} for various values of M_H/M_{CW}. For $T_1<T<T_2$ the effective potential has two minima (heavy line) and an intervening maximum (dashed line). For $M_H<M_{CW}$ the symmetric state is stable down to T_2 and metastable down to T=0. On expansion, the symmetric state will supercool practically down to T_1 or 0. On contraction, the broken-symmetry state will superheat up to T_2. (Adapted from ref. 14.)

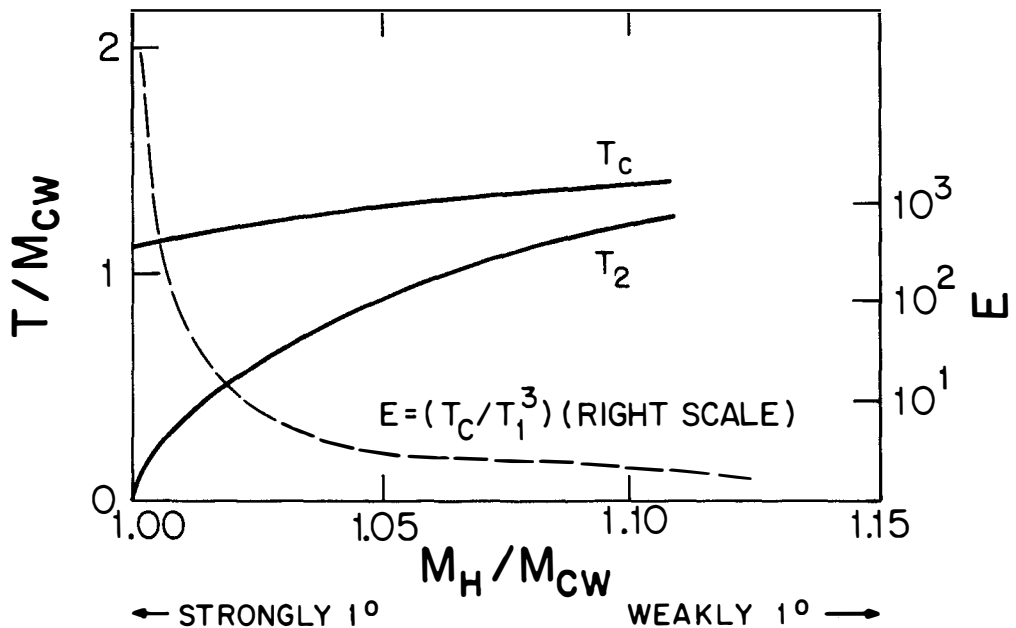


Figure 3. The supercooling temperature T_1 and the rebound temperature (heavy curves, left scale) and the entropy multiplication (dashed curve, right scale) as function of Higgs mass. (Modified from ref. 14).

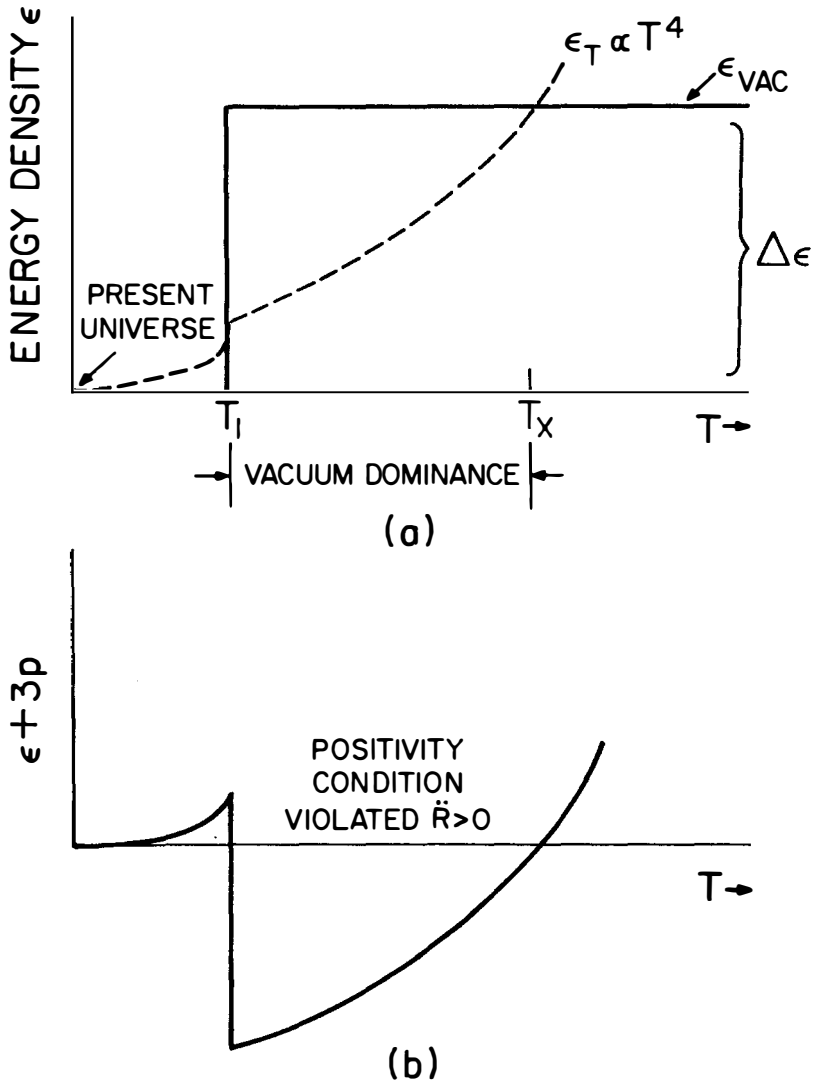


Figure 4. (a) The vacuum energy ε_o and the radiation energy ε_T as function of temperature during cooling. The vacuum energy dominates for $T_x > T > T_1$.

(b) $\varepsilon = 3p$ as function of T. During the vacuum-dominated interval $R > 0$ and R(t) will be convex upwards from the t-axis.

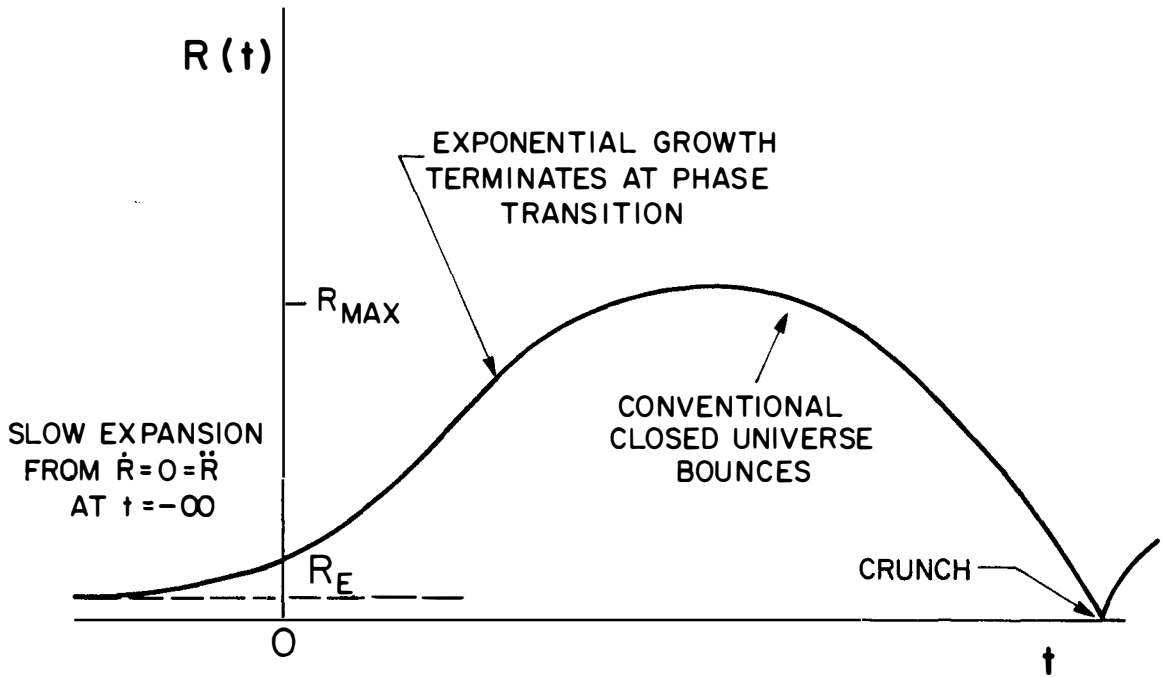


Figure 5. The Tepid Universe, which is gravitationally and thermally unstable expands exponentially from R_E with R = 0 = R at t = -∞, until it supercools down to T_1. Thereafter it oscillates between R_{max} and an infinite crunch at R = 0.

GALAXIES MAY BE SINGLE PARTICLE FLUCTUATIONS FROM AN EARLY, FALSE-VACUUM ERA

William H. Press
Department of Physics and Center for Astrophysics
Harvard University, Cambridge, MA 02138, U.S.A.

ABSTRACT: Evolving from an early, hot Friedman phase, the universe may go over to a "Guth era" of exponential expansion. The horizon size at the end of the Guth era (which can be as large as today's comoving scale of galaxies or clusters) is shown to correspond to a scale not much larger than the Compton wavelength of the matter fields at the beginning of the Guth era. Means of "freezing in" these initial fluctuations are suggested. Exit from the Guth era to a reheated Friedman universe might be mediated by the finite temperature of the cosmological event horizon: effects of field theory in curved spacetime may make the cosmological temperature decrease only as a power-law, even as the expansion proceeds exponentially.

I. INTRODUCTION

I will assume that the initial state of the universe, at least by the Planck time $t_p = (\hbar G/c^5)^{1/2} \sim 10^{-43}$ s when it first makes sense to talk about a classical spacetime metric, is smooth, isotropic and homogeneous, and described by a Robertson-Walker metric. It is a matter of taste whether to view this assumption as something which must ultimately be explained, or whether it is just a statement about what happened to be. Particle theorists tend toward the former viewpoint, while most cosmologists are not too uncomfortable with the latter view.

Many cosmologists <u>are</u> uncomfortable, however, with the idea that the perturbations from smoothness which are necessary to make galaxies might need to be postulated as initial conditions. To say that the universe is smooth has a small information content; it takes only a single sentence to describe the state completely. To say that the universe has galaxy-forming (and cluster-forming) perturbations, however, implies a very large initial information content at $t = t_p$, whole figurative reams of tabulated material describing the positions of all the galaxies that are fated to form and, indeed, all details of the initial hydrodynamic state that will form them. One should not dismiss this as just "some realization of a random function." Random functions have a high information content; when they occur in physics it is most often by the realization of some underlying complicated physical process (such as the quantum nature of matter, or stochastic mixing of deterministic equations in an appropriately large phase space). The central cosmological issue, it seems to me, is not <u>where did the smoothness come from</u>, but rather <u>where did the perturbations come from</u>.

From the Robertson-Walker metric and the Einstein equations for gravitation, one gets the well-known first integral of the expansion equation,

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G\rho}{3} - \frac{k}{R^2} \qquad (1)$$

Here R(t) is the expansion factor in the metric, scaled to make k take on the value +1, 0 or -1, corresponding to a spatially closed, marginally open, or open universe. The density ρ comprises all the

gravitating matter content of the universe, including any Lorentz-invariant vacuum energy density ("cosmological constant" term). The evolution of ρ satisfies

$$\frac{d\rho}{dR} = - \frac{3(P+\rho)}{R} \tag{2}$$

where P is the total pressure.

It is worth emphasizing that ρ and P are derived quantities from some Lagrangian that describes the underlying quantum theory of matter. From a Lagrangian L, a functional of the matter fields, one obtains the stress-energy tensor by varying with respect to the metric $g^{\mu\nu}$,

$$T_{\mu\nu} = L\, g_{\mu\nu} - 2 \frac{\delta L}{\delta g^{\mu\nu}} \tag{3}$$

(see, e.g., Lightman et al. 1975, §§21.5-21.6). The homogeneity of the metric implies a preferred Killing vector time direction ξ^{α}, in terms of which

$$\rho = \langle T_{\mu\nu}\xi^{\mu}\xi^{\nu} \rangle \tag{4a}$$

$$P = \langle \frac{1}{3}(T_{\alpha}{}^{\alpha} - T_{\mu\nu}\xi^{\mu}\xi^{\nu}) \rangle \tag{4b}$$

The angle brackets denote (or, for our purposes, at least connote) the quantum expectation values of the underlying fields in a curved spacetime metric.

In grand unified gauge theories (GUTS) the universal matter Lagrangian often contains a polynomial potential term in some scalar combination of fields ϕ,

$$L = \ldots + \ldots - V(\phi,t) \tag{5}$$

which implies a contribution to $T_{\mu\nu}$ of the form

$$T_{\mu\nu} = \ldots + \ldots - V(\phi,t) g_{\mu\nu} \ , \tag{6}$$

in other words a cosmological constant term. The potential is a function of temperature, so it depends on time, not explicitly, but through the temperature's time variation.

Guth (1981) brought together a number of threads from GUTs and cosmology to show that the nature of matter at $>10^{15}$ GeV (where the GUTs live) could have major cosmological implications. Let me here describe not exactly Guth's picture, but a close variant:

Schematically, but without reference to any particular GUT model, let us suppose that $V(\phi,t)$ has the qualitative behavior shown in Figure 1. In rough terms, the matter content of the universe at every local point has two kinds of contributions: first, thermal excitations above the minimum of the potential well in V, which represent particles of finite mass and positive pressure and density contribution; second, the height V_1 of the minimum of V above zero, which contributes an (as drawn) positive density but equally negative pressure (from eq.6).

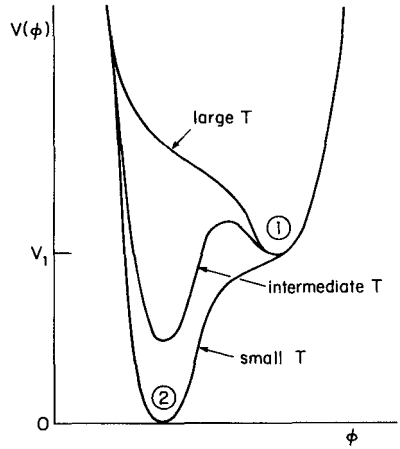


Figure 1. Assumed vacuum potential.

As the universe cools from some initially hot state, particle excitations redshift away, and the vacuum at every point in space tends to the potential minimum labeled "1" in the figure. Assume for the moment that there is an era during which the height of this minimum is relatively unchanging. Then the solution of (1) is asymptotically an exponential

$$R \sim \exp(Ht), \quad H = \left(\frac{8\pi G V_1}{3}\right)^{1/2} \tag{7}$$

We suppose that this exponential (de-Sitter) expansion proceeds through some number of e-folds before the potential passes through the intermediate temperature form in figure 1. (How many e-folds this might be is the important question that we must presently consider.) Eventually, if the potential evolves as shown, there is a loss of metastability of minimum "1". At every point in space, the vacuum comes crashing down, via a first order phase transition, into highly excited state of minimum "2", representing a now-reheated Friedman epoch which is then presumed to evolve to our present universe.

If the stage of exponential expansion (we will call it the "Guth era" of the evolution) lasts more than a few e-folds, so that the total expansion is exponentially large, then two interesting consequences follow: (1) The cosmological horizon, or domain of communication of some cosmological observer, becomes exponentially large measured in terms of today's scales. We will discuss this effect in the next section. (2) The universe becomes effectively k=0, whatever its initial value of k. To see how this happens, we need only look at how the two terms on the right hand side of eq. (1) scale with time during the Guth era. The first, varying as ρ, is constant, since ρ is dominated by the constant vacuum density V_1. The second term, varying as R^{-2}, decreases at twice the exponential rate of the expansion. It does not take many e-folds before the term is so small in magnitude that no amount of power-law recovery during the subsequent reheated phase (when $\rho \propto T^4 \propto R^{-4}$) can resurrect it in importance.

Observationally, we know that the term in k is at least not strongly dominant today (since $\rho \sim \rho_{crit}$, at least in order of magnitude, cf. Gott et al. 1974). Without a Guth era, this fact can be viewed as something of a puzzle, since the first (ρ) term in eq. (1) should have always varied as R^{-4} (radiation dominated) or R^{-2} (matter dominated) hence the ratio of the terms would have to have been immensely large at the Planck time. Particle theorists, who do not like very large unexplained numbers in their theories, thus find Guth's suggestion comforting. To my own taste, the possibility of using the Guth era's large horizon size to obtain results relating to the perturbation problem is a vastly more interesting aspect of the exponential expansion picture.

II. CAUSAL STRUCTURE OF GUTH'S COSMOLOGY

Suppose we label the time coordinate so as to make the singularity occur at some small negative value of t, the transition into the Guth era at t=0, the transition out of the Guth era at $t=t_*$, and the present epoch at $t=t_0$. Then the metric is obtained by requiring that R(t) and $\dot{R}/R$ be both continuous:

$$ds^2 = - dt^2 + R^2(t)[dr^2 + r^2(\sin^2\theta d\phi^2 + d\theta^2)] \qquad (8)$$

with

$$R(t) = \begin{cases} \exp[H(t-t_*)](1/2Ht_0)^{1/2}, & 0 \leq t \leq t_* \\ \left(\dfrac{t-t_*+1/2H}{t_0}\right)^{1/2}, & t_* \leq t \leq t_0 \end{cases} \qquad (9)$$

The value of H, the expansion rate in the deSitter phase, is given by the GUT theory through eq. (7).

One easily integrates $dr = dt/R(t)$ to find Δr, the comoving coordinate distance traversed by a signal propagating at speed c(=1). During the Guth era $0<t<t_*$ one gets

$$\Delta r_* = \frac{1}{H}(2Ht_0)^{1/2}\, e^{Ht_*}\, (1 - e^{-Ht_*}), \qquad (10)$$

while subsequently, for $t_*<t<t_0$, one has (using $t_0 >> t_*$)

$$\Delta r_0 \overset{\sim}{>} 2t_0 \qquad (11)$$

The ratio of these two distances, when $Ht_* >> 1$, is

$$\frac{\Delta r_*}{\Delta r_0} = \frac{e^{Ht_*}}{(2Ht_0)^{1/2}} \qquad (12)$$

Since H is the expansion rate of the Guth era ($>10^{24}$ s^{-1}, say) Ht_* parameterizes the dimensionless duration of that era, while Ht_0 is the present age of the universe in the same dimensionless units. The denominator in (12) is therefore a very large number; however, the numerator, being exponential, can be even larger for Ht_* only moderately big. Therefore, the domain of communication during the

Guth epoch can be very much larger than that inferred from the subsequent Friedman phase.

The nature of this causal structure is clarified by examining a conformal space-time diagram (Figure 2). Here a conformal coordinate transformation preserves light cones as 45^{o} lines, but compactifies the future infinities of the spacetime. The top diagram shows the full evolution of a radiation- or matter-dominated Friedman cosmology, with t varying from zero (singularity) at the base of the triangle to infinity at the apex (cf. Hawking and Ellis 1973, § 5.3). All matter world lines end up at this apex. The sides of the triangle are made up of points at null infinity, where all light rays go. One sees that the cosmology has no event horizons: the light cone of any worldline comes asymptotically to include the entire spacetime. On the other hand, the cosmology has particle horizons: at t=0, the light cone of one worldline includes no other worldline; only as t increases to some finite value does the past light cone of one observer first come to include any of a second observer's worldline (one "comes into the particle horizon" of the other). A few past light cones are drawn to illustrate the point. In particular, the two worldlines shown as dotted curves are supposed to be separated by the diameter of a galaxy today; they first come into each other's horizon at t ~ 1 year.

The middle diagram in Figure 2 shows a full deSitter space represented conformally. (The representation chosen is actually a semi-infinite half-plane, of which only a portion near the boundary is shown.) The key point is that timelike infinity is not a point, but rather a whole hyperplane. DeSitter space has no particle horizons, since past light cones can be extended "down" arbitrarily far to include a part of any other worldline. But deSitter space does have event horizons: asymptotically at future infinity, an observer's past light cone never comes to include another observer's worldline beyond a certain point on it. That other observer has "crossed an event horizon" as surely as if by falling into a black hole.

The lower diagram in Figure 2 is a copy of the beginning stages of a radiation dominated Friedman universe.

We are now ready to "stitch together" a Guth universe: It begins at the bottom of Figure 2c and evolves upward to the surface marked $t=t_G$. At this point it is matched smoothly to the surface labeled $t=t_G$ in Figure 2b and undergoes exponential expansion until time $t=t_*$, shown in the figure. That surface matches to the surface labelled

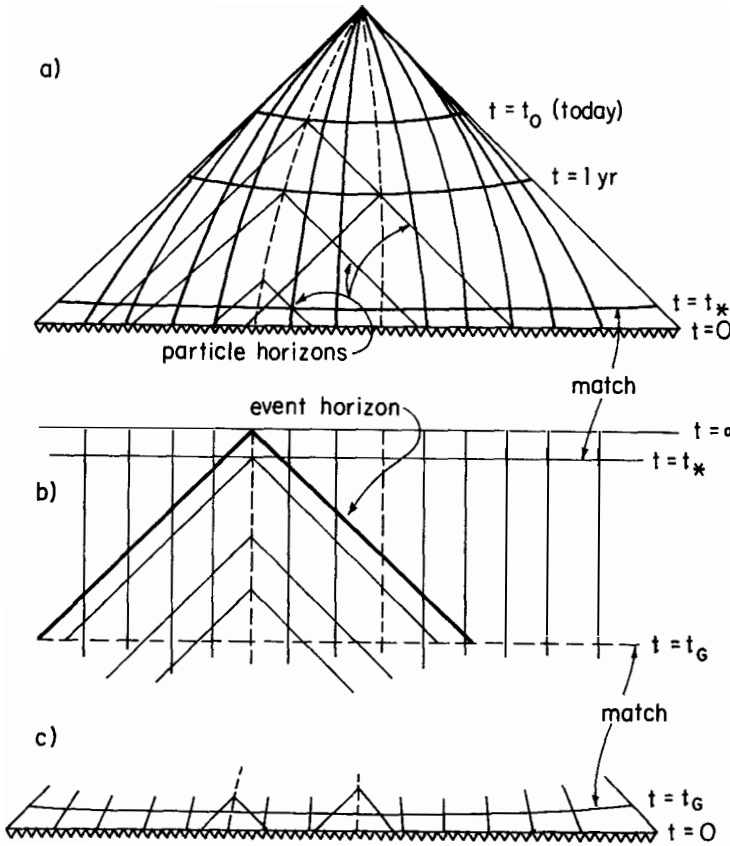


Figure 2. Conformal diagrams showing causal structure (a) of a Friedman cosmology, (b) of exponentially-expanding deSitter spacetime, (c) of an early Friedman epoch. Guth's cosmology is obtained by "sewing together" parts of the three spacetimes, as indicated. See text for details.

$t=t_*$ in Figure 2a (where we exit from the Guth era), and the evolution continues to the present.

The dotted "galaxy" worldlines are mutually incommunicado in the pre-Guth era. They come across each other's particle horizons in about the middle of Figure 2b. They exchange information (or particle interactions) at a relative redshift that first decreases from infinity to some finite value, then increases again almost to infinity (as they go out of each other's deSitter event horizon), then, in Figure 1a, finally decreases towards zero after about t = 1 yr. The crux of the matter is that when the galaxy comes within its Friedman horizon at t = 1 yr, its proper self-greeting is not "hello," but "hello again!"

III. ARE GALAXIES THE FOSSILS OF SINGLE PARTICLES?

It is interesting to convert the dimensionless number Ht_*, which parameterizes the duration of the Guth era, to a number whose interpretation is more direct today: the number of baryons (or their mass in solar masses) in a region that was able to communicate across itself during the Guth era. To use the formulae already given most conveniently, let us redefine t_0 to be that time at which "hello again" occurred for the Guth horizon mass (cf. remark at end of § II above). This is the time, therefore, which satisfies

$$\Delta r_* = \Delta r_0 \tag{13}$$

for eqs. (10) and (11), which yields (for $Ht_* >> 1$)

$$e^{Ht_*} = (2Ht_0)^{1/2} = \frac{R(t_0)}{R(t_*)} = \frac{T(t_*)}{T(t_0)} \tag{14}$$

(where eq. 9 and the inverse relation between R and T have been used). The temperature $T(t_*) \equiv T_*$ is the reheating temperature immediately after the transition back to the true vacuum. The temperature $T(t_0)$ can be related to the baryon mass within the horizon, M_H, by

$$M_H = \left(\frac{5}{2\pi^3}\right)^{1/2} \frac{m_H}{SN^{1/2}} \left[\frac{T_{planck}}{T(t_0)}\right]^3 \tag{15}$$

where S is the entropy per baryon ($\sim 10^9$), m_H is the proton mass, N is the effective number of species contributing to the thermal density,

and T_{planck} is the Planck temperature (=1.2x10^{19} GeV). Equations (14) and (15) then give

$$Ht_* = \ell n \left[\left(\frac{3.52\ M_H SN^{1/2}}{m_H} \right)^{1/3} \frac{T_*}{T_{planck}} \right] \qquad (16)$$

For $M_H = 10^{15} M_\odot$, $T^* = 10^{15}$GeV, one has $Ht_* \sim 53$, while for $M_H = 10^{21} M_\odot$ (the present horizon), $T_* = 10^{17}$GeV, one has $Ht_* = 66$. So, Guth eras which last this long (say, 60±6) have horizon sizes that are of interesting cosmological sizes today.

Let us now ask, what is the physical proper size of the comoving length Δr_* (or Δr_0) at the start of the Guth era. Evidently the answer is, from eqs. (9) and (10)

$$\Delta r_* R(0) = \frac{1}{H} (1 - e^{-Ht_*}) \approx \frac{1}{H} \qquad (17)$$

It may seem surprising that this answer is largely independent of $Ht_*(>>1)$, but Figure 2b shows why: even if $Ht_* \to \infty$, the event horizon comes to include only a finite part of the surface $t=t_G$.

Another length to estimate is the Compton wavelength λ_c associated with the temperature at the start of the Guth era. This temperature is approximately the same as T_*, the reheating temperature after the Guth era, since the energy density is constant during the exponential expansion. Therefore,

$$\lambda_c = \hbar / T_* \qquad (18)$$

and using

$$H \sim \left(\frac{8\pi G\, a\, T_*^4}{3} \right)^{1/2} \qquad (19)$$

a short calculation gives the result

$$\lambda_c \sim 15 \left(\frac{T_*}{T_{planck}} \right) \Delta r_* R(0) \qquad (20)$$

Equation (20) says that, for T_* not too much less than T_{planck}, there are only a few Compton wavelengths within the region that will subsequently inflate exponentially to become the Guth horizon mass, corresponding to a very large mass today. This is true independent of the value of that mass (which is determined by Ht_* through eq. 16).

We are led to speculate that single particle fluctuations on scale λ_c might be the primeval fluctuations that make galaxies or clusters of galaxies. For this picture, one imagines that quantum state of the matter fields at $t<t_p$ is in some sense chaotic, with all the internal variables of the matter Lagrangian uncorrelated from point to point. The first thing that happens, at $t\sim t_p$, is that the horizon size expands to exceed the Compton wavelength associated with the (then) temperature, so that it makes sense to talk about particle state and dynamical interactions between states. These dynamical interactions can be imagined to smooth out the chaotic matter field variables on scales smaller than the current horizon. Different, non-communicating, horizon-sized regions ought to be brought to statistically the same matter state; however there will be fluctuations due to the fact that, as we have seen, the Guth era horizon comes at most to include a region whose initial region contained only a few Compton wavelengths.

One has, then, a natural way of producing a preferred scale of interesting size today, with some natural amplitude of perturbation, less than unity but not much less than unity. Larger scales than the Guth horizon size have increasingly smaller perturbations, because they contain more Compton wavelengths, hence are statistically more homogeneous. Smaller scales have smaller perturbations, because--being smaller than the horizon size--they have been homegenized by physical interactions.

I am being intentionally vague about the precise internal nature of these quantum perturbations. They cannot be just adiabatic perturbations in the total matter density. Such perturbations can be shown to have only decaying modes during the exponential expansion of the Guth era (Kahn 1981). The perturbations that we need must be essentially "kinematic" in character, painting, as it were, the universe with random regions of red and green paint which (after they go out of their deSitter event horizon) expand passively through the many required e-folds. During this expansion, the regions should maintain their "redness" or "greenness." Then, just before or anytime after the reheating transition back to the true vacuum, the difference between red and green must have some way of taking on dynamical significance, e.g. by contributing differently to the equation of state. The important point is that this dynamical significance need not involve communication across the size of the regions (which are

now far outside their apparent Friedman horizons). It need only be a local property of redness or greenness.

Let me give an example, only incompletely developed from suggestions of H. Georgi and S. Coleman, of how the "red and green paint" might be included into a GUT theory. To anyone's favorite GUT theory, add two new gauge fields Φ_1^α and Φ_2^α. The α index is that of the internal group, e.g. Φ_1^α and Φ_2^α might be 5's of SU(5); the extra fields need not have any interaction at all with anything else in the theory, so no violence is done to the physics at low energies. Suppose that the Φ's contribute a term to the Lagrangian of the form

$$L(\Phi_1,\Phi_2) = (|\Phi_1|^2 + |\Phi_2|^2)^2 + m_1^2|\Phi_1|^2 + m_2^2|\Phi_2|^2 + \ldots \quad (21)$$

where m_1 and m_2 are small compared with the unification energy, and $m_1 > m_2$.

At high temperatures the mass terms (in m_1 and m_2) are invisible, since they scale only quadratically in the Φ's. There is therefore an apparent larger symmetry

$$\Phi_1 \to \Phi_1 \cos\theta + \Phi_2 \sin\theta \quad (22)$$

and every point is characterized by a direction θ in Φ_1, Φ_2 space. Interaction terms (not written down above) are supposed now to make coherent horizon-sized regions of approximately constant θ. Those with $\theta \sim 0$ are "red" (Φ_1 dominated), while those with $\theta \sim \pi/2$ are "green" (Φ_2 dominated). Once the regions go out of their deSitter horizons, their character is frozen in.

Much later, at lower temperatures, the mass terms become important. Since m_1 is larger, the Φ_1 field becomes nonrelativistic first, so its contribution to the equation of state is different: there is a pressure perturbation that is spatially congruent with the (now large) red and green regions. This pressure perturbation lasts until Φ_2 also becomes nonrelativistic. At this point we no longer care about the subsequent fate of the Φ_1 and Φ_2 fields; they presumably find decay channels into lighter particles. Their erstwhile pressure difference has now left an indelible footprint in the total stress-energy tensor.

The point is that once a pressure perturbation of fractional size δ has been maintained for about one expansion timescale, it will give

rise to a growing-mode adiabatic perturbation whose density perturbation will reach a magnitude about δ when it (much later) comes within its Friedman horizon (Press and Vishniac 1980). That this is very long after GUT physics has given way to low energy chemistry does not matter. To summarize, we obtain galaxy- or cluster-forming perturbations which were passed through six stages: (1) coming within their Guth particle horizon to get the required spatial coherency; (2) going out of their apparent deSitter horizon, to freeze in that coherency; (3) while still outside their horizon, converting internal field variables into pressure perturbations lasting at least one expansion scale; (4) undergoing transition to the true vacuum either after (3) or just before; (5) losing all traces of their internal field variables as the GUT era cools away; (6) finally coming within their Friedman horizons in the form of adiabatic density perturbations of a preferred scale.

The subsequent picture of galaxy formation might well be along the lines of Zel'dovich and collaborators (see, e.g. Doroshkevich et al. 1974, 1978, 1980).

IV. THE "GRACEFUL EXIT" PROBLEM

Up to now, we have been assuming that the transition from the false vacuum of the Guth era (with its large vacuum energy density) to the true vacuum (with its large entropy content) occurs simultaneously at all points of the spacelike hypersurface $t=t_*$ (Figure 2b). This assumption is implicit in Figure 1 also, where the transition is drawn to occur via loss of metastability of the upper state as a function of temperature T (which is constant on t = constant hypersurfaces). In fact, however, our assumption is contrary to Guth's original formulation of the "inflationary universe" and to most work that has been done on the false vacuum problem (e.g. Coleman 1977, Callan and Coleman 1977, Coleman and DeLuccia 1980).

The difficulty is that, during the exponential Guth era, the temperature decreases exponentially. After a few e-folds its value is effectively zero, so there should be no further evolution in the shape of the effective potential of Figure 1. In particular, it is not at all credible that the loss of metastability should occur after the 60 ± 6 e-folds necessary to make the transitions cosmologically interesting.

Guth's original formulation assumes a potential which maintains a metastable upper state even at zero temperature. The transition to the true vacuum then occurs not by loss of metastability, but by tunneling. The trouble with this is that the tunneling events occur at different random times at each point in space. Each produces an expanding "bubble" of true vacuum whose boundary is very nearly the future light cone of the tunneling event. Moreover, the energy density of reheated material is not deposited uniformly within the interior of the bubble, but is concentrated as an expanding null shell. (This is inevitable if the tunneling is a random event in 4-space: it must produce a Lorentz-invarient matter distribution; for exegesis on these points, see Coleman and DeLuccia 1980 and references therein.) The bubbles, then, are highly inhomogeneous. Guth has shown that there is no hope of having many overlapping bubbles restore homogeneity statistically: The statistics of bubble formation are such that the few earliest tunneling events form very massive bubble walls which are resistant to subsequent homogenization.

It seems likely to me that simultaneous loss of metastability, not barrier penetration, is the only way by which the Guth cosmology can be made to resemble our own, nearly homogeneous, one today. There must be some "clock" connected to the evolution of the potential evolution in Figure 1 that does not run down its mainspring until the requisite 60±6 e-folds of expansion have taken place. In present GUT theories, the only such clock (on which the potential depends) is the temperature. One possibility is that particle theorists will have to learn to build some additional structure into the theory, something which affects the form of the effective potential, but which changes less rapidly than an exponential as the universe expands.

Doug Eardley and I have been exploring another possibility which, though speculative at this stage, might prove able to resolve the graceful exit problem. The key idea is to note that the event horizon of deSitter space (Figure 2b) has as associated finite temperature that does not change with the expansion of the metric. The existence of this finite temperature, which is a consequence of gravitational production of quanta in a curved spacetime, was first pointed out by Gibbons and Hawking (1977) and subsequently investigated also by Lapedes (1978). The effect is closely analogous to the finite temperature of the event horizon of a black hole, which produces the celebrated "Hawking evaporation".

The apparent finite temperature of the deSitter event horizon arises from the fact that the O(4) invariant vacuum state (which can be shown, following Lapedes 1978, to be the vacuum state which is quickly approached after the transition from the early Friedman to the Guth epoch) is not the same state as that defined by local cosmological observers when they require particle (antiparticle) excitations to have positive (negative) frequencies with respect to their proper time coordinate. Actually, this latter definition of vacuum can be shown to correspond not to the O(4) global vacuum, but to a global vacuum defined by minimizing the zero point energy within the interior of the observer's event horizon only, and with zero field conditions imposed on the horizon surface. Since this is a more restrictive boundary condition than O(4) invariance, the zero point energy is lower (fewer allowed modes). Therefore, the cosmological observers locally interpret the O(4) vacuum as having a positive energy; in fact, it can be shown that they see an apparent thermal spectrum of finite temperature.

Although not so familiar in the deSitter case, the effects of discrepant local and global vacuum definitions have been studied in some detail by, e.g., DeWitt (1975), Davies and Fulling (1977), and Candelas (1980). DeWitt, in particular, notes the close analogies between this effect and the Casimir effect in quantum electrodynamics. An example even more closely related to the deSitter application is furnished by "Rindler space", which is Minkowski space populated by a self-similar congruence of uniformly accelerated observers. These observers measure the global Minkowski vacuum as having a finite temperature.

What is the stress tensor of the O(4) vacuum? Gibbons and Hawking show that the cosmological observers measure a finite temperature, so one expects that $T_{\mu\nu} \neq 0$. However, there also exist boosted congruences of observers who, by these arguments, would also measure the same finite temperature. In fact, the relation between $T_{\mu\nu}$ and local measurements of the state is indirect and depends on precisely what is meant by "measure": If the observers are only a kinematic congruence of "test" particles, so that their measurements have no dynamical significance and cause no back-reaction on the vacuum state, then their existence does not affect the underlying O(4) invariance of the problem. In this case, $T_{\mu\nu}$ must maintain its O(4) invariance, and the only possible correction due to the finite

temperature horizon must be of the form $\Lambda g_{\mu\nu}$; e.g., for a conformally invariant field it will be a pure trace anomaly or, in other words, a quantum-gravitational renormalization of the vacuum energy density. (For related references, see Brown and Cassidy 1977, Dowker and Critchley 1977, Wald 1978, Davies et al. 1977). From a macroscopic viewpoint, the finite horizon temperature contributes a pressure exactly equal to the negative of its density. It can do this because the wavelength corresponding to the peak of the blackbody spectrum is on the order of the radius of curvature of the geometry; the flat-space equation of state $P = 1/3\,\rho$ can thus be strongly affected, going over to $P = -\rho$ as the O(4) symmetry in fact demands.

On the other hand, what if the observers are not test particles, but actual dynamical systems capable of storing (in internal degrees of freedom such as particle rest mass) an energy density comparable to that of the thermal horizon flux? In this case, the observers break the O(4) invariance of the problem more than just kinematically. To the extent that they absorb thermal flux from the horizon, they can convert that flux from a Lorentz-invariant form into a form with a preferred rest-frame, their own. In other words, the macroscopic equation of state will not be $P=-\rho$, but rather

$$P = (\Gamma-1)\rho\ , \quad \Gamma\neq 0 \tag{23}$$

where the precise value of Γ will depend on details of the interactions.

There seem to be no realistic calculations of Γ in the literature, except that of Candelas and Dowker (1979) for a conformally invariant interaction, which gives $\Gamma=0$ (pure trace anomaly). One might have expected this result on the grounds that a conformally invariant interaction has no internal length scales which can form nondimensional ratios to the wavelength of the blackbody peak, therefore there is no free parameter to scale the equation of state away from the O(4) invariant form. In what follows, we will <u>assume</u> that a sufficiently complicated set of particle interactions will ultimately be shown to yield $\Gamma\neq 0$ (cf. Birrel, Davies, and Ford 1980). From this assumption one gets an intriguing scenario:

It seems possible that there can be a kind of "bootstrapping" relationship between the thermal horizon flux (produced in a state of O(4) invariance) and the complicated particle states that it interacts with: The thermal flux maintains a finite temperature which does not

redshift away exponentially. This, in turn, maintains a finite density of thermally-produced finite-mass interacting particles, created at any given time by interactions among existing particles that define a mean cosmological rest frame (break the O(4) invariance). Now these new interacting particles can continue to provide a mean rest frame, and a value of Γ different from zero, for the next stage of interaction with the thermal horizon flux. In other words, there is a kind of spontaneous breakdown of O(4) symmetry, triggered by the initial existence of finite mass particles which define a mean rest frame.

We can calculate quantitatively the (semiclassical) evolution of the cosmology during the Guth era in this scenario. If the expansion rate H is slowly varying (this can be justified self-consistently), then the instantaneous horizon temperature should just track the expansion rate,

$$T = \kappa \frac{h}{c} H = \kappa \frac{h}{c} \frac{\dot{R}}{R} \tag{24}$$

where κ is a constant of order unity. The total density is

$$\rho = \rho_{vac} + NaT^4 \tag{25}$$

where N is the number of species, a ($=\pi^2/15$ in Planck units) is the radiation constant. The total pressure is

$$P = -\rho_{vac} + (\Gamma - 1)NaT^4 \tag{26}$$

Equations (24), (25), (26), (1), and (2) are five equations for the five unknown functions of time ρ, P, T, ρ_{vac}, and R. The solution (for k=0) is easily effected by quadratures, and is

$$R(t) = \exp\left\{\frac{1}{CH_0^2}\left[\left(1 + \frac{3}{2} CH_0^2 t\right)^{2/3} - 1\right]\right\} \tag{27a}$$

$$T = T_0\left(1 + \frac{3}{2} CH_0^3 t\right)^{-1/3} \tag{27b}$$

$$\rho = \frac{3}{8\pi G}\left(H_0^{-3} + \frac{3}{2} Ct\right)^{-2/3} \tag{27c}$$

$$\frac{NaT^4}{\rho} = \frac{CH_0^2}{8}\left(1 + \frac{3}{2} CH_0^3 t\right)^{-2/3} \tag{27d}$$

Here C is a constant of order the Planck time squared,

$$C = \frac{64\pi^3}{45} \kappa^4 N(\Gamma-1) t_p^2 \tag{28}$$

and H_0 is a constant of integration that specifies the expansion rate at t=0, the beginning of the Guth era. For a Guth era beginning fairly soon after the Planck time, we have $H_0^2 C < 1$ but not $<<1$.

Equation (27a) has the asymptotic forms

$$R(t) \sim \exp(H_0 t) \ , \ t \lesssim \frac{2}{3CH_0^3} \tag{29a}$$

$$\sim \exp\left[\left(\frac{3}{2}\right)^{2/3} \frac{t^{2/3}}{C^{1/3}}\right] \ , \ t \gtrsim \frac{2}{3CH_0^3} \tag{29b}$$

One sees that the expansion will proceed for $\sim(CH_0^2)^{-2}$ e-folds as a pure exponential (eq. 29a), and with the temperature and total density about constant (eqs. 27b and 27c). The density is dominated by the vacuum energy density (cosmological constant), while the temperature is dominated by the horizon temperature. After $\sim(CH_0^2)^{-2}$ e-folds, the accumulated work of expansion on the thermal energy content (nonzero PdV work, since $\Gamma \neq 0$) starts to make itself felt, and the expansion slows from a pure exponential to an exponential of a fractional power (eq. 29b). At the same time, the horizon temperature starts falling (eq. 27b), but only as a power law in time. One therefore has some additional finite number of e-folds of expansion before the temperature has fallen sufficiently so as to cause a spatially simultaneous loss of metastability of the false vacuum. At this point, the Guth era "exits gracefully" and we have a smooth, reheated Friedman cosmology.

It may seem paradoxical that the above set of equations give a slowly decreasing ρ_{vac} (as a consequence of eqs. 27c and 27d), when ρ_{vac} is supposed to be determined by the microphysics of the effective potential V. Without a proper quantum gravitational treatment, we cannot resolve this paradox in detail; however it is easy enough to see where the resolution must lie: In eq. (4a) we already included angle brackets denoting quantum expectation values in curved space. Classically, any constant included in the Lagrangian (eq. 5) appears unmodified in eq. (4a). Quantum gravity, however, can renormalize the cosmological constant and modify the zero point. Our semiclassical

equations know about this through the requirement that total energy be conserved semiclassically (eq. 2). Therefore even without the proper quantum gravitational machinery, it is plausible that we are obtaining a correct result at semiclassical order.

A tempting conjecture is that in a proper quantum gravitational treatment the evolution should turn out to be independent of any constant added to eq. (5), i.e. that any such constant should get renormalized away in a natural way in the limit of low temperature. This would then explain a central mystery not previously mentioned in this paper, namely why the potential minimum at low temperature (Figure 1) should have the value exactly zero, i.e. why the value of the cosmological constant is observationally very small today.

The total entropy of today's visible universe (total number of particles present) is a very large number. The standard Guth scenario explains this number as the result of exponential expansion and extreme supercooling of a metastable vacuum phase; the entropy is produced when the false vacuum comes crashing down via a first order phase transition. However, it has not been clear how to obtain the benefit of this latent heat without doing violence to spatial homogeneity. When one considers the finite temperature of the event horizon, the scenario is slightly different: most of the entropy is produced during the exponential expansion, not at its termination. The source of the entropy is, in some sense, gravitational. Just as in the black-hole case, the deSitter event horizon is an object of intrinsically very high entropy content, and it shares this entropy with an exponentially expanding volume. The final phase transition to the true vacuum, in this picture, is only a nicety, put in so as to bring the cosmology to today's state of vanishing cosmological constant.

V. CONCLUSION

As a sometime pessimist, I do not think we will ever know with any confidence the detailed nature of particle microphysics at energies $\sim 10^{15}$ GeV, much less at the Planck temperature $\sim 10^{19}$ GeV. We may well come to understand, however, the generic sorts of microphysics that are allowed at those energies. There seems to be an unexpected richness in the cosmological implications of some of these sorts of microphysics, allowing for possible (albeit probably forever

speculative) explanations of such macroscopic puzzles as the origin of galaxies in the universe, and the nature of the universe's large-scale structure generally.

As a sometime optimist, I think that it would be unusual for a scientific subject to spend long at a stage of development that allows for semiquantitative speculation, but that does not seem to allow for a rigorous and decisive treatment. So perhaps the field of ultra-high-energy cosmology has surprises in store for us.

ACKNOWLEDGEMENTS

Most of this paper derives from discussions with Doug Eardley, Alan Guth, Bob Wald, Stephen Hawking, Sidney Coleman, and Howard Georgi. This work was supported in part by the US National Science Foundation, PHY 80-07351.

REFERENCES

Birrell, N.D., Davies, P.C.W., and Ford, L.H. 1980, J. Phys. A. 13, 961.
Brown, L.S. and Cassidy, J.P. 1977, Phys. Rev. D. 15, 2810.
Callan, C.G. and Coleman, S. 1977, Phys. Rev. D. 16, 1762.
Candelas, P. 1980, Phys. Rev. D. 21, 2185.
Candelas, P. and Dowker, J.S. 1979, Phys. Rev. D. 19, 2902.
Coleman, S. 1977, Phys. Rev. D. 15, 2929.
Coleman, S. and DeLuccia, F. 1980, Phys. Rev. D. 21, 3305.
Davies, P.C.W. and Fulling, S.A. 1977, Proc. Roy. Soc. A 356, 237.
Davies, P.C.W., Fulling, S.A., Christensen, S.M., and Bunch, T.S. 1977. Ann. Phys. (New York) 109, 108.
DeWitt, B.S. 1975, Phys. Reports 19, 295.
Doroshkevich, A.G., Zel'dovich, Ya. B., and Sunyaev, R.A. 1974, in Confrontation of Cosmological Theories with Observational Data, ed. M.S. Longair (Dordrecht: Reidel).
------ 1978, Astron. Zh. (USSR) 55, 913.
Doroshkevich, A.G. Zel'dovich, Ya. B., Sunyaev, R.A., and Khlopov,U. Ya., 1980, Pis'ma Astron. Zh. (USSR) 6, 457.
Dowker, J.S. and Critchley, R. 1977, Phys. Rev. D. 16, 3390.
Gibbons, G.W. and Hawking, S.W. 1977, Phys. Rev. D. 15, 2738.
Gott, J.R., Gunn, J.E., Schramm, D.N., and Tinsley, B.M. 1974, Astrophys. J. 194, 543.
Guth, A.H. 1981, Phys. Rev. D. 23, 347.
Hawking, S.W. and Ellis, G.F.R. 1973, The Large Scale Structure of Space- Time (Cambridge: Cambridge University Press).
Kahn, R. 1981, paper in preparation.
Lapedes, A.S. 1978, J. Math. Phys. 19, 2289.
Lightman, A.P., Press, W.H., Price, R.H., and Teukolsky, S.A. 1975, Problem Book in Relativity and Gravitation (Princeton: Princeton University Press.
Press, W.H. and Vishniac, E.T. 1980, Astrophys. J. 239, 1.
Wald, R.M. 1978, Phys. Rev. D. 17, 1477.

NEUTRINOS OF FINITE REST MASS IN ASTROPHYSICS AND COSMOLOGY

Ramanath Cowsik
Tata Institute of Fundamental Research, Bombay - 400005, India

Our early work on the role of neutrinos in astrophysics and cosmology indicating that 3 eV < m_ν < 35 eV and that $\tau_\nu/m_\nu > 10^{23}$ s/eV for radiative decays is presented and then extended to include the recently discovered τ-neutrino.

Des champions, ici, luttent férocement
pour la maîtrise des choses, apportant
à la bataille leurs embryons d'atomes... Celui
qui emporte l'adhésion de la plupart, fait la
loi un moment : le Chaos est arbitre,
et par ses décisions augmente encore le
désordre par lequel il règne : à ses côtés,
la Chance, arbitre suprême, gouverne le tout.

Fernand

I. INTRODUCTION

Astrophysics is usually the science in which the well-known laws of physics are assumed to hold good even on the macroscopic scale of astronomical bodies and on this premise we try to understand the nature and properties of celestial objects. However the inverse situation, where studies outside man-made laboratories, in nature, yield many insights into basic physics and, in particular, into the properties of the fundamental particles, is also quite common. The application of this general concept, especially in the study of the interplay between big-bang cosmology and the modern theories of particle physics, has become quite wide-spread and well accepted. Perhaps the very first result from such a study, which exceeded substantially in quality the results from laboratory experiments, concerns the masses of the neutrinos[1]. To-day I will review some of my early work pertaining to neutrinos of finite rest mass. I will then go on to discuss how one can use the laboratory experiments in conjunction with astrophysical and cosmological considerations to show that the newly discovered τ-neutrino is extremely stable and that $m_{\nu_\tau} < 35$ eV.

Neutrinos of finite rest mass have fascinated physicists for a long time and we are probably indebted to Markov (1964), Gerstein and Zeldovich (1967) and to Bahcall, Cabbibo and Yahil (1971) for some of the very first applications to astrophysics and cosmology. Markov and also Bludman (1974) reviewed the field until then and suggested the possibility of neutrino stars[2,3]. Gerstein and Zeldovich[4] presented the first qualitative ideas and showed that the measured age of the Universe restricts the mass of a neutrino to be less than $\sim$ 140 eV, and Bahcall et al. argued that if neutrinos had a finite rest mass, they could decay into lighter neutrinos and bosons and used this as a possible

basis for understanding the low counting rate in the 'Solar-Neutrino' experiment.[5] My own interest in neutrino induced phenomena[1,6-9] in nature and their cosmological effects has been a longstanding one (Cowsik et al. 1963, 1964, 1966, 1969) and on the particular topic of neutrinos of finite rest mass, I have considered basically three questions:

i. The effect of massive neutrinos in cosmology, leading to an upper limit[1] on the sum of their masses (Cowsik and McClelland, 1972).

ii. Clustering of neutrinos[10] and the virial mass discrepancy in the Coma-cluster of galaxies (Cowsik and McClelland, 1973) which yields a lower limit on the neutrino mass.

iii. The stability of neutrinos[11] and limits on their decay rate (Cowsik, 1977, 1979).

Topic i was also considered by Marx and Szalay[12] (1972). Topic ii was studied somewhat later, (Marx and Szalay[13], 1976) and supportive arguments for topic iii was given by Falk and Schramm[14] (1978).

After this; many important developments took place concerning the question of massive neutrinos: Lee and Weinberg[15] considered the contribution of very massive neutrinos $m_\nu > 1$ MeV/c^2 to the cosmological energy density (see also Dicus et al.[16] 1977), Gunn et al. (1978) discussed the astrophysical consequences of the existence of a very heavy neutrino[17] with $m_\nu \gtrsim 2$ GeV/c^2, Tremaine and Gunn[18] (1978) emphasized the fact that neutrinos behave like a collisionless gas and would therefore evolve without any change in density in phase space. Finally Steigman and Schramm[19] (1980) have reviewed the most recent astronomical observations which lend support to the idea that neutrinos indeed are responsible for binding the large clusters of galaxies.

The theoretical implications of neutrinos of finite rest mass have been discussed by several authors[20-23], notably by Gellmann et al. (1975), and

reviewed with extensive references by Mohapatra and Senjanovich (1980), by Pakhvasa (1980) and by Witten (1980).

With the news of a possible experimental detection of a finite rest mass for the neutrino in the range $14\ eV/c^2 \quad m_\nu < 46\ eV/c^2$ by Lyuibimov et al.[24] (1980) and oscillations of neutrinos with a mass difference $\Delta m^2 \sim 1\text{-}10\ eV^2$ by Reines et al.[25], there has been a great resurgence of interest. The theoretical ideas and the astronomical data pertaining to the question of massive neutrinos are being scrutinized carefully again by de Rejula and Glashow, Stecker, Kimble et al., Cowsik and Shipman[26-30].

In this talk, I shall mainly review my early work on the role of massive neutrinos stated in terms of the three basic questions i, ii and iii, cited above. Near the end of the talk, I shall show how the τ-neutrino could also be included in these discussions.

II. UPPER LIMIT ON THE NEUTRINO REST MASS

The hypothesis that the neutrino mass is zero, though aesthetically very appealing, has had rather limited validation by experiments performed in the laboratory. By looking for cut-offs in the spectrum of β-decay electrons and muons on a Kurie plot the following limits have been placed on the neutrino masses[31-33].

$$m_{\nu e} < 60\ eV/c^2; \quad m_{\nu\mu} < 0.5\ MeV/c, m_{\nu\tau} < 250\ MeV/c^2 \tag{1}$$

These limits, particularly for ν_μ and ν_τ, certainly need further improvement to validate the canonical assumption of masslessness. The idea behind the astrophysical method is that if one has a sufficiently large number of neutrinos, one can weigh them, as the gravitational force due to a large ensemble of neutrinos and anti-neutrinos add up. The plan is therefore to detect the net-

gravitational interaction of the large number of neutrinos in the Universe using the galaxies as test particles. In discussing this problem, we take the customary point of view that the Universe is expanding from an initially hot condensed state as envisaged in the big-bang cosmology.[34]

We will now proceed to calculate the number density of neutrinos and see what effect they have on the motion of the galaxies. In the early phase of the Universe when the temperature was greater than $\sim$1 MeV, processes of neutrino production[35] which have also been considered in the context of high temperature stellar cores would lead to a generation of various kinds of neutrinos. Some of these reactions are listed below.

$$\begin{aligned} e^+ + e^- &\rightarrow \nu + \bar{\nu} \\ e + z &\rightarrow e + z + \nu + \bar{\nu} \\ \gamma_{plasmon} &\rightarrow \nu + \bar{\nu} \end{aligned} \qquad (2)$$

Since weak interactions can also proceed through neutral currents, one could have all the kinematically allowed neutrino pairs (ν_e $\bar{\nu}_e$, ν_μ ν_μ, etc.) on the r. h. s. of the above reactions. The actual density of neutrinos then is to be obtained by balancing, in detail, the rates of such processes with the rate of destruction of neutrinos through mutual annihilation and other such processes. Fortunately, however, this horrendous task is simplified, at least for the ν_e and ν_μ (with $m < 1$ MeV/c^2), by the consideration that, in the early Universe, the density of e^+ and e^- is so high that the reactions rapidly drive the neutrinos into thermodynamic equilibrium. With this realization, we can immediately write for the number density of ν_e or ν_μ and, in fact, for that of any of the fermions and bosons to be

$$n_{Fi} = \frac{g_i}{2\pi^2\hbar^3} \int_0^\infty \frac{p^2 dp}{\exp[E/kT(z_{eq})] + 1} \tag{3}$$

and

$$n_{Bi} = \frac{g_i}{2\pi^2\hbar^3} \int_0^\infty \frac{p^2 dp}{\exp[E/kT(z_{eq})] - 1} \tag{4}$$

Here n_{Fi} = number density of fermions of the ith kind

n_{Bi} = number density of bosons of the ith kind

g_i = effective degeneracy of the spin states (~ 2)

E = $c(p^2 + m^2c^2)^{1/2}$

k = Boltzmann's constant

$T(z_{eq})$ = $T_{rad}(z_{eq}) = T_F(z_{eq}) = T_B(z_{eq})$

= common temperature of all the particles in thermal equilibrium at the epoch characterized by the red-shift z_{eq}.

Initially, to get the general idea of the method, let us consider only the light particles like ν_e and ν_μ with

$$kT(z_{eq}) \gtrsim 1 \text{ MeV} > mc^2 \tag{5}$$

Then Eqs. 3 and 4 integrate to

$$n_{Fi}(z_{eq}) \approx 0.0913\, g_i\, [T(z_{eq})/\hbar c]^3$$

and (6)

$$n_{Bi}(z_{eq}) \approx 0.122\, g_i\, [T(z_{eq})/\hbar c]^3$$

As the universe expands and cools down, the neutrinos and other such particles survive without annihilation because of the extremely small cross-sections for the annihilation of light neutrinos. Anticipating our discussion in Sec. IV, we will neglect also the decay of these particles so that their number density decreases with the Universal expansion simply as

$$n(z) = n(z_{eq}) \frac{V(z_{eq})}{V(z)} = n(z_{eq}) \frac{(1+z)^3}{(1+z_{eq})^3} \tag{7}$$

Since $(1 + z) = T_{rad}(z)/T_{rad}(0)$, the number density at the present epoch is given by

$$n_{Fi}(0) = n_{Fi}(z_{eq})(1+z_{eq})^{-3} = 0.0913\, g_i \left[\frac{T_{rad}(0)}{\hbar c}\right]^3$$

$$n_{Bi}(0) = 0.122\, g_i \left[\frac{T_{rad}(0)}{\hbar c}\right]^3 . \tag{8}$$

Using $T_{rad}(0) = 2.7^{\circ}$ K, as the present day radiation temperature of the universe, we get

$$n_{Fi}(0) \approx 150\, g_i\ cm^{-3};\ n_{Bi}(0) \approx 200\, g_i\ cm^{-3}. \tag{9}$$

These numbers are huge in comparison with the number density of baryons[36] in the Universe $\lesssim 4 \times 10^{-6}$ cm^{-3}. Consequently, even if the neutrinos have a very small rest mass, they would dominate the dynamics of the universe.

We will now proceed to set an upper limit to the energy density in the Universe, contained in all forms. The cosmological expansion of the Universe as traced by the red-shift of the galaxies is described in terms of two parameters:

a) The Hubble constant which measures the rate of change of expansion velocity with distance, designated by $H_0 \approx 50$ km/sec/Mpc and,

b) The deceleration parameter , the second derivative of the expansion velocity, designated by $q_0 \approx 0.94 \pm 0.4$. These quantities are taken from measurements by Sandage. There is a third quantity which pertains to the expansion of the Universe, namely,

c) The age of the Universe which is designated by t_0 and is shown to be

longer than 8×10^9 yr using observations of old globular clusters. These three parameters are not independent of each other but are related through the equation[34]

$$t_0 = H_0^{-1} q_0 (2q_0 - 1)^{-3/2} \left[\cos^{-1}\left(\frac{1}{q_0} - 1\right) - \frac{1}{q_0}(2q_0 - 1)^{1/2}\right] \qquad (10)$$

Equation (10) implies $q_0 < 1.2$ quite consistent with Sandage's[37] direct determination using plots of red-shifts of galaxies versus their magnitudes. The Hubble parameter H_0 and the deceleration parameter q_0 define clearly the dynamics of an isotropic, homogeneous, expanding Universe with a density, ρ_{tot}, given by the expression

$$\rho_{tot} = \frac{3 H_0^2 q_0}{4 G} < 1.3 \times 10^{-29} \text{ g cm}^{-3} < 8 \times 10^3 \text{ eV/c}^2 \text{ cm}^{-3}. \qquad (11)$$

This 8 keV cm^{-3} is the total energy density in all forms in the Universe. The weakly interacting particles alone would contribute a density ρ_{weak} given by

$$\begin{aligned} \rho_{weak} &= \Sigma\ n_{Bi} m_i + n_{Fj} m_j \\ &= \Sigma\ 200\, g_i m_i + \Sigma\ 150\, g_j m_j < \rho_{tot}. \end{aligned} \qquad (12)$$

Or, simply

$$\Sigma\ g_i m_i < 53 \text{ eV/c}^2. \qquad (13)$$

Taking $g_i = 2$ and keeping in mind that particle and antiparticle masses are the same,

$$\Sigma\ m_i < 13 \text{ eV/c}^2 \qquad (14)$$

This means that sum of the masses of all the light neutrinos and any hypothetical weakly interacting bosons is less than 13 eV/c^2. However, in the calculations performed so far, we have neglected the fact that the positrons which were

also created copiously would have annihilated and raised the temperature of the Universe[34] by a factor of ~ 1.4. This means that we should have used $T_{rad} = (2.7/1.4)^{o}K$ in the formula for the densities in Eq. 8. This leads to the slightly less restrictive limit

$$\Sigma\, m_i < 13 \quad x\,(1.4)^3 \ eV/c^2 < 35 \ eV/c^2. \tag{15}$$

If a neutrino of a particular type saturates this limit the other neutrinos should have zero mass. This limit of 35 eV/c^2 is a considerable improvement over the limit of 60 eV/c^2 for ν_e obtained from the study of tritium-decay and certainly reduces the 0.5 MeV/c^2 limit on ν_μ dramatically.

We have derived the above limit assuming that the masses of the neutrinos are smaller than $\sim$ MeV and that they were in thermodynamic equilibrium with the radiation field. If the mass of a neutrino is much larger than ~ 1 MeV as could very well be true, a priori, for ν_τ, then such neutrinos would move away from thermodynamic equilibrium and their number densities at present would become progressively much smaller with increasing mass, than the numbers predicted by Eq. 9. The actual numbers are controlled by a detailed balancing of the reactions that produce them and their annihilation in a rapidly expanding Universe. Lee and Weinberg[15] have made these estimates and were the first to show that unless m_ν is greater than 2 GeV, their cosmological number densities will be too large to be consistent with the value of ρ_{tot} defined in Eq. (11). Similar calculations[16] were also performed by Dicus et al. and their results are well exhibited in the following figure[17] taken from Gunn et al. Thus, the main results are

$$\begin{aligned} &\Sigma\, m_{\nu i} < 35 \ eV/c^2 \text{ for } m_\nu < 1 \ MeV/c^2 \\ &\quad m_{\nu i} < 2 \ GeV/c^2 \end{aligned} \tag{16}$$

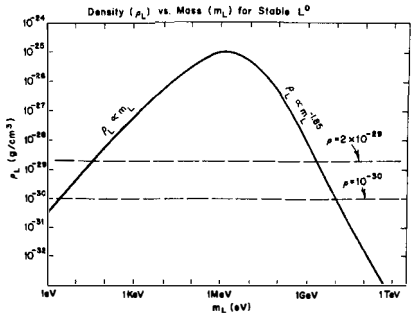


Fig. 1. The density of the Universe contributed by neutrinos of various masses.

Recently Joshi[38] has shown that a relation between the age and energy density in the Universe obtains for any globally hyperbolic space-time. This would then imply that the mass limit expressed in Eq. 16 is valid independent of the cannonical assumption of a Robertson-Walker metric to describe the universal expansion[39].

III. LOWER LIMIT TO m_ν FROM VIRIAL MASS DISCREPANCY OF RICH CLUSTERS OF GALAXIES

We have shown in Sec. II that a very large number of neutrinos produced in the big bang fill the Universe and that even if they should have a rest mass of only a few eV/c^2 they would dominate the gravitational dynamics of the Universe. This is particularly true when we consider the fact that the cosmic density of all visible matter is only $\sim \rho_{crit}/100$. One consequence of this is that through their mutual gravitation interactions, neutrinos may have triggered the initial condensations that led to the formation of clusters of galaxies. If this is true, then one may expect that in large clusters there might be substantial amounts of unseen mass in the form of neutrinos. This expectation is

indeed borne out by the observations of several clusters of galaxies[36]. Several careful analyses of the peculiar motion of individual galaxies in the Coma cluster indicate that there is a substantial discrepancy between the sum of the masses of the individual galaxies (5×10^{48} g for $H_0 \approx 50$ km s^{-1} Mpc^{-1}) and the mass required to gravitationally bind the cluster ($\sim 4 \times 10^{49}$ g). These analyses also indicate that this discrepant mass is not concentrated in a massive black hole but has a smooth distribution concentrated near the center, following the general pattern traced by the galaxies themselves.

To check, qualitatively, the hypothesis that neutrinos of finite rest mass are responsible for this binding, we construct a simple model of the Coma cluster in the form of a gravitational potential well of constant depth extending over the core region of high galactic density ($R_c \approx 0.7$ Mpc $\approx 2.1 \times 10^{24}$ cm). This well is then filled with neutrinos which are treated as a Fermi-Dirac gas at zero temperature. Notice that the assumption of zero temperature for the neutrino gas yields the minimum rest mass needed to bind the cluster; with finite temperatures one would need increasingly larger masses up to the limits derived in Sec. II. The procedure adopted here is identical to that used in developing the Thomas-Fermi model for atoms. The total number of neutrinos that can be accommodated inside the well is given by

$$N_\nu = \frac{4V}{3\pi^2} \left(\frac{2 m_\nu U_\omega}{h^2} \right)^{3/2} \tag{29}$$

Here we have assumed two helicity states for each of ν_μ, $\bar{\nu}_\mu$, ν_e and $\bar{\nu}_e$. Also, m_ν = common rest mass assumed for each of the neutrinos; $V = \frac{4}{3} \pi R_c^3$ = volume of the potential well; $U_\omega = G M_c m_\nu / R_c \approx 1.3 \times 10^{17}$ erg g^{-1} = depth of the potential well = kinetic energy needed for the neutrinos to escape from the cluster. Notice that for M_c, we have used the total discrepant mass of the cluster = 4×10^{49} g.

Now, in order that the neutrinos may be responsible for binding the cluster we set $M_c = N_\nu\, m_\nu$; substituting for U_ω in Eq. 29, we find

$$m_\nu^8 = \frac{3^4 \pi^2 \hbar^6}{2^{11} G^3 R_e^3 M_c} \approx 0.5\, \frac{\hbar^6}{G^3 R_c^3 M_c}. \tag{30}$$

Alternatively following Landau and Lifshitz[40], the self-consistent equilibrium of a gravitating cloud of Fermi gas at zero temperature yields the result

$$m_\nu^8 \approx 6\, \frac{\hbar^6}{G^3 R_c^3 M_c}. \tag{31}$$

Because of the very high power of m_ν involved in the equations (28) and (29), they both yield essentially the same results: $m > 4 \times 10^{-33}$ g ≈ 3 eV/c^2. In Fig. 2, we show the surface density calculated in the self-consistent potential with the observed distribution of galaxies[36,52].

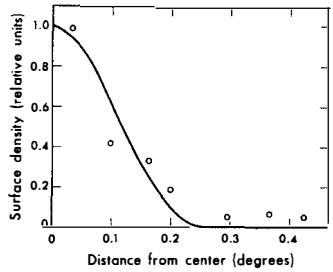


Fig. 2. The radial distribution of neutrino density compared with that of the galaxies. One has to include the effects of finite neutrino temperature to reproduce the tail in the distribution.

Notice the rather good fit between theory and observation in the central regions of the cluster. One should include effects of finite temperature of neutrinos to reproduce the long tail in the distribution of the galaxies. Tremaine and Gunn[18] have pointed out a very interesting fact, that the weakly interacting neutrinos evolve as a collisionless gas and by Liouville's theorem, the

thermal neutrinos will preserve phase space densities with occupation number 1/2 instead of 1 as implied in Eq. 29. Two points are worth noting in this regard: a) If the phase space distribution is lumpy to start with, then these lumps can be brought together through the dynamical friction with ordinary matter in the cluster which can radiate away energy and b) models with an occupation number of 1/2 are not very different from the one that we have presented. Before closing this section, it is interesting to note that the total mass in neutrinos contained within the horizon at a red-shift z behaves as

$$M_H = \frac{4 \times 10^{53} \, (m_\nu c^2/1 \text{ eV})}{(1+z)^{1.5}} \text{ g} . \qquad (32)$$

Equating M_H and the visual mass M_c of the Coma-cluster, we find $z \gtrsim 400\text{-}1000$ for $m_\nu c^2 \gtrsim 1\text{-}10$ eV. The kinetic energy of neutrinos at that time is $kT_\nu \gtrsim 2 \times 10^{-7} (1+z)^2 / m_\nu c^2/1 \text{ eV})$ and the Jeans mass M_J is of the order of M_H.

IV. LIMITS ON RADIATIVE DECAY OF NEUTRINOS

In this section, I discuss the limits that can be placed on the radiative instability of neutrinos from a variety of observations. In this regard, I am motivated by several papers[5,41-44] which have considered the possibility that neutrinos could have finite rest mass and could, therefore, decay. One particular set of these[43,44] considers the mixing of ν_e and ν_μ and predicts observable widths for the lepton-number-nonconserving decays, such as $\mu \to e + \gamma$ and $\nu_\mu \to \nu_e + \gamma$. Independent of such theoretical considerations, it is worthwhile to study the observational limits on such processes.

The astrophysical environment provides excellent possibilities for such a study of very weak processes: Pathlengths of $\sim 10^{28}$ cm are available for the decay process to take place, huge in comparison with the $\sim 10^2$ cm available in

most laboratory studies. Also, there are regions, such as the cores of very hot stars, where weak processes dominate, as the products of the competing electromagnetic channels are suppressed completely because of the enormous time scales needed for the diffusion of photons to the stellar surface. [45]

Recognizing that both neutral and charged currents are of comparable strength in weak interactions, one finds that there are several locations in nature where copious generation of ν_e and ν_μ etc., takes place. Consider, at first, the possibility that the neutrino decays solely through the channel

$$\nu \rightarrow x + \gamma \tag{17}$$

where x is any particle with a mass smaller than m_ν. The fraction of the neutrino energy carried by the photon is $\eta = (m_\nu^2 - m_x^2)/2\, m^2 \approx 0.5$ if m_x is small compared with m_ν. We can then make use of the observational limits on the photon intensities from the sources of neutrinos to place limits on the radiative stability of the neutrino. We shall proceed to derive these limits, considering sequentially, the cases where progressively longer times are available for the neutrinos to decay.

IV. 1. Studies at Particle Accelerators

Here the neutrino fluxes are generated by the decay of mesons and muons produced by a high energy proton beam interacting with a target of nuclei. The neutrino fluxes at CERN have a mean energy of $\sim$1 GeV and their decay will yield gamma rays of similar energy. These γ-rays would easily be detected in the spark chambers used in the 'neutral current' experiment. The decay length available for the neutrinos is at least several metres ($\sim 10^2$ cm). We can safely assert that the number of gamma rays that would have been produced, arriving in the direction of the beam, due to the decay of neutrinos is

less than the total number of neutral current events which have a similar signature. Equating the expected number of decays of neutrinos in a $\sim 10^2$ cm path with the number of ν-induced events ($\sigma \sim 10^{-39}$ cm^2) in a target of thickness $N_T \gtrsim 10^{25}$ nuclei/cm^2, we have

$$\tau_\nu = \frac{F_\nu}{F_\gamma} \Delta t > \frac{1}{N_T \sigma} \frac{\ell}{c} \quad \text{or} \quad \tau_\nu > 3 \times 10^6 \text{ s} . \tag{18}$$

Since $\tau_\nu = \tau_0 E_\nu / m_\nu c^2$ and $E_\nu \gtrsim 10^9$ eV

$$\tau_0 / m_\nu c^2 > 3 \times 10^{-3} \text{ s/eV}. \tag{19}$$

IV. 2. Atmospheric Neutrinos

Cosmic rays produce neutrinos[7,46] in the earth's atmosphere in a manner very similar to that described in Sec. IV. 1 and the events induced by these are recorded by instruments placed deep underground[47]. The mean energy of the neutrinos is again $\sim$ 1 GeV and the flux is $\sim 10^{-13}$ cm^{-2} s^{-1} sr^{-1}. The total event rate, R, is 10^{-13} cm^{-2} s^{-1} sr^{-1}. The decay length available at the apparatus is several metres. We can therefore write

$$\tau_\nu \frac{F_\nu}{R} \frac{\ell}{c} > 10^4 \text{s} \quad \text{or} \quad \frac{\tau_0}{m_\nu c^2} > 10^{-5} \text{ s/eV} \tag{20}$$

IV. 3. Solar Neutrinos

If we make the eminently reasonable assumption that the energy generation in the sun is due to the synthesis of protons to helium nuclei, then at the earth we expect a neutrino flux of $\sim 10^{11}$ cm^{-2} s^{-1} irrespective of any specific solar models or details of nuclear reaction chains[45]. These neutrinos have a mean energy of $\sim$ 200 keV and their decay during the $\sim$ 500 s flight from the sun to the earth would results in an intense X-ray flux. The observations indicate that the X-ray flux from the quiet sun is below the level of detectability at $\sim 10^{-4}$ cm^{-2} s^{-1} this yields

$$\tau_\nu > \frac{10^{11} \times 500}{10^{-4}} = 5 \times 10^{17} \text{ s or } \frac{\tau_0}{m_\nu c^2} > 2 \times 10^{12} \text{ s/eV} \qquad (21)$$

IV. 4. White Dwarfs and Central Stars of Planetary Nebulae

It is well-known[48,49] that these stars cool rapidly by neutrino emission. An order of magnitude estimate of the neutrino fluxes emitted by such objects can be obtained by equating the gravitational energy released in their formation $\sim GM^2/R$, with the energy carried away by the neutrinos. With the assumptions $M = M_0 = 2 \times 10^{33}$ g and $R = 10^{8.5}$ cm, a total flux of $\sim 10^{58}$ neutrinos at $E_\nu \gtrsim 100$ keV is radiated into the Universe during the formation of such an object. The rate of formation of white dwarfs is $\sim 1\ \text{yr}^{-1}$ in a galaxy. With $M_{gal} \gtrsim 10^{44}$ g and the mean density of the Universe due to galaxies, $\rho_\mu \gtrsim 10^{-31}$ g cm^{-3} one will expect an X-ray flux from space of

$$F_\gamma \gtrsim \frac{GM^2}{RE_\nu} \frac{\rho_\mu \tau_\mu R_\mu}{3 \times 10^7 \times 10^{44} \tau_\nu} . \qquad (22)$$

Here $\tau_\mu = 10^{18}$ s is the 'age' of the Universe and $R_\mu \gtrsim 10^{28}$ cm is its 'radius'. The observations of x-ray astronomers[50] yield a flux limit $F_{X\text{-ray}} < 10^{-1}$ cm^{-2} s^{-1} sr^{-1} which corresponds to

$$\tau_\nu > 10^{22} \text{ s or } \tau_0/m_\nu c^2 > 10^{17} \text{ s/eV} \qquad (23)$$

IV. 5. Supernovae

These occur in a galaxy once in a hundred years and radiate energies about a few hundred times larger than emitted in the formation of a white dwarf. The mean energy of the radiated neutrinos is ~ 10 MeV, and with the gamma-ray flux limits of $\sim 10^{-3}$ cm^{-2} s^{-1} sr^{-1} one obtains

$$\tau_\nu > 3 \times 10^{23} \text{ s or } \tau_0/m_\nu > 3 \times 10^{16} \text{ s/eV.} \qquad (24)$$

Falk and Schramm[14] have elaborated on this point lending further support to these results.

IV. 6. Neutrinos in "Big Bang" Cosmology

Noting that neutrinos of energy $\sim$100 keV live longer than 10^{22} s, one realizes that neutrinos generated during the condensed phase of the Universe will have survived until the present epoch. Thus, the limits on the neutrino masses derived in Sec. II using the dynamical effects of the cosmological neutrinos on the present day expansion of the Universe are valid. If m_ν is indeed larger than $\sim 10^{-3}$ eV/c^2, these neutrinos would have behaved as a non-relativistic gas during the expansion of the Universe at red-shifts $z < 1$ (i. e., during the relatively recent past of $\sim 3 \times 10^{10}$ yr). This expansion would have slowed the neutrinos down to non-relativistic velocities and any possible decay would yield photons with energy comparable to their rest mass. The number density of any one type of neutrinos and antineutrinos is $\sim$400 cm^{-3} and yields a photon flux of

$$F_\gamma \simeq \frac{n_\nu R_\mu}{\tau} \approx 10^{31}/\tau_0 \qquad (25)$$

(assuming $\tau_0 \approx \tau_u \approx 10^{18}$ s). The background photon flux has a maximum at the peak of the blackbody curve at 2. 7 ^{o}K, where the relic microwave is observed[51]. In order that this spectrum is not distorted substantially

$$\tau_0 > 10^{19} \text{ s} \quad \text{if} \quad m_\nu \approx 10^{-3} \text{ eV}/c^2. \qquad (26)$$

On the other hand, if $m_\nu \approx 1$ eV/c^2, their decay would yield a flux of optical photons. Using the observational limit on the background starlight flux of $\sim 3 \times 10^{8}$ cm^{-2} s^{-1} provides the limit

$$\tau_0 > 10^{23} \text{ s} \quad \text{if} \quad m_\nu \approx 1 \text{ eV}/c^2. \qquad (27)$$

Such considerations have been elaborated upon recently by De Rujula and Glashow[26], Stecker[27], Kimble et al.[28] and by Shipman and Cowsik[30]. In Table I, we summarize the limits derived in sections III. 1-6. Notice that the limits based on accelerators and atmospheric cosmic ray neutrinos are necessary to preclude the possibility that the neutrinos will decay inside the stellar sources themselves. The limits derived from the stellar sources allow us, in turn, to consider the neutrinos of cosmological origin and show that these neutrinos would have survived decay until the present epoch.

IV. 7. Effect of Competing Channels of Neutrino Decay

Let us now consider, briefly, the consequences of the existence of competing processes to the radiative decay of the neutrino. For example, the very existence of the vertex $\nu \to x + \gamma$ implies a corresponding electromagnetic interaction of ν with the Coulomb fields of nuclei: $\nu + Z \rightleftarrows Z + x$. This interaction can possibly generate a comparatively rapid "stimulated decay" chain $\nu \to x \to \nu \to x \ldots$ during the transit of the neutrino from the stellar interiors to the surface. Since the stars do evolve by rapid loss of energy in some form, one can assert that the particles do escape from the star, still retaining a good fraction of their energy. Under these circumstances, the limits derived above are not affected substantially.

There is also the possibility, though theoretically unlikely, that other relatively rapid modes of decay besides $\nu \to x + \gamma$ exist, such as

$$\nu \to x + y + z + \ldots . \tag{28}$$

In this hypothetical decay, the sum of the masses of $x,y,z,\ldots$ should be less than m_ν and none of the particles can satisfy this condition, other than neutrinos and photons of course. And yet, if any of $x,y,z,\ldots$ is easily observable, we

may try to estimate the decay rate from the flux of the decay product and discuss its importance relative to radiative decay. If however the decay products have weak interactions alone, one may present the following arguments. Using the studies of ν-induced reactions at nuclear reactors and cosmic ray experiments deep underground one can set $T_0^{\nu_e}/m_\nu c^2 > 3 \times 10^{-13}$ s/eV and $T_0^{\nu_\mu}/m_\nu c^2 > 3 \times 10^{-11}$ s/eV, where T_0 is the lifetime due to all the decay modes. These limits are not very restrictive, and if $T_0/m_\nu c^2 < 10^{-5}$ s/eV the neutrinos will decay inside the stars and the limits derived earlier will then apply to the radiative instability of the decay products x, y, z, ...

On the other hand, on general considerations, the expected lifetime for the decay of neutrinos into only weakly interacting particles is rather long. Here I wish to consider two cases: (a) When the lifetime T for the competing process is longer than or comparable to the typical times involved in the discussions in Sec. III. 3-6 the limits remain essentially unaffected (i. e. , if $T > 500$ s for the solar neutrinos and $T > 10^{18}$ s for the cases III. 4-6). (b) If 10^{18} s $< T < 1$ s, then the neutrinos will escape the stellar regions ($T/m_\nu c^2 > 10^{-5}$ s/eV) and their radiative decay can become observable depending on the relative strengths of radiative decay and the other decay modes. Arguments presented in Sec. III. 3-5 imply

$$\tau_0^{\nu_e}/T_0^{\nu_e} > 10^{15} \quad \text{or} \quad \frac{\Gamma(\nu_e \to x + \gamma)}{\Gamma\ (\text{total})} < 10^{-15} \tag{21'}$$

$$\frac{\tau_0}{T_0} > 10^4 \quad \text{or} \quad \frac{\Gamma(\nu \to x + \gamma)}{\Gamma\ (\text{total})} < 10^{-4} \ \text{(all types)} \tag{23'}$$

$$\frac{\tau_0}{T_0} > 3 \times 10^5 \quad \text{or} \quad \frac{\Gamma(\Gamma \to x + \gamma)}{\Gamma\ (\text{total})} < 3 \times 10^{-6} \ \text{(all types)} \tag{24'}$$

Table I - Limits on Neutrino Lifetimes

Source	Neutrino Density	E_ν	Observational Evidence	Lower Limit τ_o/m_ν
Accelerators	10^8/pulse	~ GeV	Total no. of events	3×10^{-3} sec/eV
Cosmic Ray Neutrinos	0.1 $cm^{-2}sec^{-1}sr^{-1}$	~ GeV	$F_\gamma \gtrsim 10^{-13}$ $cm^{-2}sec^{-1}$	10^{-5} sec/eV
Sun	10^{11} $cm^{-2}sec^{-1}$	200 keV	$F_\gamma \gtrsim 10^{-4}$ $cm^{-2}sec^{-1}$	$2\ 10^{12}$ sec/eV
White Dwarfs, etc.	10^{58}/W.D.	100 keV	$F_\gamma \gtrsim 10^{-1}$ $cm^{-2}sec^{-1}$ sr^{-1}	10^{17} sec/eV
Supernovae	$4.\ 10^{58}$/SN	10 MeV	$F_\gamma \gtrsim 10^{-3}$ $cm^{-2}sec^{-1}$ sr^{-1}	3×10^{16} sec/eV
Cosmos	600 cm^{-3}	6×10^{-4} eV	$F_\gamma \gtrsim 3\times10^{8}$ $cm^{-2}sec^{-1}$ sr^{-1} at 1 eV	$\tau_o > 10^{23}$ sec
			$F_\gamma \gtrsim 10^{13}$ $cm^{-2}sec^{-1}$ sr^{-1} at 10^{-3} eV	$\tau_o > 10^{19}$ sec

V. RADIATIVE STABILITY AND MASS OF THE TAU-NEUTRINO

Experiments performed at colliding e^+e^- beams over the last few years have established the existence of a third sequential lepton[33], τ, at a mass of 1.78 GeV/c^2 along with an associated neutrino ν_τ. Though the properties of $\tau^\pm$ are being measured with progressively greater precision[33,53,54] the only results[55,56] on ν_τ are that $m_{\nu\tau} < 250$ MeV/c^2 and its coupling has the usual strength and the V-A form of weak interactions. Further, there seems to be little hope of improving the mass limit in the foreseeable future. In order to delineate its properties, we therefore take recourse to the astrophysical methods the efficacy of which is well demonstrated for the case of the other neutrinos. But, for the astrophysical arguments to be applicable, we have to first show that the ν_τ produced in the astrophysical setting live long enough to have observable

effects. To this end we utilize two accelerator experiments to show that $\tau_0/m_{\nu T}c^2$ is longer than $\sim 10^{-6}$ s/eV. We then start the astrophysical discussions with a calculation of the effect of radiative decay on the primordial He and D abundance. Our discussion here is distinct from earlier work in that we consider the effect of photodissociation of the nuclei which indeed is the stronger and more direct effect as compared with the indirect effects of an increased radiation field discussed earlier. We then consider the emission of ν_τ from supernovae and show that their radiative lifetimes exceed the age of the Universe. Thus, the large number density of ν_τ in the Universe should also conform to the upper limits on the mass expressed in Eq. 15.

V. 1. Results from Accelerators

The generation and the decay properties of the lepton have been studied in detail by Baccino et al. [57] at SPEAR using the DELCO detector, which provides nearly a 4π coverage. They have measured the spectrum of electrons arising from $\tau^- \rightarrow e^- \bar{\nu}_e \nu_\tau$ and find an excellent fit to the theoretical expectations with standard V-A coupling for both the $\tau - \nu_\tau$ and $e - \nu_e$ vertices. If the ν_τ should decay inside their apparatus of dimensions $\sim$ 20 cm giving rise to either a gamma ray or an electron, there would be a recognizable signal and it would contaminate and disturb the fine agreement between experiment and theory. Absence of any such effects indicate that the lifetime of ν_τ is longer than $\sim 10^{-9}$ s in the lab-frame; with the mean energy of ν_τ being $\sim$500 MeV this corresponds to

$$\tau_o/m_{\nu\tau}c^2 > 2.10^{-18} \text{ s/eV}. \qquad (33)$$

This result is shown in Fig. 3 where cross-hatchings indicate forbidden regions in the $m_{\nu\tau}$-$\tau_o/m_{\nu\tau}$ plane.

We now turn our attention to the experiment performed by Heisterberg et al.[58] at Fermi-lab to measure the cross-section for the process $\nu_\mu + e^- \to \nu_\mu + e^-$. The muon-neutrino beam was produced by allowing an immense fluence of 10^{19} protons at $\sim$ 350 GeV to be incident on a target, one interaction mean free path thick. The pions and kaons which were produced in the interactions traversed a tunnel $\sim 5.10^4$ cm long ($\ell/c \gtrsim 1.7 \times 10^{-6}$ s) and the neutrinos generated in the decay of the mesons passed through a detector of cross-sectional area 10^4 cm^2 and depth $\sim 1.5 \times 10^3$ cm ($\ell/c \gtrsim 5.10^{-8}$ s). The detectors were sensitive to electrons and gamma rays and the cascades produced by the elastically scattered electrons were detected. In such a neutrino beam there would be a finite fluence of ν_τ also, generated mainly through the decay of $F\bar{F}$ mesons which are produced with roughly the same cross-section as the $D\bar{D}$ mesons viz. $\sigma_{F\bar{F}} \gtrsim \sigma_{D\bar{D}} \gtrsim 100\ \mu$b. Following the detailed discussion of Albright et al. we estimate the differential spectrum of $\nu_\tau + \bar{\nu}_\tau$ in the detector to be given by

$$\begin{aligned} F(E) &\gtrsim 3.10^{11} \exp(-E/21.6\ \text{GeV})/\text{GeV} && \text{for } E > 10\ \text{GeV} \\ &\gtrsim 0 && < 10\ \text{GeV} \end{aligned} \tag{34}$$

Now should ν_τ decay through either of the channels $\nu_\tau \to \gamma + x$ or $\nu_\tau \to e + x$, it would simulate ν_μ e-scattering events and would be recorded. An upper limit to the number of such events in their apparatus is given by the maximum number of background events of $\sim$10. This leads to the inequality

$$\int_{10}^{E_{max}} \exp\left(\frac{-1.7\ 10^6\ m_{\nu\tau}c^2}{\tau_o E}\right)\left[1 - \exp\left(\frac{5.10^{-8}\ m_{\nu\tau}c^2}{t_o E}\right)\right] F(E)dE < 10. \tag{35}$$

This translates into the result $\tau_o/m_{\nu\tau}c^2 < 4.10^{-19}$ s/eV or $\tau_o/m_{\nu\tau}c^2 > 7.10^{-7}$ s/eV, again displayed in Fig. 1. But in view of Eq. 1,

$$\tau_o / m_{\nu\tau}\, c^2 > 7.\, 10^{-7}\ \mathrm{s/eV} \qquad (36)$$

V. 2. Limits from Cosmological He and D Abundances

Within the framework for the so-called 'standard model' of the expanding Universe[34] neutrinos of mass in excess of 1 MeV go out of thermodynamic equilibrium very early and, as shown by Lee, Weinberg and Dicus et al., their density is controlled by a competition between production through a variety of channels and loss through annihilation[15,16]. Then at a temperature T_d corresponding to a red-shift Z they decouple with a density n_d and their subsequent evolution is controlled merely by expansion and possible decay. An empirical fit to the detailed calculations in the mass range 10 MeV $< m_\nu c^2 <$ 1 GeV yields

$$n_d = 10^{32} \left(\frac{mc^2}{1\ \mathrm{MeV}}\right)^{-0.45} \mathrm{cm}^{-3}$$

$$T_d = 5.\, 10^9 \left(\frac{mc^2}{1\ \mathrm{MeV}}\right)^{0.8} {}^{o}\mathrm{K} \qquad (37)$$

$$(1 + Z_d) = 2.5\ 10^9 \left(\frac{mc^2}{1\ \mathrm{MeV}}\right)^{0.8} .$$

The radiative decay of the neutrinos ($\nu \to \gamma + x$) would yield gamma-rays of energy $E_o \approx (m_\nu^2 - m_x^2)/2\, m_\nu \approx m_\nu/2$ and these gamma rays will suffer reduction in density due to effects of expansion and absorption mainly through the creation of e^+e^- pairs; their density at any time $t \gg t_d$ is given by

$$n_\gamma\ (>E_\gamma, t) = \int_{t_{min}}^{t} \frac{n_d}{\tau_o} \left\{\frac{1 + Z(t')}{1 + Z(t_d)}\right\}^3 \exp\{-(t' - t_d)/\tau_o - \sigma_\gamma c n_o \int_{t'}^{t} [1 + Z(t'')]^3\, dt''\}\, dt' \ . \qquad (38)$$

Here, t_{min} is defined through the equation $E_o\ [1 + Z(t)] = E_\gamma\ [1 + Z(t_{min})]$. The total absorption cross-sections of γ-rays, $\sigma_\gamma \gtrsim 20$ mb and, the present day mean density of hydrogen in the Universe, $n_o \leq 3.\ 10^{-6}\ cm^{-3}$. In evaluating Eq. 6, it is convenient to use the empirical fits

$$[1 + Z(t)] \gtrsim 5.\ 10^{9}\ t^{-1/2}\ \text{(relativistic)};$$

$$[1 + Z(t)] \gtrsim 10^{12}\ t^{-2/3}\ \text{(non-relativistic)} \qquad (39)$$

corresponding to the early radiation dominated era and the later times dominated by non-relativistic matter respectively.

Now $\sigma_{p.d}$ the cross-section for the photodissociation of D and He is $\sim$ 2 mb above the threshold energy E_T of $\sim$4 MeV and $\sim$20 MeV respectively[59]. These cross-sections are sufficiently large to produce substantial changes in composition before the thermalization of γ-rays. The effects of the photodissociation on the D and He abundance are indeed very complicated and have many interesting ramifications. But for the purposes of limiting the radiative decay of ν_τ we shall merely demand that the change effected in the abundance subsequent to the completion of nucleosynthesis at t_{syn} not be substantial. The logarithmic change in the abundance A of either nucleus is

$$\ell nA \gtrsim -\int_{t_{syn}}^{t_u} c\sigma_{p.d}\ n_\gamma\ (>E_T, t)\ dt \qquad (40)$$

where $t_{syn} \gtrsim 300$ s and $t_u \gtrsim 10^{18}$ s the present age of the Universe. We can make the value of the integral in Eq. 8 small enough to be acceptable either by choosing $\tau_o < t_{syn}$ or by choosing it very long, $\tau_o > t_u$. In these regions, we have

$$-\ell nA > 5.\ 10^{8}\ m^{-2.85}\ e^{-t_{syn}/\tau_o} \text{ for } \tau_o \sim t_{syn}$$

$$> 2.\,10^{24}\,\tau_o^{-1}\,M^{-2.85} \quad \text{for} \quad \tau_o > t_u \,. \tag{41}$$

Assuming that the reductions in the abundances are not substantial, Eq. 9 yields the results: for $m_{\nu\tau}c^2$ in the range 10 MeV to 250 MeV

$$\tau_o < 50 \text{ s or } \tau_o > 3.\,10^{24}\,m^{-2.85} \,. \tag{42}$$

This region of restriction is again indicated in Fig. 1.

V. 3. Limits from Studies of Supernovae

In a stellar collapse leading to a supernova explosion the core reaches temperatures of the order of 10 -20 MeV and radiates away the gravitational binding energy of $\sim 10^{53}$ ergs in the form of neutrinos[60-62]. As previously discussed by us[11] and later by Falk and Schram[14] the observations of gamma-ray astronomy and the fact that they "do evolve by rapid loss of energy in some form" allows one to further restrict the radiative decay of the neutrinos. In estimating the emission of ν_τ we should keep in mind that as with ν_μ, only the processes involving neutral currents would be operative in the production process. These would be further suppressed if $m_{\nu\tau} > 10$ MeV by a factor S which is approximately given by

$$S(m_{\nu\tau}) \approx \frac{S_{th} \int_0^\infty p^2 \left[\exp \frac{\sqrt{p^2c^2 + m^2c^4}}{kT} + 1\right]^{-1} dp}{\int_0^\infty p^2 \left[\exp \frac{pc}{kT} + 1\right]^{-1} dp} \tag{43}$$

where S_{th} is a factor which takes into account departures from thermodynamic equilibrium. Taking $S_{th} \sim 1$ we estimate $S(mc^2 < 2kT) \approx 1$, $S(mc^2 = 5kT) \approx 0.125$, $S(mc^2 = 10kT) \approx 3,5.\,10^{-3}$ and $S(mc^2 = 20kT) \approx 5.\,10^{-7}$. Now should the ν_τ decay radiatively with an extremely short lifetime then its energy will be fed back into the core and would be reradiated as ν_e and ν_μ. But if the lifetime is such that

the decay takes place in the outer mantle between the radii of $\sim 10^7$ cm and $\sim 10^{12}$ cm the decay energy will heat up the mantle and will appear as kinetic energy of the drbris. Observationally the debris has kinetic energies in the range $10^{49} - 10^{50}$ ergs and this precludes $\tau_o/m_{\nu\tau}$ in the range $\sim 10^{-12}$ s/eV to $\sim 3.10^{-3}$ s/eV for $m_{\nu\tau}$ below ~ 100 MeV. These regions are shown in Fig. 3.

Further, once $\tau_o/m_{\nu\tau}\, c^2$ is longer than $\sim 10^{-6}$ s/eV, the neutrinos emitted from the supernovae will decay sufficiently far away from the core for the γ-rays to be observable. These γ-rays will have energies $\gtrsim \frac{1}{2} m_{\nu\tau} c^2$ or $\gtrsim \frac{1}{2} kT \gtrsim 5$ MeV, whichever is larger, and our earlier discussions[9] yield the result that

$$\tau_o/m_{\nu\tau}\, c^2 > .10^{17} \text{ s/eV for } m_{\nu\tau} < 20 \text{ MeV}/c^2$$

and

$$\tau_o/m_{\nu\tau}\, c^2 > \left(\frac{m_{\nu\tau}}{1 \text{ MeV}}\right)^{1.2} S(m_{\nu\tau})10^{17} \text{ sec/eV for } m_{\nu\tau} > 20 \text{ MeV}/c^2 \qquad (44)$$

since the flux of background γ-rays falls off as $E_\gamma^{-1.2}$ at high energies.

V.4. Limits from Radiation Background Below $\sim$ 20 eV

Now, notice that our arguments so far have precluded all values of $\tau_o/m_{\nu\tau}$ shorter than $\sim 10^{17}$ s/eV. This means that if $m_{\nu\tau} > 10$ eV, it will survive longer than 10^{18} s, the present age of the Universe. Further, since the Universe is transparent to photons up to a red-shift of ~ 100 ie. $t \gtrsim 10^{15}$ s, the decay of neutrinos of $m > 10^{-2}$ eV will generate a background of radiation in the Universe. By attributing the observed background radiation to the decay of the cosmological neutrinos, we can set a lower limit to the lifetime of the neutrinos. Noticing that neutrinos of such masses would have slowed down to non-relativistic velocities during the expansion of the Universe and following the analysis of Cowsik[11], DeRujula and Glashow[26], Stecker[27] and

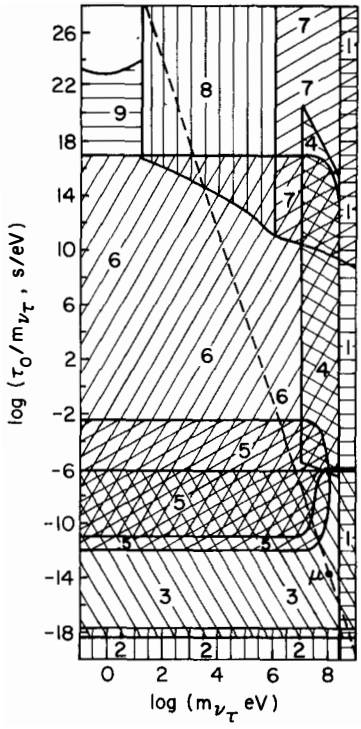


Fig. 3. Forbidden regions in the mass-lifetime plane of the τ-neutrino. (1) Decay kinematics of τ^-, (2) SPEAR experiment[57], (3) Fermilab experiment[58], (4) Photodissociation of cosmological D and He, (5) Supernova dynamics, (6) γ-rays from supernovae, (7-8) Effects on Universal expansion[1,13,15,16], and (9) Optical and UV background radiation.

Kimble et al. [28], we show in Fig. 1 the constraint[9] that $\tau_o/m_{\nu\tau} > 10^{23}$ s/eV at $m_{\nu\tau} \gtrsim 1$ eV and the limit increases as $\sim \frac{1}{m} \exp(mc^2/4\ \text{eV})$ as the radiation background decreases exponentially. The limits are shown only up to a mass of $\sim$ 35 eV/c^2 since larger masses are excluded unconditionally in any case.

V. 5. The Limits on the Mass of the Tau Neutrino

The discussions in the previous sections have established that the lifetime of ν_τ is so long that the large numbers produced in the big bang survive until today with their densities merely reduced by expansion of the Universe. As discussed extensively before, if these neutrinos are massive, they would

dominate the gravitational interactions in the Universe. Following Lee and Weinberg[15] and Dicus et al.[16] the mass region 1 MeV to 2 GeV is excluded and the region 35 eV - 1 MeV is excluded by the neutrino density estimates of Cowsik and McClelland[1] and Szalay and Marx[12]. Thus we conclude

$$m_{\nu\tau} c^2 < 35 \text{ eV} \tag{45}$$

For reasons of comparison, we have drawn a line which passes through the muon mass and lifetime and scaling as m^{-6} as expected in most models.

Thus, we have shown that the lifetime of ν_τ is longer than the age of the Universe, which enables us to use the dynamical effects of cosmologically generated neutrinos to show that $m_{\nu\tau} < 35 \text{ eV}/c^2$.

Acknowledgement

I am thankful to Dr. M. A. Pomerantz of the Bartol Research Foundation of The Franklin Institute, Newark, Delaware for the hospitality extended to me when some of this work was in progress. I am particularly indebted to Professor G. Yodh for critical discussions of every aspect of this work and to Drs. J. Kirkby and A. Skuja for discussions of their experiments at the accelerators. I am indebted to Dr. T. K. Gaisser and Dr. G. Steigman for a careful reading of the manuscript and suggesting many improvements. This work was supported in part by U.S. Department of Energy Contract DE-AS02-78ER05007.

REFERENCES

1. Cowsik, R., and McClelland, J., Phys. Rev. Lett. 29, 669 (1972).

2. Markov, M.A., Phys. Rev. Lett. 10, 122 (1964).

3. Bludman, S.S., Proc. Neutrino '74, p.284 (AIP, New York, 1974).

4. Gershtein, S.S., and Zeldovich, Ya.B., JETP Lett. 4, 407 (1966).

5. Bahcall, J.N., Cabbibo, N., and Yahil, A., Phys. Rev. Lett. 27, 757 (1971).

6. Cowsik, R., Yash Pal, Tandon, S.N., and Rengarajan, T.N., Proc. 8th Internat. Conference on Cosmic Rays 6, 211 (Jaipur, 1963).

7. Cowsik, R., Yash Pal, and Tandon, S.N., Proc. Indian Acad. Sci. 53, 217 (1966).

8. Cowsik, R., and Yash Pal, Proc. 11th Internat. Conference on Cosmic Rays (Budapest, 1969).

9. Cowsik, R., Yash Pal, and Tandon, S.N., Phys. Lett. 13, 265 (1964).

10. Cowsik, R., and McClelland, J., Astrophys. J. 180, 7 (1973).

11. Cowsik, R., Phys. Rev. Lett. 39, 168 (1977); Phys. Rev. 19D, 2219 (1979).

12. Marx, G., and Szalay, A.S. Proc. Neutrino '72, (Technoinform, Budapest, 1972).

13. Marx, G., and Szalay, A.S., Astron. and Astrophys. 49, 437 (1976).

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

15. Lee, B., and Weinberg, S., Phys. Rev. Lett. 39, 165 (1977).

16. Dicus, D.A., Kolb, E.W., and Teplitz, V.L., ibid, p. 168 (1977).

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

18. Tremaine, S., and Gunn, J.E., Phys. Rev. Lett. 42, 407 (1979).

19. Steigman, G., and Schramm, D.N., to be published in Astrophys. J.

20. Gellman, M., Ramond, P., and Slansky, R., in Supergravity (ed. P. Van Nieuvenhuizen and D.Z. Freedman, N. Holland, 1979 and references therein).

21. Mohapatra, R. N., and Senjanovic, G., Fermilab preprint pub. - 80/61 - THY (1980).

22. Pakvasa, S., XXth Internat. Conf. on High Energy Physics (Univ. of Wisconsin - Madison, 1980).

23. Witten, E., First Workshop on Grand Unification (Univ. of New Hampshire, 1980).

24. Lyubimov, V. A., Novikov, E. G., Nozik, V. Z., Trekyakov, E. F., and Kozik, V. S., ITEF - preprint No. 62 (Moscow, 1980).

25. Reines, F., Sobel, F., and Pasierb, E., preprint, 1980.

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

27. Stecker, F. W., private communication (1980).

28. Kimble, R., Bowyer, S., and Jacobson, P., private communication (1980).

29. Cowsik, R., in preparation.

30. Shipman, H., and Cowsik, R., observations on the Coma cluster yield $\tau_\nu > 10^{24}$ s at $m_\nu \gtrsim 5$ eV.

31. Bergkvist, U., Nucl. Phys. B39, 317 (1972).

32. Shrumm, E. V., and Ziok, O. H., Phys. Lett. 37B, 115 (1971).

33. Perl, M. L., in Proc. Internat. Symp. on Lepton and Photon Interactions at High Energies (ed. F. Gutbrod, DESX, Hamburg, 1977).

34. Weinberg, S., Gravitation and Cosmology (Wiley, New York, 1972).

35. Ruderman, M. A., in Topical Conf. on Weak Interactions, (CERN, Geneva, 1969).

36. Peebles, P. J. E., Physical Cosmology (Princeton, 1971).

37. Sandage, A., Astrophys. J. 173, 485 (1972).

38. Joshi, P., Gravity Foundation Essay, 1980.

39. Joshi, P., and Chitre, S. M., TIFR-preprint 1981.

40. Landau, L. D., and Lifshitz, E. M., Statistical Physics (2nd Edition, Addison-Wesley, 1969).

41. Gribov, V., and Pontecarvo, B., Phys. Rev. Lett. 28B, 493 (1969).

42. Eliezer, S., and Ross, D. A., Phys. Rev. D10, 3088 (1974).

43. Barnes, V.B., Phys. Lett. 65B, 194 (1976).

44. Barshay, Phys. Lett. 66B, 246 (1977).

45. Commins, E., Weak Interactions (McGraw-Hill, 1973).

46. Hayakawa, S., Cosmic Ray Physics (Interscience, 1969).

47. Krishnaswamy, M.R., Menon, M.G.K., Narasimham, V.S., Ito, N., Kawakami, S., and Miyake, S., Pramana 5, 59 (1975).

48. Southerland, P., Ng, J.N., Flowers, E., Ruderman, M., and Inman, C., Phys. Rev. D13, 2700 (1976).

49. Stothers, R.B., Phys. Rev. Lett. 24, 538 (1970).

50. Boldt, E., private communication.

51. Woody, D.P., Matheu, J.C., Nishioka, N.S., and Richards, P.L., Phys. Rev. Lett. 34, 1036 (1975).

52. Rood, H.J., Page, T.L., Kintner, E.G., and King, I.R., Astrophys. J. 175, 627 (1972).

53. Perl, M.O., et al., Phys. Rev. Lett. 35, 1489 (1975).

54. See papers in Proc. of Internat. Symp. on Lepton and Photon Interactions at High Energies, (ed. F. Gutbrod, Hamburg, 1977).

55. Yamada, S., in Proc. Internat. Symp. on Lepton and Photon Interactions at High Energies (ed. F. Gutbrod, Hamburg, 1977).

56. Kirkby, J., in Neutrinos '78 (ed. E.C. Fowler, Purdue, 1978).

57. Baccino, W., et al., Phys. Rev. Lett. 42, 749 (1979), and references therein.

58. Heisterberg, R.H., et al., Phys. Rev. Lett. 44, 635 (1980).

59. Fuller, E., private communication (National Bureau of Standards, compilation).

60. Arnett, W.D., Ann. N.Y. Acad. Sci. 262, 47 (1975).

61. Wilson, J.R., Cronch, R., Cochran, S., LeBlanc, J.L., and Barkat, Z., ibid, p. 54.

62. Schramm, D.N., ibid, p. 65.

CONSTRAINTS ON NEUTRINOS AND AXIONS FROM COSMOLOGY

David N. Schramm
University of Chicago

ABSTRACT

A review is made of the astrophysical arguments with regard to neutrino properties. It is shown that the best fit to the present baryon density and ^{4}He abundance is obtained with three neutrino species. It is also shown that astrophysical constraints on neutrino and axion lifetime-mass combinations rule out weakly interacting particles with lifetimes between 10^{-3} and 10^{23} sec for $M \leq 10$ MeV. For lifetimes $\geq 10^{23}$ sec the sum overall species of neutrino mass must be below 100 eV or between $\sim$ 10 GeV. There is an allowed astrophysical window for neutrinos with $M \geq 10$ MeV and $\tau \leq 1000$ sec. The possible role of massive neutrinos in the dark mass of galaxies is discussed. It is shown that the baryon density in the universe is comparable to the density obtained from the dynamics of binary galaxies. Therefore, massive neutrinos are only required if the cosmological mass density is greater than that implied by binaries and small groups of galaxies. The only objects which might imply such high densities are large clusters. For neutrinos to cluster with these large clusters requires $m_\nu \geq 3$ eV.

INTRODUCTION

This lecture reviews various constraints that can be placed on neutral weakly interacting particles from cosmology. Most of these astrophysical constraints

come from the standard Big Bang model of the universe.

The general acceptance of a hot-dense early universe (the Big Bang) began to happen with the discovery of the 3 K background radiation by Penzias and Wilson[1]. However complete acceptance did not occur until the experiments of Richards and co-workers[2] showed that the background radiation had the appropriate thermal turnover at wavelengths of $\sim$ 1 mm. The existence of this radiation tells us that the universe was at one time at least $\gtrsim 10^4$ K. At $\sim 10^4$ K hydrogen would be ionized and the free electrons would easily scatter the photons. Thus, the present observed radiation is merely the last scattered thermal radiation from $\sim 10^4$ K.

We actually have confidence that the universe was a good deal hotter than this. Gamow and his co-workers[3] predicted that there would be this thermal background on the basis of assuming that nuclear reactions occurred in the Big Bang. To have nuclear reactions requires that the temperature had to be greater than $\sim 10^9$ K. The verification of a temperature at least as hot as 10^{10} K comes from the fact that the ^{4}He abundance is about 25% by mass. This helium abundance comes as a natural consequence of the standard Big Bang if the temperature was greater than 10^{10} K (see Schramm and Wagoner[4] and references therein). This temperature is when neutrinos decouple from matter. The good agreement between the observed ^{4}He abundance and the general Big Bang predictions of 20 to 25% by mass is direct observational evidence that we have a good understanding of the universe back to the epoch when the neutrinos decoupled. We will see that our constraints on neutrino properties are thus coming from the era when the Big Bang has produced observables and not from the more mysterious early times which grand unification tests require.

THE DECOUPLING OF THE WEAK INTERACTION AND RELIC NEUTRINOS

During the early evolution of the universe, all particles, including neutrinos, were produced copiously. Neutrinos with full strength, neutral current, weak interactions were produced by reaction of the type,

$$e^+ + e^- \leftrightarrow \nu_i + \bar{\nu}_i;\ i = e,\ \mu,\ \tau. \tag{1}$$

At high temperatures ($kT > m_\nu c^2$), these neutrinos were approximately as abundant as photons.

$$n_{\nu_i}/n_\gamma = 3/8(g_{\nu_i}) \tag{2}$$

g_ν is the number of neutrino helicity states. For massless spin ½ particle would have 2 helicity states thus for $\bar{\nu} \neq \nu$, $g_\nu = 4$ however because known neutrinos appear to be only left handed then, if massive, they may be of the Majorana type ($\nu_i \approx \bar{\nu}_i$), in which case g_ν is still 2. A numerical factor of 3/4 comes from the difference between Fermi-Dirac statistics (neutrinos) and Bose-Einstein statistics (photons); the remaining factor of ½ is from the number of photon spin states ($g_\gamma = 2$).

For light neutrinos ($m_\nu \ll 1$ MeV), equilibrium was maintained until $T \approx 1$ MeV. (For massive neutrinos with $m_\nu \gtrsim 1$ MeV, they will annihilate for temperatures less than m_ν but $\gtrsim 1$ MeV, thus ν's with $m_\nu \gtrsim 1$ MeV will not be as abundant, see Gunn et al.[5] and references therein.) At lower temperatures the weak interaction rate is too slow to keep pace with the universal expansion rate so that few new neutrinos are produced and, equally important, few annihilate. Thus, for $T \approx 1$ MeV, the neutrinos decouple; at this stage their relative abundance is given above. When the temperature drops below the electron mass, electron-positron pairs annihilate heating the photons but not the decoupled neutrinos. The present ratio of neutrinos to photons must account for the extra photons produced when the $e^\pm$ pairs disappeared (c.f. Steigman[6]), because of this, the present neutrino temperature is $\sim 2°$ K rather than the photon temperature of 3° K. From the present density of photons and the above, we obtain the present number density of neutrinos. If the neutrinos have mass we can obtain the mass density, ρ_ν, by

$$\rho_\nu = \sum_i (g_{\nu_i} m_{\nu_i}/200\ h_0^2)\ (T_0/2.7)^3. \tag{3}$$

In this equation and subsequently, m_ν is in eV and the sum is over all neutrino species with $m \ll 1$ MeV, h_0 is the Hubble constant in units of 100 km/sec/Mpc.

It is implicitly assumed in the above that the neutrinos still exist today and thus have a lifetime greater than the age of the universe for decay into anything other than neutrinos.

It is interesting to compare this density to the critical density of the universe, $\rho_c = 3H_o^2/8\pi G$. The ratio Ω is defined as $\Omega = \rho/\rho_c$. For $T_o \lesssim 3K$ and $h_o < 1$ and assuming Majorana mass neutrinos with $g_\nu = 2$ we find that

$$\Omega_\nu \gtrsim 0.014\ m_{\nu_i} \tag{4}$$

We will show later that from Big Bang nucleosynthesis, the upper limit on baryon density parameter Ω_b is 0.14. Thus, there is the relationship with $g_{\nu_i} = 2$

$$\Omega_\nu/\Omega_b \gtrsim \sum_i m_{\nu_i}/1.4 \tag{5}$$

which is independent of h_o and T_o. Therefore, <u>if neutrinos have masses of the order of eV or greater then neutrinos are the dominant mass component of the universe today</u>.

CONSTRAINTS ON NEUTRINO MASSES AND LIFETIMES

From the above arguments on numbers of neutrinos produced during the Big Bang one can put constraints on allowed combinations of neutrino mass and lifetime to non-neutrino decay modes due to the limits on intensity of photon backgrounds in various energy regimes and due to limits on the mass density of the universe (c.f. Gunn et al.[5] and references therein).

Figure 1 shows a summary of astrophysical constraints on neutrino lifetimes and masses. For very long lifetimes and limits come from the limits on the cosmological mass density (Cowsik and McClelland[7] and Szalay and Marx[8]). We know the universe is not <u>too</u> closed. In fact, from the nucleochronologic[9] constraint that the age of the universe must be $\gtrsim 8.5 \times 10^9$ yr we know that $\Omega h_o^2 \lesssim 1$. Since

$$\Omega_{\nu_i} h_o^2 = \frac{g_{\nu_i} m_{\nu_i}}{200\ (\text{eV})} \left(\frac{T_o}{2.7}\right)^{2.7} \tag{6}$$

we obtain the limit that the sum of m_{ν_i}'s must be less than $\sim$ 100 eV for Majorana mass neutrinos. Gunn et al.[5] showed that since very massive neutrinos must

cluster then the density limit to use for them is matter in galaxies from which they obtain the limit that a long lived neutrino with $m_\nu \gtrsim 10$ GeV could again be allowed. Hut and Olive[10] showed that mass density again becomes excessive if $n_{lepton} \sim n_b \ll n_\gamma$ and $m_\nu \gtrsim 60$ GeV.

For ν's with finite lifetimes greater than the radiation decoupling time limits can be obtained from photon backgrounds. The most interesting constraints here come from the uv. Recent discussion of this point comes from DeRujula and Glashow[11] and Stecker[12] with the most comprehensive discussion being that of Kimbal, Bowyer and Jacobson[13] where it is shown that low mass neutrinos must have lifetimes longer than $\sim 10^{23}$ sec. For lifetimes in the regime between Big Bang nucleosynthesis ($t \sim 10^3$ sec) and decoupling ($t \sim 10^{12}$ sec) constraints come from the distortion of the 3° background (Gunn et al.[5]) and the constraints from Deuterium synthesis on n_b/n_γ (Dicus et al.[14]).

For lifetimes less than $\sim 10^3$ sec and $\gtrsim 10^{-3}$ sec Falk and Schramm[15] have ruled out all neutrinos or axions with masses less than ~ 10 MeV from supernova dynamics arguments. This argument also applies to axions and thus would constrain the particle Faissner[16] reported on at this meeting. The argument goes as follows: Supernova release from their collapsing cores the $\sim 10^{53}$ ergs of neutron star binding energy. This energy escapes via weakly interacting particles. The total energy observed in supernova outbursts is less than 10^{51} ergs, thus all but $\lesssim 1\%$ of these weakly interacting particles must escape the entire star. Because neutrinos with energy $\gtrsim 10$ MeV are trapped in the core and do not escape this argument does not apply to masses greater than 10 MeV. Similarly, because the size of the neutrino emitting core region is $\lesssim 10^{-3}$ sec. In order for the neutrinos to escape the star they would need a lifetime $\gtrsim 10^3$ sec. It is interesting to note that a weakly interacting particle which decays with $\sim 1\%$ efficiency to photons just at the edge of the core could be the driver of the supernova explosion itself. Thus, if Faissner's[16] axion had a lifetime slightly less than 10^{-3} sec most of them would not escape the core but 1% might and then

decay where they could deposit their rest mass energy.

Lifetimes less than this are probably eliminated by reactor and accelerator data. In summary, it appears that with the experimental limit that $m_{\nu_\mu} < 700$ keV and $m_{\nu_e} < 50$ eV these both must be very long lived ($> 10^{23}$ sec) and have the sum of their masses $m_{\nu_e} + m_{\nu_\mu} < 100$ eV. For m_{ν_τ} there is also this low mass possibility, however since the experimental upper limit on m_{ν_τ} is only ~ 250 MeV, it cannot be ruled out from astrophysics alone that m_{ν_τ} is between 10 and 250 MeV with a lifetime the order of seconds.

In the above arguments, it has been implicitly assumed that the neutrinos under consideration interact with the strength of the standard Weinberg-Salam weak interaction mediated by intermediate vector bosons with masses ~ 100 GeV. The same type of arguments can also be carried out for neutrinos which couple more weakly through some higher mass gauge boson. The major difference is that the weaker the interaction, the earlier the decoupling of the neutrino from the matter (e.g., Olive et al.[17] and references therein). Earlier decoupling can lead to a lower present temperature for the neutrinos relative to the 3° photon background and the 2° Weinberg-Salam neutrino background due to the annihilation of other particles (μ's, π's, etc.) heating up the ω's, ν's and the γ's after the superweak ν's decouple. A lower temperature corresponds to a lower number density and thus weakens the constraints. Since the first particles above the ω's decoupling of ~ 1 MeV have masses ~ 100 MeV. This means that the limits mentioned above will not be changed unless the superweak neutrinos decouple earlier than $\sim 10^{12}$° K which would imply gauge boson masses $\gtrsim 1$ TeV. Olive and Turner[18] have carried out a detailed description of these effects.

NEUTRINO MASSES

Since it has been shown that finite but small neutrino masses cannot be ruled out even for ν_μ and ν_τ and there may even be some theoretical justification for such masses from grand unification, let us look at possible cosmological consequences. In fact, as we will see, these cosmological arguments were used by

Schramm and Steigman[19,20] to argue in favor of finite mass neutrinos.

A long standing problem in Astronomy has been the so-called "Missing Mass" problem (or as Schramm and Steigman[19,20] say the "Missing-light" problem). This is the fact that when one measures the mass of a galaxy size object by looking at the dynamics of the object, one tends to assign a progressively larger mass to the galaxy when one looks at the dynamics on a larger scale. However, the amount of light the galaxy emits stays fixed. Table 1, based on the reviews of Faber and Gallagher[21] and Peebles[22], shows this trend. Note that while the mass associated with the actual light emitting objects (stars) is only 1 to 2 solar units, the mass associated with the central visible regions of galaxies is $\sim$ 10 times this. When that same galaxy is observed in a binary pair or in a small group and the total mass of the system is calculated and distributed among the members, it is found that the mass associated with each galaxy is 4 to 10 times that estimated from the galaxy's own internal motion. Similarly when a galaxy is in a large cluster and the virial theorem is applied to the cluster to estimate the mass of the galaxy it is found that the mass-to-light ratio is now about 400 (with a range from 100 to about 800). Thus there seems to be more and more mass on larger scales with no more light emission. This additional invisible mass is what is called the missing mass problem. However, the mass seems to really be there, so what is missing is really the light from that mass, hence Schramm and Steigman's[19,20] term "missing light."

Over the years, many things have been proposed for this missing mass (see Table 2). However as Schramm and Steigman[20] showed many of these can be eliminated <u>if</u> the missing mass truly must give an $\Omega > 0.1$ as implied if the M/L for large clusters truly give the best estimate of the mass associated with the cosmological luminoisty density.

As pointed out by Gott et al.[23], Schramm and Wagoner[4] and Yang et al.[24] and references therein, Big Bang nucleosynthesis cannot give consistent abundances for ^{4}He nor D if the density of baryons is greater than about 10% of the critical

value. Thus if Ω for the universe is $\gtrsim 0.1$ as implied by large clusters then the missing mass (light) for these clusters must not be in the form of baryons. This eliminates all of the entries in Table 2, except long lived low and high mass neutral weakly interacting particles (neutrinos, majorons,...), monopoles and black holes with masses less than $\sim 1\ M_\odot$. (Solar mass and larger black holes were in the form of baryons at the time of Big Bang nucleosynthesis.) Monopoles run into a variety of other problems such as their destruction of magnetic fields eliminating radio galaxies. Small black holes ($M \lesssim 10^{15}$ g) have the problem of blowing up via the Hawking process and thus producing too many unobserved γ-rays. Thus, all that remains are the weakly interacting neutrals and black holes with masses between $\sim 10^{15}$ g and $\sim 10^{33}$ g. Although some of the latter may be generated at the quark-hadron phase transition[25] at present we would regard such objects as very speculative. Thus, we are down to weakly interacting neutrals with non-zero rest mass. While very massive, as-yet-undiscovered, such particles may do the trick (GeV mass neutral leptons or Axion-Majorana type particles), without any experimental verification they appear extremely ad hoc. This, as pointed out by Schramm and Steigman[19,20], only leaves the low mass neutrinos first proposed by Cowsik and McClelland[7] and Marx and Szalay[8].

Massive neutrinos gravitate and they will have participated in gravitational clustering (see Gunn et al.[5]). However, since neutrinos are non-interacting, their phase space density is conserved and they will cluster only in the deepest potential wells; the slowest moving (i.e.: the heaviest) will cluster most easily (Tremaine and Gunn[26]).

Tremaine and Gunn[26] point out that neutrinos lighter than ~ 3 eV will not cluster at all. Whereas neutrinos with $m_\nu > 3$ will be trapped in large clusters, those with $m_\nu \gtrsim 10$ eV can be trapped in binaries and small groups and those with $m_\nu \gtrsim 20$ eV can be trapped with single galaxies. We have already seen that m_ν must be less than 100 eV. In fact, if neutrinos cluster efficiently on the scales of single galaxies then m_ν must be $\lesssim 20$ eV or neutrinos will contribute

too much mass on those scales. However, it may be that single galaxies, or even binaries and small groups are just inefficient at trapping neutrinos. If isothermal perturbations produce neutrino trapping on all scales then 3 eV $\lesssim m_\nu \lesssim$ 20 eV, but if adiabatic perturbations, as implied by simple GUTs[27], dominate; then $3 \lesssim m_\nu \lesssim$ 100 eV. For discussion of galaxy formation in neutrino dominated situations see Bond et al.[28], Szalay[29] and Sato[30]. The Jeans mass for neutrinos is comparable to the mass of clusters.[28] Therefore even adiabatic perturbations enable the massive ν's to cluster on that scale as long as $m_\nu \gtrsim 3$ eV.

Since neutrinos would cluster like an isothermal gas they would have a density distribution that falls off with $1/r^2$. When compared with the light from galaxies which falls off approximately like $1/r^3$ (Kron[31]) this yields M/L roughly proportional to r as implied by Table 1. Thus, neutrinos may be ideal missing mass candidates. Since they do not radiate they will not settle into the central disk regions of galaxies or into stars.

A possible way of verifying that the missing mass for galaxies is in neutrinos is to look at distant quasars through galactic halos and see if there are any gravitational lens effects due to low mass stars (Gott[32]). The possibility that one might also directly observe $\nu\bar{\nu}$ annihilation into photons or even $\nu_i\bar{\nu}_i \rightarrow \nu_j\bar{\nu}_j + \gamma$ in the density enhancements in clusters seems unlikely (Hill[33]).

Obviously, if Ω_ν were > 1 the Friedman universe would be closed by neutrinos. Current estimates put $\Omega_{cluster} < 1$ and thus Ω_ν would probably also be constrained to be < 1 however the uncertainties are sufficiently large that closure by neutrinos cannot be completely excluded. Note that with $g_\nu = 2$, if

$$\Sigma m_{\nu_i} > 100\ h_o^2\ (2.7/T_o)^3 \text{ eV} \tag{7}$$

then the Friedman universe with $\Lambda = 0$ is closed. With $h_o \geq 0.5$ and $T_o \leq 3$ we see that for a $\sum_i m_{\nu_i}$ as small as 18 eV closure is in principle conceivable. As we showed before, the limits on Ω imply $m_{\nu_i} \lesssim 20$ eV if isothermal perturbations create galaxies. The total number of such species is constrained by arguments of the types to be discussed in the next section.

It will also be shown in the next sections that not only does Big Bang nucleosynthesis provide an upper limit to the density of baryons but it also may provide a lower limit. In fact, the range on allowed baryon density will be very close to the range implied for the density of matter from the dynamics of binaries and small groups. Therefore, we do not need massive neutrinos clustering in a significant way on these scales.

BIG BANG NUCLEOSYNTHESIS

Big Bang nucleosynthesis has been described in detail in many places, (c.f. Schramm and Wagoner[4] and references therein). Thus, it will not be reviewed here other than to show how the possibility of neutrino mass effects the limits on the number of neutrino species. The basic argument was described by Gunn et al.[5] and the current details can be found in Olive et al.[17]. The point is that the greater the number of low mass ($m_\nu < 1$ MeV) neutrinos the more ^{4}He that is produced in the Big Bang.

In order to make use of these results, the mass fraction Y must be known. This is a bit of a problem since ^{4}He is made in stars as well as in the Big Bang. Stars forming now may be contaminated by as must as $\Delta Y \simeq 0.06$ "new helium." The best estimates for Y give a value of between 0.20 and 0.25 with 0.25 as an upper limit (see Yang et al.[24]). The ^{4}He abundance also has a dependence on the baryon density of the universe. If binaries and small groups of galaxies are made of baryons then a lower limit on the present density is about 2×10^{-31} g cm^{-3} (≈ 0.04 of closure density for $H_0 = 50$ kms^{-1} Mpc^{-1}). However, as noted in the previous section, it is conceivable that this mass is not from baryons but from leptons. For the present let us assume that this lower limit on $\Omega_{B,SG}$ is from baryons. We will later examine what happens if $\Omega_{B,SG}$ is in the form of leptons. A lower limit on the baryon density of this value and $Y \lesssim 0.25$ constrains the number of additional neutrino types (beyond ν_e, ν_μ and ν_τ) to $\lesssim 1$ (see figure 2). [These ^{4}He constraints can also be generalized to massless particles that couple more weakly than the usual neutrinos (c.f. Olive et al.[17])].

If one is forced to say that the ρ implied from binaries and small groups is not in baryons then we must ask what is the lower limit to ρ_b. A clear lower limit is from stellar matter but that only gives $\rho_b \geq 10^{-32}$ g/cm^3 ($\Omega \gtrsim 10^{-3}$), which gives no limit to the number of neutrinos (see Olive et al.[17]). If we note that the centers of galaxies are probably baryons then we can use internal galactic dynamics to argue that $\rho_b \gtrsim 5 \times 10^{-32}$ g/cm^3. This limit, coupled with $y \leq 0.25$ allows a total of 9 2-component neutrinos (e, μ, τ and 6 more). If the neutrinos have Dirac masses and thus 4 components, the limits becomes 4 and e, μ and τ and one more are again all that are allowed. However, this limit of 5×10^{-32} from internal galactic dynamics has some observational uncertainties. If it goes much lower then no limits are obtainable for $y < 0.25$. Hot x-ray emitting gas from clusters argues that Ω_b is probably $\gtrsim 0.02$ but there are possible loop holes in this.

It is interesting that we can get a lower limit on Ω_b from Big Bang nucleosynthesis which is almost identical with that obtainable from binaries and small groups.[24] Although normally one only concentrates on the upper limits to baryon density from big bang nucleosynthesis, there are some lower limits which are of interest. In particular, ^{3}He and D production increases dramatically with lower density. Since ^{3}He is produced not destroyed in normal stellar populations one can argue that observations (c.f. Rood et al.[34]) on ^{3}He place a lower limit to the baryon density. This argument is substantially strengthened when it is remembered that the probable way D is reduced from its big bnag calue to its present observed abundance is via $D + \rho \rightarrow {}^3He + \gamma$. Thus excess D adds to primordial ^{3}He to give the non stellar produced ^{3}He. From present limits this implies $n_b/n_\gamma \gtrsim 2 \times 10^{-10}$ or $\Omega_b > 0.02$ which yields the total number of neutrino types to be $\lesssim 5$ which is not too different from the limit obtained using $\Omega_{B,SG}$ of $\lesssim 4$. The one loop hole is the possibility that some pregalactic stellar generation burns up ^{3}He without making other excesses. (These arguments are discussed in detail by Yang et al.[24]). However, it does seem from this ^{3}He argument that binaries and small groups are probably baryon dominated so the best fit to

the number of neutrino types is probably 3 with a best fit ^{4}He abundance of ~ 0.24. It is interesting that 3 generations is also the number which best enables Ellis[35] and Nanopoulos[36] to fit the bottom quark mass.

Before leaving the subject, it is important to note that the ideas are testable by experiment. In particular, if the width of the neutral intermediate vector boson, Z°, is measured in colliding beam machines, then it will tell us the number of neutrino flavors. It is fascinating that one of the most important tests of our cosmological ideas will come from accelerators rather than telescopes.

CONCLUSION

To date, the interdisciplinary effort involving cosmologists, nuclear physicists and particle physicists has produced some exciting results. The ^{4}He abundance fixes an upper limit on neutrino and quark flavors. The grand unification ideas may resolve the puzzle of one baryon for every 10^{10} photons. If neutrinos have mass then the bulk of the mass of the universe may be in the form of leptons. In fact one might say that we have taken the Copernican principle to the extreme. Copernicus showed that the earth is not the center of the solar system, Shapley showed that the sun is the center of the galaxy and Hubble showed that our galaxy is not the center of the universe -- there is no center. Now, if neutrinos have mass, then our kind of matter may not even be the dominant matter of the universe. As our knowledge of the fundamental particles and their interactions increases, and as our determination of cosmological observables improves (or new observables are discovered) the close relationship of these two disciplines promises to continue to be an exciting one.

ACKNOWLEDGEMENTS

I would like to thank my collaborators Michael Turner, Gary Steigman, James Fry, Keith Olive, Matt Crawford, Eugene Symbalisty, Andy Jankevics, Jongmann Yang and Bob Wagoner for enabling me to draw on much jointly generated material. This work was supported by NSF grant AST 78-20402, by NASA grant NSG 7212 and by

DOE grant DE AC02-80ER10773, all at the University of Chicago.

REFERENCES

1. PENZIAS, A.A. & R.W. WILSON. 1965. Ap. J. 142: 419-421.

2. WOODY, D.P., J.C. MATHER, N.S. NISHIOKA & P.L. RICHARDS. 1975. Phys. Rev. Lett. 34: 1036-1039.

3. GAMOW, et al.

4. SCHRAMM, D.N. & R.V. WAGONER. 1977. Ann. Rev. Nucl. Sci. 27: 37-74.

5. GUNN, J.E., B.W. LEE, I. LERCHE, D.N. SCHRAMM & G. STEIGMAN. 1978. Ap. J. 223: 1015-1031.

6. STEIGMAN, G. 1979. Ann. Rev. Nucl. Part. Sci. 29: 313-337.

7. COWSIK, R. & J. McCLELLAND. 1972. Phys. Rev. Lett. 29: 669-670.

8. SZALAY, A.S. & G. MARX. 1974. Acta. Phys. Acad. Sci. Hungaricae. 35: 113-129.

9. SCHRAMM, D.N. & E.M.D. SYMBALISTY. 1981. Rep. Prog. Phys., in press.

10. HUT, P. & K.A. OLIVE. 1979. Phys. Lett. 87B: 144-146.

11. DeRUJULA, A. & S.L. GLASHOW. 1980. Phys. Rev. Lett. 45: 942-944.

12. STECKER, F.W. 1980. Phys. Rev. Lett. 45: 1460-1462.

13. KIMBALL, M., S. BOWYER & B. JACOBSON. 1981. Phys. Rev. Lett. 46: 80-83.

14. DICUS, D.A., E.W. KOLB, V.L. TEPLITZ & R.V. WAGONER. 1978. Phys. Rev. D17: 1529-1538.

15. FALK, S.W. & D.N. SCHRAMM. 1978. Phys. Lett. 79B: 511-513.

16. FAISSNER, H. 1981. Proceedings of the Moriond Astrophysics Meeting 1981, edited by Tran Thanh Van, R.M.I.E.M. Orsay, France.

17. OLIVE, K.A., D.N. SCHRAMM, G. STEIGMAN, M.S. TURNER & J. YANG. 1981. Ap. J., in press.

18. OLIVE, K.A. & M.S. TURNER. 1981, in preparation.

19. SCHRAMM, D.N. & G. STEIGMAN. 1981. Ap. J. 243: 1-7.

20. SCHRAMM, D.N. & G. STEIGMAN. 1981. Gen. Rel. & Grav., in press.

21. FABER, S.M. & J.S. GALLAGHER. 1979. Ann. Rev. Astron. Astrophys. 17: 135-187.

22. PEEBLES, P.J. 1979. A.J. 84: 730.

23. GOTT, J.R., J.E. GUNN, D.N. SCHRAMM & B.M. TINSLEY. 1974. Ap. J. 194: 543-553.

24. YANG, J., D.N. SCHRAMM, G. STEIGMAN, & R.T. ROOD. 1979. Ap. J. 227: 697-704.

25. CRAWFORD, M. & D.N. SCHRAMM. Submitted to Nature, 1981.

26. TREMAINE, S. & J.E. GUNN. 1979. Phys. Rev. Lett. 42: 407-410.

27. TURNER, M.S. & D.N. SCHRAMM. 1979. Nature 279: 303-305.

28. BOND, R., R. EFESCO & J. SILK. 1980. Phys. Rev. Lett. 45: 1980-1984.

29. SZALAY, A. 1981. In Proceedings of the 10th Texas Symposium, The New York Academy of Sciences.

30. SATO, H. 1981. In Proceedings of the 10th Texas Symposium, The New York Academy of Sciences.

31. KRON, R. Private communication.

32. GOTT, J.R. 1981. Princeton preprint.

33. HILL, C. Private communication.

34. ROOD, R.T., T.L. WILSON & G. STEIGMAN. 1979. Ap. J. 227: L97-101.

35. ELLIS, J. 1981. Proceedings of the Moriond Astrophysics Meeting 1981, edited by Tran Thanh Van, R.M.I.E.M. Orsay, France.

36. NANOPOULOS, D. 1981. Proceedings of the Moriond Astrophysics Meeting, 1981. edited by Tran Thanh Van, R.M.I.E.M. Orsay, France.

TABLE 1

Mass-to-light Ratios on Various Scales and the IMPLIED Density Parameter

Scale	M/L	Ω^*
Stars	1-2	$\sim 10^{-3}/h_o$
Inner Parts of Spiral Galaxies and Ellipticals	$10\ h_o$	$\sim 10^{-2}$
Binaries and Small Groups	$40\text{-}100\ h_o$	0.04-0.1
Large Clusters of Galaxies	$100\text{-}800\ h_o$	0.1 to 0.8

*based on multiplying M/L times the Kirshner-Oemler-Schecter[a] luminosity density

[a]KIRSHNER, R.P., A. OEMLER, JR. & P.L. SCHECTER. 1979. Ap. J. 84: 951.

TABLE 2

Possible Missing Mass (Light) Candidates

Low Mass Stars

Planetary Type Objects

Rocks, Comets or Other Forms of Semi-Solid Debris

Clumps of Gas (Smooth Gas Not Allowed Because of Absorbtion)

Low Mass Neutrinos

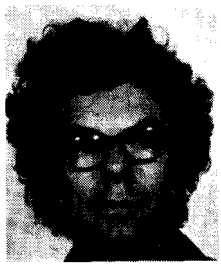
High Mass Neutral Heavy Leptons or Other Massive Weakly Interacting Particles

Monopoles

Black Holes

WHY AND HOW TO DETECT THE COSMOLOGICAL NEUTRINO BACKGROUND

J.SCHNEIDER
Observatoire 92190 MEUDON (France)



The Standard Big Bang theory predicts, parallel to the cosmic radiation background at 2.7K, a cosmological neutrino background. We successively discuss the cosmological signification and the feasibility of the detection of this background.

I - INTRODUCTION

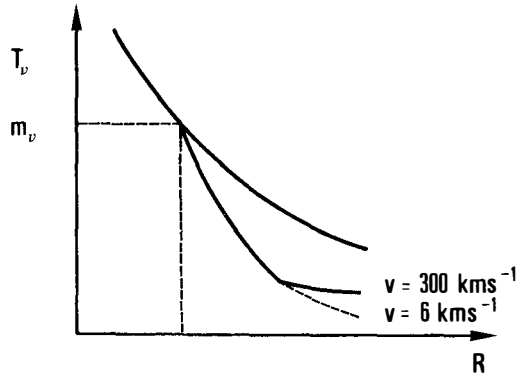
The standard Big Bang theory, whose qualitative ideas root back to the fourties had predicted the existence of a cosmic thermal radiation at a temperature of a few degrees K. This prediction was successfull since the Cosmic Radiation Background (CRB) was detected at 2.7 K. Similarly the Big Bang theory predicts a Cosmic Neutrino Background (CνB), at a temperature of 1.9 K if neutrinos are massless; if they are massive, the temperature is lower as I shall recall later. One of the challenges in cosmology today is to detect this CνB, a goal which, thanks to ideas of Opher[1] and Lewis[2], seems now not unreasonable to reach. Before discussing the feasibility of such ideas, I will recall what cosmological information the detection of the CνB could provide.

II - WHY

Let us suppose in this chapter that the detection of the CνB is possible with all desirable precision for its amplitude, angular structure etc. I will assume the existence of F families of neutrinos. Today in 1981 we have F=3 with electronic , muonic and tauic neutrinos ($\nu_e, \nu_\mu \nu_\tau$). The possible existence of other species is an open theoretical and experimental question. I will also assume that these neutrinos have infinite lifetimes and will neglect complications due to possible oscillations between the different types.

The Standard Big Bang theory[3] predicts the existence of a CνB due to the decoupling of neutrinos at a temperature of about 1 Mev in the Early Universe. Because of the reheating of the CRB by e^+e^- annihilation at a temperature of 0.5 MeV, the CνB temperature T_ν is immediatly after this annihilation lower than the CRB temperature T_R by a factor of $(11/4)^{1/3}=1.4$. The later evolution of the CνB depends on the mass of the neutrinos. If the neutrinos are massless T_ν drops like T_R, namely as R^{-1} (R is the scale factor of the Friedmann Universe) and today T_ν = 1.9 K. If the neutrinos have a mass m_ν (for convenience supposed to be same for the three types of neutrinos), when T_ν has reached m_ν it drops as R^{-2}. Today this would give $T_\nu = 510^{-5}$ K for m_ν= 30eV which would correspond to a mean velocity of neutrinos of 6 kms^{-1}. But it is an observational fact that there is a mean velocity of matter (Galaxies) in the Universe of about 200 kms^{-1}. This velocity would be communicated to neutrinos through the gravitational interaction and finally keep them at a temperature of $\sim 10^{-2}$K. This evolution of T_ν is summarized

in Figure 1.



If there is in the CνB an equal number of neutrinos and antineutrinos their number density n_ν is equal to $2.\frac{3}{8}.\frac{4}{11}.n_\gamma = 110(T_R/2.7)^3$ for each type of neutrinos, where n_γ is the photon density in the CRB. This n_ν includes neutrinos + antineutrinos. The case of an unequal number of neutrinos and antineutrinos (leptonic asymetry) will be discussed in § 3. The detection of the CνB would lead to many interesting conclusions :

1. Existence of CνB

The Standard Big Bang theory is after all only supported by the abundance of light elements and the predicted existence of CRB. These two facts are compatible with other theories, for instance models of tepid Universes[4] in which the baryon/photon ratio is about 10^{-4} at the epoch of primordial nucleosynthesis and falls at 10^{-8} at z =500 due to the formation of population III stars. In these models the neutrino/photon ratios is not 2.4/11.3/8 as in the Standard model but ot the order of 10^{-4}.

2. Chemical composition

The Standard theory predicts an equal number of neutrinos for each type. If the detector would be sensitive to the neutrinotype, we could test this second prediction. It could indeed happen that complications due to oscillations between different types lead to a desequilibrium among spacies.

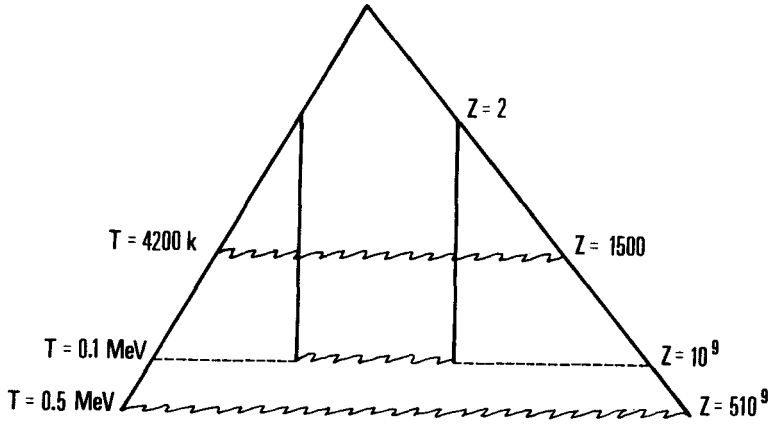
3. Neutrino asymetry

The present baryon/photon ratio = 10^{-8} is usually viewed as a remnant of the baryon asymetry A_B which occurs when $T_R = 10^{15}$Gev. A_B is defined as $(n_B - n_{\bar{B}})/n_\gamma$ where n_B and $n_{\bar{B}}$ are the baryon and antibaryon

densities. This asymetry is attributed to a combined effect of baryon non conservation ($\Delta B \neq 0$) and CP violation of Grand Unified interactions together with non equilibrium[5]. In Standard Grand Unified Theories (GUTs) based on SU(5) or SO(10) $\Delta B \neq 0$ is componsated by a lepton number non conservation ($\Delta L \neq 0$) such that $\Delta(B-L) = 0$. Consequently the neutrino asymetry A_ν of the CνB ($A_\nu = (n_\nu - n_{\bar{\nu}})/n_\gamma$) is not greater that 10^{-8} in these models. Phenomenological constraints on A_ν based on primordial nucleosynthesis[6] or on an assumed neutrino mass of 30eV[7] give only $A_\nu \leq 1$. Alternative GUTs, based for instance on $(SU(4))^4$, allow for $\Delta B=2$ and $\Delta L=2$ transitions with $\Delta(B-L) \neq 0$ [8]. Their manifestation are searched in $n\bar{n}$ [9] or $\nu\bar{\nu}$ oscillations. They are also likely to lead to a neutrino asymetry A_ν of the CνB different than A_B, and detectable if $A_\nu >> 10^{-8}$.

4. Primordial inhomogeneities

Let us assume in this section that neutrinos are massless. Then the CνN neutrinos are still relativistic and since they are decoupled they back travel on the light cone up to a T_R of 1 MeV. They could consequently give a direct access to the most primordial inhomogeneities. Figure 2 shows the different ways to test the homogeneity of the Universe at different epochs.



The small fluctuations, yet undetected, of the 2.7K CRB gives a direct measure of the homogeneity of the Universe back at a temperature of 4200 K. Its large scale homogeneity leads to the well known horizon problem [3]. Further back in time we have the light elements nucleo-synthesis at a temperature of about 0.1 MeV. Our knowledge of this epoch comes from the fossil abundances $X(He^4)$ and $X(D^2)$. $X(D^2)$ is sensitive to the baryonic density. This far extragalactic measurement of $X(D^2)$ (up to redshifts of 2 or 3) would be a means to reach some baryonic inhomogeneities at T_R=0.1 MeV. This measurement will hopefully be practicable within a near future through deuterium L_α absorption of the quasar continuum by intergalactic clouds [10]. Extragalactic He^4 does not help since $X(He^4)$ is insensitive to the matter density. With massless neutrinos we could look back at a temperature of 0.5-1 MeV (their decoupling temperature). A directional CνB detector would therefore give a direct access to small scale inhomogeneities at T_R=0.5-1 MeV; large scale isotropy at the same epoch would reenforce the horizon problem already raised by the CRB isotropy.

5. Local galactic dynamics

If on the contrary the neutrinos have a mass m_ν, they become non relativistic when T_ν drops below m_ν and they leave the light cone. They begin then to be affected by gravitational clustering. Consequently primordial geometrical information are lost. But the gravitational effects due to the present potential well of the Galaxy and/or local group (depending on the value of m_ν) can increase n_ν by factors of 10 to 10^4 [11]. The measure of n_ν would therefore be a source of information for this dynamics (discarding complications due to tepid cosmologies).

6. Finally remember Solar neutrinos

II- HOW

L.Stodolsky proposed in 1975 to detect the ν-induced spinflip of electrons or protons [12]. I will not discuss this proposal here. Another promising idea due to Opher [1] and reinvestigated by Lewis [2] is to use the coherent scattering of low energy neutrinos. Let me first recall the main points of Opher and Lewis.

The coherent scattering of a quantuum wave incident on a system with number density N of scatterers leads to a global refraction index n given by

$$n = 1+2n\,\lambda\!\!\!^{-2}\,N\,f(o) = 1+\frac{2nh^2N}{p^2}\,f(o)$$

where λ is the wavelength and f(o) the forward scattering amplitude. For neutrinos, f(o) is given by the Weinberg-Salam theory of weak interactions. The atomic scatterers being constituted by Z electrons and 3A quarks we have

$$f_e(o) = \pm \frac{GE\ (3Z-A)}{\hbar^2 c^2}$$

where $G = 10^{-49}$ ergcm^3 is the weak interactions coupling constants E is the center of mass energy. The + or - sign stands for ν or $\bar{\nu}$. Consequently the effect can discriminate ν's versus $\bar{\nu}$'s and is likely to make a large A_ν detectable. This refraction index defines a critical total reflection angle θ_c below which the ν's are totally reflected :

$$\theta_c = (n-1)^{1/2} = \sqrt{\frac{EGN(3Z-A)}{p^2c^2}}$$

The other neutrinos (ν_μ and ν_τ) interact only with quarks and for them

$$f(o) = \pm \frac{GE(Z-A)}{\hbar^2 c^2}$$

and

$$\theta_c = \sqrt{\frac{EGN(Z-A)}{p^2c^2}}$$

Thus we can discriminate ν_e versus ν_μ or ν_τ and partly measure the chemical composition of the CνB.

This total reflection phenomenon is well known and widely used for ultracold neutrons [13] and can be used to design neutrinos mirrors.

Consider then a square "mirror" of side a, thickness b and mass M. By total reflection each neutrino of momentum p_ν transfers to the mirror an impulse $\Delta p = 2p_\nu\theta_c$. The total impulse transferred to the mirror after a time T is

$$P = a^2\theta_c F\ \Delta pT$$

where F is the ν flux of the CνB through the mirror. Suppose a mechanical detector whose energy sensitivity is E_s. The detectivity condition is that E_s is smaller than the kinetic energy $P^2/2M$ acquired by the mirror, or

$$P = a^2\ \theta_c F\ \Delta pT > \sqrt{2\ E_s M}$$

Introducing the density of the mirror this finally gives a condition on T :

$$T > \frac{\sqrt{b\ Es\ \rho}}{2Fa\ \ p_\nu(n-1)}$$

This was Lewis' idea. Let us now discuss the feasibility of such an experiment by using as a mechanical detector the gravitational-wave antennas now designed. Such an antenna is characterized by its E_s and its Q-factor = $\omega \tau$, where is τ is the characteristic time of damping of vibrations of pulsation ω. Detectors are now designed to reach the "quantum limit" $E_s = h\omega$ (= 10^{-24} erg for $\omega = 10^3$ herz)[14] Q-factors of 10^{10} are not unrealistic within a few years [15]. There is even the possibility to go beyond the quantuum limit [16], although no realistic technological device is planned for the moment.

Consider then a stack of one thousend mirrors 1m wide each and 1mm thick (the value of the CνB wavelength) made of a material with $3Z-A \simeq 50$. Then $N \simeq 5 10^{23}$. From here on we have to distinguish between zero and non zero mass neutrinos, because $E = pc \simeq kT_\nu \simeq 10^{-3}$eV for massless neutrinos whereas $E \simeq m_\nu c^2$ for neutrinos with masses larger than kT/c^2 :

A. $m_\nu = o$

Then $F = n_\nu c = 10^{12} cm^{-2} sec^{-1}$

$$n-1 = \frac{GN(3Z-A)}{E} \simeq 10^{-10}$$

$$\theta_c = (n-1)^{1/2} \simeq 6"$$

$$T \gtrsim 3 10^7 \text{ sec}$$

B. $m_\nu = 30$ eV/c^2

Then $F \simeq 10^2 n_\nu v_\nu$ where v_ν is the mean neutrino velocity; we have estimated a 10^2 amplification factor due to neutrino gravitational clustering by the Galaxy or the local group. We have seen than $v_\nu \simeq 300$ kms^{-1}

Then $$n-1 = \frac{GN(3Z-A)}{p^2 c^2} m_\nu \simeq 10^{-7}$$

$$\theta_c \simeq 10'$$

$$T \gtrsim 3 10^5 \text{ sec}$$

Suppose the stack of mirrors is gravitationally coupled to a gravitational antenna with $Q = 10^{10}$. Then the vibration mode of the antenna induced by each elementary impulse communicated by an incoming neutrino will last, for a typical ω of 10^3, 10^7sec. This characteristic damping time is, at least for massive neutrinos, larger than the time determined above and it is possible to integrate the elementary impulses during this time. For massless

neutrinos, the detectivity condition is less favourable. One way of improvement would be to use the mirrors to focalize the neutrinos to amplify the flux F.

The detection of CνB is still at the stage of gedanken experiments, but work is in progress to design more workable devices.

I thank S. Bonazzola for discussions.

REFERENCES

1) R. OPHER,Astron.Astrophys. 37,135 (1974)

2) R. LEWIS,Phys. Rev. 21D, 663 (1980)

3) S. WEINBERG,Gravitation and Cosmology J. Wilay and Sons 1972 . Chap. 15

4) M.REES,Proceedings of the 5th IAU Meeting, Liège 1980 (to be published)

5) S.WEINBERG, Phys. Rev. Letters 42, 850 (1979)

6) Y.DAVID, H.REEVES, Cosmologie Physique, Ecole des Houches R. Balian Editor, North Holland (1980)

7) J.P.LUMINET, J.SCHNEIDER, Astron.Astrophys. in press

8) R. MOHAPATRA, R.MARSHAK, RIAZUDDIN, VPI-HEP. Preprint (1980)

9) C.FIDECARO, M.BALDO-CEOLIN, K.GREEN, ILL (Grenoble) proposal

10) C.LAURENT, Proceedings of "2ème Colloque Français du Télescope Spatial" P. Felenbok editor. Paris. Observatory (in press)

11) TREMAINE, J. GUNN, Phys.Rev.Letters, 42, 407

12) L.STODOLSKY, Phys. Rev.Letters 34, 110 (1975)

13) R.GOLUB, J.PENDLEBURY, Rep.Progr.Phys. 42, 439 (1979)

14) C.CAVES Caltech OAP preprint n°521 (1978)

15) S.BONAZZOLA (private communication)

16) K.THORNE, C.CAVES, V.SANDBERL, M.ZIMMERMANN . Caltech OAP preprint n° 532 (1978)

ULTRAVIOLET BACKGROUND RADIATION AND THE SEARCH FOR DECAYING NEUTRINOS

Richard C. Henry
Physics Department
The Johns Hopkins University
Baltimore, Maryland 21218

ABSTRACT

The spectrum of the observed far-ultraviolet background at high galactic latitudes provides superficial evidence of radiation from neutrino decay, but the spectrum is so uncertain that conclusions are not possible. A limit of $\sim$300 photons (cm^2 sec ster Å)$^{-1}$ is set on any non-stellar ultraviolet flux above latitude 20°. The disagreement between the Berkeley and the Johns Hopkins ultraviolet background radiation data is analysed.

I. INTRODUCTION

De Rújula and Glashow[1] point out that slowly moving massive neutrinos, remnant from the big bang, are expected to decay into lighter neutrinos and ultraviolet photons, the neutrino lifetimes being long on the Hubble scale. Stecker,[2] and Kimble, Bowyer, and Jakobsen,[3] have set limits on τ, the lifetime of a neutrino, of $\tau > 10^{23}$ sec, from the observational upper limit on the far-ultraviolet background of 285 ± 32 photons $(cm^2 \text{ sec ster } Å)^{-1}$ of Anderson, Henry, Brune, Feldman, and Fastie,[4] and various other upper limits.

The calculation of the lifetime limit from the observations is straightforward. There are about 500 3° K photons cm^{-3} relict from the big bang, and ∿150 neutrinos cm^{-3}, throughout the universe. If these neutrinos decay with lifetime τ, the volume emission of the universe is $150/\tau$ photons cm^{-3} sec^{-1}. Consider a shell of thickness dr distant r from us; its volume is $4\pi r^2 dr$, and the emission from this shell is $150 \cdot 4\pi r^2 dr/\tau$ photons sec^{-1}. This is spread over $4\pi r^2 cm^2$, resulting in an incident intensity of $150\ dr/\tau$ photons cm^{-2} sec^{-1}, or $150\ dr/4\pi\tau$ photons cm^{-2} $ster^{-1}$ sec^{-1}. Hubble's law is $v = Hr$ where $H = 3.2 \times 10^{-18}$ sec^{-1} (= 100 km sec^{-1} Mpc^{-1}), so $dr = dv/H$; but $dv = c\ d\lambda/\lambda$, so $dr = c\ d\lambda/\lambda H$ and the incident intensity per Ångstrom is $150\ c\ d\lambda/4\pi\tau\lambda H$ photons cm^{-2} sec^{-1} $ster^{-1}/d\lambda$ cm 10^8 Å cm^{-1}, or

$$150\ c/4\pi \cdot 10^8 \tau\lambda H \text{ photons } cm^{-2} \text{ sec}^{-1} \text{ ster}^{-1} \text{ Å}^{-1},$$

which units we will henceforth refer to as "units."

The only spectra of the high-galactic latitude far-ultraviolet background that have been obtained are those of Anderson *et al.*[4] and Henry, Feldman, Fastie, and Weinstein.[5] Stecker[2] points out that Anderson *et al.*[4] report a step in the spectrum of the ultraviolet background at about 1700 Å. A step of this kind (a sharp increase toward longer wavelengths) is exactly what one would expect from the decay of neutrinos distributed uniformly throughout the universe. The increase in intensity toward longer wavelengths is confirmed by Maucherat-Joubert, Cruvellier, and Deharveng.[6] The size of the step (see Figure 12 of Anderson *et al.*) is ∿300 units. Setting this equal to our expression for the expected intensity gives a lifetime for the putative neutrinos of $\tau = 2 \times 10^{23}$ sec, the neutrino mass being 14.6 eV (assuming that the neutrino into which it decays is much lighter). However, Stecker[2] does not point out that the only other observation of the spectrum that exists, that of Henry *et al.*,[5] shows a step of the *opposite* kind, namely a sharp *decrease* toward longer wavelengths, occurring at slightly shorter wavelengths. Until discords in the observations are resolved, we cannot conclude that neutrino decay has been observed.

Schramm, at this *rencontre*, summarized the constraints that exist on heavy neutrinos of various masses, m_ν, and lifetimes, τ_ν. Figure 1 is adapted from his paper; the limit from the ultraviolet background is shown as a circled dot. The far-ultraviolet background limit is of particular interest because it falls in the range of corresponding neutrino masses that is of the greatest interest for cosmology, as described in detail by Schramm and Steigman.[7] Neutrinos of mass > 3 eV might be gravitationally bound in clusters of galaxies, while more massive neutrinos could be bound up in groups of galaxies, or even in single galaxies.

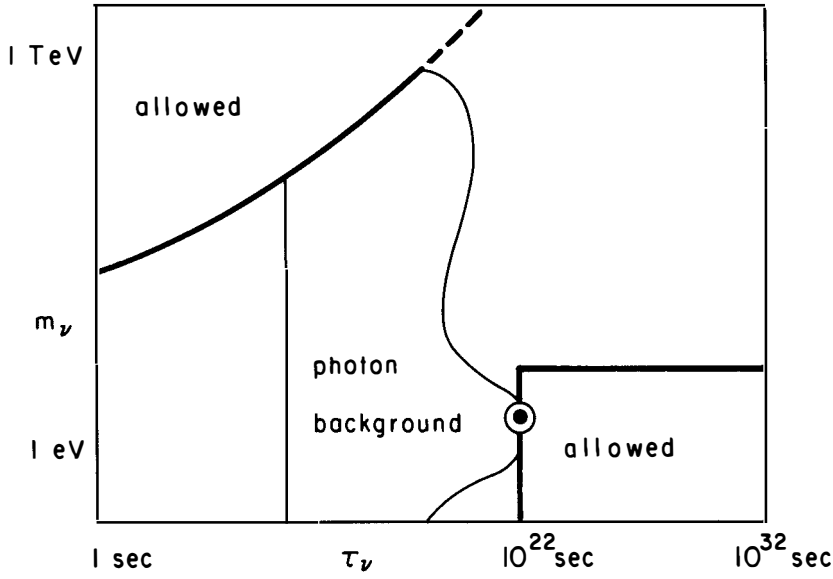


Fig. 1. The regions of the neutrino mass-lifetime domain that have not been excluded by observation are marked "allowed." The circled dot is the lower limit on the neutrino lifetime that is obtained from measurement of far-ultraviolet background radiation, as described in this paper. The ultraviolet observations are particularily important because they correspond to neutrinos having masses of appropriate value needed to gravitationally bind clusters of galaxies.

The high-galactic latitude ultraviolet background radiation observations are summarized in Figure 2. The top scale in the figure gives the corresponding (heavy) neutrino mass. The area marked with a heavy line is the Anderson *et al.*[4] spectrometer result, the filled circle being their long-wavelength photometer result, which agrees with Maucherat-Joubert *et al.*[6] (inverted triangles, lower point showing the result of a correction procedure). The open circle is the

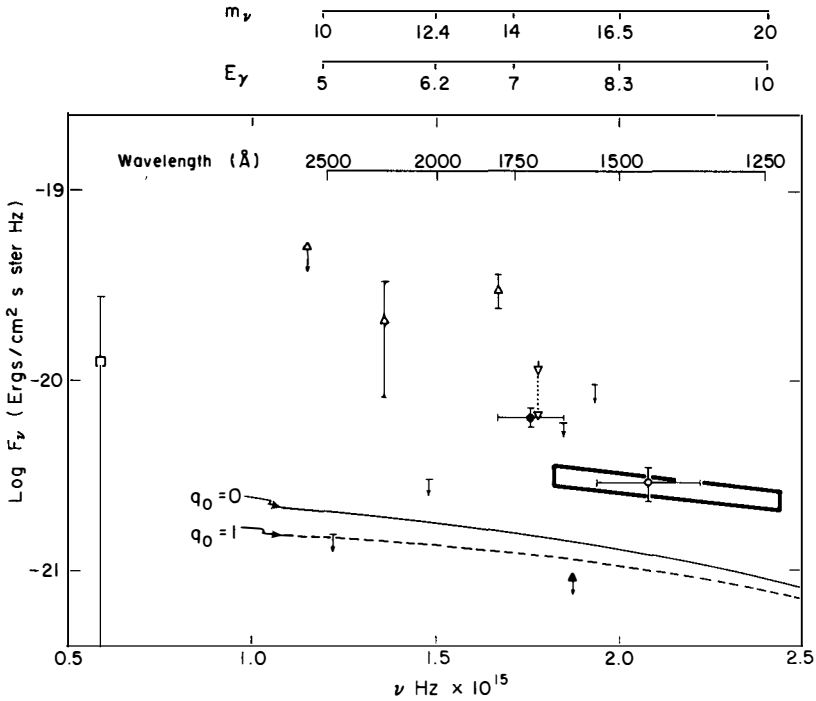


Fig. 2. This is a summary of observations of the high-galactic-latitude ultraviolet background radiation. Log intensity is plotted against frequency. Additional abscissa scales are provided giving the wavelength in Ångstroms; the energy in eV; and the rest-mass of the heavy neutrino that would decay into photons of the corresponding wavelength (assuming that the daughter neutrino is very light). The observations superfically provide evidence for neutrinos of mass ∿9 eV, 12 eV, and 14.6 eV, but known large discords among independent measurements of the background render conclusions impossible.

quoted Berkeley photometer lower limit; it should actually be raised a factor 1.33 (see below). The filled triangle represents the conflicting Apollo 17 result of Henry *et al.*[5]; the rest of their spectrum agrees precisely with Anderson *et al.* and is not shown.

The horizontal lines represent upper limits, from Lillie,[8] while the triangles are from Pitz, Leinert, Schulz, and Link.[9] The latter represents a sounding rocket flight, and the authors warn that an airglow component may be present, at least at certain wavelengths. The square represents the observation, from the ground, of Dube, Wickes, and Wilkinson.[10]

If one chooses to ignore the Apollo 17 result, a procedure that I do not recommend, and to take all of the other observations at face value, the data shown in Figure 2 could suggest the presence of emission lines from three neutrinos, masses 9 eV, 12 eV, and 14.6 eV, in the halo of our galaxy. Actually this would violate the supposition that the residual neutrino is much lighter than its parent, but it is not worth exploring further, because the data quality is such that no conclusions can be drawn at all, at this stage.

The subject of ultraviolet background radiation has been reviewed by Davidsen, Bowyer, and Lampton;[11] by Paresce and Jakobsen;[12] and by Henry.[13] All reviews have emphasized the highly discordant nature of the observations. In the remaining sections of the present paper, we re-examine one such discord, and we review the Apollo 17 observations of large regions of the sky.

II. BERKELEY AND JOHNS HOPKINS

Paresce, McKee, and Bowyer[14] have drawn attention to a specific disagreement between their measurement of the brightness of the cosmic background, and that of Anderson *et al.*[4] The disagreement is displayed in Figure 3, which is a plot of Anderson *et al.'s* counting rate (averaged over 1405 - 1605 Å) as a function of time during the Aries rocket flight (the Aries altitude is given on the top scale). The Aries spectrometer was pointed at four targets during the flight, the first three of which are marked with a horizontal line giving the level of stellar signal that is estimated to be present. The count rate that should have been observed at targets 1 and 2, if the Berkeley observations[14] are correct, is shown as horizontal lines marked B. The Berkeley photometer bandpass was ∿1380 - 1430 Å, with a long tail to longer wavelengths.[15]

The absolute calibration uncertainty claimed by Berkeley[15] is $\leq \pm 20\%$, and by Hopkins[4] $\sim \pm 10\%$. The discrepancy is therefore very significant, and warrants investigation.

Could Hopkins be missing light that is really there? This hardly seems possible, because of the in-flight check that was made at the fourth target, the star 11 CMi. The signal that is expected, on the basis of the TD-1 satellite observation[16] of this star, is shown in Figure 3. If anything, we detect too much radiation, not too little. (The extra flux that is observed is undoubtedly due to other stars at this low (18°) galactic latitude.) Another point that might occur, is the question of whether the Hopkins spectrometer has the same efficiency for the detection of diffuse radiation as it has for point sources. Henry[13] has presented detailed evidence that it does.

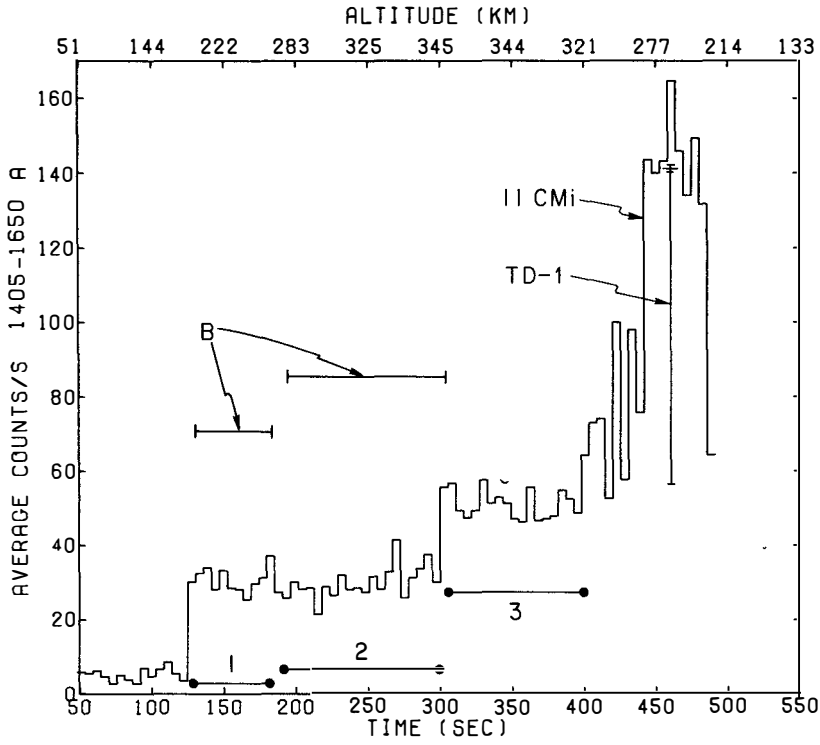


Fig. 3. This is a histogram of the 1405 - 1650 Å count rate of the spectrometer of Anderson *et al.* as a function of time (bottom scale), and of Aries rocket altitude (top scale). The estimated stellar contribution to the signal is shown by a horizontal line for the first three targets, and the contribution of the calibration star 11 CMi to the fourth target, based on its brightness as measured with the TD-1 satellite, is also indicated. The signal level that should have been observed, on the first two targets, if the Berkeley Apollo-Soyuz observations are correct, is shown by the horizontal lines marked B. The observation of 11 CMi shows that if anything, Anderson *et al.* are observing too much flux, not too little. Independent evidence is given in the text that the Berkeley data are subject to contamination of unknown origin.

We tentatively conclude, therefore, that Hopkins is not observing too low a signal, and turn to the opposite possibility, namely that Berkeley is obtaining too high a signal.

Is there any way to tell, internally, from the Berkeley data, whether contaminating signal is present? Fortunately, there is, as we now detail.

Paresce, Margon, Bowyer, and Lampton have reported[15] the discovery of many far-ultraviolet bright patches at moderate and high galactic latitudes. These are shown in Figures 4 and 5, which are adapted from Figure 3 of Paresce *et al.*[15] using the conversion from counts sec^{-1} to units that is reported in their paper. More than 200 patches brighter than 2000 units are reported, and at least 75 patches are as bright or brighter than 4000 units. These patches, if real, are of great interest and importance in the study of cosmic far-ultraviolet background radiation. No coordinates of the patches are given.

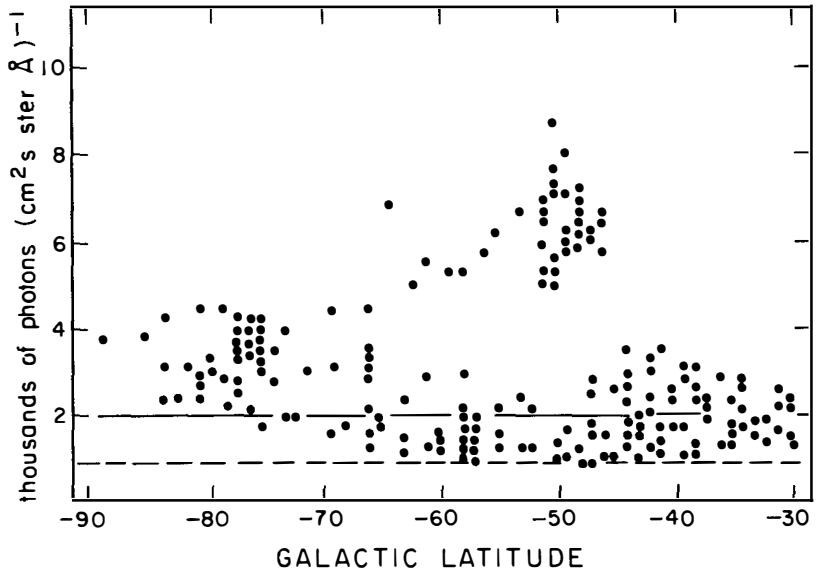


Fig. 4. An attempt to detect contamination in the Berkeley data from internal evidence. The bright patches of diffuse radiation reported in the first Berkeley paper are shown as filled circles. The second Berkeley paper involves a subset of the same data; namely, observations made from a geomagnetic latitude $|b| < 45°$, and involves more detailed stellar corrections. In the subset, there are no points above 2000 units (solid line) and the mode is as shown with the dashed line. If the striking difference between the two sets of data is due to stellar correction, then most of the solid points above the line (and presumably many of the points below the line) are not measures of diffuse background at all. Alternatively, if the difference is due to geomagnetic latitude of observation, geophysical effects are responsible for most of the points above the line (and presumably many of the points below the line), and again diffuse cosmic ultraviolet light is not the source of most of the Berkeley data.

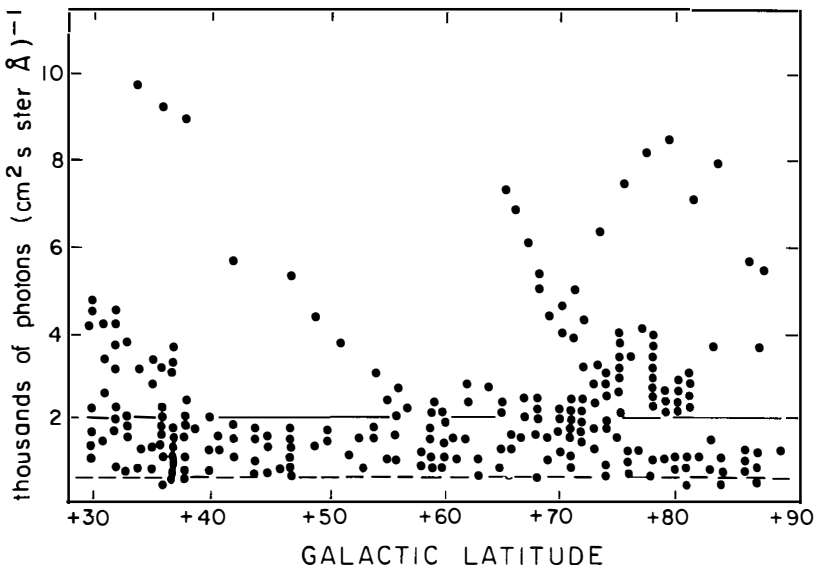


Fig. 5. The same as Figure 4, but for nothern galactic latitudes. If coordinates were provided for the many extremely bright patches, independent verification of their reality would be possible.

Paresce, McKee, and Bowyer[14] analysed a different subset from the same data source, with results that are strikingly different. The mode (most frequent intensity) that they report in the north and the south is shown as the dashed line in Figures 4 and 5. Also, they now report *no* patches brighter than 2000 units (solid line in Figures 4 and 5).

Both data samples were drawn from the 3200 data points of Figure 4 of Paresce *et al.*[15] This data set excludes data obtained "at times when the spacecraft was in daylight, in the South Atlantic Anomaly, or in the auroral zones, or when the detector internal background exceeded 3 counts s^{-1} or the zenith angle of the line of sight exceeded 90°,"[15] and also includes only data obtained when "the angle between the instrument line-of-sight and the sun was greater than 50°."[14]

The differences between the two subsets are much too large to be the result of sampling statistics, and thus must be produced by one, or both, of a) the stellar correction, and b) an exclusion of data to avoid possible geophysical effects.

a) Stellar Correction

Paresce *et al.*[15] exclude "all signals consistent with a point source transit through the instrument at a 2 σ statistically significant level" (though later they say that "variation in [diffuse background] intensity of factors of 2 or 3 over angular distances comparable to the instrument field of view . . . are common"), and also exclude data obtained when the instrument field of view contained a known star of $m_v \leq 6.5$ and spectral type A2 or earlier" (an unreddened $6^m.5$ A2 star would contribute 275 units). Note that an unreddened $6^m.6$ B3 star would contribute 6600 units.

For the other subset, Paresce *et al.*[14] remove stellar contributions by use of the SAO and TD-1 catalogs.

b) Geophysical Effects

Paresce *et al.*[15] exclude data taken in the auroral zones, while Paresce *et al.*[14] exclude data when the spacecraft magnetic latitude was > ± 45°.

One or both of these differences between the two data sets must produce the large difference which is apparent in Figures 4 and 5. It would be interesting to know which is responsible. If it is the stellar correction, then the bright patches of diffuse emission reported in the first Berkeley paper do not exist. If it is the restriction to low geomagnetic latitudes, then contamination of unknown origin, strongly dependent on geomagnetic latitude of observation, is present in the Berkeley data.

Finally, we note that the upper limit on a cosmological background of 300 units that was reported by the Berkeley group,[14, 15] is the result of subtracting 113 units from their minimum upward-looking intensity. This 113 units is the airglow brightness that they observed when looking down at earth. There is no reason to assume that the same airglow source is present *above* Apollo-Soyuz altitude (223 km), and if an upper limit on the cosmic background is to be obtained from the Berkeley data, that limit is ∿400 units.

III. APOLLO 17

The possibility exists that the far-ultraviolet signal that is detected[4,5] at high galactic latitudes originates in the light of galactic plane OB stars scattering off of high-galactic-latitude interstellar dust. The Berkeley group[14,15] advocate this as the probable origin of some of their reported moderate-galactic-latitude bright spots, and they present four separate, dif-

ferent, correlations between ultraviolet brightness and 21-cm intensity (i.e., presumably, dust) in support of this idea. The four correlations involve a total of 128 of their 3200 reported observations of the diffuse background.

In contrast, Henry, Anderson, Feldman, and Fastie[17] observed low background intensities at moderate galactic latitudes; intensities that could be accounted for completely as the radiation of stars in their field of view, with no component of dust-scattered light from galactic-plane OB stars. Henry *et al.*[17] estimated the direct stellar contribution using an ultraviolet-calibrated version of the SAO star catalog. This is not ideal, but Henry[13] has shown that in the case of sufficiently large field of view, the deduced diffuse background is very insensitive to the accuracy of the stellar correction. Henry[18] has repeated the analysis, using the TD-1 integrations of Gondhalekar, Phillips, and Wilson[19] for the stellar correction, and he finds that at 1565 Å the interstellar grains have albedo a $\sim$0.5 and scattering parameter $g > 0.7$. (In contrast, Witt[20] find $g \simeq 0.25$.) The value $g = 0.7$ represents very strong forward scattering, but even if $g = 0.7$, the residual diffuse ultraviolet background at the galactic poles could still represent the dust-scattered light of galactic-plane OB stars. Henry *et al.*[17] obtained a value $g \simeq 0.9$. This represents such strong forward scattering that only a negligible amount of light would be back-scattered at high galactic latitudes, and the observed residual is probably extragalactic.

The Henry *et al.*[17] work has been greatly extended by Anderson, Henry, and Fastie[21] and by Anderson.[22] The observations were made with the Apollo 17 far-ultraviolet spectrometer of Fastie.[23] This work is recapitulated in some detail here, allowing presentation of additional selections from the data.

The spectrometer field of view was 12° x 12°. There was no telescope; the field of view was limited by a baffle.[23] The instrument scanned from 1180 to 1680 Å every 12 seconds. Resolution was 11 Å, but for purposes of analysis, the data were gathered into six wavelength intervals, numbered 1 through 6. These intervals are detailed by Henry *et al.*[24]

The observed signal as a function of time, for sky-scan number 6, for wavelength interval four ($\lambda 4$), is displayed in Figure 6. The spectrometer looked out of the Apollo 17 service bay, and scanned around the sky as the spacecraft rotated. The lower panel in the figure gives the observed $\lambda 4$ brightness (solid line) and the star-catalog prediction (dashed line). The ordinate scale is linear (left scale) up to the solid line, and logarithmic (right scale) above it.

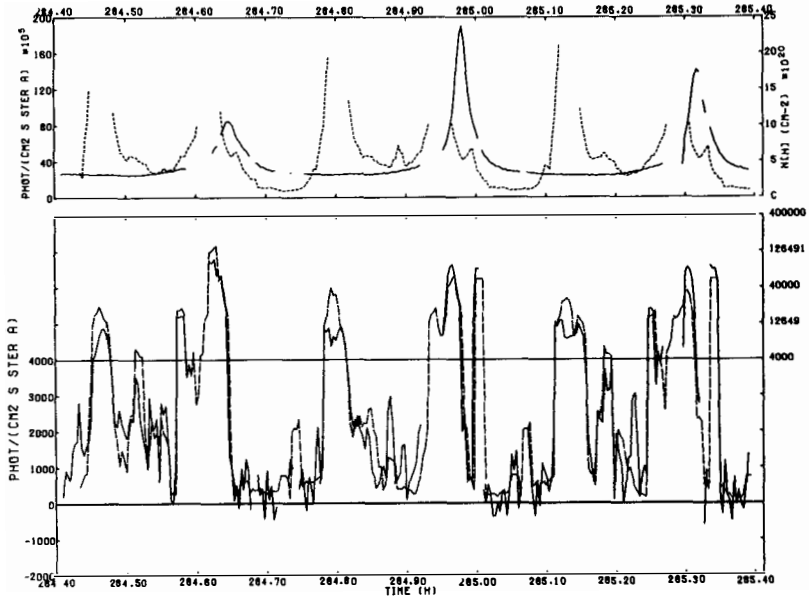


Fig. 6. The bottom panel compares the observed Apollo 17 ultraviolet signal (solid line) with the star-catalog prediction (dashed line). At the highest latitudes (lowest intensities), the observed signal is less than 1000 units, without any stellar correction at all. The top panel gives the ∿simultaneously observed Lα signal (solid line), and also the 21-cm hydrogen surface density (dashed line), which is a fair measure of the amount of dust in the field of view. The spike in Lα represents passage of the earth through the field of view. The observations were made on trans-earth coast.

The data have been corrected for a large but very well determined dark current, and for instrumentally-scattered Lα radiation (which is particularly strong when the instrument line-of-sight sweeps through earth—see top panel in the figure). Note that the average high-galactic-latitude intensity is well below 1000 units without any stellar correction at all.

The average stellar correction that is involved can be deduced from the data in Figure 7 (from Henry *et al.*[24]) which gives the average total integrated λ4 signal as a function of Gould latitude (Gould latitude differs from galactic latitude by an amount that varies from 0° to 20°, only). From 30 - 35° latitude, the average stellar correction is only ∿2000 units, so even a 30% error in the stellar correction would produce an error of only 600 units in the deduced

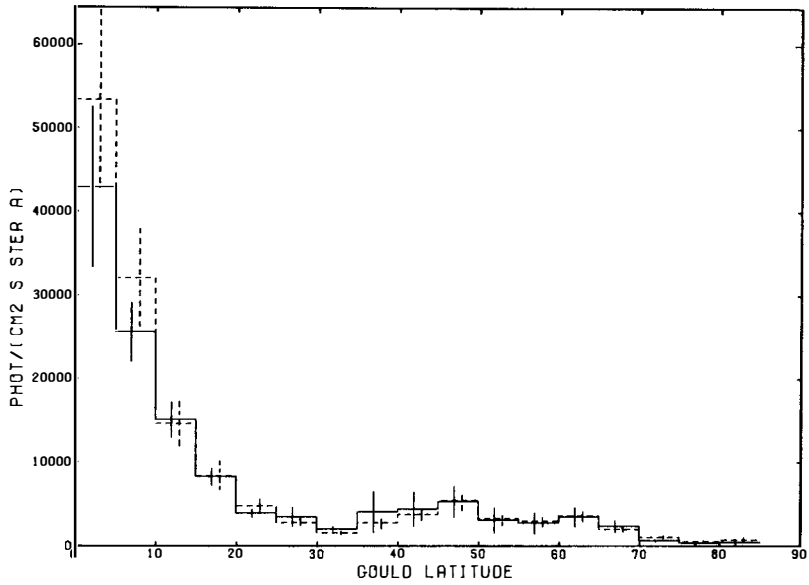


Fig. 7. The integrated average brightness of the ultraviolet sky observed from Apollo 17 (solid line), as a function of latitude. The star-catalog prediction is given by the dashed line. The integrated brightness over large regions in the Gould (∿galactic) latitude interval 30 - 35° is only ∿2000 units. The brightness is higher than this at higher latitudes because of the effects of small numbers of very bright nearby stars.

average background. The stellar correction is typically even lower at higher galactic latitudes; the average in Figure 7 is higher because of the effects of a very few extremely bright stars at mid-latitudes.

The astronauts gathered data during six intervals, making scans over six small circles on the sky. Detailed maps of the scan path are given by Henry *et al.*[24] The largest amount of data was obtained during the three astronaut sleep periods, sky scans 1 through 3. The reduced data (that is, corrected for dark-current, grating scattered Lα, and direct starlight) for one of the two galactic plane passages, for each of these three sky-scans, are given in Figures 8, 9, and 10 respectively, for λ4, 6, and 4, respectively. In the bottom panel in each figure, the residual signal for each spectral scan is plotted with a symbol, the size of which is proportional to the weight that should be given to the point, considering, primarily, the size and uncertainty of the stellar correction. The symbols are systematically smaller (i.e., lower weight) in

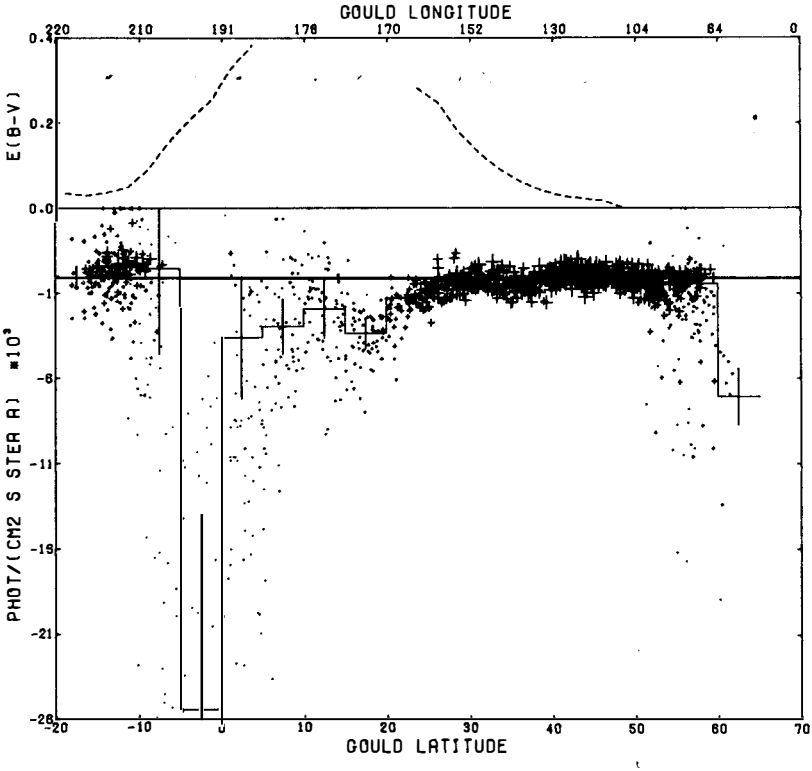


Fig. 8. The ultraviolet intensity (lower panel) during one of the two galactic-plane crossings of sky-scan number one, after correction for dark current, grating-scattered Lα, and direct starlight. The size of the plotting symbol is proportional to the weight of the observation. The weighted average is given as a histogram. The stellar correction is very large at low latitudes, so those observations have very low weight. The top panel indicates the amount of interstellar dust present in the field of view. At latitudes 25° to 35°, no signal is seen that could be attributed to the light of galactic-plane OB stars scattering off of the large amount of interstellar dust that is present at those latitudes. Note particularily that the stellar correction is not significantly larger in this latitude interval than it is at higher latitudes (otherwise, the points would be smaller; that is, would have lower weight).

Figure 9, because the correction for grating-scattered Lα was large for λ6. The top panel in each figure gives the overage color excess, E(B-V), which is a measure of the amount of interstellar dust, as estimated from the 21-cm observations

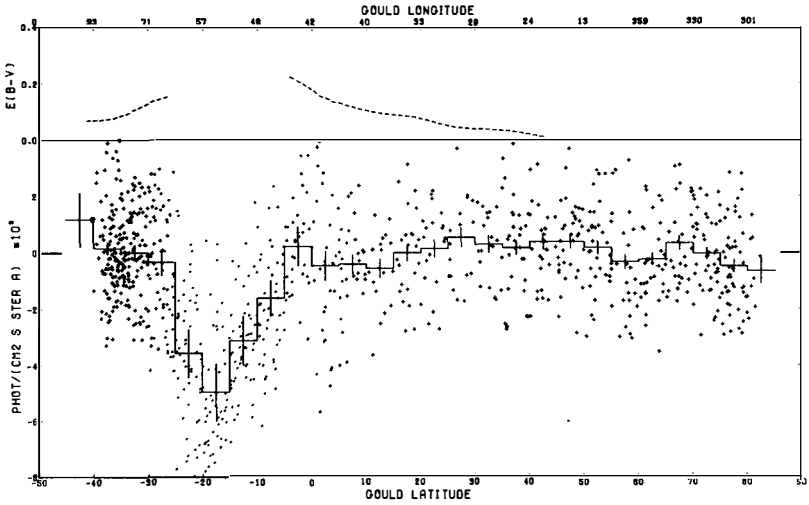


Fig. 9. The same as Figure 7, except that these are data from the second of the six sky scans. The points are systematically of lower weight because the correction for grating-scattered Lα was particularly large for the wavelength interval involved in this plot. Again, there is no evidence for light scattered from dust.

of Heiles.[25-27] Examining the figures, we see that in the case of all three sky scans, large amounts of data are obtained in latitudes where a great deal of dust is present, yet no ultraviolet signal is observed that could be attributed to the light of galactic-plane OB stars scattering from this dust. The stellar correction, as indicated by the size of the plotting symbol, is not significantly larger in these regions than it is at the highest latitudes. The ordinates in the three figures are in thousands of units. Clearly, fewer than 1000 units are present, that could be due to light scattered from dust. Complete analysis,[21,22] averaging all of the data, sets a limit of about 300 units on such light.

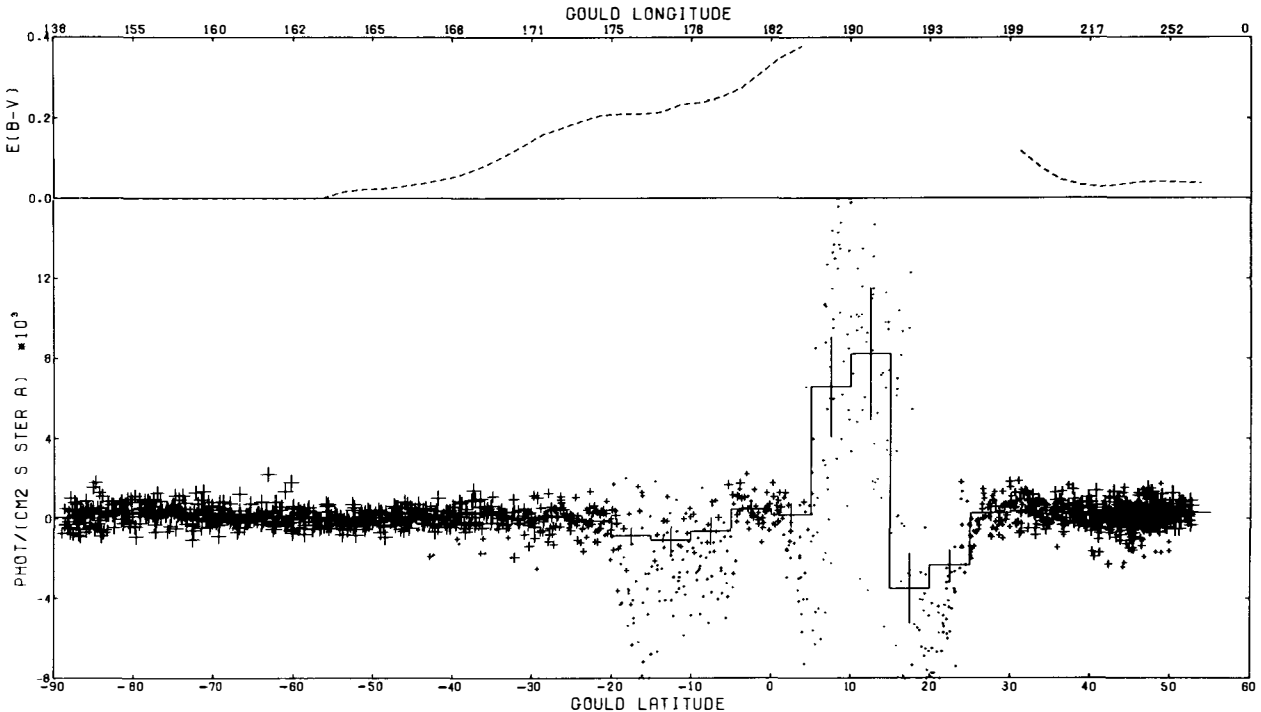


Fig. 10. The same as Figures 8 and 9, but for the third sky scan. Again, there is no evidence for light scattered from dust.

These data may be presented in another way. Figure 11 gives the residual λ6 intensity (after correction for dark current, grating scattered Lα, and direct starlight), for each spectral scan during the six sky scans (some points fall outside the plotted region), as a function of E(B-V), i.e., as a function of the amount of dust in the field of view. The weighted mean (marked with its uncertainty) is indicated by the histogram. An upper limit of 300 units for any dust-scattered light appears very reasonable, even at this wavelength where the correction for grating-scattered Lα is large.

A powerful check on the precision of the correction for grating-scattered Lα is possible. We simply repeat the analysis, correcting each scan for dark current and for starlight, but making no correction at all for grating-scattered

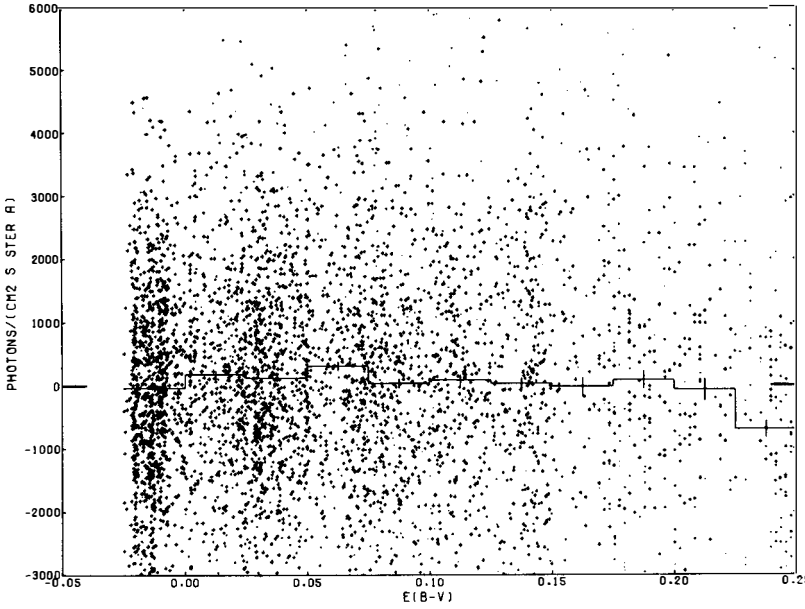


Fig. 11. The residual intensity (after correction for dark current, grating-scattered Lα, and direct starlight), for all of the spectral scans, is plotted against a parameter that is a measure of the amount of dust that is in the field of view. The weighted mean (marked with its uncertainty) is shown as a histogram. An upper limit of 300 units on any light scattered from dust in the field of view seems very reasonable.

Lα, and see how well the residual correlates with the celestial Lα signal, directly observed just a few seconds earlier. The result, for λ6, is shown in Figure 12, where the weighted mean (marked with its uncertainty) is again indicated by a histogram. I suggest that the reader lay a transparent ruler on this histogram, through zero, to verify the linearity of the correlation.

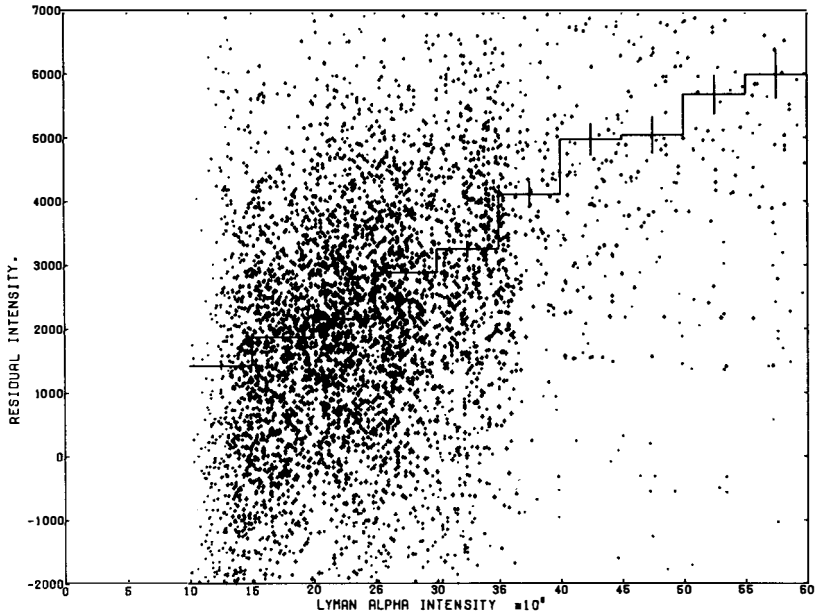


Fig. 12. The same as Figure 11; except that here *no* correction has been made for grating-scattered Lα radiation, and the residual is plotted against the observed Lα brightness of the sky at that position. Clearly, the grating is scattering Lα to longer wavelengths, and clearly we have a good measure of the *amount* of grating-scattered Lα, and can remove it from our data with precision.

A second example of this, one for λ4, is given in Figure 13, which is identical to Figure 18 of Anderson *et al.*,[21] except that here we do *not* delete the portion of the data that includes observation of the star α And. Comparing the two versions of the figure, the scans in Figure 13 that contain α And leap to the eye, at Lα ≃ 19, residual intensity = 3000 - 6000 units. The SAO catalog stellar correction is clearly in error for this one particular star, which is why Anderson *et al.*[21] deleted this portion of the data. The figure therefore supplies clear evidence for the adequacy of the SAO catalog corrections for

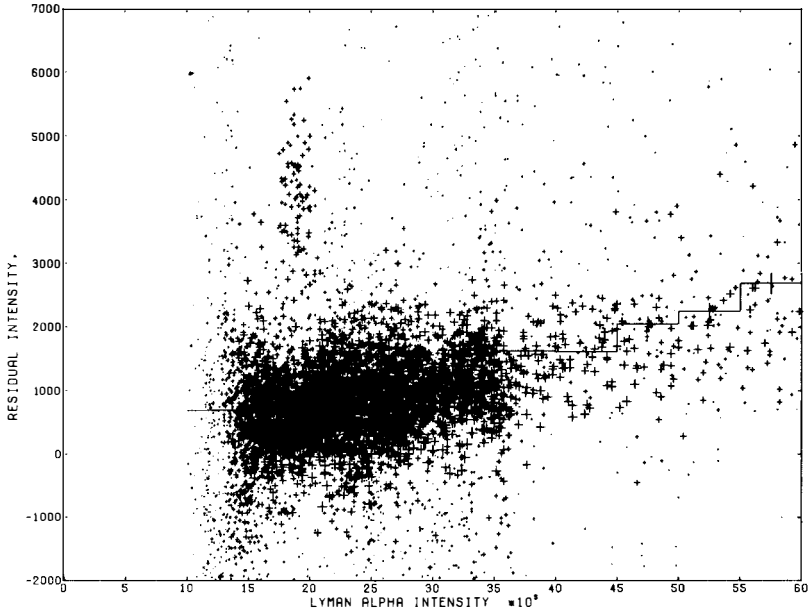


Fig. 13. As in Figure 12 (but for another wavelength), each spectrum has been corrected for dark current and stellar contribution *only*, and the residual has been plotted against the observed intensity of Lα radiation that is present. A clear correlation of residual intensity with Lα is apparent. Unlike elsewhere in the present work, the data that include the star α And have *not* been deleted—and are very apparent as an isolated group of high-residual points. The star-catalog correction for this star is clearly incorrect (which is why we delete these data)—and is equally clearly adequate for other stars.

other stars.

IV. CONCLUSION

Comparison of Johns Hopkins and Berkeley data on the ultraviolet cosmic background, and internal examination of the Berkeley data, leads to the conclusion that at latitudes above about 20°, the limit on any light scattered from interstellar dust, or indeed, the limit on light from any source other than stars, is of order 300 photons $(cm^2 \text{ sec ster } \mathring{A})^{-1}$. Examination of the spectrum of the observed background at the highest galactic latitudes provides superficial evidence for radiation from decaying neutrinos, but the spectrum is so uncertain that conclusions are not possible.

William Fastie is Principal Investigator on the Apollo 17 experiment. I am grateful to him; to Paul Feldman, and the other Co-investigators on Apollo 17; and especially to my graduate student, Rodney Anderson. This work was supported by NASA contract NAS 9-11528, and NASA grant NGR 21-001-001 to the Johns Hopkins University.

REFERENCES

1) A. De Rújula and S. L. Glashow, Phys. Rev. Lett. 45, 942 (1980).
2) F. W. Stecker, Phys. Rev. Lett. 45, 1460 (1980).
3) R. Kimble, S. Bowyer, and P. Jakobsen, Phys. Rev. Lett. 46, 80 (1980).
4) R. C. Anderson, R. C. Henry, W. H. Brune, P. D. Feldman, and W. G. Fastie, Ap. J. 234, 415 (1979).
5) R. C. Henry, P. D. Feldman, W. G. Fastie, and A. Weinstein, Ap. J. 223, 437 (1978).
6) M. Maucherat-Joubert, P. Cruvellier, and J. M. Deharveng, Astron. Astrophys. 70, 467 (1978).
7) D. N. Schramm and G. Steigman, Ap. J. 243, 1 (1981).
8) C. F. Lillie, in "The Scientific Results from the OAO-2," A. D. Code, Ed., 583; NASA: Washington, D.C. (1972).
9) E. Pitz, C. Leinert, A. Schulz, and H. Link, Astron. Astrophys. 72, 92 (1979).
10) R. R. Dube, W. C. Wickes, and D. T. Wilkinson, Ap. J. (Lett.) 215, L51 (1977).
11) A. F. Davidsen, S. Bowyer, and M. Lampton, Nature 247, 513 (1974).
12) F. Paresce and P. Jakobsen, Nature 288, 119 (1980).
13) R. C. Henry, Tenth Texas Symposium on Relativistic Astrophysics, in press.
14) F. Paresce, C. F. McKee, and S. Bowyer, Ap. J. 240, 387 (1980).
15) F. Paresce, B. Margon, S. Bowyer, and M. Lampton, Ap. J. 230, 304 (1979).
16) C. Jamar, D. Macau-Hercot, A. Monfils, G. I. Tompson, L. Houziaux, and R. Wilson, *Ultraviolet Bright-Star Spectrophotometric Catalog* (Paris: ESO) (1976).
17) R. C. Henry, R. Anderson, P. D. Feldman, and W. G. Fastie, Ap. J. 222, 902 (1978).
18) R. C. Henry, Ap. J. (Lett.) 244, L69 (1981).
19) P. M. Gondhalekar, A. P. Phillips, and R. Wilson, Astron. Astrophys. 85, 272 (1980).
20) A. N. Witt, Publ. Astron. Soc. Pacific 89, 750 (1977-78).
21) R. C. Anderson, R. C. Henry, and W. G. Fastie, Ap. J. (submitted).
22) R. C. Anderson, Ph.D. thesis, Johns Hopkins University (1979).
23) W. G. Fastie, The Moon 7, 49 (1973).
24) R. C. Henry, R. C. Anderson, and W. G. Fastie, Ap. J. 239, 859 (1980).
25) C. Heiles, Astron. Astrophys. Suppl. 20, 37 (1975).
26) C. Heiles, Ap. J. 204, 379 (1976).
27) D. Burstein and C. Heiles, Ap. J. 225, 40 (1978).

PRIMORDIAL NUCLEOSYNTHESIS

Jean Audouze[1]
Institut d'Astrophysique du CNRS
Paris - France
[1]also at Laboratoire René Bernas
Orsay - France

A quick summary of the primordial nucleosynthesis occuring during the early phases of the Universe (very often referred to as the Big Bang) is provided. The observed abundances of the light elements D, ^{3}He, ^{4}He and ^{7}Li such as the processes responsible for their formation are recalled. D and ^{7}Li can be used to probe the present density of the Universe and its dynamical properties on very large scales while the ^{4}He abundance is related to the lepton density and the rate of expansion of the Universe. Moreover the influence of the very hypothetical mass of the neutrinos on the early evolution of the Universe is mentionned.

I. INTRODUCTION

The purpose of this talk is to review very briefly current thoughts and works about the nucleosynthetic processes which should have occured during the early phases of the Universe. For more than ten years now, it is generally assumed and accepted that the lightest nuclear species like D, ^{3}He, ^{4}He and ^{7}Li are synthetized at the end of the so called Big Bang : This unique explosive and primordial eventhas induced the observed expansion of the Universe and originated the relic blackbody radiation at 2.7 K.

The analysis of the primordial nucleosynthesis processes is indeed intimately related to general astrophysical effects or consequences such as the evolution of the dilatation of the Universe (is the Universe "open", i.e. expanding for ever or can it be "closed", i.e. able to experience successive phases of expansion and contractions ?) connected itself to the value of its present density. This specific nucleosynthesis can also provide invaluable information on many aspects of the nature of the elementary particles and some of the physical laws which govern them. This paper and other contributions in these proceedings clearly establish the close relation between cosmology and elementary particle physics The recent and important progresses presently achieved in one of these fields clearly influence and are beneficial to the other.

After a short summary in Section II of the presently observed abundances of the relevant elements (D, ^{3}He, ^{4}He and ^{7}Li) the main properties of the classical (or so called canonical) Big Bang models and the characteristics of the nucleosynthetic processes which occur during the early phases are recalled in Section III. Section IV is devoted to the consequences of these nucleosynthetic aspects on the evolution of the expansion of the Universe, the present density of it and same properties of the elementary particle physics such as the number of existing neutrinos, their possible mass, etc... Section V contains our present conclusions concerning these relations existing between the nucleosynthesis, the cosmology and the elementary particle physics.

II. THE OBSERVED ABUNDANCES OF THE LIGHTEST ELEMENTS

The relevant elements are Deuterium (D) Helium 3 and 4 (^{3}He and ^{4}He) and Lithium 7 (^{7}Li). Among the many recent reviews which provide some informations on their abundances and more references the reader may consult references 1, 2 and 3.

1) Deuterium : The presently adopted D abundance is $D/H = 2 \pm 1\ 10^{-5}$ although with still very large uncertainties on such values. This abundance range is based

on meteoritical, solar wind, Jupiter and interstellar measurements. The recent determinations of the interstellar D/H ratio in the solar neighbourhood clearly show a quite large spread in this abundance according to the different lines of sight of O and B stars used for such searches. The nearby interstellar D abundance might range from D/H $\sim 2\ 10^{-6}$ up to D/H $\sim 2\ 10^{-4}$ [4]. Nevertheless various physical effects may lead to such a spread (radiation pressure affecting specifically the D atoms and chemical fractionation induced by the formation and destruction processes of DH in molecular clouds[5]). Reference 5 concludes that a proper account of these effects should restrict the possible variations of the nearby interstellar D/H within the range quoted above.

2) Helium 3 : There are still many uncertainties on the ^3He abundance. Wood et al.[6] dare only to quote an upper limit $^3He/H \leq 5\ 10^{-5}$ from their very careful search of the interstellar ^3He abundance. There are also some uncertainties on the Solar System ^3He abundance due to the transformation of the Solar System D into it. Keeping in mind these large uncertainties one can consider that the observed $^3He/H$ ratio should range from 1 to $3\ 10^{-5}$.

3) Hélium 4 : The reader is referred to Kunth (these proceedings) for a detailed analysis of the ^4He abundance deduced from the He observations from galaxies with low metallicity and high gas content (these galaxies are often referred to as blue compact galaxies or "lazy" galaxies [7]). According to this author and contrary to previous works on this subject a significant and clearcut correlation between the helium and meal abundance does not seem to exist anymore : The primordial He abundance (by mass) is equal to $Y = 0.243 \pm 0.010$ while the Solar System He abundance is about 0.27 ± 0.03.

4) Lithium 7 : The ^7Li abundance can be evaluated from carbonaceous chondrites determinations ($^7Li = 2.2 \pm 0.4\ 10^{-9}$), chondrites ($^7Li = 1.9 \pm 0.8\ 10^{-9}$), the Sun ($^7Li = 1.0 \pm 0.3\ 10^{-11}$)[1], different stars ($^7Li$ 10^{-9}) and the interstellar medium ($^7Li \sim 5\ 10^{-10}$. One adopts finally an overall $^7Li/H$ abundance of about 10^{-9}.

Finally the relevant light element abundances are :

$$\left(\frac{D}{H}\right) = \left(\frac{^3He}{H}\right) = 2 \pm 1\ 10^{-5} \left(\frac{^4He}{H}\right) = 0.08 \pm 0.01 \text{ and } \frac{^7Li}{H} = (1 \pm 0.5)10^{-9}$$

III. THE "CANONICAL" BIG BANG AND ITS NUCLEOSYNTHESIS

Although it is now fairly well established that the observed Universe is

---

1 : One should remember that in this case the ^7Li abundance is especially low because ^7Li is destroyed by thermonuclear reactions at the bottom of the external convective zone.

born from a singular very hot and dense phase, the physical conditions ruling this phase are still disputable. In cosmology one is used to call "canonical" or "standard" Big Bang the model in which the following quite simple and conservative assumptions are made like any other Big Bang model. The canonical or standard Big Bang model is constructed by assuming that the equivalence principle -which indicates that the physical laws are totally invariant, i.e. do not vary with the location and with the time- holds and that the primordial temperature has been much higher than 10^{11}K to insure the disruption of all possible nuclei and the equilibrium of weak interactions. Furthermore in the Standard Model one assumes also that :

1) the early Universe was homogeneous and isotropic. This assumption is often called the Cosmological Principle ;
2) the gravitational interactions are well described by the Einstein Relativity theory : the recent discovery of a double pulsar system such as the estimates of relativistic effects on the propagation of radar signals reflecting on the Venus or Mars surface, etc... provide very strong support in favour of this gravitation theory ;
3) the Universe is asymetric, i.e. the baryon number is positive or the amount of antimatter existing in the Universe is negligible in comparison with the amount of matter. This point is quite well established now and is a direct consequence of the Grand Unification Theories mentioned in many other chapters of the book (see the contributions of J. Ellis and D.N. Schramm) ;
4) the leptonic number -i.e. the number defined as $L = n(e^-) + n(\mu^-) - n(e^+) - n(\mu^+) + n(\nu_\mu) - n(\bar{\nu}_\mu) + n(\nu_e) - n(\bar{\nu}_e)$, where the former terms represent respectively the densities of electrons, negative muons, positrons, positive muons, muonic neutrinos, muonic antineutrinos, electronic neutrinos, electronic antineutrinos- is much smaller than the photon density ;
5) there was no unknown elementary particle during the early phases of the Universe.

The two first hypothesis are fairly well established at least in first approximation as well as the predominance of the matter over the antimatter in the observed Universe. As we will see in the next section the two last hypothesis are far less settled : There is a growing evidence in favour of very large neutrino densities which might help in solving some problems generated by the large scale dynamics of the Universe. Moreover the Grand Unification Theories (GUT) which are currently built both to unify the electromagnetic, weak and strong interactions and to offer some description of the very early phases ($t \ll 10^{-3}$ sec) of the Universe invoke some elementary particles which can hardly be considered as known in the sense of hypothesis (5) : the particle "zoo" of the GUT include for instance

Higgs bosons, gluons and many different quarks which can be unified with the leptons during these very early phases and from which the hadrons(pions,nucleons...) can be formed [10]. Therefore the Canonical Big Bang can be more and more refined by taking into account the recent progresses in elementary particle physics (and also on the observation of the large scale structure of the Universe).

As a consequence of hypothesis (1) -i.e. the cosmological principle- the metric describing the evolution of the Universe is the one of Robertson-Walker :

$$ds^2 = -dt^2 + R^2(t)\left[\frac{du^2}{1-kv^2} + v^2(d\Theta^2 + \sin^2\Theta d\rho^2)\right] \qquad (1)$$

where k is the curvature constant : the Universe is expanding for ever (open) for k = 0 and +1 and closed for k = -1. R(t) is the scale factor of the Universe. The radial (dimensionless) coordinate U and the angular coordinate Θ and ρ are the geometrical variables of this metric. This metric takes into account the isotropy of the Universe.

From the General Relativity the expansion rate is given by :

$$\frac{1}{V}\frac{dV}{dt} = \frac{3}{R}\frac{dR}{dt} = \sqrt{24\pi\ G\rho} \qquad (2)$$

where V represent a volum element. G is the gravitational constant and ρ the total density of the Universe. As we will see in the sequel the density concerns not only the nucleons but also many other particles (especially the neutrinos).

To account for an expansion slower or quicker than the one determined by this relation one can write the expansion rate as[11] :

$$\frac{1}{V}\frac{dV}{dt} = \xi\sqrt{24\pi\ G\rho}$$

where ξ = 1, for the canonical Big Bang, is < 1 for a slow expansion and > 1 for a rapid one.

The nucleosynthesis proceeds when neutrons and protons can combine to form deuterium which itself transforms in part into ^{3}He and ^{4}He through the D(p, γ)^{3}He D(D, n)^{3}He , D(D, p)T reactions. This combination between neutrons and protons occurs when the temperature of the Universe drops below kT = 0.1 MeV. In the range 0.1 < kT < 1 MeV, neutrons and protons are in equilibrium thanks to the interactions

$$n + \nu_e \rightleftharpoons \bar{e} + p$$

$$\text{and} \quad n + e^+ \rightleftharpoons \bar{\nu}_e + p$$

when the temperature is too low for these interactions to continue to proceed the n/p ratio freeze out such that

$$\frac{X_n}{X_p} = \exp - \frac{(M_n - M_p)}{kT} \qquad (4)$$

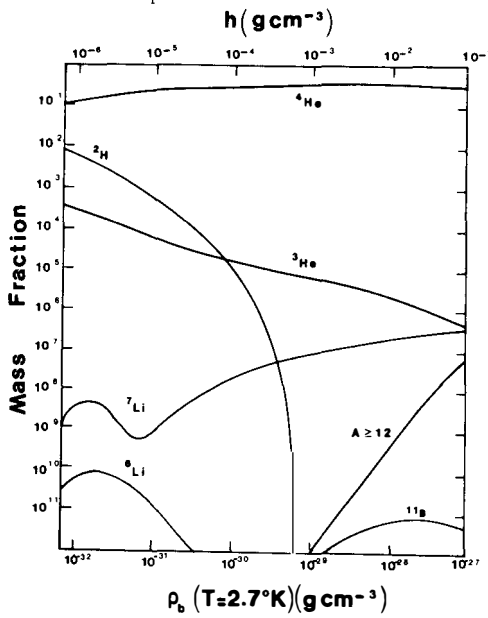


FIGURE 1 - Abundance resulting from the primordial nucleosynthesis in the frame of the canonical Big Bang[12] as a function of the present density of the Univer se.

Figure 1 shows the classical results of the Big Bang nucleosynthesis obtained in the frame of this simple model [12]. One notices the strong dependance of the D, ^{3}He and ^{7}Li abundance with the present density of the Universe in contrast with the ^{4}He abundance which is not very sensitive to this parameter. In fact :

$$X_n \sim 2 \frac{X_n/X_p}{1+(X_n/X_p)} \qquad (5)$$

For instance it is well known that too high densities for the present Universe lead to too low D abundances.

IV. ASTROPHYSICAL CONSEQUENCES

From these results one realizes that there are two different consequences : one related to the present density of the Universe and the other related to the conditions ruling the neutron-proton equilibrium.

The first consequence, i.e. the determination of an upper limit for the present density of the Universe is now very classical [1,13] : From Fig.1, one sees

that X(D) becomes much lower than 10^{-5} if $\rho_o > 10^{-30}$ g cm^{-3} which is about six times lower than the critical value $\rho_c = \frac{3H_o^2}{8G} \sim 5.7\ 10^{-30}\left(\frac{H_o}{55}\right)^2$ g cm^{-3}, where Ho, the Hubble constant, is expressed in km s^{-1} Mpc^{-1} which delineates the border between the open Universe expanding for ever ($\rho_{present} < \rho_c$) and the closed pulsating Universe ($\rho > \rho_c$). At this point I would like to stress the interesting proposal made by Austin and King[14] who pointed out that the ^{7}Li abundance can set an upper limit to the present density of the Universe as stringent as those set by the use of the D abundance. According to these authors the present density of the Universe should be $\sim 9 \pm 4\ 10^{-31}$ g cm^{-3}.

In fact, nuclear cosmologists seem to be more interested now by the implications of the He abundance which is sensitive to two related parameters :

1) the speed of the expansion of the Universe ;

2) the presence or the absence of new families of leptons...

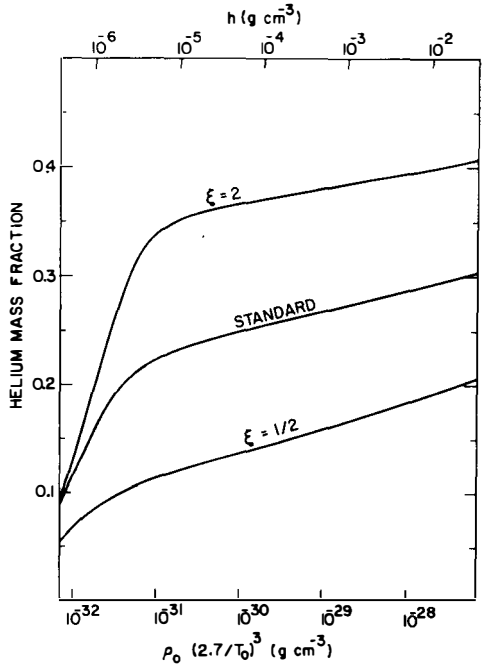


FIGURE 2 - The He abundance produced in models where the expansion rate factor (canonical Big Bang) ξ = 2, 1 and 1/2. One can notice the large influence of this parameter on this primordial abundance[12].

As it has been seen above this abundance depends on the neutron-proton ratio reached just after the freeze-out temperature. This ratio itself depends on the actual mass of the neutron M_n and on the freeze-out temperature : When the expansion rate is fast ($\xi > 1$) the freeze-out temperature remains high which leads to a high n/p ratio and therefore a high He abundance (Fig.2).

This effect of changing the expansion rate might be due not only to the presence of new families of neutrinos and/or leptons as seen below but also to inhomogeneities and/or anisotropies which might affect in some directions the gravitational effects on the overall dilatation of the Universe. The choice of another gravity theory might also influence the expansion of the Universe and therefore modify the value of the parameter ξ. The influence of the value of ξ on the primordial abundance of He has been expressed [15] as

$$Y = 0.333 + 0.02 \log h + 0.380 \log \xi \qquad (6)$$

where h is the so called baryon desity parameter defined as $h = T_9^3/\rho$ which is a constant since the expansion is adiabatic.

The expansion time scale which influence directly the primordial value of Y is very sensitive to the number of existing leptons. Up to very recently the only

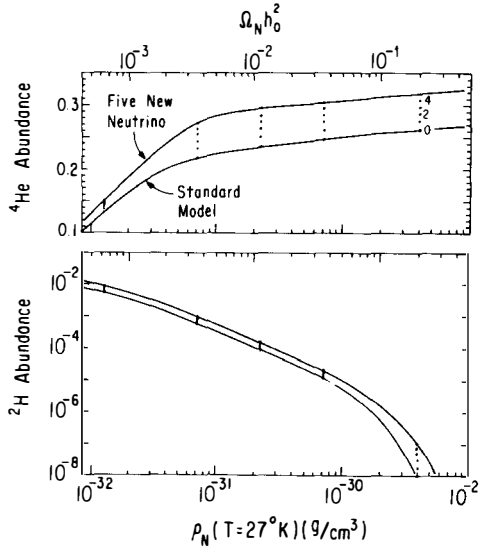


FIGURE 3 - Primordial abundances of ^{4}He and D calculated with 0 to 5 new families of leptons (to be added to the electrons and to the muons[15]).
With the results reported by Kunth in this meeting, there are only two other families of leptons which can be added to them (assuming in this case that there are no invisible metter inside the Universe which becomes less and less likely).

two observed lepton families were those of electrons and muons. A third family is that of the heavy tau lepton. Each family corresponds to a specific class of neutrinos. As shown by Yang et al. [15] and by David and Reeves [16] an increase of the number of existing leptons (therefore on existing neutrinos) increase the toal nergy density and therefore speeds up the expansion of the Universe and then leads to higher values of the primordial He abundance.

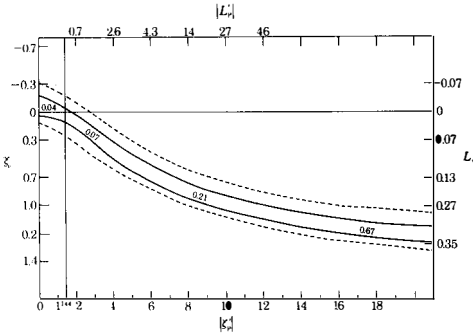


FIGURE 4 - Compatibility regions between the observed abundances and the leptonic numbers as defined by David and Reeves[16].
This diagram clearly shows that when the number of different families of leptons is allowed to increase the present density of the Universe (noted on this diagram) by the value of the cosmological parameter Ω (equal to two times the deceleration parameter q) should also be larger to make the Big Bang nucleosynthesis compatible with the observations.

In references 15 and 16 (as shown in Fig. 3 and 4), the Chicago and Saclay groups express similar conclusions according which :

a) The observed primordial He abundance can set significant constraints on the number of existing types of neutrinos and on the departure from the General Reltivity. As seen from Fig. 3, the calculations reported by Yang et al. show that the primordial He abundance determined by Kunth (this conference) seem to fix a stringent limit (less than four types of neutrinos) on the number of different families of leptons.

b) There exists a correlation between the present density of the Universe and the number of different neutrinos. From reference 16 if this number is equal to 3, it corresponds to a very open Universe ($\Omega \sim 0.05$) ; if there are 10 different neutrinos $\Omega \sim 0.08$. More than 5000 different neutrinos are needed to make the corresponding $\Omega > 0.8$.

V. SUMMARY AND CONCLUSION

This short review has the only ambition to call again the attention of the astrophysicists and the elementary particle physicists on the strong connection between the cosmology of the early Universe and some aspects of the physics of

particles. Many other contributions including those of Cowsik, Kunth, Nanopoulos and Schramm clearly show this very exciting connection.

I would like to end up this review by reminding the reader that the D and ^7Li abundances can be used to probe the present density of the Universe. Schramm (this conference) offers the exciting suggestion according which the ^3He abundance provides as lower limit ont the baryon/photon ratio $\frac{n_B}{n_\gamma} > 2\ 10^{-10}$. I would like finally to refer those who are interested bu the influence of the primordial He abundance to the contribution of Kunth and to the work of Olive et al.[17] . According to this last work which provide a clear account on this problem, they concur with the conclusion reached by many previous authors according whom the observations of the light element abundances imply that the baryons cannot close by themselves the Universe. Nevertheless if the neutrinos are found to be massive they might provide the bulk of the mass of the Universe. We may then live in a Universe where the matter is dominated by the massive neutrinos and which could be closed without being in conflict with the observations.

I would like to thank Ms Sylvie Corbin for her patience and her help in the production of this paper.

REFERENCES

1) Austin,S.A., 1981, The creation of the light elements cosmic rays and cosmology. To appear in Progress in Particle and Nuclear Physics
2) Audouze,J., 1981, Nuclear Astrophysics, ed. D. Wilkinson, Pergamon Press, p.125
3) Audouze,J., Reeves,H., Essays in Nuclear Astrophysics, ed. C.A. Barnes, D.D. Clayton and D.N. Schramm, University of Cambridge Press, to be published
4) Laurent,C., Vidal-Madjar,A., York,D.C., 1979, Ap. J., 229,923
5) Bruston,P., Audouze,J., Vidal-Madjar,A., Laurent,C., 1981, AP. J., 243,161
6) Rood,R.T., Wilson, T.L., Steigman,G., 1979, Ap. J. Lett., 227,L97
7) Kunth,D., 1981, Thèse de Doctorat ès Sciences Physiques, Université de Paris VII
8) Lequeux,J., Peimbert,M., Rayo,J.F., Serrano,A., Torres-Peimbert,S., 1979, Astron. Astrophys., 80,155
9) Steigman G., 1980, Physical Cosmology, ed. R. Balian, J. Audouze and D.N. Schramm, North Holland, p.473
10) Nanopoulos,D., 1981, Nuclear Astrophysics, ed. D. Wilkinson, Pergamon Press
11) Wagner,R.V., 1980, Physical Cosmology, ed. R. Balian, J. Audouze and D.N. Schramm, North Holland, p.398
12) Wagoner,R.V., 1973, Ap. J., 179,343
13) Reeves,H., Audouze,J., Fowler,W.A., Schramm,D.N., 1973, Ap. J., 179,909
14) Austin,S.M., King,C.H., 1977, Nature, 269,782
15) Yang,J., Schramm,D.N., Steigman,G., Rood,R.T., 1979, Ap. J., 227,697
16) David,Y., Reeves,H., 1980, Physical Cosmology, ed. R. Balian, J. Audouze andD.N. Schramm, North Holland, p.443
17) Olive,K.A., Schramm,D.N., Steigman,G., Turner,M.S., Yang,J., 1981, Ap. J., in press

PRIMORDIAL HELIUM AND EMISSION-LINE GALAXIES

D. Kunth
Institut d'Astrophysique de Paris
98 bis, Bd Arago
75014 PARIS

Recent observations of 13 low luminosity emission-line galaxies give no convincing evidence for a ΔY versus ΔZ correlation. The observed scatter is likely to result from uncontrolled parameters in the analysis rather than simple observational errors.
The results taken with the best available data in the literature suggest that $Y_p = 0.240 \pm 0.015$.
If we admit that baryons provide most of the mass which binds binary and small groups of galaxies, this mass fraction Y_p constraints the number of 2 component neutrino species to $N_\nu \leq 4$ and remove the contradictions raised by previous low Y_p determinations.

I. INTRODUCTION

The necessity to advocate a primordial helium origin has been stressed a long time ago since stellar nucleosynthesis would only account for a small fraction of the presently observed helium abundance [1].

Indeed, given the luminosity and the age of our Galaxy to be $L = 4.10^{43}$erg sec^{-1} and $t = 3.10^{17}$sec one immediately sees that with an energy production per gram of 6.10^{18}erg g^{-1} in the $H \rightarrow He$ conversion one would only obtain 2.10^{42} grams of helium.This corresponds to about $10^{9}M_{\odot}$ whereas the mass of our Galaxy is roughly $10^{11}M_{\odot}$ so that $Y \backsim 0.01$.

The observed helium abundance (by mass) is fairly uniform and comes to be $Y \backsim 0.25 \div 0.30$ in strong disagreement with the previous figure. Several explanations have been given to account for the high Y value and spatial variations :

- Galaxies may have been more luminous in the past hence a correlation between Y, Z (metals) and the luminosity would be expected [2].

- The existence of primordial temperature variations at early epoch in a Friedman Universe would account for the observed spatial variations [3].

This evocation is by no means exhaustive, and I shall admit the possibility that most of ^{4}He was produced during the early phase of an hot and dense Universe and that subsequent evolution is responsible for local variations.

The calculations of such a model were successfully performed and predicted a primordial ^{4}He abundance of 0.250 which depends little on the baryonic density but critically on the expansion rate of the Universe [4].

This later fact is becoming crucial since the discovery of the new τ lepton implying the existence of a new neutrino flavor. Adding more relativistic neutrinos to the early Universe increases the energy density hence speeds up the expansion rate and favors the ^{4}He production.

The knowledge of the amount of ^{4}He which follows from primordial nucleosynthesis can set severe constraints up in the Big Bang model or the number of neutrino flavors to be discovered.

II. THE IDEAL SITES

The helium abundance can be obtained by direct observations or indirectly in many astrophysical objects -see 5) & 6) for a review-. The problem one faces is to ascribe which net fraction is due to subsequent evolution of the site or the object under study. An ideal choice would be one for which the primordial gas has been almost unaltered.

The possible astrophysical importance of unevolved galaxies with metal-poor interstellar gas such as IZw18 and IIZw40 has been pointed out by Searle and Sargent [7]. Ever since, an increasing number of such galaxies have been discovered; They represent an homogeneous class of objects with the following properties :

$$\text{dwarf} \begin{cases} M \sim 10^9 M_\odot \\ L \sim -14 \quad -17 \end{cases}$$

$$\text{hydrogen to total mass ratio} : m_H/m_{Tot} \geq 0.1$$

$$\text{metal poor} : Z \sim Z_\odot/5 \div Z_\odot/30$$

Moreover their morphology and blue color (B-V$\sim$ 0.0, U-B$\sim$ -0.6) result from a recent burst of star formation ionizing a large fraction of the interstellar gas and producing a typical H II region like spectrum out of which a direct abundance determination of He, O, N, Ne and S can be made.

These objects are ideal for the chemical evolution models because of their high m_H/m_{Tot} ratio and the absence of dynamical structure (no spiral arms) and low mass (no accretion expected). Under such conditions the so called "instant recycling approximation" applies i.e. one can neglect the lifetime of the stars with respect to the evolution of the gas density and one expects a relation between Y and Z of the form : $Y = \alpha Z + Y_p$

where α is the proportionality coefficient between the corresponding galactic enrichments ΔY and ΔZ.

III. SEARCH FOR DWARF EMISSION-LINE GALAXIES

A search for dwarf emission-line galaxies was undertaken by this author in collaboration with W.L.W. Sargent. The aim of the search was to establish a list of objects for further detailed observations and find galaxies with more extreme properties.

In fact the later requirement was not fullfilled and among more than 50 newly discovered galaxies, no one did show any dramatic He or Z underabundance (i.e. $Y < 0.2$ and $Z < Z_\odot/30$).

Two lists were established : one from a survey among the Zwicky's lists of blue compact galaxies[8] and a recent one from candidates selected from two objective prism plates[9].

These lists were the starting point for the more detailed observations discussed below.

IV. OBSERVATIONS

1) The sample

We decided to restrict our sample to low luminosity and metal-poor objects in order to investigate the lower Z part of the (Y, Z) diagram where in principle the relation $Y = \alpha Z + Y_p$ still holds.

Moreover, objects with $Z \leqslant Z_\odot/5$ have also an observational advantage, namely the cooling of the gas by the oxygen species is less effective and results in increasing the gas temperature so that the temperature-indicative [O III] 4363 line is easily measured.

We also chose objects with an high line to continuum ratio to avoid the effect of a strong stellar contamination in the emission lines.
These requirements led us to select objects discovered from the objective-prism plates as it appeared that the objective-prism technique is highly selective towards forbidden lines whereas in the Compact Galaxies sample we found objects with more stellar background and less recent active star formation. Fig. 1 shows the [O III]/Hß ratio in both samples : it can be seen that O.P. objects exhibit an higher mean [O III]/Hß value.

2) Spectroscopy

The objects were observed at Las Campanas (Chile) with the 2.5 m telescope. We used a Boller and Chivens Spectrograph coupled with an Intensified Reticon. The dispersion 120 Å/mm resulted in an effective resolution of 3 Å after event centering and a spectral range $\lambda\lambda$3500 Å - 6800 Å. The exposure time was chosen as to get more than 5 % accuracy on faint lines such as [O III] 4363 and He I 4471. Table 1 shows the useful measured lines from wich one is able to extract relevant informations :

TABLE 1 :

line, λ	derived quantities
[O II] 3727	$n(O^+)$
[O III] 4363	Te
[O III] 4959, 5007	$n(O^{++})$
[N II] 6548, 6584	$n(N^+)$
[S II] 6717, 6730	n_e, $n(S^+)$
He I 4471, 5876	$n(He^+)$
He II 4686	$n(He^{++})$
H 12 3750 H 11 3771 H 10 3798 H 9 3835 H δ 4101 H γ 4340 H ß 4861 H α 6563	C(Hß)

The reduction procedure is very classical and allows one to analyse a final sky-subtracted spectrum removed from all kinds of observational and intrumental effects (atmospheric extinction, instrumental response...).
The relative line strengths result from the observed line ratios by applying a

correction for effects of interstellar reddening. For this purpose the reddening constant C(Hß) was deduced from the observed decrement of the Balmer lines H12 : H11 : H10 : H 9 : Hδ : Hγ : Hß. Underlying stellar absorption do affect the Balmer lines of higher serie and was accounted for.

The electronic temperature Te and the gas density n_e were obtained from the [O III] 4363 and the [S II] 6717-6780 lines.

The ionic abundances are derived from :

$$\frac{n(X^i)}{n(H^+)} = \frac{\varepsilon(H\beta)}{\varepsilon(\lambda,X^i)} \frac{I(\lambda,X^i)}{(H\beta)}$$

where X^i is the i^{th} ionization for the element X, and the ε are the emissivities. Similar expresssions hold for He^+ and He^{++}, the best measured HeI lines are λ4471 and λ5876 and we note the corresponding fractional He^+ abundances by $y^+(4471)$ and $y^+(5876)$.

To obtain the total abundances one must correct for the unseen ionization stages. A classical ionization correction scheme makes use of the coïncidences in the ionization potentials of several ionic species. This scheme follows the ionization distribution of oxygen as a guide to the ionization distribution of the other elements.

For helium the total abundance is :

$$\frac{n(He)}{n(H)} = \frac{n(He^o) + n(He^+) + n(He^{++})}{n(H^+)}$$

Since He^o is not observed a correction factor band on ionization models gives

$$i_{cf} = [1 - 0.25\ n(O^+)/n(O)]^{-1}$$

$$\text{and : } \frac{n(He^o + He^+)}{n(H^+)} = i_{cf} \frac{n(He^+)}{n(H^+)}$$

This correction is nevertheless uncertain[11] and depends critically on the electronic densities and the ionizing spectra in the objects under study. Fortunately in all our objects we have $n(O^+)/n(O) < 0.3$ and the correction factor is small, i.e. $i_{cf} \leq 1.09$.

Difficulties however are numerous and we checked against the following systematic effects :

a - Underlying stellar absorption :

We found that stellar absorption could amount 1 Å in the Balmer emission lines. Such a similar effect is expected in the He lines, the contribution from O stars would at most amount 0.6 Å and 1 Å in the λ5876 and λ4471 lines. However we checked that $y^+(4471)$ and $y^+(5876)$ did not correlate with the corresponding equivalent width W(λ4471) and W(λ5876).

b - Any systematic error in the reddening correction would differently affect the quantities $y^+(4471)$ and $y^+(5876)$. We found that no trend subsisted between the reddening constant C(Hß) and the difference $y^+(4471) - y^+(5876)$.

c - Cox and Daltabuit[12] pointed out a possible contribution due to collisional excitation and self-absorption from the metastable level 2^3S. If such effects are important the value $y^+(5876)$ could exceed that deduced from the He I 4471 line by 20 % in contradiction with our present study. Moreover the electronic densities are always smaller than 300 cm^{-3} and the opacity in the He I 3888 line is very small in our sample ($\tau < 10$).

VI. RESULTS

A recent study of 9 irregular and compact galaxies by Lequeux et al.[10] produced a positive correlation between Y and Z namely :

$$Y = (0.233 \pm 0.005) + (1.73 \pm 0.90)Z$$

These results exclude the Orion Nebula.

Other authors[13,14] have quoted from similar type of study, values as low as :

$$Y_p = 0.216$$

We have plotted in Fig.2 our results together with Lequeux et al. data. The error on Y in our data is estimated between 0.008 and 0.016 and is about 15 % for the oxygen abundance. A least square fit following the Fletcher and Powell[15] minimization procedure gives :

$$Y = (0.240 \pm 0.007) + (33 \pm 53) \frac{n(O)}{n(H)}$$

and the correlation coefficient is only r = 0.164.

Assuming that O constitutes 45 % of Z by mass this corresponds to :

$$Y = (0.240 \pm 0.007) \pm (1.25 \pm 2.0)Z$$

The same procedure applied to our data alone gives :

$$Y = (0.243 \pm 0.010) \pm (0.7 \pm 3.5)Z$$

We conclude that :

a) On the basis of our data and Lequeux et al. results, there is no convincing case for a well defined $\Delta Y/\Delta Z$ correlation.

b) The data suggest that the average Y value for low-luminosity galaxies is Y = 0.245 very close to the above Y_p = 0.240 deduced from the linear fit. Taken at face value it implies that only an upper limit for Y_p can be given which is :

$$Y_p \lesssim 0.245$$

c) Most of the scatter is rather due to uncontrolled parameters in the analysis and over simplifications in the models than to observational errors.

The larger uncertainties involve the reddening correction, the helium underlying absorption and the correction for neutral helium.

An example well illustrates this statement : II Zw40 is a best observed case by four independant observers[10,14,16,17]. From these measurements one gets :

$$Y(\text{II Zw40}) = 0.232 \pm 0.004$$

This 2 % internal accuracy is quite illusory. Indeed the ratio $I(\lambda 4471)/I(\lambda 5876)$ after reddening correction is found to be 2.0 by all observers whereas the theoretical ratio expected from the classical recombination theory is 2.76. Such a discrepancy has not yet found any satisfactory explanation.

VII. COSMOLOGICAL IMPLICATIONS

We shall admit that $Y_p \lesssim 0.245 \pm 0.010$ is a good upper limit on Big-Bang nucleosynthesis. If the energy density of the early Universe was dominated by the contribution of relativistic particles then the dynamical equations simplify and give : $\rho \propto t^{-2}$

where ρ stands for the energy density and t for the expansion time scale (age).

On the other hand, in a radiation-dominated Universe, for $T \gtrsim 10^{10}$°K, the particles which contribute to the total density (i.e. $mc^2 < kT$) are photons, electron-positron pairs, electron and muon neutrinos and the Universe is filled with a black-body spectrum, then :

$$\rho \propto gT^4$$

where g is a number depending on the number and types of relativistic particles included in the picture.

The radiation field generate electron-positron pairs which promote the following reactions :

$$n + e^+ \rightleftharpoons p + \bar{\nu}_e \qquad \text{reaction rate is} \sim \Gamma$$
$$p + e^- \rightleftharpoons n + \nu_e$$

and the spontaneous ß decay :

$$n \longleftrightarrow p + e^- + \bar{\nu}$$

at small t ($T > 10^{10}$K or $t < 1$ sec) the balance of these reactions is thermodynamic since the reaction rates $\Gamma > t^{-1}$ therefore the neutron to proton-ratio is maintained at the equilibrium value :

$$n/p = \exp(-\Delta Mc^2/kT)$$

As the expansion proceeds, the above reactions are too slow to maintain n and p to its equilibrium value and the ratio "freezes out" at :

$$(n/p)_f \# \exp(-\Delta Mc^2/kT_f)$$

where T_f is attained for $\Gamma/t \sim 1$

Typically $(n/p)_f$ is 1/6 and $T_f \approx 10^{10}$K.

Ultimately in the temperature range $T \sim 10^{10}$K to $T \sim 5\ 10^8$K an appreciable amount of deuterium accumulates and helium is produced at the end of the interval. It follows from the above picture that if one increases the total number of neutrino species[18] the resulting energy density will increase and the condition $\Gamma/t \sim 1$

occurs at a larger T_f temperature because the expansion rate is faster (since $\rho \propto t^{-2}$). As a result $(n/p)_f$ is larger and more neutrons end up in more ^4He.

At present 3 lepton types have been discovered namely the e, μ and recently the lepton[19].

From the results of the "canonical" Big Bang models the extra helium formed with the addition of ΔN_L leptons (a part from the familiar e and μ) is roughly[18] :

$$Y \approx 0.011 \Delta N_L$$

for a fixed value $\Omega_N h_o^2$ where Ω_N stands for the present mass density and h_o is the present Hubble parameter (in unit of 100 km/sec/Mpc).

The present galactic dynamics suggests that $\Omega_N h_o^2 \gtrsim 0.01$. If the mass which binds binary galaxies is primarily in the form of baryons then our result shows that at most 3 or 4 lepton types are permitted -see Fig.3- and does not run the "canonical" models into very severe difficulties. This statement is in contradiction with previous arguments since from similar studies but on much limited sample and the overoptimistic faith that $\Delta Y/\Delta Z = 3$ was well established, Y_p had been shown to be less than 22%. A recent study[20] also indicates that as many as 8 neutrino types are allowed if the binding mass is not baryonic (neutrinos of mass 10 eV). One open possibility remains however that active nucleosynthesis did occur at an early stage in the galactic or protogalactic era in the evolution of the Universe. See, for instance, in this Conference the discussion related to a population III stars and its connection to the observed deviation from a 2.7 K black-body spectrum. In this case a lower fraction of the $Y_p = 0.240$ value would be attributed to the primordial nucleosynthesis.

REFERENCES

1) Burbidge,G.R., 1958, Publ. Astron. Soc. Pacific, 70,83
2) Burbidge,E.M., Burbidge,G.R., H II regions and Related Topics, Mittlelberg, Germany, ed. T.L. Wilson and D. Downes, Lecture Notes in Physics (Springer, Berlin, 1975)
3) Silk,J., Shapiro,S.L., 1971, Ap. J.,166,249
4) Wagoner,R.V., 1973, Ap. J., 179,343
5) Danziger,I.J., 1970, Ann. Rev. Astro. Astrophys., 8,161
6) Greenstein,J.L., 1980, Physica Scripta, 21,759
7) Searle,L., Sargent,W.L.W., 1972, Astrophys. J., 173,25
8) Kunth,D., Sargent,W.L.W., 1979, Astron. Astrophys. Suppl., 36,259
9) Kunth,D., Sargent,W.L.W., Kowal,C., Astron. Astrophys. Suppl., in press
10) Lequeux,J., Peimbert,M., Rayo,J.F., Serrano,A., Torres-Peimbert,S., 1979, Astron. Astrophys., 80,155
11) Stasinska,G., 1980, Astron. Astrophys., 84,320
12) Cox,D.P., Daltabuit,E., 1971, Ap. J., 167,257
13) Peimbert,M., Torres-Peimbert,S., 1974, Ap. J., 193,327
14) French,H.B., 1980, Ap. J., 240,41
15) Fletcher,R., Powell,M.J.D., 1963, Computer Journal, 6,163
16) Kunth,D., Sargent,W.L.W., 1981, in preparation
17) Kinman,T.D., Davidson,K., 1981, Ap. J., 243,127
18) Yang,J., Schramm,D.N., Steigman,G., Rood,R.T., 1979, Ap. J., 227,697
19) Perl,M.L., et al., 1976, Physical Letters, 63B,466
20) Olive,K.A., Schramm,D.N., Steigman,G., Turner,M.S., Yang,J., 1981, Ap. J., in press

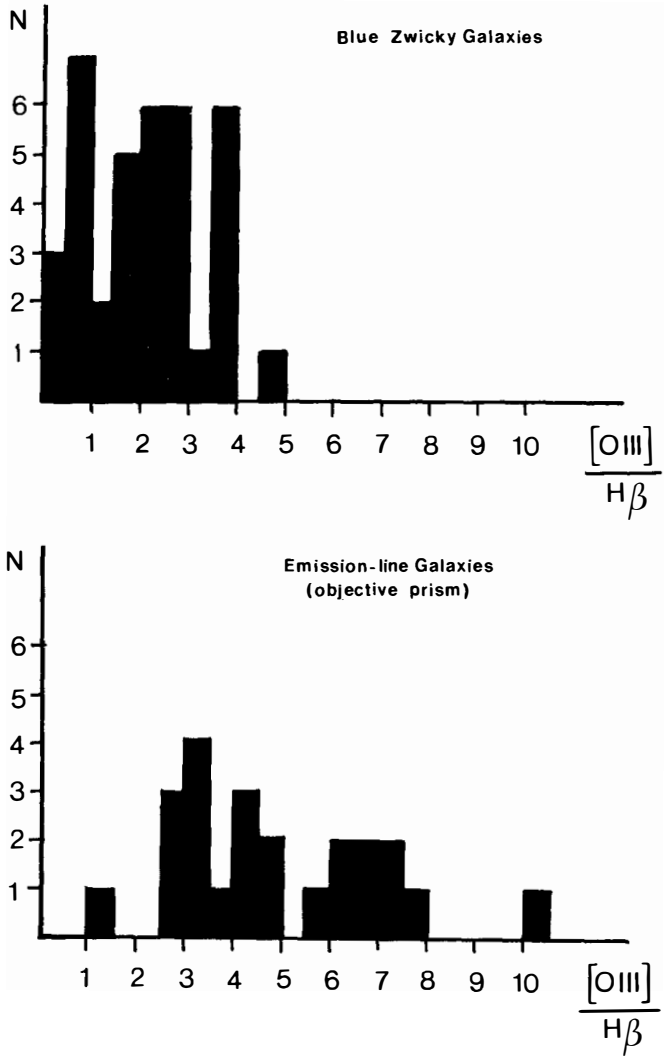


FIGURE 1 - Comparison of two emission-line galaxies samples.

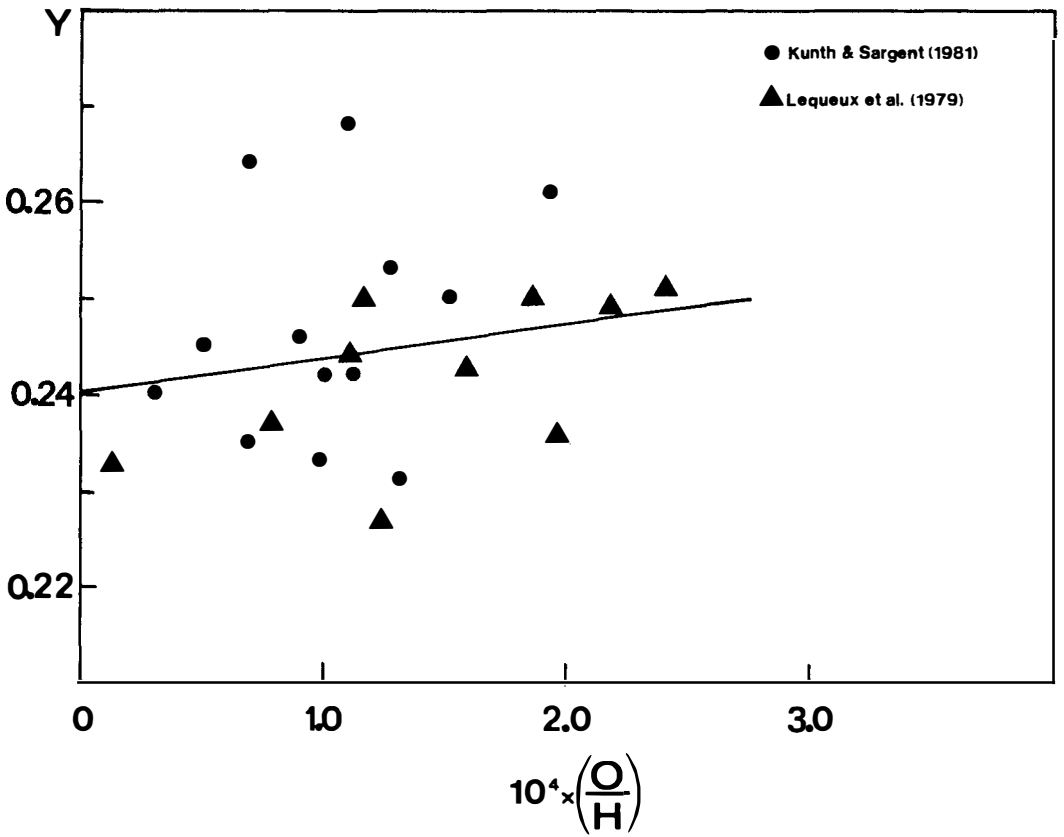


FIGURE 2 - Helium abundance versus Oxygen abundance for low-luminosity emission-line galaxies.

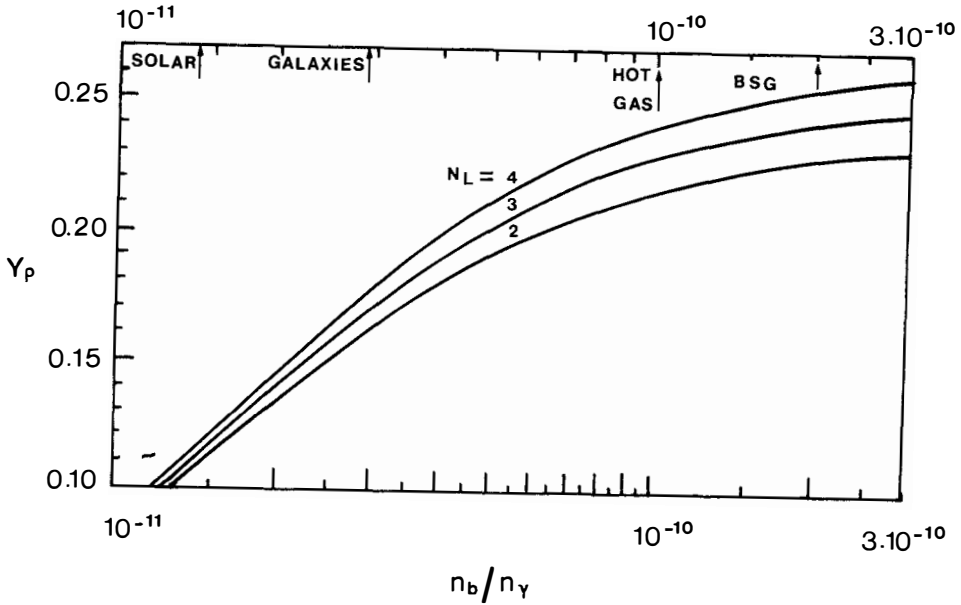


FIGURE 3 - The mass fraction of ^{4}He synthetized as a function of the baryon-to-photon ratio for N_L = 4, 3, 2 and the neutron half-time $\tau_{1/2}$ = 10.61 sec. BSG stands for binary galaxies and small groups of galaxies ($\Omega_N h_o^2 \geq 0.01$). This Figure is adapted from reference 20).

ANISOTROPY OF THE COSMIC MICROWAVE BACKGROUND RADIATION

Joseph Silk
Department of Astronomy
University of California
Berkeley, California 94720

Summary

Theoretical predictions of the angular anisotropy in the cosmic microwave background radiation on both small and large angular scales are described. The role of massive neutrinos is reviewed with regard to their effect both on the background radiation anisotropy and on the galaxy correlation function over very large scales. A brief comparison is made with recent observational data on the background radiation, ranging from angular scales of a few arc-minutes to the dipole and quadrupole components.

I. INTRODUCTION

The cosmic microwave background provides direct information on the structure of the Universe both over very large scales and at a considerable redshift that exceeds by far that of the most remote quasar. The precision of the measurements is remarkable, exceeding in accuracy by some three orders of magnitude the limits on isotropy available only from galaxy counts prior to the discovery of the background radiation in 1965.

Until very recently, only upper limits were available on possible anisotropy in the cosmic background radiation. The first evidence for anisotropy came with confirmation of a dipole moment.[1,2] A dipole contribution due to the motion of the sun and our galaxy is inevitable, and its detection provides conclusive evidence for the cosmological origin of the background radiation.

Higher order anisotropy is also expected. According to the gravitational instability theory of galaxy formation for the origin of large-scale structure from small inhomogeneities in the early Universe, angular fluctuations in the radiation are inevitable over a range of different scales. One can even hope to establish the presence of fluctuations whose scale exceeds that of the present particle horizon. The ultimate goal of the observer is to establish the nature of the initial fluctuation spectrum in the early Universe, from which the present large-scale structure evolved. By detection of fluctuations on the surface of last scattering of the radiation, when interaction with matter last occurred, one can hope to infer what may have been present initially, perhaps at the Planck epoch or as part of the structure of the initial singularity itself.

This review is organized as follows. In section II, a qualitative description is given of the generation of anisotropy. The results of more detailed computations are described in section III, and a final section com-

pares predictions and observations. This review is based on part of a summary talk given at the 10th Texas Symposium on Relativistic Astrophysics.[3]

II. QUALITATIVE DESCRIPTION

The linearized theory of gravitational instability in an expanding universe[4] indicates that after decoupling of matter and radiation at a redshift $z \sim 1000$, there are two modes for the rate of growth of the density contrast, $\delta\rho/\rho \propto t^{2/3}$ and t^{-1}. For simplicity, only the coupling of the dominant, growing mode is considered below. Prior to decoupling, density fluctuations may generally be adiabatic or isothermal. Adiabatic fluctuations are associated with finite curvature (and energy density) fluctuations in the very early Universe. Isothermal fluctuations imply a smooth radiation field in the early Universe prior to decoupling, with finite variations in the specific entropy, the fluctuations in density contrast becoming arbitrarily small as $t \to 0$.

The coupling between matter and radiation fluctuations occurs in several ways. First, in an adiabatic fluctuation, matter and radiation are coupled according to the adiabat $\rho \propto T^3$. Consequently,

$$(\delta T/T)_{t_s} = 1/3\ (\delta\rho/\rho)_{t_s}, \qquad (1)$$

where $\delta\rho/\rho$ is the fluctuation in the matter density. This equation assumes that the decoupling of matter and radiation is instantaneous. In fact, the residual level of ionization is sufficient to appreciably blur the effective surface of last scattering, which becomes much thicker for smaller-scale fluctuations. Only fluctuations on a scale containing $\gtrsim 10^{15}\ M_\odot$ approach the limit (1). Smaller scale primary adiabatic fluc tuations are effectively smoothed.

However, an important contribution on smaller scales does come from the gravitationally induced motions of fluctuations which scatter the radiation, and generate secondary temperature fluctuations of magnitude

$$(\delta T/T)_{t_s} \approx v/c, \tag{2}$$

where the velocity of the fluctuation relative to the expansion is of order

$$v \sim (\delta\rho/\rho)_{t_s}(\ell/t_s). \tag{3}$$

Here ℓ is the scale of the fluctuation and t_s is the cosmological epoch at last scattering. In fact, equation (2) yields the principal source of small angular scale anisotropy for isothermal fluctuations, and is also important for adiabatic fluctuations below scales of $\sim 10^{15}\ M_\odot$.

A third source of fluctuations can be thought of as essentially a gravitational redshift effect, due to radiation from a given direction being in effect emitted from a region of perturbed gravitational potential, and also being received in a region where large-scale density fluctuations contribute to the local potential. First computed by Sachs and Wolfe[5] for a flat background, the temperature fluctuations due to this effect amount to

$$(\delta T/T)_{t_s} \sim G\,\delta\rho\ell^2 \sim (\delta\rho/\rho)_{t_s}(\ell/ct_s)^2. \tag{4}$$

The Sachs-Wolfe fluctuations evidently are important on large scales ($\ell \gtrsim ct_s$). The particle horizon at last scattering (reshift z_s) subtends an angle

$$\sim (\Omega/z_s)^{1/2}\ \text{rad}, \tag{5}$$

and on larger angular scales (e.g. $\gtrsim 2^\circ$ if $z_s \sim 1000$), the Sachs-Wolfe fluctuations are especially significant.

In order to estimate the angular dependence of the various contributions to $\delta T/T$, it is necessary to assume some initial spectral form for $\delta\rho/\rho$. A particularly simple expression is a power-law form for the Fourier power spectrum:

$$|\delta_k|^2 \propto k^n,\ 4 > n > -3 \tag{6}$$

where $\delta_k = \int \delta\rho/\rho\ e^{-i\hat{k}\cdot\hat{x}}\, d^3x$ and the limits on n come from requiring convergence of the correlation function for fluctuations at small k ($n > -3$) and

from non-linear interactions that fix a minimal fluctuation level (and a maximum steepness, $n < 4$). One find that

$$\delta\rho/\rho \propto M^{-1/2-n/6}, \tag{7}$$

where M is the mass contained in a sphere of diameter $2\pi/k$. If this sphere also subtends an angle θ, then the three contributions to $\delta T/T$ that we have noted have an angular dependence (if $\Omega = 1$) of the form:

$$(\delta T/T)_{\text{adiabatic}} \propto \theta^{-3/2-n/2}, \tag{8}$$

$$(\delta T/T)_{\text{velocity}} \propto \theta^{-1/2-n/2}, \tag{9}$$

and

$$(\delta T/T)\text{Sachs-Wolfe} \propto \theta^{1/2-n/2}. \tag{10}$$

A complication that arises is that if the density fluctuation spectrum is of the power-law form (7) in the very early universe, it will no longer necessarily be a simple power-law after decoupling on scales less than or comparable to the horizon. Several scale-dependent effects occur: the damping of adiabatic fluctuations on scales below a critical mass ($\sim 10^{13}\ \Omega^{-5/4}\ M_\odot$ at decoupling), the continuous growth of scales above the Jeans mass immediately prior to decoupling ($\sim 10^{17}\ M_\odot$) relative to the acoustic oscillation of smaller scales, and amplification due to a velocity overshoot effect when the sound velocity drops abruptly at decoupling.[6]

The amplitude of the predicted fluctuations is readily estimated. One expects that at decoupling $(\delta\rho/\rho)_{t_s} \sim 10^{-3}\ \Omega^{-1}$, for the largest scale to have collapsed by the present epoch, about $10^{15}\ M_\odot$. The corresponding angular scale subtended by a sphere containing this mass at decoupling is $\Theta = 10(M/10^{15}\ M_\odot)^{1/3}$ arc-min, and the maximum temperature fluctuation expected on this scale is therefore of order $\delta T/T \sim 3 \times 10^{-4}\ \Omega^{-1}$. A more precise method of normalization utilizes the galaxy auto-correlation function and is described in the following section.

III. RESIDUAL RADIATION FLUCTUATIONS

A more detailed computation of the radiation anisotropy involves solving the linearized graviational field equations for the evolution of matter density perturbations, together with an equation for the radiation intensity. In the optically thick or thin limits, appropriate to either well before or well after decoupling, the linearized energy-momentum conservation equations suffice to describe the radiation field. The transition through decoupling, however, requires a more sophisticated approach, in which the linearized Boltzmann equation is used to describe the radiation field.[6,7] Collision terms involving Thomson scattering suffice to describe the physical interaction of matter and radiation.

Since the description of the evolution of matter density fluctuations is rather well-known[4], it need not be repeated here. It is useful however to write down a form of the linearized Boltzmann equation for the evolution of radiation temperature fluctuations $\Delta \equiv \delta T/T$, namely

$$\dot{\Delta} + \frac{\gamma^i}{a}\frac{\partial \Delta}{\partial x_i} = \frac{1}{2}\dot{h}_{ij}\gamma^i\gamma^j + n_e\,\sigma_T\,(\gamma^i V_i - \Delta + \ldots) \qquad (11)$$

Here h_{ij} is the perturbed metric, a is the cosmological scale factor, γ^i is the unit vector describing the radiation direction, n_e is the electron density at time t, σ_T is the Thomson scattering cross-section and V_i is the perturbed velocity (relative to the comoving frame). Inspection of (11) shows that there are three contributions to the source term for Δ. The first term on the right hand side involves the gravitational potential via the h_{ij}, and is the formal analogue of (4), whereas the second term is associated with Thomson scattering induced fluctuations, and is the analogue of (2). The third term is equivalent to (1), modified by a factor $e^{-\tau}$, where τ is the optical depth $\int_{t_s}^{t} n_e\,\sigma_T\,c\,dt$. The first term yields the large-scale anisotropy $\Delta \sim \delta(\ell/ct^2)$ whereas the second group of terms yields the small-scale angular anisotropy $\Delta \sim f\tau(\ell/ct)$ and $\Delta \sim \delta e^{-\tau}$.

For a more quantitative calculation of the anisotropy, we may evaluate the autocorrelation function for temperature fluctuations. The initial density contrast is expressed in terms of a Fourier spectrum as defined previously by

$$\delta = \int \delta_k \, e^{i\underset{\sim}{k}\cdot\underset{\sim}{x}} \, d^3k, \tag{12}$$

where for the growing mode $\delta_k \propto t^{2/3}$. The normalization is determined by requiring the correlation function $\xi(r)$ for density fluctuations to coincide with that for the observed galaxy distribution, whence

$$|\delta_k|^2 = \int r^2 \, dr \; \xi(r) \, \frac{\sin kr}{kr} \, \left(\frac{V}{16\pi^5}\right). \tag{13}$$

The normalization involves considerable uncertainty. In general, the integral (13) is not well-defined, since possible contributions to $\xi(r)$ on large scales are very uncertain. According to some studies[8], there are sizable fluctuations in $\xi(r)$ at $r \gtrsim 50$ Mpc. More generally, one has

$$\xi(r) = \left(\frac{32\pi^4}{V}\right) \int_0^\infty k^2 dk \, |\delta k|^2 \, \frac{\sin kr}{kr} \propto r^{-n-3} \quad (kr \ll 1), \tag{14}$$

and one method of normalization is to match $\xi(r)$ with observations in the linear regime where $\xi(r) < 1$.

The final step is to compute the radiation temperature fluctuations. After solving (11) for individual Fourier components $\Delta_k \, (\underset{\sim}{k}, \underset{\sim}{\gamma})$, and Fourier transforming to yield $\Delta(\underset{\sim}{x}, \underset{\sim}{\gamma})$, one evaluates the radiation correlation function in terms of the angle θ between two different directions that an observer is comparing (for example, by beam switching). Thus letting $|\underset{\sim}{\gamma_1} - \underset{\sim}{\gamma_2}| = \theta$, one has

$$(\delta T/T)^2(\theta) = \langle \Delta(x, \gamma_1) \; \Delta(x, \gamma_2) \rangle, \tag{15}$$

where the average is taken over all points x. The computed $\delta T/T$ is therefore appropriate for an average point in space, and not located in any preferred position.

IV. COMPARISON WITH OBSERVATIONS

A. Large Scale Anisotropy

Consider first the large-scale angular anisotropy. Only the growing mode is retained. Then the solution to (11), including only the first term on the right-hand side, is

$$\Delta_k(\underset{\sim}{k},\underset{\sim}{\gamma},t_o) \approx \frac{2}{9}\frac{\delta_k}{K^2}\left\{1-3i\mu K - \left(1-3i\mu K z_d^{-1/2}\right)\exp\left[-3iK\mu\left(1-z_d^{-1}\right)\right]\right\}, \quad (16)$$

where $\underset{\sim}{K} = \underset{\sim}{k}\, ct_o$, $\mu = \underset{\sim}{k}\cdot\underset{\sim}{\gamma}$, z_d is the redshift of last scattering, and t_o is the present epoch. The dimensionless wavenumber $\underset{\sim}{K}$ expresses the wavenumber of the fluctuation relative to the

present horizon. Limiting cases of (16) are

$$K << 1: \quad \Delta_k \approx \delta_k\, \mu^2 + O(K\mu^3) \quad (17)$$

and

$$K >> 1: \quad \Delta_k = -\delta_k\left\{\frac{2}{3}\frac{i\mu}{K} + \frac{2}{9K^2}\left(1-3i\mu K z_d^{-1/2}\right)e^{-3iK\mu}\right\} \quad (18)$$

Since $\theta = \cos^{-1}\mu$ represents the angular dependence of the direction of observation relative to some specified direction, we see that the $K << 1$ (scales larger than the horizon) terms result in a quadrupole (and higher order) anisotropy, whereas the $K >> 1$ limit includes a dipole term that is dominant. In general, with the dipole term subtracted,

$$|\Delta_k| \sim \frac{2}{9}\,\delta_k/K^2, \quad (19)$$

for $K << z_d^{1/2}$, with a comparable amplitude for the different multipoles. The dependence of dipole and quadrupole moments on wave number is such that the dipole contribution dominates for a wide range of n. Imposing a cut-off on the fluctuation spectrum, for example by requiring the correlation function to tend to zero at $r > r_0$, is equivalent to setting $n = 0$ at $r > r_0$, since from (13), $|\delta_k|^2 \to$ constant.

For $\xi(r)$ to be convergent, the definition (14) requires $n > -3$. Numer-

ical n-body simulations of clustering have been used to model the non-linear evolution of $\xi(r)$. The general consensus seems to be that an effective post-decoupling value of the spectral index in the range $-2 \lesssim n' \lesssim 0$ yields $\xi(r)$ with the required $r^{-1.8}$ slope in the non-linear range. For isothermal fluctuations, the simulations directly yield n. For adiabatic fluctuations, the power spectrum index on scales $\lesssim ct_d$ prior to decoupling is $n = n' + 4$, because of the suppression of growth on scales below the horizon size prior to decoupling. Other constraints on n are more speculative. For example, if $n < 1$, the universe must be closed if the fluctuation spectrum is extended to sufficiently large scales. The critical value $n = 1$ corresponds to curvature fluctuations of the same amplitude on all scales, equivalent to a fixed amplitude for density fluctuations on the scale of the horizon. In terms of the mass spectrum, one has again that

$$\delta\rho/\rho = \left[\frac{(2\pi)^3}{V} \int |\delta_k|^2 \, d^3k\right]^{1/2} \propto M^{-1/2-n/6} \tag{20}$$

and the associated metric fluctuations on scale ℓ

$$|h_{ij}| \sim G\rho\ell^2 \, (\delta\rho/\rho) \propto M^{(1-n)/6}. \tag{21}$$

Direct observation of $\xi(r)$ in the linear regime should in principle be able to determine n directly according to (14). Evolution of the fluctuation spectrum prior to and during decoupling does leave its imprint in the form of two characteristic features. One peak is associated with the maximum Jeans mass, and can be substantial if $\Omega \sim 1$ due to the substantial growth of larger-scale fluctuations after first coming within the horizon at $z \lesssim 10^4 \, \Omega$. Another feature is associated with the damping of adiabatic fluctuations below $\sim 10^{13} \, \Omega^{-5/4} \, M_\odot$: if $\Omega \lesssim 0.1$, this occurs on a scale well within the linear regime where one might expect to detect its presence in the galaxy correlation function. For small values of n, both of these features may be in conflict with observations, although the observational situation is presently not entirely clear because of apparent large-scale

inhomogeneities that are being found in redshift surveys.

The large-scale anisotropy of the microwave background radiation provides a unique probe of the spectrum of density fluctuations on large scales. The detection[10,11] of a quadrupole moment finds a natural explanation in terms of large-scale structure as predicted by (16). The alternatives involve gross deviations from the standard model (such as invoking an anisotropic cosmological model, or a universe filled with very long wavelength gravitational waves). With the normalization fixed by comparison with the galaxy correlation function, the remaining free parameters are n and Ω if we adopt a power-law initial fluctuation spectrum. The results are shown in figure 1 of a computation[9] of the dipole, quadrupole, and higher moments of the radiation anisotropy in a spatially flat universe.

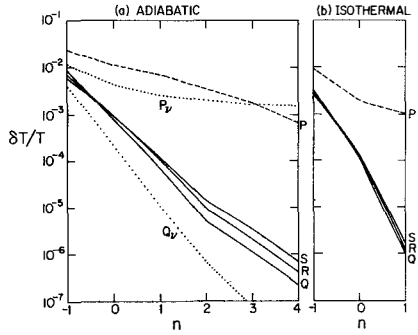


Figure 1. Predicted large-scale dipole (P); quadrupole (Q), octopole (R), and hexadecapole (S) anisotropy (a) for adiabatic and (b) for isothermal density fluctuations.[9] Dotted lines are dipole and quadrupole anisotropies for adiabatic fluctuations in a neutrino-dominated (Ω = 1) universe.[17]

Key points to note are that similar amplitudes are found for both adiabatic and isothermal modes, because the dominant effect is due to the interaction of radiation with the gravitational potential associated with the fluctuations. The amplitudes are spatial averages, and should be regarded as uncertain to a factor 3 or 4. The quadrupole and dipole moments are independent averages, and have no preferred relative alignment. The large dipole effect is due to Fourier components on scales from ~ 100 Mpc to the horizon size. In addition, there will be a contribution to the dipole anisotropy from our peculiar motion induced by the non-linear fluctuation that we are in, associated with the Virgo supercluster. The quadrupole and higher moments are due to fluctuations on the horizon scale and of greater wavelength. Comparison with the measured values of dipole[1,2,10] and quadrupole[10,11] anisotropy indicates that an isothermal fluctuation spectrum with $n \approx 0$ satisfactorily accounts for both. Adiabatic fluctuations are in some difficulty. The dipole anisotropy constrains $n \gtrsim 2$; however in this range, the quadrupole anisotropy is far too small. Thus an alternative explanation for the quadrupole anisotropy must be sought if adiabatic fluctuations are to survive as a viable theory.

Appealing to rescattering of the radiation after the decoupling epoch does not alleviate the difficulty with adiabatic fluctuations. Reheating and scattering only erases temperature fluctuations on an angular scale less than the horizon at the epoch of last scattering, or about $5(\Omega x)^{1/3}$ degrees, where x (assumed $>> 10^{-3}\ \Omega^{-1}$) is the fractional ionization. Variation of Ω complicates matters somewhat, but the results of figure 1 are relatively insensitive to Ω.

The most drastic change in these results occurs if neutrinos have a finite mass, in particular of ~ 30eV as suggested by a recent experiment. In a neutrino-dominated universe, which this value of the neutrino mass would imply, the amplitude of adiabatic density fluctuations at decoupling

is reduced by about an order of magnitude. The neutrino fluctuations experience an earlier period of growth once the neutrinos become non-relativistic, while the matter fluctuations are still undergoing damping by photon diffusion and acoustic oscillations. It is the neutrino fluctuations on scales above the neutrino Jeans mass that determine the final amplitude of the density fluctuations after the decoupling epoch[12,13]. In this case the dipole anisotropy is reduced by a factor 2, while the quadrupole anisotropy is reduced by a factor of 2-10 for $-1 \lesssim n \lesssim 2$. This reduction does help; it enables $0 \lesssim n \lesssim 1/2$ adiabatic fluctuations to provide a possible model that accounts (within a factor ~ 3) for both the dipole and quadrupole moments.

While constant curvature fluctuations ($n = 1$) apparently yield too low a value for the quadrupole moment, however, another difficulty with adiabatic fluctuations is alleviated in a massive neutrino-dominated universe. The galaxy autocorrelation function is consistent with $n \approx 1$ in this case, because the baryon mass within the maximum neutrino Jeans mass ($\sim 10^{14}\ M_\odot$) produces the dominant peak in the post-decoupling density fluctuation spectrum, as opposed to the much larger mass scale ($\sim 10^{17}\ M_\odot$) associated with the maximum matter Jeans mass in a standard model with $\Omega = 1$. Results of computations[17] of the galaxy autocorrelation function in the linear regime as given by equation (14) are shown in Figure 2 for three cases: with neutrinos ($\Omega_\nu = 0.98$, $\Omega_b = 0.02$) and without neutrinos ($\Omega_b = 1.0$ and 0.1). It should be noted that a value of $n < 1$ does lead to a divergence in the metric fluctuations on large scales, and a long wavelength cut-off or increase in n to a value in excess of 1 must be specified.

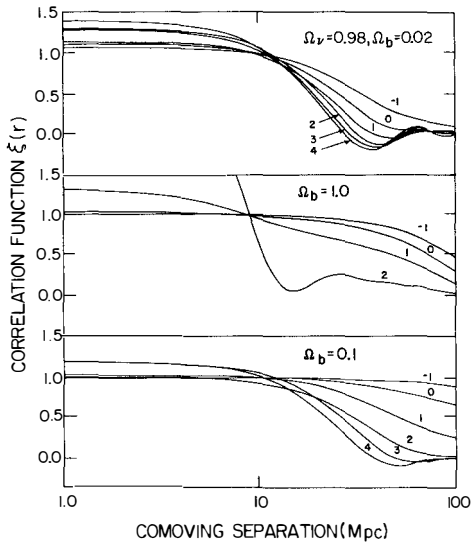


Figure 2. The galaxy autocorrelation function as a function of comoving separation in the linear regime for a neutrino-dominated universe (Ω_ν = 0.98, with baryon content Ω_b = 0.02), and for two cosmological models without neutrinos (Ω_b = 1.0 and 0.1). For each cosmological model, the correlation function is plotted for several values of the fluctuation spectral index n, as indicated.

B. Small-scale Anisotropy

The small angular scale structure of the background radiation (that is to say, with the dipole and quadrupole components subtracted off) is best described in terms of the autocorrelation function of the radiation distribution given by (15). To compare with observations, the antenna response, modelled by a Gaussian distribution $f(\Theta,\sigma)$ with beam width σ must be folded in, yielding a "smeared" correlation function

$$F(\theta,\sigma) \approx \frac{1}{2\sigma^2} \int_0^\infty \psi \, d\psi (\delta T/T)(\psi) \exp\left\{\frac{-1}{4\sigma^2}(\theta^2 + \psi^2)\right\} I_0\left(\frac{\theta\psi}{2\sigma^2}\right). \qquad (22)$$

For a point beam ($\sigma \to 0$), the correlation function reduces to $F(\theta,0) = (\delta T/T)(\theta)$. The mean square temperature fluctuation for a beam switched through an angle θ is

$$(\delta T/T)^2 (\Theta,\sigma) = F(0,\sigma) - F(\Theta,\sigma). \qquad (23)$$

The most significant observation to date is that reported by Fabbri et al.[14] on an angular scale of 6 degrees. Their claimed detection of $\delta T/T = 3(\pm 0.7) \times 10^{-5}$ with a beam $\sigma \approx 3°$ is compared in figure 3 with theoretical predictions for adiabatic and isothermal fluctuations.

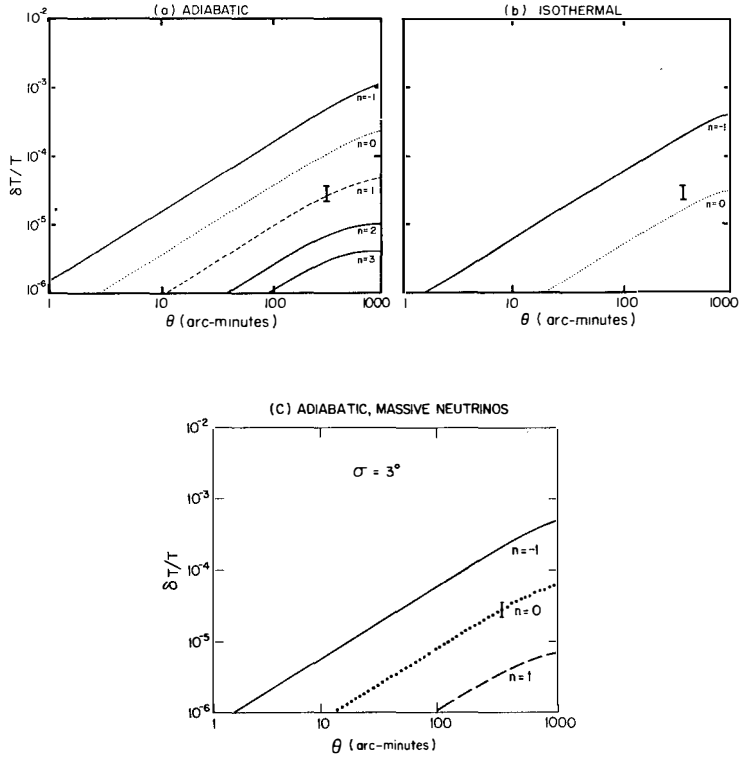


Figure 3. Temperature fluctuations for an antenna of beam width 3° and switched through an angle θ. Observation shown is due to Fabbri et al.[14]. Predictions[15,17] are (a) for adiabatic fluctuations, (b) for isothermal fluctuations, and (c) for adiabatic fluctuations in a neutrino-dominated (Ω = 1) universe.

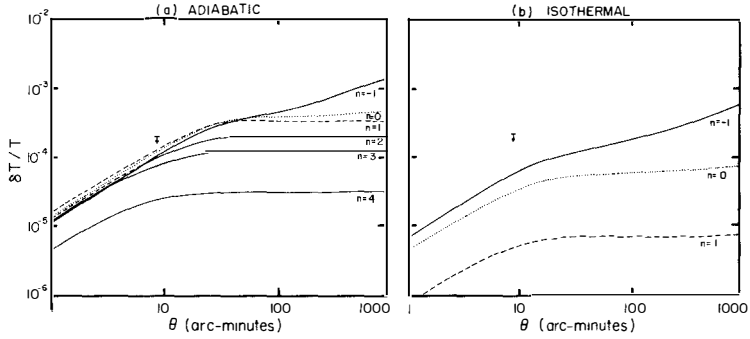
The calculations[15] are for a spatially flat model and cover a range of values of n. An adiabatic spectrum with $n \approx 1$ or an isothermal spectrum with $n \approx 0$ fits the reported fluctuations. If conservatively taken as an upper limit on $\delta T/T$, one concludes that $n \gtrsim 1$ for adiabatic fluctuations or $n \gtrsim 0$ for isothermal fluctuations, provided that $\Omega \approx 1$. It should be noted that an adiabatic fluctuation spectrum with $n > 1$ diverges on small-scales, and a small-scale cut-off or change in n to a value less than unity must be specified.

The predicted small-scale anisotropy scales as $\sim \Omega^{-1}$, and consequently $\Omega \lesssim 0.1$ constrains $n > 3$ (adiabatic) or $n > 1$ (isothermal), effectively ruling out the possibility of isothermal fluctuations (since cluster simulations require $n \lesssim 0$). On the other hand, a neutrino-dominated universe raises Ω to a value near unity, and in addition, reduces the predicted level of small-scale anisotropy by a factor ~ 10 for a neutrino mass of 30 eV (Figure 3), thereby allowing either model. The 6° observation is especially significant because reionization will not be able to appreciably attenuate fluctuations on this scale.

On smaller angular scales, a representative limit is that of Partridge[16], who finds

$$\delta T/T < 2 \times 10^{-4}, \ \theta = 9', \ \sigma = 1.8.$$

Figure 4 presents a comparison of predicted fluctuations for the beam used in this experiment with this upper limit. Provided that there has been no appreciable rescattering of the radiation since decoupling, and this is at least a direct implication of the adiabatic theory in which galaxy formation occurs at a late epoch, one concludes that adiabatic fluctuations can be reconciled with observation only if $\Omega \sim 1$. If $\Omega \sim 0.1$, the adiabatic



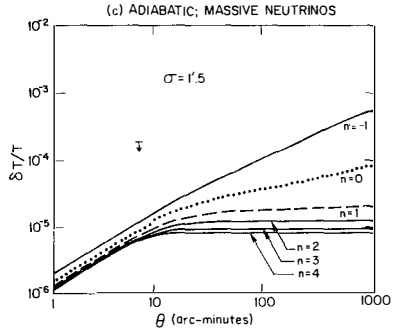


Figure 4. Temperature fluctuations as in figure 4 but for an antenna of beam width 1.5. Observational upper limit shown is due to Partridge.[16]

fluctuation theory is unacceptable for any value of n ($\lesssim 4$). However isothermal fluctuations are consistent with the upper limit on $\delta T/T$ if $n \lesssim 0$ and $\Omega \gtrsim 0.1$. As before, these conclusions about small-scale anisotropy are drastically revised in a neutrino-dominated universe, where the predicted level of anisotropy is reduced by a factor $\sim 10(m_\nu/30\text{eV})$.

In summary, one may conclude that astronomers are on the verge of verifying the gravitational instability theory of galaxy formation, one of the greatest outstanding challenges posed by a viable cosmology. The seeds of the observed large-scale structure in the Universe may already have been detected: it seems likely that as the sensitivity of future experiments is improved, the current indications of angular anisotropy will, at some level, be confirmed. If the current observations are accepted, it is the quadrupole anisotropy that provides the greatest constraint on theory. The fact that gravitational potential fluctuations are of order $G\delta\rho.L^2$ indicates that small amplitude but sufficiently large-scale density fluctuations, both at the present epoch and on the surface of last scattering, can produce significant large angular scale variations in the radiation temperature. Accounting for the quadrupole anisotropy requires, in the standard model, an isothermal fluctuation spectrum resembling white noise. An adiabatic fluctuation theory is viable only if the neutrino has a rest-mass in excess of a few eV, in which case $n \approx 1/2$. Perhaps the most important point is simply that the quadrupole moment is most simply and elegantly interpreted in terms of the density fluctuations on very large scales whose presence is inferred from the requirement that an initial fluctuation spectrum is required in order for structure to develop.

I am indebted to M.L. Wilson for providing indispensible advice on the anisotropy results. This approach has been supported in part by NASA under grant NGR 05-003-578.

REFERENCES

1. Smoot, G.F., Gorenstein, M.V., and Muller, R.A. 1977, Phys. Rev. Letters, 34, 898.

Smoot, G.F. and Lubin, P.M. 1979, Ap. J. (Letters), 234, L83.

2. Cheng, R.S., Saulson, P.R., Wilkinson, D.T., and Coey, B.E. 1979, Ap. J. (Letters), 232, L139.

3. Silk, J. 1980, in the 10th Texas Symposium Relativistic Astrophysics, Ann. N.Y. Acad. Sci. (in press).

4. Peebles, P.J.E. 1980, The Large Scale Structure of the Universe (Princeton: Princeton University Press).

5. Sachs, R.K. and Wolfe, A.M. 1967, Ap. J., 147, 73.

6. Silk, J. and Wilson, M.L. 1980, Physics Scripta, 21, 708.

7. Peebles, P.J.E. and Yu., J. 1970, Ap. J., 162, 815.

8. Kirshner, R., Oemler, A., and Schechter, P. 1979, A. J., 84, 951.

9. Silk, J. and Wilson, M.L. 1981, Ap. J. Letters, 244, 37.

10. Fabbri, R., Guidi, I., Melchiorri, F. and Natale, V. 1980, Phys. Rev. Letters, 44, 1563.

11. Boughn, S.P., Cheng, E.S., and Wilkinson, D.T. 1981, Ap. J. (Letters), 243, L113.

12. Bond, J.R., Efstathiou, G., and Silk, J. 1980, Phys. Rev. Lett., 45, 1980.

13. Doroshkevich, A.G., Zel'dovich, Yu. B., Sunyaev, R.A., and Khlopov, M. Yu. 1980, Sov. Astron. Letters, 6, 457.

14. Fabbri, R., Guidi, I., Melchiorri, F., and Natale, V. 1981, Proc. 2nd Marcel Grossmann Meeting (in press).

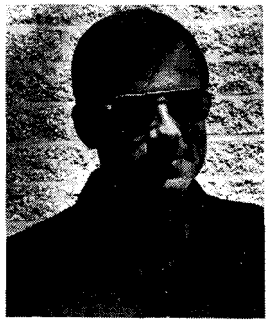
15. Wilson, M.L. and Silk, J. 1981, Ap. J., 243, 14.

16. Partridge, R.B. 1980, Ap. J., 235, 681.

17. Wilson, M.L. 1981 (in preparation).

THE EVOLUTION OF STRUCTURE IN THE UNIVERSE: OBSERVATIONAL CONSIDERATIONS

R. B. Partridge
Haverford College, Haverford, Pa., U.S.A.



Large scale structure in the Universe--galaxies and clusters, for instance--is believed to arise from small density perturbations generated at an early epoch in the expanding Universe. This paper treats the observational evidence which bears on the nature, spectrum and initial amplitude of such density perturbations. Special attention is devoted to studies of the microwave background radiation. The observational results, while not conclusive, favor isothermal over adiabatic perturbations, and suggest a perturbation spectrum $\Delta\rho/\bar{\rho} \propto M^{-1/2}$.

I. STRUCTURE IN THE UNIVERSE

Except on the largest scales, matter in the Universe is not uniformly distributed. Luminous matter--consisting of stars and galaxies--is clumped, as any deep photograph of the night sky will show. The most prominent features in the present mass distribution of the Universe are galaxies, with a typical mass of 10^9-10^{12} $M_\odot$, and clusters of galaxies, with masses 10^{13}-10^{15} $M_\odot$ ($1\ M_\odot = 2\times10^{30}$ Kg). The analysis of galaxy counts by Peebles and his collaborators[1] and the recent redshift survey by Davis[2] (discussed in this volume by Press) show that inhomogeneity in the distribution of mass extends over a wide range of mass and length scales.

While the existence of inhomogeneities in the Universe has been recognized for some time, only recently are explanations for them being offered. Those of us interested in the large scale structure in the Universe are faced with a number of difficult questions:--(1) What mechanism or process causes matter in the Universe to be clumped? (2) Why do the more obvious structures--galaxies and clusters--have the mean densities that they do? (3) Why is structure prominent on certain mass scales and not others? (4) To what extent is the inhomogeneity we see now due to "initial" conditions as opposed to subsequent physical processes which have modified them?

As an observational astronomer, I shall be concerned primarily with the observational results which bear on these questions. Since the observations are quite diverse, however, some theoretical framework seems advisable. What follows is a brief supplement to the paper of Silk in this volume. Both he and I base our remarks on the hot Big Bang model for the Universe. The crucial feature of this model is a hot dense phase early in the history of the Universe, when conditions ensured that thermal equilibrium obtained. Evidence for such a phase is provided by the (largely) thermal spectrum of the cosmic microwave background (see Silk and Rowan-Robinson, this volume, for further discussion).

A second property of this model is expansion. As the Universe expands, it cools. At an epoch some 10^6 years after the Big Bang, the temperature of the matter and radiation drops to $\sim 10^3$ K, and the primeval plasma recombines. The sudden disappearance of free electrons sharply reduces the (Thomson scattering) interaction between radiation and matter and these two components of the Universe decouple. Matter and radiation continue to cool,

but at different rates. A second consequence of sudden decoupling at $t \sim 10^6$ y is that the Universe becomes transparent, at least in the absence of any subsequent and universal re-ionization of the matter content of the Universe. The latter point in particular has important observational consequences, as we shall see.

The cooling of the Universe, the decrease in its mean density, and the evolution of other properties of an expanding Universe are most easily parametrized by the redshift $(z + 1)$. Using subscript zero to represent the present values of wavelength, temperature, density, random velocities and epoch, we have

$$\lambda_o = \lambda(z + 1)$$
$$T = T_o(z + 1)$$
$$\rho = \rho_o(z + 1)^3$$
$$\bar{v} = \bar{v}_o(z + 1)$$
$$t = \begin{cases} t_o(z + 1)^{-1} & \text{open (empty) model} \\ t_o(z + 1)^{-3/2} & \text{flat space model} \end{cases}$$

Redshift is used rather than epoch, since the relation between the two is model dependent (see, for instance, Weinberg[3]). The redshift of decoupling is $z_d \sim 1100$. Representative values for present cosmological parameters are $T_o = 2.8$ K, $\rho_o \sim 10^{-30}$ gm/cm^3 and $t_o = 10\text{-}20 \times 10^9$ y: the last two are model-dependent.

Of particular interest to us is the evolution of density inhomogeneities in an expanding Universe. Consider a small region with density $\rho = \bar{\rho} + \Delta\rho$ slightly above the average density. The amplitude of the density perturbation $\Delta\rho/\bar{\rho}$ will increase, but slowly, as shown a generation ago by Lifschitz[4]:

$$\Delta\rho/\bar{\rho} \propto (z + 1)^{-1}. \qquad (1)$$

This result applies in the linear regime, or for $\Delta\rho/\bar{\rho} \lesssim 1$. Because the power-law growth is slow, the structure we now see cannot develop from arbitrarily small perturbations.

Nor can the structure we now see develop from purely random fluctuations; the masses are too great. We thus must face a choice:--either some non-random perturbations in the density were present <u>ab initio</u> or some physical process operating in an initially homogeneous Universe produced density perturbations. The former option is unpalatable to many, especially when we consider just how special the initial conditions must have been to give us the precise scales and densities we now observe in subunits of the Universe. The latter option is much more fruitful, and is an area of cosmology on which the growing collaboration between astrophysi-

cists and elementary particle theorists is beginning to shed some light.

II. CLASSIFICATION OF PERTURBATIONS

Density perturbations can be characterized by three parameters. The first, loosely speaking, is their type; the second, the amplitude, $\Delta\rho/\bar{\rho}$; and the third, a scale, generally given as the mass of the perturbation, M.

Four major categories of perturbations have been discussed in the literature (for recent reviews, see Silk's article in this volume, Sunyaev[5], Ozernoi[6], and references therein).

In adiabatic perturbations, both radiation and matter is perturbed, thus keeping the specific entropy n_γ/n_{p+} constant (Sunyaev and Zel'dovich[7], Peebles and Yu[8]). In isothermal perturbations, as the name suggests, the temperature and hence radiation density are constant; only the matter is clumped (Sunyaev[5]; Gott and Rees[9]). Vortex or "whirl" perturbation involve large-scale motions of the matter and radiation; this form of perturbation has been discussed by Ozernoi[6],[10], Chibisov[11] and their collaborators. Finally, there is the possibility of source-free gravitational waves (see Dautcourt[12]).

The two other important parameters, mass and fractional amplitude, may be expressed by giving the index and normalizing coefficient of a mass spectrum

$$\Delta\rho/\bar{\rho} = k\left(\frac{M}{M_o}\right)^{-\alpha} \qquad (2)$$

The index α can be related to the spectral power index n defined by $(\Delta\rho/\bar{\rho})^2_\ell \propto \ell^{-n}$ as follows:

$$\alpha = \frac{1}{2} + \frac{n}{6} .$$

If $\Delta\rho/\bar{\rho}$ is time dependent, the mass spectrum will be also.

A variety of theoretical arguments have been made to assign values to k and α. One suggestion with considerable appeal is due to Zel'dovich[13]: $\alpha = 2/3$ ensures that the perturbations are of constant curvature as they enter the horizon (see Silk, this volume). As we shall see, this requires $k \gtrsim 10^{-4}$ for $M_o \sim 10^{15}\ M_\odot$ at z_d. Other arguments suggest $\alpha = 1/3$ (Gott and Rees[9]), $\alpha = 1/2$ (Hogan[14]) or $\alpha = 5/6$ (Press, cited by Gott[15]) with values of k several orders of magnitude below or above 10^{-4}, respectively. A point of special relevance to this conference is the link between particle physics and the index α. Predictions, rather than assump-

tions, can be made about the index α based on the particle physics of the early Universe (as discussed here by Silk, Press and Nanopoulous, and elsewhere by Gott[15]). Particle physics arguments may also bear on the type of perturbations generated early in the Universe: for instance, explanations of the baryon number of the Universe based on GUT's require adiabatic fluctuations.

As I shall try to show later, we may be able to determine the spectrum and even the type of primordial density perturbations from astronomical observations. The possibility of contributing to fundamental physics this way is an exciting prospect for an astronomer.

Unfortunately, while the initial mass spectrum holds the most interest for fundamental physics, we will probably not be able to observe it directly, except possibly at the large mass end. The reason is that a variety of physical processes affect the perturbations before the earliest epoch at which we can hope to observe them. Some of these processes are mass dependent, and hence will alter the mass spectrum. Fortunately, many of them are well understood.

To sketch the effect of these processes, I would like to focus attention on perturbations of three masses, 10^{11} $M_\odot$, 10^{15} $M_\odot$ and 10^{19} $M_\odot$, which are, respectively, the representative mass of a galaxy, the largest aggregates of matter with $\Delta\rho/\bar{\rho} \gtrsim 1$ now, and a rough estimate of the mass of the Universe at the important epoch of decoupling.

10^{11} $M_\odot$ Perturbations of this mass cross the horizon* or become causally connected about 1 year after the origin of the Universe. Prior to that time, the perturbation amplitude can increase. Almost immediately after entering the horizon, a 10^{11} $M_\odot$ perturbation will stop growing and enter a phase of oscillation (since its mass has fallen below the Jeans mass, which in turn approximately equals the mass within the horizon--Rees[16]). The oscillatory phase lasts until the epoch of decoupling when the Jeans mass suddenly drops to $\sim 10^5$ $M_\odot$[17]. During this oscillatory phase, adiabatic perturbations of mass $\lesssim 10^{13}$ $M_\odot$ are strongly damped by photon drag, a process described first by Silk[18]. On the other hand, this process does not affect isothermal perturbations; for

*At any epoch, t, the radius of the Universe is $\sim$ ct. As the Universe grows older, its radius and mass increase. In this case, the mass within a radius ct was $\sim 10^{11}$ $M_\odot$ for t $\sim$ 1 yr: larger masses were not causally connected then or earlier.

these $\Delta\rho/\bar{\rho}$ remains constant throughout the oscillatory phase. Note the sharply divergent predictions about the amplitude of fluctuations on a scale of $\sim 10^{11}\ M_\odot$ depending on the type of perturbation (figure 1).

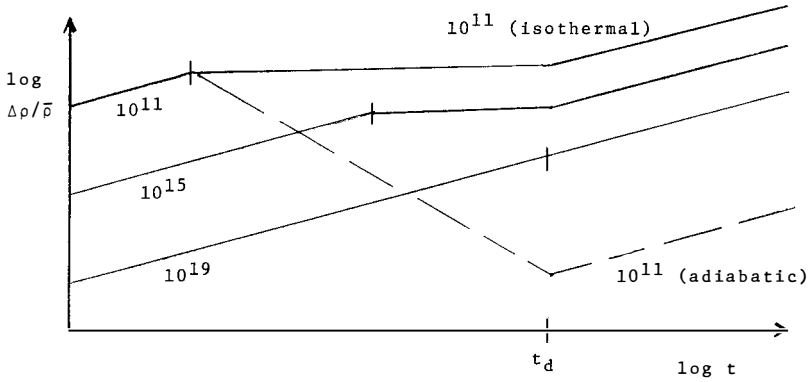


Figure 1. Schematic representation of the evolution of density perturbations of different mass. The (baryonic) mass in the perturbation is given in $M_\odot$. Vertical ticks mark the epoch at which the Jeans mass equals each.

The fate of vortex perturbations of this mass has been considered by Chibisov[11] and reviewed by Jones[19]. Readers are referred to these papers for details.

$10^{15}\ M_\odot$. In many ways, perturbations in this mass range have the simplest life histories. They enter the horizon well before the epoch of decoupling, then oscillate without appreciable damping until the epoch of decoupling. Observations of perturbations on this scale and larger, then, can tell us directly about the initial mass spectrum.

$10^{19}\ M_\odot$. This value was chosen to be the mass just crossing the horizon at the decoupling epoch. Slightly less massive perturbations may or may not oscillate depending on the relation between the Jeans mass and the horizon mass. This in turn is model dependent; the Jeans mass will approximately equal the horizon mass if the Universe is radiation dominated until decoupling. For this, it is necessary to have the energy density in radiation larger than $\rho_m c^2$ until t_d. Here ρ_m is the density of matter alone. In the open, low density, cosmological models now favored, this condition does hold until decoupling.

Physical processes at and after t_d. Physical processes during the epoch of decoupling, when the fractional ionization is changing rapidly, are complicated and still not fully understood. In general, they work to reduce $\Delta\rho/\bar{\rho}$. Detailed discussions and calculations are given by Peebles and Yu[8] and Doroshkevitch et al[20].

After decoupling, any perturbation with a mass greater than the Jeans mass of $\sim 10^5\ M_\odot$ will grow[17]. However, as Lifschitz[4] showed many years ago, the rate of growth is slow

$$\Delta\rho/\bar{\rho} \propto (z + 1)^{-1}$$

as well as mass-independent. This phase lasts until $\Delta\rho/\bar{\rho}$ begins to exceed unity, and non-linear effects appear, including possibly dissipation and fragmentation*.

Note that pure Lifschitz growth would produce a spectrum of masses in the Universe today having the same mass-dependence as the perturbations at decoupling. This fact presents a problem for those favoring adiabatic perturbations, since the Universe today is clumpier on galactic mass scales than on scales $\gtrsim 10^{13}\ M_\odot$. The answer, supplied by Zel'dovich[21] among others, is that pure Lifshitz growth of adiabatic perturbations with masses in excess of $\sim 10^{13}\ M_\odot$ is followed by collapse and fragmentation. "Pancakes" of $\sim 10^{15}\ M_\odot$ collapse along a single axis and fragment to produce galaxies. In this scenario, the present mass spectrum reflects the mass spectrum at decoupling only for masses $\gtrsim 10^{15}\ M_\odot$.

An alternative model is provided by Press and Schechter[22]: gravitational instability. In this scenario, larger and larger masses form by gravitational "agglomeration" of smaller masses. At present this process has produced density perturbations $\Delta\rho/\bar{\rho} \sim 1$ on a scale of $\sim 10^{15}\ M_\odot$. This scenario requires the presence of smaller mass "particles" which clump to form larger units, and thus appears to be consistent only with isothermal perturbations.

Once again, note the clear distinction between the two scenarios. If the perturbations at t_d are primarily adiabatic, only masses $\gtrsim 10^{13}\ M_\odot$ will be represented; the pancake theory or some variant of it is favored; and galaxy formation is a secondary process. If the perturbations are primarily isothermal, the predominant mass scales will be $\sim 10^5\ M_\odot$ and larger; and galaxies and

*Or, in open models, until the epoch corresponding to $z = (\rho_c/\rho_o) - 1$ (see III, ¶4, below).

clusters will form, in that order, by gravitational instability. Can observations help distinguish between these two cases?

III. ASTRONOMICAL EVIDENCE BEARING ON DENSITY PERTURBATIONS

The density perturbations we have been discussing are responsible for the structure we see in the Universe today. It follows that a careful study of the distribution of matter now may reveal some of the properties of the spectrum of density perturbations which were present at a much earlier epoch. In this section I will discuss some of the relevant observational evidence, both direct and indirect, which may help answer the questions raised in section I.

The epoch of galaxy formation. From direct observations[23], we know galaxies exist at redshifts of order unity. Thus they must have formed--that is, contracted and fragmented to form stars[24],[25]--by an early epoch.

Several attempts[26] have been made to observe primeval galaxies--galaxies in the act of forming their first generation of presumably highly luminous stars. None of these searches has been successful. If the models of galaxy formation[25] which initially motivated them are correct, then the epoch of galaxy formation can be pushed still further into the past. Very roughly, the redshift of galaxy formation is limited to $z_f \gtrsim 7$ (see [26]).

Several indirect arguments are in accord with $z_f \sim 3$-10. First, there are the number counts of QSO and radio sources, which suggest that some objects of roughly galactic mass were forming at $z \sim 3$. Next, one can argue[25] from the presently observed density of galaxies that they must have formed at redshifts of order 10, unless very significant dissipation was involved. Finally, in open cosmological models, growth in density perturbations following the $(z + 1)^{-1}$ law of Lifschitz[4] ceases as the Universe enters its linear expansion phase[27]. Thus galaxy (or cluster) formation must be complete by an epoch $z \sim \frac{\rho_c}{\rho_o} - 1$ where ρ_c is the so-called critical density $3H_o^2/8\pi G$, and ρ_o is the present density. Many cosmologists[28] now favor $\rho_o \sim 0.1\rho_c$; hence $z_f \gtrsim 9$.

These limits on z_f set constraints on the amplitude of perturbations at the decoupling era. If we assume $\Delta\rho/\bar{\rho} > 1$ is required to initiate the process of collapse and fragmentation, we have from relation (1)

$$(\Delta\rho/\bar{\rho})_d = 1\left(\frac{z_f + 1}{z_d + 1}\right).$$

With values for z_f in the range 3-10, we see that perturbations of 0.3-1% are required.

So far, this argument has assumed that galaxies form directly from perturbations of characteristic galactic mass, say 10^{11} $M_\odot$. If so, we require $\Delta\rho/\bar{\rho} \gtrsim .003$ on a mass scale of 10^{11} $M_\odot$ at t_d. Galaxies may, however, form as fragments in the collapse of larger structures, such as the "pancakes" of Zel'dovich[21]. In this scenario, pancakes of $\sim 10^{15}$ $M_\odot$ have values of $\Delta\rho/\bar{\rho} \sim 1$ at $z \sim 10$, so we see that perturbations of order 0.3-1% are required, but on a larger scale than for direct galaxy formation.

More detailed arguments[5], [29] suggest somewhat smaller lower limits on $(\Delta\rho/\bar{\rho})_d$. Nevertheless, the crucial conclusion still holds--substantial density perturbations at t_d are required to produce the structure we now see in the Universe. Equally, the perturbation amplitude cannot have been too high; otherwise presently observed systems would have much higher densities than they do. These results help pin down the normalizing coefficient k in equation (2).

Correlations. The important work of Peebles and his collaborators[1] has shown that correlations exist between the positions of galaxies over distances in the range 0.1-10 Mpc at least. Following Peebles, we define a dimensionless correlation function $\xi(r)$ by

$$\xi(r) = \langle\rho(\underline{r}_1)\rho(\underline{r}_1 + \underline{r})\rangle/\langle\rho\rangle^2 - 1. \qquad (3a)$$

His work on a variety of galaxy samples shows

$$\xi(r) = (r/r_o)^{-1.77} \qquad (3b)$$

with $r_o \sim 5\left(\frac{100}{H_o}\right)$ Mpc. This value of r_o corresponds approximately to 10^{15} $M_\odot$. Here we have, in principle, a means of determining the index α of equation (2). This problem has been addressed both by Peebles[30], [1] and by Gott and Rees[9]. Three questions arise in these papers: what index α is favored by the observed relation (3); does the relation (3) hold for perturbations on the largest scales $r \gtrsim 10$ Mpc; and is some cutoff needed at either large scales or small scales to prevent unacceptable physical consequences? To this writer, at least, these questions do not seem to have been finally decided. The redshift survey of Davis[2] reveals substantial structure on scales $\gtrsim 10$ Mpc; it appears to be consistent with

an extrapolation of (3) rather than to fall below it (Davis, private communication). This may tend to weaken Peebles' argument[30] that the initial mass spectrum must be steep. In [15] it is suggested that either a low density model with $\alpha \sim 1/3$ or a critical density model with $\alpha = \frac{1}{2}$ provide as good a fit to Peebles' data as do models for adiabatic perturbations with $\alpha \gtrsim 1$ (as favored in [30]). If $\alpha \gtrsim 1$, some cutoff at short wavelengths (small masses) is required to prevent the formation of Black Holes by gravitational collapse of perturbations of small mass--see the following paragraphs. The situation is less crisply defined than we might like.

Indirect constraints: limits on perturbation amplitudes at epochs $\ll t_d$. Up to this point in Section III, we have been considering constraints on perturbations which can be set at t_d or after. There are, however, several arguments which allow us to limit $\Delta\rho/\bar{\rho}$ at much earlier epochs. I have already alluded to the first: if $\alpha > 2/3$, perturbations do not enter the horizon with constant curvature. Instead, at epochs closer to $t = 0$, when the horizon mass was smaller, the curvature perturbations were larger. In other words, $\alpha > 2/3$ implies an initially chaotic Universe with large curvature perturbations; these could easily have collapsed to form Black Holes for which there is no astrophysical evidence.

Likewise, if $\alpha > 2/3$, or k is too large, perturbations present at $t \sim 100$ sec would have altered primordial nucleosynthesis (Audouze, this volume; Wagoner[31]), particularly He production[32]. Since the observed He abundance appears to be in acceptable agreement with that predicted for homogeneous, isotropic Friedman models (see Kunth, this volume; Greenstein[33]), we must conclude that $\alpha \lesssim 2/3$ and that k was not too large, or that some cutoff operated to prevent large amplitude perturbations on small scales.

While these constraints are interesting, they are by no means conclusive. Fortunately, there is another approach to the questions raised in Section I that is now being actively exploited—careful study of the angular distribution of the cosmic microwave background.

IV. FLUCTUATIONS IN THE COSMIC MICROWAVE BACKGROUND

For more than a decade, it has been recognized that the microwave background will be affected by non-uniform velocity fields or inhomogeneous distribution of matter on the surface of last scattering[7], [34]. The effect is to introduce intensity or

temperature fluctuations $\Delta T/T$ in the microwave background. Thus a study of the fine scale anisotropies in the microwave background provides a "snapshot" of the distribution of matter at a much earlier epoch, the epoch of last scattering of the microwave photons. In principle, such a "snapshot" would permit us to determine both the slope and amplitude of the perturbation spectrum (2).

To proceed further, we need to know what the redshift of the epoch of last scattering is. Certainly $z_s \lesssim z_d$; before the primeval plasma recombined, Thomson scattering kept the mean free path of photons low. If there was no subsequent re-ionization of the matter content of the Universe, we may take $z_s = z_d \sim 1100$. We will make this assumption for the remainder of this Section, then relax it at the end.

Assuming that the surface of last scattering is at a large redshift, $z_d \sim 1100$ permits us, following Weinberg[3], to write down a relation between the mass of a perturbation, assumed approximately spherical in shape, and its observed angular diameter

$$\theta = \frac{\rho_o H_o}{2 c \rho_c} \left(\frac{6M}{\pi \rho_o} \right)^{1/3} \qquad (4)$$

where, as before, ρ_c is the critical density $3H_o^2/8\pi G$ and ρ_o is the actual mass density of the Universe.

Equation (4) above establishes a straightforward link between an observational parameter and the mass scale of a density perturbation. Unfortunately, the link between the amplitude of density perturbations and the fluctuation level $\Delta T/T$ is much more complicated. In the first place, fluctuations in the cosmic microwave background are produced by a number of physical processes, the more important of which will be described briefly below. In addition, because the recombination of the primeval plasma is not instantaneous, it follows that in many cases the observable fluctuations in the microwave background are averaged out[5], [7].

The most straightforward connection between $\Delta\rho/\bar{\rho}$ and $\Delta T/T$ holds for purely adiabatic fluctuations: here $\Delta T/T = 1/3\ \Delta\rho/\bar{\rho}$. Unfortunately, this simple relationship between temperature fluctuations and density perturbations will apply only for adiabatic perturbations of mass $\gtrsim 10^{15}\ M_\odot$, since smaller mass perturbations will be optically thin through the epoch of decoupling so that their temperature fluctuations must be averaged over a line of sight, and thus the observed level of $\Delta T/T$ will be reduced[5].

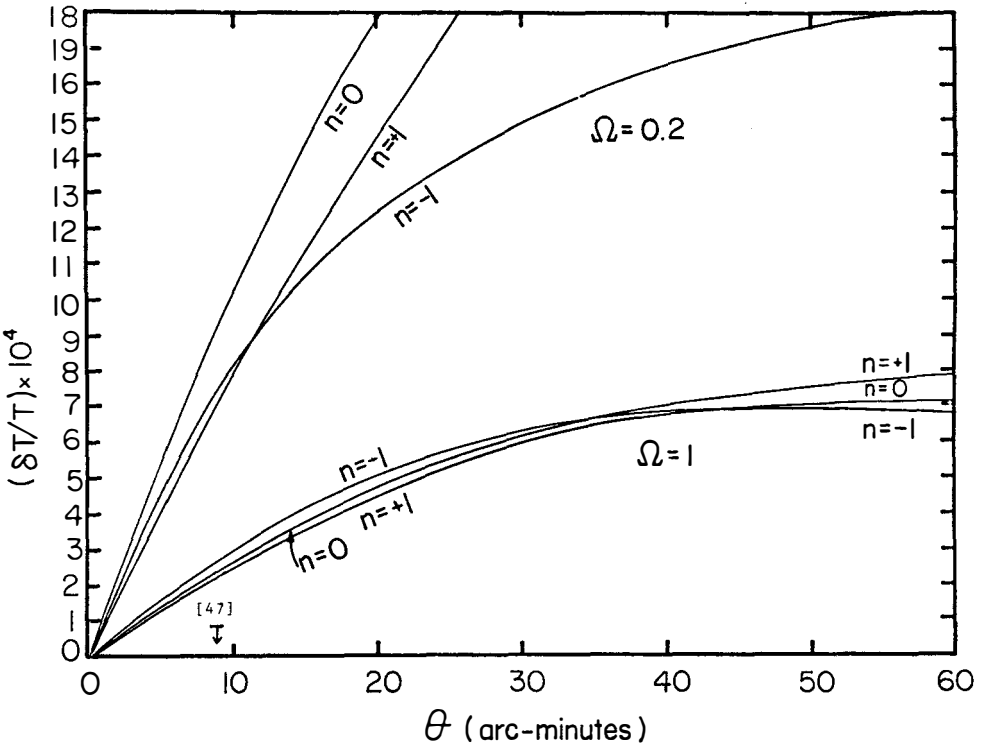


Figure 2. Calculated fluctuations assuming adiabatic density perturbations; from [29]

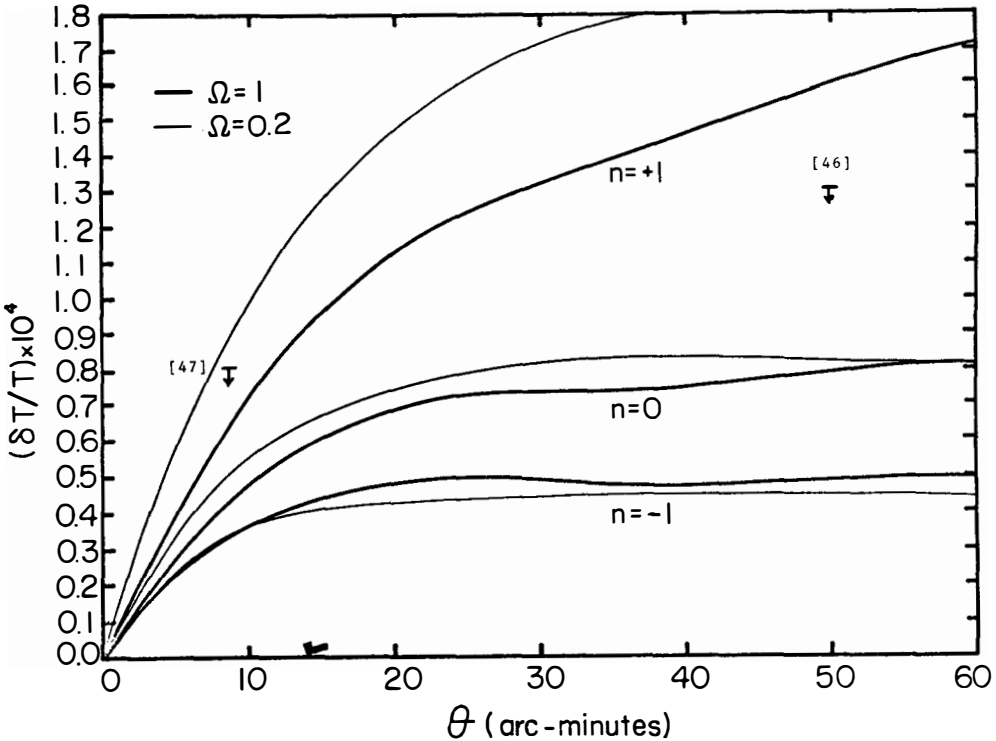


Figure 3. Calculated fluctuations assuming isothermal density perturbations; from [29]

For lower mass perturbations of all types, most of the observable temperature fluctuations are generated by the Doppler effect arising either from the motion of matter caused by the gravity of the perturbations at t_d[5], [29], or directly from vortex perturbations[11], [35]. The latter will be dominant if present.

Calculations of the amplitude of observable fluctuations as a function of mass (or angular scale) have been made by several groups both analytically[5], [7] and numerically[8], [20], [29], for many values of α. While there is some disagreement on the predicted amplitudes, especially for isothermal fluctuations[20], [29], for all reasonable values of α, $\Delta T/T \sim 10^{-5}$-10^{-3}. In the case of adiabatic perturbations, larger values of $\Delta T/T$ result for the lower values of ρ_o/ρ_c now favored. As a representative example of calculated values of $\Delta T/T$, we reproduce the results of Silk and Wilson[29] as figures 2 and 3.

Calculations of $\Delta T/T$ for vortex perturbations[35] give substantially higher values, because of the large velocity fields at t_d.

A different and more conservative approach to this question is to ask what is the <u>minimum</u> permissible value of $\Delta T/T$. Davis and Boynton[36] take this tack. Assuming conservatively that only isothermal perturbations are present, they show that $\Delta T/T$ must have a lower bound of a few times 10^{-6} in order to be consistent with the observed distribution of galaxies now.

The possibility that massive neutrinos make up much of the mass of the Universe (see Cowsik and also Schramm, this volume) still further complicates and attenuates the connection between $\Delta T/T$ and $\Delta\rho/\bar{\rho}$. This question is being actively investigated by Silk, among others--see his contribution here. Quantitatively, the effect of adding massive neutrinos or other non-baryonic matter is to reduce $\Delta T/T$ for a given $\Delta\rho/\bar{\rho}$.

Finally, as we prepare to turn to the observational results, it is worth noting that predicted values of $\Delta T/T$ resulting from perturbations at t_d reach a maximum on angular scales $\gtrsim 10'$, corresponding approximately to a mass of $10^{15}\ M_\odot$. For smaller masses and scales, either the perturbations are damped out before t_d, or the observable fluctuation level is reduced by averaging along the line of sight through the decoupling epoch, as mentioned above.

Observational results. The opportunity provided by the microwave background to study the distribution of matter at an early epoch has led many groups to search for temperature fluctuations. Two general remarks may be made about the results to date:--(1) No one has reliably reported detection of fluctuations which are indisputably primordial; only upper limits on $\Delta T/T$ on various angular scales are available. (2) These upper limits, however, are tight enough to constrain the spectrum of density perturbations, and even to eliminate certain types of perturbations entirely.

Observational upper limits available in mid-1979 are summarized and reviewed critically in [37]. Table 1 is reproduced from [37]; the corrections referred to in the last column were made by the author, not by the observers, and include corrections for telescope efficiency and conversion to the 95% confidence level[37].

Table 1. A list of all published upper limits on the small-scale anisotropy of the cosmic microwave background. The results are generally expressed as limits at the 95% confidence level. The corrections are my own estimates, and may be ignored.

Observers	Wavelength cm	Angular scale	Reported or published upper limit $\Delta T/T$	Corrected upper limit $\Delta T/T$
Conklin et al. [38]	2.80	10'	1.8×10^{-3}	3.5×10^{-3},*
Penzias et al. [39]	0.35	2'	9.0×10^{-3}	2.0×10^{-2}
Boynton et al. [40]	0.35	$\sim$ 1.5	3.7×10^{-3}	1.8×10^{-3}
Carpenter et al. [41]	3.60	2'-1°	7.0×10^{-4}	
Parijskij [42]	2.8	3'-1°	3.0×10^{-5}	4.0×10^{-4}
Parijskij [43]	4.0	$\sim$ 12'×40'	5.0×10^{-5}	1.6×10^{-4}*
Stankevich [44]	11.0	8'-20'	1.5×10^{-4}	4.0×10^{-4}*
Caderni et al. [45]	0.13	30'	1.2×10^{-4}	
Parijskij et al. [46]	4.0	5' to 150'	8.0×10^{-5} to 1.3×10^{-5}	4.0×10^{-4} to 7.5×10^{-5}
Partridge [47]	0.9	$\sim$ 7'	8.0×10^{-5}	

*Not corrected for possible errors in statistical analysis, but only for telescope efficiency. Also converted to 2σ values.

The results of Parijskij et al.[46] and Partridge[47] set the most stringent limits near the two ends of the range of angular scale covered by the measurements. When these results are compared with figures 2 and 3*, it is clear that they fall well below the predictions of the models for adiabatic perturbations, and effectively

*Note that experiment [47] employed beam switching; hence, as shown by Boynton[48],[36], the true angular scale sampled was the beam-switch angle of 9', not the actual size of either beam. The upper limit has been plotted at θ = 9'.

limit the index α to $\lesssim 1/2$ for isothermal perturbations. In addition, if we put these limits into the context established by Chibisov and Ozernoi[11] and Anile et al[35], we see that models based on vortex or "whirl" perturbations are also ruled out.

Since 1979, additional upper limits have been published or reported[49]-[52]. While some of these do not appreciably change the picture summarized in Table 1, the very low upper limit[52] on fluctuations on a scale of $\sim 5°$ is of great importance: at submillimeter wavelengths, $\Delta T/T \lesssim 2\text{-}3\times10^{-5}$. It seems likely that at least some of the observed fluctuation on this angular scale at this short wavelength is produced by warm, emitting, dust clouds in our Galaxy. Thus the true level of primordial fluctuations on this scale must be lower still. From a theoretical point of view the importance of this result lies in the fact that the corresponding mass scale is $> 10^{19}\ M_\odot$, the horizon mass scale at t_d (see Section II). Thus this observation samples the mass spectrum at t_d directly, uninfluenced by any causal physical process such as damping. From an observer's point of view, the results show the power of bolometric receivers with their very wide bandwidth[53]. A group of us[54] is attempting to repeat such measurements on a large angular scale (in our case $\sim 3°$), using conventional ground-based superheterodyne receivers at $\lambda \sim 3$ cm. At present our results are more than an order of magnitude less sensitive than the Italian results[52]. It seems clear to me that the best future observations on 1/2°-20° scales will be made with bolometers, and generally not ground-based.

On the other hand, on smaller angular scales, 3'-30', conventional radio astronomical techniques can be pushed further. For instance, it appears that a new receiver at the 140-foot telescope operated by the National Radio Astronomy Observatory in the U.S. should permit observers[55] to reduce the upper limits on $\Delta T/T$ by a factor of 2-3 on a scale of $\sim 12'$.

On still smaller angular scales, filled aperture antennas cannot be used--their resolution is $\gtrsim 1'$ because of diffraction. Hence observers[56], [57] have turned to radio interferometry or aperture synthesis. In addition to permitting observations on angular scales as small as 1" or less, aperture synthesis provides in principle a two-dimensional map of the temperature fluctuations. Unfortunately, aperture synthesis also has a strong disadvantage built in[57]--much lower sensitivity for a given integrating time

than a filled aperture telescope. For instance, 24 hours of observation with the Very Large Array has produced a tentative upper limit $\Delta T/T \lesssim 1\times10^{-2}$ on a scale of $\sim$ 6" (see Table 2 below): this same integrating time resulted in an upper limit of 8×10^{-5} at $\sim$ 7' (see [47]). The use of shorter baselines at the VLA will improve the limits on $\Delta T/T$, but not by a large factor.

The results of the aperture synthesis observations to date appear in Table 2. Given both the poor sensitivity and the fact that predicted values of $\Delta T/T$ discussed above drop rapidly as θ falls below $\sim$ 10', it is reasonable to ask why people are pursuing these observations. One answer is that the theoretical predictions are model dependent and not always in perfect agreement; it is prudent to check them. A much more compelling reason is laid out in the following paragraphs.

The effect of re-ionization at epochs >> t_d. Up to this point, we have been assuming that the epoch of last scattering--the epoch we "see" when we observe the microwave background--coincided with the epoch of decoupling at $z_d \sim 1100$. A number of scenarios for the formation of galaxies and clusters suggest that this may not be so (see [58], [59], and Rowan-Robinson and Puget, this volume). If the matter in the Universe is re-ionized at an epoch corresponding to $1100 > z \gtrsim 10$, perhaps by energy released in the process of galaxy formation, Thomson scattering from the free electrons will shift the surface of last scattering to a much lower redshift $z_s \sim 10\text{-}20$ (depending on the density ρ_o). On the other hand, re-ionization after an epoch of $z \sim 10\text{-}20$ will have no effect; the Universe will remain transparent so that we may again take $z_d \sim 1100$ as the surface of last scattering.

If matter is re-ionized at $z > 10\text{-}20$, three important observational consequences follow:--(1) Obviously, observations of temperature fluctuations will no longer tell us anything directly about the initial mass spectrum. (2) While density perturbations on the surface of last scattering will still produce temperature fluctuations by the Doppler mechanism discussed above, the angular scale at which $\Delta T/T$ rises to its maximum value will be shifted from $\sim$ 10' to $\sim$ 3° (see [60]). (3) Perhaps the most intriguing new element introduced is the possibility of wavelength-dependent fluctuation amplitudes[59]. Pure Thomson scattering (at these energies) is entirely independent of wavelength. However, there are other scattering and absorbing processes which may come into play during the

epoch of galaxy formation, and many of these are wavelength-dependent. Extinction by dust, discussed in this volume by Puget and Rowan-Robinson, is one example. Hogan[59] has looked at these processes in some detail, and has predicted values of $\Delta T/T$ for a specific model. His results appear as figure 4 here. Note, first, the wavelength dependence of the contours of $\Delta T/T$, and, second, the large values of $\Delta T/T$ that these scattering processes produce on small angular scales. Indeed, comparison of Table 2 with figure 4 suggests that the available interferometric observations, insensitive as they are, already put the squeeze on Hogan's models. It seems likely to me that tighter constraints on the small scale anisotropy of the cosmic microwave background, combined with more precise measurements of its spectrum, can provide a critical test for those theories of galaxy formation which call for large releases of energy at epochs much later than t_d.

Table 2. Aperture synthesis (interferometric) searches for fluctuations in the cosmic microwave background. See [56] for references to the original papers.

Observers	Wavelength cm	Angular scale of synthesized beam	$\Delta T/T$
Goldstein et al.	21	$\sim$ 17"	$\lesssim 3\times10^{-2}$
Martin et al.	11	13"	$(7\pm4)\times10^{-2}$*
" "	3.7	4"	$< 4\times10^{-2}$
Partridge et al.	6.0	6"	$\lesssim 1\times10^{-2}$**

*Detected fluctuations probably due to discrete sources[56].
**Preliminary result.

Very large angular scale anisotropy. It is well known that there is a dipole anisotropy of $\sim 10^{-3}$ in the microwave background[61]. Only recently have these difficult experiments been pushed hard enough to measure higher moments of the distribution[62]. Although measurements by different groups are not in perfect accord, there is clear evidence for a quadrupole component. While a quadrupole component can be introduced by anisotropy in a purely homogeneous cosmology[63], it is also possible to regard the quadrupole signal as a fluctuation on a particularly large scale. This approach has been taken by Silk and Wilson[64], and is discussed in more detail by Silk in this volume. Here I shall mention only that no adiabatic perturbation model appears consistent with the observations of both the dipole and the quadrupole moments (unless massive neutrinos are added to the picture). In general an adiabatic model produces too large a dipole component and too small a quadrupole component. A model based on purely isothermal

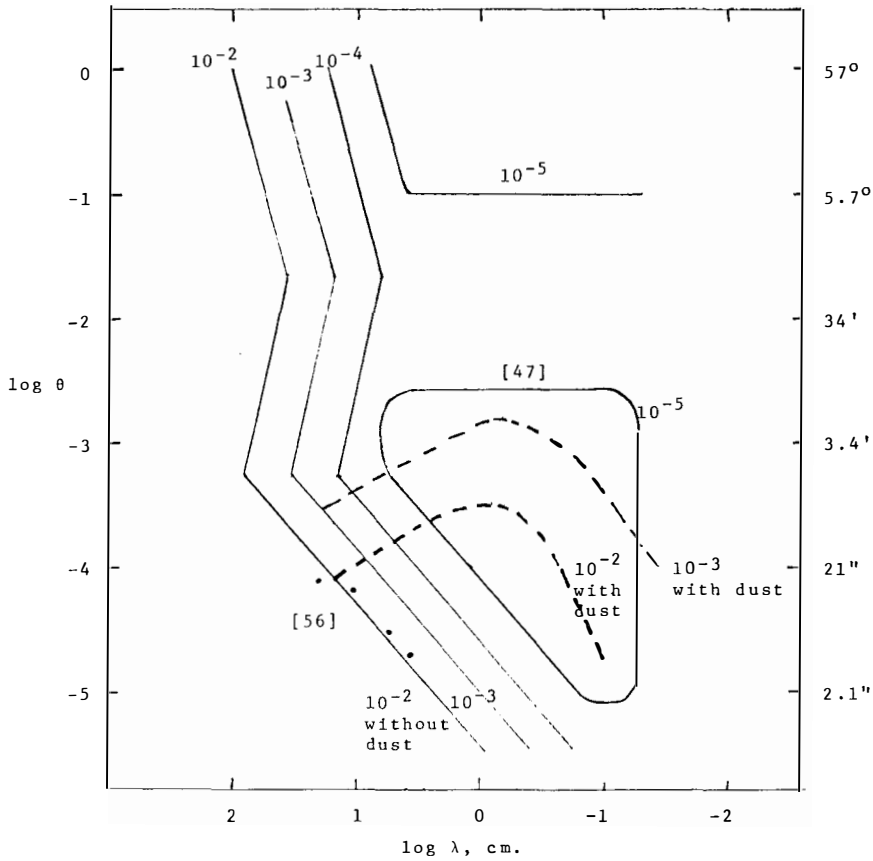


Figure 4. Calculated contours of $\Delta T/T$, with and without dust. Adapted from [59]. Parameters of the interferometric [56] and one filled-aperture [47] observations are indicated.

perturbations with index $\alpha \sim \frac{1}{2}$ does appear to fit the available dipole and quadrupole observations (and is consistent with the smaller angular scale upper limits discussed above).

Perhaps we are in fact on the verge of determining the nature of the primordial perturbations which gave rise, directly or indirectly, to all the structure we see about us in the Universe.

V. DEDICATION AND ACKNOWLEDGEMENTS

Five years ago, after a colloquium at Yale on a related topic, I sat down with Beatrice Tinsley to discuss how measurements of the kind discussed here might best be planned and made. That hour's conversation had an important bearing on my work over this half decade. I know that there are many colleagues--in many branches of astronomy--who benefitted in similar ways from working with or even just talking with Beatrice Tinsley. All of us lament her death, which occurred during the Rencontre, and all of us admire her courage during the time it closed in on her. To her memory, then, this review is dedicated.

It should be clear from my frequent references to their papers, that I benefitted greatly from the chance to talk with other participants at the Sixteenth Rencontre, especially Joe Silk. I am also indebted to Paul Boynton and George Lake for useful comments.

Many thanks also to my wife, Jane Widseth, who took charge of our son John while I joined my colleagues in sorting out the early history of the Universe; it is not clear whether her task or ours was the more challenging! I am grateful to the organizers of the Astrophysics Meeting of the Sixteenth Rencontre, Jean Audouze, Phil Crane, Tom Gaisser, and Dennis Hegyi, and to the staff of the meeting. Finally, thanks to the anonymous sous-chef of pastry who contributed so much to my enjoyment of Les Arcs.

This paper is a somewhat abbreviated version of a longer review which will appear later in the proceedings of the summer school on the origin and evolution of galaxies, held at the Scuola Ettore Majorana in Erice, Sicily, in May 1981.

VI. REFERENCES

1. Peebles, P. J. E., The Large-Scale Structure of the Universe, Princeton Univ. Press, Princeton, N.J. (1981).
2. Davis, M., A Complete Redshift Survey (preprint)--discussed here by Press (1981).

3. Weinberg, S., Gravitation and Cosmology, John Wiley Co., New York (1972).

4. Lifschitz, E., Journal Physics U.S.S.R., 10, 116 (1946).

5. Sunyaev, R. A., I.A.U. Symposium 79 (ed. M. S. Longair), Reidel Publ. Co., Dordrecht, Holland (1978).

6. Ozernoi, L. M., ibid.

7. Sunyaev, R. A., and Zel'dovich, Ya. B., Astrophys. and Space Science, 7, 3 (1970)

8. Peebles, P. J. E., and Yu, J. T., Astrophys. Journal, 162, 815 (1970).

9. Gott, J. R., and Rees, M. J., Astron. and Astrophys., 45, 365 (1975).

10. Ozernoi, L. M., and Chibisov, G. V., Astron. Zh., 47, 769 (1971)--English version, Sov. Astron. Journal, 14, 615 (1971).

11. Chibisov, G. V., Astron. Zh., 49, 286 (1972)--English version, Sov. Astron. Journal, 16, 235 (1972). See also Chibisov, G. V., and Ozernoi, L. M., Astrophysics Letters, 3, 189(1969).

12. Dautcourt, G., Physica Scripta, 21, 714 (1980).

13. Zel'dovich, Ya. B., Monthly Notices Royal Astron. Soc., 160, 1p (1972).

14. Hogan, C. J., Nature, 286, 360 (1980).

15. Gott, J. R., in Physical Cosmology Les Houches Session XXXII, 564 (1979).

16. Rees, M. J., Proceedings of the International School of Physics Enrico Fermi, Course 47 (ed. R. K. Sachs), Academic Press, New York (1971).

17. Dicke, R. H., and Peebles, P. J. E., Astrophys. Journal, 154, 891 (1968). See also Field, G. B., Stars and Stellar Systems, Vol. 9, Chapter 10, University of Chicago Press, Chicago (1975).

18. Silk, J., Astrophys. Journal, 151, 459 (1968) and I.A.U. Symposium 63, 175 (ed. M. S. Longair), Reidel Publ. Co., Dordrecht, Holland (1974).

19. Jones, B. T., Rev. Modern Phys., 48, 107 (1976).

20. Doroshkevich, A. G., Zel'dovich, Ya. B., and Sunyaev, R. A., Astron. Zh., 55, 913 (1977)--English version, Sov. Astron. Journal, 22, 523 (1978).

21. Zel'dovich, Ya. B., I.A.U. Symposium 79, 409 (ed. M. S. Longair and J. Einasto), Reidel Publ. Co., Dordrecht, Holland (1978).

22. Press, W. H., and Schechter, P., Astrophys. Journal, 187, 425 (1974). See also [1] above.

23. H. Spinrad has reported detection of galactic redshifts of $z \sim 1$ (to be published).

24. Larson, R. B., Monthly Notices Roy. Astron. Soc., 166, 585 (1974).

25. Partridge, R. B., and Peebles, P. J. E., Astrophys. Journal, 147, 868 (1967); ibid 148, 377 (1967).

26. Davis, M., and Wilkinson, D. T., Astrophys. Journal, 192, 251 (1974); Partridge, R. B., Astrophys. Journal, 192, 241 (1974); Koo, D. C. and Kron, R. G., Publ. Astron. Soc. Pacific, 92, 537 (1980); Kron, R. G., Physica Scripta, 21, 652 (1980); and Tinsley, B. M., Philos. Trans. Roy. Soc. London, Ser. A, 296, 303 (1980).

27. Sunyaev, R. A., Astron. and Astrophys., 12, 190 (1971).

28. Gott, J. R., Gunn, J. E., Schramm, D. N., and Tinsley, B. M., Astrophys. Journal, 194, 543 (1974).

29. Silk, J., and Wilson, M. L., Physica Scripta, 21, 708 (1980). See also Wilson, M. L., and Silk, J., "On the Anisotropy of the Cosmological Background I" (preprint).

30. Peebles, P. J. E., "Primeval Adiabatic Perturbations: Constraints from the Mass Distribution" (preprint; submitted to Astrophys. Journal).

31. Wagoner, R. V., Astrophys. Journal, 179, 343 (1973). See also Peebles, Physical Cosmology, Chap. VIII, Princeton University Press, Princeton, N.J. (1971).

32. Barrow, J., Nature 272, 211 (1978). See also Olson, D. W., and Silk, J., Astrophys. Journal, 226, 50 (1978).

33. Greenstein, J. L., Physica Scripta, 21, 759 (1980).

34. Silk, J., Astrophys. Journal, 151, 459 (1968).

35. Anile, A. M., Danese, L., De Zotti, G., and Motta, S., Astrophys. Journal Letters, 205, L59 (1976).

36. Davis, M., and Boynton, P. E., Astrophys. Journal, 237, 365 (1980).

37. Partridge, R. B., Physica Scripta, 21, 624 (1980).

38. Conklin, E. K., and Bracewell, R. N., Nature, 216, 777 (1967).

39. Penzias, A. A., Schraml, J., and Wilson, R. W., Astrophys. Journal Letters, 157, L49 (1969).

40. Boynton, P. E., and Partridge, R. B., Astrophys. Journal, 181, 243 (1973).

41. Carpenter, R. L., Gulkis, S., and Sato, T., Astrophys. Journal Letters, 182, L61 (1973).

42. Parijskij, Yu. N., Astrophys. Journal Letters, 180, L47 (1973).

43. Parijskij, Yu. N., Ast. Zh., 50, 453; English version in Sov. Astron. Journal, 17, 291 (1973).

44. Stankevich, K. S., Ast. Zh., 51, 216; English version in Sov. Astron. Journal, 18, 126 (1974).

45. Caderni, N., Fabbri, R., De Cosimo, V., Melchiorri, B., Melchiorri, F. and Natale, V., Phys. Rev., D16, 2424 (1977).

46. Parijskij, Yu. N., Petrov, Z. E. and Cherkov, L. N., Pis'ma Ast. Zh., 3, 483; English version in Sov. Astron. Journal Letters, 3, 263 (1977).

47. Partridge, R. B., Astrophys. Journal, 235, 681 (1980).

48. Boynton, P. E., I.A.U. Symposium 92, 293 (ed. G. O. Abell and P. J. E. Peebles) Reidel Publ. Co., Dordrecht, Holland (1980).

49. Birkinshaw, M., Gull, S. F., and Northover, K. J. E., "Measurements of the Gas Contents of Clusters II" (preprint; submitted to Monthly Notices Roy. Astron. Soc.).

50. Ledden, J. E., Broderick, J. J., and Condon, J. J., Astron. Journal, 85, 780 (1980).

51. Seielstad, G. A., Masson, C. R., and Berge, G. L., Astrophys. Journal, 244, 717 (1981).

52. Melchiorri, F., Melchiorri, B., Ceccarelli, C., and Pietranera, L., preprint (1981). See also Fabbri et al., Proc. Second Marcel Grossmann Meeting, Trieste (in press) (1979).

53. Boynton, P. E., I.A.U. Symposium 79, 317 (ed. M. S. Longair and J. Einasto), Reidel Publ. Co., Dordrecht, Holland (1978).

54. Solheim, J. E., Mandolesi, N., Sironi, G., and Partridge, R. B., Physica Scripta, 21, 629 (1980); also Sironi et al. submitted to the Paris conference on cosmic rays (1981).

55. Various groups have proposed to use the new 2-cm receiver at the 140-foot telescope to reach $\Delta T/T \lesssim 3\times10^{-5}$.

56. Goldstein, S. J., Marscher, A. M., and Rood, R. T., Astrophys. Journal, 210, 321 (1976); Goldstein, S. J., Turner, K. C., and Rood, R. T., Astrophys. Journal, 229, 455 (1979); Martin, H. M., Partridge, R. B., and Rood, R. T., Astrophys. Journal Letters, 240, L79 (1980); Partridge and Rattner, in preparation.

57. Martin, H. M., in preparation (theory of interferometric observations of fluctuations in the cosmic microwave background).

58. Rees, M. J., in The Evolution of Galaxies and Stellar Populations (ed. B. M. Tinsley and R. B. Larson), Yale University Press, New Haven (1977); Nature 275, 35 (1978).

59. Hogan, C. J., Monthly Notices Roy. Astron. Soc., 192, 891 (1980).

60. Davis, M., Physica Scripta, 21, 717 (1980).

61. Cheng, E.S., Saulson, P. R., Wilkinson, D. T., and Corey, B. E., Astrophys. Journal Letters, 232, L139 (1979); and Gorenstein, M. V., and Smoot, G. F., Astrophys. Journal, 244, 361 (1981); and references therein.

62. Fabbri, R., Guidi, I., Melchiorri, F., and Natale, V., Phys. Rev. Letters 44, 1563 (1980) and ibid 45, 401 (1980); and Boughn, S. P., Cheng, E. S., and Wilkinson, D. T., Astrophys. Journal Letters, 243, L113 (1981).

63. Novikov, I. D., I.A.U. Symposium 63, 273 (ed. M. S. Longair), Reidel Publ. Co., Dordrecht, Holland (1974).

64. Silk, J., and Wilson, M. L., Astrophys. Journal Letters, 244, L37 (1981). See also Peebles, P. J. E., Astrophys. Journal Letters, 243, L119 (1981).

POPULATION III OBJECTS AND THE SHAPE OF THE COSMOLOGICAL BACKGROUND RADIATION

J. Heyvaerts
Observatoire de Meudon - 92190 Meudon
J-L. Puget
Institut d'Astrophysique, 98 bis Bd Arago - 75014 Paris

The spectrum of the cosmological radiation may keep track of non thermal processes having followed the decoupling era, in the form of departures from a strictly Planckian spectrum. In this paper we examine the consequences of energy and metals release by a population of pregalactic objects. The latter are assumed to condense into dust. The formation of the universal spectrum under these conditions is described in a self-consistent manner. It is shown that a good enough agreement can be obtained with presently available data. The conditions necessary for this are discussed and confronted to other observational evidences. It is concluded that the observed spectrum can be explained if the star burst occured before the epoch $z \simeq 30$ and after $z \simeq 300$, with a release varying between 0.3 MeV/nucleon for $z = 30$ and 2 MeV/nucleon at $z = 200$, while the mass fraction in grains vary from 10^{-4} to 10^{-6}. We reach conclusions partly similar to those of an independent work by Negroponte et al.[1] ; our results however point to the possibility that the population III even might have occured recently, at $z \simeq 30$ to 50, this being still consistent with all the considered constraints.

I. INTRODUCTION

Recent measurements of the cosmological radiation by Woody and Richards[2,3] in the millimeter and submillimeter ranges using He^3 cooled polarizing interferometers, have recently added new detailed data to the set of measurements previously made in the centimeter range. These measurements have been described in great detail, and are well known. Their quality is limited at the low frequency end by diffraction phenomena in the entrance aperture of the instrument, roughly below 5 cm^{-1} [3] and at high frequency by the difficulty of substracting a very important atmosphere contribution above, say, 12 cm^{-1}, though the model used is impressively accurate to leave an almost exactly vanishing result when substracted from the data at high frequencies[3]. On the other hand, the discrepancy at 2.64 mm[4] with the CN point clearly shows that more observations are needed below this frequency. Centimeter wave data, on the other hand, are still not very accurate at present because the error bars are often large and not always overlapping The Rayleigh Jeans temperature is approximately 2.72°K ± 0.08[5]. By contrast the Woody Richards data indicate a significantly higher temperature, near 2.96°K. Though it may be premature to claim that a distortion is definitely present, it is still timely to examine which processes may leave such a spectral signature. It is the aim of this paper to examine one of them. A full examination of the problem would require a compilation of all the sources of opacity and energy release in various forms which may feature the "photosphere of the Universe" which we may tentatively define as the period between decoupling and now. A recent review of the subject has been given Sunyaev and Zeldovich[6]. In this paper we take the view that spectral features are radiation transfer effects affecting the universal radiation, while according to certain authors, it could rather be primordial and due to a fundamental deviation from the Planck law[7]. Such ideas seem to imply a reduction in the He^4 production in the Big Bang[8]. In this framework the emissive part in the universal radiation transfer is the almost entirely ununderstood. It is believed that it could be due either to turbulence dissipation, either to the turn on of various emissive objects like quasars, early generations of stars...Condensation of such objects might and perhaps should begin just after decoupling[9]. As far as opacity is concerned, more possibilities have been considered : hot free electrons produced by ionizing radiation of any early emitting objects, like quasars, are a source of Compton opacity and Bremsstrahlung. If heavy elements are present in the universal matter at an early epochthey may condense into dust or assemble in molecules which may also be a very effective source of opacity[9].

The Compton interaction of cosmic photons with hot electrons is a problem

which has by now be considered at length[10,11,12,13,14,15]. It has been concluded from these studies[11] that, if a blackbody at temperature T_R interacts with electrons at a much higher temperature, the brightness temperature decreases in the Rayleigh Jeans part, and increases in the Wien part. The distorsion grows in proportion to a parameter y:

$$y = \int_{t_o}^{t} \frac{k\, T_e(t)}{m_e\, c^2}\, n_e(t)\, \sigma_T c\, dt \qquad (1)$$

and the radiation temperature turns out to be an ever increasing function of frequency. For this reason, the energy fed by the hot electrons into the distortion grows fairly rapidly with y, as e^{4y}[13]. Even for y = 0.1, e^{4y} = 1.5, which means that the reservoir of hot electrons should have lost an energy equal to 50 % of the cosmological photon energy in the process. Feedback effects might then be important, and to our knowledge have not been considered in detail. In a dusty Universe, radiation may be absorbed by dust as well as by atoms, and can be partly converted into infrared photons. Dust dominates only if a fairly high degree of ionization is already realized. Actually,

$$\frac{n_G \sigma_G}{n_{H_o} \sigma_H} = \frac{n_{H_o} + n_{H^+}}{n_{H_o}} \left(\frac{f_G}{10^{-5}}\right) a^{-1} \left(\frac{0.8\ 10^{-25}}{\sigma_H}\right) \qquad (2)$$

σ_H is of the order of the square of a Bohr radius, and then the above ratio is usually very small unless the ionization is quite complete. $a \simeq 10^{-2}$ micron, $\sigma_H \simeq 10^{-16}$ cm^2, $f_G \simeq 10^{-3}$, we obtain :

$$n_G \sigma_G \Big/ n_H \sigma_H \quad \frac{(n_H + n_{H^+})}{n_H}\, 10^{-5}$$

It is then to be expected that most ionizing photons would normally be used in an ionization process, and then converted partly into heat and partly into lower energy fluorescence photons. Nevertheless the opacity of the Universe to these photons is normally higher due to grains than to Thomson scattering. Actually, by the same sort of argument, we get

$$\frac{n_G \sigma_G}{n_H \sigma_T} = \left(\frac{n_{H_o} + n_{H^+}}{n_{H^+}}\right) \left(\frac{f_G}{10^{-5}}\right) a^{-1} \left(\frac{0.8\ 10^{-25}}{\sigma_T}\right) \qquad (3)$$

which may be of the order of 10^4 if $n_{H_o} \ll n_{H^+}$. ($f_G \simeq 10^{-3}$, $a \simeq 10^{-2}$)

II. PLAUSIBILITY OF POPULATION III OBJECTS AND PREGALACTIC DUST

How likely is it that dust exists in the Universe at early redshifts ? Following Rees[9] the arguments in favour of the existence of so called population III objects (a first generation of stars which might have produced some nucleosynthesis prior to galaxy formation) may be listed as follows :

1/ Logical necessity : the isothermal fluctuation spectrum which may have led to galaxy formation should have had a small mass tail which should have collapsed.

2/ Dead remnants of population III might provide the hidden mass in galaxies and clusters. The existence of such a hidden mass seems now behind doubt. Recent observational work indicates that in galaxies this mass is not due to very low luminosity red halo stars[16]. Other explanations might be massive neutrinos or strictly non luminous bodies, like meteorites (but this may pose a severe heavy elements problem). Wesson and Lermann[18] give observational arguments against a significant dielectric dust mass.

3/ Measurements of the X ray spectrum of intergalactic gas in galaxy clusters have shown a feature at 7 KeV typical of iron, which is then deduced to have approximately solar abundance. This cluster gas may have originated either in galaxies or be primordial. It is then important to discuss whether the galactic mass loss may be responsible for the accumulation of this gas in the cluster. The amount of intergalactic matter in a cluster is quite variable from cluster to cluster. To within a factor ten, it seems at present that this mass is comparable to the total mass of galaxies in the cluster. This then raises the question : is it conceivable that galaxies may have exchanged mass with the intergalactic medium at an average rate of 10 $M_\odot$/year ? Moreover this amount of matter should have been processed by stars. The first question is then to know whether stars actually return to the interstellar medium a comparable amount of mass. In the present state of our Galaxy, such a gas ejection seems to be in the realm of possibilities. Cataclysmic objects, like supernovae (1 every 50 years) as well as the some 150 planetary nebulae known cannot provide the required rate. Similarly, the mass loss of a typical star like sun is 10^{-14} $M_\odot$/year, so that gas from normal stars is not more than 10^{-3} $M_\odot year^{-1}$. However high mass loss stars, like giant stars, may have higher production (10^{-8} $M_\odot$/year) but they are much less numerous. Estimating that 1 star among 600 is a giant, we may expect a gas yield of 3 $M_\odot$/year. G and K supergiants produce winds which may be 10 to 1000 times stronger, but they are approximately 1 for 1000 giant stars. Then they would at best double the previous figure. Of course, this does not tell anything concerning the fraction of this mass which may eventually be exchanged with intergalactic medium, but it shows that even in normal galaxies stellar populations can face a high rate of mass exchange. Concerning the galactic mass loss (or mass gain) itself, estimations are quite uncertain. According to Oort[19], high velocity hydrogen

clouds observed at high altitudes are accreting on to the Galaxy, but correspond only to a rate of mass increase of 0.2 $M_\odot$/year, while conversely Mathiews and Bregman[20] insist on the possibility of having galactic winds. In a discussion of galactic winds Bregman put forward the idea that in elliptical and SO galaxies the expected mass accumulated from star loss being absent, a galactic wind should be the main agent removing this gas. Actually intense extraction should make the Galaxy more bluish than observed except if very low mass stars are formed, ablation is ineffective according to him, and cataclysmic ejections at the galactic size are not observed. The wind is produced if the interstellar gas, or part of it, acquires a high enough energy density to escape out of the potential well. If SN rates were much higher for younger galaxies, the existence of such winds would be more probable. Then, present knowledge seems to be consistent with the idea that cluster gas originates in galaxies, but they imply that at the epoch of their formation galaxies have had a higher rate of mass exchange with the intergalactic medium[21].

4/ One may add to this list any evidence for the possible existence of heavy elements in pregalactic gas. Very important in this respect are the observations of metal abundance in very old galactic stars. Metal abundances in globular clusters of the outer halo have been studied by Searle and Zinn[22]. These authors conclude that metal abundances do not show any gradient with distance in the outer halo ($d > 8$ kpc), and in particular do not seem to approach zero. This may be interpreted in terms of a primordial gas containing already some 1 % at most of the solar metal abundance. However these observations certainly do not imply necessarily such an interpretation. They shed however some light on the early period of Galaxy formation, and in particular on the possibility that remote globular clusters could have been formed during a rather extended period of time, of the order of 10^9 years and should have exchanged mass with the universal medium at a high rate. We refer the reader to the original paper for details of the argumentation. We discuss later the observational and physical limitations on the existence and amount of "pregalactic" dust, and translate this into restrictions on the possible values of the parameters of our model.

III. MODELIZATION OF THE INTERACTION OF COSMOLOGICAL RADIATION WITH COSMIC DUST

In this paper, we do not consider the possibility[23] that the cosmological radiation owes its origin entirely to some population III event though this might have some appealing consequences, like giving a natural explanation of the ratio of the numbers of photons and baryons[23]. It is worth mentioning here that this may lose some of its appeal if the photon/baryon ratio finds some fundamental explanation in newly developped thesis of baryogenesis (for example, Weinberg[24])

. Nevertheless the very idea that early collapsed objects may leave a signature in the cosmological radiation in an otherwise standard Big Bang Universe is worth investigating. We then consider the Universe at an epoch when it is still closely Friedmanian, and consider only the case $\Omega \leq 1$. This simplifies the time-redshift relation, which can be written :

(4)

with $m = 5/2$ for $\Omega = 1$ and $m = 2$ for $\Omega < 1$ and $(1 + z) \ll \Omega^{-1}$. After the decoupling between matter and radiation, the Universe is assumed to be filled with perfect blackbody radiation at temperature $T(1 + z)$, with $T = 2.72°K$. We assume the Universe to remain transparent up to some epoch when it becomes filled with dust created by "population III". At the same time, or just after, radiation is released in the Universe. This may be the result of the nuclear reactions which have necessarily taken place. It may nevertheless be checked[25] that the nucleosynthetic energy self-consistently associated to the expected amount of dust is not sufficient to explain the extra energy of the distortion. More material may have been processed, which has remained locked in dead remnants not visible by now, or else these remnants may have accreted matter very soon, and have radiated a substantial part of the released gravitational energy. We do not investigate the exact form in which this energy is produced, but assume that it interacts efficiently with dust. We call it below "star radiation" for the sake of definiteness. We assume that the even is relatively brief so that the details of the time history of the energy release are not critical. The physical reason for the formation of a spectral feature under these conditions is the extra heating of dust, submitted to both cosmological and "optical" radition, which will be converted into infrared radiation in a wavelength dependent manner, provided the Universe is neither completely opaque nor transparent at these wavelengths[25].

To sum up, we enumerate and describe below the parameters of our models :

a - Parameters Ω, H_o (and then m) of the model Universe.

b - A parameter f_G characterizes the amount of dust released by population III.

$$f_G = \frac{\text{number of nucleons in grains}}{\text{total number of nucleons}} \qquad (5)$$

With this definition, if a is the size of grains, f their density, the number density of baryons n_o at present epoch, we shall have at epoch z a number density of grains :

$$n_g(Z) = (1 + Z)^3 \frac{f_G n_o m_p}{\frac{4}{3} n a^3 \rho_G} \qquad (6)$$

c - Parameters describe the amount and time dependence of the population III energy release. Let the power released at epoch z (or equivalently x = (1 + z) per unit volume be : q(x)dt. If this radiant energy had propagated freely up to now, it would contribute a diffuse radiant energy density :

$$\varepsilon_o = \int_1^\infty \frac{q(x)\, dt}{x^4} = \int_1^\infty \frac{q(x)}{x^4} \frac{dx}{\dot{x}} \qquad (7)$$

We express ε_o as a fraction of the present baryon rest mass energy density

$$\varepsilon_o = n_o f m_p c^2$$

The parameter f describes the amplitude of the energy release and q(x) its history. We can write, considering (7) and (4) :

$$q(x) = n_o f m_p c^2 H_o x^{4+m} l(x)$$

where l(x) is a function with integral unity, which we took to be a Gaussian centered on xIII = (1 + zIII) with width Δz.

Of course q(x) is an emissivity. To get the actual "starlight" radiant energy density in the Universe, we have to solve a transfer problem.

d - Parameters are necessary to describe the properties of dust. We already mentioned grain size and density. We also need to define optical properties. Unfortunately the actual shape of the distorsion may depend on the details of these properties, which are of course unknown because the very nature of the hypothetical dust is not. As we could not think of any reason why they should be of a peculiar composition, we assumed these properties to be similar to those of interstellar dust as compiled by Puget and Serra[26]. Actually the Universe at $z \simeq 100$ ressembles to some extent present interstellar clouds[23]. The absorption efficiency, $Q(\nu)$ scales as λ^{-s} between 30 μ and several mm, with $s \simeq 2$. For "starlight" we assumed the geometrical cross section i.e. $Q(\nu) = 1$. The exact behaviour of $Q(\nu)$ in the optical thickness is substancial. The restriction of considering power law absorption efficiencies limits the validity of our investigation to redshifts such that the peak of cosmological radiation did not exceed 30 , i.e. to

$$z_{III} < 70 \qquad (10)$$

The distortion is energetically less demanding if produced at small redshifts, and for this reason favour "low redshift models". The case of high redshift ($\simeq$ 150 - 250) necessitates the consideration of detailed optical properties of dust and the discussion of various grain nature. These aspects have been considered by Negroponte, Silk and Rowan-Robinson[1], and we refer the reader to their publication for any appreciation of the effect of irregular dust optical properties.

IV. CALCULATION OF DUST THERMAL HISTORY AND DISTORTION

The basic physics of the problem is contained in the transfer equations for cosmological and "starlight" radiation, and the heat balance equation for dust. Let ν^* be the frequency of some radiation I_{ν^*} its specific intensity and T_D^* the temperature of dust. The equation of radiative transfer can be written :

$$\dot{x}\frac{\partial I\nu}{\partial x}+\frac{\nu^*}{x}\dot{x}\frac{\partial I\nu^*}{\partial \nu^*}-3\frac{\dot{x}}{x}I\nu^*=c\pi a^2 n_g(x)Q(\nu^*)\cdot\left(B(\nu^*,T_D^*)-I(\nu^*)\right)+c\frac{q(x,\nu^*)}{4\pi} \quad (11)$$

where $Q(\nu^*)$ is the grain absorption efficiency and $q(x, \nu^*)$ the volume spectral power emissivity of population III objects. The heat balance equation can be written, neglecting dust heat capacity :

$$\int d\nu^*\, 4\pi a^2 Q(\nu^*)\left(\pi I(\nu^*)-\pi B(\nu^*,T_D^*)\right)=0 \quad (12)$$

In these equations $I(\nu^*)$ is meant to represent the intensity at any wavelength, both infrared or of "starlight". It is nevertheless convenient to separate these very different wavelength domains both in their contribution to eq. (11) and (12). Concerning the starlight part, we are interested only in the total output, and may integrate over ν^* in this range in both equations. Let I_s be the integrated specific intensity. We obtain for I_s the equation :

$$\dot{x}\frac{\partial I_s}{\partial x}-4\frac{\dot{x}}{x}I_s=\frac{c\,q(x)}{4\pi}-c\pi a^2 n_g(x)I_s \quad (13)$$

which integrates to :

$$I_s(x)=\frac{c}{4\pi H_0}x^4\int_x^\infty dx'\frac{q(x')}{(x')^{4+m}}e^{-\frac{\tau_0}{4-m}(x'^{4-m}-x^{4-m})} \quad (14)$$

where τ_0 is the standard of optical thickness

$$\tau_0 = c\, n_{g0}\, \pi a^2 / H_0 \qquad (15)$$

The infrared transfer equation is simplified using "comoving" variables :

$$I(\nu^*) = J(\nu)\, x^3 \qquad \text{with } \nu^* = x.\nu \qquad (16)$$

$$B(\nu^*, T_0^*) = x^3\, B(\nu, T_0) \qquad \text{with } T_D^* = x.T_D$$

With this change, the transfer equation can be formally solved for $J(\nu, x)$ in terms of $T_D(x)$. Writing :

$J(\nu, x) = B(\nu, T)\ (1 + \Psi(\nu, x))$

where T is the reference temperature defined above (2.7°K), we obtain :

$$\Psi(\nu, x) = \tau_0 \int_x^\infty q\nu^s\, \frac{e^{h\nu/kT} - e^{h\nu/kT_D(x)}}{e^{h\nu/kT_D(x)} - 1}\, x^{3+s-m} \exp\left[-\tau_0 q\nu^s \frac{(x')^{4+s-m} - x^{4+s-m}}{4+s-m}\right] dx' \qquad (18)$$

This expression can be reported in the heat balance equation and leaves only one integral equation for the dust thermal history, the only one which one needs to solve numerically, putting :

$$u = \frac{h\nu}{kT} \; ; \quad y = \frac{T_D}{T} \; ; \quad \nu_{th} = \frac{kT}{h} \; ; \quad \tau = \tau_0 q \nu_{th}^s$$

$$\beta = \tau/(4+s-m)\, n_s \; ; \quad n_s = (s+3)! \sum_1^\infty n^{-(4+s)} \; ; \quad \mu = \frac{x^{4+s-m}}{4+s-m}$$

This equation can be written :

$$[y(\mu)]^{4+s} = 1 + \alpha(x(\mu)) + \beta \int_\mu^\infty d\mu' [y(\mu')]^{4+2s}\, \phi(y(\mu'))\, \frac{(\mu' - \mu)\tau}{4+s-m}$$

$$\alpha(x) = \frac{c^2}{kT} I_s(x)\, x^{-(4+s)} (2 n_s q \nu_{th}^{3+s})^{-1}$$

$$\phi(y, a) = \int_0^\infty dw\, \frac{w^{3+2s}}{e^w - 1}\, \frac{1 - e^{-w(y-1)}}{1 - e^{-wy}} \exp(-a y^s w^s)$$

ϕ has been computed numerically, and eq. (20) discretized and solved as a non linear triangular system. Once $T_D(x)$ has been so obtained, (18) gives the relati-

ve distortion. A useful and stringent check to the numerical precision of the solution is the conservation theorem, an exact consequence of the equations expressing that energy entered as starlight emerges as distortion :

$$\int d\nu \frac{2h\nu^3}{c^2} \frac{\Psi(x,\nu)}{\exp(h\nu/kT)-1} = \frac{n_0 f m_p c^3}{4\pi} \int_x^\infty q(x')\,dx' \qquad (21)$$

This necessary test is satisfied to within some percent, giving an estimation of numerical precision.

V. RESULTS

A set of ("comoving") dust temperature profile, more exactly of ratio T_D/T versus $x = (1 + z)$ is shown in Fig.1 for a model Universe with $\Omega = 1$, and various values of zIII and optical thickness τ_{III}, taken at the thermal energy ν_{th} between the epoch of the population III burst and now. For large optical thickness temperature profiles look very much like step functions with a slight overshoot after zIII. This is due to the usually high "starlight" optical thickness of the burst. The released energy is then quickly transferred to dust. It takes a certain time for it to achieve equilibrium with cosmological radiation, whence the overshoot. Not surprisingly we observe departures of this behaviour for small optical-optical thicknesses.

The resulting spectral distortion is easily calculated and represented in terms of brightness temperature versus λ^{-1} in Fig.2 for several values of the parameter τ . The general features of these curves, as expected, is to run from T (2.7°K) at long wavelengths because the opacity there is very small up to a value which approaches the equilibrium temperature at small wavelengths, because due to the ν^s dependence of opacity, these are optically thick. The larger, the more extended the frequency band which is at high brightness temperature. The production of a non Planckian spectrum then requires $\tau \approx 1$, because if it is too small, radiation is negligible and if it is too large, the spectrum is thermalized. This is actually the major condition to fulfill. Fig.3 shows that, at constant τ_{III}, the other parameters have very little influence on the shape of the distortion, even if they are varied in large proportions ; the parameter f determines essentially the amplitude of the distortion. Its value is almost uniquely fixed by observations at a value near 10^{-5}, as we discussed below. One satisfactory fit to the data is shown in Fig. 4. As can be seen, the agreement is still not perfect ; this may of course be due to the simplications made in choosing this dust absorption law on one part, and on the other part, to the quality of the data, which has been already discussed. The effect of considering more special laws for Q

has been investigated in some detail by Negroponte et al.[1], who, following the idea of an earlier paper[27], advocate that the distortion mainly reflects the optical properties of dust at higher frequencies, in particular the absorption feature around 10 μ and a decrease of absorption below 10 μ. It is obvious that such a feature creates in the spectrum a region where the brightness temperature returns near 2.7°K, and then mimics the bump seen in the Berkeley group data at higher frequencies. Consistent with their view, Negroponte et al.[1] insist that the population III phenomenon must occur at z = 150 - 200 and that the dust should consist of amorphous silicates or basaltic glasses. Though we do not dispute this conclusion, we disagree with them that $z_{III} > 100$ is compelling, because our present treatment, which is now free of the absence of selfconsistency of our previous paper[28], shows that agreement with data may be reached for z_{III} smaller, of the order of 30, say. This is perhaps our only point of disagreement. Nevertheless, this issue is of some importance. We discuss below the likeliness of model parameters. Before turning to this, it is perhaps useful to remark that in all cases where the relative dust temperature profile, $y(x) = T_D(x)/T(x)$, is like a step function, i.e. almost always except for quite small values of τ_{III}, the resulting spectrum can be calculated analytically from (17, 18). Its expression is given in terms of the (constant) comoving dust temperature T_D after epoch xIII and xIII itself :

$$I(\nu) = \frac{2h\nu^3}{c^2}\frac{1}{e^{h\nu/kT}-1} - \left[\frac{1}{e^{h\nu/kT_D}-1} - \frac{1}{e^{h\nu/kT}-1}\right] \cdot \frac{2h\nu^3}{c^2} e^{-\left(x_{III}^{4+s-m}-1\right)\frac{\tau_0}{4+s-m}} g\nu^s \left(\frac{h\nu}{kT}\right)^{4-m}$$

This expression correctly converges to Planck spectra at T and T_D when $\nu \to 0$ and infinity respectively. T_D is implicitely related to the parameter f by the energy conservation theorem (21). We checked that the above approximation is an excellent one to the results of numerical calculations for $\tau_{III} \geq 0.05$. It does not contain the effect of the small overheating sometimes observed, which happens during a short period corresponding to an optically thin layer and does not affect very much the spectrum.

VI. CONSTRAINTS ON THE MODEL PARAMETERS

It is appropriate now to discuss the limits put by known observational data on the model parameters.

a) Energy

It should first be noted that the parameter f should have quite a well defi-

ned value. Actually the present energy density of cosmological radiation is 0.36 eV cm^{-3}, while that of blackbody radiation at 2.7°K is 0.22 eV cm^{-3}. The 0.14 eV cm^{-3} of the distortion are, according to the definition of f and energy conservation theorem, equal to $n_o f m_p c^2$. Numerically we then have to choose f and Ω in such a way that

$$\Omega f = 2.5\ 10^{-5} \qquad (23)$$

It should be stressed that, though f is independant of the epoch x_{III}, the energy requirements for producing the distorsion do depend on x_{III} because radiative and rest mass energy do not fade out with expansion at the same rate. More precisely an energy :

$$\mathcal{E} = f x_{III} m_p c^2 \qquad (24)$$

is needed per universal baryon at epoch x_{III} to produce the presently observed excess. If nucleosynthetic activity of population III is solely responsible for this, and if H - He transformation is the energy source, this means that, if Ω = 1, a fraction $1.3\ 10^{-3} x_{III}$ of all nucleons should enter into the population III. With x_{III} = 30, this makes 4 % of all nucleons. If x_{III} = 200, this makes one quarter of the universal mass. We argue below that for $x_{III} \approx 30$ a proportion $f_G \leq 2\ 10^{-4}$ should now be visible as dust, which implies that a fraction $\sim 5.10^{-3}$ returns to the gas in the case x_{III} = 30, while for x_{III} = 200 f_G should be $\sim 10^{-5}$ (see below) and a fraction $4\ 10^{-5}$ only will not be locked in dead remnants ; if we admit that the metal yield of these objects is 10^{-2}, this means that a fraction $\sim$ 50 % of the mass should return to the gas in the former case, and $4\ 10^{-3}$ in the other case. Supernovae expell 10 % of their mass. In this specific hypothesis, again, one can estimate that the remnant mass at present epoch should be of the order of $2.8\ 10^8\ x_{III}\ M_\odot$ per $(Mpc)^3$ or else of the order of $x_{III}\ 10^{-32}$ g cm^{-3}. Actually the mass in remnants could be much smaller if accretion powers the energy emission. With a 10 % emission efficiency, we find that a fraction $10\ f x_{III}$ of universal nucleons should be accreted on an undetermined number of dense objects (enough to produce the necessary dust however) namely $3\ 10^{-3}$ for x_{III} = 30 and $2\ 10^{-2}$ for x_{III} = 200. None of these figures seems implausible, though it is noted that, in any case, the hidden mass so produced is not large. The drastic difference of mass and/or metal yield which distinguishes $x_{III} \sim 30$ from $x_{III} \sim 200$ can then unfortunately not be used to judge their respective likeliness.

b) Constraints fixed by the condition $\tau_{III} \approx 1$

The expression for $\tau_{III}(\nu_{th})$ is :

$$\tau_{III} = \tau_o Q(\nu_{th}) \frac{x_{III}^{4+s-m} - 1}{4+s-m} = \frac{q}{3\pi} \frac{1}{4+s-m} \frac{c H_o}{a G \rho_g} q \nu_{H_1}^s \Omega_1 f_g \left(x_{III}^{4+s-m} - 1\right)$$

The main parameters which enter in τ_{III} are f_G, of course, and a high power of x_{III}. For $\Omega = 1$, $4 + s - m = 3.5$. $\tau_{III} \sim 1$ if a non Planckian spectrum is to be observed. This fixes broadly the possible values of x_{III}.

c) Constraints concerning the melting of dust

This is a loose constraint. Assuming that dust melts near 1000°, this puts a limit on the temperature of the Universe, and then on x_{III}, namely

$$x_{III} \leq 330 \left(\frac{T_{melt}}{1000K} \right)$$

d) Compatibility with extreme population III metallicity

In the study of Searle and Zinn[22], the metallicity in globular clusters of the outer halo have been measured. It is remarkable that the metallicity does not show any gradient in this outer halo. As explained by these authors, this may be the result of the conditions of formation of these clusters. They find virtually no value of $\left[\frac{Fe}{H}\right]$ below -2, even for clusters as far as 50 kpc. This may be consistent with a very small amount of primordial metallicity, but it does not imply it. Estimating f_G to be well represented by these figures, we obtain the constraint

$$f_G \lesssim 10^{-2}\ Z_o \sim 2.10^{-4} \qquad (27)$$

e) Absence of noticeable direction dependent dust radio emission

Wesson and Lermann[18] demonstrated the detectability of dust immerged in a hot plasma, as is the case for intergalactic dust, if it exists. Due to cluster formation, it is probable that this dust should have acquired by now an inhomogeneous repartition.It is kwown that electrons captured onto a grain maintain a small asymmetric charge distribution if the grain is dielectric, while grains have some random distribution of their angular momentum. The rotation of the electric dipole causes centimetric radio emission[29], whose inhomogeneity should be as small as 10^{-3}. The argument can be used to derive an upper limit on intergalactic dielectric dust. According to them, the mass density in this form of dust in the intergalactic space cannot exceed several 10^{-33} g cm^{-3} at the very best. Compared to the critical density 10^{-29} g cm^{-3}, this sets again a limit $f_G \lesssim$ some 10^{-4}.

f) Visibility of remote quasars

The most distant quasar observed to date is at a redshift z of approximately 3.5, i.e. x = 4.5. The amount of dust in the Universe should not obscure the optical observations up to that distance. A severe limit is obtained assuming that the absorption cross section in this domain is the geometrical one. The optical-

optical thickness between now and x under these conditions is at most :

$$\tau(x) = \frac{1}{4-m} \frac{q}{32\pi} \frac{cH_0}{aG\rho_4} \Omega f_4 (x^{4-m} - 1)$$

Simultaneously τ_{III} should be approximately 1, which may be written :

$$1 \sim \frac{10^{-4}}{4+s-m} \frac{q}{32\pi} \frac{cH_0}{aG\rho_4} \Omega f_4 (x_{III}^{4+s-m} - 1)$$

Taking the quotient of those expressing $\tau(4.5) \leq 1$, we obtain a condition on x_{III} which depends virtually on nothing (except the dust opacity law, implicit in these relations)

$$x_{III} \geqslant 32.6 \qquad (30)$$

If it turned out that quasar detectability is limited by extinction, this would provide a measurement of f_G and then of x_{III}.

g) Limit of validity of the present calculations

As mentioned above, these calculations cannot be pushed behind $x_{III} \approx 70$, for then the power law approximation for Q_{abs} breaks down. This however does not imply that the whole idea breaks down, on the contrary[1].

Most of this discussion can be summarized by a regioning of the f_G-x_{III} parameter plane showing which regions contradict some observational evidence, as well as those regions where a reasonable representation of the distortion is to be expected (Fig. 5).

VII. CONCLUSIONS

From our own present and preceeding work, as well as from that made independently by Negroponte, Rowan-Robinson and Silk, we conclude that the idea of cosmological radiation distortion by early dust makes sense. Negroponte et al. have shown that the effect of silicate bands at z near 100 or 200 can explain it, and we stress here that even a smoothe absorption law can lead to acceptable predictions, thus allowing the population III burst to occur more recently in the past. These conclusions are now based on consistent mathematics and in our case on a stringent test of numerical precision. We believe that the parameters involved in these models do not conflict with any observational constraint known to us, as we demonstrated in the last paragraph.

REFERENCES

1) Negroponte,J., Rowan-Robinson,M., Silk,J., 1980, preprint
2) Woody,D.P., Richards,P.L., 1979, Phys. Rev. Letters, 42,295
3) Woody,D.P., Richards,P.L., 1980, preprint Lawrence Berkeley Lab., 10342, submitted to Ap. J.

4) Thaddeus,P., 1972, An. Rev. Astron. Astrophys., 10,305
5) Wilkinson,D., 1980, Physica scripta, 21,606
6) Sunyaev,R.A., Zeldovitch,Y.B., 1980, An. Rev. Astron. Astrophys., 18,537
7) Jacobsen,H.P., Kon,M., Segal,J.E., 1979, Phys. Rev. Letters, 42,1788
8) Mathews,G.J., Alhassid,Y., Fuller,G., 1980, Preprint Kellog Lab, OAP 606
9) Rees,M.J., 1978, Nature, 275,35
10) Weyman,M., 1966, Ap. J., 145,560
11) Zeldovitch,Y.B., Sunyaev,R.A., 1969, Astrophys. Sp. Sc., 4,301
12) Zeldovitch,Y.B., Sunyaev,R.A., 1970, Astrophys. Sp. Sc., 7,20
13) Zeldovitch,Y.B., Illarionov,A.F., Sunyaev,R.A., 1972, JETP, 35,643
14) Chan,K.L., Jones,B.T., 1975a, Ap. J., 195,1
15) Chan,K.L., Jones,B.T., 1975b, Ap. J., 198,245
16) Hegyi,D.J., 1980, Moriond Meeting proceedings
17) Peebles,P.J.E., 1971, Physical Cosmology, Princeton U press
18) Wesson,P.S., Lermann,A., 1976, Astron. Astrophys., 53,383
19) Mihalas,D., 1970, Stellar Atmospheres, W.H. Freeman, San Francisco
20) Oort,J.H., 1970, Astron. Astrophys. 7,381
21) Mathews, W.G., Baker,J.C., 1971, Ap. J., 170,241
22) Bregman,J.N., 1978, Ap. J., 224,768
23) Vigroux,L., 1977, Astron. Astrophys. Letter, 56,473
24) Searle,L., Zinn,R., Ap. J., 225,357
25) Rees,M.J., 1980, Physica Scripta, 21,614
26) Weinberg,S., 1980, Physica Scripta, 21,773
27) Puget,J.L., Heyvaerts,J., 1980, Astron. Astrophys., 83,L10
28) Audouze,J., Lequeux,J., Masnou,J.L., Puget,J.L., 1980, Astron. Astrophys., 80,276
29) Puget,J.L., Serra,G., 1981, in preparation
30) Spitzer,L., 1968, Diffuse Matter in Space, J. Wiley Inc., N.Y.

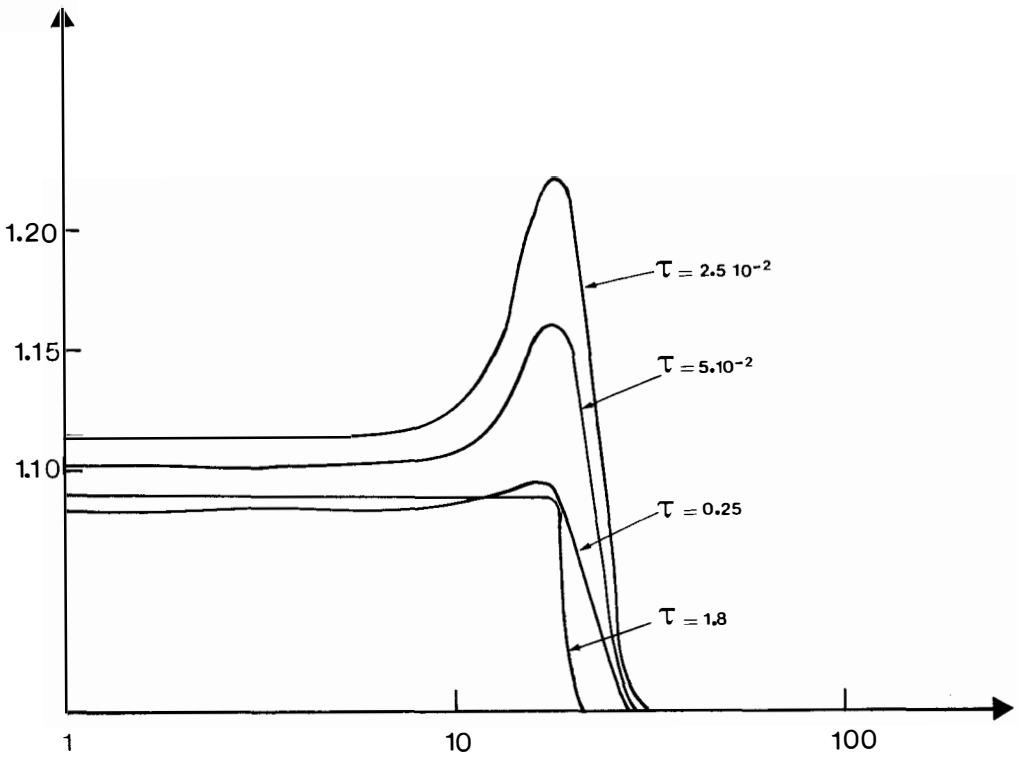


Fig 1

FIGURE 1 - The ratio of the dust temperature to the radiation temperature is given as a function of red-shift for different values of the optical depth between the present time and the population III era $_{III}$.
z_{III} = 20, f = 1.5 10^{-5} and Z = 5

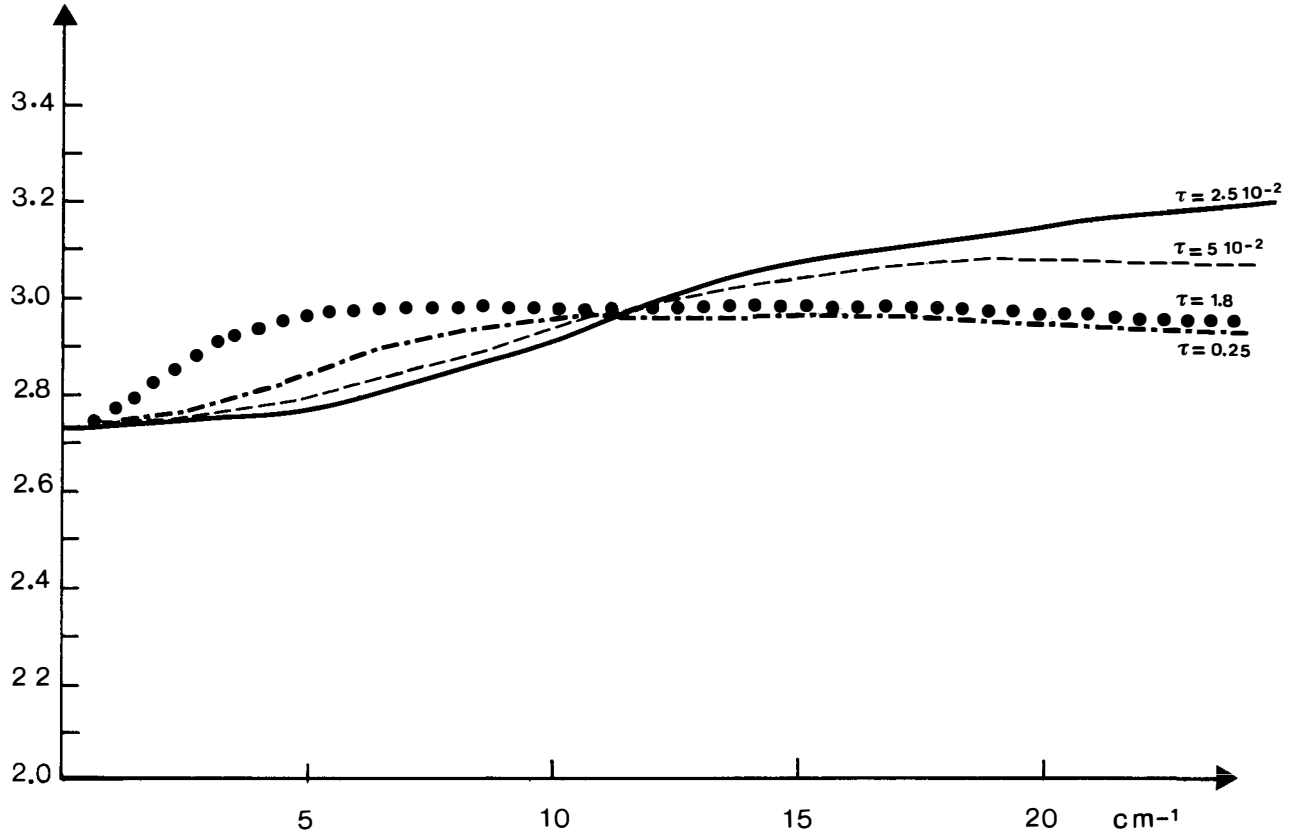


Fig 2

FIGURE 2 - The brightness temperature of the distorted spectrum is given as a function of wave number for the same value of $_{III}$ as in Fig. 1.

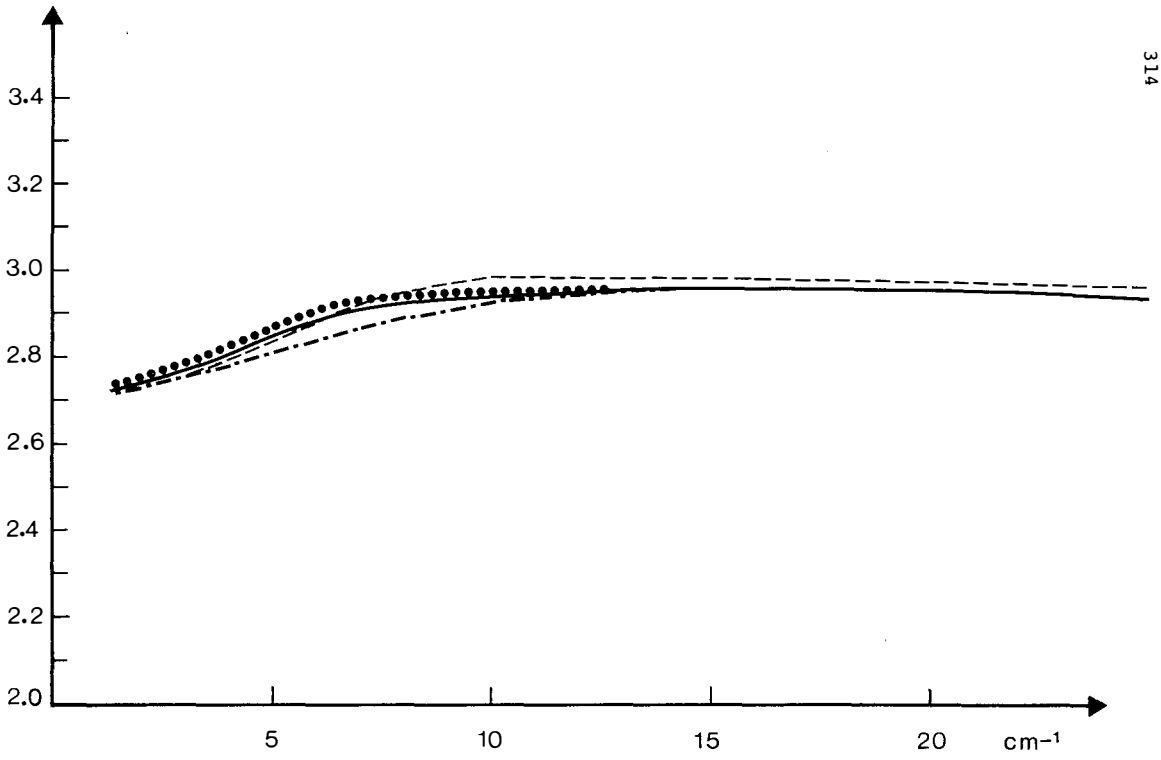


Fig 3

FIGURE 3 - To show the effects of the other parameters on the spectrum three cases are compared with the "nominal" case : full line :

full line : $_{III} = 0.25$, $s = 2$, $Z_{III} = 20$, $Z = 5$
dashed line : $_{III} = 0.25$, $s = 3$, $Z_{III} = 20$, $Z = 5$
dotted line : $_{III} = 0.25$, $s = 2$, $Z_{III} = 20$, $Z = 2$
dot-dash line : $_{III} = 0.25$, $s = 2$, $Z_{III} = 20$, $Z = 10$

Changing Z_{III} but leaving $_{III}$, s and unchanged gives negligeable changes in the final spectrum.

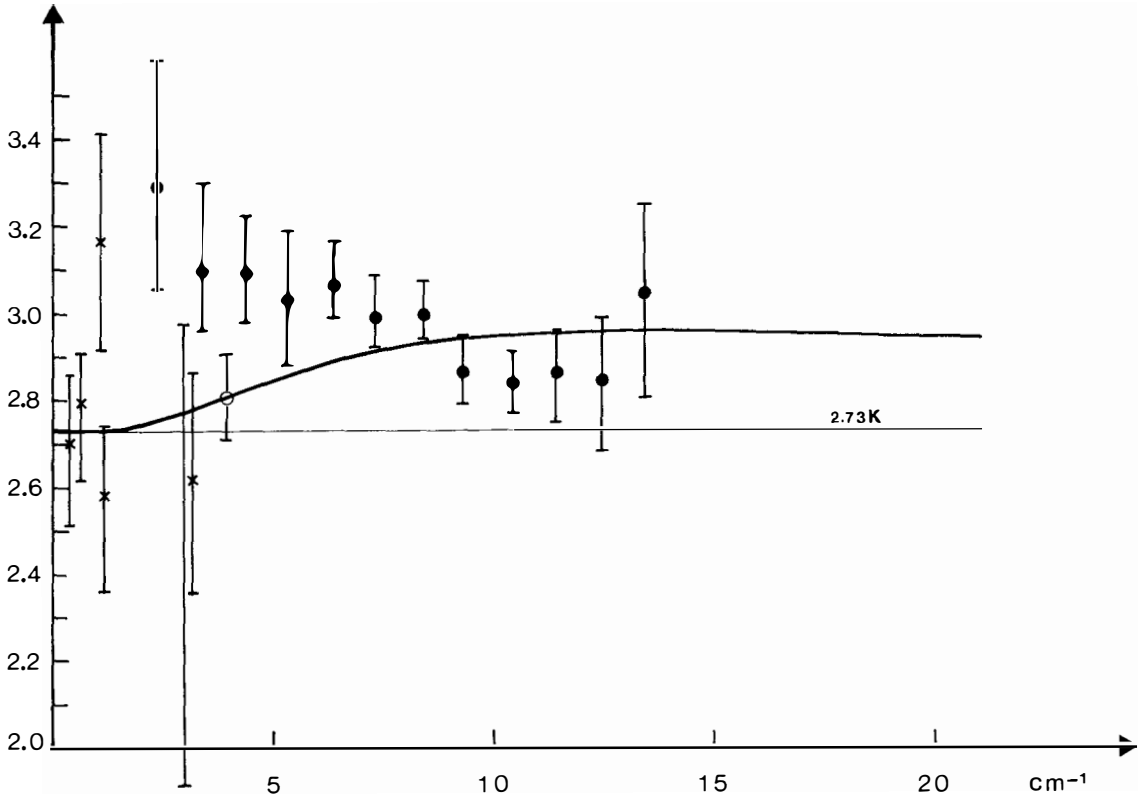


Fig 4

FIGURE 4 - The nominal case $\tau_{III} = 0.25$, $z_{III} = 20$, $Z = 5$, $s = 2$ is compared with the data.
full dots : balloon born spectrometer data of Woody and Richards
open circle : CN excitation
crosses : ground based centimetric data taken from the compilation of Woody and Richards.

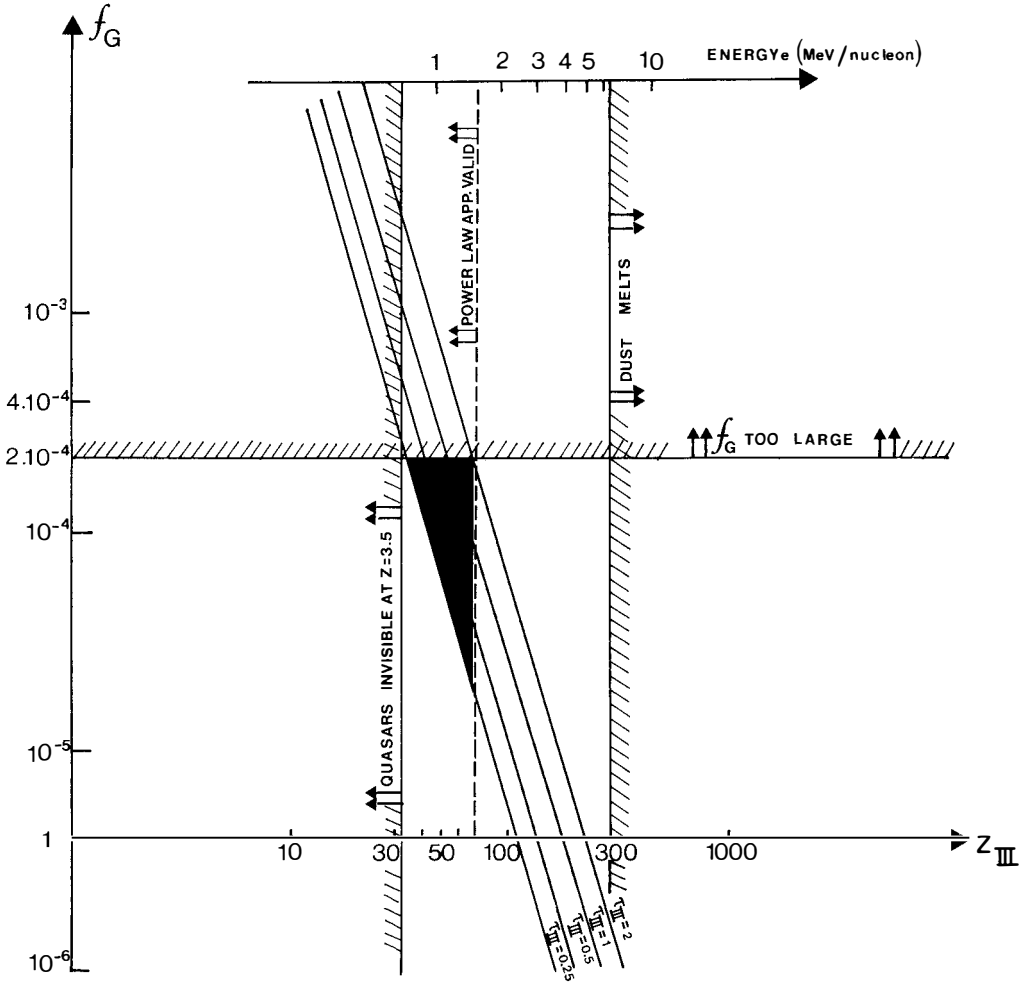

f_G
1
2
3
4
5
10
ENERGYe (MeV/nucleon)
POWER LAW APP. VALID
DUST MELTS
f_G TOO LARGE
QUASARS INVISIBLE AT Z=3.5
10^{-3}
4.10^{-4}
2.10^{-4}
10^{-4}
10^{-5}
10^{-6}
1
10
30
50
100
300
1000
Z_{III}
$\tau_{III}=0.25$
$\tau_{III}=0.5$
$\tau_{III}=1$
$\tau_{III}=2$

Fig 5

DISTORTION OF THE MICROWAVE BACKGROUND BY DUST

Michael Rowan-Robinson
Department of Applied Mathematics, Queen Mary College,
Mile End Road, London E1 4NS.

ABSTRACT

The Woody and Richards distortion of the microwave background has a natural explanation within the framework of the isothermal density fluctuation picture. A pregalactic generation of "stars" makes light and metals. The latter are able to condense into dust grains at a redshift $\sim$ 150-225, which then absorb the starlight and reradiate it in the infrared. At the present epoch we see this emission redshifted into the millimetre range of the spectrum.

When the Woody and Richards (1979) spectrum of the cosmic microwave background was first announced it was extremely puzzling. The distortion from a blackbody spectrum was quite unlike what had been expected due to heat input before the epoch of recombination. This heat input would arise, for example, from the dissipation of lower mass-scale perturbations in the adiabatic fluctuation picture. However it turns out the distortions have a very natural explanation within the isothermal fluctuation scenario.

If at recombination a power-law spectrum of isothermal density fluctuations is present, there is no reason why this spectrum should not extend down to masses far below that of galaxies. Fluctuations on a scale slightly larger than the Jeans mass at recombination ($\sim 10^6 M_\odot$) will collapse rapidly ($\sim 10^6$ years). It is natural to suppose that these will fragment and form an early generation of "stars", Population III. The "stars" might be normal stars of up to 100 $M_\odot$, supermassive stars ($M > 100\ M_\odot$) or black holes. All three types of object may be expected to produce visible and ultraviolet radiation. The normal and supermassive stars will give a yield of helium and heavy elements. This yield is hard to estimate because of the dominant effect mass-loss can have on the evolution of massive stars (Humphreys and Davidson, 1979, de Loore et al, 1977). Clearly the net yield of helium has to be less than or equal to the observed primordial value, $Y \leq 0.25$ ($<< 0.25$ if we stick to the conventional Big Bang nucleosynthesis picture). The yield of metals should be $Z \lesssim 10^{-4}$ for consistency with Population II metals, and would have to be $\lesssim 2 \times 10^{-5}$ if the metals are in the form of dust grains spread uniformly through intergalactic space, for consistency with upper limits on quasar reddening (McKee and Petrosian, 1976).

The approach taken by Rowan-Robinson et al (1979) is to assume that Population III makes light and dust on a reasonably short time-scale ($<< 10^8$ years) and then to consider the consequences for the microwave background spectrum. We suppose that there is a substantial enhancement of the energy-density of the cosmic background radiation due to Population III (by 20-40%, say, to correspond to the energy in the Woody and Richards distortion). This requires that a substantial fraction of the matter in the universe has to undergo processing in Population III, unless the efficiency with which matter is converted to radiation is very much greater than the $\sim 1\%$ expected from thermonuclear processes. A second consequence of the high energy-density enhancement factor, β, is that dust grains cannot form until an epoch, z_f, which depends on the absorption efficiency and melting temperature of the grains, but is typically in the range 150-225.

Remarkably, this is just the range of redshift needed to shift the 10μ silicate feature into the millimetre region of the spectrum and produce the type of distortion observed by Woody and Richards. Our model does not depend on the

details of how the radiation is produced, provided it is produced by epoch z_f, or soon after. In detailed calculations, in which the radiative transfer equation is solved exactly, Negroponte et al (1981) find that if the period during which Population III radiation is produced extends more than about 10^7 years after the epoch z_f then the distortion starts to be washed out.

Once the dust grains form they rapidly absorb the starlight and re-emit in the infrared. At any particular locality we would see the grain formation front moving outwards at the speed of light. Once the dust optical depth to the horizon is greater than unity the intensity of the visible and ultraviolet background light is quickly quenched. The temperature of the dust drops rapidly (Fig. 1) to a value slightly higher than that of the primordial background and thereafter obeys

$$T_d(z) = T_d(0)\ (1 + z)$$

due to the expansion of the universe.

Negroponte et al (1981) have calculated the temperature history and resulting distortion of the background spectrum for several types of silicate for which optical constants are available (obsidian, basaltic glass, amorphous and dirty silicates). The parameter β was chosen to give $T_d(0) = 3.0$, which ensures a distortion of amplitude similar to that seen by Woody and Richards (1979). Z is then chosen to give the largest possible value of z_f consistent with an assumed grain condensation temperature of 1500 K. The primordial background is assumed to have a temperature 2.7 K at the present epoch. For $\Omega = 1$, we find $\beta \simeq 0.2 - 0.4$, $Z \simeq 2.5 - 7.5 \times 10^{-6}$, $z_f = 150 - 225$. For $\Omega = 0.1$ the parameters are similar, except that Z is about a factor 4 higher. An example of the resulting background spectrum is shown in Fig. 2, a plot of the brightness temperature against wavelength. The broken curve shows the $\pm 1\sigma$ range of the Woody and Richards (1979) measurements. The predicted distortion agrees well with that observed. The steep rise in brightness temperature at short wave-

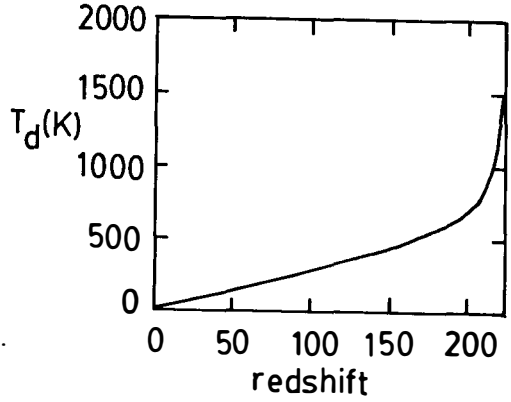


Fig.1: Variation of grain temperature with epoch for obsidian grains (from Negroponte et al, 1981).

lengths is due to the Rayleigh-Jeans tail of the background from Population III stars. This is unlikely to be observable due to foreground radiation from Galactic dust. A natural question to ask is whether the whole background could be produced in this way (Rees 1978, Carr and Silk 1981). The main problem is the need to thermalisize the radiation out to a wavelength of at least 20 cm. If this is to be done by dust we need $Z \geq 10^{-3}$ and for consistency with extreme Population II metal abundances we must assume that some Population II stars were formed before Population III (Rees, 1981, personal communication). The optical depth in dust at 1 mm would be ~ 100 so there would be no prospect of explaining a Woody and Richards type of distortion. Carr (1980) suggests thermalisation by free-free absorption, but the effectiveness of this has yet to be demonstrated in a radiative transfer calculation.

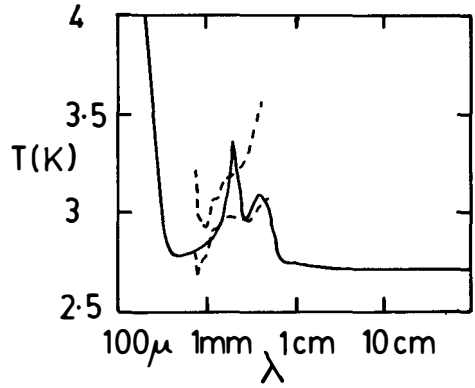


Fig.2: Microwave background spectrum for obsidian grain model ($\Omega = 1$, $1+z_f = 225$, $Z = 5 \times 10^{-6}$, $\beta = 0.42$).

Finally we can ask what the Population III stars must have been like for our model to work. Stars of mass $\lesssim 20\ M_\odot$ do not generate their light on a short enough time-scale to produce the observed distortion. A rather small mass-fraction of $20 - 100\ M_\odot$ stars would suffice to give the low yield of metals required. Thus most of the light must probably be generated by stars of mass $> 100\ M_\odot$ or black holes.

References

Carr, B.J. and Silk, J., 1981, preprint.
Humphreys, R.M. and Davidson, K., 1979, Astrophys.J., 232, 409.
de Loore, C., de Greve, J.P. and Vanbeveren, D., 1978, Astron.Astrophys., 67, 373.
McKee, C.F. and Petrosian, K., 1976, Astrophys.J., 189, 17.
Negroponte, J., Rowan-Robinson, M. and Silk, J., 1981, Astrophys.J. (in press).
Rees, M., 1978, Nature, 275, 35.
Rowan-Robinson, M., Negroponte, J. and Silk, J., 1979, Nature, 281, 635.
Woody, D.P. and Richards, P.L., 1979, Phys.Rev.Lett., 42, 925.

A LIMIT ON THE STELLAR POPULATION OF MASSIVE HALOS

Dennis J. Hegyi
Physics Department
University of Michigan
Ann Arbor, MI 48109, USA

ABSTRACT

The measured rotational velocities of the edge-on spiral galaxy NGC 4565, coupled with several arguments supporting the spherical symmetry of halos, can be used to determine the space density of the halo mass. We show that if the halo mass surrounding NGC 4565 were contained in a population of M5 stars, the minimum expected surface brightness would exceed our measured halo surface brightness. These observations were made in the I Kron band with the annular scanning photometer.

I. Introduction

Over the last ten years evidence has been accumulating to substantiate the claim that spiral galaxies are surrounded by massive spherical halos.[1),2)] While the evidence to support the existence of massive halos is now quite compelling, little if any direct progress has been made towards the determination of the nature of the halo mass.

We shall show, based on optical observations made with the annular scanning photometer,[19)] that the halo of the spiral galaxy NGC 4565 is not primarily composed of M5 nor of more massive stars. The conclusion is based on the photometry of the faintest M5 subdwarf[3)] available with large parallax (>0.10 arc seconds). If fainter M5 subdwarfs are subsequently discovered then this conclusion will have to be modified.

Our derived limit on the allowed stellar population of halos is interesting because the mass of an M5 star, 0.14 $M_\odot$, is close to the lower mass limit for nuclear burning stars, ~.08 $M_\odot$.[4)] If the halo is composed of 0.14 $M_\odot$ nuclear burning stars, then there cannot be more than a factor of ~2 variation in mass in the nuclear burning portion of the initial mass function of the halo stars. Otherwise, the halo surface brightness would exceed our measured surface brightness. Such an initial mass function for the halo would be quite different from the presently observed mass function in the solar neighborhood.

II. Evidence for Halos

The strongest evidence for massive halos surrounding spiral galaxies comes from direct measurements of the rotation curves of spiral galaxies. Radio[5),6),7)] and optical[8)] observations of over 50 spiral galaxies show that the rotation curves of spiral galaxies are flat or increase slightly outward to the limits of detectability of the HI or optical disk. The fact that the rotation curves are symmetric about the optical center of the galaxies is a demonstration of the stability of the rotation and supports the equilibrium condition,

$$\frac{GM_r}{r^2} = \frac{v^2}{r} \quad , \tag{1}$$

that gravity provides the centripetal force which keeps the observed particles in stable circular orbits. In Equation (1), M_r is the mass within galactic radius r, and v is the observed circular velocity. We have assumed that M_r is spherically symmetric and shall justify that assumption below.

Equation (1) should accurately describe the dynamics on a galactic scale. The left hand side of that Equation, Newton's Law of Gravitation, is at least as accurate as a part in 10^{10} on the scale of the solar system. On the very largest scale, cH_o^{-1}, the radius of the visible universe, Newton's Law of Gravitation also gives order of magnitude agreement. Therefore, it may be expected to work on the intermediate scale of galaxies as well. The right hand side of Equation (1) depends only on the geometrical properties of rotation. Solving Equation (1) for M_r, we find that

$$M_r = 1.45\times10^{12} \left(\frac{v}{250 \text{ km/s}}\right)^2 \left(\frac{r}{100 \text{ kpc}}\right) M_\odot \ . \tag{2}$$

Many spiral rotation curves have been measured out to ~50 kpc and one optical rotation curve extends out to 120 kpc.[8] To obtain information about the halo mass distribution beyond 50 kpc, binary galaxies have been used. The interpretation of the binary data is difficult because of selection effects introduced in the statistical reduction of the data. Nevertheless, the data are consistent with the halo hypothesis. The Turner[9] and Peterson[10] binary samples, which have median projected separations of 50 and 110 kpc, respectively, yield M/L ratios which are proportional to the median separations. This would be expected if the light were centrally concentrated, which it is, and if the halo mass increased linearly with galactic radius.

There is only one argument which appears to contradict the existence of large massive halos. White and Sharpe[11] have numerical models of pairs of interacting galaxies showing that if the centers of the galaxies are closer than the radius containing half the mass, the galaxies merge in less than one orbital period. Based on the statistics of binary galaxies, they argue that binaries cannot be merging so rapidly, so that perhaps halos are not very large or not universally present. It is possible to construct a simple analytical model of a galaxy moving through the halo of a companion which gives similar results, so there is little reason to doubt the calculation. But one may question whether the initial conditions used for the models accurately describe the actual physical situation. It is the relative velocity of one galaxy streaming through the halo of the second which gives rise to dynamical friction. But it is possible that binary galaxies are formed with zero relative velocity between the core of one galaxy and the halo of the other, analagous to rigid body rotation. If binary galaxies formed with these initial conditions, then the White and Sharpe calculation would be expected to overestimate the merger rate. Of course, this possibility requires further support before it may be said to offer a solution to the apparent contradiction with the universality of large halos. In summary, we believe that, taken together, the evidence supporting halos out to at least 100 kpc is much stronger than the opposing evidence.

We have described the halo mass distribution as spherically symmetric. Arguments based on the persistence of the warps in the disks of spiral galaxies have shown that the spiral disks are imbedded in a spherically symmetric potential.[12] A more recent argument due to Van der Kruit,[13] also supporting a spherically symmetric mass distribution, is based on the variation of the scale height of stars perpendicular to the disks of spirals and on the velocity dispersion in the z direction. A third argument based on star counts by Monet, Richstone and Schechter[14] requires that at least one half the mass within the solar circle resides in a spherically distributed component.

We shall use the result that the halo is spherically symmetric in our search for halo optical emission.

III. Optical Observations of Halos

In this section we shall describe the calculation of the surface brightness expected from a halo of M5 stars and compare these results with our observations. We shall use NGC 4565 as a candidate galaxy because it is an edge-on spiral at high galactic latitude and has an accurately measured rotation curve.

Krumm and Salpeter[7] have found a flat rotation curve for NGC 4565 with velocity 253 km s^{-1} out to 11.6 arc minutes. Using Equation (1) and the result that the halo mass is spherically distributed, the mass density at radius r, ρ_r, is

$$\rho_r = \frac{v^2}{4\pi G}\frac{1}{r^2} \quad . \tag{3}$$

The mass per unit area, λ_r, integrated along a line of sight which passes a distance r from the galactic center can be seen to be

$$\lambda_r = 2\int_0^{(R_{max}^2 - r^2)^{1/2}} \rho_r dz = \frac{v^2}{2\pi G}\frac{1}{r}\tan^{-1}\sqrt{(R_{max}/r)^2 - 1} \quad , \tag{4}$$

where R_{max} is the maximum radius of the halo. The number of stars per unit area at radius r, if each star has mass m_*, is

$$N_r = \lambda_r / m_* . \tag{5}$$

Finally, the surface brightness, measured on a linear scale, of a halo composed of stars of absolute magnitude, M_*, and mass, m_*, at distance, d Mpc, plotted in units of the number of m_o magnitude stars per arc sec^2 is

$$\sigma_r = 10^{\frac{m_o - M_* - 11.34}{2.5}} \times \frac{\left(\frac{v}{250\ \text{kms}^{-1}}\right)^2 \frac{2}{\pi} \tan^{-1}\sqrt{(\theta_{max}/\theta)^2 - 1}}{\theta\ d\ m_*} \qquad (6)$$

Here, θ is in arc minutes and θ_{max} is R_{max} expressed in arc minutes.

Perhaps the most difficult part of evaluating Equation (6) is the determination of the absolute magnitude M_* of a characteristic halo star. The problem is that there is an apparent correlation between stellar luminosity and metal abundance. A sequence of decreasing metal abundances going from disk to old disk to halo stars, is a sequence of decreasing luminosities. Most observations[15),16),17)] support this trend but Eggen has counterexamples at ~.6 $M_\odot$ for which he shows that the luminosity of stars of different metal abundances are similar. However, most of the data including that of Eggen's support the correlation between metal abundance and luminosity.

The faintest M5 $[(R-I)_K = 1.29]$ star that we have been able to find has an absolute luminosity $M_I = 10.84$ in Kron's photometric system. It is GL 299.[3)] Using Hoxie theoretical fit to the lower main sequence, we obtain the corresponding mass, $m_* = .14M_\odot$, which is in good agreement with the available binary star data.

We have tried to determine whether future observations might reveal less luminous M5 stars. GL 299 lies about 2 magnitudes below the Hyades main sequence at a distance modulus of 3.1. Figure 2 of Eggen[17)] is an HR diagram showing the spheroid and disk populations of the Galaxy. GL 299 lies on the lower edge of an extrapolation of Eggen's spheroid luminosity distribution, which adds confidence to our tentative conclusion that GL 299 is a faint M5 star. One may ask whether the luminosity of GL 299 is characteristic of M5 halo stars with low metal abundances. There is some unpublished work by Vanden Berg[18)] indicating that halo stars with metal abundances as low as 10^{-5} $Z_\odot$ would have luminosities which are similar to the luminosity of faint stars

from the spheroid of the Galaxy. At log T_{eff} = 3.5, which is close to M5, changing the metal abundance from $Z_\odot$ to 10^{-5} $Z_\odot$ results in a 2 magnitude decrease in luminosity, about the same as the spread observed by Eggen. Based on Vanden Berg's models, it appears that we may be able to estimate Population III luminosities reasonably well.

Using GL 299 as our standard M5 star we have plotted in Figure 1 the expected surface brightness for a halo surrounding NGC 4565 in which all of the interior mass is contained in M5 stars. Beyond 4 arc minutes a small fraction of the total mass is contained in the visible disk. The dashed line in Figure 1 is a plot of the expected halo surface brightness for a halo extending out to 150 kpc, or 28 arc minutes, while the solid curve corresponds to a halo of 62 kpc or 11.6 arc minutes. The halo must extend out to at least 11.6 arc minutes since Krumm and Salpeter have measured a rotation curve for NGC 4565 which is flat out to that distance. We have taken the galaxy to be at 18.4 Mpc[1] based on its group association and an H_o = 50 kms^{-1} Mpc^{-1}.

Also plotted in Figure 1 are observations of the halo of NGC 4565 taken with the annular scanning photometer. The data were taken with the McGraw-Hill 1.3 m telescope in February and March 1979. The centers of the circular scan paths C and D are 10 arc minutes from the galaxy's center along a line perpendicular to the galactic disk, which is at position angle 134°. Position C is to the northeast, and D is to the southwest. The radius of the scan is 9.0 arc minutes and a 30 arc second circular aperture was used. The data in Figure 1 containing 51,000 scans, has been normalized using a relatively starless sky area close to the galaxy to correct for sky gradients and telescopic effects. This procedure has been discussed in earlier work.[19),20)] The curve fit to the data is the de Vaucouleurs surface brightness law.

Returning to Figure 1 again, the abscissa is the galactic radius and the ordinate is halo surface brightness in the Kron I band plotted in units of the number of stars of 25.34 magnitude/arc sec^2. On the right side of the figure is the tick mark labelled 10^{-3}. The sky brightness averaged 19.7

magnitudes/arc sec^2 in I Kron during our observing run, and the tick mark refers to 10^{-3} of the sky brightness. Data points have been chosen from regions which are uncontaminated by foreground stars and background galaxies. The error bars are due to photon statistics.

As may be judged from a comparison of the expected surface brightness for a halo of M5 stars and the data, it is highly unlikely that a significant fraction of the halo mass could be contained in M5 stars. No systematic effects that appear plausible would decrease our measured surface brightness though many systematic effects would tend to increase the measured halo signal. Our plotted data has been corrected for known systematics. While our observations exclude the faintest M5 stars, M7 and M8 stars are sufficiently faint so that they cannot be ruled out.

These results may be compared with those of other workers. Hohlfeld and Krumm[21] have ruled out M0 or more massive stars based on J band observations of NGC 4565. Dekel and Shaham[22] have calculated the surface brightness expected for a halo of NGC 4565 under a variety of assumptions. However, as they point out, it is difficult to deduce firm, model independent conclusions. Bahcall and Soneira[23] have used star counts to obtain information about the Galaxy's halo. Though their results may be model dependent, they rule out a halo of M6 or brighter stars.

Some effort has been expended searching for color gradients in the halo as a way of placing limits on a stellar halo component. The problem with this method is that a few giants located in the spheroidal component of a galaxy easily dominate the surface brightness emitted by a faint halo composed of low mass stars. Consequently, searching for color gradients[20),24)] is a less satisfactory way of obtaining information about a stellar halo component. For example, our earlier color gradient measurements[20)] were not able to rule out M5 stars.

I wish to thank Garth Gerber for his assistance with the observations and data reduction and the AFOSR for partial financial support.

REFERENCES

1. Faber, S.M. and Gallagher, J.S. 1979 Ann. Rev. Astron. Astrophys. 17, 135.

2. Peebles, P.J.E. in "Physical Cosmology", ed. R. Balian, J. Audouze, D. Schramm (North Holland, NY 1980).

3. Veeder, G.J. 1975 A. J. 1056.

4. Straka, W.C. 1971 Ap. J. 165, 109.

5. Rogstad, D.H. and Shostak, G.S. 1972 Ap. J. 176, 315.

6. Bosma, A. 1978 PhD thesis Groningen Univ., Groningen.

7. Krumm, N. and Salpeter, E.E. 1979 A. J. 84, 1138.

8. Rubin, V.C., Ford, W.K., Jr. and Thonnard, N. 1978 Ap. J. 225, L107.

9. Turner, E.L. 1976 Ap. J. 208, 304.

10. Peterson, S. 1978 PhD thesis Cornell University, Ithaca.

11. White, S.D.M. and Sharpe, N.A. 1977 Nature 269, 395.

12. Saar, E.M. 1978 In IAU Symposium 84, "The Large-Scale Characteristics of the Galaxy, ed. W.B. Burton.

13. Van der Kruit, P.C. (preprint).

14. Monet, D.G., Richstone, D.O. and Schechter, P.L. 1981 Ap. J. (to be

15. Hartwick, F.D.A., Crampton, D. and Cowley, A.P. 1976, Ap. J. 208, 776.

16. Mould, J.R. and McElroy, D.B. 1978 Ap. J. 220, 935.

17. Eggen, O.J. 1979 Ap. J. 230, 786.

18. Vanden Berg, D. (private communication).

19. Hegyi, D.J. and Gerber, G.L. 1977 Ap. J. (Letters) 218, L7.

20. Hegyi, D.J. and Gerber, G.L. 1979, Proceedings of a Conference on Photometry, Kinematics and Dynamics of Galaxies, ed. D. Evans, Univ. of Texas, Austin.

21. Hohlfeld, R.G. and Krumm, N. (preprint).

22. Dekel, A. and Shaham, J. 1978, Astron. Astrophys. 74, 186.

23. Bahcall, J.N. and Soneira, R.M. 1980 Ap.J. (Letters) 238, L17.

24. Spinrad, H. Ostriker, J.P., Stone, R.P.S., Chin, L-T.G., and Gustavo, B.A. 1978 Ap. J. 225, 56.

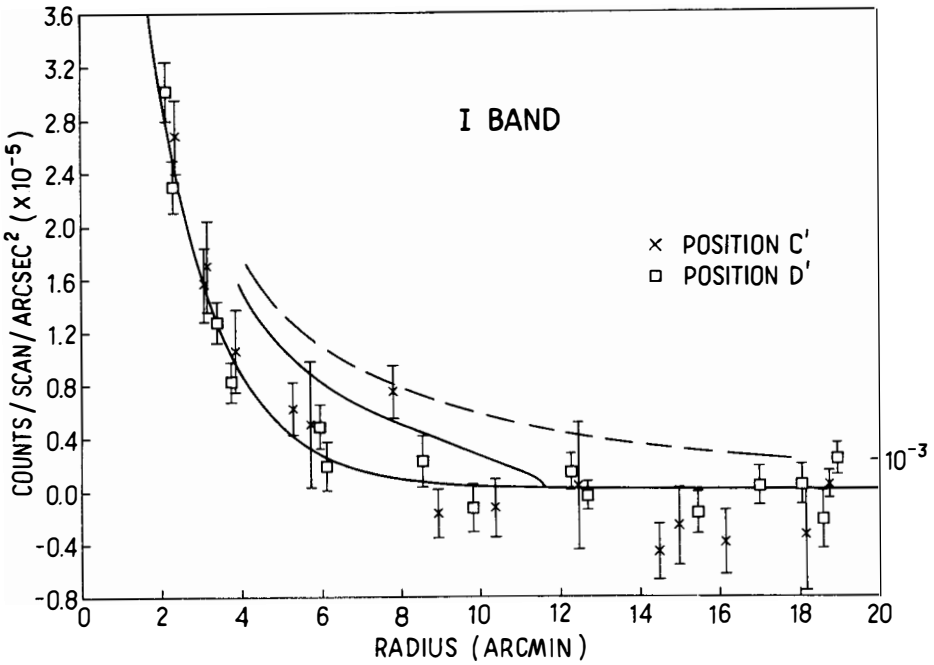
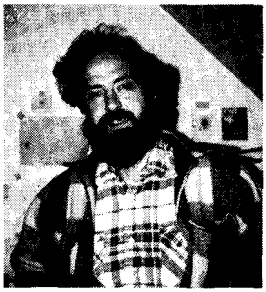


Figure 1. The calculated surface brightness for a halo of M5 subdwarfs and the observations plotted versus galactic radius. The data, in the I Kron band, is for NGC 4565. The dashed line is for a halo of 150 kpc radius, while the solid curve is for a 62 kpc halo. The curve fitted to the data is the de Vaucouleurs's surface brightness law.

DARK MATTER

George Lake
Institute of Astronomy
Cambridge University



ABSTRACT

I review what is known about the form and distribution of dark matter as deduced from the internal dynamics and clustering of galaxies. From their internal dynamics, there are indications that later type galaxies have relatively more dark mass. It is shown that all available evidence argues for the continuity of the galaxian two-point covariance function over a factor of roughly 10^3 in radius ($5h^{-1}$kpc to $5h^{-1}$Mpc) and a factor of 100 in luminosity. Using velocity data on scales from $50h^{-1}$kpc to 10 Mpc and the Cosmic Virial Equation, the deduced value of Ω is found to consistently lie in the range of 0.08 - 0.12. It is further argued that massive neutrinos are not consistent with this result as their maximum Jeans mass (the scale below which perturbations are erased by Landau damping) is $\sim$ 1000 galaxy masses even for a neutrino mass as large as 75eV. Instead of explaining observed features of galaxies, they prevent their formation by any known process.

I. Introduction

A theory of galaxy formation from first principles is greatly hampered by the elusiveness of the dark matter which pervades galactic haloes and clusters of galaxies. As if this isn't bad enough, there is remarkably little information about conditions in the universe at even a moderate redshift of one or two. This makes it exceedingly simple to abandon the standard model and obtain all the observational consequences of the hot Universe theory from cold initial conditions (for example Zel'dovich and Starobinsky 1976, Carr 1977).

The standard model has unmistakably aesthetic appeal and concrete successes - predictions of the microwave background, the helium abundance and perhaps now the entropy per baryon. Even more recently, the possibility that the neutrino has a finite mass of order 10^{-8} m_p has led to numerous discussions of its viability as the dark matter.

I will review what is known about the form and the distribution of the dark matter from studies of the internal dynamics of galaxies and the clustering of galaxies. I think these features make it exceedingly implausible that the dark mass is neutrinos, though they do indicate that the dark mass is universal and perhaps linked to a very early cosmological epoch.

II. Dark Matter in Galaxies

a) Evidence for larger mass-to-light ratios

The luminosity profiles of spiral galaxies betray two basic components; a central bulge with a profile much like that of an elliptical galaxy surrounded by a disk with an exponential surface density (see Kormendy 1980 for a thorough review). An examination of rotation curves prior to 1973 shows the peak expected from the central bulge and the broad maximum characteristic of the disk surface density. They were extrapolated (with dashed lines) as Keplerian rotation profiles.

Ostriker and Peebles (1973) found that rotationally-supported self-gravitating systems are subject to large-scale bar instabilities. When applied to Sc galaxies with large disk/bulge ratios, this indicated that a large fraction of the mass inside the observed axisymmetric disk must be in a pressure-supported stabilizing "halo".

Rotation curves out to 5 disk scale lengths now exist for a large number of galaxies, and show constant or weakly rising velocity with radius (see Faber and Gallagher 1979; hereafter F6).

In Sc galaxies, this yields a mass-to-light ratio (M/L_B) of ~ 5, whereas stellar population models (Larson and Tinsley 1978; hereafter LT) suggest a value of 2.

b) Where is the Mass ?

One alternative for the observed increase in mass with radius has been a changing mass-to-light radius in the disk. The absence of strong color gradients have long suggested that the mass-to-light ratio should be constant with radius (see for example, Freeman 1970), but there are even better dynamical arguments.

The outer regions of galactic disks are often observed to be warped, appearing as hat brims when viewed edge-on. To prevent the precession time from being too short requires a potential which must be more spherical than that resulting from a disk (Saar 1978, Tubbs and Sanders 1979). A more detailed analysis of bending waves (Bertin and Mark 1980, Lake and Mark 1980) indicates that the mass-to-light ratio is a constant. This is derived from a relationship between the height of bending and the surface density applied to our galaxy, NGC 2841 and M33; the three galaxies with adequate data for a comparison.

Van der Kruit (1980) has examined the scale height of the neutral hydrogen in the edge-on galaxy NGC 891. The observed constancy of velocity dispersion with radius in face-on galaxies enables a calculation of the local surface density once the scale height is known. Van der Kruit finds a constant mass-to-light ratio (M/L_B) for the disk population of 3, about the same as the stellar population of LT but considerably different from the total M/L_B of 8.2 (FG).

A final method of constraining mass-to-light ratios in the disk population is to study the bar instability. The work of Ostriker and Peebles (1973) has been challenged on the grounds that it neglects the effects of resonances, anisotropic dispersion velocities and strongly varying velocity dispersions with radius (for recent discussions see Toomre 1977 and Lake and Ostriker 1980). Efstathiou, Negroponte and I (1981) have avoided most of these complications by simulating disks with exponential profiles in halos which reproduce the observed rotation curves. Our results yield a limit to the mass fraction which is in the disk (this is an upper limit for galaxies without bars, a lower limit for barred galaxies), which with adequate photometry yields a limit to the mass-to-light ratio for the disk. We find $M/L_B \sim 1$ for the disks of Sd galaxies, in agreement with LT.

c) Universality of Dark Matter

Are all galaxies comprised of similar admixtures of luminous and dark material? Galaxies with rotation velocities between roughly 80 km/sec and 275 km/sec come in both barred and unbarred flavors. This indicates that, in the mean, they are only marginally stable which would imply that M(bulge & halo)/ M(disk) is constant inside a few disk scale lengths. Since the later type

galaxies of this group (Sbc, Sc, Scd) have larger disk/bulge ratios, this may be an indicator that they have relatively more dark mass.

Galaxies of still later Hubble type (Sd, Sdm, Sm) with rotation velocities $\sim$ 75-80 km/sec are nearly all barred. This may be due to a lesser halo mass, or due to a larger gas fraction; the criterion for gas stability being slightly different than for stars. A further ambiguity arises in that these smaller galaxies are nearly all satellites of larger galaxies.

Flat rotation curves have also been observed in SO galaxies (cf. Illingworth 1981 and references therein). The use of absorption lines for stellar rotation velocities greatly limits the radial extent of the measurements; but is indicative of a changing mass-to-light ratio with distance.

In elliptical galaxies, there is still greater ambiguity. Davies (1981) finds velocity dispersions decreasing with radius, indicating a constant mass-to-ratio. Strangely, it seems that the most flattened ellipticals observed (e.g. NGC 4473 and NGC 4697) have flat velocity dispersion profiles. Perhaps anisotropies are accentuated in a collapse through a halo in the same manner as rotation (Fall and Efstathiou 1980).

In the large dominant galaxies in clusters of galaxies (designated cD galaxies), enormous changes in the mass-to-light ratios are seen, (see Matthews (1978) and Dressler (1980) for discussions on M87 and the cD in Abell 2029, respectively). The results on these objects are more indicative of cluster dynamics (stripping, cannibalism) than of any intrinsic trend of M/L with luminosity.

Tinsley (1980) has also concluded that the relative amount of dark mass must be greater in late-type (bluer) galaxies. Her results depend on stellar population models and rather uncertain corrections for the mass in gas. She corrects for gas mass by examining the quantity $M_{luminous}/(M_{tot} - 2M_{HI})$, where M_{HI} is the mass of netural hydrogen, the 2 is a rough correction factor for molecular gas and $M_{luminous}$ is calculated from stellar population models and total luminosity. If one examines instead $(M_{luminous} + 2M_{HI})/M_{tot}$, the trend seen by Tinsley nearly vanishes. This seems like a more reasonable quantity, since the gas mass is distributed through the disk rather than like the dark matter. The above discussion makes it likely that the general trend is there, only difficult to discern through the uncertainty of population synthesis. (For a discussion of the cause of the observed correlation of M_{HI} with type, see Efstathiou and Lake (1981).)

III Clustering

a) Continuity of the two-point correlation function

The use of the galaxy covariance function $\xi(r)$ (Totsuji and Kihara 1969; Peebles 1974) has provided a simple and powerful discription of the large-scale

distribution of matter (cf. Fall 1979; Peebles 1980). Peebles and Hauser (1974) found that $\xi(r)$ (defined as the fractional increase in the density of galaxies caused by the presence of a galaxy a distance r away) is well-represented by a power law,

$$\xi(r) = (r_c/r)^{\gamma} . \qquad (1)$$

Over distances of $50h^{-1}$ kpc < r < $5h^{-1}$ Mpc (h is the Hubble constant in units of 100 km/s/Mpc), $\gamma = 1.8$ (Peebles and Hauser 1974) and $r_c = (4.2 \pm 0.3)h^{-1}$Mpc (Kirschner, Oemler and Schecter 1979; Peebles 1979).

The limits quoted above for the power law behavior of $\xi(r)$ are set on the large scale by the depth, size, sky coverage and homogeneity of the catalog, and on the small scale by the positional accuracy of the measurements and the sample size (i.e. number of close pairs).

Gott and Turner (1979) have used the accurate binary separation measurements of the Zwicky Catalog (Turner 1976) to provide an estimate of $\xi(r)$ down to separations of $\sim 5h^{-1}$kpc. They find no evidence for a deviation from a power law on this scale. Lake and Tremaine (1980) have used data gathered by Holmberg (1969) on the distribution of companion galaxies around 115 spiral galaxies with known redshifts. The results indicate that the power law found by Peebles and Hauser (1974) is still valid on scales of 5 - 40 h^{-1}kpc and at low luminosities.

In summary, all available evidence is consistent with the remarkably simple model that the galaxy covariance function is a power law over a factor of roughly 10^3 in radius ($5h^{-1}$kpc to $5h^{-1}$Mpc) and is constant over a factor of at least 100 in luminosity.

b) Velocity data - measuring Ω

This is the section which will invite attacks from all sides.

In reading the recent literature one finds arguments for linearly increasing mass-to-light ratios out to 1 Mpc (Ostriker, Peebles and Yahil 1974) and even $\sim$ 10 Mpc (Davis *et al.* 1980). On the other hand, equation (1) and the velocities observed in galactic halos and deep surveys (Kirschner, Oemler and Schechter 1979, Efstathiou *et al.* 1981) are argued to be further evidence for the continuity of clustering from $50h^{-1}$kpc to a few h^{-1}Mpc; the implicit assumption being made that the dark and luminous mass must cluster together (see, for example, Peebles 1981 p. 283-284).

The major problem seems to be the flip-flops between describing systems with relaxed single particle distribution functions and considering the sub-clustering of a clustering hierarchy. A detailed calculation (Davis and Peebles 1977) of the latter yields the cosmic virial equation (CVE):

$$\Omega_G = \frac{<\Delta V^2(r)>}{(715 \text{ km s}^{-1})^2} \left(\frac{4.3 \text{ Mpc}}{hr_c}\right)^{\gamma} \left(\frac{1 \text{ Mpc}}{hr}\right)^{2-\gamma} \quad (2)$$

where V is the one-dimensional velocity dispersion, hr_c is the normalization of the covariance function, r is the relevant scale of observation over Ω_G is the fraction of closure density in clustered objects. I have neglected a factor Q is the equation, which is the ratio of the three-point correlation function to products of the two-point function. Q is found to have values between $\sim 0.8 - 1.1$ depending on the particular sample.

We can apply equation (2) on any scale. On the smallest scale the mean value of the rotation velocity of the galaxies listed by FJ is 210 km s^{-1}. This corresponds to a one-dimensional velocity dispersion of 150 km s^{-1} on a scale of $\sim 50\ h^{-1}$kpc, leading to:

$$\Omega_G = 0.08 \text{ for } r \sim 50\ h^{-1} \text{ kpc}. \quad (3)$$

This number would be decreased by about 15% if I had used the velocity corresponding to an M^* galaxy from Aaronson, Huchra and Mould (1980).

Deep redshift surveys are a powerful method of determining Ω_G. Davis, Geller and Huchra (1978) found $<\Delta V^2>^{\frac{1}{2}} = 320$ km s^{-1} for bright Southern Galaxies, $m_B < 13$. This yields:

$$\Omega_G = 0.15 \pm 0.04 \text{ for } r \sim 1 - 4 \text{ Mpc}. \quad (4)$$

Peebles (1979) analyzing the Kirschner, Oemler and Schechter (1978, hereafter KOS) finds $<\Delta V^2>^{\frac{1}{2}} = 500 \pm 200$ km/sec, yielding:

$$\Omega_G = 0.36 \pm 0.22 \text{ for } r \sim 1 - 4 \text{ Mpc}. \quad (5)$$

Virgocentric flow models may also be used to determine (see Davis et al. 1980). Rescaling the results of Davis et al. to the new more accurate determination of the flow velocity due to Schechter's (1980) analysis of the data of Mould et al. (1980) leads to:

$$0.08 < \Omega < 0.12 \quad \text{for } r \sim 10 \text{ Mpc}. \quad (6)$$

Here the uncertainty is due mainly to the determination of the overdensity in the local supercluster.

Finally one can treat the velocity associated with the anisotropy in the microwave background (Smoot and Lubin 1979) as a peculiar velocity with respect to the correlation length $r \sim 4$ Mpc. Using the CVE results in

$$\Omega_G = 0.08 \pm 0.02 \quad \text{for} \quad r \sim 4 \text{ Mpc}. \quad (7)$$

There is a remarkable uniformity to these values of Ω calculated from data on scales differing by over two orders of magnitude. The value calculated from the KOS sample (5) is high, leading one to suspect that larger deep surveys will yield smaller numbers. If one reanalyzes the KOS sample without the rich field MP4, the result is:

$$\Omega_G \sim 0.15 \qquad \text{for} \quad r \sim 1 - 4 \text{ Mpc} \tag{8}$$

The Durham group's southern deep survey indicates a somewhat smaller value ($\Omega \leq 0.1$), but their final results are not yet in.

In short, on scales larger than that associated with gas dynamical processes (Rees and Ostriker 1977), the assumption of the continuity of clustering works quite well. Other authors have reached quite different conclusions, normally from switching between the assumption of relaxed single particle distribution functions (which yields small numbers on small scales and large numbers on large scales) and the clustering hierarchy.

IV What is Dark Matter?

a) Two populations or one?

Dekel and Shaham (1979) have modeled the observations of NGC 4565 (Heygi and Gerber 1977; and Heygi, this conference) with a stellar mass function which varies with radius. This model is not a dynamical model and would quickly phase mix. There is a simple argument that the halo is a completely different population than the bulge (not just one population varying slightly with radius).

A dynamical system in equilibrium is described by a distribution function $f(x, y, z, v_x, v_y, v_z, m, t)\, dxdydzdv_x dv_y dv_z dm$ where x, y, z are the three spatial coordinates, v_x, v_y, v_z are the corresponding velocities and m is the mass of a particular species. The distribution function, f, represents the number of stars at a time, t, of a mass in the range m to m + dm found in the six-dimensional box located at x, y, z, v_x, v_y, v_z with dimensions dx, dy, dz, dv_x, dv_y, dv_z. To find the density, f is integrated over velocities and species.

Following the motion of stars in time, since the time evolution operator is a canonical transformation, the total number of stars in a given element is constant. If the system is stationary, this reduces to:

$$[H, f] = 0 \tag{9}$$

where $[\ ]$ denotes Poisson brackets and H is the Hamiltonian.

Equation (9) is the requirement that f be a function of the integrals of the motion. In a spherical system, this means $f = f(E, J^2, m)$ where $E = \frac{v^2}{2} + \psi$ and J^2 are the energy and the square of the total angular momentum, and ψ is the potential. Neglecting the angular momentum for the moment and taking $\psi \sim \ln(r)$ as a good approximation to the potential outside the core of the system, we find:

$$f = f(\frac{v^2}{2} + \ln(r), m). \qquad (10)$$

Equation (10) leads to some simple conclusions. A particle whose energy is precisely defined is expected to be found over several decades in radius. If the mass function is slowly varying with energy, color gradients will be logarithmic; as is observed in elliptical galaxies. Finally, if one wants to explain the difference between the light profile ($\sim r^{-3}$) and the density profile ($\sim r^{-2}$) in spiral galaxies using a smoothly varying mass function, the variation must be exponential. Including the angular momentum only makes this worse. Since galaxies have formed by collapse, their velocity dispersions should be biased in the radial direction; thus requiring even more drastic changes in the mass function with energy to produce the same spatial gradients. With such exponential variations, one could even make the disk and bulge appear to be a single population. The conclusion is that the bulge and the halo are dynamically distinct entities.

b) No ν's please

Cowsik and McClelland (1973) and Tremaine and Gunn (1979) have noted interesting effects of neutrinos with modest rest masses less than 100 eV. Tremaine and Gunn used the conservation of phase density to set lower bounds on the masses of neutrinos needed to comprise the halos of galaxies and the unseen mass of great clusters of galaxies. They then use constraints on Ω to rule out the possibility that neutrinos are the dark matter.

Subsequently, numerous authors have argued that the aforementioned constraints on Ω depend in part on the assumption that the mass is in baryons. They have then concluded that the masses of neutrinos needed for galactic halos is acceptable and open to experimental verification. The horde of writers who have stampeded this point have missed the importance of Tremaine and Gunn's argument. Their deduced lower limits were excellent for their purposes, but are actually very poor lower limits. The best description of large-scale neutrino clustering is that of Bond et al. (1980) They point out that neutrino perturbations are Landau damped on scales smaller than the peak value of the Jean's length (defined as that point at which the perturbations have decayed by 1/e), in the same manner as collisionless baryons (Gilbert 1965).

For a typical galactic halo (an isothermal with $v = 270$ km s^{-1} and a core radius of 1 kpc), the neutrino mass required is ~ 75 eV. A protogalaxy of neutrinos with this mass does not come inside its Jeans length until a redshift of order 100. The maximum Jeans mass is about 1000 galaxy masses. Thus all neutrino perturbations on the scale of galaxies should be entirely erased. If the baryonic perturbations were adiabatic, even they have no hope of growing. (Numerous authors have appealed to the "pancaking" of adiabatic neutrino perturbations. The neutrinos crash through caustics at several thousands of kilometers per second, with no source of dissipation to leave galaxies behind.) Even isothermal perturbations will be inhibited, as their growth depends on $\delta\rho/\rho$; and the neutrinos dilute the perturbations by contributing only to the denominator.

In short, if neutrinos have an astrophysically interesting mass, it is probably only astrophysically annoying. A mass of $\gtrsim 10$ eV would almost certainly mean that the universe we see was a tepid, rather than hot, big bang. Earlier at this conference, Dr Schramm has argued that nucleosynthesis dictates a baryonic density of $\Omega \sim 0.05 - 0.10$. As this is the range deduced from studies of large-scale clustering, I conclude that neutrinos have masses too small to be of interest for large-scale clustering.

c) Phase transitions

One promising method for the generation of perturbations and the origin of the dark mass is via phase transitions (see Linde 1979) at an early epoch. In the standard hot model this would occur as super-cooling at a temperature corresponding to the Grand Unification energy or at a Weinberg-Salam time. In the first case the lumps would be $\sim$ grams and decay quickly by the Hawking process. This in turn might produce $\sqrt{n}$ fluctuations of much smaller amplitude growing (in scale)nearly as the horizon. In the second case the lumps would be roughly planetary size, with the standard $\delta_K \sim k^2$ spectrum on scales larger than this.

In a cold universe super-heating or shattering during the transition from pion condensate to nuclear gas could produce fragments of the order of a Chandrasekhar mass (Hogan 1980).

d) Hydrodynamic fluctuations

One final possibility worth mentioning is that the dark mass and perturbations leading to galaxies may be generated at late epochs. The dark matter is then burned out stars from an early epoch and galaxies result from blast waves (Ostriker and Cowie 1980) or ionization fronts (Rees 1981). These schemes have the disadvantage of abandoning some standard precepts, but are more easily constrained by observation (cf. Thorstensen and Partridge 1975).

V Summary

The clustering of galaxies is remarkably continuous on scales from 50 h^{-1}kpc to 5 h^{-1}Mpc. Mass-to-light ratios are useful for studying the distribution of dark vs. light matter inside galaxies, but become ambiguous on large-scales if relaxed single particle distributions are adopted and proper account of sub-clustering is not taken. Numerous schemes may lead to these observed properties, with massive neutrinos being a notable failure.

References

Aaronson, M., Mould, J., Huchra, J., Sullivan, W., Schommer, R. and Bothum, G. 1980. Ap.J., 239, 12.

Bertin, G. and Mark, J.W-K. 1980. A. & A., 88, 289.

Bond, J.R., Efstathiou, G. and Silk, J. 1980. Phys.Rev.Lett., 45, 1980.

Carr, B.J., 1977. M.N.R.A.S., 181, 293.

Cowsik, R. and McClelland, J., 1973. Ap.J., 180, 7.

Davies, R. 1981. M.N.R.A.S., 194, 876.

Davis, M. and Peebles, P.J.E. 1977. Ap.J.Suppl., 34, 425.

Davis, M., Geller, M. and Huchra, J. 1978. Ap.J., 221, 1.

Davis, M., Tonry, J., Huchra, J. and Latham, D. 1980. Ap.J. (Letters), 238, L113.

Dressler, A. 1979. Ap.J., 231, 659.

Efstathiou, G. and Lake, G. 1981, in preparation

Faber, S.M. and Gallagher, G. 1979. Ann.Rev.Ast.Ap., 17, 135.

Fall, S.M. 1979. Rev.Mod.Phys., 51, 21.

Fall, S.M. and Efstathiou, G. 1981. M.N.R.A.S., 193, 189.

Freeman, K. 1970. Ap.J., 160, 811.

Gilbert, I.H., 1965. Ap.J., 144, 233.

Gott, J.R. and Turner, E. 1979. Ap.J. (Letters), 232, L79.

Hogan, C., 1980. Nature, 286, 360.

Holmberg, E. 1969. Ark.fur Astr., 5, 305.

Illingworth, G. 1981. in NATO conference on Normal Galaxies, ed. S.M. Fall. and D. Lynden-Bell, in press.

Kirschner, R.P., Oemler, A. and Schechter, P.L., 2978. A.J., 83, 1549.

Kormendy, J. 1980, in "Two-Dimensional Photometry". Geneva: ESO.

Kruit, P. van der 1980, preprint.

Lake, G. and Mark, J.W.R. 1980. Nature, 287, 705.

Lake, G. and Ostriker, J. 1980, unpublished.

Lake, G. and Tremaine, S. 1980. Ap.J. (Letters), 238, L13.

Lake, G., Efstathiou, G. and Negroponte, J. 1981, in preparation.

Linde, A.P. 1979. Rept. Prog. Phys., 42, 389.

Ostriker, J.P. and Peebles, P.J.E. 1973. Ap.J., 186, 467.

Ostriker, J.P., Peebles, P.J.E. and Yahil, A. 1974. Ap.J. (Letters), 193, L1.

Ostriker, J.P. and Cowie, L. 1980, preprint.

Matthews, W.G. 1978. Ap.J., 219, 413.

Mould, J., Aaronson, J. and Huchra, J. 1980. Ap.J., 238, 458.

Peebles, P.J.E. 1971. Physical Cosmology. Princeton: Princeton University Press.

Peebles, P.J.E. 1979. A.J. 84, 730.

Peebles, P.J.E. 1980. The Large-Scale Structure of the Universe. Princeton: Princeton University Press.

Peebles, P.J.E. and Hauser, M.G. 1974. Ap.J. Suppl., 28, 19.

Rees, M.J., 1978. Nature, 275, 35.

Rees, M.J., 1981, in preparation.

Rees, M.J., and Ostriker, J.P., 1977. M.N.R.A.S., 179, 541.

Saar, E.M., 1978 in IAU Symposium 84 "The Large-Scale Characteristics of the Galaxy", ed. W.B. Burton.

Schechter, P.L. 1980, delivered at Princeton meeting on the Virgocentric flow.

Smoot, G.F. and Lubin, P.M. 1979. Ap.J. (Letters), 234, L83.

Tinsley, B. 1980, M.N.R.A.S., 194, 63.

Toomre, A. 1977. Ann.Rev.Ast.Ap., 15, 437.

Totsuji, H. and Kihara, T. 1969. Pub.Astr.Soc. Japan, 21, 221.

Thorstensen, J.R. and Partridge, R.B. 1975. Ap.J., 200, 527.

Tremaine, S. and Gunn, J.E., 1979. Phys.Rev.Lett., 42, 407.

Tubbs, A. and Sanders, A.H. 1979. Ap.J., 230, 736.

Turner, E.L. 1976. Ap.J., 208, 20.

Zel'dovich, Ya.B. and Starobinsky, A.A. 1976. Zh.Eksp.Tear.Fiz.Pis.Red. 24, 616.

LIST OF PARTICIPANTS

ALY Jean-Jacques	Dept. de Phys. des Part. Elm. CEN Saclay 91191 GIF-SUR-YVETTE, France
AUDOUZE Jean	Institut d'Astrophysique 98 bis, Bd. Arago 75014 PARIS, France
BLUDMAN Sidney	Dept. of Physics Univ. of Pennsylvania PHILADELPHIA PA 19104, USA
COWSICK Ramanath	Tata Inst. of Fundamental Research BOMBAY - 400005 India
CRANE Philippe	European Southern Observ. Karl-Schwarzschild-str.2 8046 GARCHING - MUNCHEN, Germany
EICHENDORF J.W.	European Southern Observ. Karl-Schwarzschild-str.2 8046 GARCHING - MUNCHEN, Germany
ELBERT J.W.	Physics Dept. University of Utah SALT LAKE CITY UT 84112, USA
GAISSER Thomas	Bartol Research Fondation University of Delaware NEWARK DE 19711, USA
HEGYI Dennis	Randal Laboratory University of Michigan ANN ARBOR MI 48109, USA
HENRY Richard C.	Physics Dept. Johns Hopkins Univ. BALTIMORE MD 21218, USA
HEYVAERTS Jean	D.A.F. Observatoire 92190 MEUDON, France
KAISER Nicholas	The Inst. of Astronomy Madingley Road CAMBRIDGE CB3 OHE, England
KUNTH D.	Institut d'Astrophysique 98 bis, bd Arago 75014 PARIS, France

LAKE George	The Institute of Astronomy Madingley Road CAMBRIDGE CB3 0HE, England
NANOPOULOS Demetrios	CERN Theory Division 1211 GENEVA 23 Switzerland
PARTRIDGE Bruce	Haverford College HAVERFORD PA 19041 USA
PRESS William	Harvard College Observatory 60 Garden Street CAMBRIDGE MA 02138, USA
PUGET Jean-Louis	Institut d'Astrophysique 98 bis, bd Arago 75014 PARIS, France
ROWAN-ROBINSON M.	Dept. of Applied Mathematics Mile End Road LONDON E1 4NS, England
SCHAEFFER Richard	Dept. de Phys. Theor. CEN Saclay 91191 GIF SUR YVETTE, France
SCHNEIDER Jean	D.A.F. Observatoire 92190 MEUDON, France
SCHRAMM David	Univ. of Chicago AAC-100 j 5640 S. Ellis CHICAGO IL 60637, USA
SILK Joseph	Astronomy Dept. Univ. of California BERKELEY CA 94720, USA
TORNQVIST Nils A.	Research Inst. Theor. Phys. Siltavuorenpenger 20C 0017 HELSINKI 17, Finland
WATSON Alan	Dept. of Physics Univ. of Leeds LEEDS 2, England
YODH Gaurang B.	Dept. of Physics and Astronomy University of Maryland COLLEGE PARK MD 20742, USA