Particle Physics and Cosmology: Dark Matter

Editor

Mark SREDNICKI

Department of Physics, University of California
Santa Barbara, CA 93106, USA



1990

NORTH-HOLLAND
AMSTERDAM · OXFORD · NEW YORK · TOKYO

© Elsevier Science Publishers B.V., 1990

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the written permission of the publisher, Elsevier Science Publishers B.V., P.O. Box 211, 1000 AE Amsterdam, The Netherlands.
Special regulations for readers in the USA: This publication has been registered with the Copyright Clearance Center Inc. (CCC), Salem, Massachusetts. Information can be obtained from the CCC about conditions under which photocopies of parts of this publication may be made in the USA.
All other copyright questions, including photocopying outside of the USA, should be referred to the publisher.

No responsibility is assumed by the Publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

ISBN: 0 444 88595 1 (hardbound)
0 444 88596 X (paperback)

North-Holland
Elsevier Science Publishers B.V.
P.O. Box 211
1000 AE Amsterdam
The Netherlands

Sole distributors for the USA and Canada:

Elsevier Science Publishing Company, Inc.
655 Avenue of the Americas
New York, NY 10010
USA

Library of Congress Cataloging-in-Publication Data

Particle physics and cosmology : dark matter / editor, Mark Srednicki.
p. cm. -- (Current physics--sources and comments ; 6)
Includes bibliographical references.
ISBN 0-444-88595-1. -- ISBN 0-444-88596-X (pbk.)
1. Dark matter (Astronomy) I. Srednicki, Mark Allen.
II. Series.
QB791.3.P37 1989
523.1'12--dc20 89-70996
CIP

Printed in The Netherlands

Preface

The general subject of dark matter is vast, addressed in a great variety of papers ranging from presentation of detailed astronomical observations to wild speculations in high-energy theory. It is clearly impossible to cover all of this ground in a single volume of reprinted papers; some organizing thread is necessary. I have, therefore, decided to concentrate on the case for a universe with $\Omega = 1$, dominated by dissipationless cold dark matter with a Harrison–Zel'dovich spectrum of initial density fluctuations. Readers who are new to this subject should not fear: all of this phraseology will be explained in the introduction. Expert readers who favor some of the alternatives – baryonic dark matter, hot dark matter, decaying dark matter, non-Gaussian fluctuations, isothermal fluctuations, etc. – will perhaps appreciate a compact presentation of the case for the scenario which, to date, is best developed and most clearly not in conflict with observation. This is not to say that the cold dark matter model – as specified in more detail above – is free of controversy, or can blithely reproduce all of the observational data. However, it has met most of the tests to which it has been put, and those with which it still has difficulty cannot yet be regarded as evidence of definitive failure, but rather as points in need of further study and evaluation. Furthermore, high-energy theorists have provided us with two excellent candidates for cold dark matter: the axion and the lightest supersymmetric particle. Significantly, each of these hypothetical particles was proposed in response to certain problems in elementary-particle theory, rather than being specifically concocted to account for dark matter.

In any single reprint volume, the majority of deserving papers are not included, for the simple reason that there is not enough room for all of them. I have tried to choose papers which, collectively, will give the reader a sense of the diversity of ideas which underlies this subject. The issues addressed are still very much the subject of active research and expert discussion. My hope is that readers of this book will be prepared to participate in that discussion.

I am grateful to Marc Davis, Keith Olive, Joel Primack, and Joe Silk, among many others, for teaching me much of what I know about dark matter. I would also like to thank Eloïse Uranga for patience and support during the course of this project. This work was partially supported by NSF Grant No. PHY-86-14185, and by a Fellowship from the Alfred P. Sloan Foundation.

Acknowledgements

The following articles have been reprinted by kind permission of the publisher, The American Institute of Physics:
From Physical Review:
M. Goodman and E. Witten, Detectability of Certain Dark-Matter Candidates, Phys. Rev. D 31 (1985) 3059–3063; I. Wasserman, Possibility of Detecting Heavy Neutral Fermions in the Galaxy, Phys. Rev. D 33 (1986) 2071–2078; K. Griest, Cross Sections, Relic Abundance, and Detection Rates for Neutralino Dark Matter, Phys. Rev. D 38 (1988) 2357–2375.
From Physical Review Letters:
J.R. Bond, G. Efstathiou and J. Silk, Massive Neutrinos and the Large-Scale Structure of the Universe, Phys. Rev. Lett. 45 (1980) 1980–1984; S. Tremaine and J.E. Gunn, Dynamical Role of Light Neutral Leptons in Cosmology, Phys. Rev. Lett. 42 (1979) 407–410; B.W. Lee and S. Weinberg, Cosmological Lower Bound on Heavy-Neutrino Masses, Phys. Rev. Lett. 39 (1977) 165–168; H. Goldberg, Constraint on the Photino Mass from Cosmology, Phys. Rev. Lett. 50 (1983) 1419–1422; J. Silk and M. Srednicki, Cosmic-Ray Antiprotons as a Probe of a Photino-Dominated Universe, Phys. Rev. Lett. 53 (1984) 624–627; S.P. Ahlen, S. Barwick, J.J. Beatty, C.R. Bower, G. Gerbier, R.M. Heinz, D. Lowder, S. McKee, S. Mufson, J.A. Musser, P.B. Price, M.H. Salamon, G. Tarlé, A. Tomasch and B. Zhou, New Limit on the Low-Energy Antiproton/Proton Ratio in the Galactic Cosmic Radiation, Phys. Rev. Lett. 61 (1988) 145–148; J. Silk, K. Olive and M. Srednicki, The Photino, the Sun, and High-Energy Neutrinos, Phys. Rev. Lett. 55 (1985) 257–259; P. Sikivie, Experimental Tests of the 'Invisible' Axion, Phys. Rev. Lett. 51 (1983) 1415–1417, Erratum, 52 (1984) 695; D.O. Caldwell, R.M. Eisberg, D.M. Grumm, M.S. Witherell, B. Sadoulet, F.S. Goulding and A.R. Smith, Laboratory Limits on Galactic Cold Dark Matter, Phys. Rev. Lett. 61 (1988) 510–513.
From Astronomical Journal:
G. Steigman, C.L. Sarazin, H. Quintana and J. Faulkner, Dynamical Interactions and Astrophysical Effects of Stable Heavy Neutrinos, Astron. J. 83 (1978) 1050–1061.
From JETP Letters:
S.S. Gershtein and Ya.B. Zel'dovich, Rest Mass of Muonic Neutrino and Cosmology, JETP Lett. 4 (1966) 120–122; M.I. Vysotskiĭ, A.D. Dolgov and Ya.B. Zel'dovich, Cosmological Limits on the Masses of Neutral Leptons, JETP Lett. 26 (1977) 188–190.

The following articles have been reprinted by permission from Nature:
J. Einasto, A. Kaasik and E. Saar, Dynamical Evidence on Massive Coronas of Galaxies, Nature 250, 309–310, copyright © 1974 Macmillan Magazines Ltd; G.R. Blumenthal, S.M. Faber, J.R. Primack and M.J. Rees, Formation of Galaxies and Large-Scale Structure with Cold Dark Matter, Nature 311, 517–525, copyright © 1984 Macmillan Magazines Ltd; J.M. Uson and D.T. Wilkinson, Improved Limits on Small-Scale Anisotropy in Cosmic Microwave Background, Nature 312, 427–429, copyright © 1984 Macmillan Magazines Ltd.

The following articles have been reprinted from Astrophysical Journal by kind permission of the publishers, University of Chicago Press:
J. Yang, M.S. Turner, G. Steigman, D.N. Schramm and K.A. Olive, Primordial Nucleosynthe-

sis: a Critical Comparison of Theory and Observation, Astrophys. J. 281 (1984) 493–511; M. Davis, G. Efstathiou, C.S. Frenk and S.D.M. White, The Evolution of Large-Scale Structure in a Universe Dominated by Cold Dark Matter, Astrophys. J. 292 (1985) 371–394; S.D.M. White, C.S. Frenk, M. Davis and G. Efstathiou, Clusters, Filaments, and Voids in a Universe Dominated by Cold Dark Matter, Astrophys. J. 313 (1987) 505–516; J.E. Gunn, B.W. Lee, I. Lerche, D.N. Schramm and G. Steigman, Some Astrophysical Consequence of the Existence of a Heavy Stable Neutral Lepton, Astrophys. J. 223 (1978) 1015–1031; W.H. Press and D.N. Spergel, Capture by the Sun of a Galactic Population of Weakly Interacting, Massive Particles, Astrophys. J. 296 (1985) 679–684. J.P. Ostriker, P.J.E. Peebles and A. Yahil, The Size and Mass of Galaxies, and the Mass of the Universe, Astrophys. J. 193 (1974) L1–4; V.C. Rubin, W.K. Ford Jr. and N. Thonnard, Extended Rotation Curves of High-Luminosity Spiral Galaxies. IV. Systematic Dynamical Properties, Sa → Sc, Astrophys. J. 225 (1978) L107–111; M. Aaronson, Accurate Radial Velocities for Carbon Stars in Draco and Ursa Minor: the First Hint of Dwarf Spheroidal Mass-to-Light Ratio, Astrophys. J. 266 (1983) L11–15; D.N.C. Lin and S.M. Faber, Some Implications of Nonluminous Matter in Dwarf Spheroidal Galaxies, Astrophys. J. 266 (1983) L21–25; S.D.M. White, C.S. Frenk and M. Davis. Clustering in a Neutrino-Dominated Universe, Astrophys. J. 274 (1983) L1–5; P.J.E. Peebles, Large-Scale Background Temperature and Mass Fluctuations due to Scale-Invariant Primeval Perturbations, Astrophys. J. 263 (1982) L1–5; L.F. Abbott and M.B. Wise, Large-Scale Anisotropy of the Microwave Background and the Amplitude of Energy Density Fluctuations in the Early Universe, Astrophys. J. 282 (1984) L47–50; N. Vittorio and J. Silk, Fine-Scale Anisotropy of the Cosmic Microwave Background in a Universe Dominated by Cold Dark Matter, Astrophys. J. 285 (1984) L39–43; J.R. Bond and G. Efstathiou, Cosmic Background Radiation Anisotropies in Universes Dominated by Nonbaryonic Dark Matter, Astrophys. J. 285 (1984) L45–50; P. Lubin, T. Villela, G. Epstein and G. Smoot, A Map of the Comic Background Radiation at 3 millimeters, Astrophys. J. 298 (1985) L1–5.

The following article has been reprinted from Clusters and Groups of Galaxies (eds F. Mardirossian, G. Giuricin and M. Mezzetti) by kind permission of the publishers, D. Reidel Publishing Co.:
J.R. Primack and G.R. Blumenthal, Growth of Perturbations between Horizon Crossing and Matter Dominance: Implications for Galaxy Formation, p. 435–440.

The following articles have been reprinted by kind permission of the publishers, Editions Frontières:
From New and Exotic Phenomena; Proceedings of the Seventh Moriond Workshop (eds J. Trân Thanh Vân and O. Fackler). Frejus Collaboration, Search for Dark Matter and Neutrino Sources with the Frejus Detector, p. 215–220.
From '88 Electroweak Interactions and Unified Theories; Proceedings of the XXIIIrd rencontres de Moriond, Leptonic session (ed. J. Trân Thanh Vân), Kamioka Collaboration, Search for High Energy Neutrinos from Dark Matter Annihilation in the Sun, p. 459–466.

1. Introduction

This introduction is not intended to be a comprehensive summary of the reprinted papers. My goal is to introduce the basic concepts which are needed to understand them and to provide a unified point of view. Excellent introductions to cosmology and the physics of galaxy formation with complete references to earlier work can be found in refs. [1,2]. There are also a number of more recent reviews of cosmology and astrophysics with dark matter, some of which are refs. [3–6]. Two recent conference proceedings [7,8] contain many reports on a wide variety of subjects connected with dark matter.

Let us begin by discussing the Robertson–Walker metric which applies to any homogeneous, isotropic cosmology,

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right]. \tag{1}$$

Here, $a(t)$ is the cosmic scale factor; the sign of k determines whether the universe is open or closed. *Why* the universe is homogeneous and isotropic can be explained by inflationary models of earlier epochs [9], with which we will not be directly concerned. The inflationary models also predict that $k = 0$ (or, more precisely, that it is small enough to ignore). Einstein's equations for the metric reduce to

$$\left(\frac{\dot{a}}{a}\right)^2 = \tfrac{8}{3}\pi G\rho - \frac{k}{a^2}, \tag{2}$$

where ρ is the total mass–energy density, G is Newton's constant, the dot denotes a time derivative, and I am setting $c = 2.998 \times 10^{10}$ cm/s $= 1$, $\hbar c = 1.973 \times 10^{-14}$ GeV cm $= 1$, and $k_{\text{Boltzmann}} = 8.617 \times 10^{-5}$ eV/K $= 1$. Newton's constant can then be expressed as $G = 1/M^2_{\text{Planck}}$ with $M_{\text{Planck}} = 1.221 \times 10^{19}$ GeV.

Today, the value of $\dot{a}/a$ is the Hubble parameter,

$$(\dot{a}/a)_0 = H_0 = 100h \text{ km s}^{-1}\text{Mpc}^{-1},$$

where one megaparsec is 3.086×10^{19} km, and $0.5 \lesssim h \lesssim 1$. We see that, in inflationary models, the mass density today, ρ_0, must equal the critical density

$$\begin{aligned}\rho_{\text{crit}} &\equiv \frac{3H_0^2}{8\pi G} \\ &= 1.879 \times 10^{-29} h^2 \text{ g cm}^{-3} \\ &= 1.054 \times 10^{-5} h^2 \text{ GeV cm}^{-3} \\ &= 2.775 \times 10^{-7} h^2 M_\odot \text{ pc}^{-3},\end{aligned}$$

where $M_\odot = 1.116 \times 10^{57}$ GeV is the mass of the sun. It is convenient to define

$$\Omega \equiv \rho_0/\rho_{\rm crit}.$$

Direct observational data restricts Ω to the range $0.01 \lesssim \Omega \lesssim 2$, with the lower bound from estimates of visible matter and the upper from measurements of the deceleration parameter, $q_0 = -(a\ddot{a}/\dot{a}^2)_0$. However, there are many indirect arguments that further restrict Ω.

In the very early universe, all particles were in thermal equilibrium at a temperature T, where we are quite safe in taking T to be at least a few million electron volts. (Since the advent of grand unified theories and inflation, it is common to discuss temperatures up to 10^{19} GeV, where quantum gravitational effects are expected to set in.) The energy density of the universe at these early times is dominated by relativistic particles, and we say that the universe is "radiation dominated". The energy density is given by

$$\rho = \frac{\pi^2}{30} g T^4, \quad g = g_{\rm B} + \tfrac{7}{8} g_{\rm F}, \tag{3}$$

where $g_{\rm B}$ is the number of spin degrees of freedom of relativistic bosons and $g_{\rm F}$ is the number of spin degrees of freedom of relativistic fermions. Photons contribute $+2$ to $g_{\rm B}$, electrons and positrons (when relativistic) $+4$ to $g_{\rm F}$, and each species of light neutrinos $+2$ to $g_{\rm F}$. The entropy density is always dominated by relativistic particles, even when the energy density is not, and is given by

$$s = \tfrac{2}{45}\pi^2 g T^3.$$

Conservation of entropy implies that sa^3 is constant, and this in turn implies that $\dot{T}/T = -\dot{a}/a$ during periods of time when the number of relativistic-particle species is constant. From eq. (2) we then see that $T \sim t^{-2}$ when $\rho \sim T^4$. At later times, the energy density is dominated by the rest mass of nonrelativistic particles, and we say that the universe is "matter dominated". In this case, $\rho \sim T^3$ and $T \sim t^{-3/2}$.

For $T \lesssim 100$ MeV, the only relativistic particles are photons, neutrinos, electrons, and positrons. When T drops below ~ 2 MeV, the weak interactions become so weak that the neutrinos effectively cease to interact; they are said to "decouple". After this time, the entropy of neutrinos and the entropy of photons plus electrons and positrons are separately conserved. A short time later, the temperature drops below the electron mass. Electrons and positrons then annihilate into photons, but not neutrinos. Conservation of photon–electron–positron entropy causes the photon temperature T to increase by a factor of $[\frac{1}{2}(2 + \frac{7}{2})]^{1/3} = (\frac{11}{4})^{1/3}$ over what it would otherwise be. The neutrino temperature T_ν is unchanged. The energy density of the universe, now dominated by photons and neutrinos, is

$$\rho = \frac{\pi^2}{30}(2T^4 + \tfrac{7}{4} N_\nu T_\nu^4)$$

$$= \frac{\pi^2}{30}\left(2 + \tfrac{7}{4}(\tfrac{4}{11})^{4/3} N_\nu\right) T^4,$$

where $N_\nu \geq 3$ is the number of light neutrino species. From here on, I will take $N_\nu = 3$; then $\rho = \sigma T^4$ with $\sigma = 1.106$.

At temperatures of order 0.1 MeV, the neutrons which have not yet decayed (roughly one for every seven protons) are bound into ^{4}He nuclei, with small amounts of ^{2}D, ^{3}He, and ^{7}Li as well. The detailed results [paper 5] of this process of primordial nucleosynthesis can be calculated to very good precision, and compared with observations of the relative abundances of the elements. From this we get constraints on the parameters Ω_b and N_ν, where Ω_b is the mass density of baryons (protons and neutrons) divided by the critical density. We get information on N_ν since the expansion rate of the universe depends on it through the energy density. The results are

$$0.011 \leq \Omega_b h^2 \leq 0.048, \qquad N_\nu \leq 4.$$

Since $h \gtrsim 0.5$, we see that $\Omega_b \leq 0.19$, and so there is not enough mass in baryons to make up to critical density. Hence, inflationary models demand dark matter.

At much lower temperatures, the energy density of nonrelativistic matter comes to dominate that of the photons and neutrinos. The total energy density of nonrelativistic matter (baryonic plus dark) is proportional to T^3, and today is equal to $\Omega\rho_{crit}$. It is now convenient to introduce another common measure of time and temperature, the redshift z;

$$1 + z \equiv \frac{a(t_0)}{a(t)} = \frac{T}{T_0},$$

where t_0 is the present time and $T_0 = 2.7\ \text{K} = 2.3 \times 10^{-4}$ eV is the present photon temperature. The second equality is valid after electron–positron annihilation. In terms of z, the energy density of nonrelativistic matter is $(1+z)^3\Omega\rho_{crit}$, and the energy density of photons and neutrinos (after electron–positron annihilation) is $(1+z)^4\sigma T_0^4$. These are equal for $z = z_{eq}$, where

$$1 + z_{eq} = \frac{\Omega\rho_{crit}}{\sigma T_0^4} = 2.5 \times 10^4 \Omega h^2, \tag{4}$$

corresponding to a temperature of $T_{eq} = (1 + z_{eq})T_0 = 5.8\Omega h^2$ eV. At a temperature of $T_{rec} \sim 0.4$ eV [or $(1 + z_{rec}) \sim 1500$], the remaining electrons and baryons combine to form hydrogen and helium atoms. (For obscure reasons this is called *re*combination, although the electrons and protons were never previously combined.) At this time, the mean free path for photons becomes essentially infinite; photons and baryons decouple.

The universe today exhibits structure on many scales. Galaxies range in mass from 10^6 to $10^{13}\ M_\odot$, with most bright galaxies having masses of 10^{11}–$10^{12} M_\odot$. Galaxies themselves are found in pairs, small groups of three to ten, clusters of ten to a thousand, and superclusters of two or more clusters. The mass of a supercluster is typically $10^{16} M_\odot$. Also, there are voids, large regions devoid of bright galaxies. If filled with matter at critical density, some of these regions would have masses comparable to superclusters.

To account for all of this structure we need to allow for fluctuations in the energy density. Let us write

$$\rho(\boldsymbol{r}, t) = \rho(t)[1 + \delta(\boldsymbol{r}, t)], \tag{5}$$

where we take $\delta(\boldsymbol{r}, t) \ll 1$ and where $\boldsymbol{r}$ is the coordinate that appears in the metric of eq. (1). Actual physical distances are measured by $a(t)\boldsymbol{r}$. For perturbations on scales much less than

the horizon size, $l_H \sim t \sim (G\rho)^{-1/2}$, and for temperatures below T_{eq} when the universe is matter dominated, we can use Newtonian physics to get an equation for δ. We treat the matter as a nonrelativistic gas with fluid velocity $\dot{a}\boldsymbol{r} + \delta\boldsymbol{v}$, where $\delta\boldsymbol{v}$ is a small perturbation needed to give the gas nonzero temperature and pressure. To be as general as possible we will *not* assume that the matter temperature T_m is the same as photon temperature T. Thermodynamically, we take the perturbation to be adiabatic; this is the prediciton of most models of the generation of fluctuations, which generally do not distinguish between photons and other particles. Using conservation of mass and momentum, it is straightforward (if slightly tedious) to show that

$$\left[\frac{\partial^2}{\partial t^2} + 2\left(\frac{\dot{a}}{a}\right)\frac{\partial}{\partial t} - 4\pi G\rho\left(1 - \frac{\lambda_J^2}{\lambda^2}\right)\right]\tilde{\delta}(\boldsymbol{k}, T) = 0. \tag{6}$$

Here $\tilde{\delta}(\boldsymbol{k}, t)$ is the Fourier transform of $\delta(\boldsymbol{r}, t)$ with respect to $\boldsymbol{r}$, and λ is the corresponding (time dependent) proper wavelength,

$$\lambda = \frac{2\pi}{k}(1+z)^{-1}. \tag{7}$$

The proper wavelength λ is the physical size of the perturbation at time t; today that distance has grown (through the universal expansion) to $2\pi/k$. The (time dpendent) Jeans length λ_J is given by

$$\lambda_J = \left(\frac{\pi v_s^2}{G\rho}\right)^{1/2},$$

where v_s is the speed of sound in the gas. This appears in eq. (6) through the relation $\nabla P = v_s^2 \nabla\rho$, where P is the pressure and ∇ is the spatial gradient. The Jeans length is roughly the distance traveled in a gravitational free-fall time, $\sim (G\rho)^{-1/2}$, when moving at the speed of sound v_s. We will compute λ_J in a few interesting cases shortly.

To solve eq. (6), it is helpful to trade time depence for photon-temperature dependence via $t \sim T^{-3/2}$; we get

$$\left[T^2\frac{\partial^2}{\partial T^2} + \tfrac{1}{2}T\frac{\partial}{\partial T} - \tfrac{3}{2}\left(1 - \frac{\lambda_J^2}{\lambda^2}\right)\right]\tilde{\delta}(\boldsymbol{k}, T) = 0.$$

From this equation, we see that only perturbations with $\lambda > \lambda_J$ can grow as T decreases. Perturbations with $\lambda < \lambda_J$ act as oscillating sound waves. For $\lambda \gg \lambda_J$, the solutions can easily be found explicitly,

$$\tilde{\delta}(\boldsymbol{k}, T) = \tilde{A}(\boldsymbol{k})T^{-1} + \tilde{B}(\boldsymbol{k})T^{3/2}, \tag{8}$$

where $\tilde{A}(\boldsymbol{k})$ and $\tilde{B}(\boldsymbol{k})$ depend on the initial spectrum of fluctuations. If Ω is less than one, this result is valid as long as $\Omega(1+z) \gtrsim 1$; when $\Omega(1+z) \lesssim 1$, the universe becomes "curvature dominated", and perturbations stop growing. Before the period of matter dominance, $v_s = 1/\sqrt{3}$; then the Jeans length is essentially the same as the horizon size, and perturbations inside the horizon do not grow.

We now turn to the evaluation of λ_J. Actually, since mass is conserved but size is not, it is helpful to consider instead the Jeans mass M_J, defined as the mass contained within a sphere of diameter λ_J,

$$M_J = \frac{4\pi}{3}\rho\left(\frac{\lambda_J}{2}\right)^3.$$

There are three separate cases of interest.

(1) The baryon–electron–photon fluid with $T_{rec} < T < T_{eq}$.
The pressure is dominated by the photons, so $P = \frac{1}{3}\rho_\gamma = (\pi^2/45)T^4$, and the mass density in baryons is $\rho_b = \Omega_b\rho_{crit}(1+z)^3$. Since $P \sim T^4$ and $\rho_b \sim T^3$, we have $3\delta P/P = 4\delta\rho_b/\rho_b$ which yields $v_s^2 = 4P/3\rho_b$. We then find that the Jeans mass in baryons is

$$M_J \sim (1.8\times 10^{16} M_\odot)(\Omega_b h^2)^{-2},$$

independent of z. This is bigger than the mass of a supercluster. (Notice that the factor of $(\Omega_b h^2)^{-2}$ has an allowed range of 430 to 8300.) We have found, therefore, that without dark matter only structure on very large scales can grow before recombination.

There is another effect which must be taken into account as well. Adiabataic fluctuations in the baryon density which are smaller than a certain mass (the Silk mass) can be damped out by photon diffusion. To estimate the Silk mass, we first need to compute the mean free path of a photon. Photons can scatter off electrons with the Thompson cross section $\sigma_T = 8\pi\alpha^2/3m_e^2 = 6.65\times 10^{-25}\,\mathrm{cm}^2$, where m_e is the electron mass and $\alpha = \frac{1}{137}$ is the fine-structure constant. The mean free path for a photon is then $L_\gamma \sim 1/\sigma_T n_e$, where n_e is the number density of electrons. The number density of electrons equals the number density of protons (because of charge conservation), and this in turn equals $0.86\Omega_b\rho_{crit}(1+z)^3/m_p$, where m_p is the proton mass. The factor of 0.86 accounts for the fact that approximately 14% of baryons are neutrons in ^{4}He nuclei. In a time t a photon will travel a distance $\lambda_S \sim (tL_\gamma)^{1/2}$ via a random walk. This means that by time t a perturbation of size λ_S will have diffused away. For $z \lesssim z_{eq}$ but $\Omega z \gg 1$, time is related to redshift via

$$t \simeq \tfrac{2}{3}H_0^{-1}\Omega^{-1/2}(1+z)^{-3/2}. \tag{9}$$

We then find

$$\lambda_S \sim (1.0\times 10^4\ \mathrm{Mpc})(\Omega h^2)^{-1/4}(\Omega_b h^2)^{-1/2}(1+z)^{-9/4}.$$

The Silk mass is defined as the baryonic mass contained within a sphere of radius λ_S, and works out to be

$$M_S \sim (1.4\times 10^{12} M_\odot)(\Omega h^2)^{-3/4}(\Omega_b h^2)^{-1/2}\left(\frac{1+z_{rec}}{1+z}\right)^{15/4}.$$

We see that the Silk mass is always smaller than the Jeans mass. Perturbations in the baryons smaller than the Silk mass at redshift z have damped out by that time, and no longer exist.

Without dark matter, $\Omega = \Omega_b$, and by the time of recombination all baryonic fluctuations smaller then $M_S \sim 10^{14} M_\odot$ have disappeared.

(2) The baryon–electron–photon fluit with $T < T_{rec}$.
The photons are now decoupled from the baryons and electrons, which have formed atomic hydrogen and helium. We can treat the fluid as an ideal monatomic gas undergoing adiabatic expansion. Elementary thermodynamics then gives $P \sim \rho_b^{5/3}$. Since $\rho_b \sim T^3$, we must have $P \sim T^5$, and so $v_s^2 = 5P/3\rho \sim T^2 \sim (1+z)^2$. The ideal gas equation of state is $P = T_m \rho_b / \mu$, which then implies $v_s^2 = 5T_m/3\mu$. Here $\mu = 1.2 m_p$ is the mean atomic weight. Note the distinction being made between the photon temperature T and the hydrogen temperature T_m; they are equal only at the time of recombination. Since $v_s^2 \sim (1+z)^2$, we have $v_s^2 = (5T_{rec}/3\mu) \times (1+z_{rec})^{-2}(1+z)^2$. Putting this into the formula for λ_J we find

$$M_J \sim (1.1 \times 10^5 M_\odot)(\Omega_b h^2)^{-1/2} \left(\frac{1+z}{1+z_{rec}} \right)^{3/2},$$

a drop of thirteen orders of magnitude! Fluctuations larger than $M_J(z)$ can begin to grow at redshift z. However, without dark matter, Silk damping has already removed everything below $\sim 10^{14} M_\odot$. The problem of Silk damping would not apply to isothermal fluctuations, which do not entail a perturbation in the photon density and, therefore, cannot be damped by having photons diffuse away.

(3) Dark Matter.
We imagine that there is some form of matter which may have been in thermal equilibrium with the photons at some early time, but which has become completely decoupled and noninteracting by the time of matter domination. This means that out previous treatment of hydrogen does not apply. In order for a gas to undergo adiabatic expansion, there must be collisions in the gas to redistribute the energy and change its temperature. If there are no collisions, the Jeans length is still the relevant scale, but v_s^2 must be replaced with the velocity dispersion $\langle v^2 \rangle$. In this situation, the Jeans length is called the free streaming length λ_{fs} instead. We must modify eq. (6) by dropping the λ_J^2/λ^2 term; however, for $\lambda \lesssim \lambda_J$, the nonlinear corrections to eq. (6) cannot be neglected [2,10]. Physically, perturbations smaller than λ_{fs} are not gravitationally bound, and the particles just drift (or "free stream") away. This differs from the behavior of an interacting gas; as mentioned earlier, in this case perturbations smaller than the Jeans length act as oscillating sound waves. Nevertheless, the Jeans length and the free streaming length both arise as a typical fluid velocity multiplied by the time scale for gravitational collapse.

For temperatures at which the dark-matter particles are relativistic, clearly $\langle v^2 \rangle = 1$ (since the particles are all moving at the speed of light), and the free streaming length is of order the horizon size: $\lambda_{fs} \sim (G\rho)^{-1/2}$. In all cases of interest, the dark matter becomes nonrelavitistic at a time when the universe is radiation dominated, and $\rho = \sigma T^4$. At this time, the dark matter has a temperature T_{dm}, which is either equal to the photon temperature T (if the dark-matter particles are still in equilibrium with the photons) or is some fixed fraction of it (if the dark-matter particles are no longer in equilibrium and the entropy of some now annihilated particle species has gone into the photons). For $T_{dm} \gtrsim m$, the energy density of dark matter is given roughly by the fully relativistic formula of eq. (3), with g specified by the number of spin

degrees of freedom of the dark-matter particle. The free-streaming mass now has its maximum value; ignoring all numerical factors, it is

$$\begin{aligned} M_{\mathrm{fs,max}} &\sim \rho_{\mathrm{dm}} \lambda_{\mathrm{fs}}^3 \\ &\sim T_{\mathrm{dm}}^4 (GT^4)^{-3/2} \\ &\sim \frac{M_{\mathrm{Planck}}^3}{m^2} \left(\frac{T_{\mathrm{dm}}}{T} \right)^6 \\ &\sim (2.9 \times 10^{16} M_\odot)(m/30\ \mathrm{eV})^{-2} (T_{\mathrm{dm}}/T)^6, \end{aligned}$$

where I have set $T_{\mathrm{dm}} \sim m$ in the third line, and used numbers in the last line which agree with the more precise calculation of paper 7 for a light neutrino with $T_{\mathrm{dm}}^3 = \frac{4}{11} T^3$ (see also refs. [10,11]). We see that if the dark-matter particle has a mass of a few GeV, the free-streaming mass is always much smaller than any scale of interest for galaxy formation. If, however, we consider a light neutrino with a mass of ~ 30 eV as dark matter, the maximum free-streaming mass is about the same as a supercluster mass. It is then of interest to know how the free-streaming mass behaves as a function of time for later times. Light neutrinos decouple from photons at a temperature $T_{\mathrm{dec}} = T_0(1 + z_{\mathrm{dec}})$ of order a few MeV, when they are relativistic. At this time, they have a Fermi–Dirac distribution of momenta, with a typical momentum of order T_{dec}. After decoupling, the momenta simply redshift, so that a typical momentum is of order $T_0(1 + z)$, and $\langle v^2 \rangle \sim (T_0^2/m^2)(1 + z)^2$. We also have $\rho \sim (1 + z)^4$ and $\rho_{\mathrm{dm}} \sim (1 + z)^3$, so we find that M_{fs} is independent of z when the universe is radiation dominated. However, recall that $T_{\mathrm{eq}} = 5.8 \Omega h^2$ eV is not much smaller than m, so we quickly enter the matter dominated era in which $\rho \sim (1 + z)^3$, which leads to $M_{\mathrm{fs}} \sim (1 + z)^{3/2}$.

With this information, we can begin to compare the possibilities for galaxy and structure formation with and without dark matter. We have seen that, without dark matter, we have the following situation: $\Omega = \Omega_{\mathrm{b}} < 0.2$; baryonic fluctuations smaller than $\sim 10^{19} M_\odot$ cannot begin growing until after recombination; fluctuations smaller than $\sim 10^{14} M_\odot$ have been wiped out by Silk damping. Furthermore, growth stops at $z \sim \Omega_{\mathrm{b}}^{-1} > 5$, which leaves only a redshift factor of ~ 300 for perturbations on interesting scales to grow. Since fluctuations which produced galaxies and clusters are by now evolving nonlinearly, the magnitude of fluctuations at recombination must have been of order $\delta\rho/\rho \sim \frac{1}{300}$. Since $\delta\rho/\rho = 3\delta T/T$, we must have $\delta T/T \gtrsim 10^{-3}$ at recombination. These fluctuations would survive in the present microwave background radiation, since the photons which comprise it have not scattered since recombination. However, the observational limits generally restrict $T/T \lesssim 10^{-4}$ on both small and large angular scales ([papers 20,21]; for a recent review, see ref. [12]).

The story is very different with dark matter. Now, fluctuations can begin to grow at the beginning of the matter dominated era. Also, we can have $\Omega = 1$, which gives still more time for linear growth. Fluctuations in the baryons below the Silk mass were wiped out, but not those in dark matter. After recombination, when the baryonic Jeans mass drops, the baryons can fall into the gravitational potential wells created by the dark matter. It is possible to have fluctations large enough to form galaxies and still small enough to be consistent with limits on the microwave background [papers 18,19].

The remaining pieces of needed information are the value of the free-streaming mass, and the initial spectrum of fluctuations. We have already estimated the free streaming mass. To

specify the initial spectrum of fluctuations, we treat the spatial Fourier transform of the density fluctuation, $\tilde{\delta}(\boldsymbol{k}, t)$, as a Gaussian random variable with two-point correlation function

$$\langle \tilde{\delta}^*(\boldsymbol{k}, t)\tilde{\delta}(\boldsymbol{q}, t)\rangle = P(k, t)\delta^{(3)}(\boldsymbol{k}-\boldsymbol{q}),$$

where $\delta^{(3)}$ is the three-dimensional Dirac delta function, and $P(k, t)$ is called the power spectrum; $P(k, t)$ is a function of the magnitude of $\boldsymbol{k}$ and the time t. Inflation predicts [9] that the initial power spectrum, $P(k, 0)$, has the the Harrison–Zel'dovich form, $P(k, 0) \sim k$. We must then compute the spectrum at later times. A study of the growth of perturbations, using linearized general relativity, reveals the following. During the radiation dominated era (and in frames in which the universal expansion is spatially homogeneous), perturbations $\tilde{\delta}(\boldsymbol{k}, t)$ grow like $(1+z)^{-2}$ if the physical wavelength $\lambda = 2\pi k^{-1}(1+z)^{-1}$ is larger than the horizon size $l_{\rm H} \sim (G\rho)^{-1/2} \sim (GT_0^4)^{-1/2}(1+z)^{-2}$. However, once $\lambda \gtrsim l_{\rm H}$, $\tilde{\delta}(\boldsymbol{k}, t)$ stops growing. During the matter dominated era, all perturbations, whether smaller or larger than the horizon, grow like $(1+z)^{-1}$. [We have already seen this for perturbations smaller than the horizon in eq. (8).] Thus, perturbations which entered the horizon during the matter dominated era have all experienced the same growth factor, and the shape of $P(k, 0)$ is unchanged. However, perturbations which entered the horizon during the radiation dominated era grew like $(1+z)^{-2}$ only until $\lambda \sim l_{\rm H}$, when $(1+z)^{-2} \sim GT_0^4/k^2$. Thus, the shape of the power spectrum of these perturbations is altered to $k^{-4}P(k)$, and there is a "knee" in the power spectrum at $\lambda \sim l_{\rm H,eq} \sim t_{\rm eq}$. Using eqs. (4), (7) and (9), we find that the "knee"appears at $2\pi/k_{\rm eq} \simeq 25(\Omega h^2)^{-1}$ Mpc, corresponding to a mass of $M_{\rm eq} \simeq (4\pi/3)\rho_{\rm crit}(\pi/k_{\rm eq})^3 \simeq (2.3\times 10^{15}M_\odot)(\Omega h^2)^{-2}$. Using the inflationary prediction $P(k, 0) \sim k$, we see that for perturbations with $M \lesssim M_{\rm eq}$, the effective power spectrum is $P(k) \sim k^{-3}$. This is the only form for $P(k)$ which does not introduce a dimensionful parameter; since $\delta(\boldsymbol{r}, t)$ is dimensionless by definition [see eq. (5)], $\tilde{\delta}(\boldsymbol{k}, t)$ has dimensions of length cubed and, therefore, so does $P(k)$. If there is no dimensionful parameter, $P(k)$ must be proportional to k^{-3}. Another way to see the significance of this power spectrum is to compute the root-mean-square fluctuation in mass inside a randomly placed sphere with diameter $2\pi/k$; this is given by $(\delta M/M)^2 \sim k^3P(k)$, where the factor of k^3 comes from a phase space integral. Thus, $P(k) \sim k^{-3}$ predicts that fluctuations in mass all have the same amplitude when they cross the horizon. The value of this amplitude must be $\lesssim 10^{-5}$ in order not to violate bounds on the fluctuations in the microwave background [papers 18–20]. It is also necessary to appreciate that the preceding discussion has been rather crude. In fact, the transition from $P(k) \sim k$ to $P(k) \sim k^{-3}$ occurs over several decades of k, and the asymptotic behavior of $\delta M/M$ for $M \ll M_{\rm eq}$ is $[\ln(M_{\rm eq}/M)]^{3/2}$, rather than a constant. The distinctions are important when making detailed comparisons with the observed structure of galaxies and clusters ([papers 12–14] and refs. [13,14]).

This picture is largely complete if the free-streaming mass is smaller than about $10^6M_\odot$, the mass scale of the smallest galaxies. In this case we say [10] that the dark matter is "cold," since the particles which comprise it became nonrelativistic at an early time. Conversely, if the free-streaming mass is large we say that the dark matter is "hot." The most plausible dark-matter candidates are either clearly cold (with $m \gtrsim 1$ GeV) or clearly hot (with $m \sim 30$ eV). For hot dark matter with $m \sim 30$ eV, the free-streaming mass is of the order of a supercluster mass, and there is an exponential drop in the power spectrum for smaller fluctuations (see [paper 7] and refs. [10,11]). In this case, galaxies must form through fragmentation during the collapse of larger structures which form first.

Which (if either) of these scenarios can reproduce the observed galaxy distribution can best be decided with numerical simulations of the nonlinear regime ([papers 11,15,16,] refs. [15,16],

and references therein). The "cold" scenario seems to do much better, overall, than the "hot" scenario, although the issue is perhaps not yet completely settled [16]. Also, the presence of cosmic strings [17] may change the conclusions. The viability of the cold scenario with $\Omega = 1$ requires the galaxy distribution to differ somewhat from the mass (dark-matter) distribution, which leads to the necessity of "biased" galaxy formation. For example, perhaps galaxies only form when the density fluctuation is larger than some threshold. This leads to a galaxy distribution with stronger correlations than the mass distribution, and this improves the agreement between observation and the numerical simulations. The most often quoted measure of the galaxy distribution is the galaxy–galaxy correlation function $\xi(r)$, which is defined as follows: if there is a galaxy at $r = 0$, then the probability to find another galaxy at a distance r is $1 + \xi(r)$ times what it would be if the galaxies were randomly distributed. Other correlation functions, such as the cluster–cluster and cluster–galaxy, are similarly defined. When fit to a power law, the galaxy–galaxy correlation function is roughly given by $\xi_{gg} \sim [r/(5h^{-1}\ \text{Mpc})]^{-1.8}$ for $0.1h^{-1}\ \text{Mpc} \lesssim r \lesssim 9h^{-1}\ \text{Mpc}$.

There is also direct and convincing observational evidence for dark matter ([papers 1–3] and refs. [3–6]). The most persuasive comes from measurements of the "rotation curves" of spiral galaxies. The rotation curve is simply the velocity v_{rot} of test body in circular orbit around the center of the galaxy at a distance r. In practice, the "test body" can be either a star or hydrogen gas, and the velocity is measured via Doppler shifts. (Since we are not often fortunate enough to be in the plane of rotation of another galaxy, extracting the rotation velocity from Doppler measurements is not entirely straightforward.) It is easy to predict the result in the absence of dark matter. Most of the mass of a spiral galaxy would be in the visible stars at the core, so that for $r \gtrsim 3$ kpc, little mass is being added as we increase the distance from the center. Then Newton's second law tells us that $v_{rot}^2/r = GM/r^2$, where M is the mass of the galaxy, so that $v_{rot} \sim r^{-1/2}$. In fact, the observations yield an essentially constant value of v_{rot} (usually in the range of 100–300 km/s) for $r \gtrsim 5$ kpc out to the limits of measurement, indicating the presence of dark matter. Assuming that the mass of the galaxy is spherically distributed, a constant rotation velocity would mean $M(r) = v_{rot}^2 r/G$, where $M(r)$ is the mass contained within radius r. This then means that the mass density $\rho(r) = v_{rot}^2/4\pi G r^2$, although this obviously must be modified for both small r (since we cannot have an infinite density) and large r (since we cannot have an infinite total mass). This sphere of dark matter must be held up against gravitational collapse by internal pressure. (Actually, if the dark-matter particles are noninteracting, it is really conservation of the energy and angular momentum of each particle that prevents collapse. Nevertheless, it is helpful to think in terms of the pressure the gas would exert on walls of a container.) The gravitational force on a fluid element with volume dV at radius r is $GM(r)\rho(r)\,dV/r^2$, and this must be balanced by the force due to the pressure gradient, $P'(r)\,dV$, where $P(r)$ is the pressure and the prime denotes its radial derivative. If we further describe the dark matter as a nonrelativistic ideal gas, we can use the equation of state $P(r) = T(r)\rho(r)/m$, and solve for $T(r)$. One solution is $T(r) = \frac{1}{2}mv_{rot}^2$, independent of r. This mass distribution is, therefore, called an isothermal sphere, and it produces the observed flat rotation curve. Isothermal spheres arise generically from gravitational collapse of a density fluctuation $\delta\rho$ in a noninteracting gas; the process is called "violent relaxation", violent because it occurs quickly, on a time scale $\sim (G\delta\rho)^{-1/2}$. If we include baryons, they also undergo collapse, but since they can lose energy by radiation, they tend to form either a dense core of stars, or a disk of stars and gas, the latter arising due to conservation of angular momentum.

A standard measure of the amount of dark matter present in some particular region is the mass-to-light ratio, M/L. This is simply the total mass in that region divided by the power it

emits in visible light; both M and L are usually given in solar units, with $M_\odot = 2.0 \times 10^{33}$ g and $L_\odot = 3.9 \times 10^{33}$ erg/s. The mean luminosity density for the universe is approximately $1.8 \times 10^8 h L_\odot$ Mpc^{-3}. The mean mass-to-light ratio of the universe is gotten by dividing $\Omega\rho_{\rm crit}$ by the mean luminosity, and results in $M/L \simeq 1500\Omega h$ in solar units. Cores of galaxies have $M/L \simeq 14h$, indicating that we need a halo roughly 100 times more massive than the core in order to have enough dark matter in galactic halos to get $\Omega = 1$. This factor of 100 is consistent with measurements of M/L for clusters of galaxies. Their total mass is estimated using the virial theorem to convert velocity dispersions to gravitational potential energy.

Clearly, the most pressing question is the nature of the dark matter. We already have three arguments that it cannot be baryonic: there is no workable theory of galaxy formation with adiabatic fluctions but without nonbaryonic dark matter*; the existence of galactic halos which have not collapsed argues for a composition of dissipationless, noninteracting particles; and $\Omega = \Omega_{\rm b} < 0.2$ is inconsistent with inflationary models. However, there is further evidence: it is difficult to make galactic halos (in particular, the halo of our own galaxy) out of baryonic matter in any form without leaving behind easily observable signatures. For example, hydrogen at a temperature of $T \sim m_{\rm p} v_{\rm rot}^2 \sim 10^6$ K would be ionized, and would copiously emit unobserved X-rays [paper 4]. The most viable possibilities are failed stars which never began nuclear burning, and supermassive black holes which accreted all material ejected from the stars which formed them. Neither possibility seems likely, and would require star formation in the halo to be quite different from star formation in the disk and core. Another possibility is black holes left over from the Big Bang, but these are not really baryonic; they do not count as baryons for nucleosynthesis calculations, for example.

If we agree that some form of nonbaryonic dark matter is necessary, what should it be? The most obvious candidate is one of the three known neutrinos, endowed with a rest mass of ~ 30 eV ([paper 6] and ref. [20]). This is not in conflict with any known principles of particle physics, and indeed is predicted by some grand unified models, although the most natural scale for a neutrino mass in a grand unified theory is $m_{\rm lepton}^2/M_{\rm GUT} \sim (1\ {\rm GeV})^2/(10^{14}\ {\rm GeV}) \sim 10^{-5}$ eV. (For reviews of grand unified models, see ref. [21].) The number 30 eV arises as follows. The present number density of each species of neutrino is $n_\nu = (3\zeta_3/2\pi^2)T_\nu^3$, where $T_\nu^3 = \frac{4}{11}T_0^3$, $\zeta_3 = 1.202$ is the Riemann zeta function of three, and I have assumed that the neutrinos still have two spin degrees of freedom each. This means that the neutrino mass is of the Majorana type rather then the Dirac type. This is far more natural from a particle-physics point of view. A Dirac mass would require introducing two new spin degrees of freedom for each massive neutrino; these would be new particles, requiring new quantum fields, with no particular reason to exist. Furthermore, the new particles could themselves be given arbitrarily large Majorana masses, leaving the original neutrinos with small Majorana masses. (This, in fact, is the origin of the formula $m_\nu \sim m_{\rm lepton}^2/M_{\rm GUT}$.) Finally, the new spin states would count in the value of N_ν, which is restricted by the nucleosysthesis calculations, although the increase depends on the temperature at which the new states decoupled. We, therefore, assume Majorana masses in the following discussion. Of the three neutrinos, the heavier would be unstable against decay into the lighter, but with lifetimes of at least 10^{16} years (see, e.g., ref. [22]). We can thus treat all three as stable. If we define $\overline{m}$ as the average of the masses of the neutrinos, we find that the mass density in neutrinos is $\rho_\nu = 3\overline{m}_\nu n_\nu$, which yields $\Omega_\nu h^2 = 0.93\overline{m}_{30}$, where $\overline{m}_{30} = \overline{m}/(30\ {\rm eV})$. This then leads to a hot dark-matter scenario ([paper 7] and refs. [10,11]).

* However, it may be possible to get a consistent picture with only baryons if the fluctuations are isothermal [18]; these are not subject to Silk damping. Isothermal fluctuations are not produced in the usual scenarios of inflation and baryon number generation, but it is not impossible to get them (see ref. [19] and references therein).

No cold dark matter candidate has ever been seen in the laboratory. However, theorists have concocted a very large zoo of hypothetical particles, many of which have the right properties to serve as cold dark matter. Before peeking into any of their cages, let us study these properties in more detail. We need a heretofore undiscovered particle which is either stable or has a lifetime which is at least the present age of the universe; otherwise the dark matter would have already disappeared*. There is a very short list of particles known to be absolutely stable: the photon, the graviton, the electron, and the lightest spin one-half particle (which is most likely, but not necessarily, one of the presently known neutrinos). The stability of these four is guaranteed by, respectively, electromagnetic gauge invariance, general coordinate invariance, conservation of electric charge, and conservation of angular momentum. While it is always possible to imagine that *any* currently held principle of physics will eventually be found to be in error, these four are as sound as anything we know. The moral is that absolute stability only occurs when there is a very good reason.

The standards are a little lower for a long lived particle: here we can add the three known neutrinos and the proton. If the three neutrinos have different masses (and if they have masses at all, there is no reason for them to be the same), then the two heavier neutrinos are almost certainly unstable, with a decay rate of order $G_F^3 m_\tau^2 m_\nu^5$, where $G_F = 1.2 \times 10^{-5}$ GeV^{-2} is the Fermi constant and $m_\tau = 1.8$ GeV is the tau lepton mass [22]. This is negligibly small because it contains a factor of m_ν^5, and m_ν is small compared to the masses of the particles (the W boson and the tau lepton) which appear as virtual states during the decay. The proton is long lived because, in the Standard Model of elementary particles, there is no particle which can mediate a process that would allow the proton to decay. If we introduce such a particle, with mass M, the proton would then decay at a rate of order $(m_p/M)^n m_p$, where n is an integer that depends on the details of the new particle and its interactions. In grand unified theories, there is such a particle, with $M = M_{GUT} \sim 10^{14}$ GeV and $n = 2$, giving the proton a predicted lifetime of order 10^{30} years. Of course, this prediction is sensitive to the exact value of M_{GUT} and other details of the theory; the limit on the proton lifetime is 2.5×10^{32} years for the $p \to e^+ \pi^0$ mode [24]. The moral here is that if we want a long-lived particle, its decay rate should involve ratios of small to large masses. It is, of course, possible to suppress decay rates by assuming (with no particular justification) that appropiate coupling constants are unusually small. Although we cannot say that such assumptions are necessarily wrong, we can shorten our time in the zoo by restricting our attention to particles which are long lived for better reasons.

The two best reasons suggested to date are supersymmetry, and time reversal invariance of the strong interactions. Let us discuss each in turn.

Supersymmetry relates bosons to fermions, and so predicts that for every known particle, there is a supersymmetric partner, differing in spin by one-half. (For a review of supersymmetry, see ref. [25].) Thus for each type of quark there is a spin zero squark, for each type of lepton there is a spin zero slepton, for the photon a spin one-half photino, for the W boson a spin one-half wino, for the Higgs boson a spin one-half higgsino, etc. The couplings of these new particles to the old ones are restricted by supersymmetry; in the most staightforward supersymmetric extension of the Standard Model, the new particles always couple to the old ones in pairs. This means that one of them must be stable, since in any Feynman diagram if one new particle goes in, at least one (possibly three, five, ...) comes out. Unfortunately, the most straightforward extension is not the only one; it is possible to add couplings which mix the higgsino and the ordinary neutrinos, resulting in decay of the lightest superparticle to neutrinos. The same type of couplings also lead to far too short a lifetime for the proton, the

* However, there have been studies of particles which decayed to other stable particles at some late time [23].

decay being mediated by a virtual squark. These disasterous couplings can be plausibly forbidden in certain grand unified versions of the model, but their possibility does detract from the power of the basic idea of supersymmetry as a justification for a new long-lived particle. If we assume they are absent, the lightest superparticle has a lifetime of the same order of magnitude as the proton. In order to serve as dark matter (a possibility first suggested in ref. [26]), the lightest superparticle must be colorless and electrically neutral; otherwise they would not form the noninteracting, dissipationless gas that we need [paper 31]. This limits the possibilities to the lightest neutralino (a linear combination of the photino, higgsinos, and zino which has a definite Majorana mass; see [papers 30–32,45]) and the lightest sneutrino (spin zero partner of a neutrino; see ref. [27]). The mass of either of these dark-matter candidates is likely to be at least a few GeV, making them cold dark matter. They would interact with each other and with ordinary matter by exchange of other superparticles such as squarks, with characteristic cross sections of order $\sigma \sim 8\pi\alpha^2 m^2/m_{\rm sp}^4$, where m is the mass of a dark-matter particle and $m_{\rm sp}$ is the mass of some other superparticle which is exchanged. We know that $m_{\rm squark}$ and $m_{\rm gluino}$ are at least ~ 60 GeV [28], and generally expect that $m_{\rm sp} \gtrsim M_{\rm W}$ [25].

It is, however, not enough to simply postulate a heavy, long-lived particle. For illustration, let us suppose that a nearly pure photino is the lightest superparticle. At temperatures above the photino mass, the photinos would be in thermal equilibrium, with their mass density fixed according to eq. (3). As the temperature drops below their mass, their number density decreases due to Boltzmann suppression (see [papers 25–27,30–32] and refs. [29–31]). Eventually, however, their number cannot be further reduced because they are being pulled apart by the universal expansion faster than they can annihilate. At this time, their self-interactions "freeze out", and afterward their number density decreases only due to the universal expansion; there is no further Boltzmann suppression. Freeze-out occurs at a temperature $T_{\rm fr}$ when $(\dot{a}/a)n \sim \sigma v n^2$, where n is the number density of photinos, σ is their annihilation cross section, and v is the relative velocity of an annihilating pair. For nonrelativistic particles, σv is essentially independent of v, and is given by the speed of light times the characteristic formula for σ mentioned above. Using the Boltzmann formula $n = 2(2\pi mT)^{-3/2}\exp(-T/m)$, and $\dot{a}/a = (4\pi^3 Gg/45)^{1/2}T^2$ from eqs. (2) and (3), we find a transcendental equation for $T_{\rm fr}$ which yields an essentially logarithmic dependence of $T_{\rm fr}/m$ on σv and m; for cross sections as described above, typically $T_{\rm fr} \simeq \frac{1}{20}m$. The final mass density of photinos today, $\rho_{\tilde{\gamma}}$, ends up proportional to $(\sigma v)^{-1}$, with no explicit dependence on the photino mass m. The rule-of-thumb result is that $\rho_{\tilde{\gamma}} \simeq \rho_{\rm crit}$ for $\sigma v \simeq 10^{-26}$ cm^3/s $\simeq 10^{-9}$ GeV^{-2}. This is exactly what we get from the formula $\sigma v \sim 8\pi\alpha^2 m^2/m_{\rm sp}^4$ with $m_{\rm sp} \sim 100$ GeV and $m \sim 10$ GeV. These numbers are well in line with the general expectations of high-energy theorists. This coincidence of the value of the cross section needed for $\Omega = 1$ with the value generally expected on other grounds may be a profound clue to the nature of dark matter. It should also be noted that the same numbers apply to other particles such as heavy neutrinos (with $M_{\rm W}$ playing the role of $m_{\rm sp}$); these were the first cold dark-matter candidates to be discussed [papers 25–29]. Again, however, we have no good reason for either the existence or (if it exists) the stability of such a particle.

The other physics principle which could yield a long-lived particle is time reversal invariance of the strong interactions. Generically, a theory will be time reversal invariant if all of its coupling constants are real. The connection can be seen heuristically from the Schrödinger equation, $i(\partial/\partial t)\,|\,\psi\rangle = H\,|\,\psi\rangle$; If H has no complex parameters, complex conjugation (as opposed to hermitian conjugation) is the same as $t \to -t$. However, this definition is a little too glib, since sometimes a complex parameter can be made real by absorbing its phase into the definition of one or more of the quantum fields. Quantum chromodynamics (QCD), the theory of interacting quarks and gluons, was originally thought to be time reversal invariant. The

quark masses could in principle be complex, but their phases could apparently be absorbed into the quark fields. This turns out to be incorrect; QCD embodies a complicated field theoretic effect, the result of which is to forbid changing the phase of the determinant of the quark mass matrix. (For a careful explanation of this, see ref. [32] and references therein.) This phase becomes a new parameter of the theory, known as the "vacuum angle" θ. QCD is time reversal invariant only if $\theta = 0$. The limit on the magnitude of the electric dipole moment of the neutron constrains θ to be less than 10^{-8} [33]. The problem is to understand why θ, which in principle could be anywhere from zero to 2π, is so close to zero. An explanation is provided by axion models [34], in which θ is determined by the expectation value of a scalar field. The potential energy of this field is minimized when $\theta = 0$; this does not require any special adjustment of the parameters of the theory. The quantum particle associated with the field is called the axion, and it has a mass m_a of order $(100\ \text{MeV})^2/f$, where f is an arbitrary parameter with dimensions of mass known as the axion decay constant. (For a review of axions, see ref. [35].) The coupling of axions to ordinary matter is proportional to $1/f$; the decay rate of the axion (into two photons) is $\sim \alpha^2 m_a^3/f^2$, and is tiny because it involves ratios of small to large masses. The most stringent limits on f and m_a come from supernova 1987A. The observance of the neutrino signal at the expected level means that not much (if any) energy was emitted in the form of axions, and this can be translated into the limits $f \gtrsim 10^{11}$ GeV and $m_a \lesssim 10^{-3}$ eV [36]. This small value of m_a would apparently mean that the axion is a hot dark matter candidate. However, it is possible to end up with a lot of nonrelativistic axions through the following process [papers 22–24]. After inflation, our universe has a random value of θ, and the axion field has stored potential energy. This potential energy decreases through the universal expansion and through certain damping effects which drive θ towards zero. One finds that the potential-energy density today equals the critical density for $f \sim 10^{12}$ GeV, and that the effective value of θ today is 10^{-21}. Thus, we need a value of f just above the lower limit inferred from the supernova. However, it must also be noted that before the appearance of papers 22–24, the most popular value for f was the grand unification scale $M_{\text{GUT}} \sim 10^{14}$ GeV; there is no convincing reason (other than a desire for an axion dominated universe) to introduce a scale of 10^{12} GeV. Quantum mechanically, the remaining potential energy consists of axions at rest. These axions never come into thermal equilibrium and never acquire large velocity dispersions; in other words, they are nonrelativistic, and behave as cold dark matter.

If the dark matter is indeed composed of some as yet undiscovered elementary particle, an obvious and vitally important problem is to detect and identify it. Methods for doing so can be divided into two broad classes, direct and indirect. Direct detection implies that a dark-matter particle passing through one's laboratory interacts with a detector in some way. Indirect detection implies looking for products of the annihilation of two dark-matter particles. Preliminary experimental results (all negative) are available for both direct and indirect detection.

Indirect detection was first discussed in the context of heavy neutrinos as dark matter. We know the annihilation cross section of the heavy neutrinos (or photinos, or sneutrinos, or whatever) because that is what determines their present relic density. We can, therefore, estimate the rate of production of gamma rays and cosmic rays due to continuing annihilation in our own galactic halo. This was first done for gamma rays in paper 28 and ref. [37], and then extended to antiprotons and positrons (for photinos as dark matter) in paper 33. Much more theoretical work on this subject was done subsequently [38]. A recent measurement of the low-energy antiproton flux [paper 34] has set some modest limits on the properties of cold dark matter candidates.

Another possibility for indirect detection involves looking for high-energy neutrinos among

the annihilation products. For a generic annihilation in the halo this is hopeless, but there may be a significant enhancement in the dark-matter density inside the sun or the earth, which would greatly increase the annihilation rate. Of course, among the resulting particles only neutrinos could escape and reach the earth's surface. The possibility of such a density enhancement was first suggested in paper 29 (although the specific numerical results have recently been criticized [39]), and then rediscovered by Spergel and Press [40], who also suggested that dark-matter trapping would be an ongoing process [paper 35]. The possibility of observing high-energy neutrinos from the sun was first pointed out in paper 36, and from the earth in ref. [41]. Again, much more theoretical work on this subject was done subsequently [42,31], represented here by papers 40 and 41. The experimental results of the IBM, Frejus, and Kamioka collaborations [papers 37–39] place strong limits on sneutrinos and heavy Dirac neutrinos as dark matter, and modest limits on photinos and higgsinos. (The limits on Dirac neutrinos apply only if the dark matter consists of equal numbers of particles and antiparticles, so that they can annihilate. This is always the case for sneutrinos, which dump any excess lepton number into ordinary neutrinos [31], and for Majorana fermions, which are their own antiparticles.) Regrettably, there are no indirect methods for detecting axions, which are so weakly coupled that they essentially never annihilate.

Direct detection of dark-matter particles in the laboratory was first proposed for axions by Sikivie [paper 42] and for heavy neutral particles by Goodman and Witten [paper 43] and Wasserman [paper 44]. Sikivie's method involves looking for axion to photon conversion in a microwave cavity in the presence of a magnetic field. A preliminary version of this experiment has been performed [43]; the results do not constrain any parameters. The direct detection schemes of papers 42 and 43 involve looking for the small amount of energy deposited when a heavy dark-matter particle collides with a nucleus in a laboratory detector. Several methods for observing this energy are possible; for a thorough summary, see ref. [44]. This sort of direct detection is easiest for sneutrinos and heavy Dirac neutrinos; there are already strong experimental limits on these particles [papers 46,47] which apply even in the case of Dirac neutrinos with an excess of particles over antiparticles. Detecting photinos and similar Majorana fermions is considerably more challenging.

Another related subject has received a great deal of attention recently. As originally noted in paper 29, heavy neutrinos in the sun would be an extra agent of thermal transport which could cool the solar core, and so ameliorate the solar neutrino problem. This idea was revived in ref. [40]. It was later found, however, that none of the dark-matter candidates so far discussed have a large enough interaction cross section to cool the core appropriately [45], although study of this possibility continued [46] and soon enough particles (now generically called cosmions) with the needed propensities were conconted [47]. The enhanced interaction cross sections of these particles also means enhanced detectability, and a new round of experiments will either find cosmions or rule them out [48]. I have, therefore, placed my bets and chosen not to include papers on cosmions in this volume.

The papers which are reprinted are loosely organized into sections. Section 3.1 is on the observational evidence for dark matter. Dark matter was first reported in 1933 by Zwicky [49], who noticed that galaxies in rich clusters have velocity dispersions which require at least ten times more mass than is visible in order to have the cluster be gravitationally bound. Although not taken seriously at the time, evidence for dark matter gradually mounted until, in 1974, two seminal papers were published [papers 1,2] suggesting that galaxies have large dark halos which stabilize clusters and have sufficient mass to make $\Omega \simeq 0.2$. Subsequent measurements of rotation curves (here represented by paper 3) strengthened the conclusion. A detailed and thorough review of the evidence to date has recently been given by Trimble [5].

The evidence against baryons as dark matter is presented in section 3.2. Paper 4 summarizes the problems with baryons in any form comprising truly invisible halos, and paper 5 contains the most thorough calculation of primordial nucleosynthesis, which severely restricts the value of Ω_b. As discussed earlier, another strong case against baryons can be made from the theory of galaxy formation; this is briefly recounted in paper 14.

Although a great deal of work has been done on the subject of section 3.3, light neutrinos as dark matter, the case against this idea (at least without cosmic strings) is now persuasive, and is presented in section 3.4. Included here is the 1966 paper of Gershtein and Zel'dovich, which was the first to point out the existence of a cosmological limit on the neutrino mass [paper 6]. This fact was rediscovered and more carefully analyzed a few years later [20]. The idea that light neutrinos are good dark-matter candidate seems to have occured to many people at once ([paper 7] and ref. [11]); only one of these papers is included here, in view of the results of papers 8–11. In paper 8, Tremaine and Gunn used conservation of phase space density to derive a lower limit on the mass of light neutrinos in galactic halos. The limit is strongest for dwarf galaxies, hence the relevance of observations concerning them ([papers 9,10], for more recent results, see ref. [50]). Further, the numerical simulations of White, Frenk, and Davis [paper 11] are difficult to reconcile with observed structure. Other arguments against hot dark matter are given in papers 13 and 14.

Section 3.5 is concerned with the development of the theory of a universe dominated by cold dark matter. The key problems are the computation of the spectrum of fluctuations [papers 12,13] and identifying mechanisms for the formation of galaxies consistent with observations [papers 13–16]. The numerical simulations presented in papers 15 and 16 are the best evidence so far that the cold dark-matter scenario is workable. Section 3.6 consists of papers on fluctuations in the microwave background radiation. These are closely connected with the fluctuations which result in galaxies. Included here is theoretical work on both large- [paper 17] and small- [papers 18,19] scale fluctuations. Also included is the important experimental result of Uson and Wilkinson [paper 20], which directly inspired papers 18 and 19. Paper 21 is a representative paper on measurements of large-scale fluctuations; for a thorough review of the experimental situation, see ref. [12].

The axion appeared as a surprise cold dark matter candidate in papers 22–24 of section 3.7. Section 3.8 takes up the subject of heavy neutrinos and supersymmetric particles as dark matter. Six years earlier, three groups independently and almost simultaneously noticed the existence of a cosmological lower bound on the mass of a heavy neutrino [papers 25–27]. If the bound is approached, the heavy neutrinos would contribute significantly to the mass density of the universe. The consequences were explored in papers 28 and 29, which contain the first discussions of structure formation in the presence of cold dark matter. The first careful treatment of the cosmological bounds on a stable supersymmetric particle is due to Goldberg [paper 30], although the possibity of a superparticle dominating the mass of the universe had been mentioned previously [26]. Goldberg made the important discovery that Majorana and Dirac fermions have quite different annihilation cross sections. A general neutralino was analyzed in paper 31, and the sneutrino in ref. [27]. A number of other studies were subsequently carried out [30,31]. Precise results for pure photinos and higgsinos (as well as heavy Majorana and Dirac neutrinos) are given in paper 32, and general neutralinos are again studied in paper 45.

Section 3.9 consists of papers on indirect detection, beginning with the idea of using low-energy antiprotons as a signature of dark-matter annihilation [papers 33,34]. Papers 35–41 discuss high-energy neutrinos from the sun due to annihilating dark matter; see also refs. [41,42]. Paper 42 suggests a possibility for axion detection. Section 3.10 takes up the subject of

direct detection, first suggested in papers 43 and 44; for a complete review and references, see ref. [44]. Paper 45 contains a detailed analysis of the possibility of detecting dark matter composed of neutralinos. Finally, the volume closes with the only direct experimental constraints on cold dark matter: papers 46 and 47 present results which rule out sneutrinos and Dirac neutrinos with mass greater than ~12 GeV, if they are present in the halo with the expected local density and velocity dispersion.

2. References

[1] S. Weinberg, Gravitation and Cosmology (Wiley, New York, 1972).
[2] P.J.E. Peebles, The Large-Scale Structure of the Universe (Princeton Univ. Press, Princeton, 1980).
[3] J. Primack, in: Proceedings of the International School of Physics "Enrico Fermi", Course XCII, ed. N. Cabibbo (North-Holland, Amsterdam, 1987) 137.
[4] A. Dekel, S.M. Faber and M. Davis, in From the Planck Scale to the Weak Scale: Toward a Theory of the Universe; Proceedings of the 1986 Theoretical Advanced Study Institute, ed. H.E. Haber (World Scientific, Singapore, 1987) 787.
[5] V. Trimble, Annu. Rev. Astron. Astrophys. 25 (1987) 425.
[6] G.R. Blumenthal, in: The Santa Fe TASI-87; Proceedings of the 1987 Theoretical Advanced Study Institute, eds R. Slansky and G. West (World Scientific, Singapore, 1988).
[7] Dark Matter in the Universe: IAU Symposium No. 117, eds J. Kormendy and G.R. Knapp (Reidel, Dordrecht, 1986).
[8] Nearly Normal Galaxies: From the Planck Time to the Present, ed. S.M. Faber (Springer, Berlin, 1987).
[9] Inflationary Cosmology, eds L.F. Abbott and So-Young Pi (World Scientific, Singapore, 1986).
[10] J. Bond and A. Szalay, Astrophys. J. 274 (1983) 443.
[11] G.S. Bisnovatyĭ-Kogan and I.D. Novikov, Sov. Astron. 24 (1980) 516.
H. Sato and F. Takahara, Prog. Theor. Phys. (Japan) 64 (1980) 2029.
F.R. Klinkhamer and C.A. Norman, Astrophys. J. 243 (1981) L1.
A.G. Doroshkevich, Ya.B. Zel'dovich, R.A. Syunyaev and M.Yu. Khlopov, Sov. Astron. Lett. 6 (1980) 252.
[12] R.B. Partridge, Rep. Prog. Phys. 51 (1988) 647.
[13] M.S. Turner, F. Wilczek and A. Zee, Phys. Lett. B 125 (1983) 35; erratum B 125 (1983) 519.
[14] J.R. Primack and G.R. Blumenthal, in: Formation and Evolution of Galaxies and Large Structures in the Universe, eds J. Audouze and J. Trân Thanh Vân (Reidel, Dordrecht, 1983); and in Fourth Workshop on Grand Unification, eds H.A. Weldon, P. Langacker and P. Steinhardt (Birkhäuser, Basel, 1983).
[15] C.S. Frenk, S.D.M. White, M. Davis and G. Efstathiou, Astrophys. J. 327 (1988) 507.
[16] J.M. Centrella, J.S. Gallagher III, A.L. Melott and H.A. Bushnose, Astrophys. J. 333 (1988) 24.
[17] A. Vilenkin, Phys. Rep. 121 (1985) 263.
[18] P.J.E. Peebles, Astrophys. J. 315 (1987) L73.
[19] M.S. Turner, A.G. Cohen and D.B. Kaplan, Phys. Lett. B 216 (1989) 20.
[20] R. Cowsik and J. McClelland, Phys. Rev. Lett. 29 (1972) 669.
G. Marx and A.S. Szalay, in Neutrino '72, eds A. Frenkel and G. Marx (Technoinform, Budapest, 1972).
A.S. Szalay and G. Marx, Astron. Astrophys. 49 (1976) 437.
[21] P. Langacker, Phys. Rep. 72 (1981) 85.
M. Srednicki, in From the Planck Scale to the Weak Scale: Toward a Theory of the Universe; Proceedings of the 1986 Theoretical Advanced Study Institute, ed. H.E. Haber (World Scientific, Singapore, 1987).
[22] E.D. Commins and P.H. Bucksbaum, Weak Interations of Leptons and Quarks (Cambridge Univ. Press, Cambridge, 1983).
[23] M. Davis, M. Lecar, C. Pryor and E. Witten, Astrophys J. 250 (1981) 423.
M.S. Turner, G. Steigman and L.M. Krauss, Phys. Rev. Lett. 52 (1984) 2090.
G. Gelmini, D.N. Schramm and J.F.W. Valle, Phys. Lett. B 146 (1984) 311.
K.A. Olive, D. Seckel and E. Vishniac, Astrophys. J. 292 (1985) 1.
K.A. Olive, D.N. Schramm and M. Srednicki, Nucl. Phys. B 255 (1985) 495.
[24] IMB Collaboration: G. Blewitt et al., Phys. Rev. Lett. 55 (1985) 2114.
[25] H.P. Nilles, Phys. Rep. 110 (1984) 1.
H.E. Haber and G.L. Kane, Phys. Rep. 117 (1985) 75.
[26] N. Cabibbo, G. Farrar and L. Maiani, Phys. Lett. B 105 (1981) 525.
S. Weinberg, Phys. Rev. Lett. 50 (1983) 387.
[27] J.S. Hagelin, G.L. Kane and S. Raby, Nucl. Phys. B 241 (1984) 638.
L.E. Ibanez, Phys. Lett. B 137 (1984) 160.
[28] UA1 Collaboration: C. Albajar et al., Phys. Lett. B 198 (1987) 261.
[29] Ya.B. Zel'dovich, Adv. Astron. Astrophys. 3 (1965) 241.
H.Y. Chiu, Phys. Rev. Lett. 17 (1966) 712.

[30] L.M. Krauss, Nucl. Phys. B 227 (1983 556; Phys. Lett. B 128 (1983) 37.
J. Bernstein, L.S. Brown and G. Feinberg, Phys. Rev. D 32 (1985) 3261.
E.W. Kolb and K.A. Olive, Phys. Rev. D 33 (1986) 1202; erratum 34 (1986) 2531.
R.J. Scherrer and M.S. Turner, Phys. Rev. D 33 (1986) 1585.
[31] K. Griest and D. Seckel, Nucl. Phys. B 283 (1987) 681; erratum B 296 (1988) 1034.
[32] S. Coleman, Aspects of Symmetry (Cambridge Univ. Press, Cambridge, 1985).
[33] V. Baluni, Phys. Rev. D 19 (1979) 2227.
R.J. Crewther, P. DiVecchia, G. Veneziano and E. Witten, Phys. Lett. B 88 (1979) 123.
[34] R.D. Peccei and H.R. Quinn, Phys. Rev. Lett. 38 (1977) 1440; Phys. Rev D 16 (1977) 1791.
S. Weinberg, Phys. Rev. Lett. 40 (1978) 223.
F. Wilczek, Phys. Rev. Lett. 40 (1978) 279.
J.E. Kim, Phys. Rev. Lett. 43 (1979) 103.
M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, Nucl. Phys. B 166 (1980) 493.
A.P. Zhitnitskii, Sov. J. Nucl. Phys. 31 (1980) 260.
M. Dine, W. Fischler and M. Srednicki, Phys. Lett. B 104 (1981) 199.
[35] J.E. Kim, Phys. Rep. 150 (1987) 1.
[36] J. Ellis and K.A. Olive, Phys. Lett. B 193 (1987) 525.
R. Mayle, J.R. Wilson, J. Ellis, K. Olive, D.N. Schramm and G. Steigman, Phys. Lett. B 203 (1988) 188.
T. Hatsuda and M. Yoshimura, Phys. Lett. B 203 (1988) 188.
G.G. Raffelt and D. Seckel, Phys. Rev. Lett. 60 (1988) 1793.
M.S. Turner, Phys. Rev. Lett. 60 (1988) 1797.
R.P. Brinkman and M.S. Turner, Phys. Rev. D 38 (1988) 2338.
[37] F. Stecker, Astrophys. J. 223 (1978) 1032.
[38] F.W. Stecker, S. Rudaz and T.F. Walsh, Phys. Rev. Lett. 55 (1985) 2622.
J.S. Hagelin and G.L. Kane, Nucl. Phys. B 263 (1986) 399.
M. Srednicki, S. Theisen and J. Silk, Phys. Rev. Lett. 56 (1986) 263; erratum 56 (1986) 1883.
S. Rudaz, Phys. Rev. Lett. 56 (1986) 2128.
J. Silk and J. Bloemen, Astrophys. J. 313 (1987) L47.
S. Rudaz and F.W. Stecker, Astrophys. J. 325 (1988) 16.
F.W. Stecker and A.J. Tylka, Astrophys. J. 336 (1989) L51.
[39] K. Griest, Phys. Rev. D 37 (1987) 2703.
[40] D.N. Spergel and W.H. Press, Astrophys. J. 294 (1985) 663.
[41] L.M. Krauss, M. Srednicki and F. Wilczek, Phys. Rev. D 33 (1986) 2079.
K. Freese, Phys. Lett. B 167 (1986) 295.
[42] M. Srednicki, K.A. Olive and J. Silk, Nucl. Phys. B 279 (1987) 804.
T.K. Gaisser, G. Steigman and S.Z. Tilav, Phys. Rev. D 34 (1986) 2206.
J.S. Hagelin, K.-W. Ng and K.A. Olive, Phys. Lett. B 180 (1987) 375.
A. Gould, Astrophys. J. 231 (1987) 560; 571.
S. Ritz and D. Seckel, Nucl. Phys. B 304 (1987) 877.
K.-W. Ng, K.A. Olive and M. Srednicki, Phys. Lett. B 188 (1987) 138.
[43] S. DePanfilis, A.C. Melissinos, B.E. Moskowitz, J.T. Rogers, Y.K. Semertzidis, W.U. Wuensch, H.J. Halama, A.G. Prodell, W.B. Fowler and F.A. Nezrick, Phys. Rev. Lett. 59 (1987) 839.
[44] J.R. Primack, D. Seckel and B. Sadoulet, Annu. Rev. Nucl. & Part. Sci. 38 (1988) 751.
[45] L.M. Krauss, K. Freese, W.H. Press and D.N. Spergel, Astrophys. J. 299 (1985) 1001.
[46] J. Faulkner and R.L. Gilliland, Astrophys. J. 299 (1985) 994.
R.L. Gilliland, J. Faulkner, W.H. Press and D.N. Spergel, Astrophys. J. 306 (1986) 703.
M. Nauenberg, Phys. Rev. D 36 (1987) 1080.
A. Renzini, Astron. Astrophys. 171 (1987) 121.
[47] G.B. Gelmini, L.J. Hall and M.L. Lin, Nucl. Phys. B 281 (1987) 726.
S. Raby and G. West, Phys. Lett. B 194 (1987) 557; Nucl. Phys. B 292 (1987) 793; Phys. Lett. B 200 (1988) 547; Phys. Lett. B 202 (1988) 47.
G.G. Ross and G.C. Segre, Phys. Lett. B 197 (1987) 45.
[48] B. Sadoulet, J. Rich, M. Spiro and D.O. Caldwell, Astrophys. J. 324 (1988) L75.
[49] F. Zwicky, Helv. Phys. Acta 6 (1933) 110.
[50] M. Aaronson and E. Olszewski in ref. [7] 153.
J. Kormendy in ref. [7] 139.
J. Madsen and R. Epstein in ref. [7] 163.
S. Tremaine in ref. [8] 76.

3. Papers

3.1. The observational evidence for dark matter

1. J. Einasto, A. Kaasik and E. Saar, Dynamical evidence on massive coronas of galaxies, Nature 250 (1974) 309–310
2. J.P. Ostriker, P.J.E. Peebles and A. Yahil, The size and mass of galaxies, and the mass of the universe, Astrophys. J. 193 (1974) L1–L4
3. V.C. Rubin, W.K. Ford and N. Thonnard, Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa → Sc, Astrophys. J. 225 (1978) L107–L111

3.2. The case against baryons as dark matter

4. D.J. Hegyi and K.A. Olive, Can galactic halos be made of baryons?, Phys. Lett. B 126 (1983) 28–32
5. J. Yang, M.S. Turner, G. Steigman, D.N. Schramm and K.A. Olive, Primordial nucleosynthesis: a critical comparison of theory and observation, Astrophys. J. 281 (1984) 493–511

3.3. Light neutrinos as dark matter

6. S.S. Gershtein and Ya.B. Zel'dovich, Rest mass of muonic neutrino and cosmology, JETP Lett. 4 (1966) 120–122
7. J.R. Bond, G. Efstathiou and J. Silk, Massive neutrinos and the large-scale structure of the universe, Phys. Rev. Lett. 45 (1980) 1980–1984

3.4. The case against light neutrinos as dark matter

8. S. Tremaine and J.E. Gunn, Dynamical role of light neutral leptons in cosmology, Phys. Rev. Lett. 42 (1979) 407–410
9. M. Aaronson, Accurate radial velocities for carbon stars in Draco and Ursa Minor: the first hint of a dwarf spheroidal mass-to-light ratio, Astrophys. J. 266 (1983) L11–L15
10. D.N.C. Lin and S.M. Faber, Some implications of nonluminous matter in dwarf spheroidal galaxies, Astrophys. J. 266 (1983) L21–L25
11. S.D.M. White, C.S. Frenk and M.Davis, Clustering in a neutrino-dominated universe, Astrophys. J. 274 (1983) L1–L5

3.5. The case for cold dark matter

12. P.J.E. Peebles, Large-scale background temperature and mass fluctuations due to scale-invariant primeval perturbations, Astrophys. J. 263 (1982) L1–L5
13. J.R. Primack and G.R. Blumenthal, Growth of perturbations between horizon crossing and matter dominance: implications for galaxy formation, in Clusters and Groups of Galaxies, eds F. Mardirossian, G. Giuricin and M. Mezzetti (Reidel, Dordrecht, 1984) p. 435–440
14. G.R. Blumenthal, S.M. Faber, J.R. Primack and M.J. Rees, Formation of galaxies and large-scale structure with cold dark matter, Nature 311 (1984) 517–525
15. M. Davis, G. Efstathiou, C.S. Frenk and S.D.M. White, The evolution of large-scale structure in a universe dominated by cold dark matter, Astrophys. J. 292 (1985) 371–394
16. S.D.M. White, C.S. Frenk, M. Davis and G. Efstathiou, Clusters, filaments, and voids in a universe dominated by cold dark matter, Astrophys. J. 313 (1987) 505–516

3.6. The microwave background: theory and experiment

17. L.F. Abbott and M.B. Wise, Large-scale anisotropy of the microwave background and the amplitude of energy density fluctuations in the early universe, Astrophys. J. 282 (1984) L47–L50
18. N. Vittorio and J. Silk, Fine-scale anisotropy of the cosmic microwave background in a universe dominated by cold dark matter, Astrophys. J. 285 (1984) L39–L43

19. J.R. Bond and G. Efstathiou, Cosmic background radiation anisotropies in universes dominated by nonbaryonic dark matter, Astrophys. J. 285 (1984) L45–L50
20. J.M. Uson and D.T. Wilkinson, Improved limits on small-scale anisotropy in cosmic microwave background, Nature 312 (1984) 427–429
21. P. Lubin, T. Villela, G. Epstein and G. Smoot, A map of the cosmic background radiation at 3 millimeters, Astrophys. J. 298 (1985) L1–L5

3.7. Axions as dark matter

22. J. Preskill, M.B. Wise and F. Wilczek, Cosmology of the invisible axion, Phys. Lett. B 120 (1983) 127–132
23. L.F. Abbott and P. Sikivie, A cosmological bound on the invisible axion, Phys. Lett. B 120 (1983) 133–136
24. M. Dine and W. Fischler, The not-so-harmless axion, Phys. Lett. B 120 (1983) 137–141

3.8. Heavy neutrinos and supersymmetric pasticles as dark matter

25. P. Hut, Limits on masses and number of neutral weakly interacting particles, Phys. Lett. B 69 (1977) 85–88
26. B.W. Lee and S. Weinberg, Cosmological lower bound on heavy-neutrino masses, Phys. Rev. Lett. 39 (1977) 165–168
27. M.I. Vysotskiĭ, A.D. Dolgov and Ya. B. Zel'dovich, Cosmological limits on the masses of neutral leptons, JETP Lett. 26 (1977) 188–190
28. J.E. Gunn, B.W. Lee, I. Lerche, D.N. Schramm and G. Steigman, Some astrophysical consequences of the existence of a heavy stable neutral lepton, Astrophys. J. 223 (1978) 1015–1031
29. G. Steigman, C.L. Sarazin, H. Quintana and J. Faulkner, Dynamical interactions and astrophysical effects of stable heavy neutrinos, Astron. J. 83 (1978) 1050–1061
30. H. Goldberg, Constraint on the photino mass from cosmology, Phys. Rev. Lett. 50 (1983) 1419–1422
31. J. Ellis, J.S. Hagelin, D.V. Nanopoulos, K. Olive and M. Srednicki, Supersymmetric relics from the big bang, Nucl. Phys. B 238 (1984) 453–476
32. M. Srednicki, R. Watkins and K.A. Olive, Calculations of relic densities in the early universe, Nucl. Phys. B 310 (1988) 693–713; erratum

3.9. Indirect detection of cold dark matter

33. J. Silk and M. Srednicki, Cosmic-ray antiprotons as a probe of a photino-dominated universe, Phys. Rev. Lett. 53 (1984) 624–627
34. S.P. Ahlen, S. Barwick, J.J. Beatty, C.R. Bower, G. Gerbier, R.M. Heinz, D. Lowder, S. McKee, S. Mufson, J.A. Musser, P.B. Price, M.H. Salamon, G. Tarlé, A. Tomasch and B. Zhou, New limit on the low-energy antiproton/proton ratio in the galactic cosmic radiation, Phys. Rev. Lett. 61 (1988) 145–148
35. W.H. Press and D.N. Spergel, Capture by the sun of a galactic population of weakly interacting massive particles, Astrophys. J. 296 (1985) 679–684
36. J. Silk, K. Olive and M. Srednicki, The photino, the sun, and high-energy neutrinos, Phys. Rev. Lett. 55 (1985) 257–259
37. IMB Collaboration: J. LoSecco et al., Limits on the flux of energetic neutrinos from the sun, Phys. Lett. B 188 (1987) 388–392
38. Frejus Collaboration, Search for dark matter and neutrino sources with the Frejus detector, in: New and Exotic Phenomena; Proceedings of the Seventh Moriond Workshop, eds O. Fackler and J. Trân Thanh Vân (Editions Frontières, Les Arcs, 1987) p. 215–220
39. Kamioka Collaboration, Search for high energy neutrinos from dark matter annihilation in the sun, in: '88 Electroweak Interactions and Unified Theories; Proceedings of the XXIIIrd Rencontre de Moriond, ed. J. Trân Thanh Vân (Editions Frontières, Les Arcs, 1988) p. 459–466
40. J. Ellis, R.A. Flores and S. Ritz, Implications for dark matter particles of searches for energetic solar neutrinos, Phys. Lett. B 198 (1987) 393–402
41. K.A. Olive and M. Srednicki, Solar neutrino searches and cold dark matter, Phys. Lett. B 205 (1988) 553–558

3.10. Direct detection of cold dark matter

42. P. Sikivie, Experimental tests of the 'invisible' axion, Phys. Rev. Lett. 51 (1983) 1415–1417; erratum 52 (1984) 695
43. M. Goodman and E. Witten, Detectability of certain dark-matter candidates, Phys. Rev. D 31 (1985) 3059–3063

44. I. Wasserman, Possibility of detecting heavy neutral fermions in the Galaxy, Phys. Rev. D 33 (1986) 2071–2078
45. K. Griest, Cross sections, relic abundance, and detection rates for neutralino dark matter, Phys. Rev. D 38 (1988) 2357–2375; erratum
46. S.P. Ahlen, F.T. Avigone III, R.L. Brodzinski, A.K. Drukier, G. Gelmini and D.N. Spergel, Limits on cold dark matter candidates from an ultralow background germanium spectrometer, Phys. Lett. B 195 (1987) 603–608
47. D.O. Caldwell, R.M. Eisberg, D.M. Grumm, M.S. Witherell, B. Sadoulet, F.S. Goulding and A.R. Smith, Laboratory limits on galatic cold dark matter, Phys. Rev. Lett. 61 (1988) 510–513

Dynamic evidence on massive coronas of galaxies

A LONGSTANDING unresolved problem in galactic astronomy is the mass discrepancy observed in clusters of galaxies. The virial mass of the cluster per galaxy and the mass–luminosity ratio are considerably larger than the corresponding quantities for individual galaxies. This discrepancy cannot be a result of expansion or be because of the recent origin of clusters: these ideas contradict our present knowledge of the physical evolution and ages of galaxies[1]. Therefore it is necessary to adopt an alternative hypothesis: that the clusters of galaxies are stabilised by hidden matter.

The discovery of the emission of X rays from clusters of galaxies can be explained in terms of hot intracluster gas. The mass of this gas, however, is insufficient to stabilise clusters[2].

Another possible way of removing the mass discrepancy is to suppose that the masses of individual galaxies have been underestimated. This hypothesis is supported by the fact that there is a very slow decrease in the rotational velocity of spiral galaxies, which indicates the pressure of large amounts of matter in their outer regions[3]. On the other hand, the discovery of large halos in elliptical galaxies[4,5] indicate that these also contain previously unseen matter.

There have been several estimates of the mass of galaxies, based on the new observational data mentioned[6,7]. A critical analysis shows, however, that these estimates are very uncertain.

We here attempt to obtain more exact information on the distribution of mass in the outer regions of galaxies.

Consider a body moving in a circular orbit of radius R, with a velocity V_c, around the centre of the galaxy. If mass distribution is spherical, then

$$M(R) = RV_c^2/G$$

is the mass of the galaxy inside the sphere of radius R, where G is the gravity constant. The same formula can be used for a nonspherical mass distribution, neglecting an error of about 10%.

Spiral galaxies contain neutral hydrogen and HII regions moving in the plane of the galaxy in nearly circular orbits. Observations of these objects permit us to calculate the function $M(R)$ in some galaxies at rather large distances from the centre. We have calculated this for five galaxies of different mass. The results for one galaxy (IC342) are given in Fig. 1.

For all galaxies we have also found the mass distribution $M_s(R)$ of known stellar populations—the bulge, the halo and the disk. We assume that these populations are physically homogeneous, in particular that the mass–luminosity ratio of stars is constant for the whole population. The relevant parameters (luminosity, L; effective radius R_0; mass–luminosity ratio, f; axial ratio of equidensity ellipsoids, ϵ) for all of the populations, have been determined from combined photometric, spectrophotometric, kinematical and radio observations, using previously described methods[7,8].

The resulting mass distribution of known stars, $M_s(R)$, deviates considerably from the mass distribution, $M(R)$, derived from the rotation curve. Rejecting the possibility of large systematic deviations from the circular velocity in the outer regions of galaxies, we conclude that galaxies contain

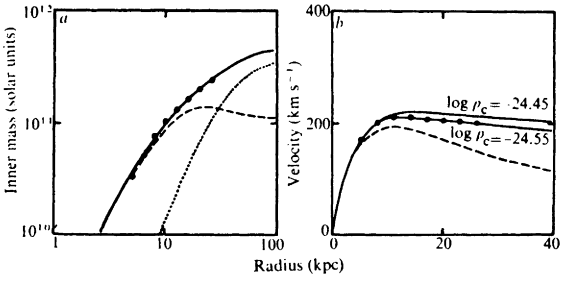


Fig. 1 The distributions of (*a*) the inner mass, $M(R)$, and (*b*) the circular velocity, V_c, in the galaxy IC342. Dots, observed values; dashed lines, model functions for known stellar populations; dotted lines, distributions for the corona; solid lines, total distributions. In Fig. 1*b* two variants of the total velocity distribution are given to demonstrate its dependence on the central density of the corona.

a previously unrecognised, massive population. It is convenient to call this population a 'corona'[7]. We have estimated the parameters of coronas for the five galaxies (Table 1) adopting, for simplicity, an exponential density law

$$\rho(R) = \rho_c \exp(-R/R_0)$$

with spherical symmetry. (A small axial ratio, ϵ, would result in a large part of the kinetic energy being in the form of rotation, although the results of Ostriker and Peebles do not indicate this[9].)

The available data permit us to derive the central density of the corona, ρ_c, with great accuracy. The length of the observed rotation curve, however, is insufficient to allow any determination of the mass and extent of the corona.

To determine total masses and dimensions of galactic coronas, more distant test bodies are needed. For this purpose isolated pairs of galaxies can be used, with secondary galaxies serving as test bodies for determining the mass distribution of primary galaxies[10].

In the case of double galaxies we know the radial component of differential velocity, $v_r = \Delta V_r$. The projection factor, $p = V_c^2/v_r^2$, which depends on the orientation of the velocity vector and on the shape of the orbit, remains unknown, but its expected value, $\langle p \rangle$, can be calculated. Therefore, the mass distribution function can be found only statistically:

$$\langle M(R) \rangle = \langle p \rangle \langle Rv_r^2 \rangle / G,$$

where $\langle Rv_r^2 \rangle$ is the corresponding mean value for a sample of pairs of galaxies.

We have collected data on 105 pairs of galaxies with known types, radial velocities and estimates of magnitudes and dimensions[11,12]. To prevent the sample from contamination by optical pairs, only those pairs with signs of interaction or with a small differential motion, $\Delta V_r/V_r < 0.23$, have been considered. The distances have been determined from the mean recession velocity, V_r, using the Hubble constant, 50 km s^{-1} Mpc^{-1}. For some nearby pairs other distance indicators have been used. The pairs have been divided into five equal groups according to the radius, R. The resulting function $\langle M(R) \rangle$ is shown in Fig. 2.

We have also calculated the average mass distribution of known stellar populations, $M_s(R)$, and have determined the distribution of mass in the corona, $M_c(R)$ (Fig. 2). The resulting central density agrees well with our previous estimates from individual galaxies. The use of distant satellities as test bodies enables us to derive masses and effective radii of coronas with a sufficient accuracy. There are enough pairs in our sample to allow an estimation of the masses of coronas separately for primary elliptical galaxies and also for bright and intermediate spiral galaxies. The estimated parameters are given in Table 2. In the case of primary elliptical galaxies the total masses of coronas may be underestimated.

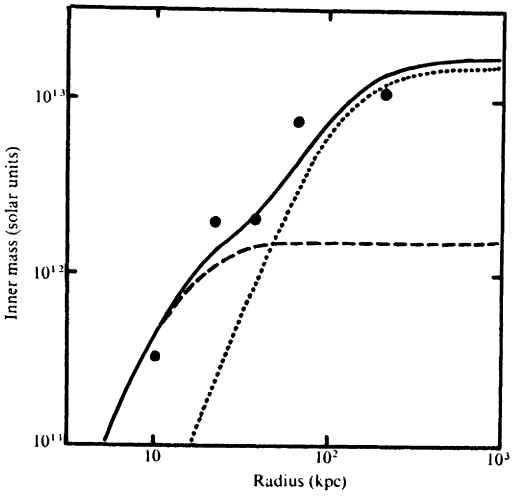


Fig. 2 The distribution of the mean inner mass, $\langle M(R) \rangle$, obtained from 105 pairs of galaxies. Symbols as in Fig. 1.

The mass of galactic coronas exceeds the mass of populations of known stars by one order of magnitude, as do the effective dimensions, by a similar factor. The central density of coronas is surprisingly constant: $\log \rho_0 = -24.5 \pm 0.1$ (g cm^{-3}).

The presence of massive coronas in galaxies considerably reduces (if not removes) the virial mass discrepancy in clusters of galaxies. The mass–luminosity ratio rises to $f \approx 100$ for spiral and $f \geqslant 120$ for elliptical galaxies. With $H = 50$ km s^{-1} Mpc^{-1} this ratio for the Coma cluster is 170 (ref. 1).

According to new estimates the total mass density of matter in galaxies is 20% of the critical cosmological density.

The possible physical nature of galactic coronas will be discussed in a separate article.

We are indebted to I. D. Karachentsev, L. M. Ozernoi, R. A. Sunyaev, and A. D. Chernin, for valuable discussions, and to Mrs U. Rümmel for her help in the preparation of data.

JAAN EINASTO
ANTS KAASIK
ENN SAAR

W. Struve Astrophysical Observatory,
202444 Tôravere, Estonia, USSR

Received April 10, 1974.

[1] Rood, H. J., Page, T. L., Kintner, E. C., and King, I. R., *Astrophys. J.*, **175**, 627 (1972).
[2] Holberg, J., Bowyer, S., and Lampton, M., *Astrophys. J. Lett.*, **180**, L55 (1973).
[3] Roberts, M. S., in *Stars and Stellar Systems* (edit. by Sandage, A.), 9 (in the press).
[4] de Vaucouleurs, G., *Astrophys. Lett.*, **4**, 17 (1969).
[5] Arp, H., and Bertola, F., *Astrophys. Lett.*, **4**, 23 (1969).
[6] Bertola, F., and Capaccioli, M., *Asiago Contr. No. 234* (1970).
[7] Einasto, J., *Tartu astr. Obs. Teated*, **40**, 3 (1972).
[8] Einasto, J., and Kaasik, A., *Astr. Circ. USSR*, **790**, 1 (1973).
[9] Ostriker, J. P., and Peebles, P. J. E., *Astrophys. J.*, **186**, 467 (1973).
[10] Einasto, J., Saar, E., Kaasik, A., and Traat, P., *Astr. Circ. USSR*, **811** (1974).
[11] Karachentsev, I. D., *Astrofizika*, **2**, 81 (1966).
[12] Page, T., *Astrophys. J.*, **159**, 791 (1970).
[13] Karachentsev, I. D., *Soobshch. spec. astr. Obs.*, **7**, 3 (1972).

THE ASTROPHYSICAL JOURNAL, 193:L1–L4, 1974 October 1
© 1974. The American Astronomical Society. All rights reserved. Printed in U.S.A.

THE SIZE AND MASS OF GALAXIES, AND THE MASS OF THE UNIVERSE

J. P. OSTRIKER
Princeton University Observatory

P. J. E. PEEBLES
Joseph Henry Laboratories, Princeton University

AND

A. YAHIL
Princeton University Observatory; and Department of Physics, Tel-Aviv University
Received 1974 May 28; revised 1974 July 15

ABSTRACT

Currently available observations strongly indicate that the mass of spiral galaxies increases almost linearly with radius to nearly 1 Mpc. This means that the total mass per giant spiral is of the order of $10^{12}\ M_\odot$, and that the ratio of this mass to the photographic light within the Holberg radius, f, is $\sim 200\ (M/L)_\odot$. Using this value of f and the luminosity function of surveyed galaxies, we determine a local mean cosmological mass density $\approx 2 \times 10^{-30}$ g cm^{-3} corresponding to $\Omega \equiv \rho/\rho_{crit} \approx 0.2$. The uncertainty in this result is not less than a factor of 3.

Subject headings: cosmology — galactic structure

I. THE ARGUMENT

There are reasons, increasing in number and quality, to believe that the masses of ordinary galaxies may have been underestimated by a factor of 10 or more. Since the mean density of the Universe is computed by multiplying the observed number density of galaxies by the typical mass per galaxy, the mean mass density of the Universe would have been underestimated by the same factor. Finally, the current estimate (Shapiro 1971) for the ratio of gravitational energy to kinetic energy in the Universe is about $\Omega = 0.01$. If we increase the estimated mass of each galaxy by a factor well in excess of 10, we increase this ratio by the same amount and conclude that observations may be consistent with a Universe which is "just closed" ($\Omega = 1$)—a conclusion believed strongly by some (cf. Wheeler 1973) for essentially nonexperimental reasons.

There are two very important facts about astronomical measurements of mass that must be kept in mind. First, almost all *measurements* of mass are based on the requirement that gravitational forces of attraction balance inertial forces in a system of mutually interacting gravitating masses which is not steadily expanding or contracting: masses found by observed light multiplied by *assumed* mass-to-light ratios are not measured. Second, no gravitational information can reach an observer from matter which is distributed homogeneously on spherical shells external to the observer. Thus, mass measurements using two gravitationally interacting objects (pairs of galaxies, or one gas cloud in orbit about a galaxy) provide no knowledge about spherically distributed mass, if any, surrounding the two objects. This result, although arising from a quirk of symmetry and the inverse square law, remains approximately valid for moderately large departures from spherical symmetry. Thus, the rotation curve interior to the Sun's orbit gives little information about the disk mass exterior to the Sun ($r > 10$ kpc) and almost no information about the exterior halo mass. Similarly the studies by Page (1961) of galaxies' masses in binary pairs contain little information concerning distributed mass with a scale greater than the typically $\sim$30 kpc separation.

This ignorance would give us little cause for concern if we had reason to believe from measurements with $r < 20$ kpc that the mass $M(r)$ had "converged" to some limiting value. However, this is not the case; available evidence, summarized below, indicates that the measured masses of galaxies diverge with increasing distance even though the luminosities of the flattened components of spirals and S0 galaxies do appear to be convergent. The estimated rate of divergence of M with r varies between weak (logarithmic) to strong (linear), depending on the method of measurement. The best evidence suggests that within local giant spiral galaxies $M(r) \propto r$ for 20 kpc $\lesssim r \lesssim$ 500 kpc.

The implied density distribution thus appears to be similar to that in the outer parts of isothermal gas spheres. If the observed galaxies are in fact embedded in enormous isothermal spheres comprised of optically faint mass points, the latter may be the result of the collapse phase (cf. Eggen, Lynden-Bell, and Sandage 1963) and subsequent violent relaxation (Lynden-Bell 1967) within an original assemblage of stars, globular clusters, etc., that separated from other parts of the Universe at the period of galaxy formation and now exists as a halo with a very high mass-to-light ratio surrounding the more visible, largely second-generation part of the galaxy.

Regardless of the validity of the speculative scenario outlined in the last paragraph, the dynamically observed masses of the best studied galaxies together with their luminosity can be used to determine a mass-to-light ratio. We shall show that, for giant spirals (M/L)

$\simeq 200\ t_{10}^{-1}(M_\odot/L_\odot)$, where M is the mass within ~ 0.5 Mpc, L is the easily measured light within ~ 0.02 Mpc of the center, and t_{10} is the age of the Universe in units of 10^{10} years. This value for M/L, combined with the observed galaxy luminosity-function within several Mpc, allows us to estimate the local mass density at $2 \times 10^{-30}\ h^1 t_{10}^{-1}$ g cm^{-3} in spiral galaxies giving $\Omega \equiv (\rho/\rho_{\rm crit}) \approx 0.2 \times 10^{0\pm 0.5}$ [for $h \equiv$ Hubble constant/(100 km s^{-1} Mpc^{-1}) = 0.5]. A similar conclusion, based on similar arguments, has recently been reached by Einasto, Kaasik, and Saar (1974).

II. THE MASS AND MASS-TO-LIGHT RATIO OF SPIRAL GALAXIES

Since we are explicitly noting that the mass determined for a galaxy will depend on the position of the probe measuring the mass, we will need to consider a variety of methods to obtain the function $M(r)$. The following discussion, based largely on the best observed local giant spiral galaxies, is summarized in table 1 and figure 1.

For 10 kpc $< r <$ 30 kpc, the rotational velocity of neutral hydrogen provides the best measure (optical measurements pertain primarily to gas with $r < 10$ kpc). In a recent study of three local giant spirals (M31, M81, M101) Roberts and Rots (1973), whose results provide entries 2, 5, and 6 of table 1, pointed out that in the outer parts "the three galaxian rotation curves decline slowly, if at all" indicating that $M \propto r$. A similar result was found by Rogstad and Shostak (1972) for local Sc galaxies.

Double galaxies, whose masses may be estimated with the virial theorem, provide another estimate of mass at $r \approx 30$ kpc. The standard study by Page (1961) gives values (entry 3) which are surprisingly an order of magnitude smaller than the rotational mass estimates at the same radius. Another sample of apparent pairs in the Virgo cluster studied by van den Bergh (1960), however, gives considerably larger estimates for M. We have (conservatively) removed three pairs from the latter study, which, from their anomalously large values of $(\Delta v)^2$, are likely to be "optical" pairs rather than bound systems, and have recalculated the mean mass (entry 4). Recently Jones (1973) has also examined the binary galaxy problem, focusing particularly on wide separations; from his figure 1 we have selected the nine pairs having separation closest to 100 kpc and calculated the mass in the conventional manner (entry 8). We plot in figure 1 the points for $M(r)$ from binaries assuming $h = 1$; the heads of the three arrows indicate the positions of the corresponding points if $h = 0.5$. It is clear that the discrepancy between Page's binaries and the rotationally derived masses is not due to an uncertainty in h; it may arise from the particular selection rules used to determine the sample of pairs examined.[1]

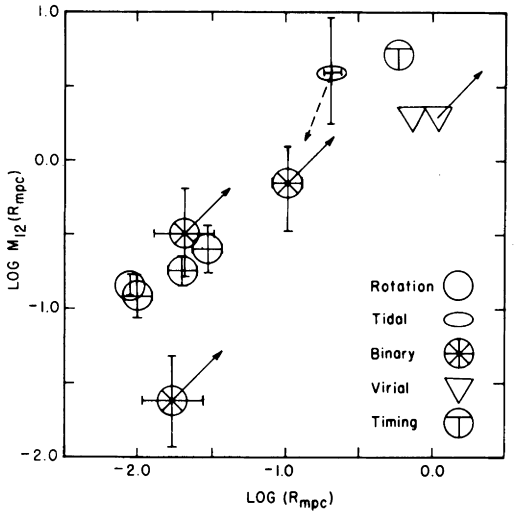


FIG. 1.—Mass (unit $10^{12}\ M_\odot$) of local giant spiral galaxies within a distance (R/1 Mpc) of their centers, as determined by various methods.

Dwarf spheroidal companions of our Galaxy are tidally limited, as noted by Hodge (1966). From the

[1] In a subsequent assay at the problem, Page (1962) revised the masses of spirals *downward* slightly thus increasing the discrepancy; we did not use this later work because galaxy separations were not quoted.

TABLE 1

$M(r)$ FOR SPIRAL GALAXIES

Object	#	$\log_{10}(R_{\rm Mpc})$	$\log_{10}(M_{12})$	Method
1. Galaxy	1	$-2.05\pm$?	-0.84 ± 0.07	Rotation
2. Local giants	3	-2.00 ± 0.10	-0.92 ± 0.14	Rotation
3. Field doubles	20	-1.77 ± 0.20	-1.62 ± 0.30	Binary
4. Virgo doubles	23	-1.68 ± 0.20	$-0.5\ \pm0.3$	Binary
5. Local giants	3	-1.69 ± 0.10	-0.74 ± 0.10	Rotation
6. Local giants	3	-1.52 ± 0.10	-0.60 ± 0.16	Rotation
7. Dwarf spheroidals	3	-1.14 ± 0.04	-0.60 ± 0.55	Tidal
8. Field doubles	9	-0.98 ± 0.11	-0.15 ± 0.32	Binary
9. Dwarf spheroidals	3	-0.69 ± 0.03	$+0.59\pm0.35$	Tidal
10. M81 group	7	$-0.15\pm$?	$+0.30\pm$?	Virial
11. M31 group	6	$-0.22\pm$?	$+0.72\pm$?	Timing
12. Local groups	200	$+0.04\pm$?	$+0.30\pm$?	Virial

straightforward tidal theory one can derive the mass of our Galaxy interior to the perigalacticon of the tidally truncated companion galaxy. The six galaxies, considered in two groups, thus measure the mass of our galaxy at (roughly) two radii (see entries 7 and 9). Only the distant group is plotted in figure 1 due to the large error in the estimate (determined from quoted errors in Hodge 1966) from the other group. The ellipse at the tail of the arrow is for a circular orbit; the head points to the determination if, in fact, the orbits are highly eccentric with perigalacticon at one-half the present radius for each companion galaxy.

At large separation three further values of $M(r)$ are given. The mass for the M81 group (entry 10) is based primarily on the two outlying members. If these two are omitted, the radius and mass decrease by a factor 10. The Local (M31) group mass determination (entry 11) is based on the requirement that an initially separating pair of galaxies reverse and, after 10^{10} years, achieve the relative velocity and position reached by our Galaxy and Andromeda; for the method see Peebles (1971) and references therein. The numerical result given in table 1 was based on a distance to Andromeda of 700 kpc, an approach velocity of 300 km s^{-1} (heliocentric), the standard solar motion, and an assumed local galactic rotation rate of 230 km s^{-1}, the quoted mass scales roughly as t_{10}^{-1}, due to dependence on the allowed orbital time. Gunn (1973), using the identical method but somewhat different kinematical parameters, derived a smaller mass ($\log M_{12} = 0.46$) for the Local Group. Kahn and Woltjer (1959) and Oort (1970) with similar methods have derived masses in the range $(1\text{–}10) \times 10^{12}$ $M_\odot$. Finally, Geller and Peebles's (1973) study of wide groups (entry 12) was designed to statistically compensate for projection effects.

Although the values of $M(r)$ given in table 1 are among the best that could be obtained from the current literature, the uncertainties, which have been roughly indicated by error bars, are obvious. However, the general trend seen in figure 1, of significantly increasing mass with increasing radius, is almost certainly real. It has been found in other studies of galaxy masses (cf. Rood, Rothman, and Turnrose 1970; Field and Saslaw 1971; Rood 1974), where it was sometimes expressed as an increase in the "virial discrepancy" with increasing size of group.

Let us designate by M/L the ratio of the total galactic mass to the photographic light within the easily measured radius of $\sim$20 kpc (essentially the Holmberg radius) in solar units. Then, for the Local Group and the M81 group we have $f_{sp} \equiv (M/L)_{spiral} \approx (5 \times 10^{12})/(2.5 \times 10^{10}) = (2 \times 10^{12})/(1 \times 10^{10}) = 200$. Since these mass determinations do not depend significantly on the Hubble constant but the result for the Local Group does depend on t_{10}, we can write this

$$f_{sp} \simeq 200 h^0 t_{10}^{-1} \qquad (1)$$

compared with the mass-to-light ratio of elliptical galaxies

$$f_{el} \simeq 300 h^1$$

derived by Rood *et al.* (1972) from their study of the Coma cluster.

The very large mass-to-light ratio and the very great extent of spiral galaxies can perhaps most plausibly be understood as due to a giant halo of faint stars. Such a structure, which appears superficially so improbable, has been proposed recently (Ostriker and Peebles 1973) for quite different reasons concerning dynamical stability. The least troublesome way of ensuring the stability of a cold disk of stars against nonaxisymmetric disturbances is to suppose that the spherically distributed (halo) mass *interior* to $r \sim 10$ kpc is substantial. This requires the interior halo to have a large mass-to-light ratio ($f \sim$ 50–100). If, further, the spherical halo *exterior* to 10 kpc is as extensive as the halo of an elliptical (cf. here Kormendy and Bahcall 1974 with Arp and Bertola 1971) and the mass-to-light ratio stays large, then the mass associated with the outer parts of spiral galaxies must be large on these grounds.

III. COSMOLOGICAL IMPLICATIONS

The total space density associated with galaxies is

$$\rho_{gal} = f_{sp} j_{sp} + f_{el} j_{el}$$
$$= 4.0 \times 10^{-30} h^1 t_{10}^{-1} + 0.4 \times 10^{-30} h^2 \text{ g cm}^{-3}, \qquad (2)$$

where for j_{sp}, j_{el}, the mean volume emissivities of the light from spiral and elliptical galaxies, we have taken 3×10^8 $hL_\odot$ Mpc^{-3} and 2×10^7 $hL_\odot$ Mpc^{-3}, respectively, from Shapiro's (1971) analysis of the de Vaucouleurs catalog. Since the critical density is 1.9×10^{-29} h^2 g cm^{-3}, we have

$$\Omega_{gal} = (\rho_{gal}/\rho_{crit}) = \frac{0.21}{ht_{10}} + 0.02 \geq 0.2 \qquad (3)$$

since $ht_{10} < 1$.

The uncertainty of this result, due to inaccuracy in f_s alone, is clear from the preceding discussion. The total uncertainty, which is not less than a factor of 3, contains an important contribution due to our poor knowledge of j_s as well. Gunn (1973) has pointed out that, since the de Vaucouleurs catalog is based on the neighborhood of our Galaxy, it is influenced by the local supercluster and may overestimate j_s by as much as $10^{0.5}$. However, estimates of $(j_{sp} + j_{el})$, based on counts of galaxies in deeper surveys, are consistent within a factor $\sim$1.5 of the values adopted here. In particular, Oort's (1958) analysis of the mean emissivity within a sphere of radius $\sim 300\, h^{-1}$ Mpc also gives $j \approx 3 \times 10^8$ h, the value we have adopted from the survey of our local region. Furthermore, the brightness of the night sky (see references in Peebles 1971 and Mattila 1973), which integrates over all galaxies including those of low surface brightness or small angular size that might be missed in the standard survey, would permit $(j_{sp} + j_{el})$ larger than the adopted value by a factor 3–10; however, if the uncatalogued galaxies have the same spatial distribution as the counted ones, then Shectman's (1973) analysis of the fluctuations in the background indicates an emissivity of $j \lesssim 5 \times 10^8$ h.

There are various other cosmological arguments which can be given for either a low- or high-density Universe. Most are well known and will not be repeated here. Two new points seem especially significant. The *Copernicus* determination of interstellar deuterium (Rogerson and York 1973) coupled with conventional (or unconventional) big-bang nuclear calculations (cf. Wagoner 1973) strongly indicates an empty ($\Omega = 0.01$) Universe *unless* subsequent production of deuterium has been significant (Colgate 1973; Hoyle and Fowler 1973). Finally, the great extent of rich clusters of galaxies (Yahil 1974; Rood *et al.* 1972) and the correlations observed among galaxy positions, when interpreted from the viewpoint of gravitational instability (Peebles 1974), appear to indicate that $\Omega \approx 1$.

The arguments presented above indicate that *within the current observational uncertainties* the masses associated with ordinary spiral galaxies may make a cosmologically interesting contribution. Further observational work on (*a*) the nightsky background light due to unresolved galaxies, (*b*) optical or radio searches for giant halos surrounding ordinary galaxies, and (*c*) the puzzling problem of binary mass determinations would be especially rewarding at this time.

We would like to express our thanks to Drs. J. E. Gunn, J. H. Oort, M. Roberts, J. Silk, and L. Spitzer for helpful conversations concerning this manuscript, to thank the National Science Foundation (grants GP34888X and GP30799X2) for material support, and the Astronomy Department of the University of California, Berkeley, for hospitality in the spring of 1973 when this work was begun.

REFERENCES

Arp, H., and Bertola, F. 1971, *Ap. J.*, **163**, 195.
Colgate, S. 1973, *Ap. J.* (*Letters*), **181**, L53.
Eggen, O. J., Lynden-Bell, D., and Sandage, A. R. 1963, *Ap. J.*, **136**, 748.
Einasto, J., Kaasik, A., and Saar, E. 1974, Preprint 1 of Tartu Observatory.
Field, G. B., and Saslaw, W. C. 1971, *Ap. J.*, **170**, 199.
Geller, M., and Peebles, P. J. E. 1973, *Ap. J.*, **184**, 329.
Gunn, J. E. 1973, Orange Aid Preprint 338.
Hénon, M. 1968, *Bull. Astr. Series 3*, **3**, 241.
Hodge, P. W. 1966, *Ap. J.*, **144**, 869.
Hoyle, F., and Fowler, W. A. 1973, *Nature*, **241**, 384.
Jones, B. T. 1973, preprint.
Kahn, F. D., and Woltjer, L. 1959, *Ap. J.*, **130**, 105.
Kormendy, J., and Bahcall, J. N. 1974, *A.J.*, **79**, 671.
Lynden-Bell, D. 1967, *M.N.R.A.S.*, **136**, 101.
Mattila, K. 1973, preprint.
Oort, J. H. 1958, *Distribution of Galaxies and the Density in the Universe*, Institut International de Physique Solvay, Brussels.
———. 1970, *Astr. and Ap.*, **7**, 381.
Ostriker, J. P., and Peebles, P. J. E. 1973, *Ap. J.*, **186**, 467.
Page, T. 1961, *The Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Berkeley: University of California Press), p. 277.
———. 1962, *Ap. J.*, **136**, 685.
Peebles, P. J. E. 1971, *Physical Cosmology* (Princeton: Princeton University Press).
———. 1974, *Ap. J.* (*Letters*), **189**, L51.
Roberts, M. S., and Rots, A. H. 1973, *Astr. and Ap.*, **26**, 483.
Rogerson, J. B., and York, D. G. 1973, *Ap. J.* (*Letters*), **186**, L95.
Rogstad, D. H., and Shostak, G. S. 1972, *Ap. J.*, **176**, 315.
Rood, H. J. 1974, *Ap. J.*, **188**, 451.
Rood, H. J., Page, T. L., Kintner, E. C., and King, I. R. 1972, *Ap. J.*, **175**, 627.
Rood, H. J., Rothman, V. C. A., and Turnrose, B. E. 1970, *Ap. J.*, **162**, 411.
Shane, C. D., and Wirtanen, C. A. 1967, *Pub. Lick Obs.*, **22**, 1.
Shapiro, S. L. 1971, *A.J.*, **76**, 291.
Shectman, S. A. 1974, *Ap. J.*, **188**, 233.
van den Bergh, S. 1960, *M.N.R.A.S.*, **121**, 387.
Wagoner, R. 1973, *Ap. J.*, **179**, 343.
Wheeler, J. A. 1973, "The Universe as a Home for Man," presented in Washington, D.C., on the occasion of the 500th anniversary of the birth of Copernicus, to be published.
Yahil, A. 1974, *Ap. J.*, **191**, 623.

J. P. OSTRIKER: Princeton University Observatory, Peyton Hall, Princeton, NJ 08540

P. J. E. PEEBLES: Joseph Henry Laboratories, Jadwin Hall, Princeton University, Princeton NJ 08540

A. YAHIL: Department of Physics, Tel-Aviv University, Tel-Aviv, Israel.

THE ASTROPHYSICAL JOURNAL, 225:L107–L111, 1978 November 1
© 1978. The American Astronomical Society. All rights reserved. Printed in U.S.A.

EXTENDED ROTATION CURVES OF HIGH-LUMINOSITY SPIRAL GALAXIES. IV. SYSTEMATIC DYNAMICAL PROPERTIES, Sa→Sc

VERA C. RUBIN,*† W. KENT FORD, JR.,* AND NORBERT THONNARD
Department of Terrestrial Magnetism, Carnegie Institution of Washington
Received 1978 June 7; accepted 1978 July 18

ABSTRACT

For a sample of 10 high-luminosity spiral galaxies, Sa through Sc, we have obtained accurate rotation curves which extend to about 80% of the de Vaucouleurs radii. For this sample: (1) All rotation curves are approximately flat, to distances as great as $r = 50$ kpc. Secondary velocity undulations indicate rotational velocities lower by about 20 km s^{-1} on the inner edges than on the outer edges of spiral features. (2) V_{max} is correlated with Hubble type, and total mass M is a function of both V_{max} and radius. At equal radii, M(Sa) > M(Sc). Hence, surface mass density decreases systematically along the Hubble sequence. (3) V_{max} is not correlated with luminosity or with radius. Galaxies with similar V_{max} have radii and luminosities which differ by factors of 2 and 3. This implies a large intrinsic scatter in the Tully-Fisher relation. (4) Masses are a few times $10^{11}M_{\odot}$ out to the de Vaucouleurs radius, and some masses approach $10^{12}M_{\odot}$ out to the Holmberg radius. M/L_B ratios are low, near 3.5. There is a weak suggestion that M/L_B is higher for early-type galaxies. While this is contrary to the accepted result of Roberts, masses for early-type galaxies are systematically low in his sample, due to an extrapolation procedure based on falling rotation curves.

Subject headings: galaxies: individual — galaxies: internal motions — galaxies: structure

I. INTRODUCTION

In the 50 years since Hubble (1926) introduced his classification sequence for galaxies, few systematic observational programs have attempted to study dynamical properties of galaxies as a function of Hubble type (HT). The constraints have been principally instrumental. Optical rotation curves (e.g., Burbidge and Burbidge 1975) have furnished valuable dynamical information, but generally only for the inner regions of late-type spirals. Neutral hydrogen observations have revealed integral properties of gas-rich systems (Roberts 1975), but with limited spatial resolution.

Available optical instrumentation now permits the detection of emission across a very large portion of the disks of spirals, well beyond the "turnover point" in the rotation curves. We have initiated a program to obtain spectra of spiral galaxies at high velocity resolution and at a large spatial scale, in order to study their properties as a function of HT. We now have velocities for 10 spirals, Sa through Sc, of high intrinsic luminosity. The galaxies, chosen with extreme care, have: angular diameters near 3′ or 4′ to match the KPNO/CTIO spectrograph slit lengths; high inclination so that uncertainties in inclinations produce little effect on rotational velocities and hence masses; high luminosity as indicated by the widths of their 21 cm profiles (if available); and large linear diameters. Two of the galaxies, NGC 4378 and NGC 7217, were studied originally for other reasons. These, plus NGC 3672, have been discussed separately (NGC 3672, Rubin, Thonnard, and Ford 1977 [Paper I]; NGC 4378, Rubin *et al.* 1978 [Paper II]; NGC 7217, Peterson *et al.* 1978 [Paper III]). Distances are determined from the corrected velocities, and we adopt $H = 50$ km s^{-1} Mpc^{-1} throughout.

* Visiting Astronomer, Kitt Peak National Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc., under contract with the National Science Foundation.

† Visiting Astronomer, Cerro Tololo Inter-American Observatory, which is supported by the National Science Foundation under contract AST 74-04128.

II. THE ROTATION CURVES

Optical spectra were obtained with the Kitt Peak and Cerro Tololo 4 m spectrographs plus Carnegie image tube, usually at 25 Å mm^{-1}. Errors in the rotational velocities (measuring errors plus projection uncertainties) are generally less than ±8 km s^{-1} per point. Reproductions of eight spectra are shown in Figure 1 (Plate L3) (Papers II and III show the other two). Many striking features can be observed directly from the spectra. (1) *All rotation curves are approximately flat*, with only a slight rise or fall following the initial steep gradient. (2) Secondary velocity undulations occur at positions of spiral features. Rotational velocities are ~20 km s^{-1} lower on the inner edges of the arms than on the outer edges. This is especially noticeable in NGC 2998, whose velocities are plotted in Figure 2. While this confirms a prediction of the density wave theory, velocity gradients will exist also in any gravitational model with a mass concentration in the arms. (3) Nuclear emission is often strong, often exhibits a large velocity gradient, and often remains intense several kiloparsecs beyond the nucleus. There follows

a region of weaker emission with a shallow velocity minimum. Beyond this minimum the emission increases in intensity, with a corresponding velocity increase where the first spiral feature is encountered. (4) Star formation is proceeding vigorously over great distances in a single galaxy. In NGC 801, emission extends to a radius of 49 kpc. (5) Nuclear spectra come in a wide variety; emission lines can be narrow with a large velocity gradient (NGC 801), or broad, with the entire gradient within the nucleus (NGC 7541).

Parameters for these galaxies discussed are listed in Table 1. Classifications come from de Vaucouleurs, de Vaucouleurs, and Corwin (1975) (hereafter RC2) and/or Sandage (1978). Three galaxies not listed in

TABLE 1

DATA FOR PROGRAM GALAXIES

NGC (1)	Hubble Class (2)	Source (3)	Heliocentric Velocity Opt. (km s^{-1}) (4)	Heliocentric Velocity 21 cm (km s^{-1}) (5)	V_0 Opt. (km s^{-1}) (6)	Distance (Mpc) (7)	i (degrees) (8)	Radius r_{dev} (kpc) (9)	Nuclear Distance of Last Velocity (kpc) (10)	Fraction r_{dev} Observed (11)	Rotation Curve at Large r (12)
4378..	Sa I	deV;S	2540±12	2536	2428	48.6	35	23.4	22.0	0.94	Falling
4594..	Sa	deV	1076±10		911	18.2	84	23.6	15.	0.64	Rising
7217..	Sb-Sab III	deV;S	955±6	954±10	1236	24.7	35	13.8	11.0	0.80	Falling
2590..	Sb	N	4990±25	4992±20	4793	95.9	75	34.8	17.4	0.50	Rising
1620..	Sbc	deV	3496±10	3509±8	3424	68.5	70	30.8	21.9	0.71	Rising
3145..	Sbc I	deV;S	3650±15	3648±20	3414	68.3	61	30.8	25.3	0.82	Flat
801...	Sbc-Sc	N	5763±10	5764±20	5948	119.	86	57.1	49.1	0.86	Flat or rising
7541..	Sbc-Sc III	deV;S	2685±10	2665:	2873	57.5	72	28.6	23.2	0.81	Rising
7664..	Sbc-Sc	N	3464±10	3481±15	3709	74.2	58	35.6	28.1	0.79	Flat or rising
2998..	Sc I	deV;S	4767±10	4777±5	4781	95.6	62	39.1	34.0	0.87	Flat
3672..	Sc I-II	deV;S	1857±10	1867±10	1655	33.1	70	19.8	17.6	0.89	Flat

NGC (1)	Hubble Class (2)	V_{max} Opt. (km s^{-1}) (13)	V_{max} 21 cm (km s^{-1}) (14)	Disk Mass to Last Velocity (10^{11} $M_\odot$) (15)	Mass to r_{dev} (10^{11} $M_\odot$) (16)	$\int SdV$ (Jy km s^{-1}) (17)	H I (10^{10} $M_\odot$) (18)	H I/ $M(r_{dev})$ (19)	m_B (mag) (20)	$L_B^{b,i}$ (10^{10} $L_\odot$) (21)	$M(r_{dev})/L_B^{b,i}$ (22)
4378..	Sa I	320	323	3.05	3.24/3.89	6.6	0.37	0.01	12.28 S	4.53±0.4	8.6±2
4594..	Sa	344		2.8							>5
7217..	Sb-Sab III	290	313	1.13	1.41/1.55	11.7	0.17	0.01	11.08 S	4.41±0.4	3.5±1
2590..	Sb	262	251	1.74	4.11/4.52	18.3	8.1	0.18	13.4 P	10.5±2	4.3±2
1620..	Sbc	248	231	1.95	2.96	27.7	4.7	0.16	12.62 P	9.20±2	3.2±1
3145..	Sbc I	251	273	2.55	3.11	37.7	4.1	0.13	12.35 S	10.5±1	3.0±1
801...	Sbc-Sc	248	224	3.73	4.40	15.8	11.	0.25	13.1 Z	23.8±9	1.8±2
7541..	Sbc-Sc III	238	236	1.92	2.46				12.45 S	7.05±0.7	3.5±1
7664..	Sbc-Sc	208	204	1.77	2.36	29.7	4.0	0.17	12.9 Z	7.38±3	3.2±1
2998..	Sc I	211	213	2.38	2.74	25.1	5.4	0.20	12.65 S	14.9±1.4	1.8±1
3672..	Sc I-II	208	223	1.07	1.22	66.3	2.7	0.22	11.66 S	4.45±0.4	2.7±1

NOTES TO TABLE 1

Col. (3).—N, Nilson 1973. S, Sandage 1978. deV, de Vaucouleurs *et al.* 1976.

Col. (5).—Thonnard *et al.* 1978 except: NGC 4378, Krumm and Salpeter 1976; NGC 7217, Paper III; NGC 7451, Shostak 1978, contaminated by companion, not included in optical–21 cm comparison.

Col. (6).—$V_0 = V_H + 300 \sin l \cos b$.

Col. (7).—Distance $= V_0/H$; $H = 50$ km s^{-1} Mpc^{-1}.

Col. (8).—$\cos^2 i = [1.042\, b^2/a^2 - 0.042]$; b/a from RC2 or Nilson 1973.

Col. (9).—Radius to 25 mag arcsec^{-2} from de Vaucouleurs *et al.* 1976 or Nilson 1973.

Col. (13).—V_{max} from smooth rotation curve.

Col. (14).—$\Delta V/2 \sin i$; ΔV read at 25% of mean flux, corrected for filter width; see n. (5).

Col. (15).—Factor of 1.1 (Brandt 1960) included.

Col. (16).—/Mass × 1.2 (Sa). /Mass × 1.1 (Sab, Sb). $\epsilon_M = \pm 25\%$; except NGC 2590, $\epsilon_M = 50\%$.

Col. (17).—See n. (5).

Col. (18).—Corrected for optical depth effects at 21 cm after Roberts 1975.

Col. (19).—See n. (18).

Col. (20).—P, Peterson unpublished; $\sigma_m = \pm 0.4$ mag. S, Sandage 1978; $\sigma_m = \pm 0.1$ mag. Z, Zwicky *et al.* 1961–68, corrected by $\Delta m = -0.4$ mag. Correction established by 4 galaxies in common with Sandage 1978, $\sigma_m = \pm 0.4$ mag.

Col. (21).—Corrected for: $A_b = 0.13$ (csc $b - 1$); $A_i = 0.8 \log a/b$ (RC 2); σ_L comes from σ_m (n. 20).

Col. (22).—Error estimate from ϵ_M and σ_L.

NOTE.—All data for NGC 4594 are from Schweizer 1978.

RC2 have Nilson (1973) classifications. Sources for other data and corrections applied are indicated in the notes. Differences between the optical and the 21 cm velocities are small: $\langle V_{21} - V_{opt} \rangle = 5$ km s^{-1}. However, the mean velocity difference for four galaxies with velocities in RC2 is (ours−catalog) = 77 km s^{-1}. Velocity errors this large still permeate the best available catalogs. Columns (9), (10), and (11) give the galaxy radius (25 mag arcsec^{-2}; RC2), the radius of the last measured velocity, and the ratio of the two. In the mean, our velocities extend over 80% of the galaxy radius. V_{max}, the peak velocity of the optical rotation curve (in the plane of the galaxy), is listed in column (13). Data for NGC 4594 obtained by Schweizer (1978) with the same CTIO equipment are also included.

Rotation curves are plotted in Figures 3 and 4. The general flatness of the curves, and the pronounced increase in V_{max} with earlier HT, are notable. A plot of V_{max} versus HT, Figure 5, shows this tight correlation. A correlation between V_{max} and HT found earlier by Brosche (1971) lies about 50 km s^{-1} below that indicated in Figure 5, and is defined principally from galaxies with types later than Sbc. The correlation indicated in Figure 5 may represent an upper envelope defined by high-luminosity galaxies.

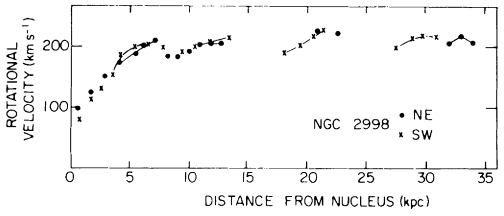


FIG. 2.—Rotational velocities in NGC 2998, as a function of distance from nucleus. Velocities for strongest emission regions are connected with lines. Note fairly good velocity agreement between velocities from NE and SW major axes, and positive velocity gradient across each arm.

III. MASSES AND MASS-TO-LIGHT RATIOS

Masses and mass distributions have been determined from the rotation curves by two procedures: (1) disk-modeled galaxies, which give lower limits, and (2)

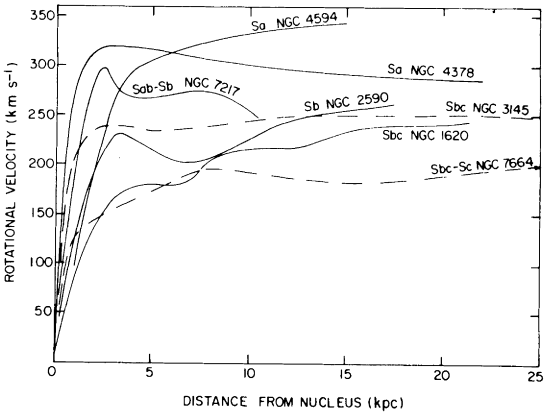


FIG. 3.—Rotational velocities for seven galaxies, as a function of distance from nucleus. Curves have been smoothed to remove velocity undulations across arms and small differences between major-axis velocities on each side of nucleus. Early-type galaxies consistently have higher peak velocities than later types.

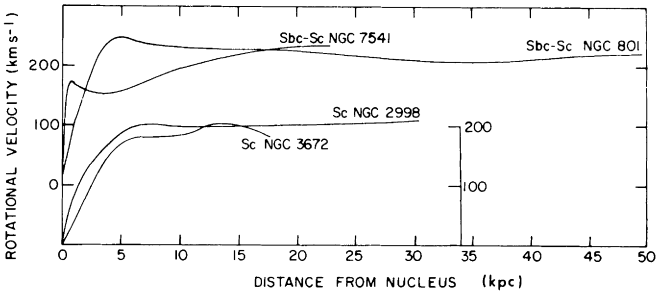


FIG. 4.—Rotation curves for two pairs of galaxies, which illustrate the lack of Tully-Fisher relation. NGC 7541 and NGC 801, both Sbc–Sc, have V_{max} values of 238 and 248 km s^{-1}. However, their luminosities (7.05 ± 0.7 and 23.8 ± 9 × 10^{10} $L_\odot$) and radii (23.2 and 49.1 kpc) differ by factors of 3 and 2. Similarly, the Sc galaxies NGC 2998 and NGC 3672 have V_{max} of 211 and 208 km s^{-1}, but luminosities 14.9 ± 1.4; 4.45 ± 0.4 × 10^{10} $L_\odot$ and radii 34.0 and 17.6 kpc.

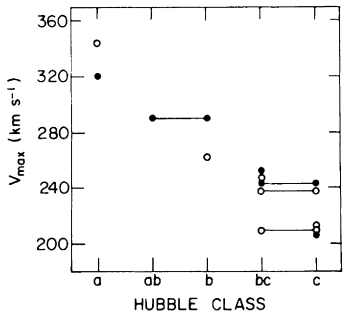


FIG. 5.—Maximum rotational velocity, in the plane of the galaxy, as a function of Hubble type. Galaxies with two classifications are entered twice with a connecting line. Open circles denote values of V_{max} which come from the last measured velocity (i.e., rising rotation curve).

spherical galaxies, which give upper limits to the masses. Both models are parameter independent. For the thin disk approximation (Kuz'min 1952), increased by 1.1 (Brandt 1960), integral masses as a function of radius (mass within a disk of radius r) are shown in Figure 6. The linear increase of mass with r is a consequence of the flat rotation curve: $M(r) = (2/G\pi)$ $V^2(r)r$ for disk models, and V^2 is approximately constant. Because total mass varies both with r and V^2, and V is correlated with HT (Figure 5), total mass is a function of HT and of radius. While early-type galaxies have steeper mass gradients, dM/dr, large late-type galaxies (NGC 801, Figure 6) will have larger total masses than small early-type (NGC 4378). The *surface mass density decreases systematically along the Hubble sequence.*

The mass calculated to the last observed velocity, and the mass extrapolated to the RC2 radius, are listed in columns (15) and (16) of Table 1. If we had chosen to determine the masses from disk-plus-bulge models, then these masses would have been increased by about 1.2 (Sa) and 1.1 (Sab) times over the disk-molded masses. These refinements are included in column (16) and are discussed in Papers II and III. Masses to the RC2 radii lie in the range $1.2 < M(r_{dev}) < 4.5 \times 10^{11} M_\odot$; random errors should not exceed 25%. The expression $M(r_{dev}) = 1.5 \times 10^5 V^2 r$ (V in km s^{-1}, r in kpc) reproduces all 11 masses to within 20%, and seven of them to within 10%. Total masses calculated from integrated 21 cm profiles (Roberts 1975) are in good agreement with those calculated from the detailed optical rotation curves, but are about 20% smaller, due to the falling rotation curves adopted in the 21 cm models (Roberts 1969). If the exponent in the adopted Brandt (1960) curve is changed from 3 to 2, the 21 cm masses will agree with the disk-modeled masses determined from optical rotation curves.

For galaxies modeled as spheres, $M(r) = G^{-1}V^2r$, or $\frac{1}{2}\pi$ (1/1.1) times that used above. The integrated mass out to any r is about 40% larger than the mass determined from the same rotation curve for a disk model. For spherical models, and flat rotation curves, $M_{sph} = 2.1 \times 10^5 V^2 r$ is a good approximation to the mass distribution. For our sample, spherical models give $2 < M(r_{dev}) < 7 \times 10^{11} M_\odot$; out to their Holmberg radii (26.5 mag arcsec^{-2}), masses for some must approach $10^{12} M_\odot$. However, both disk and spherical mass distributions can reproduce the observed rotation curves.

By choosing to observe galaxies of high inclinations, we have maximized the accuracy of the rotational velocities and masses, but the uncertain internal extinction corrections pose a problem for the total magnitudes. The calculated blue luminosities (following RC2) and mass-to-blue-luminosity ratios are tabulated in columns (21) and (22). To the limits of the RC2 radii, integrated mass-to-luminosity ratios, M/L_B, are consistently small for this sample of high-luminosity spirals, near 3.5 for disk mass models, and about 5 for spherical mass models. Uncertainties in M/L_B are of the order of 1 or 2 in solar units, excluding systematic effects due to the choice of models and corrections applied.

There is a weak suggestion that M/L is higher for the earlier HT (col. [22]) a result in conflict with the widely quoted result (Roberts 1969) that M/L ratios are constant across the Hubble spiral sequence. However, five of the eight early-type galaxies in Robert's study had no 21 cm detection, so total masses come from limited optical measures, with extrapolation based on falling rotation curves. Hence, the masses for these early-type galaxies are too low by factors of from 2.5 to 5. Thus the earlier conclusions is probably not valid.

Previous correlations (Holmberg 1964) between high M/L and early HT have been questioned because of the procedure for obtaining masses (Sandage, Freeman, and

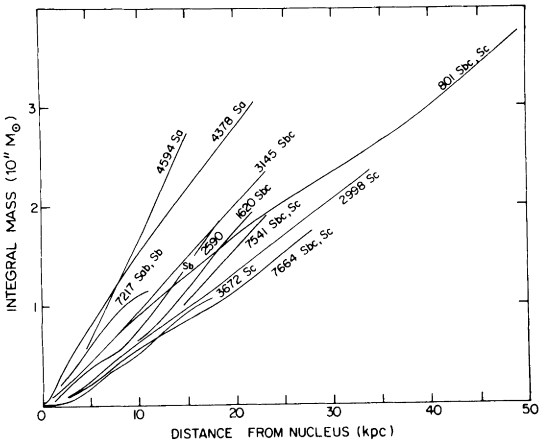


FIG. 6.—Integral mass within disk of radius r, as a function of r, for 11 galaxies, Sa through Sc, out to last measured velocity. Scale gives mass for disk models; masses for spherical galaxies are 1.4 times larger. Linear increase of mass with radius is a consequence of flat rotation curves. Mass gradient, dM/dr, is greater for early types, but total mass is a function both of V^2 and r. Surface mass density decreases systematically with Hubble type.

Stokes 1970). A correlation of M/L with HT might be expected from flat rotation curves ($M \propto V^2 r$), at least for galaxies of approximately equal radii, because of the relation between HT and V, and the tight correlation between luminosity and radius ($L \propto r^{2.4}$, Holmberg 1975). We prefer to keep an open mind on the question of whether M/L varies across the Hubble spiral sequence. At least some early-type galaxies have high M/L ratios.

IV. 21 CENTIMETER PROPERTIES

The 21 cm integrated flux densities, the hydrogen masses, and the hydrogen-mass-to-total-mass ratios are listed in columns (17), (18), and (19). H I constitutes 1% of the mass for our single Sa and Sab galaxies, but about 20% of the mass for the Sb and Sc galaxies.

From the 21 cm profile width, ΔV, we determine the peak velocity of the rotation curve, $V_{max} = \Delta V/(2 \sin i)$ (col. [14]). This value is in good agreement with V_{max} (optical), found from the (resolved) optical rotation curve. The mean difference, $\langle V_{max}$ (opt) $- V_{max}$ (21 cm)$\rangle$, is 1 ± 6 km s^{-1}. A major surprise is the lack of a correlation of V_{max} (optical or 21 cm) with luminosity or with radius, i.e., a Tully-Fisher (1977) relation. The pairs of rotation curves in Figure 5 show that we do not observe such a correlation. Galaxies of the same HT and same V_{max} have radii which differ by factors of 2, and luminosities which differ by factors of 3.

V. CONCLUSIONS

The major result of this work is the observation that rotation curves of high-luminosity spiral galaxies are flat, at nuclear distances as great as $r = 50$ kpc. Roberts and his collaborators (Roberts 1976) deserve credit for first calling attention to flat rotation curves. Recent 21 cm observations by Krumm and Salpeter (1976, 1977) have strengthened this conclusion. These results take on added importance in conjunction with the suggestion of Einasto, Kaasik, and Saar (1974), and Ostriker, Peebles, and Yahil (1974) that galaxies contain massive halos extending to large r. Such models imply that the galaxy mass increases significantly with increasing r which in turn requires that rotational velocities remain high for large r. The observations presented here are thus a necessary but not sufficient condition for massive halos. As shown above, mass distributions from disk models or spherical models adequately reproduce the observed velocities. The choice between spherical and disk models is not constrained by these observations.

We thank the directors of Kitt Peak National Observatory and Cerro Tololo Inter-American Observatory for observing time, Drs. J. Ostriker, M. S. Roberts, R. J. Rubin, and F. Schweizer for valuable comments, and A. Sandage for making available his galaxy classifications before publication. V. R. thanks the staff of CTIO for their hospitality during her stay as an invited Visiting Astronomer, where these results were compiled.

REFERENCES

Brandt, J. C. 1960, *Ap. J.*, **131**, 293.
Brosche, P. 1971, *Astr. Ap.*, **13**, 293.
Burbidge, E. M., and Burbidge, G. R. 1975, in *Galaxies and the Universe*, ed. A. and M. Sandage and J. Kristian (Chicago: University of Chicago Press).
de Vaucouleurs, G., de Vaucouleurs, A., and Corwin, H. G. 1976, *Second Reference Catalogue of Bright Galaxies* (Austin: University of Texas).
Einasto, J., Kaasik, A., and Saar, E. 1974, *Nature*, **250**, 309.
Holmberg, E. 1964, *Uppsala Astr. Obs. Medd*, No. 148.
———. 1975, in *Galaxies and the Universe*, ed. A. and M. Sandage and J. Kristian (Chicago: University of Chicago Press).
Hubble, E. 1926, *Ap. J.*, **64**, 321.
Krumm, N., and Salpeter, E. E. 1976, *Ap. J.* (*Letters*), **208**, L7.
———. 1977, *Astr. Ap.*, **56**, 465.
Kuz'min, G. G. 1952, *Pub. Tartu Astr. Obs.*, **32**, 211.
Nilson, P. 1973, *Uppsala Obs. Ann.*, Vol. **6**.
Ostriker, J. P., Peebles, P. J. E., and Yahil, A. 1974, *Ap. J.* (*Letters*), **193**, L1.
Peterson, C. J., Rubin, V. C., Ford, W. K., Jr., and Roberts, M. S. 1978, *Ap. J.*, Vol. **226**, in press (Paper III).
Roberts, M. S. 1969, *A.J.*, **74**, 859.
———. 1975, in *Galaxies and the Universe*, ed. A. and M. Sandage and J. Kristian (Chicago: University of Chicago Press).
———. 1976, *Comments Ap.*, **6**, 105.
Rubin, V. C., Ford, W. K., Jr., Strom, K. S., Strom, S. E., and Romanishin, W. 1978, *Ap. J.*, **224**, 782 (Paper II).
Rubin, V. C., Thonnard, N., and Ford, W. K., Jr. 1977, *Ap. J.* (*Letters*), **217**, L1 (Paper I).
Sandage, A. 1978, in preparation.
Sandage, A., Freeman, K. C., and Stokes, N. R. 1970, *Ap. J.*, **160**, 831.
Schweizer, F. 1978, *Ap. J.*, **220**, 98.
Shostak, S. 1978, *Astr. Ap.*, submitted.
Thonnard, N., Rubin, V. C., Ford, W. K., Jr., and Roberts, M. S. 1978, *A.J.*, submitted.
Tully, R. B., and Fisher, J. R. 1977, *Astr. Ap.*, **54**, 661.
Zwicky, F., Herzog, E., Wild, P., Karpowicz, M., and Kowal, C. T. 1961–1968, *Catalogue of Galaxies and Clusters of Galaxies* (Pasadena: California Institute of Technology), Vols. 1–6.

W. KENT FORD, JR., VERA C. RUBIN, and NORBERT THONNARD: Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road, N.W., Washington, DC 20015

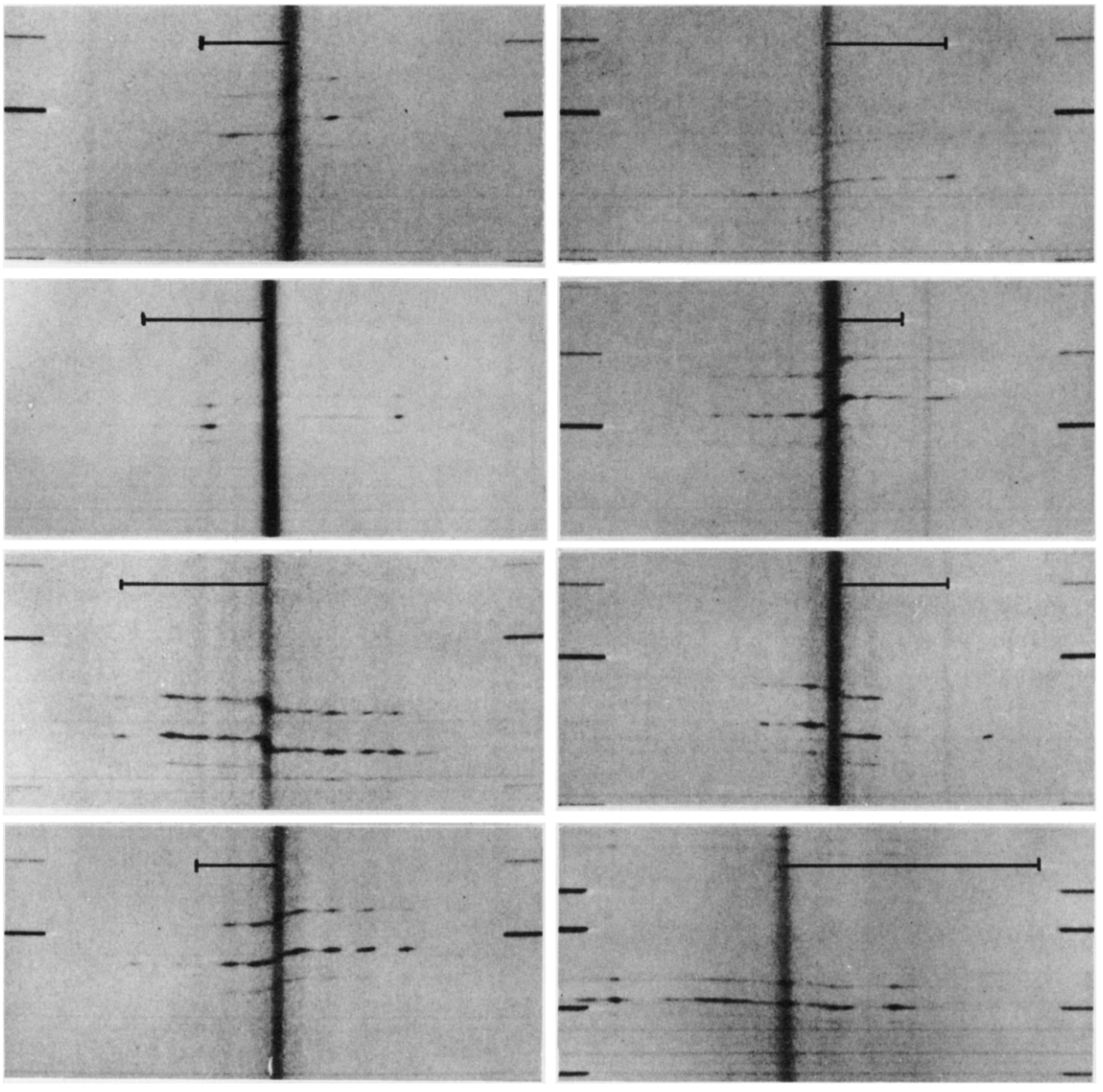


FIG. 1.—Hα region major axis spectra for galaxies of different Hubble types, taken with the 4 m RC spectrograph plus Carnegie image tube plus preflashed IIIa-J plate. Plates are (*) H_2 treated, 26 Å mm^{-1}, KPNO; or (†) N_2 baked, 52 Å mm^{-1}, CTIO. For all spectra, scale perpendicular to the dispersion is 24″ mm^{-1}, and transfer optics are f/2. (*a*)*NGC 2590, Sb, exposure 120 minutes. (*b*)*NGC 1620, Sbc, exposure 129 minutes. (*c*)†NGC 3145, Sbc I, exposure 90 minutes. (*d*)*NGC 801, Sbc–Sc, exposure 150 minutes. (*e*)*NGC 7541, Sbc–Sc, exposure 114 minutes. (*f*)*NGC 7664, Sbc–Sc, exposure 119 minutes. (*g*)*NGC 2998, Sc I, exposure 200 minutes. (*h*)†NGC 3672, Sc I–II, exposure 120 minutes. On each spectrum, Hα is strongest emission line; [N II] λ6583 is at longer λ (*up* on print). Vertical stripe is continuum from stars in nucleus. Solid horizontal line on each spectrum indicates 20 kpc in plane of galaxy. Linear extent of spectra varies from a radius of $r = 17.4$ kpc (NGC 2590) to $r = 49$ kpc (NGC 801). Note that velocity is often not constant across emission regions (spiral features) but is lower at inner edge and higher at outer edge, especially apparent in NGC 2998. The letters (*a*) and (*b*) refer to the upper left and upper right, respectively.

RUBIN *et al.* (*see* page L107)

CAN GALACTIC HALOS BE MADE OF BARYONS?

Dennis J. HEGYI
Department of Physics, University of Michigan, Ann Arbor, MI, USA

and

Keith A. OLIVE
CERN, Geneva, Switzerland

Received 29 March 1983

Several arguments are presented indicating that the apparently non-luminous matter forming massive halos of spiral galaxies is not baryonic. There are difficulties with a halo dominated by gas, snowballs, dust and rocks, jupiters, low mass stars, dead stars and neutron stars. Also, halos may not be composed of black holes unless they are either extremely efficiently accreting or primordial. Consequently, it appears that a significant fraction of the universe may be in the form of massive neutrinos, gravitinos, monopoles, etc.

There appears to be no escape from the conclusion that spiral galaxies are surrounded by a large amount of sub-luminous or non-luminous mass [1]; the conclusion is based on an application of simple newtonian mechanics to a system in equilibrium. The central issue at the present time concerns the nature of the halo mass. We have assembled several arguments pointing out the problems associated with baryonic halos thus suggesting that halos may not be made of baryons. These problems force one to seriously consider the possibility that halos which comprise a significant fraction of the mass of the universe, are either composed of massive neutrinos, photinos, gravitinos, etc., or of very efficiently accreting or primordial black holes.

The strongest evidence for the existence of massive galactic halos is based on the measured velocity of matter in stable circular orbit in the disks of spiral galaxies. The corresponding centripetal acceleration, calculated from the measured velocities, is larger than can be understood from the gravitational field arising from the visible galactic matter; additional matter, a massive halo, is postulated to explain the observed centripetal acceleration. Measurements [2,3] on more than 50 spiral galaxies for which symmetric rotation curves have been obtained support the equilibrium condition,

$$M_r = (K/G)\, v^2 r, \qquad (1)$$

where K is a constant of order unity, G is the gravitational constant, v is the observed rotational velocity, at galactic radius r, and M_r is the mass contained within radius r. One would expect that if the bulk of the mass were contained in the luminous central regions, the rotational velocity would fall as $v \propto r^{-1/2}$. Instead, one finds that v is independent of r indicating that $M_r \propto r$, or $\mathrm{d}M/\mathrm{d}r = \text{const}$. Thus, there is an equal amount of mass per increment of galactic radius unlike the galaxy luminosity distribution which is centrally concentrated.

Other evidence supporting massive halos, though not as compelling as the galaxy rotation curves, includes binary galaxies [4,5] and the stability of spiral arms [6]. Investigations [7,8] pursuing independent directions have argued that the unseen mass is distributed spherically symmetrically.

Let us now explore the possibility that massive galactic halos are composed of baryons by considering a halo made of gas. Since we will require the halo to be both stable and static we assume that it must be in hydrostatic equilibrium, i.e., the halo is maintained only if it is at a temperature T_{eq}. For temperatures less than T_{eq}, the halo would collapse in a gravitational timescale $\tau_c = (3\pi/32G\rho)^{1/2} \approx 5 \times 10^8$ yr, much less than the age of the galaxy. To determine T_{eq}, we use

the following

$$dP_r/dr = -GM_r\rho_r/r^2 \quad (2)$$

where P_r and ρ_r are the pressure and density, respectively, at radius r and

$$P_r = (2\rho_r/m_p)kT \quad (3)$$

is the equation of state. Solving eqs. (2) and (3) for T_{eq} with $M_r \propto r$ and $\rho_r \propto r^{-2}$ we have,

$$T_{eq} = (Gm_p/2k)(M_r/r) \approx 1.9 \times 10^6 \text{ K} \quad (4)$$

for $M_r \approx 10^{12}\, M_\odot$ at $r \approx 100$ kpc.

A hot gas halo at temperature T_{eq}, however, would be emitting a thermal X-ray background which is in excess of the limits of the observed X-ray background. To see this, we compare the quantity $\delta C\Omega^2$, which is a measure of the X-ray emissivity of the halo matter evaluated at T_{eq}, with the observed upper limit derived from the X-ray background. Here $\delta \approx h_0^3$ where $h_0 = H_0/(100 \text{ km s}^{-1}\text{Mpc}^{-1})$, C is a clumpiness parameter, and $\Omega = \rho/\rho_c$. The quantity, ρ, is the average density of matter in the universe, ρ_H is the average halo density, 1.7×10^{-26} g cm^{-3} corresponding to our standard halo containing $10^{12}\, M_\odot$ in 100 Kpc, and $\rho_c \equiv 3H_0^2/8\pi G = 1.88 \times 10^{-29} h_0^2$ g cm^{-3}. To evaluate $\delta C\Omega^2$ for halos we use [1] $\Omega \geqslant \Omega_H = [M/L]_{spirals}/[M/L]_{critical} \geqslant 35/700 = 0.05$ and $C\Omega = \rho_H/\rho_c$ which follows directly from the definition, $C \equiv \langle\rho^2\rangle/\langle\rho\rangle^2$. With $h_0 = \frac{1}{2}$ to obtain a lower limit on the halo X-ray emission, we find that for halos $\delta C\Omega^2 \geqslant 22$. Silk [9], based on observations of the X-ray background, finds that at $T = 2 \times 10^6$ K, $\delta C\Omega^2 \lesssim 1$. A halo of hot gas violates the upper limits by a factor of at least 20.

Another possibility for the halo matter is low mass stars [10] or jupiters, that is, non-nuclear burning objects with mass $< 0.08\, M_\odot$. We shall show that if one uses a single power law for the distribution function describing the number of stars and jupiters of a given mass as a function of mass, where the index in the power law is chosen to match the observed stellar (nuclear burning) distribution function, such a halo will radiate much more light than is observed.

To draw this conclusion it is necessary to use an upper limit on the index, x, in the stellar mass distribution function which is defined by

$$\phi(m) = Am^{-(1+x)}, \quad (5)$$

where $\phi(m)$ is the number of stars per unit volume per unit mass. Constraints on x which have been obtained from observation of narrow band spectral features [11] and stellar data in the solar neighborhood [12] for stars of mass $< 0.8\, M_\odot$ require $x \leqslant 1$ at the 2σ level. Other data in the infrared [13,14], though intrinsically less sensitive as a discriminant, are satisfied by the weaker constraint, $x \leqslant 1.35$.

Our conclusion is also sensitive to a second quantity, the smallest mass object which can gravitationally collapse, m_{min}, also known as the Jeans' mass. Arguments on the fragmentation of gas clouds [15–17] place a lower limit on the Jeans' mass of $0.007\, M_\odot$ for a collapsing cloud consisting of hydrogen and helium. Because there are complications handling the radiative transfer in this mass range, we shall adopt $m_{min} \geqslant 0.004\, M_\odot$ since this limit was determined using an optically thin cloud which avoids these complications and should be a lower limit on m_{min} [18]. In fact, others argue [19] that the first stars to form may have masses $\gtrsim 1500\, M_\odot$.

We shall summarize the calculation of the surface brightness of the halo here and shall present the details elsewhere. Using the data of Tinsley and Gunn [11] for both dwarfs in the mass range $0.08\, M_\odot$ to $0.8\, M_\odot$ and for giants with $m_{giant} = 0.8\, M_\odot$ we have calculated the ratio of mass-to-luminosity in the I and K Johnson spectral bands for the halo of NGC 4565. The rate of evolution of giants from the main sequence has been corrected for low metal abundance ($Z = 10^{-5}$) based on model calculations [20]. The calculated surface brightness must satisfy the observational data in the I and K band. The 2σ lower limits on the mass-to-light ratio, M/L, for the halo of NGC 4565 expressed in solar units corresponding to the I data [21] and the K data [22] are,

$$M/L_I > 60\, M_\odot/L_{\odot,I} \quad (6a)$$

$$M/L_K > 38\, M_\odot/L_{\odot,K}. \quad (6b)$$

If we choose $m_{min} = 0.004\, M_\odot$ and determine the power law index, x, necessary to satisfy the observational data, we find $x > 1.6$ and $x > 1.7$ for the I and K bands, respectively.

Alternatively, rather than solve for a lower limit on x, we can choose x and try to solve for m_{min}. For $x < 1$, there is no solution for m_{min} consistent with the observations; it is not possible to put enough mass in the halo to satisfy the rotation curve and the surface

brightness constraints. If we relax the constraint on x and use the Salpeter value, $x = 1.35$, we find that x_{min} must be less than $2 \times 10^{-4}\ M_\odot$, more than a factor of 20 below the calculated lower limit on m_{min}. On this basis we find a halo filled with either low mass stars or jupiters inconsistent with present observations.

A recent observation [23] suggests that we are being much more conservative than may be necessary. It appears that the initial mass function turns over for $m \gtrsim 0.1\ M_\odot$. This result was based on stars with a metal abundance similar to that of the sun. For stars without metals, m_{min} is expected to be larger since metals allow a collapsing star to cool and fragment more effectively.

Next, we consider a halo composed of stars with an initial mass greater than $2\ M_\odot$ which are now either white dwarfs or neutron stars. The maximum mass of white dwarfs is $\sim 1.4\ M_\odot$ [24], which coincidentally, appears to be about equal to the mass of neutron stars. Because the two most precisely determined neutron star masses are the same to within the 1% measuring uncertainty [25], and because all neutron star mass determinations are consistent with $1.4\ M_\odot$, we shall take this value for the mass of all neutron stars. Since stars with masses larger than $2\ M_\odot$ must evolve to a remnant of only $1.4\ M_\odot$, at least 40 percent of the present halo mass contained in these remnants must have been ejected. The problem with massive stars is to find a suitable place to hide their ejecta. The ejected mass cannot be in hot gas because of previous arguments and it cannot be in cool gas or it will quickly collapse into the disk. Depending on the exact distribution of stellar masses, there would be an additional problem. A significant amount [26] ($\gtrsim 10\%$) of the ejected mass would be in the form of helium and metals.

One might expect an interstellar medium with metal in the halo to lead to a second generation of stars with a mass function similar to the solar neighborhood yielding a halo which is too bright. A second problem involves the contamination due to the ejected metals of the oldest stars in the galaxy. Observations show there exist old low mass stars in the nuclear bulge [27] with a metal abundance of $Z \sim 10^{-5}$. The metal abundance of these stars should be characteristic of the interstellar medium at the epoch when the stars formed. Unless the nuclear bulge formed substantially later than the massive halo (in which case the above arguments still apply) it is difficult to understand how the oldest stars would not be contaminated at a level which is larger than $Z \sim 10^{-5}$.

A similar argument applies to a halo composed primarily of metals, including rocks, dust or grains, etc. If the halo is composed primarily of metals then $\Omega_{metals} \geqslant 0.05$. Since the mass in the disk and nuclear bulge of the galaxy corresponds to an $\Omega \sim 10^{-3}$, then it would be impossible to find stars with a metal abundance as small as $Z \sim 10^{-5}$ if more than one part in 5×10^6 of the halo material contaminated the first generation of stars. The above arguments for the evolution of stars with initial masses greater than $2\ M_\odot$ also rule out a halo of neutron stars for the same reasons.

We now consider a halo composed of snowballs, objects consisting primarily of condensed hydrogen which are held together by electrostatic forces. Snowballs, consequently, are not ruled out by arguments based on the Jeans' mass which only involve gravitational binding. We shall show, nevertheless, that though we cannot envision any way in which snowballs might form, they would rapidly evaporate into a gaseous state, if they did form.

For our standard halo, $10^{12}\ M_\odot$ in 100 Kpc radius, snowballs of density equal to that of solid hydrogen, $\rho_s \sim 7 \times 10^{-2}$ g cm^{-3}, must have a radius greater than 1 cm to avoid all collisions during their lifetime, 10^{10} yr. At typical halo velocities $\sim$250 km s^{-1} or 500 eV per hydrogen atom, kinetic energies are so much larger than typical binding energies, $\sim$1 eV, that a single collision would destroy a snowball. Consequently, under such conditions, they must form without dissipation at a density about equal to the present halo density, $\rho_H \sim 1.7 \times 10^{-26}$ g cm^{-3} in a timescale $\tau_c \sim (3\pi/32 G\rho_H)^{1/2}$.

The snowballs will be most stable against evaporation in a universe with the lowest temperature cosmic background radiation. Assuming that $\Omega \leqslant 2$, this will occur with $\Omega = 2$ and $h_0 = 1$. Because of the lack of dissipation during the formation of snowballs, they must form when the universe has an age equal to $2\tau_c$. Under these most favorable conditions the temperature of the cosmic background radiation is $\sim$7.6 K. At this temperature, in equilibrium, the vapor pressure of solid hydrogen is 0.045 mm [28]. This is sufficiently high so that there is no equilibrium between the gaseous and solid phases of hydrogen; all the hydrogen is in the gaseous state. It may be shown using standard two phase thermodynamic equilibrium methods [29] that, in fact,

the halo density would have to increase by a factor $\sim 10^{19}$ before any solid hydrogen could remain in equilibrium with the gas.

Having established the ultimate equilibrium state, we consider the rate at which evaporation takes place. The time for evaporation of an H_2 molecule, t_{ev}, should be [30]

$$t_{ev} \sim \nu_0^{-1} \exp(b/kT).$$

where $\nu_0 \sim 10^{12}\ s^{-1}$ is the characteristic vibration frequency of a solid hydrogen lattice and b is the binding energy of a hydrogen molecule to the surface of the solid. It is possible to estimate b by using the principle of corresponding states [31]. For a Lennard Jones potential the depth of the well, which we shall use as an approximate value of b, is, $b \sim 0.77\ T_c k$. T_c is the critical temperature, equal to 33.3 K for H_2 and k is Boltzmann's constant. With these numbers, $t_{ev} \sim 10^{-11}$ s, so we may conclude that evaporation proceeds rapidly. It is also possible to show that snowballs would evaporate at 2.7 K.

The final possibility is that the halo is composed of black holes. While it is not possible to rule them out, particularly massive black holes with masses larger than $10^2\ M_\odot$, some general statements can be made. Consistent with our argument, however, we point out though black holes are certainly of interest here, they have no well defined baryon number.

The initial mass function which evolved into a halo of black holes cannot have had many low mass stars ($< 1\ M_\odot$); otherwise, the stars would be observable today. One might expect a halo composed of black holes of mass $\sim 10\ M_\odot$ to have evolved from similar mass stars. But since the evolution of stars in this mass range appears to reliably lead to mass loss, black holes in this mass range are not favored. However, any black hole which can very efficiently accrete its ejecta remains a possibility. While this may not appear likely for low mass black holes, no such statement can be made about supermassive black holes because so little is known about them. Also, it is not possible to rule out primordial black holes.

In conclusion, we have argued that there may be significant amounts of stable non-baryonic matter on galactic scales. In addition, it does not appear likely that the halo is composed of black holes, unless they are very efficiently accreting or primordial. Consequently, other constituents for the halo should be seriously considered. Popular suggestions include massive neutrinos [32,33] and, from supersymmetric theories, massive gravitinos [34] or photinos. Depending on their type and velocity, magnetic monopoles could provide the halo mass; however, because of the constraints on monopoles [35–39], they appear to be less likely.

We would like to thank J. Ellis, G.W. Ford, F. Palla and D. Schramm for helpful discussions.

References

[1] S.M. Faber and J.S. Gallagher, Ann. Rev. Astron. Astrophys. 17 (1979) 135.
[2] N. Krumm and E.E. Salpeter, Astronom. J. 84 (1979) 1138.
[3] V.C. Rubin, W.K. Ford Jr. and N. Thonnard, Astrophys. J. 225 (1978) L107.
[4] E.L. Turner, Astrophys. J. 208 (1976) 304.
[5] S.D. Peterson, Astrophys. J. 232 (1979) 20.
[6] J.P. Ostriker and P.J.E. Peebles, Astrophys. J. 186 (1973) 467.
[7] E.M. Saar, in: IAU Symposium 84, The large-scale characteristics of the galaxy, ed. W.B. Burton (1978).
[8] D.G. Monet, D.O. Richstone and P.L. Schechter, Astrophys. J. 245 (1981) 454.
[9] J. Silk, Ann. Rev. Astron. Astrophys. 11 (1973) 269.
[10] A. Dekel and J. Shaham, Astron. Astrophys. 74 (1979) 186.
[11] B.M. Tinsley and J.E. Gunn, Astrophys. J. 203 (1976) 52.
[12] G.E. Miller and J.M. Scalo, Astrophys. J. Suppl. 41 (1979) 513.
[13] M. Aaronson, J.G. Cohen, J. Mould and M. Malkan, Astrophys. J. 223 (1978) 824.
[14] J.A. Frogel, S.E. Persson and J.G. Cohen, Astrophys. J. 240 (1980) 785.
[15] C. Low and D. Lynden-Bell, Mon. Not. R. Astron, Soc. 176 (1976) 367.
[16] M.J. Rees, Mon. Not. R. Astron. Soc. 176 (1976) 483.
[17] J. Silk, Astrophys. J. 256 (1982) 514.
[18] F. Palla, E.E. Salpeter and S.W. Stahler, preprint.
[19] J.F. Tohline, Astrophys. J. 239 (1980) 417.
[20] A.V. Sweigart and P.G. Gross, Astrophys. J. Suppl. 36 (1978) 405.
[21] D.J. Hegyi, in: Proc. Moriond Astrophysics Meeting, eds. J. Audouze et al. (Frontieres, Oreux, 1981).
[22] S.P. Boughn, P.R. Saulson and M. Seldner, Astrophys. J. 250 (1981) L15.
[23] R.G. Probst and R.W. O'Connell, Astrophys. J. 252 (1982) L69.
[24] S. Chandrasekhar, Mon. Not. R. Astron. Soc. 95 (1935) 207.
[25] J.H. Taylor and J.M. Weisberg, Astrophys. J. 253 (1982) 908.

[26] W.D. Arnett, Astrophys. J. 219 (1978) 1008.
[27] H.E. Bond, Astrophys. J. 248 (1981) 606.
[28] V.J. Johnson, A. compendium of the properties of materials at low temperature (Phase I) U.S. Air Force (1960).
[29] See, e.g. F.W. Sears, An introduction to thermodynamics, the kinetic theory of gases and statistical mechanics (Addison-Wesley, Reading, MA, 1950) Ch. 6.
[30] D. Hollenbach and E.E. Salpeter, Astrophys. J. 163 (1971) 155.
[31] J.O. Hirschfelder, C.F. Curtiss and R.B. Bird, Molecular theory of gases and liquids (Wiley, New York, 1954) Ch. 4.
[32] A.L. Melott, Nature 296 (1982) 721.
[33] J.R. Bond and A.S. Szalay, Proc. Neutrino 81 (Hawaii, 1981) p. 59.
[34] H. Pagels and J.R. Primack, Phys. Rev. Lett. 48 (1982) 223.
[35] E.N. Parker, Astrophys. J. 160 (1960) 383.
[36] M.S. Turner, E.N. Parker and J.T. Bogban, Phys. Rev. D. 26 (1982) 1296.
[37] E.W. Kolb, S.A. Colgate and J.A. Harvey, Phys. Rev. Lett. 49 (1982) 1373.
[38] J. Ellis, D.V. Nanopoulos and K.A. Olive, Phys. Lett. 116B (1982) 127.
[39] F.A. Bais, J. Ellis, D.V. Nanopoulos and K.A. Olive, CERN preprint TH-3383.

THE ASTROPHYSICAL JOURNAL, 281:493–511, 1984 June 15
© 1984. The American Astronomical Society. All rights reserved. Printed in U.S.A.

PRIMORDIAL NUCLEOSYNTHESIS: A CRITICAL COMPARISON OF THEORY AND OBSERVATION

J. YANG,[1,2] M. S. TURNER,[2,3] G. STEIGMAN,[4] D. N. SCHRAMM,[2,3] AND K. A. OLIVE[3]
Received 1983 August 25; accepted 1983 December 20

ABSTRACT

Primordial nucleosynthesis is reexamined in the context of a detailed comparison of theory and observation. A new argument is presented to show how the observed abundances of D and ^{3}He can be used to derive a *lower* bound to the nucleon density. In concert with the previously known upper bound from D alone, we define a conservative ("safe bet") range for the nucleon-to-photon ratio: $\eta = (3\text{–}10) \times 10^{-10}$. New observations of ^{7}Li are consistent with the abundances of D and ^{3}He and help us to define a reasonable ("best bet") range: $\eta = (4\text{–}7) \times 10^{-10}$. In either of these ranges the predicted and observed abundances of D, ^{3}He, and ^{7}Li are all in concordance. The upper bounds correspond to $\Omega_N \leq 0.14\text{–}0.19$, and we conclude that nucleons fail to close the universe by at least a factor of 5–7. We review the recent observational data on ^{4}He and conclude that there is complete consistency between the predicted abundance of ^{4}He and those of the other light elements. In particular, for the standard model ($N_\nu = 3$, $10.4 \leq \tau_{1/2} \leq 10.8$ minutes) we find that $Y_p \leq 0.25$ for $\eta \lesssim 7 \times 10^{-10}$. We predict a lower bound to the ^{4}He mass fraction of $Y_p \geq 0.24$ if $N_\nu \geq 3$ and $\tau_{1/2} \geq 10.4$ minutes. If ν_τ is not light (i.e., $m \gtrsim 1$ MeV), $N_\nu \geq 2$, and Y_p can be as low as 0.226; if Y_p is unambiguously determined to be less than this value, the standard model will be in trouble. Only one additional light, two-component neutrino species ($N_\nu = 4$) is marginally permitted: for $N_\nu = 4$, $\tau_{1/2} \geq 10.4$ minutes, and $\eta \geq 3 \times 10^{-10}$, $Y_p \geq 0.253$.

Subject headings: cosmology — early universe — nucleosynthesis

I. INTRODUCTION

The almost universal acceptance of the standard (i.e., simplest) hot big-bang model (Friedmann-Robertson-Walker cosmology) rests, in large part, on the success of this model in accounting for the abundances of the light elements, particularly helium-4 and deuterium. The pioneering calculations of Peebles (1966) and Wagoner, Fowler, and Hoyle (1967) demonstrated that an epoch of nucleosynthesis, occurring within the first few minutes of the evolution of the universe, would produce a large amount of ^{4}He ($Y \approx 0.25$; $Y = {}^4$He mass fraction) as well as an abundance of D now known to be comparable to that observed in our Galaxy today (D/H $\approx$ few $\times 10^{-5}$). To date, there are no other astrophysical scenarios for producing such large amounts of both ^{4}He and D. In addition, it was noticed that astrophysically interesting amounts of ^{3}He and ^{7}Li were also produced during primordial nucleosynthesis. As a result of the general agreement between the predictions of the standard model and the observed abundances of the light elements (for a recent comparison see Olive *et al.* 1981, hereafter OSSTY), big-bang nucleosynthesis has been exploited as a probe of the early universe. As such it has placed useful and important constraints on both particle physics and cosmology. For example, primordial nucleosynthesis has been used to infer an upper limit to the energy density at early epochs (Shvartsman 1969), which leads to constraints on the number of light ($m \lesssim 1$ MeV) neutrino (or, other light) species (Steigman, Schramm, and Gunn 1977; Yang *et al.* 1979, hereafter YSSR; OSSTY; Szalay 1981), on the number of intermediate-mass ($m \approx 0.1\text{–}10$ MeV) neutrino species (Kolb and Scherrer 1982), and on the number of superweakly interacting particles (Steigman, Olive, and Schramm 1979; Olive, Schramm, and Steigman 1981). From the perspective of cosmology, primordial nucleosynthesis serves to constrain the universal density of nucleons (Wagoner, Fowler, and Hoyle 1967; Reeves *et al.* 1973; Gott *et al.* 1974; YYSR; OSSTY) and to limit the amplitude of density inhomogeneities and anisotropy during the early universe (Wagoner 1973; Gisler, Harrison, and Rees 1974; Epstein and Petrosian 1975; Barrow 1976; Matzner and Rothman 1982). For a review of this earlier work, see Schramm and Wagoner (1977).

To provide the proper context for the new results to be presented in this article, we shall briefly review the conclusions reached in previous work. The abundances of ^{4}He and D are the primary probes of primordial nucleosynthesis. The mass fraction of ^{4}He produced primordially, Y_p, is an increasing function of the nucleon-to-photon ratio $\eta\,(=n_N/n_\gamma)$, the neutron half-life $\tau_{1/2}$, and the number of light species present, parameterized by the number of equivalent neutrino species N_ν (N_ν provides a measure of the "speedup" of the universal expansion rate; for sufficiently fast expansion, Y_p eventually decreases with increasing N_ν [see OSSTY]). For an *upper* limit to Y_p of 0.25 and *lower* limits to $\tau_{1/2}$ (10.4 minutes) and η (2×10^{-10} as inferred from the dynamics of binaries and small groups of galaxies), one obtains an *upper* limit to the number of light neutrino species (OSSTY): $N_\nu \leq 4$. As Kolb and Scherrer (1982) have shown, intermediate-mass neutrinos contribute one unit or slightly more than one unit to N_ν depending on their mass. If, however, one entertains the possibility that something other than nucleons (e.g., massive neutrinos, gravitinos) dominates the mass on scales of binaries and small groups (BSG), then η may be smaller than the lower bound of 2×10^{-10}. From the mass contained within the luminous parts of galaxies (Holmberg radius) a more certain but less restrictive lower bound to the nucleon

[1] Division of Geological and Planetary Sciences, California Institute of Technology.
[2] The University of Chicago.
[3] Fermilab Astrophysics Group, Fermi National Accelerator Laboratory.
[4] Bartol Research Foundation.

abundance may be obtained: $\eta > 0.3 \times 10^{-10}$; in this case there is no constraint to N_ν from primordial nucleosynthesis (OSSTY). One of the main results of this article is to remedy this situation by deriving a new, restrictive lower bound to η from primordial nucleosynthesis alone.

The abundance of D produced primordially decreases very rapidly with increasing η. The requirement that big-bang production account for the minimum observed abundance of deuterium, $(\mathrm{D/H})_o > 1 \times 10^{-5}$, leads to the constraint that $\eta < 10^{-9}$ (YSSR). The fraction of the critical density contributed by nucleons, Ω_N, is related to the nucleon abundance η by

$$\Omega_N = 3.53 \times 10^{-3} h_o^{-2} (T_o/2.7\ \mathrm{K})^3 \eta_{10}\ , \qquad (1)$$

where the present value of the Hubble parameter is $H_0 = 100 h_o$ km s^{-1} Mpc^{-1} ($\frac{1}{2} < h_o < 1$), T_o is the present temperature of the microwave background (2.7 K $< T_o <$ 3.0 K), and $\eta_{10} = 10^{10}\eta$. The upper bound on η_{10} ($\lesssim 10$) from the deutrium abundance leads to the upper limit $\Omega_N \lesssim 0.2$. Thus, nucleons alone fall a factor of at least 5 short of providing the mass density needed if the universe is to be closed ($\Omega_o > 1$).

In the last few years there have been new observations which provide data of value in our attempts to infer the primordial abundances of the light elements. The quantity and quality of the ^{4}He data, in particular, have improved significantly (see references in § IV). Observations of ^{3}He in several galactic H II regions (Wilson, Rood, and Bania 1983) are helping to clarify the issue of galactic versus primordial production. Very recently, Spite and Spite (1982*a*, *b*) have reported observations of ^{7}Li in halo and old disk stars which may provide a measure of the primordial abundance of ^{7}Li. In view of the potential significance of the new data, we have carried out a thorough reexamination of primordial nucleosynthesis with several goals in mind. Our primary aim is to test the consistency of the standard hot big-bang model. That is, do the predicted abundances of the light elements agree with the primordial abundances inferred from current observations? We answer this question affirmatively and then proceed to deal with more specific issues in greater detail. In particular, we use the observed abundances of D and ^{3}He to establish a *lower* bound to the nucleon abundance η (based solely on primordial nucleosynthesis). From D and ^{7}Li we are able to place upper bounds on η. Our very conservative ("safe bet") range for η is $(3\text{–}10) \times 10^{-10}$, while our reasonable ("best bet") range is $(4\text{–}7) \times 10^{-10}$ The restrictive range for η allowed by the results of primordial nucleosynthesis has important consequences for cosmology and particle physics; that there *is* an allowed range establishes the concordance of the standard model. Since the standard model does provide a good fit to the observed abundances, we are able to investigate the effects of deviations from the standard model. The new lower bound on η permits us to reestablish a firm upper limit to the number of light, two-component neutrino species: $N_\nu \leq 4$. We also constrain the amplitude of isothermal density perturbations (fluctuations in η) during the epoch of primordial nucleosynthesis: $\delta\eta/\eta < O(2)$. Finally, keeping in mind the existing uncertainties in the abundances derived from the observational data, we delimit the range of the abundances for which the predictions of big-bang nucleosynthesis are in concordance with the observations.

The outline of this article is as follows. In § II we briefly review the physics of big-bang nucleosynthesis with the purpose of elucidating the dependences of the predicted abundances on the various parameters (η, $\tau_{1/2}$, N_ν). In § III we concentrate on the abundances of D, ^{3}He, and ^{7}Li. Here we derive the "standard" upper bound on η from the lower bound to D/H, and we find a more restrictive upper bound when we also consider ^{7}Li. Most importantly, we use D plus ^{3}He to provide an upper bound to the primordial abundance of D. This leads to the new *lower* bound to η. In § IV we review the recent observational work relating to the primordial abundance of ^{4}He, both the values derived and their uncertainties. We show that the predicted and inferred abundances of ^{4}He agree, and we obtain an upper bound to N_ν. In § V we discuss the consistency of the standard model. The cosmological consequences are outlined in § VI, and in § VII we derive limits to isothermal density perturbations. Our results and conclusions are summarized in § VIII. Uncertainties in the numerical code for calculating the primordial abundances are discussed in Appendix A. In Appendix B we discuss the survival of ^{3}He in the course of galactic evolution.

II. REVIEW OF PRIMORDIAL NUCLEOSYNTHESIS

During the early evolution of the universe all interactions proceed sufficiently rapidly (compared with the universal expansion rate) to ensure that thermal equilibrium is established (even among weakly interacting particles). For this reason (equilibrium) we may begin our discussion of big-bang nucleosynthesis relatively late in the early evolution of the universe. For example, when $t \approx 10^{-2}$ s, the temperature is ~ 10 MeV, and the universe is dominated by a relativistic gas of photons, electron-positron pairs, and light neutrinos. At such relatively late times and low temperatures the cosmology and the physics are well understood. For example, the average interparticle separations exceed ~ 10 fm. This relativistic gas is contaminated by a trace amount of nucleons ($\eta \approx 10^{-10}$ to 10^{-9}); the internucleon separation exceeds $\sim 10^4$ fm. At a temperature $T \approx 10$ MeV the neutral-current weak interactions (for the e-neutrino the charged-current interactions also contribute),

$$e^+ + e^- \leftrightarrow \nu_i + \bar{\nu}_i \quad (i = e, \mu, \tau, \ldots)\ , \qquad (2)$$

are sufficiently rapid to ensure that all light ($m \ll T$) neutrino species are in thermal equilibrium ($T_\nu = T_e = T_\gamma$). The charged-current weak interactions,

$$p + e^- \leftrightarrow n + \nu_e\ , \quad n + e^+ \leftrightarrow p + \bar{\nu}_e\ , \quad n \leftrightarrow p + e^- + \bar{\nu}_e\ , \qquad (3)$$

maintain chemical equilibrium between neutrons and protons, and

$$n/p = \exp(-\Delta m/T)\ . \qquad (4)$$

Although the rate for neutrons and protons to collide and form deuterons is rapid,

$$n + p \leftrightarrow d + \gamma\ , \qquad (5)$$

the equilibrium abundance of deuterons is very small [$n_D/n_N \propto \eta \exp(2.2\ \mathrm{MeV}/T)$] because the background density of photons is enormous. The weak binding of the deuteron results in a bottleneck to nucleosynthesis until the number of photons capable of dissociating the deuteron becomes sufficiently small. Since the number of such photons per nucleon varies as $\eta^{-1} \exp(-2.2\ \mathrm{MeV}/T)$, nucleosynthesis is delayed until $T \approx 0.1$ MeV; for smaller η the bottleneck persists to a lower temperature.

The light neutrinos decouple from the other particles at $T \approx$ few MeV when the weak rates (2) become slow compared with the universal expansion rate ($T \approx 3.5$ MeV for ν_μ, ν_τ, ..., and $T \approx 2$ MeV for ν_e). Subsequently, the "free" gas of neutrinos cools adiabatically as the universe expands. At a slightly lower temperature (which depends on the strength of the charged-current weak interaction [3]), interaction rates (3) become slow compared with the universal expansion rate, and reactions (3) can no longer maintain the neutron-to-proton ratio at its equilibrium value (4). At this point the neutron-to-proton ratio is said to "freeze out" although, in fact, the ratio continues to decrease but at a rate much slower than if equilibrium were maintained. When $T \lesssim m_e/3$, the electron-positron pairs annihilate, heating the photons but *not* the already decoupled neutrinos, so that

$$\left(\frac{T_\nu}{T_\gamma}\right)^3_{T \ll m_e} = \frac{4}{11}\left(\frac{T_\nu}{T_\gamma}\right)^3_{T \gg m_e} = \frac{4}{11} . \quad (6)$$

Finally, for $T \lesssim 0.1$ MeV (with the precise value of T depending logarithmically on η), the density of photodissociating photons is sufficiently small to permit the buildup of deuterium. As soon as the deuterium abundance becomes significant, further n, p, and d reactions occur leading to the synthesis of ^{3}H, ^{3}He, ^{4}He, and a small amount of ^{7}Li.

The gaps at mass-5 and mass-8 (no stable nuclei) provide bottlenecks to the synthesis of heavier elements. Furthermore, as the temperature decreases it is increasingly (i.e., exponentially!) difficult to penetrate the Coulomb barriers in nuclear collisions. As a result, most of the neutrons which were present when nucleosynthesis began in earnest find themselves incorporated into the most tightly bound light nucleus, ^{4}He, with trace abundances of D, ^{3}He, and ^{7}Li being produced.

With the above review as background we can now understand how the predicted abundances depend on the various parameters: η, $\tau_{1/2}$, N_ν. Let us first consider ^{4}He. Since most neutrons are incorporated into ^{4}He, the primordial abundance (by mass) Y_p is essentially determined by the neutron-to-proton ratio at the time nucleosynthesis starts: $Y_p \approx 2n/p(1 + n/p)^{-1}$. The n/p ratio is determined by the competition between the weak interaction rate (which varies as $\tau_{1/2}{}^{-1}$) and the universal expansion rate. For example, for a longer half-life, interactions (3) will freeze out at a higher temperature, when there are more neutrons relative to protons (see eq. [4]), leading to the production of more ^{4}He. In Figure 1, we show the predicted primordial abundance of ^{4}He for three choices of the neutron half-life (Christensen *et al.* 1972; Kugler, Paul, and Trinks 1978, 1979; Wilkinson 1979; Byrne *et al.* 1980; Byrne 1982).

When the universe is radiation dominated, the expansion rate is determined by the effective number of relativistic degrees of freedom:

$$t(\mathrm{s}) = 2.42 g_{\mathrm{eff}}{}^{-1/2} T_{\mathrm{MeV}}{}^{-2} , \quad (7)$$

where

$$g_{\mathrm{eff}} = \sum_B g_B (T_B/T_\gamma)^4 + \tfrac{7}{8} \sum_F g_F (T_F/T_\gamma)^4 . \quad (8)$$

In equation (8), $g_B(g_F)$ are the number of boson (fermion) helicity states, and $T_B(T_F)$ are the temperatures ($\leq T_\gamma$) of the various species. For example, with three two-component neutrino species we have (for $T > m_e$) $g_{\mathrm{eff}} = 2 + \frac{7}{8}(4 + 3 \times 2) = 43/4$. It is clear from equations (7) and (8) that the addition of new, light particles increases g_{eff} and, therefore, leads to a faster expansion. A faster expansion results in an earlier freezing out of the weak interactions, leaving behind more neutrons, which results in more ^{4}He. The effect of $N_\nu = 2, 3, 4$ neutrino species is also shown in Figure 1.

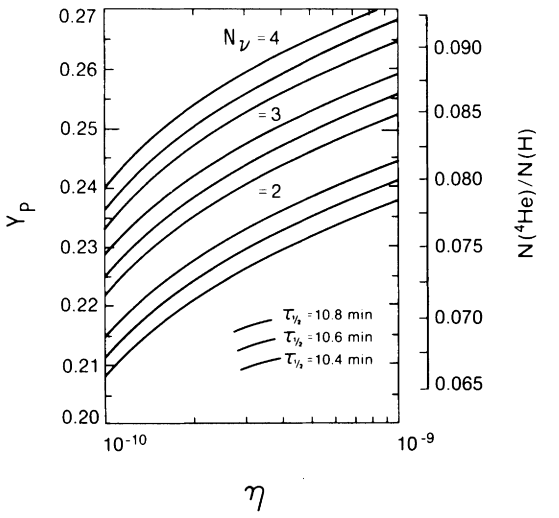


FIG. 1.—The abundance of ^{4}He (by mass and by number) as a function of the nucleon-to-photon ratio (η) for $N_\nu = 2, 3, 4$ species of light, two-component neutrinos and for three choices for the neutron half-life ($\tau_{1/2} = 10.4$, 10.6, 10.8 minutes).

Finally, it should be noted that although the predicted abundance of ^{4}He is rather insensitive to the nucleon abundance η, Y_p does increase slowly with increasing η. Basically, a larger η means that the "deuterium bottleneck" can be overcome earlier, at a higher temperature, when the neutron-to-proton ratio is higher. To summarize (see Fig. 1), the primordial abundance of ^{4}He depends on η, $\tau_{1/2}$, and N_ν, increasing with increasing values of each of these parameters. An upper limit (from observational data) to Y_p, coupled to lower limits to any of the three parameters, will yield an upper limit to the third. We will exploit this in § IV to obtain upper limits to η and to N_ν.

Now let us turn to D and ^{3}He. Since ^{4}He is the most tightly bound of the light nuclei, nuclear reactions tend to burn D and ^{3}He to ^{4}He. The surviving abundances of these elements depend on the competition between the reaction rates (which depend on η) and the expansion rate. The higher the nucleon abundance, the faster D and ^{3}He are destroyed. The predicted primordial abundances (by number relative to hydrogen) of D and ^{3}He are shown in Figure 2 as a function of η. A faster expansion rate will permit more D and ^{3}He to survive. Since the effect of changing the expansion rate is equivalent to a very small change in η (see YSSR), we show the results in Figure 2 for $N_\nu = 3$ and $\tau_{1/2} = 10.6$ minutes. The rapid decrease in D/H with increasing η means that a *lower* limit to the primordial abundance of D will lead to a firm *upper* limit to η. If we can bound D/H from above [in fact, in § III we will bound (D + ^{3}He)/H from above], we may obtain a *lower* bound to η. This will be discussed in § III.

Finally, consider the synthesis of ^{7}Li. As mentioned earlier, the gaps at mass-5 and mass-8 inhibit the production of nuclei heavier than ^{4}He. Nevertheless, some ^{7}Li is synthesized; see

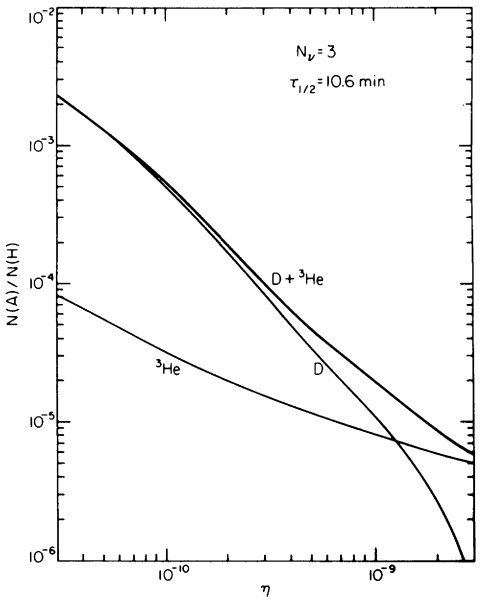


FIG. 2.—The abundance (by number relative to H) of D, ^{3}He, and their sum as a function of η for $N_\nu = 3$ and $\tau_{1/2} = 10.6$ minutes.

Figure 3. For low nucleon abundance ($\eta \lesssim 3 \times 10^{-10}$), ^{7}Li is produced mainly via $^4\mathrm{He} + {}^3\mathrm{H} \rightarrow {}^7\mathrm{Li} + \gamma$. For higher nucleon abundance, most of the ^{7}Li comes from the decay of ^{7}Be produced via $^3\mathrm{He} + {}^4\mathrm{He} \rightarrow {}^7\mathrm{Be} + \gamma$. Here too, the effect of a speedup in the expansion rate is indistinguishable from a slight change in η (see YSSR). The primordial production of ^{7}Li is shown in Figure 3 for $N_\nu = 3$, $\tau_{1/2} = 10.6$ minutes. We will return to ^{7}Li in § III when we use the new results of Spite and Spite (1982*a*, *b*) to constrain η.

To facilitate the comparison between theory and observation, Table 1 provides a detailed summary of the predicted abundances of D, ^{3}He, and ^{7}Li for $1 \leq \eta_{10} \leq 20$.

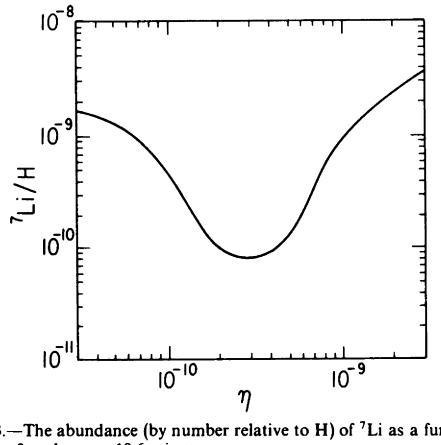


FIG. 3.—The abundance (by number relative to H) of ^{7}Li as a function of η for $N_\nu = 3$ and $\tau_{1/2} = 10.6$ minutes.

TABLE 1

PRIMORDIAL ABUNDANCES OF D, ^{3}He, AND ^{7}Li

$10^{10}\eta$	10^5(D/H)	10^5(D + ^{3}He)/H	10^{10}(^{7}Li/H)
1	49	53	4.4
1.5	25	28	1.8
2	16	18	1.1
3	8.1	9.7	0.76
4	5.1	6.5	1.0
5	3.6	4.8	1.7
6	2.7	3.8	2.7
7	2.1	3.1	3.9
8	1.7	2.6	5.3
9	1.4	2.3	6.9
10	1.1	2.0	8.6
15	0.48	1.2	17
20	0.23	0.87	25

III. DEUTERIUM, HELIUM-3, AND LITHIUM-7

Since the barely bound deuteron is so easily destroyed, it is hard to find an astrophysical site, other than the big bang, where it can be produced in its observed abundance (Epstein, Lattimer, and Schramm 1976). It is generally agreed, therefore, that the observed abundance of D provides a *lower* limit to the primordial abundance.

Observations of deuterated molecules in the atmosphere of Jupiter provide a possible estimate of the protosolar deuterium abundance. As is well known from interstellar studies, chemical fractionation may distort the true D/H ratio; observations of deuterated molecules in the atmospheres of the giant planets may provide more information about the chemistry and physics of planetary atmospheres and less about protosolar abundances. A possible way to minimize the model-dependent fractionation effects is to concentrate on those molecules which contain most of the H and most of the D. Using observations of the $P_4(1)$ line of HD and the $S_4(1)$ line of H_2, Trauger *et al.* (1973) derived: $(\mathrm{D/H})_{\mathrm{Pre}\odot} = (2.1 \pm 0.4) \times 10^{-5}$. McKellar, Goetz, and Ramsey (1976) used new laboratory parameters for the $P_4(1)$ line to revise the Trauger *et al.* (1973) value to $(\mathrm{D/H})_{\mathrm{Pre}\odot} = (5.6 \pm 1.4) \times 10^{-5}$. Then Trauger, Roesler, and Mickelson (1977) remeasured the $P_4(1)$ line and found $(\mathrm{D/H})_{\mathrm{Pre}\odot} = (5.1 \pm 0.7) \times 10^{-5}$. The Trauger, Roesler, and Mickelson (1977) observations were reanalyzed by Combes and Encrenaz (1979), who compared HD with CH_4 and, after making a necessary—and uncertain—correction for the solar C/H ratio, derived $(\mathrm{D/H})_{\mathrm{Pre}\odot} = (1.3 \pm 1.0) \times 10^{-5}$. More recently, Encrenaz and Combes (1982) have revised their previous result to $(\mathrm{D/H})_{\mathrm{Pre}\odot} = (1.2\text{–}3.1) \times 10^{-5}$. A still more model-dependent, recent estimate by Kunde *et al.* (1982) is based on CH_3D: $(\mathrm{D/H})_{\mathrm{Pre}\odot} = 3.6^{+1.0}_{-1.4} \times 10^{-5}$.

An indirect estimate of the protosolar abundance of deuterium can be inferred from the meteoritic and solar wind data on ^{3}He (to be discussed below) (Black 1971, 1972; Geiss and Reeves 1972). Using this approach, the abundance relative to ^{4}He is found to be $(\mathrm{D}/{}^4\mathrm{He})_{\mathrm{Pre}\odot} = (2.8 \pm 0.8) \times 10^{-4}$. If we adopt a lower bound to the ^{4}He abundance, $({}^4\mathrm{He/H})_\odot \geq 0.08$, then $(\mathrm{D/H})_{\mathrm{Pre}\odot} \geq 1.6 \times 10^{-5}$ is derived. This result is consistent with the estimates from deuterated molecules on Jupiter, and, taken together, all the solar system data suggest a protosolar abundance in excess of $1\text{–}2 \times 10^{-5}$.

Ultraviolet absorption studies (York and Rogerson 1976; Vidal-Madjar *et al.* 1977; Vidal-Madjar *et al.* 1983) have led to a determination of the deuterium abundance in the inter-

stellar gas in the vicinity ($\lesssim 1$ kpc) of the Sun. Although an average abundance $\langle D/H\rangle_{ISM} = 2 \times 10^{-5}$ has been derived, individual abundance determinations span the range $(D/H)_{ISM} \approx (\frac{1}{4}\text{–}4) \times 10^{-5}$ (see the discussion in YSSR and Bruston *et al.* 1981). Since the weak deuterium lines are in the wings of the much stronger hydrogen lines, complex structure in the interstellar medium (e.g., many clouds at various velocities) and in the circumstellar medium (Laurent 1983) may complicate the analysis. York (1983) notes that for the "cleanest" lines of sight there is no evidence for a variation in the deuterium abundance, and he suggests that $\langle D/H\rangle_{ISM} = 2 \times 10^{-5}$ and $(D/H)_o \geq 1 \times 10^{-5}$ are reliable inferences from the present data. We note that the solar system and interstellar estimates of $D/H \approx 2 \times 10^{-5}$ are consistent. To perhaps err on the side of caution, we shall adopt as a *lower* limit to the present (or presolar) abundance $(D/H)_o \geq 1 \times 10^{-5}$. On the assumption that the present abundance provides a lower limit to the primordial production (no post-big-bang production of D), we may derive (see Fig. 2 and Table 1) a safe-bet upper limit to the nucleon abundance:

$$(D/H)_p \geq (D/H)_o \geq 1 \times 10^{-5} \Rightarrow \eta_{10} \leq 10 . \quad (9)$$

Note that if we had adopted the best-bet value of $(D/H)_p \gtrsim 2 \times 10^{-5}$, the limit on the nucleon abundance would have been $\eta_{10} \lesssim 7$.

Gas-rich meteorites, lunar soil, and the foil placed on the Moon by the *Apollo* astronauts provide samples of solar wind–implanted ^{3}He. Since deuterium was burned to helium-3 during the Sun's approach to the main sequence, solar wind ^{3}He is the sum of presolar D and ^{3}He. From gas-rich meteorites, Black (1972) found $[(D + {}^3He)/{}^4He]_{Pre\odot} = (3.9 \pm 0.3) \times 10^{-4}$. Analyzing ^{3}He implanted on the lunar foil, Geiss and Reeves (1972) concluded that $[(D + {}^3He)/{}^4He]_{Pre\odot} = (4.3 \pm 0.3) \times 10^{-4}$. These results are consistent with the abundance derived from spectroscopic studies of ^{3}He in the wings of ^{4}He in quiescent solar prominences (Hall 1975): $[(D + {}^3He)/{}^4He]_{Pre\odot} = (4 \pm 2) \times 10^{-4}$. We therefore adopt for an estimate of the presolar abundance of D plus ^{3}He

$$[(D + {}^3He)/{}^4He]_{Pre\odot} = (4.1 \pm 0.2) \times 10^{-4} . \quad (10)$$

In contrast to the gas-rich meteorites, carbonaceous chondrites provide a sample of the primitive solar nebula in which the deuterium is unprocessed. Several components of ^{3}He are found in carbonaceous chondrites; that with the smallest ^{3}He/^{4}He ratio is identified with the presolar ^{3}He. Anders, Heymann, and Mazor (1970) find $({}^3He/{}^4He)_{Pre\odot} = (1.25 \pm 0.76) \times 10^{-4}$, a result consistent with Black's (1972) estimate of $({}^3He/{}^4He)_{Pre\odot} = (1.5 \pm 1.0) \times 10^{-4}$. For our safe-bet estimate of the presolar abundance of ^{3}He we adopt

$$({}^3He/{}^4He)_{Pre\odot} = (1.3 \pm 0.6) \times 10^{-4} . \quad (11)$$

From equations (10) and (11) we may derive an indirect estimate of the presolar abundance of deuterium:

$$(D/{}^4He)_{Pre\odot} = (2.8 \pm 0.8) \times 10^{-4} . \quad (12)$$

To convert the above abundance estimates relative to ^{4}He to abundances relative to H, we need to know the solar abundance of ^{4}He. Direct measurements of ^{4}He line intensities in solar prominences (Heasley and Milkey 1978) yield $y_\odot = ({}^4He/H)_\odot = 0.10 \pm 0.025$. Indirect estimates of $y_\odot$ follow from a comparison of solar models with observations. Bahcall *et al.* (1982) find[5] $Y_\odot = 0.25 \pm 0.01$ (3σ); this corresponds to $y_\odot = 0.086 \pm 0.005$. This estimate agrees with the result obtained by Christensen-Dalsgaard and Gough (1980) from a detailed comparison of the observed periods of solar surface oscillations with model calculations. If we assume that $({}^4He/H)_\odot$ lies in the range $0.08 \lesssim y_\odot \lesssim 0.10$, we find from equations (10)–(12) that

$$[(D + {}^3He)/H]_{Pre\odot} \lesssim 4.3 \times 10^{-5} , \quad (13a)$$

$$[{}^3He/H]_{Pre\odot} \lesssim 1.9 \times 10^{-5} , \quad (13b)$$

$$[D/H]_{Pre\odot} \gtrsim 1.6 \times 10^{-5} . \quad (13c)$$

We now use these estimates to place an *upper* limit on the primordial abundance of D plus ^{3}He and, hence, a *lower* limit to η (see Fig. 2 and Table 1). Since deuterium is easily destroyed, the primordial abundance may exceed the present abundance (or, the presolar abundance) by a large factor. Although models of galactic evolution (Truran and Cameron 1971; Audouze and Tinsley 1976) suggest astration of no more than a factor of 2–3, large destruction may be possible (Pagel 1982). In stars, deuterium is burned to ^{3}He, which is more difficult to destroy. Some ^{3}He will survive stellar processing; the observed ^{3}He therefore provides a constraint on the primordial abundance of deuterium (i.e., too much primordial D will result in too much D and/or ^{3}He in the presolar nebula). In Appendix B we discuss the survival of ^{3}He in some detail. If g is the fraction of ^{3}He which survives processing through stars, then (see eq. [B3′])

$$\left(\frac{D + {}^3He}{H}\right)_p \leq \left(\frac{D}{H}\right)_{Pre\odot} + \frac{1}{g}\left(\frac{{}^3He}{H}\right)_{Pre\odot} = \left(\frac{D + {}^3He}{H}\right)_{Pre\odot} + \left(\frac{1}{g} - 1\right)\left(\frac{{}^3He}{H}\right)_{Pre\odot} . \quad (14)$$

Using the upper limits from equations (13a) and (13b), we find

$$\left(\frac{D + {}^3He}{H}\right)_p \leq \left(2.4 + \frac{1.9}{g}\right) \times 10^{-5} . \quad (15)$$

From the analysis in Appendix B we are led to a safe-bet lower limit to g of $g \geq \frac{1}{4}$; a best-bet lower limit is $g \geq \frac{1}{2}$. For the safe-bet value of $g \geq \frac{1}{4}$,

$$[(D + {}^3He)/H]_p \leq 10 \times 10^{-5} \Rightarrow \eta_{10} \geq 3 , \quad (16)$$

whereas for the best-bet value of $g \geq \frac{1}{2}$,

$$[(D + {}^3He)/H]_p \leq 6.2 \times 10^{-5} \Rightarrow \eta_{10} \geq 4 . \quad (17)$$

If the arguments concerning the survival of ^{3}He outlined in Appendix B are correct, it is expected that stellar production of ^{3}He in the last 4.6×10^9 yr has enhanced the interstellar abundance relative to the presolar abundance (Rood, Steigman, and Tinsley 1976). Recent observations of the $^3He^+$ hyperfine line in three H II regions (W3, W51, and W43) by Wilson, Rood, and Bania (1983) are in apparent agreement with this prediction.

Finally, let us turn to the observations of ^{7}Li. Until very recently our information concerning ^{7}Li was limited to observations of meteorites (Mason 1979), the interstellar medium (Vanden Bout *et al.* 1978), and the atmospheres of some

[5] The helium-4 abundance by mass Y, is related to the ^{4}He/H ratio y by $Y = 4y(1 + 4y)^{-1}(1 - Z)$, where Z is the heavy-element abundance by mass. For the Sun we have assumed $Z_\odot = 0.02$.

Population I stars (Zappala 1972; Duncan 1981). All these observations are consistent with a present abundance of $(^7\mathrm{Li/H})_o$ within a factor of 2 of 1×10^{-9}. Recently, Spite and Spite (1982*a*, *b*) have observed ^{7}Li in the atmospheres of 12 halo and old disk stars. From their data Spite and Spite derive an abundance

$$(^7\mathrm{Li/H})_{\mathrm{Pop\ II}} = (1.12 \pm 0.38) \times 10^{-10} . \quad (18)$$

If we refer to Figure 3, we see that this estimate for the primordial ($\sim$ Population II) abundance puts us in the valley of the ^{7}Li/H versus η curve: $2 \lesssim \eta_{10} \lesssim 5$. It is, however, possible that there was some astration of ^{7}Li prior to the formation of these stars:

$$\left(\frac{^7\mathrm{Li}}{\mathrm{H}}\right)_{\mathrm{Pop\ II}} = f_7 \left(\frac{^7\mathrm{Li}}{\mathrm{H}}\right)_p ; \quad f_7 \leq 1 . \quad (19)$$

But, since deuterium is more fragile than lithium, D too would be astrated from its primordial value:

$$\left(\frac{\mathrm{D}}{\mathrm{H}}\right)_o \approx \left(\frac{\mathrm{D}}{\mathrm{H}}\right)_{\mathrm{Pre}\odot} = f_2 \left(\frac{\mathrm{D}}{\mathrm{H}}\right)_p ; \quad f_2 \leq f_7 . \quad (20)$$

Combining equations (19) and (20), we find a bound on the primordial ratio of ^{7}Li to D:

$$\left(\frac{^7\mathrm{Li}}{\mathrm{D}}\right)_p \leq \frac{(^7\mathrm{Li/H})_{\mathrm{Pop\ II}}}{(\mathrm{D/H})_o} \lesssim 1.5 \times 10^{-5} . \quad (21)$$

From Figures 2 and 3 and Table 1, it follows that $\eta_{10} \lesssim 7$. In obtaining this best-bet limit to η we have compared the lithium abundance in old stars with the deuterium abundance at present or in the presolar nebula. We therefore expect the inequality $f_2 \leq f_7$ to hold not only because deuterium is more fragile, but also because we have allowed more time for its destruction. On the other hand, as we discuss in Appendix A, uncertainties in the key rates, which determine the predicted abundance of ^{7}Li, lead to uncertainties in the abundance of a factor of 2. If we allow for no astration of ^{7}Li ($f_7 = 1$), but allow the predicted abundance to be uncertain by a factor of 2, we also obtain the limit $\eta_{10} \lesssim 7$.

So, to summarize, from observations of D, ^{3}He, and ^{7}Li, limits to the primordial abundances of these elements have been derived, and these have led to upper ($\eta_{10} \lesssim 7$–10) and lower ($\eta_{10} \gtrsim 3$–4) limits to the ratio of nucleons to photons.

IV. PRIMORDIAL HELIUM-4

Of the elements produced during primordial nucleosynthesis, helium-4 provides the potentially most stringent test of the consistency of the standard model and is the most powerful probe of the conditions in the early universe. There are several reasons for the special position occupied by ^{4}He. Next to hydrogen, helium is the most abundant element in the universe. The large observed abundance of ^{4}He cannot be accounted for by stellar production alone. It is a significant achievement of the standard model that primordial nucleosynthesis can account for the observed abundance. Furthermore, since ^{4}He abundances may be determined with relatively high precision, very detailed comparisons between theory and observation are possible. ^{4}He thus differs markedly from the other light elements, whose abundances are known, at best, only to within a factor of 2. In addition, data on the abundance of ^{4}He exist not only for our own Galaxy, but for many nearby and distant galaxies as well. There is, therefore, the possibility that the observational data may lead to an accurate estimate of the primordial abundance of ^{4}He.

There are, of course, some problems. Since ^{4}He is synthesized in stars, some of the observed ^{4}He is not primordial. Because ^{4}He is not easily destroyed, the observed abundances may only provided an upper limit to the primordial abundance ($Y_o \geq Y_p$). Since an accurate estimate of Y_p is required, the crucial problem is to derive Y_p from the data on Y_o. In the attempt to infer from Y_o, different astrophysical objects offer different advantages and pose different problems. For example, nearby, bright galactic H II regions can be resolved spatially, and line strengths can be measured to high accuracy. However, galactic H II regions are significantly contaminated with stellar debris; thus some of the observed ^{4}He has been synthesized in stars. In contrast, the ^{4}He is more nearly primordial in "gassy" galaxies which have low abundances of the heavy elements. However, most such objects are distant and unresolved; what is observed may be the superposition of many H II regions. Such circumstances complicate the analysis necessary for an accurate determination of the ^{4}He abundance.

If ^{4}He is to provide a probe of cosmology and particle physics, an accurate estimate of the helium abundance $y_p = (\mathrm{He/H})_p$ is required. Before discussing the theoretical implications of the observational data, we will attempt to critically review the present observational situation, paying particular attention to the uncertainties. Previously, the observational situation was reviewed by YSSR and OSSTY. More recently, in an excellent review of the abundances of the light elements, Pagel (1982) has presented a comprehensive and critical discussion of ^{4}He. There has been a general consensus that all current data are consistent with a primordial abundance no greater than $Y_p \lesssim 0.25$ ($y_p \lesssim 0.083$).[6] Here, we focus on two issues: (i) Can a primordial abundance be derived from the data? (ii) If so, what are the uncertainties associated with such an estimate?

Pagel (1982) has discussed the estimates of Y_p derived from stellar observations and notes, "... the uncertainties just about cover the range of interest [$0.2 < Y_p < 0.3$] which is encouraging in a way, but is of no help in estimating Y_p." We have nothing more to add regarding stellar determinations of ^{4}He and refer the reader to Pagel (1982).

In Table 2 we have gathered together the *average* ^{4}He abundance determined from emission lines from H II regions in a wide variety of objects; the data covers a modest range in the heavy-element abundance. For objects with roughly "galactic" abundance [$\mathrm{O/H} \approx (4 \pm 2) \times 10^{-4}$], the helium abundance is in the range $Y \approx 0.30 \pm 0.02$. We note that from this perspective, the solar helium abundance ($0.23 \lesssim Y_\odot \lesssim 0.27$; see § III) seems anomalously small. This hint and the scatter in Table 2 suggest that the estimate, $\Delta Y \approx 0.02$, might be too conservative. It does seem clear from Table 2 that the helium abundance is generally lower where the oxygen abundance is lower: for $\mathrm{O/H} \lesssim 10^{-4}$, $Y \approx 0.25 \pm 0.02$.

To study this trend further we have collected in Table 3 helium abundances determined from observations of individual H II regions with oxygen abundances which are estimated to be about an order of magnitude below solar ($\mathrm{O/H} \lesssim 7 \times 10^{-5}$). Even with this restricted group the scatter is large, $0.22 \lesssim Y \lesssim 0.30$. For the data in Table 3, $\langle y \rangle = 0.086$ ($\langle Y \rangle = 0.26$). From Tables 2 and 3 it is clear that all the ^{4}He abundance

[6] For the primordial abundance ($Z_p = 0$) the number ratio and the mass fraction are related by $Y_p = 4y_p(1 + 4y_p)^{-1}$.

TABLE 2

AVERAGE HELIUM ABUNDANCES

Object	$10^4\langle O/H\rangle$	He/H (Y)[a]	Reference
⟨Orion⟩	5.6	0.100 (0.29)	PTP III
	4.2	0.101 (0.29)	Dufour 1975
	4.2	0.114 (0.31)	Hawley 1978
⟨Gal. H II reg.⟩	4.0	0.117 (0.32)	Hawley 1978
	2.6	0.127 (0.34)	Talent and Dufour 1979
	2.5	0.097 (0.28)	Peimbert, Torres-Peimbert, and Rayo 1978
⟨LMC⟩	3.8	0.084 (0.25)	PTP I
	2.7	0.103 (0.29)	Dufour 1975
⟨SMC⟩	1.1	0.078 (0.24)	PTP II
	1.1	0.094 (0.27)	Dufour 1975
	1.0	0.081 (0.24)	Dufour and Harlow 1977
⟨DELG⟩[b]	1.4	0.092 (0.27)	French 1980
	1.1	0.080 (0.24)	Lequeux *et al.* 1979
	1.0	0.082 (0.25)	Kunth 1982
	0.95	≥0.082[c] (≥0.25)	Kinman and Davidson 1981
	0.58	0.078 (0.24)	Tully *et al.* 1981

[a] $Y = 4y(1 + 4y)^{-1}$, y = He/H.
[b] DELG = extragalactic H II regions, dwarf emission-line galaxies, blue compact galaxies, etc.
[c] Kinman and Davidson 1981 make no corrections for neutral helium.

determinations are compatible with $Y = 0.26 \pm 0.04$. Although this result is reassuring, smaller uncertainties are required if significant constraints on cosmology and particle physics are to be derived from primordial nucleosynthesis. The results presented in Tables 2 and 3 represent averages over many H II regions and/or parts of H II regions in several galaxies. To obtain a better sense of the range in y and the uncertainty in y, it is valuable to consider the data in more detail.

It cannot be emphasized too strongly that abundances are not observed, emission lines are. There are numerous steps to be taken, each involving uncertainties, in going from the observational data to the derived abundances. The strengths of various recombination lines of He^+ and He^{++} are measured and compared with those of H^+ to derive y^+ and y^{++}. In making this comparison one relies on the accuracy of standard star observations to obtain relative line fluxes, and in addition, a correction (usually small) for reddening must be made. To derive the total helium abundance, a correction for neutral helium must also be made; the fact that the He and H Strömgren zones may be of unequal size introduces uncertainty into this correction. For example, selective absorption by dust within the H II region and/or line blanketing in the exciting star(s) can reduce the He^+ zone relative to the H^+ zone. The He^+ zone will be smaller than the H^+ zone if the effective temperature of the ionizing radiation is less than $\sim 37 \times 10^3$ K. In these cases the measured value of y^+ will also depend on the beam size and the position in the H II region where the observation is made (Shaver *et al.* 1983). Since these effects tend to reduce the observed value of y^+, Shaver *et al.* (1983) suggest that the *upper envelope* of the y^+ versus O/H distribution may provide a better estimate of the helium abundance. These problems of deriving accurate abundances from accurate observational data are exacerbated if the emission nebulae are inhomogeneous and are ionized by radiation from a distribution (in space as well as in spectral type) of stars. In general, they are. There is ample evidence in the literature which suggests that accurate abundances are difficult to obtain.

By studying nearby, bright, large H II regions, many of the difficulties mentioned above may be mitigated. The exciting stars may be identified; position-by-position reddening and ionization corrections may be made; abundances derived from several different lines may be intercompared. Even so, problems still persist. For example, Hawley (1978) has studied 14 galactic H II regions in detail. Even within a given H II region he finds

TABLE 3

HELIUM ABUNDANCES IN H II REGIONS OF LOW METAL ABUNDANCE

Object	10^4O/H	He/H (Y)	Reference
POX 120	0.71	0.090 (0.26)	Kunth 1982
POX 105	0.70	0.077 (0.24)	Kunth 1982
Mrk 600	0.69	≥0.098[a] (≥0.28)	Kinman and Davidson 1981
Mrk 36	0.69	0.089 (0.26)	French 1980
DDO 64	0.68	≥0.083[a] (≥0.25)	Kinman and Davidson 1981
PHL 293B	0.60	0.095 (0.28)	French 1980
VII Zw 403	0.58	0.078 (0.24)	Tully *et al.* 1981
POX 186	0.53	0.081 (0.25)	Kunth 1982
A1228+12	0.44	≥0.095[a] (≥0.28)	Kinman and Davidson 1981
CG 1116+51	0.43	0.107 (0.30)	French 1980
TOL 65	0.34	0.079 (0.24)	Kunth 1982
I Zw 18	0.15	0.076 (0.23)	Lequeux *et al.* 1979
I Zw 18	0.14	≥0.069[a] (≥0.22)	Kinman and Davidson 1981

[a] Kinman and Davidson 1981 make no correction for neutral He.

variations in y^+ and y (corrected for neutral He) which far exceeds the internal uncertainties. For example, for position 1 in M16, Hawley (1978) finds $y^+ = 0.079$ and derives $y = 0.100$, whereas for position 2: $y^+ = 0.063$ and $y = 0.154$. For M20, he finds $y_1{}^+ = y_2{}^+ = 0.068$, whereas $y_1 = 0.134$ and $y_2 = 0.110$. For galactic H II regions the ionization correction is often large (and, clearly, uncertain). If these same H II regions were distant and unresolved, "average" ionization and reddening corrections would have to be made. From Hawley's work it is clear that such average corrections are uncertain.

Hawley's results are borne out by other investigations. For example, in their study of the Orion nebula, Peimbert and Torres-Peimbert (1977, hereafter PTP III) have data from 12 different positions within the nebula. They find a wide spread in He^+ ($0.060 < y^+ < 0.090$). A significant and variable ionization correction reduces the spread in He ($0.091 < y < 0.109$, $0.27 < Y < 0.30$). Similar dispersion is found by Talent and Dufour (1979) in their study of four galactic H II regions, by Dufour (1975) in his study of SMC and LMC H II regions, and by Dufour and Harlow (1977) in their study of 10 SMC H II regions. Furthermore, Dufour and Harlow's (1977) data show that different lines lead to different estimates of y^+; the dispersion is larger than would be expected on statistical grounds alone. This is also found by Rayo, Peimbert, and Torres-Peimbert (1982, hereafter RPTP) in M101. Indeed, the large dispersion of estimates for y^+ from different lines leads Tully *et al.* (1981), in their study of VII Zw 403, to conclude, "... the scatter in abundance estimates from the individual lines is significantly greater than what was anticipated based on the internal errors ...." Tully *et al.* (1981) discuss a number of reasons why the helium abundance may be underestimated and, echoed by Shaver *et al.* (1983), note that, "in the absence of very detailed spatial coverage on well-resolved emission regions, a better estimate of the helium abundance may come from the *upper envelope* in the helium versus metal abundance plots." For VII Zw 403, Tully *et al.* (1981) derive a helium abundance $Y = 0.24 \pm 0.04$; with such a large uncertainty, this measurement is of no cosmological use.

It is often claimed that many of the sources of uncertainty which affect the derivation of the helium abundance from the observational data are reduced or become negligible if one concentrates on emission nebulae with low heavy-element abundance. Ionization and reddening corrections are reduced for objects where line blanketing, selective absorption, etc., are less important. However, it must be noted that the "low" metal abundance emission nebulae studied so far (with the exception of I Zw 18, the most metal deficient H II region yet discovered) have roughly an order-of-magnitude higher "metal" abundance than the globular clusters; clearly, we are not yet seeing primordial material. If a precision of greater than 5%–10% or so is desired ($\Delta Y \leq \pm 0.01$–0.02), ionization corrections cannot be neglected (Tully *et al.* 1981; Kunth 1982). Since most low-Z emission nebulae are unresolved, the potential advantages they offer are, to some extent, compensated for by the effects of unresolved inhomogeneities. The dispersion in helium abundances derived for different H II regions with comparable, but low, abundances of the heavy elements lends some support to this suggestion. For example, Lequeux *et al.* (1979) find similar metal abundances for II Zw 40, II Zw 70, and IC 10(1) ($Z \approx 0.2\ Z_\odot$) but find helium abundances which range from $y = 0.074$ ($Y = 0.23$) to $y = 0.084$ ($Y = 0.25$).

Even if very accurate estimates of the present helium abundance were available, one would still have to account for the contamination due to stellar production since the big bang. We have already noted that the abundance of ^{4}He is somewhat lower where the heavy-element (usually oxygen) abundance is lower. If a linear correlation between Y and Z could be established, the relation could be extrapolated to $Z = 0$ to obtain the primordial abundance Y_p. The present observational situation is murky at best. As already noted, there is a dispersion in Y even for objects with similar values of Z. As a result, most studies of Y versus Z (or O/H) yield a scatter diagram. Evidence for a Y versus Z correlation ($\Delta Y/\Delta Z \approx 3$) is found by Peimbert and Torres-Peimbert and their collaborators (Peimbert and Torres-Peimbert 1974, 1976, hereafter PTP I, PTP II; Peimbert, Torres-Peimbert, and Rayo 1978; Lequeux *et al.* 1979). Although French (1980) claimed evidence for a Y versus Z correlation ($\Delta Y/\Delta Z \approx 3.2$), his data were contaminated by an anomalously low value of $Y = 0.17$ for I Zw 18.[7] In contrast, many other observers find no evidence in their data to support a linear Y versus Z correlation (Dufour 1975; Smith 1975; Hawley 1978; Talent and Dufour 1979; Kinman and Davidson 1981; Kunth 1982; Shaver *et al.* 1983). For example, although Peimbert, Torres-Peimbert, and Rayo (1978) find $\Delta Y/\Delta Z = 2.0$ from a study of galactic H II regions, Shaver *et al.* (1983), in a recent study of galactic H II regions, find no evidence for a correlation and place the limit at $\Delta Y/\Delta Z \leq 0.8$.

It must be kept in mind that very accurate data are required if a Y versus Z correlation is to be revealed. For example, Kunth's (1982) data are perhaps the most accurate and extensive to date, and although he finds no Y versus Z correlation, his limit, $\Delta Y/\Delta Z = 1.2 \pm 3.2$, cannot entirely exclude a trend in Y versus Z. At present then, the observational situation is ambiguous. We do expect that Y will increase with Z (stars *do* produce ^{4}He), and the data in Tables 2 and 3 do suggest that "solar" metal abundance H II regions are enhanced in ^{4}He relative to "low" metal abundance H II regions. However, the uncertainties in Y are apparently still too large to conclude that there exists a convincing linear relation between Y and Z.

There is little succour to be found in turning to theory. Early models of galactic evolution (which ignored mass loss) found little helium enrichment. For example, Arnett (1978) concluded that $\Delta Y/\Delta Z \approx 0.73$. In contrast, Dearborn and Trimble (1978) found that the helium enrichment increased dramatically ($\Delta Y/\Delta Z \approx 2.0$–$3.3$) when mass loss was included. Similarly, Chiosi (1979) and Chiosi and Caimmi (1979) derive $\Delta Y/\Delta Z \approx 1.8$–$3.3$ from models including mass loss. However, these results are contradicted by Maeder (1980), who finds, for evolution with mass loss included, $\Delta Y/\Delta Z \approx 0.73$–$0.77$. The discrepancy is apparently due to different choices for the initial mass function (IMF) (cf., discussion at the ESO Workshop on Primordial Helium, Munich, 1983 February). Clearly, there are large uncertainties in the theory too (e.g., the IMF, the evolution of very low Z stars, metal-enhanced star formation, etc.).

All the data of which we are aware are consistent with a universal abundance of helium in the range $Y = 0.26 \pm 0.04$. We should be pleased that the helium abundance is known so well (uncertainties as small as 15% are extremely rare for the estimates of the abundance of any other element) and reassured

[7] According to French (1981), the helium abundance obtained for I Zw 18 was in error, being too low because of contamination from a night-sky line.

that the abundance derived from observations falls nicely in the range predicted by primordial nucleosynthesis (see Fig. 1). The difficulties arise if one attempts a more detailed comparison between the predicted (primordial) and observed abundances. For such a comparison to yield significant constraints requires an accurate ($\Delta Y \lesssim 0.01$) estimate of the primordial abundance Y_p. Although we have found no convincing evidence for a linear Y versus Z relation, we do believe that the estimates of Y for low-Z objects provide a reasonable estimate of the upper limit to Y_p. From the data assembled in Table 2 it is not unreasonable to assert that $Y_p < 0.26$.

It is, however, much more difficult to bound Y_p from below. Most of the low estimates of Y_p (<0.22–0.23) are derived from extrapolating an assumed ΔY versus ΔZ relation. As already emphasized, this procedure is suspect. There are also individual objects for which a low value of Y has been derived. I Zw 18 is one example. However, disregarding French's (1980) result, different observers have derived $0.22 < Y(\text{I Zw } 18) < 0.25$. Thus we are hesitant to base a determination of Y_p on the observations of a single object. Another case in point is the recent study of M101 (RPTP). RPTP studied five H II regions and the nucleus of M101. Only for three of the H II regions were the data of sufficiently high quality to permit them to derive a helium abundance. However, they found that different helium lines yield different abundances y^+; thus, half the lines were not considered because they might be strongly affected by stellar emission or absorption. To determine the appropriate ionization correction factor (to account for y^0) they compared their observations with Stasinska's (1980) models. Only for NGC 5471 did they find any fit, and then they derived $y = 0.0721 \pm 0.0009$ ($Y = 0.224 \pm 0.002$). We suspect that their estimate of the uncertainty is artificially small. Indeed, under similar circumstances in their study of VII Zw 403, Tully *et al.* (1981) concluded that the spread in derived abundances was more representative of the true uncertainty than the formal internal errors. For comparison, Smith (1975) found for NGC 5471, $y \approx 0.074$–0.078 ($Y \approx 0.23$–0.24). Clearly, caution is suggested in dealing with any estimate of the primordial helium abundance derived from the study of one object.

Recently, Kunth (1982) has studied 13 blue compact galaxies with $Z \lesssim 0.2\, Z_\odot$. For this sample Kunth finds $0.23 < Y < 0.27$ and concludes that there is no evidence for a Y versus Z correlation. From a weighted average of his own data he derives a *primordial* abundance $Y_p = 0.245 \pm 0.003$. Allowing for 3 σ uncertainties, Kunth's data suggest $0.236 \leq Y_p \leq 0.254$.

In previous studies of the helium abundance (OSSTY; YSSR; Pagel 1982) it was concluded that $Y_p \lesssim 0.25$. Nothing in the foregoing discussion leads to a contradiction of this conclusion. From the currently accumulated data, $Y_p \gtrsim 0.23$ is reasonable, but Y_p as low as 0.22 may also be consistent with the data. We find no convincing evidence that Y_p is less than 0.22. Keeping these conclusions (especially the uncertainties) in mind, we will now compare the derived primordial abundance with the theoretical predictions and discuss the implications.

We shall utilize Figure 1 for our comparisons between theory and observation. Figure 1 has been corrected for the systematic decrease of 0.003 in Y calculated recently by Dicus *et al.* (1982). To facilitate these comparisons we present the relevant numerical data in Tables 4 and 5. Recall that in § III we used the data on D, D + ^{3}He, and ^{7}Li to constrain the nucleon abundance to the range $3\text{–}4 \lesssim \eta_{10} \lesssim 7\text{–}10$. Consistency among D, ^{3}He, ^{7}Li, and ^{4}He requires that we limit our attention to this range. The following conclusions follow from Figure 1 and Tables 4 and 5 (and the assumption that $\tau_{1/2} \geq 10.4$ minutes).

1. If, indeed, $\eta_{10} \gtrsim 3$, then $N_\nu = 4$ is permitted only if $Y_p \gtrsim 0.253$. Although we believe that $Y_p \lesssim 0.25$, to exclude $N_\nu = 4$ requires third-decimal place accuracy in the adopted upper limit to Y_p.

2. The mass fraction $Y_p \leq 0.25$ is consistent with $N_\nu \leq 3$ provided that $\eta_{10} \leq 7.2$. Note, however, if we use Kunth's

TABLE 4

CONSTRAINTS ON η FROM UPPER LIMITS TO Y_p FOR $\tau_{1/2} \geq 10.4$ MINUTES

$Y_p \leq$	N_ν	$\eta_{10} \leq$
0.26	2	[a]
	3	[a]
	4	6
0.25	2	[a]
	3	8
	4	2
0.24	2	[a]
	3	3
	4	<2
0.23	2	4
	3	<2
	4	<1
0.22	2	<2

[a] A more restrictive constraint, $\eta_{10} < 10$, follows from the requirement that $(\text{D/H})_p > 1 \times 10^{-5}$.

TABLE 5

LOWER LIMIT TO Y_p AS A FUNCTION OF η FOR $\tau_{1/2} \geq 10.4$ MINUTES

η_{10}	N_ν	$Y_p \geq$[a]
2	2	0.221
	3	0.235
	4	0.247
3	2	0.226
	3	0.240
	4	0.253
4	2	0.230
	3	0.244
	4	0.256
5	2	0.232
	3	0.246
	4	0.259
6	2	0.234
	3	0.248
	4	0.260
7	2	0.235
	3	0.250
	4	0.261
8	2	0.236
	3	0.251
	4	0.263
9	2	0.237
	3	0.252
	4	0.264
10	2	0.238
	3	0.252
	4	0.265

[a] These values for Y_p are accurate to ± 0.002.

(1982) 3 σ upper limit, $Y_p \leq 0.254$, $N_\nu \leq 3$ is allowed for all $\eta_{10} \leq 10$; $N_\nu = 4$ is just barely consistent ($\eta_{10} \geq 3$) with this limit.

3. For $\eta_{10} \geq 3$ and $N_\nu \geq 3$, $Y_p \geq 0.24$. If future observations were to imply that Y_p is less than 0.24, the standard model would remain consistent only if $N_\nu < 3$. This is possible if the τ-neutrino is heavy (> 10 MeV) and unstable.

4. Since e- and μ-neutrinos are known to be light, $N_\nu \geq 2$. For $\eta_{10} \geq 3(4)$ and $N_\nu \geq 2$, $Y_p \geq 0.226(0.230)$, so that inconsistency would arise if future observations yield $Y_p \lesssim 0.22$–0.23.

V. CONSISTENCY OF THE STANDARD MODEL

The observed abundances of deuterium and helium-4 (the two elements which require a primordial origin) differ by some four orders of magnitude. Big-bang nucleosynthesis predicts a unique relationship (depending weakly on the assumed values of $\tau_{1/2}$ and N_ν) between the primordial abundances of D (or D plus ^{3}He) and ^{4}He. This connection is shown in Figure 4, where the anticorrelation between ^{4}He and D + ^{3}He is clearly seen. The predicted primordial abundances *do* differ by some four orders of magnitude. Indeed, for $\tau_{1/2} = 10.6 \pm 0.2$ minutes and $N_\nu = 3 \pm 1$, the predicted abundances of D, ^{3}He, ^{4}He, and ^{7}Li are in excellent agreement with the observational data (see Fig. 5). This success is a significant achievement of the standard, hot big-bang model.

As Figure 4 illustrates, consistency among the abundances of D, ^{3}He, and ^{4}He permits us to constrain N_ν to $N_\nu \leq 4$. This is to be compared with the analysis presented in OSSTY, where it was noted that uncertainties in bounding η from below might permit $N_\nu > 4$. The constraint $N_\nu \leq 4$ has been reestablished here by requiring that D + ^{3}He not be overproduced primordially; this requirement corresponds to the constraint $\eta_{10} \gtrsim 3$–4.

There are claims in the literature that the standard model is inconsistent with current data (Stecker 1980, 1981; Rana 1982; Gautier and Owen 1983). We believe that such conclusions have been reached by considering a carefully selected subset of the available data (e.g., see the discussion in Olive and Turner 1981) and by ignoring the uncertainties in the abundances derived from the observational data. We have already noted that the standard model with $N_\nu = 3$ is consistent provided

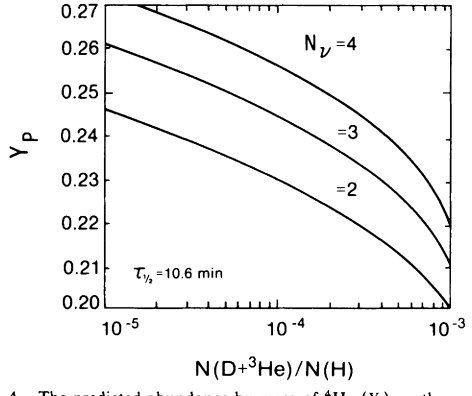


FIG. 4.—The predicted abundance by mass of ^{4}He (Y_p) vs. the predicted abundance (by number relative to H) of D plus ^{3}He for $N_\nu = 2$, 3, 4 ($\tau_{1/2} = 10.6$ minutes).

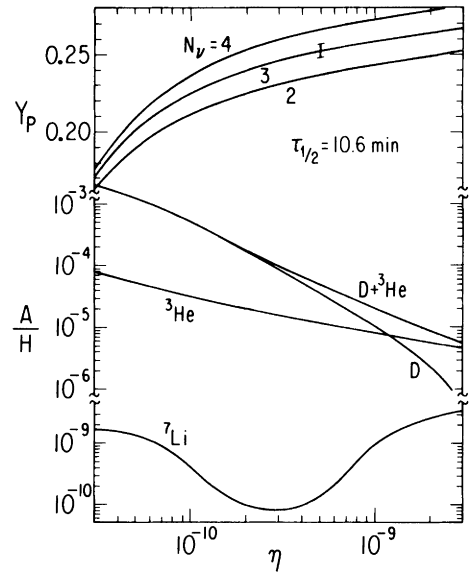


FIG. 5.—The predicted primordial abundances of ^{4}He (by mass), D, ^{3}He, and ^{7}Li (by number relative to H) as a function of η for $\tau_{1/2} = 10.6$ minutes; for ^{4}He the predictions for $N_\nu = 2$, 3, 4 are shown, and the size of the "error" bar shows the range in Y_p which corresponds to $10.4 < \tau_{1/2} < 10.8$ minutes. Note the changes in the abundance scales.

that $Y_p \geq 0.24$; if ν_τ is heavy ($N_\nu = 2$), then consistency remains for $Y_p \geq 0.22$–0.23. From our review of the observational literature we find that Y_p may be as large as 0.25. In a very recent discussion of the data presented by Kunth (1982), Kunth and Sargent (1983) reaffirm Kunth's estimate of the primordial abundance $Y_p = 0.245 \pm 0.003$ and note that this result is consistent, within the errors, with all previous high-quality results.

The data at present suggest that $Y_p \gtrsim 0.23$–0.24. What options would be available if future observations should lead to the conclusion that $Y_p \lesssim 0.22$? One possibility has already been addressed by OSSTY. If the lower bound to η derived from the abundances of D + ^{3}He is ignored, the universal abundance of nucleons may be quite small: $\eta_{10} \gtrsim 0.1$–0.3 (OSSTY). For such a low nucleon-to-photon ratio, primordial nucleosynthesis of ^{4}He yields a low abundance: $Y_p \gtrsim 0.11$–0.15. The restrictive bound on η ($\gtrsim 3 \times 10^{-10}$) imposed by D and ^{3}He may be avoided by entertaining the possibility that a generation (Population III) of exotic stars might succeed in destroying D *and* ^{3}He without the concomitant overproduction of ^{4}He or the heavy elements. Or, perhaps, the current data, being restricted to the solar system and the Galaxy, are not a fair sample of "universal" abundances. Extragalactic observations of (or, limits to) D and ^{3}He would be very valuable.

A more radical option would be to modify the standard cosmology. For example, neutrino degeneracy (Wagoner, Fowler, and Hoyle 1967; Beaudet and Goret 1976; Yahil and Beaudet 1976; David and Reeves 1980) will change the relative abundances of D, ^{3}He, and ^{4}He. Anisotropy may be another panacea. Although small anisotropy will increase the ^{4}He abundance by speeding up the expansion rate (Hawking and Tayler 1966; Thorne 1967; Barrow 1976), Matzner and Rothman (1982) suggest that larger anisotropy may decrease

the abundance of the ^{4}He without changing that of D significantly.

In extremis, of course, the standard cosmology could be discarded in favor of more exotic models such as a cold (or tepid) bang (Rees 1978) or even a matter-antimatter symmetric cosmology (Stecker 1980). Indeed, Population III objects have been proposed (Wagoner 1969; Talbot and Arnett 1971; Negroponte, Rowan-Robinson, and Silk 1981; Tarbet and Rowan-Robinson 1982; Bond, Arnett, and Carr 1984) as the source of the observed (pregalactic) helium. Although these hypothetical objects are capable of producing large amounts of helium, the predicted yield depends on a large number of adjustable parameters such as the IMF, stellar rotation, and magnetic fields. In general, it is not possible to produce deuterium in its observed abundance in such Population III objects. It is, therefore, our view that such proposals are extremely ad hoc and have little to recommend themselves when compared with the simplicity of big-bang nucleosynthesis and the success of the standard model in fitting the observed abundances of the light elements with so few free parameters.

VI. CONSTRAINTS FROM PRIMORDIAL NUCLEOSYNTHESIS

Our analysis has used the observational data on the abundances of D, ^{3}He, and ^{7}Li to derive bounds to the nucleon abundance. For convenience, these constraints are summarized in Table 6, and all the predicted abundances are shown together in Figure 5. The nucleon contribution to the total mass density, measured by the density parameter Ω_N, is related to the nucleon abundance η by equation (1). For $\frac{1}{2} \leq h_o \leq 1$ and $2.7 \leq T_o \leq 3.0$ K,

$$0.0035\eta_{10} \leq \Omega_N \leq 0.019\eta_{10} . \quad (22)$$

From Table 6, it is a "safe bet" that the nucleon abundance is in the range $3 \lesssim \eta_{10} \lesssim 10$. For this safe-bet range, $0.011 \lesssim \Omega_N < 0.19$. If we accept the inferences based on the recent observations of ^{7}Li (Spite and Spite 1982*a*, *b*), the upper limit may be reduced to $\eta_{10} \lesssim 7$; this limit is consistent with the average deuterium abundances which suggest that $(\mathrm{D/H})_p \gtrsim 2 \times 10^{-5}$. The lower limit to η may be increased somewhat if ^{3}He proves more resilient, as recent estimates (Brunish and Truran 1984) suggest; for $[(\mathrm{D} + {}^3\mathrm{He})/\mathrm{H}]_p < 6 \times 10^{-5}$, $\eta_{10} > 4$. Our best-bet range for η, based on the current data, is $4 \lesssim \eta_{10} \lesssim 7$. For the best-bet range, $0.014 \lesssim \Omega_N \lesssim 0.14$.

It is clear that the constraints from primordial nucleosynthesis point to a low (nucleon) density universe ($\Omega_N \lesssim 0.14$–0.19). How do our bounds on Ω_N compare with other estimates of the universal mass density? Dynamical estimates rely on measurements of the mass (direct and indirect) and the light on various scales. A critical mass-to-light ratio follows from the average luminosity density. The recent deep survey of Kirshner *et al.* (1983) yields a value $(M/L)_c \approx 1200h_o$ (solar units). By comparing the mass and light on various scales, the average mass-to-light ratio is sought. This dynamical approach relates Ω_o to $\langle M/L \rangle$,

$$\Omega_o = \langle M/L \rangle / \langle M/L \rangle_c \approx (1200h_o)^{-1} \langle M/L \rangle . \quad (23)$$

For the inner, luminous parts of galaxies, $(M/L)_{\mathrm{Gal}} \approx (8\text{–}20)h_o$ (Faber and Gallagher 1979), so that $\Omega_{\mathrm{Gal}} \approx 0.007$–$0.017$. It is clear that nucleons can account for all the mass inferred from the dynamics of the luminous parts of galaxies. To see if nucleons are capable of accounting for all the mass in the universe requires a good estimate of the average mass-to-light ratio. It is expected that dynamics on the largest scales will provide the data necessary to infer $\langle M/L \rangle$ and Ω_o.

Although past studies of large-scale dynamics have tended to favor a moderately high density, more recent analyses seem to favor lower values. For example, from an application of the "cosmic virial theorem," Davis, Geller, and Huchra (1978) find $0.2 \leq \Omega_o \leq 0.7$, and Peebles (1979), using the data of Kirshner, Oemler, and Schechter (1979), obtains $0.2 \leq \Omega_o \leq 0.6$. A more recent study of clusters by Press and Davis (1982), however, permits a lower density, $\Omega_o \geq 0.07$, although they conclude that values as high as $\Omega_o \approx 0.6$ cannot be excluded. Studies of the Virgocentric flow have followed the same pattern. Davis *et al.* (1980) find $0.3 \leq \Omega_o \leq 0.5$, and Davis and Huchra (1981) claim $0.4 \leq \Omega_o \leq 0.5$, whereas Yahil, Sandage, and Tammann (1980) find $0.03 \leq \Omega_o \leq 0.2$, and Aaronson *et al.* (1982) derive $0.06 \leq \Omega_o \leq 0.13$. On the largest scale yet probed—superclusters—Ford *et al.* (1981) find, by modeling the dynamics, a preferred range $0.06 \leq \Omega_o \leq 0.16$.

As we have already noted, the nucleon contribution to the present universal mass density may be as large as $\Omega_N \approx 0.14$ (for $\eta_{10} = 7$) or even $\Omega_N \approx 0.19$ (for $\eta_{10} = 10$). Unless it is reliably established that $\Omega_o > 0.2$, nucleons alone may account for the bulk of the mass of the universe. Only for $\Omega_o > 0.2$ are massive neutrinos or other exotic particles required (Schramm and Steigman 1981).

Even so, there are other hints that nucleons may not be the whole story. The problem has to do with the growth of perturbations in a low-density universe. Assume for the moment that nucleons do dominate the current mass density, and that there are three species of neutrinos which are massless or so light as to be relativistic during the epochs to be considered. The energy density in nucleons is

$$\epsilon_N = \Omega_N \epsilon_c = 37(T_o/2.7\ \mathrm{K})^3 \eta_{10}\ \mathrm{eV\ cm^{-3}} , \quad (24)$$

and that in relativistic particles is $\epsilon_R = \epsilon_\gamma + \epsilon_\nu = \epsilon_\gamma[1 + 3 \times \frac{7}{8}(T_\nu/T_\gamma)^4] \approx 0.42(T_o/2.7\ \mathrm{K})^4\ \mathrm{eV\ cm^{-3}}$. In equation (24), $\epsilon_c = 1.1 \times 10^4 h_o^2\ \mathrm{eV\ cm^{-3}}$ is the critical energy density. The universe is "radiation dominated" for epochs whose redshift z exceeds z_{eq}, where

$$1 + z_{\mathrm{eq}} = \epsilon_N/\epsilon_R = 89(2.7\ \mathrm{K}/T_o)\eta_{10} . \quad (25)$$

Since $\eta_{10} \lesssim 7$–10, for a nucleon-dominated universe, $z_{\mathrm{eq}} \lesssim 600\text{–}900 \lesssim z_{\mathrm{rec}} \approx 10^3$. A nucleon-dominated universe is radiation dominated at recombination. After decoupling and before z_{eq}, perturbations only grow by a factor of 0(2.5) (Mészáros 1974). On the other hand, perturbations in the linear regime cease growing when $z < z_*$, where $1 + z_* = \Omega_o^{-1}$. Thus, in a nucleon-dominated universe perturbations must become nonlinear ($\delta\rho/\rho > 1$) before z_* ($z > z_*$). Here

$$1 + z_* = \Omega_N^{-1} = 283h_o^2(2.7\ \mathrm{K}/T_o)^3\eta_{10}^{-1} . \quad (26)$$

TABLE 6

SUMMARY OF CONSTRAINTS

Abundance Constraint	Bound on η
$\mathrm{D/H} \gtrsim 1 \times 10^{-5}$	$\eta_{10} \lesssim 10$
$\mathrm{D/H} \gtrsim 2 \times 10^{-5}$	$\eta_{10} \lesssim 7$
$^7\mathrm{Li/D} \lesssim 1.5 \times 10^{-5}$	$\eta_{10} \lesssim 7$
$^7\mathrm{Li/H} \approx 1 \times 10^{-10}$	$2 \lesssim \eta_{10} \lesssim 5$
$(\mathrm{D} + {}^3\mathrm{He})/\mathrm{H} \lesssim 6 \times 10^{-5}$	$\eta_{10} \gtrsim 4$
$(\mathrm{D} + {}^3\mathrm{He})/\mathrm{H} \lesssim 10 \times 10^{-5}$	$\eta_{10} \gtrsim 3$

Between z_* and z_{eq}, growth is linear [$\delta\rho/\rho \propto (1+z)^{-1}$], so that the perturbations which are to become nonlinear before z_* must be large at z_{eq}:

$$\left(\frac{\delta\rho}{\rho}\right)_{eq} \gtrsim \frac{1+z_*}{1+z_{eq}} = 3.2\left[\left(\frac{h_o}{\eta_{10}}\right)\left(\frac{2.7\text{ K}}{T_o}\right)\right]^2 \gtrsim \frac{0.65}{\eta_{10}{}^2}. \quad (27)$$

Even if we allow $\eta_{10} \approx 10(\Omega_N \lesssim 0.2)$, we infer $(\delta\rho/\rho)_{eq} \gtrsim 6.5 \times 10^{-3}$. Allowing for a factor of 0(2.5) growth between decoupling and z_{eq}, it follows that $(\delta\rho/\rho)_{dec} \gtrsim 2.6 \times 10^{-3}$. Adiabatic perturbations of this magnitude are not reflected in the microwave radiation (Wilson and Silk 1981*a*, *b*; Peebles 1981; Wilkinson 1982).

Finally, we turn to a consideration of the entropy of the universe. For three species of light neutrinos, the specific entropy density s/k is related to the number density of photons by $s/k = 7.04n_\gamma$; each additional light-neutrino species would increase s/k by $1.1n_\gamma$. Our bounds on the nucleon abundance, $3\text{–}4 \le \eta_{10} \le 7\text{–}10$, restrict the baryon number-to-specific entropy ratio, $kn_B/s \sim 0.14\eta$, to the interval

$$(4\text{–}6) \times 10^{-11} \le kn_B/s \le (10\text{–}14) \times 10^{-11}. \quad (28)$$

If the universal expansion has been adiabatic, and if baryon number is conserved (likely to be an excellent approximation for $kT \ll 10^{14}$ GeV), then the baryon-to-entropy ratio is a constant of the expansion. If the universal baryon asymmetry is produced during the very early evolution of the universe ($t \approx 10^{-34}$ s, $kT \approx 10^{14}$ GeV), then this ratio is a measure of the asymmetry which must be generated in order to account for the present ratio of nucleons to photons. Since entropy could have been produced since the epoch of baryogenesis, this ratio represents the *minimum* asymmetry which must have been produced.

Because our constraints on η and kn_B/s are, more precisely, constraints on η and kn_B/s during the epoch of nucleosynthesis, we can use the presently observed baryon-to-entropy ratio to constrain entropy production between the epoch of nucleosynthesis and the present epoch:

$$S(\text{now})/S(\text{nuc}) = [kn_B/s(\text{nuc})]/[0.14\eta(\text{now})], \quad (29)$$

where S(now), S(nuc) are, respectively, the entropy of the universe now and at the epoch of big-bang nucleosynthesis. In order to obtain a bound on the amount of entropy production since primordial nucleosynthesis, i.e., $S(\text{now})/S(\text{nuc})$, we need a *lower* bound on η(now) which does not involve primordial nucleosynthesis. Since nucleons certainly dominate the mass-to-light ratio in the solar neighborhood, $\eta(\text{now}) > 0.14 \times 10^{-10}$ (OSSTY), and a conservative conclusion is that $S(\text{now}) < (50\text{–}70)S(\text{nuc})$, since $kn_B/s(\text{nuc}) < (10\text{–}14) \times 10^{-11}$. It is most likely that the inner, luminous parts of galaxies are also dominated by nucleons, so that $\eta(\text{now}) > 0.3 \times 10^{-10}$ (OSSTY). This limit also receives support from studies of the X-ray emitting gas in many rich clusters of galaxies. With this less conservative but still reasonable limit to η(now), and with the previous limits to kn_B/s(nuc), we find $S(\text{now}) < (24\text{–}33)S(\text{nuc})$. [To summarize, the standard model can tolerate at most a factor of $O(24\text{–}70)$ entropy production since nucleosynthesis. In fact, if the mass of binaries and small groups of galaxies is dominated by baryons, implying that $\eta \gtrsim 2 \times 10^{-10}$, then only a factor of $O(5)$ increase in the entropy can be tolerated.]

We conclude this section by noting the amusing consequences of a somewhat unconventional interpretation of our results. Equation (1) may be thought of as relating the microwave temperature T_o to the nucleon density ($\Omega_N h_o{}^2$) and the nucleon-to-photon ratio η:

$$T_o = 17.7(\Omega_N h_o{}^2 \eta_{10}{}^{-1})^{1/3}\text{ K}. \quad (30)$$

For $0.01 < \Omega_N h_o{}^2 < 1$ and $3 < \eta_{10} < 10$, we would "predict" for the present temperature of the relic photon background $1.8 < T_o < 12$ K.

VII. LIMITS TO INHOMOGENEITIES IN THE EARLY UNIVERSE

We have seen that primordial nucleosynthesis in the standard model yields abundances consistent with bounds to the primordial abundances inferred from current observations. What of departures from the standard model? In this section we relax the assumption of perfect homogeneity and investigate the constraint which primordial nucleosynthesis places upon the amplitude of isothermal density perturbations: $\delta = \delta\eta/\bar{\eta}$. The effect of adiabatic perturbations has previously been analyzed by Gisler, Harrison, and Rees (1974).

To constrain δ we will utilize the abundances of D, ^{3}He, and, most importantly, ^{4}He. In general, the helium abundances averaged over many regions with differing local values of η, $\langle Y_p \rangle$, exceeds the helium abundance corresponding to the average nucleon abundance $\bar{\eta}$, $Y_p(\bar{\eta})$ (see Fig. 6). Therefore, to obtain an upper limit to δ, the minimum value of $\langle Y_p \rangle$ predicted for a given value of $\bar{\eta}$ and δ must be compared with the upper limit to $\langle Y_p \rangle$ derived from observations, and thus we shall use $\tau_{1/2} = 10.4$ minutes and $N_\nu = 3$.

Our treatment of the effects of inhomogeneities on primordial nucleosynthesis follows that of Epstein and Petrosian (1975). Define the volume distribution of nucleon abundance, $f(\eta)$, so that $f(\eta)d\eta$ is the fraction of space for

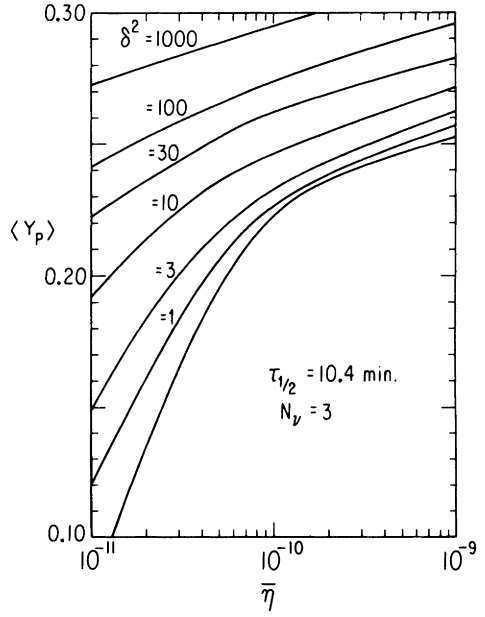


FIG. 6.—The average primordial abundance of ^{4}He (by mass), $\langle Y_p \rangle$, as a function of $\bar{\eta}$ for a range of density perturbations δ ($\equiv \delta\eta/\bar{\eta}$). Since we are interested in the *minimum* value of $\langle Y_p \rangle$, we have chosen $\tau_{1/2} = 10.4$ minutes. The results are shown for $N_\nu = 3$.

which η is in the interval $(\eta, \eta + d\eta)$. For simplicity, Epstein and Petrosian (1975) chose

$$f(\eta) \propto \eta^{a-1} \exp(-a\eta/\bar{\eta}) . \quad (31)$$

For the above choice, the perturbations are

$$\delta^2 = (\delta\eta/\bar{\eta})^2 = (\langle\eta^2\rangle - \bar{\eta}^2)/\bar{\eta}^2 = a^{-1} . \quad (32)$$

Fluctuations in η would, on sufficiently small scales, have been smoothed by nucleon diffusion. At the epoch of nucleosynthesis the diffusion scale corresponds to a (nucleonic) mass scale of $O(10^{-20} M_\odot)$. On larger scales, fluctuations will not have been erased prior to nucleosynthesis, and the synthesis of the elements would have preceeded independently in each fluctuation volume, with the various yields determined by the local value of η.

The relevant abundances are derived from observations of regions (e.g., the local interstellar medium, galactic and extragalactic H II regions, metal-poor galaxies) which, at the epoch of nucleosynthesis, contained many diffusion volumes; indeed, they contain many horizon volumes as well. As a result, the data refer to well-mixed volumes and, therefore, provide average abundances

$$\langle X_i \rangle = \int_0^\infty X_i(\eta)\eta f(\eta)d\eta \Big/ \int_0^\infty \eta f(\eta)d\eta . \quad (33)$$

With the form adopted for $f(\eta)$,

$$\langle X_i \rangle = \frac{1}{a!}\int_0^\infty X_i(a^{-1}\bar{\eta}z)z^a e^{-z}dz . \quad (34)$$

For $\eta \gtrsim 2 \times 10^{-10}$, $N_\nu = 3$, and $\tau_{1/2} = 10.4$ minutes, the predicted abundance of ^{4}He is well fitted by

$$Y_p(\eta) \approx 0.23 + 0.01 \ln \eta_{10} . \quad (35)$$

Using equation (35) in equation (34) yields

$$\langle Y_p \rangle - Y_p(\bar{\eta}) \approx 0.01 \left[\frac{d}{da}(\ln a!) - \ln a\right] . \quad (36)$$

In the limit $\delta^2 = a^{-1} \gg 1$, this reduces to

$$\langle Y_p \rangle - Y_p(\bar{\eta}) \approx 0.01 \ln \delta^2 . \quad (37)$$

For $\eta \lesssim 2 \times 10^{-10}$, equation (35) no longer provides an adequate fit to Y_p. However, for $\delta^2 \gg 1$, most of the contribution to $\langle Y_p \rangle$ is from those regions with $\eta > 2 \times 10^{-10}$ (because of the weighting factor in eq. [34]). In this case ($\eta < 2 \times 10^{-10}$, $\delta^2 \gg 1$), $\langle Y_p \rangle$ can still be approximated as

$$\langle Y_p \rangle \approx 0.23 + 0.01 \ln(\bar{\eta}_{10}\delta^2) . \quad (38)$$

The integral in equation (34) for $\langle Y_p \rangle$ has been done numerically, and the results as a function of $\bar{\eta}$ and δ^2 are shown in Figure 6.

Now consider the abundances of D and ^{3}He, keeping in mind that observations place an upper limit on the sum of the primordial abundances. Define $X_{23} \equiv$ (D + ^{3}He)/H; for $0.2 < \eta_{10} < 30$, $X_{23}(\eta)$ is well represented by

$$X_{23}(\eta) \approx 5.0 \times 10^{-4} \eta_{10}^{-1.4} . \quad (39)$$

Because of the rapid increase of X_{23} with decreasing values of η, for $\delta^2 \gg 1$, small values of η will contribute very significantly to $\langle X_{23} \rangle$. However, for $\eta_{10} < 0.2$, the dependence of X_{23} on η changes, becoming much less steep. We have numerically integrated equation (34), and the results for

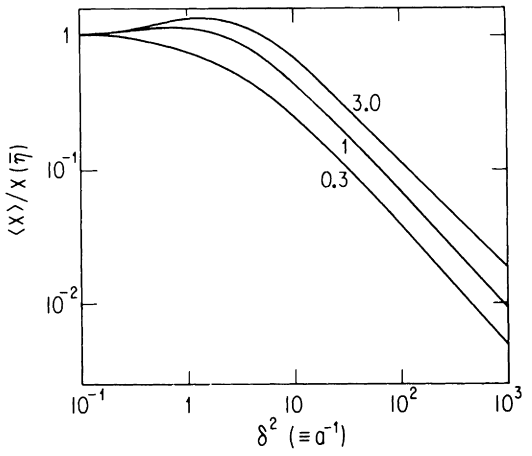


FIG. 7.—The ratio of the average abundance (by number relative to H) of D plus ^{3}He ($\langle X_{23} \rangle$) to the D plus ^{3}He abundance evaluated at the average nucleon-to-photon ratio [$X_{23}(\bar{\eta})$] as a function of the amplitude of the density perturbation δ ($\equiv \delta\eta/\bar{\eta}$). The results are shown for three values of the average nucleon-to-photon ratio: $\bar{\eta}_{10} = 0.3, 1, 3$.

$\langle X_{23} \rangle / X_{23}(\bar{\eta})$ as a function of δ^2 are shown in Figure 7 for $\bar{\eta} = 0.3, 1, 3.0$. Note that for small values of δ^2, $\langle X_{23} \rangle / X_{23}(\bar{\eta})$ is nearly independent of $\bar{\eta}$ and slightly greater than 1. In contrast, for large values of δ^2, $\langle X_{23} \rangle / X_{23}(\bar{\eta})$ is *less* than 1 and depends on $\bar{\eta}$. For $\delta^2 \gg 1$ and $0.3 \lesssim \eta_{10} \lesssim 3$, $\langle X_{23} \rangle / X_{23}(\bar{\eta}) \approx 3\bar{\eta}_{10}^{1/2}(\delta^2)^{-0.85}$

With the results presented in Figures 6 and 7, the observed abundances of D, ^{3}He, and ^{4}He can be used to constrain δ. First consider ^{4}He and the requirement that $\langle Y_p \rangle < 0.25$; of course, corresponding constraints will follow from different choices for $\langle Y_p \rangle_{max}$. To use ^{4}He alone requires that we adopt a *lower* limit to $\bar{\eta}$; from OSSTY we take $\bar{\eta} > 3 \times 10^{-11}$. Using Figure 6, $\delta^2 < 50$ follows from the requirements that $\langle Y_p \rangle < 0.25$ and $\bar{\eta} > 3 \times 10^{-11}$. If we had required $\langle Y_p \rangle < 0.24$, we would find $\delta^2 < 25$ for $\bar{\eta}_{10} > 0.3$; in contrast, for $\langle Y_p \rangle < 0.25$ and $\bar{\eta}_{10} > 0.1$, $\delta^2 < O(100)$.

Now let us consider the effect of imposing the further requirement that D + ^{3}He not be overproduced; recall that $\langle X_{23} \rangle_p \lesssim (0.6\text{–}1.0) \times 10^{-4}$. Above, only considering ^{4}He and insisting that $\langle Y_p \rangle \leq 0.25$ and $\bar{\eta}_{10} \geq 0.3$, we found that $\delta^2 \approx 50$ is allowed. For these values of $\bar{\eta}$ and δ^2 we find, from Figures 2 and 7, $\langle X_{23} \rangle \approx 1.7 \times 10^{-4}$; D + ^{3}He is overproduced! By referring to Figures 2, 6, and 7 (and Table 1), we find that the joint constraints $\langle Y_p \rangle \leq 0.25$ and $\langle X_{23} \rangle_p \leq 1 \times 10^{-4}$ can only be satisfied for $\bar{\eta}_{10} > 3$ and $\delta^2 < 3$. For $\langle Y_p \rangle \leq 0.25$ and $\langle X_{23} \rangle_p \leq 6 \times 10^{-5}$, even tighter restrictions are obtained: $\bar{\eta}_{10} > 4$, $\delta^2 < 2$.

To summarize, if we ignore the lower bound on $\bar{\eta}$ imposed by D + ^{3}He and only require $\langle Y_p \rangle < 0.25$, we find that for $\bar{\eta}_{10} > 0.1(0.3)$, $\delta < 14(7)$; for $\langle Y_p \rangle < 0.24$, these limits become $\delta < 9(5)$. More restrictive bounds are obtained when the limits from D + ^{3}He are included. For $\langle Y_p \rangle < 0.25$ and $\langle X_{23} \rangle_p < (0.6\text{–}1.0) \times 10^{-4}$, it follows that $\bar{\eta}_{10} > 3\text{–}4$ and $\delta < 1.4\text{–}1.7$. Recently, Barrow and Morgan (1983) have also used the abundances of D and ^{4}He to constrain δ; they find $\delta < O(3)$.

VIII. SUMMARY AND CONCLUSIONS

If the full potential of primordial nucleosynthesis to test the standard model and to probe the early evolution of the universe

is to be exploited, a careful and detailed comparison between theory and observation is required. A critical confrontation of the predicted abundances with the presently available data has been the goal of this paper. The standard model (isotropic and homogeneous Friedmann-Robertson-Walker cosmology; conventional particle physics with three species of light, two-component neutrinos) emerges from this comparison with flying colors. Consistent upper and lower bounds to the nucleon abundance are obtained ($\eta_{10} < 7$–10, $\eta_{10} > 3$–4), and a restrictive limit to the number of light neutrino flavors is derived: $N_\nu \lesssim 4$ ($N_\nu = 4$ is excluded if $Y_p < 0.250$). Although small-scale deviations from homogeneity cannot be excluded, the amplitude of such deviations is restricted to $\delta\eta/\eta < O(2)$.

With the possible exception of ^{7}Li, the accuracy of the predicted abundances of the light elements far exceeds that of the primordial abundances inferred from currently available data. Clearly, more—and more accurate—observational data coupled to more severe constraints on chemical evolution (astration, stellar production, etc.) would be very valuable. The conclusions of this study are no more certain than the data on which they are based. If the ranges we have used for the abundances should prove to be unduly restrictive, then our bounds, also, will be too constraining. On the other hand, unless the central values for the abundances change dramatically (by an amount larger than the currently adopted uncertainty), our confirmation of the concordance of the standard model and observations will be unaltered.

A major result of our investigation is that upper limits to the presolar abundances of deuterium and helium-3 provides an upper limit to the sum of the primordial abundances of these elements. If the fraction of ^{3}He which survives stellar processing in the material returned to the interstellar medium exceeds $\frac{1}{4}$ (see Appendix B), then $[(\mathrm{D} + {}^3\mathrm{He})/\mathrm{H}]_p < 10 \times 10^{-5}$, implying $\eta_{10} > 3$. If this fraction exceeds $\frac{1}{2}$, then $[(\mathrm{D} + {}^3\mathrm{He})/\mathrm{H}]_p < 6 \times 10^{-5}$, so that $\eta_{10} > 4$.

The presolar and interstellar estimates of the abundance of deuterium suggest that the primordial abundance exceeds $(\mathrm{D/H})_p > (1\text{–}2) \times 10^{-5}$. This leads to an upper bound to η: $\eta_{10} < 7$–10. The recent data (Spite and Spite 1982*a*, *b*) on ^{7}Li in Population II stars suggest $\eta_{10} < 7$ and are consistent with $2 < \eta_{10} < 5$.

We cannot help but notice that the upper and lower bounds on η seem to be converging. For $\eta_{10} = 5$, the predicted primordial abundances, $(\mathrm{D/H})_p = 3.6 \times 10^{-5}$, $({}^3\mathrm{He/H})_p = 1.2 \times 10^{-5}$, and $({}^7\mathrm{Li/H})_p = 1.7 \times 10^{-10}$, are remarkably consistent with the observational data. For $\eta_{10} = 5$ and $N_\nu = 3$, the predicted abundance of ^{4}He ranges from $Y_p(\tau_{1/2} = 10.4) = 0.246$ to $Y_p(\tau_{1/2} = 10.8) = 0.253$; for $N_\nu = 2$, $Y_p < 0.238$; for $N_\nu = 4$, $Y_p > 0.259$.

Nucleons alone fail to close the universe by at least a factor of 5. For $3\text{–}4 < \eta_{10} < 7\text{–}10$, the nucleon contribution to the universal mass density lies in the range $0.011\text{–}0.014 < \Omega_N < 0.14\text{–}0.19$. Although a nucleon-dominated universe is a low-density universe, nucleons can account for the mass inferred from dynamics of the luminous parts of galaxies ($\Omega_{\mathrm{Gal}} \approx 0.01$); they may also account for the dark mass inferred on larger scales.

We have come a long way in utilizing primordial nucleosynthesis to test the standard model and to provide constraints on crucial parameters in cosmology (η) and particle physics (N_ν). Further progress in testing the standard model and in constraining the allowed range of parameters relies on progress in observational and theoretical astronomy and in laboratory experiments. From the nuclear physics laboratories, reduced uncertainties in the neutron half-life and in the rates of the reactions $^4\mathrm{He}(^3\mathrm{He}, \gamma)^7\mathrm{Be}$ and $^7\mathrm{Li}(p, \alpha)^4\mathrm{He}$ would be very valuable. High-energy physics experiments may lead to interesting constraints on N_ν from the measured width of the Z^0. [The width of the Z^0 boson can be used to determine the number of neutrino species lighter than $O(10$ GeV).] From the observational side, more and better data on the abundances of the light elements as well as an improved understanding of possible evolutionary effects would be very valuable.

In the course of this investigation we have benefited from the valuable advice and constructive criticism of very many of our colleagues. We are particularly grateful to J. Audouze, K. Davidson, H. French, J. S. Gallagher, D. Gautier, L. Hobbs, R. Kirshner, E. W. Kolb, D. Kunth, A. Oemler, B. Pagel, H. Reeves, R. T. Rood, P. Shaver, J. W. Truran, R. V. Wagoner, R. E. Williams, T. Wilson, and D. G. York. We especially thank Bernard Pagel for helpful comments on an earlier version of this manuscript. This work was supported at the Bartol Research Foundation by DOE DE-AC02-78ER-05007 and at Chicago by DOE AC02-80ER-10773-A003 and NSF AST 81-16750 and at Caltech by NSF EAR 78-16873.

APPENDIX A

UNCERTAINTIES IN THE CALCULATED ABUNDANCES

If primordial nucleosynthesis is to serve as a stringent test of the standard model and as a powerful probe of the early universe, it is important to reexamine the precision of the predictions. This is particularly important if, as is anticipated, the accuracy to which the abundances of the light elements are determined improves. We have already emphasized that the predicted yields are most sensitive to the values of η, N_ν, and $\tau_{1/2}$. Here we consider their dependence on the adopted values of relevant reaction rates.

Since Wagoner (1973) last updated his code, there have been some changes in a few of the relevant reaction rates (Fowler, Caughlan, and Zimmerman 1975; Caughlan and Fowler 1980). In Table 7 we list those rates which differ from Wagoner (1973). We have incorporated these changes in the code and find relative changes of less than 1% in the primordial abundances of D, ^{3}He, and ^{4}He. In contrast, the yield of ^{7}Li has increased by approximately a factor of 3 (OSSTY). The reason for this large change is as follows: Nucleosynthesis begins in earnest when the temperature drops to $O(0.1$ MeV). First, deuterium is formed via $n(p, \gamma)d$, and then tritium by $d(d, p)t$. The tritium is converted to ^{4}He by $t(d, n)^4$He, and at the same time, small amounts of ^{7}Li are produced via $^4\mathrm{He}(t, \gamma)^7\mathrm{Li}$. Since 1973, the rate for $t(d, n)^4$He decreased by a factor of ~ 3 in the relevant temperature range. This decrease results in more t surviving for synthesis into ^{7}Li. In addition, among those reactions which destroy ^{7}Li, the rate for $^7\mathrm{Li}(\alpha, \gamma)^{11}\mathrm{B}$ decreased by $\sim 20\%$, resulting in more ^{7}Li surviving.

How sensitive are the predicted yields of primordial nucleo-

TABLE 7

NEW REACTION RATES

Reaction	Analytic Fit to the Reaction Rate[a]
$^3H(d, n)^4He$	$8.09 \times 10^{10} T^{-2/3} \exp[-4.524 T^{-1/3} - (T/0.120)^2] \times (1 + 0.092 T^{1/3} + 1.80 T^{2/3} + 1.16 T + 10.52 T^{4/3} + 17.2 T^{5/3}) + 8.73 \times 10^8 T^{-2/3} \exp(-0.523/T)$
$^3He(n, p)^3H$	$7.06 \times 10^8 (1 - 0.15 T^{1/2} + 0.098 T)$
$^3He(\alpha, \gamma)^7Be$	$5.40 \times 10^6 T^{-2/3} \exp(-12.826 T^{-1/3})(1 + 0.033 T^{1/3} - 0.350 T^{2/3} - 0.080 T + 0.056 T^{4/3} + 0.033 T^{5/3})$
$^3He(d, p)^4He$	$6.67 \times 10^{10} T^{-2/3} \exp[-7.181 T^{-1/3} - (T/0.315)^2] \times (1 + 0.058 T^{1/3} - 1.14 T^{2/3} - 0.464 T + 3.08 T^{4/3} + 3.18 T^{5/3}) + 4.36 \times 10^8 T^{-1/2} \exp(-1.72/T)$
$^4He(2\alpha, \gamma)^{12}C$	$2.79 \times 10^{-8} T^{-3} \exp(-4.403/T) + 1.35 \times 10^{-8} T^{-3/2} \exp(-24.811/T)$
$^6Li(p, \gamma)^7Be$	$5.87 \times 10^5 T^{-2/3} \exp(-8.413 T^{-1/3})$
$^6Li(n, \alpha)^3H$	$2.54 \times 10^9 T^{-3/2} \exp(-2.53/T) + 1.68 \times 10^8 [1 - 0.261(1 + 49.18 T)^{-3/2}]$
$^6Li(p, \alpha)^3He$	$3.80 \times 10^{10} T^{-2/3} (1 + 0.095 T)^{-5/6} \exp[-8.41 T^{-1/3} (1 + 0.095 T)^{1/3}] + 5.97 \times 10^9 T^{-3/2} \exp(-18.30/T)$
$^6Li(\alpha, \gamma)^{10}B$	$1.27 \times 10^7 T^{-2/3} \exp[-18.79 T^{-1/3} - (T/1.327)^2] \times (1 + 0.022 T^{1/3} + 1.00 T^{2/3} + 0.156 T + 1.39 T^{4/3} + 0.549 T^{5/3}) + 3.18 \times 10^3 T^{-3/2} \exp(-3.484/T) + 1.54 \times 10^4 T^{-1} \exp(-7.212/T)$
$^7Li(p, \alpha)^4He$	$8.04 \times 10^8 T^{-2/3} (1 + 0.049 T^{1/3} + 0.230 T^{2/3} + 0.079 T - 0.027 T^{4/3} - 0.023 T^{5/3}) \exp[-8.471 T^{-1/3} - (T/30.07)^2] + 1.07 \times 10^{10} T^{-3/2} \exp(-30.443/T) + 1.54 \times 10^6 T^{-3/2} \exp(-4.479/T)$
$^7Li(\alpha, \gamma)^{11}B$	$2.26 \times 10^8 T^{-2/3} \exp[-19.161 T^{-1/3} - (T/0.991)^2] \times (1 + 0.022 T^{1/3} + 0.382 T^{2/3} + 0.058 T + 0.571 T^{4/3} + 0.221 T^{5/3}) + 1.51 \times 10^3 T^{-3/2} \exp(-3.03/T) + 3.45 \times 10^4 T^{-1/2} \exp(-5.35/T)$
$^7Be(n, p)^7Li$	$6.76 \times 10^9 (1 - 0.903 T^{1/2} + 0.218 T)$
$^7Be(p, \gamma)^8B$	$4.09 \times 10^5 T^{-2/3} \exp(-10.262 T^{-1/3}) + 3.30 \times 10^3 T^{-3/2} \exp(-7.306/T)$
$^7Be(\alpha, \gamma)^{11}C$	$2.12 \times 10^8 T^{-2/3} \exp[-23.212 T^{-1/3} - (T/0.658)^2] \times (1 + 0.018 T^{1/3} + 0.228 T^{2/3} + 0.029 T + 0.548 T^{4/3} + 0.175 T^{5/3}) + 7.27 \times 10^4 T^{-3/2} \exp(-6.53/T) + 1.23 \times 10^5 T^{-1} \exp(-9.742/T)$
$^9Be(p, \alpha)^6Li$	$2.11 \times 10^{11} T^{-2/3} \exp[-10.36 T^{-1/3} - (T/0.520)^2] \times (1 + 0.040 T^{1/3} + 1.09 T^{2/3} + 0.307 T + 3.21 T^{4/3} + 2.30 T^{5/3}) + 4.51 \times 10^8 T^{-1} \exp(-3.046/T) + 6.70 \times 10^8 T^{-3/4} \exp(-5.160/T)$
$^9Be(p, d)2^4He$	$2.11 \times 10^{11} T^{-2/3} \exp[-10.36 T^{-1/3} - (T/0.520)^2] \times (1 + 0.040 T^{1/3} + 1.09 T^{2/3} + 0.307 T + 3.21 T^{4/3} + 2.30 T^{5/3}) + 5.79 \times 10^8 T^{-1} \exp(-3.046/T) + 8.50 \times 10^8 T^{-3/4} \exp(-5.800/T)$
$^9Be(p, \gamma)^{10}B$	$1.33 \times 10^7 T^{-2/3} (1 + 0.04 T^{1/3} + 1.52 T^{2/3} + 0.428 T + 2.15 T^{4/3} + 1.54 T^{5/3}) \exp[-10.359 T^{-1/3} - (T/0.846)^2] + 9.64 \times 10^4 T^{-3/2} \exp(-3.445/T) + 2.72 \times 10^6 T^{-3/2} \exp(-10.62/T)$
$^{10}B(p, \gamma)^{11}C$	$4.61 \times 10^5 T^{-2/3} \exp[-12.062 T^{-1/3} - (T/4.402)^2] \times (1 + 0.035 T^{1/3} + 0.426 T^{2/3} + 0.103 T + 0.281 T^{4/3} + 0.173 T^{5/3}) + 1.93 \times 10^5 T^{-3/2} \exp(-12.041/T) + 1.14 \times 10^4 T^{-3/2} \exp(-16.164/T)$
$^{10}B(p, \alpha)^7Be$	$1.26 \times 10^{11} T^{-2/3} \exp[-12.062 T^{-1/3} - (T/4.402)^2] \times (1 + 0.035 T^{1/3} - 0.498 T^{2/3} - 0.121 T + 0.300 T^{4/3} + 0.184 T^{5/3}) + 2.59 \times 10^9 T^{-1} \exp(-12.26/T)$
$^{11}B(p, \gamma)^{12}C$	$4.62 \times 10^7 T^{-2/3} (1 + 0.035 T^{1/3} + 3.00 T^{2/3} + 0.723 T + 9.91 T^{4/3} + 6.07 T^{5/3}) \exp[-12.095 T^{-1/3} - (T/0.239)^2] + 7.89 \times 10^3 T^{-3/2} \exp(-1.733/T) + 9.68 \times 10^4 T^{-1/5} \exp(-5.617/T)$
$^{11}B(p, \alpha)2^4He$	$2.59 \times 10^{11} T^{-2/3} \exp[-12.10 T^{-1/3} - (T/2.02)^2] \times (1 + 0.035 T^{1/3} + 1.22 T^{2/3} + 0.295 T + 2.15 T^{4/3} + 1.32 T^{5/3}) + 7.40 \times 10^6 T^{-3/2} \exp(-1.733/T) + 8.14 \times 10^9 T^{-3/2} \exp(-7.177/T) + 1.71 \times 10^9 T^{-2/3} \exp(-12.70/T)$

NOTE.—Those rates not listed in this table either have not changed since Wagoner 1973 or do not noticeably change our calculated abundances. These new rates are from Fowler, Caughlan, and Zimmerman 1975 and Caughlan and Fowler 1980.

[a] The rates are $N_A \langle \sigma V \rangle$ in $cm^3\ s^{-1}\ g^{-1}$; the temperature is in units of 10^9 K.

TABLE 8

THE SENSITIVITY OF THE YIELDS OF PRIMORDIAL NUCLEOSYNTHESIS TO UNCERTAINTIES (ESTIMATED 1 σ) IN NINE KEY REACTION RATES

REACTION	ADOPTED RATE / NOMINAL RATE	$\eta_{10} = 1$ X_2	$\eta_{10} = 1$ X_3	$\eta_{10} = 1$ X_7	$\eta_{10} = 10$ X_2	$\eta_{10} = 10$ X_3	$\eta_{10} = 10$ X_7
$p(n, \gamma)d$	1.1	93	101	91	99	101	109
	0.9	108	101	110	102	99	90
$d(d, n)^3\mathrm{He}$	1.1	96	103	101	95	101	105
	0.9	104	97	99	106	99	94
$d(d, p)t$	1.1	96	95	102	95	98	100
	0.9	104	105	98	105	102	100
$t(d, n)^4\mathrm{He}$	1.1	100	100	91	100	100	100
	0.9	100	100	110	100	100	100
$^3\mathrm{He}(d, p)^4\mathrm{He}$	1.1	100	100	100	100	99	100
	0.9	100	100	100	100	101	100
$^4\mathrm{He}(t, \gamma)^7\mathrm{Li}$	1.1	100	100	110	100	100	100
	0.9	100	100	90	100	100	100
$^4\mathrm{He}(^3\mathrm{He}, \gamma)^7\mathrm{Be}$	2.0	100	100	102	100	100	196
	1.29	100	100	101	100	100	128
	1.07	100	100	100	100	100	107
	0.86	100	100	99	100	100	86
	0.5	100	100	98	100	100	51
$^7\mathrm{Li}(p, \alpha)^4\mathrm{He}$	2.0	100	100	38	100	100	99
	0.5	100	100	274	100	100	101
$^7\mathrm{Be}(n, p)^7\mathrm{Li}$	1.13	100	100	100	100	100	94
	0.94	100	100	100	100	100	103
	0.75	100	100	100	100	100	115

NOTE.—The yields of D, ^{3}He, and ^{7}Li are normalized to 100 for the nominal values of the rates. For these calculations, $N_\nu = 3$ and $\tau_{1/2} = 10.6$ minutes were adopted. In all cases, Y_p changes by less than 1%.

synthesis to possible future changes in the reaction rates? In an attempt to answer this question quantitatively, we varied nine of the key reaction rates by a 1 σ uncertainty (adopted from Caughlan and Fowler 1980). For Y_p the relative change was less than 1%, and the relevant changes in D and ^{3}He were less than 10%. In contrast, changes in the predicted abundance of ^{7}Li as large as a factor of 2 were found when the rates for $^7\mathrm{Li}(p, \alpha)^4\mathrm{He}$ and $^4\mathrm{He}(^3\mathrm{He}, \gamma)^7\mathrm{Be}$ were varied by 1 σ. More recent data on this latter reaction (Harris *et al.* 1983) suggest a central rate close to the central rate we have adopted. Fowler (1983) prefers a rate 1.07 times our adopted rate and estimates the 1 σ uncertainty to be $\pm 20\%$. For completeness, we have in Table 8 shown the results for the "new" and the "old" standard rates and for the "new" and the "old" estimated uncertainties. To understand this result, note that for large values of $\eta(> 3 \times 10^{-10})$ ^{7}Li is produced mainly via $^4\mathrm{He}(^3\mathrm{He}, \gamma)^7\mathrm{Be}$ followed by ^{7}Be beta-decaying to ^{7}Li. For $\eta > 3 \times 10^{-10}$, doubling this rate (the estimated 1 σ uncertainty) doubles the yield of ^{7}Li. For smaller values of $\eta(< 3 \times 10^{-10})$, ^{7}Li is mainly produced directly by $^4\mathrm{He}(t, \gamma)^7\mathrm{Li}$. In this regime $(\eta < 3 \times 10^{-10})$, the amount of ^{7}Li which survives is most sensitive to the rate of the reaction $^7\mathrm{Li}(p, \alpha)^4\mathrm{He}$, which has a large uncertainty. As Beaudet and Reeves (1983) have noted, the ^{7}Li yield is also sensitive to the rate of the $^7\mathrm{Be}(n, p)^7\mathrm{Li}$ reaction. At high values of η the ^{7}Be converted to ^{7}Li will be destroyed and will not, therefore, be available to free decay to ^{7}Li later—when the ^{7}Li can survive. In contrast to Beaudet and Reeves (1983), who assign a factor of 2 uncertainty to this reaction, Fowler (1983) estimates an uncertainty no bigger than $\pm 20\%$ about a "new" standard value which is 6% below that in Table 7. The effect on the ^{7}Li yield of the "new" rate and its uncertainty is shown in Table 8. In Table 8 we summarize the sensitivity of the predicted abundances of D, ^{3}He, and ^{7}Li to the nine key reactions.

Recently, Dicus *et al.* (1982) (see also Cambier, Primack, and Sher 1983) have calculated the corrections to the predicted abundance of ^{4}He resulting from radiative, Coulomb, and finite temperature effects in the rates for the weak reactions: $n \leftrightarrow p + e^- + \bar{\nu}_e$, $n + e^+ \leftrightarrow p + \bar{\nu}_e$, $n + \nu_e \leftrightarrow p + e^-$. For the following range of parameters, $\eta_{10} = 0.3$–30, $N_\nu = 2$–10, $\tau_{1/2} =$ 10.1–11.1 minutes, they find a *systematic decrease* in Y_p of ~ 0.003; they also find relative changes of about 1%–2% in the abundances of D, ^{3}He, and ^{7}Li. The values of Y_p presented in this paper *include* the -0.003 correction of Dicus *et al.* (1982).

In summary, for a specific choice of values of η, N_ν, and $\tau_{1/2}$, the calculated abundance of ^{4}He should be accurate to $\sim 1\%$ $(\Delta Y \approx \pm 0.002)$, those of D and ^{3}He should be accurate to $\sim 10\%$. The predicted abundance of ^{7}Li is not more reliable than a factor of 2; for ^{7}Li, then, the accuracy of the observational data may exceed that of the predicted abundance.

APPENDIX B

THE SURVIVAL OF ^{3}He DURING GALACTIC EVOLUTION

Deuterium is burned to ^{3}He via $\mathrm{D}(p, \gamma)^3\mathrm{He}$ whenever the temperature exceeds $\sim 6 \times 10^5$ K. Since pre-main-sequence collapse is convective, sufficiently high temperatures will be reached to convert any primordial deuterium to helium-3 in the material that has been cycled through stars. Helium-3 can be destroyed by $^3\mathrm{He}(^3\mathrm{He}, ^4\mathrm{He})2p$ and $^3\mathrm{He}(\alpha, \gamma)^7\mathrm{Be}$. In addition, however, ^{3}He is produced during proton-proton burning by the sequence of reactions $p + p \rightarrow \mathrm{D} + e^+ + \nu_e$, $\mathrm{D} + p \rightarrow {}^3\mathrm{He} + \gamma$. At high temperatures, as the ^{3}He production and destruction rates drive the ^{3}He abundance toward its equilibrium value, the ^{3}He abundance will be very small. At low temperatures, however, the time to achieve equilibrium exceeds the lifetime of the star, and the ^{3}He abundance can be large. For temperatures above $\sim 7 \times 10^6$ K, ^{3}He destruction will be faster than ^{3}He production. Above $\sim 10^8$ K, ^{4}He is burned to ^{12}C and ^{16}O by the 3α reaction and by $^{12}\mathrm{C}(\alpha, \gamma)^{16}\mathrm{O}$. The following general picture emerges: Virtually all the prestellar D is burned to ^{3}He either during the pre-main-sequence collapse or in those zones of the star with temperatures which exceed $\geq 6 \times 10^5$ K (since the base of the outer convective zones of low-mass stars is at least this hot, D is burned to ^{3}He even in low-mass stars); ^{3}He will survive (and even be *produced* by *p-p* burning) in those zones where $T \lesssim 7 \times 10^6$ K; for those zones where $7 \times 10^6 \lesssim T \lesssim 10^8$ K, ^{3}He is burned to ^{4}He; for those zones where $T > 10^8$ K, ^{4}He is converted to ^{12}C and ^{16}O.

Since *some* ^{3}He will survive stellar processing, an estimate of the fraction that survives, g, will enable us to place an upper limit on the primordial abundance of D plus ^{3}He. To quantify this effect, suppose that f is the fraction of the interstellar gas, at any time t, that has never been through stars (more precisely, the fraction that has not been at temperatures hotter than 6×10^5 K).

Then, the primordial abundance of D is related to that at time t by

$$(\mathrm{D/H})_t = f(\mathrm{D/H})_p \,. \tag{B1}$$

Since some ^{3}He survives stellar processing, the abundance of ^{3}He at time t is

$$(^3\mathrm{He/H})_t \geq f(^3\mathrm{He/H})_p + g(1-f)[(\mathrm{D} + {}^3\mathrm{He})/\mathrm{H}]_p \,. \tag{B2}$$

The inequality in equation (B2) arises from the neglect of any newly synthesized ^{3}He from p-p burning. Equations (B1) and (B2) may be combined to yield

$$\left(\frac{\mathrm{D} + {}^3\mathrm{He}}{\mathrm{H}}\right)_p \leq \left(\frac{\mathrm{D}}{\mathrm{H}}\right)_t + \frac{1}{g(^3\mathrm{He/H})_t} - \left(\frac{1-g}{g}\right) f \left(\frac{^3\mathrm{He}}{\mathrm{H}}\right)_p \,. \tag{B3}$$

Since the last term on the right-hand side of equation (B3) is negative (and, in general, small), we may reinforce the inequality by neglecting it,

$$\left(\frac{\mathrm{D} + {}^3\mathrm{He}}{\mathrm{H}}\right)_p \leq \left(\frac{\mathrm{D}}{\mathrm{H}}\right)_t + \frac{1}{g(^3\mathrm{He/H})_t} \,. \tag{B3'}$$

Although the constraint on the primordial abundances in equation (B3′) is weaker than that in equation (B3), equation (B3′) permits us to avoid specifying the model-dependent quantity f. An estimate of (or, lower limit to) g permits us to bound from above the primordial abundance of D plus ^{3}He. In constraining g from below, we will proceed from more general to more model (of galactic chemical evolution) dependent results.

The models of Iben (1967*a*, *b*) and Rood (1972) show that low-mass stars ($M < 2\,M_\odot$) are net producers of ^{3}He. Of the material these stars return to the interstellar medium, more has been processed at low temperatures ($< 7 \times 10^6$ K) than at high temperatures. These results were used by Rood, Steigman, and Tinsley (1976) to predict that the ^{3}He abundance should have increased since the formation of the solar system. Detections of $^3\mathrm{He}^+$ by Wilson, Rood, and Bania (1983) in several H II regions seem to support this prediction and provide some confidence in the models on which it was based. For stars less massive than 8 $M_\odot$, Iben and Truran (1978) find a relationship between the abundance of ^{3}He returned to the interstellar medium and the D plus ^{3}He abundance in the gas from which the star formed (their eq. [18]),

$$(^3\mathrm{He/H})_f = 0.7[(\mathrm{D} + {}^3\mathrm{He})/\mathrm{H}]_i + 18 \times 10^{-5} M^{-2} \,. \tag{B4}$$

In equation (B4), M is the mass of the star ($M < 8$) in units of the solar mass. Thus, for low-mass stars, we may conclude that $g \gtrsim 0.7$.

Now let us turn to high-mass stars ($8 < M < 100$). The massive star models (with mass loss) computed by Dearborn *et al.* (1978) were examined to determine the survival of ^{3}He. According to Dearborn (1983), for stars in this mass range, more than one-fourth (by mass) of the star is in a zone which has never been hotter than 7×10^6 K. Using the Snow and Morton (1976) mass loss rates, at least this much mass is lost prior to the traverse of the supernova shock wave. For massive stars then, we conclude that $g \gtrsim \frac{1}{4}$. This conclusion is supported by the recent calculations (without mass loss) of Brunish and Truran (1984), who find for $10 < M < 50$, $g \approx 0.25$.

Since we are ignorant of the IMF for the earliest generations of stars (Population III), our first estimate of the survival of ^{3}He should be on a star-by-star basis. For "normal" stars with $M < 100$, the model calculations discussed above suggest that $g > \frac{1}{4}$. Unless exotic stars are postulated for early galactic evolution, we consider this constraint to be conservative.

Most IMFs give greater weight to low-mass stars in the contribution to material returned to the interstellar medium. For a Salpeter (1955) IMF, Brunish and Truran (1984) find for a single generation of stars ($0.3 < M < 50$)

$$(^3\mathrm{He/H})_f \approx 1 \times 10^{-5} + 0.5[(\mathrm{D} + {}^3\mathrm{He})/\mathrm{H}]_i \,. \tag{B5}$$

If most of the interstellar material has, on average, been cycled through at most one generation of stars (i.e., the fraction which has been through several generations is small), then it is likely that $g \gtrsim \frac{1}{2}$. This conclusion receives some support when we consider the buildup of ^{4}He and the heavier elements in the course of galactic chemical evolution.

Of the material returned by stars, some has come from regions which are sufficiently cool that ^{3}He survives; in the rest of the material, ^{4}He and heavier elements have been produced. Since observations constrain the amount of ^{4}He and heavier nuclei synthesized in the course of galactic evolution, we may derive a limit to the destruction of ^{3}He. For the fraction by mass of ^{4}He (heavier nuclei) produced from the matter which has been cycled through stars, write $\Delta Y_*(\Delta Z_*)$. The abundances at time t are

$$Y_t = Y_p + (1-f)\Delta Y_* \,; \qquad Z_t = (1-f)\Delta Z_* \,. \tag{B6}$$

The amount of stellar-produced ^{4}He is $\Delta Y_t = Y_t - Y_p = (1-f)\Delta Y_* = Z_t(\Delta Y_*/\Delta Z_*)$. Note that ΔZ_* and ΔY_* could be large (e.g., $\Delta Y_* + \Delta Z_* \approx 1$) but ΔY_t and Z_t small if $1 - f$, the fraction of interstellar matter which has been cycled through stars, is small ($f \approx 1$). Now, recall that g is the fraction of the matter cycled through stars in which ^{3}He survives so that

$$g \approx 1 - (\Delta Y_* + \Delta Z_*) = 1 - (1-f)^{-1}(\Delta Y_t + Z_t) \,. \tag{B7}$$

In the above it has been assumed that where ^{3}He has been destroyed, all the material has been burned completely to ^{4}He and/or heavier nuclei. Indeed, stellar models indicate that stars spend sufficient time on the main sequence that those regions which are

hot enough to burn away the ^{3}He will convert any initial hydrogen to ^{4}He and beyond. However, to allow for the possibility of incomplete burning (i.e., only partial burning to ^{4}He and beyond in those zones where ^{3}He has been destroyed) we write

$$g \approx 1 - (1 + \alpha)(\Delta Y_* + \Delta Z_*) = 1 - (1 - f)^{-1}(1 + \alpha)(\Delta Y_t + Z_t) . \quad \text{(B8)}$$

Equation (B8) may be rewritten to relate the unknown quantities f and g to the potentially observable quantities ΔY_t and Z_t:

$$f + (1 - f)g = 1 - (1 + \alpha)(\Delta Y_t + Z_t) . \quad \text{(B9)}$$

If we take the sum of equations (B1) and (B2), we find

$$\left(\frac{\mathrm{D} + {}^3\mathrm{He}}{\mathrm{H}}\right)_t \geq [f + (1 - f)g]\left(\frac{\mathrm{D} + {}^3\mathrm{He}}{\mathrm{H}}\right)_p , \quad \text{(B10)}$$

so that

$$\left(\frac{\mathrm{D} + {}^3\mathrm{He}}{\mathrm{H}}\right)_t \geq [1 - (1 + \alpha)(\Delta Y_t + Z_t)]\left(\frac{\mathrm{D} + {}^3\mathrm{He}}{\mathrm{H}}\right)_p . \quad \text{(B11)}$$

At present or at the time of the formation of the solar system, it is likely that $\Delta Y + Z \approx 0.05$–$0.1$. As a result, we expect (unless α is large) that the sum of the primordial abundances of D and ^{3}He is close to the present (or presolar) sum of abundances.

To summarize, some ^{3}He survives the evolution of all standard stars ($M < 100$). Even for a pregalactic generation of stars, star by star (independent of the IMF), we expect that $g > \frac{1}{4}$. Low-mass stars preserve more of their initial ^{3}He and may even be net producers of ^{3}He. For most IMFs, which give significant weight to the low-mass end, we expect that $g > \frac{1}{2}$. Indeed, even more ^{3}He might survive than these estimates suggest. Since ^{4}He and heavier nuclei are produced in those zones of a star where ^{3}He is destroyed, the present abundances of ^{4}He and the heavier nuclei may constrain even more severely the destruction of ^{3}He. For $\Delta Y_0 + Z_0 < 0.1$ and $\alpha < 1$, more than 80% of the primordial D plus ^{3}He should be present today.

REFERENCES

Aaronson, M., Huchra, J., Mould, J., Schechter, P. L., and Tully, R. B. 1982, *Ap. J.*, **258**, 64.
Anders, E., Heymann, D., and Mazor, E. 1970, *Geochim. Cosmochim. Acta*, **34**, 127.
Arnett, W. D. 1978, *Ap. J.*, **219**, 1008.
Audouze, J., and Tinsley, B. M. 1976, *Ann. Rev. Astr. Ap.*, **14**, 43.
Bahcall, J. N., Huebner, W. F., Lubow, S. H., Parker, P. D., and Ulrich, R. K. 1982, *Rev. Mod. Phys.*, **54**, 767.
Barrow, J. D. 1976, *M.N.R.A.S.*, **175**, 359.
Barrow, J. D., and Morgan, J. 1983, *M.N.R.A.S.*, **203**, 393.
Beaudet, G., and Goret, P. 1976, *Astr. Ap.*, **49**, 415.
Beaudet, G., and Reeves, H. 1983, in *Proceedings of the ESO Workshop on Primordial Helium*, ed. P. A. Shaver, D. Kunth, and K. Kjär (Garching: European Southern Observatory), p. 53.
Black, D. C. 1971, *Nature Phys. Sci.*, **234**, 148.
———. 1972, *Geochim. Cosmochim. Acta*, **36**, 347.
Bond, J. R., Arnett, W. D., and Carr, B. J. 1984, *Ap. J.*, in press.
Brunish, W., and Truran, J. W. 1984, in preparation.
Bruston, P., Audouze, J., Vidal-Madjar, A., and Laurent, C. 1981, *Ap. J.*, **243**, 161.
Byrne, J. 1982, paper presented at conference, The Neutron and Its Applications, Cambridge.
Byrne, J., Morse, J., Smith, K. F., Shaikh, F., Green, K., and Greene, F. L. 1980 *Phys. Letters*, **92B**, 274.
Cambier, J.-L., Primack, J. R., and Sher, M. 1983, *Nucl. Phys.*, **B209**, 372.
Caughlan, G. R., and Fowler, W. A. 1980, private communication.
Chiosi, C. 1979, *Astr. Ap.*, **80**, 252.
Chiosi, C., and Caimmi, R. 1979, *Astr. Ap.*, **80**, 234.
Christensen, C. J., Nielsen, A., Bahnsen, A., Brown, W. K., and Rustad, B. M. 1972, *Phys. Rev.*, **D5**, 1628.
Christensen-Dalsgaard, J., and Gough, D. O. 1980, *Nature*, **288**, 544.
Combes, M., and Encrenaz, T. 1979, *Icarus*, **39**, 1.
David, Y., and Reeves, H. 1980, in *Physical Cosmology*, ed. R. Balian, J. Audouze, and D. N. Schramm, (Amsterdam: North-Holland), p. 443.
Davis, M., Geller, M. J., and Huchra, J. 1978, *Ap. J.*, **221**, 1.
Davis, M., and Huchra, J. 1981, *Ap. J.*, **254**, 437.
Davis, M., Tonry, J., Huchra, J., and Latham, D. W. 1980, *Ap. J.* (*Letters*), **238**, L113.
Dearborn, D. S. P. 1983, private communication.
Dearborn, D. S. P., Blake, J. B., Hainebach, K. L., and Schramm, D. N. 1978, *Ap. J.*, **223**, 552.
Dearborn, D. S. P., and Trimble, V. 1978, in *IAU Symposium 80, The HR Diagram*, ed. A. G. D. Philip and D. S. Hayes, (Dordrecht: Reidel), p. 375.
Dicus, D. A., Kolb, E. W., Gleeson, A. M., Sudarshan, E. C. G., Teplitz, V. L., and Turner, M. S. 1982, *Phys. Rev.*, **D26**, 2694.
Dufour, R. J. 1975, *Ap. J.*, **195**, 315.
Dufour, R. J., and Harlow, W. V. 1977, *Ap. J.*, **216**, 706.
Duncan, D. K. 1981, *Ap. J.*, **248**, 651.
Encrenaz, T., and Combes, M. 1982, *Icarus*, **52**, 54.
Epstein, R., Lattimer, J., and Schramm, D. N. 1976, *Nature*, **263**, 198.
Epstein, R. I., and Petrosian, V. 1975, *Ap. J.*, **197**, 281.
Faber, S. M., and Gallagher, J. S. 1979, *Ann. Rev. Astr. Ap.*, **17**, 135.
Ford, H. C., Harms, R. J., Ciardullo, R., and Bartok, F. 1981, *Ap. J.* (*Letters*), **245**, L53.
Fowler, W. A. 1983, private communication.
Fowler, W. A., Caughlan, G. R., and Zimmerman, B. A. 1975, *Ann. Rev. Astr. Ap.*, **13**, 69.
French, H. B. 1980, *Ap. J.*, **240**, 41.
———. 1981, private communication to G. S. and M. S. T.
Gautier, D., and Owen, T. 1983, *Nature*, **302**, 215.
Geiss, J., and Reeves, H. 1972, *Astr. Ap.*, **18**, 126.
Gisler, G. R., Harrison, E. R., and Rees, M. J. 1974, *M.N.R.A.S.*, **166**, 663.
Gott, J. R., Gunn, J. E., Schramm, D. N., and Tinsley, B. M. 1974, *Ap. J.*, **194**, 543.
Hall, D. N. B. 1975, *Ap. J.*, **197**, 509.
Harris, M. J., Fowler, W. A., Caughlan, G. R., and Zimmerman, B. A. 1983, *Ann. Rev. Astr. Ap.*, **21**, 165.
Hawking, S. W., and Tayler, R. J. 1966, *Nature*, **209**, 1278.
Hawley, S. A. 1978, *Ap. J.*, **224**, 417.
Heasley, J. N., and Milkey, R. W. 1978, *Ap. J.*, **243**, 127.
Iben, I. 1967*a*, *Ap. J.*, **147**, 624.
———. 1967*b*, *Ap. J.*, **147**, 650.
Iben, I., and Truran, J. W. 1978, *Ap. J.*, **220**, 980.
Kinman, T. D., and Davidson, K. 1981, *Ap. J.*, **243**, 127.
Kirshner, R. P., Oemler, A., and Schechter, P. L. 1979, *A.J.*, **84**, 951.
Kirshner, R. P., Oemler, A., Schechter, P. L., and Shectman, S. 1983, *A.J.*, in press.
Kolb, E. W., and Scherrer, R. J. 1982, *Phys. Rev.*, **D25**, 1481.
Kugler, K. J., Paul, W., and Trinks, U. 1978, *Phys. Letters*, **72B**, 422.
———. 1979, *IEEE Trans.*, **26**, 3152.
Kunde, V., *et al.* 1982, *Ap. J.*, **263**, 443.
Kunth, D. 1982, Ph.D. thesis, Université Paris VII.
Kunth, D., and Sargent, W. L. W. 1983, *Ap. J.*, **273**, 81.
Laurent, C. 1983, in *Proceedings of the ESO Workshop on Primordial Helium*, ed. P. A. Shaver, D. Kunth, and K. Kjär, (Garching: European Southern Observatory), p. 335.
Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., and Torres-Peimbert, S. 1979, *Astr. Ap.*, **80**, 155.
Maeder, A. 1980, *Astr. Ap.*, **92**, 101.
Mason, B. 1979, "Cosmochemistry, Part I, Meteorites," chap. B, in *Data of Geochemistry*, ed. M. Fleischer (6th ed.: Washington, D.C.: US Government Printing Office).
Matzner, R., and Rothman, T. 1982, *Phys. Rev. Letters*, **48**, 1565.
McKellar, A. R. W. Goetz, W., and Ramsey, D. A. 1976, *Ap. J.*, **207**, 663.
Mészáros, P. 1974, *Astr. Ap.*, **37**, 225.
Negroponte, J., Rowan-Robinson, M., and Silk, J. 1981, *Ap. J.*, **248**, 38.
Olive, K. A., Schramm, D. N., and Steigman, G. 1981, *Nucl. Phys.*, **B180**, 497.
Olive, K. A., Schramm, D. N., Steigman, G., Turner, M. S., and Yang, J. 1981, *Ap. J.*, **246**, 557 (OSSTY).
Olive, K. A., and Turner, M. S. 1981, *Phys. Rev. Letters*, **46**, 516.

Pagel, B. E. J. 1982, *Phil. Trans. Roy. Soc. London, A*, **307**, 19.
Peebles, P. J. E. 1966, *Ap. J.*, **146**, 542.
———. 1979, *A.J.*, **84**, 730.
———. 1981, *Ap. J. (Letters)*, **243**, L119.
Peimbert, M., and Torres-Peimbert, S. 1974, *Ap. J.*, **193**, 327 (PTP I).
———. 1976, *Ap. J.*, **203**, 581 (PTP II).
———. 1977, *M.N.R.A.S.*, **179**, 217 (PTP III).
Peimbert, M., Torres-Peimbert, S., and Rayo, J. F. 1978, *Ap. J.*, **220**, 516.
Press, W. A., and Davis, M. 1982, *Ap. J.*, **259**, 449.
Rana, N. 1982, *Phys. Rev. Letters*, **48**, 209.
Rayo, J., Peimbert, M., and Torres-Peimbert, S. 1982, *Ap. J.*, **255**, 1 (RPTP).
Rees, M. J. 1978, *Nature*, **275**, 35.
Reeves, H., Audouze, J., Fowler, W. A., and Schramm, D. N. 1973, *Ap. J.*, **179**, 909.
Rood, R. T. 1972, *Ap. J.*, **177**, 681.
Rood, R. T., Steigman, G., and Tinsley, B. M. 1976, *Ap. J. (Letters)*, **207**, L57.
Salpeter, E. E. 1955, *Ap. J.*, **121**, 161.
Schramm, D. N., and Steigman, G. 1981, *Ap. J.*, **243**, 1.
Schramm, D. N., and Wagoner, R. V. 1977, *Ann. Rev. Nucl. Part. Sci.*, **27**, 37.
Shaver, P. A., McGee, M. X., Newton, L. M., Danks, A. C., and Pottasch, S. R. 1983, *M.N.R.A.S.*, **204**, 53.
Shvartsman, V. F. 1969, *Soviet Phys.—JETP Letters*, **9**, 184.
Smith, H. E. 1975, *Ap. J.*, **199**, 591.
Snow, T. P., and Morton, D. C. 1976, *Ap. J. Suppl.*, **32**, 429.
Spite, F., and Spite, M. 1982*a*, *Astr. Ap.*, **115**, 357.
———. 1982*b*, *Nature*, **297**, 483.
Stasinska, G. 1980, *Astr. Ap.*, **84**, 320.
Stecker, F. W. 1980, *Phys. Rev. Letters*, **44**, 1237.
———. 1981, *Phys. Rev. Letters*, **46**, 517.
Steigman, G., Olive, K. A., and Schramm, D. N. 1979, *Phys. Rev. Letters*, **43**, 239.
Steigman, G., Schramm, D. N., and Gunn, J. E. 1977, *Phys. Letters*, **B66**, 202.
Szalay, A. S. 1981, *Phys. Letters*, **B101**, 453.
Talbot, R. J., and Arnett, W. D. 1971, *Nature Phys. Sci.*, **229**, 150.
Talent, D. L., and Dufour, R. J. 1979, *Ap. J.*, **233**, 888.
Tarbet, P. W., and Rowan-Robinson, M. 1982, *Nature*, **298**, 711.
Thorne, K. S. 1967, *Ap. J.*, **148**, 51.
Trauger, J. T., Roesler, F. L., Carleton, N. P., and Traub, W. A. 1973, *Ap. J. (Letters)*, **184**, L137.
Trauger, J. T., Roesler, F. L., and Mickelson, M. E. 1977, *Bull. AAS*, **9**, 516.
Truran, J. W., and Cameron, A. G. W. 1971, *Ap. Space Sci.*, **5**, 117.
Tully, R. B., Boesgaard, A. M., Dyck, H. M., and Schempp, W. V. 1981, *Ap. J.*, **246**, 38.
Vanden Bout, P. A., Snell, R. L., Vogt, S. S., and Tull, R. G. 1978, *Ap. J.*, **221**, 598.
Vidal-Madjar, A., Laurent, C., Bonnet, R. M., and York, D. G. 1977, *Ap. J.*, **211**, 91.
Vidal-Madjar, A., Laurent, C., Gry, C., Bruston, P., Ferlet, R., and York, D. G. 1983, *Astr. Ap.*, **120**, 58.
Wagoner, R. V. 1969, *Ap. J. Suppl.*, **18** (No. **162**), 247.
———. 1973, *Ap. J.*, **179**, 343.
Wagoner, R. V., Fowler, W. A., and Hoyle, F. 1967, *Ap. J.*, **148**, 3.
Wilkinson, D. H. 1979, *Progr. Part. Nucl. Phys.*, **6**, 328.
Wilkinson, D. T. 1982, in *The Birth of the Universe*, ed. J. Audouze and J. Tranh Van, (Editions Frontières), p. 159.
Wilson, M., and Silk, J. 1981*a*, *Ap. J.*, **243**, 14.
———. 1981*b*, *Ap. J. (Letters)*, **244**, L37.
Wilson, T. L., Rood, R. T., and Bania, T. M. 1983, in *Proceedings of the ESO Workshop on Primordial Helium*, ed. P. A. Shaver, D. Kunth, and K. Kjär, (Garching: European Southern Observatory), p. 379.
Yahil, A., and Beaudet, G. 1976, *Ap. J.*, **206**, 26.
Yahil, A., Sandage, A., and Tammann, G. A. 1980, in *Physical Cosmology*, ed. R. Balian, J. Audouze, and D. N. Schramm, (Amsterdam: North-Holland), p. 127.
Yang, J., Schramm, D. N., Steigman, G., and Rood, R. T. 1979, *Ap. J.*, **227**, 697 (YSSR).
York, D. G. 1983, private communication.
York, D. G., and Rogerson, J. B., Jr. 1976, *Ap. J.*, **203**, 378.
Zappala, R. A. 1972, *Ap. J.*, **172**, 57.

K. A. OLIVE: Fermilab Astrophysics Group, Fermilab, P.O. Box 500, Batavia, IL 60510

D. N. SCHRAMM and M. S. TURNER: Astronomy and Astrophysics Center, University of Chicago, Chicago, IL 60637

G. STEIGMAN: Bartol Research Foundation, University of Delaware, Newark, DE 19716

J. YANG: Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125

REST MASS OF MUONIC NEUTRINO AND COSMOLOGY

S. S. Gershtein and Ya. B. Zel'dovich
Submitted 4 June 1966
ZhETF Pis'ma 4, No. 5, 174-177, 1 September 1966

Low-accuracy experimental estimates of the rest mass of the neutrino [1] yield $m(\nu_e) < 200$ eV/c^2 for the electronic neutrino and $m(\nu_\mu) < 2.5 \times 10^6$ eV/c^2 for the muonic neutrino.

Cosmological considerations connected with the hot model of the Universe [2] make it possible to strengthen greatly the second inequality. Just as in the paper by Ya. B. Zel'dovich and Ya. A. Smorodinskii [3], let us consider the gravitational effect of the neutrinos on the dynamics of the expanding Universe. The age of the known astronomical objects is not smaller than 5×10^9 years, and Hubble's constant H is not smaller than 75 km/sec-Mparsec $= (13 \times 10^9 \text{ years})^{-1}$. It follows therefore that the density of all types of matter in the Universe is at the present time [1)]

$$\rho < 2 \times 10^{-28} \text{ g/cm}^3.$$

The space surrounding us is filled presently with equilibrium radiation of temperature 3°K [4]. It is proposed that this is "relict" radiation and is proof of the high temperature possessed by the plasma during the pre-stellar high-density period.

At a temperature of the order of 3 MeV for ν_e and of the order of 15 MeV for ν_μ, complete thermodynamic equilibrium existed between ν, γ, e^+, and e^-. The number of other particles in this equilibrium is small, except perhaps gravitons, which, however, have no effect on the arguments that follow. In thermodynamic equilibrium, the ratio of the number of fermions and antifermions with spin 1/2 to the number of quanta is

$$[\nu_e] + [\bar{\nu}_e] = [\nu_\mu] + [\bar{\nu}_\mu] = [e^+] + [e^-] = 2\,\frac{\int (e^x + 1)^{-1} x^2 dx}{\int (e^x - 1)^{-1} x^2 dx}[\gamma] = 1.5[\gamma].$$

However, during the course of the cooling from $T > m_e c^2$ (for which these relations are written) to the present time, when $T \ll m_e c^2$, these relations change, since the annihilation

of the e^+e^- increases the number of quanta without changing the number of neutrinos per unit of co-moving volume [5]. At the present time we can expect

$$[e^+] + [e^-] = 0, \quad [\nu_\mu] + [\bar{\nu}_\mu] = [\nu_e] + [\bar{\nu}_e] = 0.5[\gamma].$$

At 3°K we have $[\gamma] = 550$ g/cm^3, from which we obtain for the neutrino at the present time

$$[\nu_\mu] + [\bar{\nu}_\mu] = [\nu_e] + [\bar{\nu}_e] = 300 \text{ cm}^{-3}.$$

Comparing with the density limit given above, we obtain

$$m_0(\nu_\mu) < 7 \times 10^{-31} \text{ g} = 400 \text{ eV}/c^2$$

and the same for $m_0(\nu_e)$. Thus, we obtain no new information for the electronic neutrino; for the muonic neutrino, on the other hand, the cosmological considerations reduce the upper limit of the rest mass by three orders of magnitude.

In considering the question of the possible mass of the neutrino we have, naturally, used statistical formulas for the four-component ($m \neq 0$) particles. We know, however, that in accordance with the (V - A) theory, neutrinos having a definite polarization participate predominantly in weak interactions. Equilibrium for neutrinos for opposite polarization is established only at a higher temperature. This, incidentally, can change the limit of the mass by not more than a factor of 2.

A neutrino with non-zero rest mass can become annihilated in accordance with the diagram

$$\nu_\mu + \bar{\nu}_\mu \xrightarrow[\text{(weak)}]{} \mu^+ + \mu^- \xrightarrow[\text{(el.-mag.)}]{} \gamma + \gamma$$

if $m(\nu_\mu) > m(e^\pm)$, and also either into 3γ or, in the square of the weak interaction, into a $\nu_e + \bar{\nu}_e$ pair, if it is assumed the $m(\nu_\mu) > m(\nu_e)$. When $v < c$ the annihilation cross section behaves like $1/v$. Estimates show, however, that there is no time for noticeable annihilation to take place during the course of the cosmological expansion.

The momentum of the interaction particles changes during the course of expansion like $1/R$, where R is the linear scale, independently of the presence and magnitude of the particle rest mass. At the present time the neutrino momentum should be of the same order (somewhat smaller) as the momentum of the relict quanta, i.e., $\bar{p} \approx 5 \times 10^{-4}$ eV/c.

If the neutrinos have a rest mass, then their velocity and the speed of sound in the neutrino gas are of the order of p/m, i.e, say 30 km/sec at $m = 5$ eV/c^2 and 3 km/sec at $m = 50$ eV/c^2. Strong gravitational perturbations should be produced in such a gas by the galaxies. It is possible that a more detailed analysis of these processes will allow us to lower the foregoing estimate of the upper limit of the neutrino mass.

This note is the result of the stimulating circumstances of the Summer School in Balaton-vilagose, and we use this opportunity to express gratitutde to the organizers of the school.

[1] A. H. Rosenfeld, A. Barbaro Galtieri, W. H. Barkas, P. L. Bastien, I. Kirz, and M. Roos, Tables, Revs. Mod. Phys. 37, 633 (1965).

[2] G. Gamow, Phys. Rev. 70, 572 (1946); 74, 505 (1948); Revs. Mod. Phys. 21, 367 (1949); G. Gamow, Vistas in Astronomy 2, 1726 (1956); R. Dicke, P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson, Astrophys. J. 142, 414 (1965); Ya. B. Zel'dovich, UFN 89, 647 (1966), Soviet Phys. Uspekhi 9, in press.

[3] Ya. B. Zel'dovich and Ya. A. Smorodinskii, JETP 41, 907 (1961), Soviet Phys. JETP 14, 647 (1962).

[4] A. A. Penzias and R. W. Wilson, Astrophys. J. 142, 419 (1965).

[5] P. J. E. Peebles, Phys. Rev. Lett. 16, 410 (1966).

1) We use the asymptotic formula

$$T = \pi/2H\sqrt{\rho/\rho_c} \; ; \quad \rho_c = 3H^2/8\pi\sigma \; ; \quad \rho = 3\pi/32\sigma T^2.$$

Other more complicated estimates based on an investigation of remote objects give a similar result:

$$q_0 = \rho/2\rho_c < 2.5; \quad H \leq 120 \text{ km/sec.Mparsec},$$

$$\rho_c \leq 2.5 \times 10^{-29} \text{ g/cm}^3, \quad \rho < 1.25 \times 10^{-28} \text{ g/cm}^3.$$

Massive Neutrinos and the Large-Scale Structure of the Universe

J. R. Bond, G. Efstathiou, and J. Silk
Astronomy Department, University of California, Berkeley, California 94720
(Received 9 July 1980)

If neutrinos dominate the mass density of the Universe, they play a critical role in the gravitational instability theory of galaxy formation. For neutrinos of mass $30m_{30}$ eV, a maximum Jeans mass $M_{\nu m} \approx 4\times 10^{15} m_{30}^{-2} M_{\odot}$ is derived. On smaller scales, neutrino fluctuations are damped, and the growth of baryon fluctuations is greatly inhibited. Structures on scales $> M_{\nu m}$ may collapse, forming galaxies as in Zel'dovich's "pancake" theory.

PACS numbers: 98.50.Eb, 14.60.Gh, 95.30.Cq, 98.80.Ft

The idea that massive neutrinos may solve the missing-mass problem in galaxies and in clusters of galaxies, as well as dominating the mass content of the Universe, has been explored by many authors.[1,2] The recent experimental reports of electron-neutrino mass detection[3] and neutrino oscillations,[4] coupled with the natural appearance of a neutrino-flavor mass spectrum with splittings of order of those between the three lepto-quark families in extensions of grand unified theories to SO(10),[5] suggest that we take seriously the astrophysical implications of a neutrino

© 1980 The American Physical Society

mass. Here, we consider the role of light massive neutrinos in the early Universe and in galaxy formation.

The mean occupation number of the neutrino background in the standard big-bang model as a function of momentum p and time t is $n(p,t) = \{\exp[pc/T_\nu(t)]+1\}^{-1}$, which is valid even in the nonrelativistic (NR) regime, where it is neither Fermi-Dirac nor degenerate. The neutrino "temperature," $T_\nu(t)$, for left-handed neutrinos (ν_L) is now 1.9 K, red shifted from 1 MeV when they decoupled. If right-handed neutrinos (ν_R) couple only as in the Weinberg-Salam model extended to include a neutrino Dirac mass, rates for reactions such as $e^+e^- \to \nu_R\bar{\nu}_L$ [$\sim(m_\nu/T_\nu)^2$ times the $e^+e^- \to \nu_L\bar{\nu}_L$ rate] were always substantially less than the expansion rate.[6] The resulting number ratio of right-handed to left-handed neutrinos is $g-1 \lesssim 10^{-4}$, valid for all ν flavors. Subsequently, however, left and right helicities can mix through deflections in a gravitational field, though spin projection along a fixed axis is conserved.

We then obtain a neutrino plus antineutrino mass density in the NR epoch (in units of the cosmological closure density) of $\Omega_\nu = 0.93\bar{m}_{30}h^{-2}$, where h is Hubble's constant in units of 100 km s^{-1} Mpc^{-1}. (Mpc is 10^6 parsec.) If the average neutrino mass $\bar{m}_{30} = \frac{1}{3}(\sum g m_{30})$ (where the sum is over ν flavors) exceeds $\sim h^2$, the Universe is closed. Here m_{30} is the neutrino mass in 30-eV units. Observations of the deceleration parameter[7] suggest $\Omega_\nu \lesssim 2$, which slightly modifies the well-known constraint on light-neutrino masses,[1] $\bar{m}_{30} \lesssim 2h^2$.

Recent estimates[8] of the mean luminosity density in the Universe imply baryons in galaxies contribute a fraction $\Omega_B = (M/L)_G/2315h \sim 0.004$–$0.007$ to the mean density required for closure. The mass-to-blue-light ratio of a typical galaxy, $(M/L)_G$ (in solar units), due to "conventional" stars probably lies in the range $9h$–$16h$,[9] yielding the above spread. This gives the baryon density in luminous forms of matter, which is smaller than the mass density in neutrinos by the factor $\Omega_\nu/\Omega_B \sim (140$–$240)\bar{m}_{30}h^{-2}$. Neutrinos need only have $\bar{m} \gtrsim 0.2h^2$ eV to dominate, unless most baryons reside in dark forms of matter, such as very-low-mass stars or black holes.[10]

In the primordial nucleosynthesis era, the neutrinos are extremely relativistic (ER), and, as discussed above, g is *not* 2, though there are two spin states, but ≈ 1. Thus, light element production remains essentially *unchanged* from the standard picture. This is not true if right-handed neutrinos couple, for example, through weak vector bosons[11] at a strength of $G_R \gtrsim 10^{-3}G_F$, where G_F is the usual Fermi coupling constant for left-handed neutrinos. The right decoupling temperature is then lower than the temperature at which copious quark-antiquark production occurs. Its red shift then results in the right- and left-handed neutrino temperatures maintaining approximate equality through the nuclear burning era. Thus, $g \approx 2$, which increases the expansion rate, and consequently raises the final helium and deuterium abundances above those of the standard model.[6] To get the observed deuterium abundance in the standard model with $g=1$ requires[12] $0.01 \lesssim \Omega_B h^2 \lesssim 0.02$; to explain low helium measurements may require $\Omega_B h^2$ to be as low as 0.003.[13]

When the red shift falls below $z_{NR} \approx 1.8\times 10^5 m_{30}$, the neutrinos go NR. The neutrino distribution function then results in a velocity dispersion of the left-handed background $\langle v^2\rangle^{1/2} = (12\eta_5/\eta_3)^{1/2}T_\nu/m_\nu c \simeq 6m_{30}^{-1}(1+z)$ km s^{-1}, where η_k is the Riemann eta function. This is now smaller than the dispersion in rich clusters ($\sim$1500 km/s) and in galaxies ($\sim$250 km/s). Assume now that one neutrino flavor carries almost all of the mass[5]; our general conclusions do not depend upon this assumption. A slight modification of the Tremaine and Gunn[2] argument to one neutrino flavor shows that nondegenerate neutrinos cannot form the dark matter in galactic halos unless $m_\nu > 25$ eV. Such a high value is certainly not ruled out by cosmology; therefore it is possible that massive neutrinos are the dark matter in the halos of spiral galaxies and the "missing mass" in clusters. We thus consider the role which neutrinos play in large scale clustering.

The velocity dispersion and the mass density imply a Jeans mass of

$$M_{J\nu} = M_{\nu m}x^{-3/2}f(x), \quad x = (1+z_{NR})/(1+z),$$
$$f(x) = 26.6\left(\frac{x}{24\eta_5}\int_0^\infty \frac{y^4\,dy}{(e^y+1)(y^2+x^2)^{1/2}}\right)^{3/2}\left(\frac{1}{2x\eta_3}\int_0^\infty \frac{y^2\,dy\,(y^2+x^2)^{1/2}}{e^y+1}\right)^{-2}, \tag{1}$$

which peaks at the maximum Jeans mass $M_{\nu m} = 3.9\times 10^{15}m_{30}^{-2}M_\odot$ attained at the red shift $z_m = 42\,900 m_{30}$. The function $f(x)$ involves integrals of the distribution function which we evaluate numerically in the semirelativistic regime. In the ER regime, $f = 0.32x^{7/2}$, and so the Jeans mass rises as $(1+z)^{-2}$ as z

drops. In the NR regime, $f = 26.6$, and the Jeans mass falls as $(1+z)^{3/2}$ through the hierarchy of structural mass scales observed in the Universe until the present epoch, when it is $1.4\times10^{9} m_{30}^{-7/2} \times M_{\odot}$. (See Fig. 1.)

When neutrinos are ER, but decoupled from matter, perturbations decay on all length scales smaller than the horizon.[14] Further, in the NR regime, neutrinos Landau damp on all scales smaller than the Jeans length,[15] in the same manner as collisionless baryons.[16] We conclude that *all primordial neutrino perturbations will be erased on mass scales smaller than the peak value* $M_{\nu m}$ and on all scales smaller than the associated length $\lambda_{\nu m}(z) \approx m_{30}^{-2}(1+z_m)/(1+z)$ kpc. Neutrino damping does not drive the damping of baryon fluctuations.[14] However, adiabatic perturbations do damp prior to and during the recombination epoch by the viscous coupling of the baryons to the photons on all scales below the Silk mass in baryons,[17,18] $M_{BS} \approx 3\times10^{13}\,\Omega_B^{-1/2} \times \Omega_\nu^{-3/4} h^{-5/2} M_{\odot}$, $\sim 10^{15} M_{\odot}$ for an Einstein–de Sitter universe with $\Omega_B = 0.01$ and $m_{30} = 1$. The maximum neutrino Jeans mass encloses a mass in baryons in our standard case $M_{Bm} \approx (\Omega_B/\Omega_\nu) M_{\nu m}$ $\approx 1.2\times10^{14} m_{30}^{-3} (\Omega_B h^2/0.01) M_{\odot}$. Depending upon parameter choice, M_{BS} may be greater or less than M_{Bm}, but they are of same order.

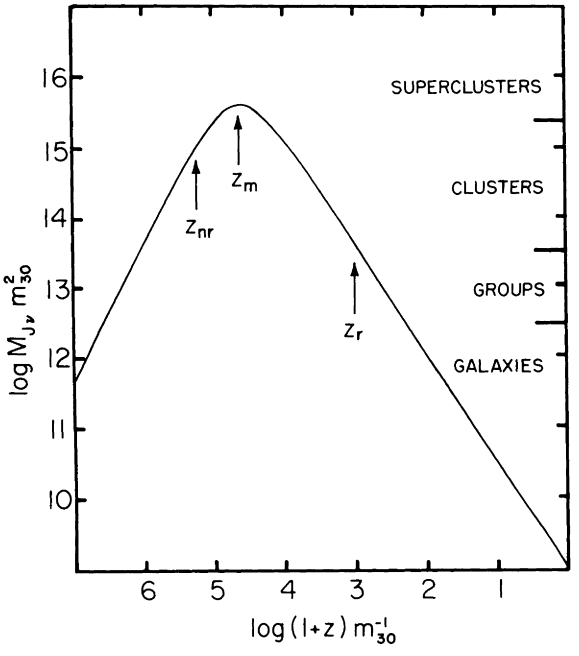


FIG. 1. The Jeans mass in solar mass units for one neutrino flavor vs red shift, along with approximate ranges for various large-scale structures. The neutrino mass in units of 30 eV is m_{30}. Perturbations on scales smaller than the maximum value of the Jeans mass are damped.

In the conventional picture of galaxy formation,[19] isothermal fluctuations (which do not suffer viscous damping) and all surviving adiabatic fluctuations grow at a rate $\sim(1+z)^{-1}$ in an Einstein–de Sitter universe after photon decoupling. In the presence of massive neutrinos, this result is drastically altered: *Baryon fluctuations smaller than the critical mass* M_{Bm} *grow extremely slowly in the linear regime until they first exceed the instantaneous Jeans mass.* The growth of perturbations of wavelength λ is described by the linearized Einstein field equations perturbed about a Friedmann background. After recombination, these reduce to an equation for the spatial Fourier transform of the fractional baryon and neutrino density fluctuations $\delta_B(\vec{k}, t)$ and $\delta_\nu(\vec{k}, t)$:

$$\frac{d^2\delta_B}{dt^2} + \frac{2\dot{a}}{a}\frac{d\delta_B}{dt} = 4\pi G\rho_T\left(\frac{\rho_B}{\rho_T}\delta_B + \frac{\rho_\nu}{\rho_T}\delta_\nu\right) \qquad (2)$$

which is valid in the linear regime, $\delta_B \ll 1$, $\delta_\nu \ll 1$, when matter pressure is neglected. Here, $a(t)$ is the Robertson-Walker scale factor and $\rho_T = \rho_\nu + \rho_B$. This is coupled to an integral equation for δ_ν obtained from the linearized Vlasov equation for the collisionless neutrino distribution function describing growth for $\lambda \geq \lambda_{J\nu}$ and damping for $\lambda \leq \lambda_{J\nu}$, where $\lambda_{J\nu}(z)$ is the instantaneous Jeans length.[15] Growth of neutrino density fluctuations occurs on scales $\lambda > \lambda_{\nu m}$ for $z < z_m$, but is suppressed in the radiation-dominated era.[20] Not until the epoch of photon decoupling can the baryons respond to the neutrino perturbations. Prior to recombination, no growth of baryon or photon fluctuations occurs on scales below the photon Jeans length, which is approximately the horizon size. After decoupling, δ_B rapidly responds to the neutrino perturbation (even if $\delta_B = 0$ initially), until $\delta_B \approx \delta_\nu$ on scales $\lambda > \lambda_{\nu m}$. Growth continues until $\delta_\nu \sim (1+z)^{-1}$ exceeds unity, when the enhanced self-gravity enables the first structures of mass $\sim M_{\nu m}$ to condense from the expanding background.

To consider scales $\lambda \leq \lambda_{\nu m}$, we take δ_ν to be zero below $\lambda_{\nu m}$ as our initial condition. For wavelengths $< \lambda_{J\nu}(z)$, we find the solutions in the Einstein–de Sitter case: $\delta_B \sim t^{p_\pm}$, where $p_\pm = \pm\frac{1}{6}[(1 + 24\rho_B/\rho_T)^{1/2} \mp 1]$. Only the p_- mode grows. If baryons dominate the mass density, $p_+ = \frac{2}{3}$, yielding the usual growing mode. If neutrinos dominate, p_+ is very small, $\sim\frac{1}{20}$ for $\Omega_B h^2 = 0.01$, and

$m_{30}=1$: Growth is negligible.

The growing part of the solution for $\lambda_{J\nu}(z) \leq \lambda \leq \lambda_{J\nu}(z_r)$ yields the fluctuation spectrum $\delta\rho_B/\rho_B(z) = \frac{3}{5}[\delta\rho_B/\rho_B(z_r)](\Omega_B/\Omega_\nu)[M/M_{J\nu}(z)]^{2/3}$. We write the fluctuation spectrum at z_r as $\delta\rho_B/\rho_B(z_r) = \delta\rho_B/\rho_B(z_m) = (M_c/M)^{n/6+1/2}$ which corresponds to a Fourier power spectrum $|\delta_B|^2 \propto k^n$. Note that the mass spectrum flattens by $M^{-1/3}$ at $M < M_{J\nu}(z_r)$, which corresponds to a power spectrum $|\delta_B|^2 \propto k^{n-4}$. If $n<1$, fluctuations on scale $M_{\nu m}$ are the first to condense after the recombination epoch.

We now show that condensation of galactic or subgalactic masses requires that either galaxies were formed in the nonlinear neutrino regime, $\delta_\nu \gtrsim 1$, or the baryon fluctuations were substantially nonlinear on galactic scales, $\delta_B \gtrsim 1$, at recombination. The nonlinear evolution of neutrino structures of mass $M_{\nu m}$ is likely to resemble a variant of Zel'dovich's "pancake" theory of galaxy formation from adiabatic fluctuations.[21] According to this theory, the generic type of collapse is asymmetric, with preferential collapse along one axis occurring at red shift z_{nl}. The Zel'dovich theory faces a serious constraint due to the relatively large amplitudes required for the initial adiabatic fluctuations: $\delta_B \gtrsim 10^{-3}(1+z_{nl})$ at decoupling. Since neutrino fluctuations grow between z_m and z_r, the value of δ_B at decoupling is reduced to $\sim\delta_\nu(z_m) \lesssim 5\times10^{-5}(1+z_{nl})m_{30}^{-1}$. The associated temperature fluctuations in the cosmic background radiation are[18] $\lesssim \frac{1}{3}\delta_B(z_r) \approx \frac{1}{4}\delta_\nu(z_m)$ over scales $\sim\lambda_{\nu m}$ and are below recent observational upper limits ($\Delta T/T \lesssim 2\times10^{-4}$ over scales $\sim 4'$).[22]

A difficulty appears in the galaxy distribution where significant power would be expected over large scales, corresponding to M_{Bm}. The two-point galaxy correlation function[19] has the form $\xi(r)=(r_0/r)^{1.8}$ with $r_0 \approx 4h^{-1}$ Mpc, whence the mass scale below which clustering is nonlinear is $M_{Bnl} \approx 5\times10^{14}\Omega_B h^{-1}[r_0 h/(4\text{ Mpc})]^3 M_\odot$. There is, however, tentative evidence for larger-scale nonlinear structure.[23] If we require M_{Bnl} to exceed M_{Bm}, we require $m_{30} > 2.9[(4\text{ Mpc})/r_0]$. This requires $\Omega_\nu h^2 \sim 1$ and $M_{Bm} \approx 4.6\times10^{12}(\Omega_B h^2/0.01)M_\odot$. This mass scale corresponds to a group of ~ 80 galaxies.[19]

The mass limit deduced using the correlation function disagrees with the neutrino mass inferred from measured mass-to-light ratios if we assume that the luminous baryonic matter has not segregated from the dark neutrino matter. Since $M_{\nu m}/M_{Bm} \approx 800(M/L)^{-1}m_{30}h^{-1}$, and $M/L \approx 500h$ for rich clusters, we require $m_{30} \approx 0.6h^2$ and hence $M_{\nu m} \approx 10^{16}h^{-4}M_\odot$. The problem is still more severe in the case of groups of galaxies, where $M/L < 500h$. We conclude that substantial dissipation on cluster scales is required in order to segregate the luminous matter from the neutrinos. We cannot rule out this possibility. Since in such a picture, galaxy formation occurs after the collapse of the pancake, it is difficult to see how galaxy halos composed of neutrinos could form given the high velocity dispersion expected of this cluster material [$\sim 1300 m_{30}^{-1}(1+z_{nl})$ km/s].

The alternative possibility that $\delta_\nu \sim 0$ initially constrains the initial amplitudes of isothermal baryon density fluctuations. To achieve $\delta\rho_B/\rho_B \sim 1$ now on scales $\sim M_{Bnl}$, we infer that $\delta\rho_B/\rho_B \sim 0.02(0.01/\Omega_B)[(4\text{ Mpc})/r_0 h]^2 m_{30}^{-2}[M_{Bnl}/M_B]^{n/6+1/2}$ at z_r on scales corresponding to a mass in baryons of M_B. Hence, on galactic scales $\delta\rho_B/\rho_B \approx m_{30}^{-2}$ for $1 \lesssim n \lesssim 4$. A value of n in this range is required for isothermal fluctuations to account for the galaxy correlation function.[19] Such a theory seems unattractive (galaxies exist because galaxies have always existed).

In summary, the ratio of right- to left-handed neutrinos is likely to be $\ll 1$ and thus primordial nucleosynthesis is unaffected by the presence of a neutrino mass in the standard big-bang model. To avoid conflict with q_0 determinations implies that $m_\nu \lesssim 190$ eV, otherwise $\Omega_\nu > 2$. If the dominant mass density of the Universe is in massive neutrinos (e.g., $m_\nu > 1$ eV if $\Omega_B h^2 \sim 0.01$), the gravitational growth of density fluctuations is inhibited in the linear regime on mass scales $< M_{\nu m}$. Moreover, the large-scale structure displays a prominent feature at $M_{\nu m} \sim 4\times10^{15}m_{30}^{-2}M_\odot$, in which case comparison with the galaxy distribution requires $m_\nu > 85$ eV. Alternatively, nonlinear structure on galaxy scales must be present in the very early Universe. We conclude that confirmation of a neutrino mass will require a significant revision of the gravitational instability theory of galaxy formation as outlined here.

We thank our colleagues at Berkeley, especially G. Lake and M. Wilson, for numerous discussions. This work was supported by the National Science Foundation under Grants No. AST-79-15244 and No. AST-79-23243.

[1]R. Cowsik and J. McClelland, Phys. Rev. Lett. 29, 669 (1972); J. E. Gunn, B. W. Lee, I. Lerche, D. N. Schramm, and G. Steigman, Astrophys. J. 223, 1015 (1978), and references therein.

[2]S. Tremaine and J. E. Gunn, Phys. Rev. Lett. 42,

407 (1979).
[3]V. A. Lyubimov, E. G. Novikov, V. Z. Nozik, E. F. Tretyakov, and V. S. Kowsik, to be published.
[4]F. Reines, S. W. Sobel, and E. Pasierb, Phys. Rev. Lett. 45, 1307 (1980).
[5]P. Ramond, California Institute of Technology Report No. CALT-68-709, 1979 (to be published); E. Witten, Phys. Lett. 91B, 81 (1980).
[6]See also S. L. Shapiro, S. A. Teukolsky, and I. Wasserman, Phys. Rev. Lett. 45, 669 (1980).
[7]A. Sandage, Astrophys. J. 173, 485 (1972).
[8]M. Davis, J. J. Geller, and J. Huchra, Astrophys. J. 221, 1 (1978).
[9]S. M. Faber and J. S. Gallagher, Annu. Rev. Astron. Astrophys. 17, 135 (1979).
[10]See also D. N. Schramm and G. Steigman, to be published.
[11]M. A. B. Beğ, R. V. Budny, R. Mohapatra, and A. Sirlin, Phys. Rev. Lett. 38, 1252 (1977).
[12]J. Yang, D. N. Schramm, G. Steigman, and R. T. Rood, Astrophys. J. 277, 697 (1979).
[13]F. W. Stecker, Phys. Rev. Lett. 44, 1237 (1980).
[14]P. J. E. Peebles, Astrophys. J. 180, 1 (1973).
[15]J. R. Bond, G. Efstathiou, and J. Silk, to be published.
[16]I. H. Gilbert, Astrophys. J. 144, 233 (1965).
[17]B. J. T. Jones, Rev. Mod. Phys. 48, 107 (1976).
[18]J. Silk and M. Wilson, Phys. Scr. 21, 708 (1980).
[19]S. M. Fall, Rev. Mod. Phys. 51, 21 (1979), and references therein.
[20]P. Mészáros, Astron. Astrophys. 37, 225 (1974).
[21]Ya. B. Zel'dovich, in *Large Scale Structure of the Universe*, I.A.U. Symposium No. 79, edited by M. Longair and J. Einasto (Reidel, Dordrecht, 1977), p. 409, and references therein.
[22]R. B. Partridge, Astrophys. J. 235, 681 (1980).
[23]R. P. Kirshner, A. G. Oemler, and P. L. Schechter, Astron. J. 84, 951 (1979).

Dynamical Role of Light Neutral Leptons in Cosmology

Scott Tremaine
W. K. Kellogg Radiation Laboratory, California Institute of Technology, Pasadena, California 91125, and Institute of Astronomy, Cambridge, England

and

James E. Gunn
W. K. Kellogg Radiation Laboratory, Hale Observatories, California Institute of Technology, Pasadena, California 91125, and Carnegie Institution of Washington, Washington, D. C. 20005
(Received 31 May 1978)

Using the Vlasov equation, we show that massive galactic halos cannot be composed of stable neutral leptons of mass $\lesssim 1$ MeV. Since most of the mass in clusters of galaxies probably consists of stripped halos, we conclude that the "missing mass" in clusters does not consist of leptons of mass $\lesssim 1$ MeV (e.g., muon or electron neutrinos). Lee and Weinberg's hypothetical heavy leptons (mass ≈ 1 GeV) are not ruled out by this argument.

Observations of rich clusters of galaxies[1] show that most of their mass is in a form which emits very little light: the mass-to-blue-light ratio $M/L_B \approx 200(H/50)$ in solar units,[1] where H(km s^{-1} Mpc^{-1}) is the Hubble constant. This is the classical "missing-mass" problem, although really it is the light and not the mass which is missing. The problem persists in smaller systems. Groups of galaxies have $M/L_B \approx 140(H/50)$.[2] The Local Group has $M/L_B \approx 60(H/50)$ if $q_0 = 0$,[3] binary galaxies have $M/L_B \approx 65(H/50)$,[4] and even in our own Galaxy and $M31$ most of the mass probably lies at $r \gtrsim 10$ kpc, outside the visible regions.[5-7] We conclude that even for systems with radii as small as ~ 20 kpc, most of the mass lies in dark material unlike the stellar population with $M/L_B \approx 7(H/50)$,[8] which dominates the mass in the central regions of galaxies. It seems likely that the dark mass in binaries, groups, and clusters is material from the halos of galaxies that has been stripped off by tidal interactions.[9,10] More detailed and quantitative comparisons[11] based on expected values of M/L_B from stellar content yield the same conclusion; viz., that the ratio of "missing" to visible mass is similar for all systems large enough that separation would not occur by dissipative processes in the time available—and thus that the dark material is likely to be the same everywhere. These arguments and Occam's razor lead us to the assumption that the composition of the dark material is the same in galactic halos, binary galaxies, groups, and clusters.

In some gauge theories the electron and muon neutrinos have a small, nonzero rest mass,[12] and it has been proposed that the dark material consists of such particles.[13] Shortly after the big bang the neutrinos are relativistic and in thermal equilibrium, and so their momentum distribution is

$$n_{\nu_e}(\vec{p})\,d\vec{p} = n_{\bar{\nu}_e}(\vec{p})\,d\vec{p} = n_{\nu_\mu}(\vec{p})\,d\vec{p} = n_{\bar{\nu}_\mu}(\vec{p})\,d\vec{p} = \frac{g_\nu}{h^3}\,d\vec{p}\left\{\exp\left[\frac{p}{kT_\nu(z)}\right]+1\right\}^{-1}. \quad (1)$$

© 1979 The American Physical Society

In this equation $c=1$, $T_\nu(z)=T_\gamma(z)$ is the common temperature of the neutrinos and the radiation, and g_ν is the number of allowed helicity states. For neutrinos of nonzero rest mass two helicities are allowed, but the cross section for production of positive-helicity neutrinos is down by $\sim(m/kT_\nu)^2$ from the cross section for negative-helicity neutrinos, and as a result, the former decouple at a very high temperature, where the evolution of the universe is not well understood. Thus it is not clear whether both helicity states should be included in g_ν, and we only know that $1\leq g_\nu\leq 2$. Both helicity states are decoupled by $kT_\nu\approx 1$ MeV,[14] and, since the electron and muon neutrinos are known to be less massive than 1 MeV, they are still relativistic when they decouple. Thereafter their momenta red shift like $1+z$, so that Eq. (1) continues to hold with $T_\nu(z)\propto 1+z$. But $T_\gamma(z)$ is also $\propto 1+z$ except for a jump of $(\frac{11}{4})^{1/3}$ at e^+e^- recombination,[14] and so we have

$$T_\nu(0)=(\tfrac{4}{11})^{1/3}T_\gamma(0)=1.9\ \mathrm{K}. \tag{2}$$

From Eqs. (1) and (2) the present mass density in neutrinos is

$$\rho_\nu=[(m_{\nu_e}+m_{\bar\nu_e})g_{\nu_e}+(m_{\nu_\mu}+m_{\bar\nu_\mu})g_{\nu_\mu}]6\pi\zeta(3)\left(\frac{kT_\nu(0)}{h}\right)^3=(1.9\times 10^{-31}\ \mathrm{g\ cm^{-3}})\left(\frac{m_{\nu_e}g_{\nu_e}+m_{\nu_\mu}g_{\nu_\mu}}{1\ \mathrm{eV}}\right). \tag{3}$$

In terms of the critical density $\rho_c=3H^2/8\pi G$,

$$\Omega_\nu=\frac{\rho_\nu}{\rho_c}=0.04\left(\frac{50}{H}\right)^2\frac{m_{\nu_e}g_{\nu_e}+m_{\nu_\mu}g_{\nu_\mu}}{1\ \mathrm{eV}}. \tag{4}$$

Neutrinos with a cosmologically interesting mass will have very small velocities: For $m\gtrsim 1$ eV, $v\approx kT_\nu(0)/m\lesssim 50$ km s^{-1}. Because this velocity is smaller than the typical velocity dispersion within galaxies or clusters, and because galaxies and clusters collapse in a dynamical time, the neutrinos will participate in the galaxy clustering. The ratio of the density distributed like the galaxies to ρ_c is called Ω^*. The best observational estimate of Ω^* is roughly 0.05.[15] Since $\Omega_\nu<\Omega^*$, we have

$$m_{\nu_e}g_{\nu_e}+m_{\nu_\mu}g_{\nu_\mu}\lesssim(1.2\ \mathrm{eV})(H/50)^2. \tag{5}$$

This mass is well below the current experimental upper limits on m_{ν_μ} and m_{ν_e} (cf. Ref. 13). This is the chain of argument which led Cowsik and McClelland[13] to the suggestion that massive neutrinos could supply the missing mass. If we look further, however, we can find a flaw. Because the neutrinos are noninteracting, the density of a fluid element in phase space is conserved (the Vlasov equation). Thus, the maximum fine-grained phase-space density is conserved. The maximum coarse-grained phase-space density is less well constrained: We know only that it must decrease with time, because the fluid can become "frothy" in phase space,[16] i.e., the regions of maximum fine-grained density may be mixed in with lower-density regions.

While neutrino background is homogeneous the maximum phase-space density is $2g_\nu h^{-3}$ from the four kinds of neutrinos, assuming $g_{\nu_e}=g_{\nu_\mu}\equiv g_\nu$ [cf. Eq. (1)]. We assume that the central regions of the bound systems formed by the neutrinos resemble isothermal gas spheres; then their velocity distribution is Maxwellian and the maximum phase-space density is $\rho_0 m_\nu^{-4}(2\pi\sigma^2)^{-3/2}$. Here ρ_0 is the central density and σ is the one-dimensional velocity dispersion, and we assume that $m_{\nu_e}=m_{\nu_\mu}\equiv m_\nu$. The core radius is defined by $r_c^{\,2}=9\sigma^2/4\pi G\rho_0$, and the requirement that the maximum phase-space density has decreased becomes

$$m_\nu{}^4>\frac{9h^3}{4(2\pi)^{5/2}g_\nu G\sigma r_c{}^2},\quad m_\nu>(101\ \mathrm{eV})\left(\frac{100\ \mathrm{km\ s^{-1}}}{\sigma}\right)^{1/4}\left(\frac{1\ \mathrm{kpc}}{r_c}\right)^{1/4}g_\nu{}^{-1/4}. \tag{6}$$

In comparison, the Pauli principle requires only that the present maximum phase-space density is $<4g_\nu h^{-3}$, giving a limit on m_ν which is similar to Eq. (6) but less severe by a factor $2^{1/4}$. Note, however, that the principle that the maximum phase-space density decreases is quite distinct from the Pauli principle. For example, our argument would also work for any hypothetical noninteracting Maxwell-Boltzmann particles. It does not work for bosons because their equilibrium phase-space density does not have a maximum.

For rich clusters of galaxies[1] $r_c\simeq 0.25(50/H)$ Mpc, $\sigma\simeq 10^3$ km s^{-1}. Thus, $m_\nu\gtrsim(3.6\ \mathrm{eV})(H/50)^{1/2}g_\nu{}^{-1/4}$, roughly consistent with the limit (5). Binary galaxies, however, have $r_c\lesssim 100\ (50/H)$ kpc (the typical separation of the components) and $\sigma\approx 100$ km s^{-1} (Ref. 4), and inequalities (5)

and (6) yield $m_\nu \lesssim (0.6\text{ eV})(H/50)^2 g_\nu^{-1}$ and $m_\nu \gtrsim (10\text{ eV})(H/50)^{1/2} g_\nu^{-1/4}$, which are inconsistent. The inconsistency is worse still for galactic halos, where $r_c \lesssim 20$ kpc, $\sigma \approx 150$ km s^{-1}, and $m_\nu \gtrsim (20\text{ eV}) g_\nu^{-1/4}$. We conclude that massive muon or electron neutrinos cannot supply the dark material in galactic halos or binary galaxies or, by extension, in clusters of galaxies.

The contradiction cannot be removed by assuming that the luminous parts of the galaxies condense first and suck the neutrinos into their potensial well, because we have not assumed that the neutrinos supply all the gravitational potential.

Notice that this argument is not based on the Jeans mass. In fact, the present Jeans mass for the neutrinos is

$$M_J \approx \rho[\pi v_s^2/G\rho]^{3/2} \approx 3\times 10^{13} M_\odot [m_\nu/1\text{ eV}]^{-7/2} g_\nu^{-1/2}, \quad (7)$$

where we have replaced the "sound speed" v_s by $[kT_\nu(0)/m_\nu]^{1/2}$. Thus systems of galactic or larger mass could form if $m_\nu \lesssim 4$ eV; the phase-space argument shows, however, that they would be larger than the systems we observe.

Are there any loopholes? The most uncertain parameter is Ω^*. If we regard these arguments as setting a lower limit on Ω^*, then the most conservative limit is obtained when $g_\nu = 1$ and one species of neutrinos is much more massive than the other (say, $m_{\nu_\mu} \gg m_{\nu_e}$). Then the values $r_c \lesssim 20$ kpc, $\sigma \approx 150$ km s^{-1} appropriate for galactic halos require $\Omega^* \gtrsim 1.0(50/H)^2$, $m_\nu \gtrsim 24$ eV, if massive neutrinos are to make up the halos. This result is well outside the acceptable limits to Ω^* and H (cf. Ref. 15). A very simple argument is worth noting at this point. Particles as heavy as 24 eV behave completely classically at the phase-space densities in the great clusters and would cluster on these scales precisely as the galaxies do.[17] If Ω^* were as large as unity in such particles, the ratio of unseen to visible matter in the clusters would be larger than it is, corresponding to mass-to-light ratios in excess of 1500—a value never approached in observation.

In a cosmology with nonzero muon or electron lepton number the problem only gets worse. The reason is that the maximum phase-space density of neutrinos plus antineutrinos is independent of chemical potential μ_ν, while the spatial density of neutrinos plus antineutrinos increases if μ_ν is nonzero. Thus inequality (6) is unchanged but inequality (5) requires smaller neutrino masses. (If m_ν is very small or zero, the neutrinos do not cluster with the galaxies; in this case, if $\mu_\nu kT_\nu \gg 1$, neutrinos could close the universe, but they could not supply the missing mass in clusters. Cosmological helium production is also impossible in such models.[14])

Finally, we consider the possibility that there are other stable neutral leptons besides the electron and muon neutrino. If the lepton mass is $\lesssim 1$ MeV all of our arguments are still valid [if there are *more* than two species with comparable masses, the inconsistency of inequalities (5) and (6) is still worse]. If their mass is $\gtrsim 1$ MeV, the assumption that the leptons are relativistic when they decouple is no longer valid; in this case Lee and Weinberg[18] have shown that leptons with mass $\gtrsim 1$ GeV can contribute $\Omega_\nu \approx \Omega^*$. Such heavy leptons, although reprehensible on etymological grounds, are not ruled out by our phase-phase-density arguments (see also Ref. 17).

In summary, we know that there is some dark material which dominates the dynamics of binary galaxies, and makes up most of the mass in the outer halos of isolated galaxies. We have shown that this material cannot be muon or electron neutrinos of nonzero rest mass, or any stable neutral lepton less massive than 1 MeV. Several arguments suggest that the dark material is universal, so that the same restrictions apply to the unknown component which provides most of the mass in groups and clusters of galaxies.

We feel that the most likely remaining possibilities are small black holes or very low-mass stars, or perhaps some much heavier stable neutral particles.

One of us (S.T.) thanks J. P. Ostriker and P. J. E. Peebles for discussions. This research was supported in part by the National Science Foundation Grant No. AST76-80801A01 and by the Alfred P. Sloan Foundation.

[1]The most recent review is N. A. Bahcall, Annu. Rev. Astron. Astrophys. 15, 505 (1977).

[2]J. R. Gott and E. L. Turner, Astrophys. J. 213, 309 (1977).

[3]J. E. Gunn, Comments Astrophys. Space Sci. 6, 7 (1975).

[4]E. L. Turner, Astrophys. J 208, 304 (1976).

[5]W. L. W. Sargent, in *The Evolution of Galaxies and Stellar Populations*, edited by B. M. Tinsley and R. B. Larson (Yale Univ. Observatory, New Haven, 1977), p. 427.

[6]D. N. C. Lin and D. Lynden-Bell, Mon. Not. Roy. Astron. Soc. 181, 59 (1977).

[7]M. S. Roberts and R. N. Whitehurst, Astrophys. J. 201, 327 (1975).
[8]S. M. Faber and R. E. Jackson, Astrophys. J. 204, 668 (1976).
[9]D. O. Richstone, Astrophys. J. 204, 642 (1976).
[10]J. E. Gunn, Astrophys. J. 218, 592 (1977).
[11]J. E. Gunn, J. Phys. (Paris), Colloq. 37, C6-183 (1976).
[12]H. Fritzsch, M. Gell-Mann, and P. Minkowski, Phys. Lett. 59B, 256 (1975).
[13]R. Cowsik and J. McClelland, Phys. Rev. Lett. 29, 669 (1972), and Astrophys. J. 180, 7 (1973).
[14]A general reference is S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972), Chap. 15, or P. J. E. Peebles, *Physical Cosmology* (Princeton Univ. Press, Princeton, N. J., 1971), Chap. 8.
[15]J. R. Gott, J. E. Gunn, D. N. Schramm, and B. M. Tinsley, Astrophys. J. 194, 543 (1974).
[16]D. Lynden-Bell, Mon. Not. Roy. Astron. Soc. 136, 101 (1967).
[17]J. E. Gunn, B. L. Lee, I. Lerche, D. N. Schramm, and G. Steigman, "Some Astrophysical Consequences of the Evolution of a Heavy Stable Neutral Lepton" (to be published).
[18]B. W. Lee and S. Weinberg, Phys. Rev. Lett. 39, 165 (1977).

THE ASTROPHYSICAL JOURNAL, 266:L11–L15, 1983 March 1
© 1983. The American Astronomical Society. All rights reserved. Printed in U.S.A.

ACCURATE RADIAL VELOCITIES FOR CARBON STARS IN DRACO AND URSA MINOR: THE FIRST HINT OF A DWARF SPHEROIDAL MASS-TO-LIGHT RATIO[1]

MARC AARONSON
Steward Observatory, University of Arizona
Received 1982 August 6; accepted 1982 October 8

ABSTRACT

Velocities accurate to ~ 1 km s^{-1} have been obtained with the Multiple Mirror Telescope and echelle spectrograph for three carbon stars in the Draco dwarf galaxy and one carbon star in the Ursa Minor dwarf. These observations demonstrate that measurement of radial velocities having such high precision is quite feasible for stars as faint as $V \sim 18$ mag. The data presented here are of importance for understanding the dynamical history of the dwarf systems. In addition, they provide a first and tantalizing hint of the velocity dispersion in a dwarf spheroidal and suggest that Draco may have a mass-to-light ratio an order of magnitude greater than that found for galactic globulars. If confirmed, this result would support the existence of a massive halo about the Galaxy. It would furthermore rule out the possibility that neutrinos could provide a solution to the missing mass problem, if the dark matter on small and large scales is similar.

Subject headings: galaxies: general — mass-luminosity relation — stars: carbon

I. INTRODUCTION

The outer halo of the Milky Way contains seven ghostly collections of stars known as dwarf spheroidals. The absence of recent star formation in these systems has so far limited our knowledge of their dynamical properties. Simple radial velocities have only become available for all seven objects within the last year, and even many of these are quite uncertain. This is because in general such velocities can only be obtained from spectra of individual stars at the tip of an old giant branch, for which $V \gtrsim 17$ mag.

In this *Letter* highly accurate radial velocities are presented for stars in two dwarf spheroidals, Draco and Ursa Minor. The precision of these observations is believed to be ~ 1 km s^{-1}. The stars in question range between $V = 17.1$–17.7 mag, which is $\gtrsim 2$ mag fainter than any previously published stellar velocities accurate to ~ 1 km s^{-1} of which the author is aware.

Such velocity measurements have a number of applications. From them, orbital parameters of the dwarfs can be derived (e.g., Hodge 1966). They also can be used to model the mass of the Galaxy (Hartwick and Sargent 1978). They furthermore provide an important constraint for tidal interaction theories (Lynden-Bell 1976). However, the real benefit of very accurate velocities is that determination of the velocity dispersion then becomes feasible, from which direct measurement of the mass and mass-to-luminosity (M/L) ratio follows. The major question in this regard is whether or not M/L for dwarf spheroidals is similar to that of the galactic globulars they superficially resemble in stellar content. The latter appear to have $M/L \sim 2$, with little evidence for the presence of hidden mass (Illingworth 1976; Gunn and Griffin 1979).

There is some reason for believing that the dwarf spheroidals may in fact have substantially greater M/L values (Faber and Lin 1983). Why this might be so is easy to see. King (1962) gives the following relation between tidal radius, r_t; perigalactic distance, R_P; dwarf mass, M; Milky Way mass, M_g; and orbital eccentricity, e:

$$r_t = R_P\left[M/M_g(3+e)\right]^{1/3}. \quad (1)$$

Because the dwarf spheroidals provide such an excellent fit to tidally truncated King models (Hodge 1966), it seems plausible that their radii have indeed been determined by a close encounter with the Galaxy. Let us take Draco as an example, for which $r_t = 26'$ (Hodge 1971). Then, with $R_P = 55$ kpc (equal to the Large Magellanic Cloud distance, as compared with a current distance of 76 kpc from Stetson 1979), $M_g = 1 \times 10^{12}$ $M_\odot$ (i.e., a galaxy with a massive halo), and $e = 0.5$, we find

$$M/L = 29, \quad (2)$$

in units of M/L for the Sun, adopting $M_V = -8$ for Draco (Zinn 1980). Similar estimates can be obtained

[1]Research reported here used the Multiple Mirror Telescope Observatory, a facility operated jointly by the University of Arizona and the Smithsonian Institution.

for the other dwarf spheroidals as well, because if the Galaxy does indeed possess a massive halo, the dwarfs can only maintain their large tidal radii if they themselves contain a copious quantity of dark matter.

In the above example (the result for which has of course a large uncertainty), the amount of hidden to visible mass derived is comparable to the ratio thought to apply to our Galaxy (see Gunn 1982). Even with an order of magnitude increase in mass, the central escape velocity just scales as $(M)^{1/2}$ and will only range from a few km s^{-1} for Draco to $\sim$ 20 km s^{-1} for the Fornax dwarf. Since the observed gas expansion rate for globular cluster giants undergoing mass loss is typically $\sim$ 50 km s^{-1} (Mallia and Pagel 1978), there does not seem to be any conflict with the current upper limits on dwarf gas found by Knapp, Kerr, and Bowers (1978).

The observations presented in § II suggest that equation (2) may be rather close to the truth. Some further implications of such a result are discussed in § III.

II. OBSERVATIONS

The measurements in this *Letter* were originally motivated by the discovery of carbon stars in Draco and Ursa Minor (Aaronson, Liebert, and Stocke 1982; Aaronson, Olszewski, and Hodge 1983). The optical spectra of these stars are characterized by very prominent band heads of the C_2 Swan system, and it was thought that high-resolution spectroscopy centered on one of the band heads could yield an accurate radial velocity even with spectra of very low signal-to-noise ratio. This expectation proved to be correct.

For several reasons, the λ5636 band head was selected for observation. First, the band head sits in a relatively clean region away from other prominent features (particularly isotopic ones) which might degrade its "sharpness." Second, the stars in question are red, generally having $B - V \gtrsim 1.5$ mag. Finally, a red-sensitive image-tube/Reticon package ("Red Hot") having a very low dark count (2 counts s^{-1} per array on a cold night) became available.

Observations of four dwarf spheroidal carbon stars were obtained with the Multiple Mirror Telescope (MMT) and echelle spectrograph/Reticon combination during the spring of 1982 (Table 1, Fig. 1). The image stacker described by Chaffee and Latham (1982) was also employed. The spectral coverage was $\sim$ 50 Å, with a resolution as judged from the width of the Th-Ar comparison lines of $\sim$ 1/6 Å. Comparison lamps were taken before, during, and after all exposures in order to contain possible flexure effects which, however, proved to be almost negligible. Radial velocities were determined using the Smithsonian Astrophysical Observatory (SAO) cross-correlation software (Tonry and Davis 1979; Latham 1982).

The information given in Table 1 includes the object name, V-magnitude, date of observation, exposure time, height of the cross-correlation peak (a value of 0.4 being considered quite acceptable), heliocentric radial velocity, and error returned by the SAO software. Appropriate integration times were not *a priori* apparent, and so in general the spectrum was built up until the λ5636 band head became very clearly visible. As it turned out (see below), most exposure times were about twice as long as necessary for achieving adequate results. One star (Draco C1) with an unusual emission-line spectrum which might indicate its membership in a binary system (see Aaronson, Liebert, and Stocke 1982) was observed twice, but no significant change in radial velocity was found.

Two radial velocity templates were employed, having been selected from a list of galactic R and CH stars very kindly provided by McClure (1982), who has been monitoring such objects for radial velocity variations. The two templates chosen were HD 156074 and HD 197604; these have been observed by McClure 11 and 5 times, respectively, and have velocities with a mean standard deviation of $\sim$ 0.2 km s^{-1}, with no evidence for any

TABLE 1

RADIAL VELOCITIES OF DWARF SPHEROIDAL CARBON STARS

Name	V[a] (mag)	Date	Exposure Time (min)	Height	Velocity (km s^{-1})	σ (km s^{-1})
Draco C1	17.1	1982 Apr 19	105[b]	0.3	−298.7	0.9
		1982 May 16	90	0.5	−297.6	0.6
Draco C2	17.3	1982 May 30	80	0.5	−300.2	0.5
Draco C3	17.5	1982 May 30	30	0.3	−279.7	0.9
Ursa Minor 335 ...	17.7	1982 May 30	105	0.5	−260.7	0.6

[a] V-magnitude references are as follows: Dra C1 and C2, Aaronson, Liebert, and Stocke 1982; UMa 335, Cudworth, Olszewski, and Schommer 1983; Dra C3, photoelectric measurement obtained by J. Mould with the Kitt Peak 1.3 m telescope.

[b] Observed through cirrus.

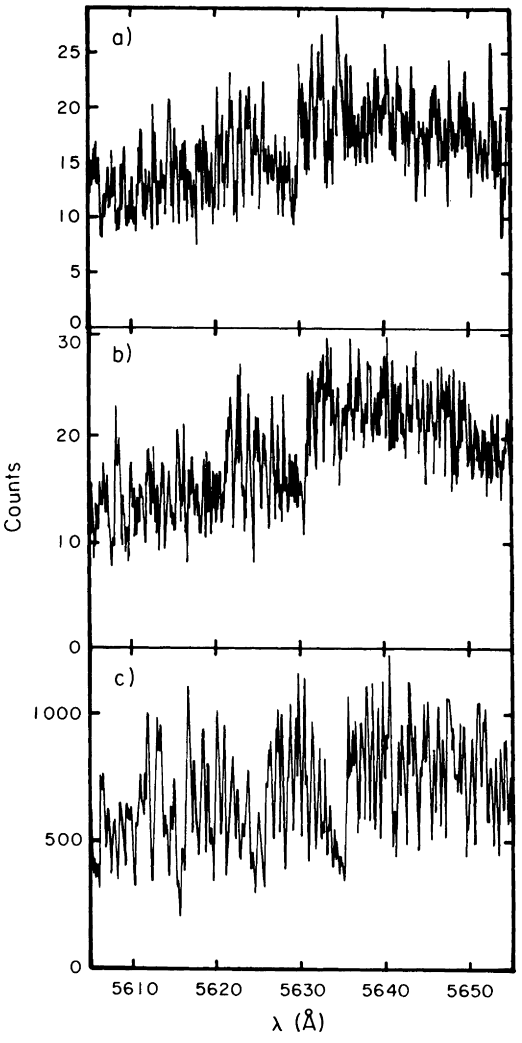


FIG. 1.—MMT echelle spectroscopy of (*a*) Draco C2, (*b*) Ursa Minor 335, and (*c*) the template star HD 197604. The most prominent absorption break in these spectra is the λ5636 C_2 Swan band head. The spectra in (*a*) and (*b*) have been Hanning smoothed, and the dark count level in them is roughly at 7 and 9 respectively.

systematic variations larger than 0.5 km s^{-1}. Any such variations are, of course, a concern only for the absolute velocity scale and are not relevant to the velocity dispersion problem. The two templates yielded velocities agreeing well within the estimated errors, and we have listed in Table 1 only the results obtained with HD 197604.

For several reasons, the errors given in Table 1 may be underestimated, but probably by no more than 30%. First, the error coefficient in the software is adjusted for the Mount Hopkins 1.5 m Tillinghast spectrograph and not the MMT echelle (see Tonry and Davis 1979). Several additional causes for concern are the correction from observed to heliocentric velocity (thought to be accurate to $\lesssim 0.2$ km s^{-1}), guiding differences between program and template objects, and systematic offsets between comparison lamp and atmospheric lines. However, the findings presented in Table 2 suggest that the last effects are of little importance. The results of cross-correlating the template stars against each other are given here. It is quite gratifying (and possibly a bit lucky) that all of the returned values are within 0.1 km s^{-1} of that expected.

The longer observations are mostly composed of 30 minute chunks, and, to test the effects of signal-to-noise ratio on the final velocities, some of these shorter exposures were by themselves cross-correlated. In all cases having a correlation peak height greater than 0.15 the results in Table 1 were reproduced to within 1 km s^{-1}. Owing to the strength of the λ5636 band head, it is in fact a simple matter to recover the values in Table 1 to within a few km s^{-1} from raw spectral plots using only a pencil and ruler. We are therefore confident that the errors in Table 1 are realistic, particularly for Draco C3, for which only one 30 minute exposure was obtained.

Our Draco radial velocities are in good agreement with the value of -277 ± 35 km s^{-1} given by Hartwick and Sargent (1978) and also with the value of -292 ± 6 km s^{-1} measured for Draco C1 by Aaronson, Liebert, and Stocke (1982) from this star's emission lines. Our Ursa Minor radial velocity is close to one of the two velocities found by Hartwick and Sargent. These authors obtained -244 km s^{-1} for star 230, but only -105 km s^{-1} for star 177. Hartwick and Sargent suggested that the latter star could be an interloper based on the presence of stronger metallic lines. However, both stars appear to be proper motion members (Cudworth, Olszewski, and Schommer 1983).

III. DISCUSSION

The results in Table 1 provide the first clue about the velocity dispersion for a dwarf spheroidal. Of course, any dispersion calculated from just three measurements is necessarily crude, so we instead ask the following question: What is the minimum velocity dispersion allowable by a χ^2 test at the 5% level? For 2 degrees of freedom, this corresponds to that dispersion which yields $\chi^2 = 3.0$. The answer is 6.5 km s^{-1}.

To see what such a dispersion implies, we start with the relation between mass, M (in solar units), core radius, r_c (in parsecs), dimensionless mass parameter, μ, and dispersion squared, $\langle V_r^2 \rangle$, in (km s^{-1})^2 given by Illingworth (1976):

$$M = 167\, r_c \mu \langle V_r^2 \rangle. \quad (3)$$

TABLE 2

TEMPLATE SELF - CORRELATIONS

TEMPLATE USED	ADOPTED VELOCITY	VELOCITY RETURNED BY SOFTWARE			
		HD 197604 1982 Apr 19	HD 197604 1982 May 16	HD 156074 1982 May 16	HD 197604 1982 May 30
HD 156074[a] ...	−12.7	16.6	16.5	−12.7[b]	16.6
HD 197604[a] ...	16.5	16.6	16.6[b]	−12.7	16.6

[a] Template constructed from 1982 May 16 observation.
[b] Template cross-correlated against original spectrum.

For Draco we have $r_c = 6'.5$ and $\mu = 4.1$ (King 1966; Hodge 1971), so that

$$M = 9.9 \times 10^4 \langle V_r^2 \rangle \quad (4)$$

and

$$M/L = 0.72 \langle V_r^2 \rangle . \quad (5)$$

Equations (4) and (5) should yield lower limits because the observed dispersion requires correction upward for its expected decrease with radius, owing to the decrease in escape velocity far from the center. (All three carbon stars in fact lie within r_c, although two only just so.) As an alternative approach to estimating M/L, we might instead use the $r^{1/4}$ law/virial theorem method (Illingworth 1976). From Table 2 of Hodge (1966), Draco's effective radius, r_e, is estimated to be 9′, and the coefficient in equation (6) would then become 0.98. However, the King model seems preferable owing to the far superior fit of Draco's surface brightness profile to it than to the $r^{1/4}$ relation. For $\langle V_r^2 \rangle = [6.5 \text{ (km s}^{-1})]^2$, we then find $M/L = 31$, a value similar to the prediction of equation (2). Hence, we might conclude that, at the 95% confidence level, Draco's mass-to-light ratio is at least an order of magnitude more than that of the typical galactic globular.

Is there another explanation for the large range in the Draco carbon star velocities? Two possibilities are that the stars have large atmospheric motions or that they are members of binary systems. However, none of the carbon stars (either in Draco or other spheroidals) are known to be variables. Furthermore, McClure's (1982) preliminary results indicate a low frequency of binary membership among galactic R and CH stars: only one of 12 stars shows variations greater than 3 km s^{-1}. Most of the remaining stars appear to exhibit velocity changes $\lesssim 2$ km s^{-1}, and it seems likely that these do arise from small-scale atmospheric motions. Similar low-amplitude motions have been observed for the giant stars in M3 by Gunn and Griffin (1979).

One further explanation for the results here could be that Draco, the closest of the dwarf spheroidals, is just now being tidally disrupted. The time scale for this would be short, $\lesssim 2 \times 10^8$ yr (see Hodge and Michie 1969). Such an argument therefore seems ad hoc but cannot be fully ruled out at present. The idea will become untenable, though, if other, more distant dwarfs such as Fornax can be shown to have large values of M/L as well.

Several interesting implications follow if Draco's mass-to-light ratio is really similar to that estimated above. First, in some fundamental sense, the dwarf spheroidals, puny as they are, would be confirmed as true galaxies and not just extended versions of globular clusters. Second, considerable support would be given to the notion that the Galaxy possesses a massive halo. A final conclusion can be made relating to the nature of the hidden matter itself. Tremaine and Gunn (1979) have used a confinement argument to limit the mass of any collisionless particle that is relativistic at decoupling (such as a neutrino) which might provide a source of dark matter. The constraint depends on the scale size such that mass $m \propto r^{-1/2}$. The smallest scale size considered by Tremaine and Gunn was that of a galactic halo for which, however, the limits are weak, yielding for a neutrino mass $m_\nu \gtrsim 24$ eV and a mass density parameter $\Omega \gtrsim 0.25\ h^2$, where h is the Hubble constant in units of $100/H_0$. Applied to Draco, these limits are considerably more stringent: for $\langle V_r^2 \rangle = [6.5 \text{ (km s}^{-1})]^2$ and neutrino helicity $g_\nu = 1$, we find $m_\nu \gtrsim 530$ eV and $\Omega \gtrsim 5h^2$.

Tremaine and Gunn (1979) argue quite plausibly that the hidden mass, whatever its nature, is similar on all scales. However, an alternative picture has been presented by Schramm and Steigman (1981). To account for the observed increase of M/L with scale size, these authors suggest that, while low-mass neutrinos dominate in the great clusters, nonluminous baryonic matter may contribute significantly to the mass in halos and small groups. Until the current experimental controversy about neutrino mass is resolved, a choice between these two hypotheses can only be made on philosophical grounds, but if the former view is correct, then the presence of a substantial amount of hidden matter in Draco would appear to eliminate the possibility that neutrinos could provide a solution to the missing mass problem (see also Lin and Faber 1983).

It is clearly of far-reaching significance to obtain radial velocities for additional stars in Draco in order to confirm the velocity dispersion suggested above, especially since the present case for a large dispersion rests mainly with the observation for one star (Draco C3). This will require measurement of accurate velocities for noncarbon stars. With a careful selection of the spectral region, such observations appear quite feasible with the MMT and echelle spectograph. A window centered on the Mg band at $\sim \lambda 5200$ looks rather promising in this regard (Latham 1982). Extending these measurements to other dwarf spheroidals is also very important. Fornax seems a particularly promising object to work on as, in addition to bright globular clusters, it contains a large number of luminous carbon stars. Perhaps the most intriguing system, though, is Ursa Minor. This dwarf has roughly the same luminosity and distance as Draco, but apparently twice the tidal radius, although the latter value of $59' \pm 25'$ is very uncertain (Hodge 1971). It may be, nevertheless, that Ursa Minor either is currently on its final approach to the Galaxy and is now being pulled apart (Hodge 1971), or has a mass-to-light ratio considerably larger than Draco's.

It is a pleasure to thank Paul Schechter, Garth Illingworth, and Nick Suntzeff for helpful discussion; Fred Chaffee, John Huchra, Dave Latham, Ruth Peterson, and Bill Wyatt for putting up with innumerable questions about the SAO reduction software; Janet Robinson, Sharon Perpignani-Huth, Fred Chaffee, Ray Weymann, and Jacques Beckers for assistance at the telescope; and Bob McClure and Jeremy Mould for communication of data before publication. It is especially a pleasure to thank Doug Lin and Sandy Faber for their strong encouragement and for first presenting to the author some of the ideas discussed in this *Letter*. This work was partially supported with funds from NSF grant AST 81-17365.

REFERENCES

Aaronson, M., Liebert, J., and Stocke, J. 1982, *Ap. J.*, **254**, 507.
Aaronson, M., Olszewski, E. W., and Hodge, P. W. 1983, *Ap. J.*, **267**, in press.
Chaffee, F. H. Jr., and Latham, D. W. 1982, *Pub. A.S.P.*, **94**, 386.
Cudworth, K. M., Olszewski, E. W., and Schommer, R. A. 1983, in preparation.
Faber, S. M., and Lin, D. N. C. 1983, *Ap. J. (Letters)*, **266**, L17.
Gunn, J. E. 1982, in *Astrophysical Cosmology: Proceedings of the Vatican Study Week on Cosmology and Fundamental Physics*, ed. H. A. Bruck, G. V. Coyne, and M. S. Longair (Vatican: Pontif. Acad. Sci.), p. 233.
Gunn, J. E., and Griffin, R. F. 1979, *A.J.*, **84**, 752.
Hartwick, F. D. A., and Sargent, W. L. W. 1978, *Ap. J.*, **220**, 453.
Hodge, P. W. 1966, *Ap. J.*, **144**, 869.
_____. 1971, *Ann. Rev. Astr. Ap.*, **9**, 35.
Hodge, P. W., and Michie, R. W. 1969, *A.J.*, **74**, 587.
Illingworth, G. 1976, *Ap. J.*, **204**, 73.
King, I. R. 1962, *A.J.*, **67**, 471.
_____. 1966, *A.J.*, **71**, 64.
Knapp, G. R., Kerr, F. J., and Bowers, P. F. 1978, *A.J.*, **83**, 360.
Latham, D. W. 1982, private communication.
Lin, D. N. C., and Faber, S. M. 1983, *Ap. J. (Letters)*, **266**, L21.
Lynden-Bell, D. 1976, *M.N.R.A.S.*, **174**, 695.
Mallia, E. A., and Pagel, B. E. J. 1978, *M.N.R.A.S.*, **184**, 55P.
McClure, R. D. 1982, private communication.
Schramm, D. N., and Steigman, G. 1981, *Ap. J.*, **243**, 1.
Stetson, P. B. 1979, *A.J.*, **84**, 1149.
Tonry, J., and Davis, M. 1979, *A.J.*, **84**, 1511.
Tremaine, S. D., and Gunn, J. E. 1979, *Phys. Rev. Letters*, **42**, 467.
Zinn, R. 1980, in *Globular Clusters*, ed. D. Hanes and B. Madore (Cambridge: University of Cambridge Press), p. 191.

Note added in proof.—On 1982 October 10, a radio velocity for a fourth (noncarbon) Draco star was secured with the MMT and echelle using an order centered at $\sim \lambda 5200$. The result obtained for star no. 75-473 ($V = 17.6$ mag) was -285.6 ± 1.1 km s^{-1}, adding considerable support to the case for a large Draco velocity dispersion. Both the twilight sky and the carbon star HD 197604 were used as templates and yielded velocities agreeing to within 0.2 km s^{-1}, so there does not appear to be any significant zero-point difference between the carbon and the noncarbon velocity systems.

MARC AARONSON: Steward Observatory, University of Arizona, Tucson, AZ 85721

THE ASTROPHYSICAL JOURNAL, 266:L21–L25, 1983 March 1
© 1983. The American Astronomical Society. All rights reserved. Printed in U.S.A.

SOME IMPLICATIONS OF NONLUMINOUS MATTER IN DWARF SPHEROIDAL GALAXIES[1]

D. N. C. LIN AND S. M. FABER
Lick Observatory and Board of Studies in Astronomy and Astrophysics, University of California, Santa Cruz
Received 1982 August 19; accepted 1982 November 4

ABSTRACT

Strong constraints are placed on the nature of nonluminous matter if it dominates the gravitational potential of dwarf spheroidal galaxies. In particular, the phase-space constraint in small galaxies sets a lower limit of several hundred eV on particle mass if the dark matter consists of noninteracting fermions. This limit is sufficiently strong to rule out neutrinos. Difficulties of a fundamental nature are also encountered even with more massive noninteracting particles. In scenarios where black holes comprise the nonluminous matter, they have to be less massive than $\sim 100\ M_\odot$.

The quasi-exponential light profiles and dark matter in dwarf spheroidal galaxies may indicate a possible evolutionary link to dwarf irregular galaxies. Evidence is presented which supports the view that dwarf spheroidal galaxies are former dwarf irregular galaxies which lost their gas near the Milky Way via ram pressure sweeping.

Subject headings: cosmology — galaxies: formation — galaxies: structure — neutrinos

I. INTRODUCTION

There are seven known dwarf spheroidal (DS) galaxies in the neighborhood of the Milky Way (Hodge 1971; Cannon, Niss, and Nørgaard-Nielsen 1981). Preliminary evidence indicates that these DS satellites may contain appreciable amounts of nonluminous matter and that their luminosity profiles are as well fitted by exponentials as they are by tidally truncated models (Faber and Lin 1983). Although the evidence for these findings is still tentative, we discuss their implications briefly here. Our main purpose is to emphasize the importance of dwarf galaxies to our understanding of galaxy formation and nonluminous matter.

II. ORIGIN OF DWARF SPHEROIDAL GALAXIES

Although dwarf spheroidal galaxies have been classified as members of the elliptical family of galaxies, their nearly exponential profiles resemble those of disk systems. Since their masses are clearly small, a parallel with dwarf irregular (DI) galaxies is appropriate. In this section, we explore the hypothesis that DS galaxies are former DI galaxies that somehow lost their gas. Although DI galaxies appear irregular in blue light, their underlying old stars are smoothly distributed (Hodge 1977, 1978) and their axially averaged luminosity profiles are closely exponential (Hodge 1971). If gas could somehow be removed from a DI galaxy, the young stellar population would soon fade, leaving behind a smooth, low surface brightness galaxy not unlike a DS galaxy. Related views on the origin of DS galaxies have been previously expressed by Einasto *et al.* (1974), Frogel *et al.* (1982), and Aaronson and Mould (1980). To assess this picture for the origin of DS galaxies, we address the following four questions.

1. *Are the structural parameters of DS galaxies consistent with their previously having been DI galaxies*? To examine this question, we have selected three nearby DI galaxies as prototypes. The upper section of Table 1 shows their current properties; the second section shows their hypothetical properties a few billion years after star formation ceases. Fading factors for the stellar population are based on M/L_V versus color from the population models of Larson and Tinsley (1978), plus a final assumed globular cluster M/L_V of 2.5 (Illingworth 1976; Gunn and Griffin 1979).

Comparing the final properties with those of actual DS galaxies in Table 1, we see that two of the DI galaxies (NGC 6822 and IC 1613) are too large to match any DS galaxy, while the third (GR 8) is too small. The extrapolated DI properties nicely bracket the observed DS range, however, and it seems probable that DI galaxies of intermediate brightness would provide a good match.

If only central surface brightnesses are compared, the DI sample can be enlarged to five galaxies (see notes to Table 1). The final central surface brightness after fading based on this sample is 24.4 μ_V, a good match to the observed mean DS value of 23.9 μ_V. Final isophotal radii and internal velocities have also been estimated, as described in Table 1 and the notes, and these quantities again compare favorably. Finally, we note that the ellipticity distribution of the local DS galaxies appears to be consistent, within the errors, with the ellipticity

[1] *Lick Observatory Bulletin*, No. 941.

TABLE 1

COMPARISON OF DWARF IRREGULAR AND DWARF SPHEROIDAL GALAXIES

Galaxy	Central μ_V (mag arcsec^{-2})	Abs. V mag	Radius at 25.0 μ_V (pc)	Δv (km s^{-1})
Dwarf Irregular Galaxies: Now				
GR 8 (DDO 155) ...	21.6	−10.6	200	15
IC 1613	22.7	−15.4	1700	50
NGC 6822.....	20.8	−16.6	1300	40
Mean of five	22.4	...	...	...
Dwarf Irregular Galaxies: After Fading				
GR 8 (DDO 155) ...	25.8	−6.3	too faint	15
IC 1613	24.2	−14.6	~ 1700	50
NGC 6822.....	22.6	−14.9	$\lesssim$ 1300	40
Mean of five	24.4	...	...	...
Dwarf Spheroidal Galaxies				
Sculptor	23.9*	−10.9	240	14
Fornax	23.3*	−13.6	1100	34
Leo I	21.5:	−11.4	530	21
Leo II.....	23.9*	−9.8	160	12
Draco.....	24.6*	−8.6	30	6.4
Ursa Minor	25.0:	−8.8	too faint	4.5
Mean of values with asterisk	23.9	...	...	...

NOTES.—All photometry has been corrected for Galactic reddening and extinction. Basic data for three DI galaxies are from Hodge (1967, 1977, 1978). Objects in the mean of five are IC 1613, NGC 6822, Ho I, Ho II, and A1008−04. Central surface brightness for the last three was scaled from m_{25} in the RC2 (de Vaucouleurs, de Vaucouleurs, and Corwin 1976) using an additive correction based on IC 1613 and NGC 6822. Sources for internal motions (Δv − HWHM of H I profile) are: GR 8 (Fisher and Tully 1975); IC 1613 (Epstein 1964); NGC 6822 (Rogstad, Rougoor, and Whiteoak 1967). Basic data on DS galaxies are from Hodge (1961*a*, *b*, 1962, 1963, 1964*a*, *b*, 1966, 1971), Hodge and Smith (1974), and Hodge (1983). Central brightnesses for Fornax, Sculptor, and Leo II are photoelectric values; Draco's is inferred from star counts scaled to M3 (see Hodge 1964*b*). The other three are eye estimates from the PSS. Isophotal radii are derived from photoelectric profiles (Fornax and Sculptor) or from star counts normalized to the tabulated central surface brightness (all others). Internal motions are $\Delta v'$ from Table 1 in Faber and Lin (1983).

distributions for DI galaxies as a whole (Thuan and Seitzer 1979).

2. *Do DI galaxies themselves contain enough nonluminous matter to yield DS galaxies with high mass-to-light ratios*? That DI galaxies do appear to contain nonluminous matter has been noted previously by Tully *et al.* (1978) and by Tinsley (1981). To assess the amount quantitatively, we test whether M/L_V will have the right value after stripping and fading. We start with a prototypical DI galaxy, for which $M_{\rm gas}/M_{\rm tot}$ and $(M/L_V)_{\rm tot}$ are both known from 21 cm studies. The value of $M_*/M_{\rm tot}$ before stripping is obtained from a model stellar mass-to-light ratio (taken to be 0.4) based on *UBV* colors (Larson and Tinsley 1978). $M_{\rm non}/M_{\rm tot}$ is then the fractional remaining mass. The results in Table 2 indicate that nearly two-thirds of the total mass in DI galaxies within the Holmberg radius appears to be in nonluminous form. We further find that M/L_V after fading should be about 28 (Table 2, row [2]), consistent both with the tidal limits in Faber and Lin (1983) and with Aaronson's recent lower limit for Draco of 30 (Aaronson 1983). (The value in Table 2 has been rounded to 30 and used to calculate the internal velocities, $\Delta v'$, in Table 1 of Faber and Lin 1983.)

3. *Is it physically likely that small DI galaxies in the neighborhood of the Milky Way would have had their gas removed by ram pressure stripping*? The rate of mass loss from dwarf galaxies may be estimated quantitatively with a standard formula (Frank and Gisler 1976). For a typical DI galaxy with a mass of $2.4 \times 10^7\ M_\odot$ (Table 1, Faber and Lin 1983) and a gas radius of ~ 1 kpc, an ambient number density of $\sim 10^{-6}$ cm^{-3} is required to

TABLE 2

PREDICTED M/L_V RATIOS OF DWARF SPHEROIDAL GALAXIES

Galaxy	M_{tot} (1)	M_{gas} (2)	M_* (3)	M_{non} (4)	L_V (5)	$(M/L_V)_*$ (6)	$(M/L_V)_{tot}$ (7)
Original dwarf irregular ...	1.00	0.22[a]	0.069[b]	0.71	0.158[c]	0.40[d]	5.8[e]
Final dwarf spheroidal	0.78	0.00	0.069	0.71	0.028[c]	2.5[f]	27.9[g]

[a]Mean of $M_{H\,I}/M_{tot}$ by Fisher and Tully 1975, Thuan and Seitzer 1979, and Tully *et al.* (1978), corrected to $H_0 = 75$ km s^{-1}.
[b]Col. (6) divided by col. (7).
[c]Col. (3) divided by col. (6).
[d]Larson and Tinsley 1978 for $\langle B - V \rangle = 0.40$, $\langle U - B \rangle = -0.20$.
[e]Mean of Tinsley 1981, Faber and Gallagher 1979, and Thuan and Seitzer 1979, corrected to $H_0 = 75$ km s^{-1}.
[f]Typical globular cluster M/L_V of 2.5 assumed (Illingworth 1976; Gunn and Griffin 1979).
[g]Col. (1) divided by col. (5).

strip the gas over a Hubble time. Gas of this density must pervade most of the halo of the Milky Way at a typical distance of roughly 100 kpc.

Such a density is too low to be detected in either our own or other galaxies, even if the gas is quite hot. However, densities of about this order might be expected on indirect arguments. For example, supernovae may provide a nearly hydrostatic galactic corona with a density of $\sim 10^{-4}$ cm^{-3} a few kiloparsecs above the plane (Bregman 1980; de Boer and Savage 1982). For a constant gas temperature of $\sim 10^6$ K (the dynamical temperature of the isothermal halo), one infers a density of $\sim 10^{-6}$ cm^{-3} at 100 kpc. A similar result may be extrapolated from the recent possible detection of an X-ray corona toward the Galactic bulge (Garmire and Nugent 1981). Both of these estimates are uncertain since the density falloff is very sensitive to the actual gas temperature. However, gas densities of the required magnitude do seem plausible, though marginal.

4. *Do the stellar populations of DS galaxies show any signs of relatively recent transition from DI to DS status*? If the condition for gas removal from DS galaxies is indeed marginal, the star-formation histories of DS galaxies would be variable, depending on their orbits, gravitational potentials, and the ambient gas density they encounter. Averaged over all DS galaxies, however, the stellar populations should look considerably younger than those in globular clusters.

Evidence for youthful stars in DS galaxies is well known and includes the presence of carbon stars (see Aaronson, Olszewski, and Hodge 1983, and references therein), anomalous Cepheid variables (Zinn and Searle 1976), and red horizontal branches. Star-formation scenarios based on the bolometric luminosity distribution of carbon stars furthermore indicate widely differing star-forming histories, from galaxies like Carina (Mould *et al.* 1982) and Fornax (Aaronson and Mould 1980), which apparently underwent star formation up until just a few billion years ago, to Sculptor (Frogel *et al.* 1982), Draco (Aaronson, Liebert, and Stocke 1982), and Ursa Minor (Aaronson, Olszewski, and Hodge 1983), which have been dormant a longer period of time. There also seems to be a slight trend for more massive DS galaxies to have more youthful stars (Frogel *et al.* 1982; Aaronson, Liebert, and Stocke 1982). Such a trend would be consistent with gas removal via stripping, since larger galaxies should retain their gas longer on the average and thus have younger stars.

To conclude this discussion of DS galaxies, the evidence as it presently stands on structure, mass-to-light ratios, stellar populations, composition, and halo gas density seems to be moderately consistent with a picture in which DS galaxies were created from DI galaxies via ram pressure sweeping.

III. NEW CONSTRAINTS ON THE NATURE OF NONLUMINOUS MATTER

We next briefly sketch out several constraints that would be placed on nonluminous matter were it to exist in large quantities within dwarf galaxies. One possible constituent of nonluminous matter may be massive neutrinos. Tremaine and Gunn (1978) have derived a limit on the neutrino mass based on the constancy of their phase-space density since cosmological times. Applying their formula (6) with $\sigma = \Delta v/\sqrt{2}$ from Table 1 and r_c from Hodge (1971)[2] yields the following lower limits on m_ν: 150 eV (Fornax), 500 eV (Ursa Minor), 580 eV (Carina), and 600 eV (Draco). These limits are based on an assumed M/L_V of 30. Aaronson's (1983) actual dispersion measurement of 11.3 km s^{-1} for Draco would imply $m_\nu \gtrsim 500$ eV.

These lower limits on m_ν are in strong contradiction to the most conservative limit of 30–100 eV set by the overall mass density of the universe (see Schramm and Steigman 1981). The proven existence of large amounts

[2]For Carina, $2\bar{r}_1$ from Faber and Lin (1983) was used.

of nonluminous matter in small galaxies would therefore suffice to rule out massive neutrinos as a major constituent of nonluminous matter in the universe. Aaronson (1983) has reached similar conclusions.

Of great importance furthermore is the *precise ratio* of ordinary to nonluminous matter in dwarf galaxies. In the fading picture, DS galaxies start out with the same fractional amount of ordinary matter as DI galaxies, or $\sim 30\%$ within the H I radius. Since the nonluminous halo doubtless extends somewhat beyond this point, the ratio of ordinary to nonluminous matter averaged over the *whole* galaxy is probably comparable to that in bigger galaxies and in the universe as a whole, namely $\sim 7\%$ (Faber 1982; Gunn 1982).

The above result places strong constraints on galaxy formation models. The standard particle physics model for the early universe sets an upper limit of ~ 1 keV on the mass of any noninteracting particle (Blumenthal Pagels, and Primack 1982). At 1 keV, the Jeans mass is $\sim 10^{12}\ M_\odot$. Below this mass, only slowly moving, nonrelativistic particles would be bound to density perturbations, and the trapping of nonluminous matter in small galaxies should therefore be drastically reduced. In order to have a Jeans mass comparable to that of the dwarf galaxies, say $\lesssim 10^8\ M_\odot$, a particle mass of $\gtrsim 10$ keV is required (see Fig. 1 in Blumenthal *et al.*). However, the existence of such massive noninteracting particles would necessitate fundamental changes to the standard particle physics model (Blumenthal *et al.*).

Alternatively, noninteracting particles may be trapped indirectly by baryonic inhomogeneities in galaxies after the collapse of larger entities, e.g., superclusters (Doroshkevich *et al.* 1981). Current numerical experiments to evaluate the efficiency of trapping by large galaxies within superclusters put it at roughly 10% (J. R. Bond, private communication). Since the efficiency for satellite dwarf galaxies would probably be even smaller, the above scenario may be inadequate to account for the fact that dwarf galaxies seem to be *just as efficient* at trapping nonluminous matter as are more massive objects.

Dwarf galaxies also place important constraints on theories in which nonluminous matter is made up of discrete, massive objects such as black holes. In this case, core collapse would occur on a time scale comparable to the relaxation time (Spitzer and Hart 1971*b*; Lightman and Fall 1978). Inserting values for the mass, velocity dispersion, and density of a typical DS galaxy into a standard formula (Spitzer and Hart 1971*a*), we find that the core collapse time is longer than the Hubble time only if the halo consists mainly of black holes which are more massive than $\sim 100\ M_\odot$. Since the smooth central luminosity profiles of DS galaxies show no sign of core collapse, black holes or other condensed remnants, if they exist, cannot exceed this mass.

IV. CONCLUSIONS

We conclude with a resumé of badly needed observations on dwarf galaxies. Most important is conclusive evidence one way or the other for the existence of nonluminous matter in these objects. Such evidence would come from the study of 21 cm rotation curves or of stellar velocity dispersion. Next in importance is the dependence of dwarf galaxy properties on mass and environment. In regions where the ambient gas density is high or where galaxies are moving rapidly, gas may be swept from even relatively massive galaxies. DS galaxies in such regions might therefore have high luminosities, high surface brightnesses, and perhaps even show distinct nuclei. The overall DI galaxy population should also be reduced. There actually exists preliminary evidence for such trends among dwarf companions of nearby massive spirals (Einasto *et al.* 1974). Finally, more profile measurements of dwarfs are needed to test whether similar profiles and ellipticity distributions typify both DS and DI classes. Many of these observations are already underway and should provide important new information on dwarf galaxies in the not too distant future.

In conclusion, we point out that our discussion here and in Faber and Lin (1983) is oriented toward the actual existence of nonluminous matter in dwarf galaxies. It should be emphasized, however, that an eventual finding of *little or no* nonluminous matter in dwarf galaxies would have essentially an equal importance to theories of cosmology and galaxy formation.

It is a pleasure to thank the following colleagues for helpful discussions: George Blumenthal, Paul Hodge, Joel Primack, Edwin Salpeter, and Renzo Sancisi. Robert Kraft shared valuable information on the nature of stellar populations in dwarf spheroidal galaxies. Special thanks go to Allan Sandage for valuable discussions and for permission to view his extraordinary photographs of dwarf galaxies in the Virgo Cluster. We are also greatly indebted to Marc Aaronson, whose first measurement of the velocity dispersion in Draco added considerably to the interest of these *Letters*. Last, we would like to pay a special tribute to Paul Hodge, whose painstaking, 20 year program of observations on dwarf galaxies provided the stimulus for most of this research and the fundamental data for many of our conclusions.

REFERENCES

Aaronson, M. 1983, *Ap. J.* (*Letters*), **266**, L11.
Aaronson, M., Liebert, J., and Stocke, J. 1982, *Ap. J.*, **254**, 507.
Aaronson, M., and Mould, J. 1980, *Ap. J.*, **240**, 804.
Aaronson, M., Olszewski, E. W., and Hodge, P. W. 1983, *Ap. J.*, submitted.

Blumenthal, G. R., Pagels, H., and Primack, J. R. 1982, *Nature*, **299**, 37.
Bregman, J. N. 1980, *Ap. J.*, **236**, 577.
Cannon, R. D., Niss, B., Nørgaard-Nielsen, H. U., 1981, *M.N.R.A.S.*, **196**, 1P.
de Boer, K. S., and Savage, B. P. 1982, preprint.
de Vaucouleurs, G., de Vaucouleurs, A., and Corwin, H. 1976, *Second Reference Catalogue of Bright Galaxies* (Austin: University Texas Press).
Doroshkevich, A. G., Khlopov, M. Yu., Sunyaev, R. A., Szalay, A. S., and Zeldovich, Ya. B. 1981, *Ann. NY Acad. Sci.*, **375**, 32.
Einasto, J., Saar, E., Kaasik, A., and Chernin, A. D. 1974, *Nature*, **252**, 111.
Epstein, E. E. 1964, *A.J.*, **69**, 1964.
Faber, S. M. 1982, in *Astrophysical Cosmology: Proceedings of the Vatican Study Week on Cosmology and Fundamental Physics*, ed. H. A. Brück, G. V. Coyne, and M. S. Longair (Vatican City: Pontifical Scientific Academy), p. 191.
Faber, S. M., and Gallagher, J. S. 1979, *Ann. Rev. Astr. Ap.*, **17**, 135.
Faber, S. M., and Lin, D. N. C. 1983, *Ap. J. (Letters)*, **266**, L17.
Fisher, J. R., and Tully, R. B. 1975, *Astr. Ap.*, **44**, 151.
Frank, J., and Gisler, G. 1976, *M.N.R.A.S.*, **176**, 533.
Frogel, J. A., Blanco, V. M., McCarthy, M. F., and Cohen, J. G. 1982, *Ap. J.*, **252**, 133.
Garmire, G. P., and Nugent, J. J. 1981, *Bull. AAS*, **13**, 786.
Gunn, J. E. 1982, in *Astrophysical Cosmology: Proceedings of the Vatican Study Week on Cosmology and Fundamental Physics*, ed. H. A. Brück, E. V. Coyne, and M. S. Longair (Vatican City: Pontifical Scientific Academy), p. 233.
Gunn, J. E., and Griffin, R. F. 1979, *A.J.*, **84**, 752.
Hodge, P. W. 1961*a*, *A.J.*, **66**, 249.
_____. 1961*b*, *A.J.*, **66**, 384.
_____. 1962, *A.J.*, **67**, 125.
_____. 1963, *A.J.*, **68**, 470.
_____. 1964*a*, *A.J.*, **69**, 438.
_____. 1964*b*, *A.J.*, **69**, 853.
_____. 1966, *A.J.*, **71**, 204.
_____. 1967, *Ap. J.*, **148**, 719.
_____. 1971, *Ann. Rev. Astr. Ap.*, **9**, 35.
_____. 1977, *Ap. J. Suppl.*, **33**, 69.
_____. 1978, *Ap. J. Suppl.*, **37**, 145.
_____. 1983, preprint.
Hodge, P. W., and Smith, D. W. 1974, *Ap. J.*, **188**, 19.
Illingworth, G. 1976, *Ap. J.*, **204**, 73.
Larson, R. B., and Tinsley, B. M. 1978, *Ap. J.*, **219**, 46.
Lightman, A. P., and Fall, S. M. 1978, *Ap. J.*, **221**, 567.
Mould, J. R., Cannon, R. D., Aaronson, M., and Frogel, J. A. 1982, *Ap. J.*, **254**, 500.
Rogstad, D. H., Rougoor, G. W., and Whiteoak, J. B. 1967, *Ap. J.*, **150**, 9.
Schramm, D. N., and Steigman, G. 1981, *Ap. J.*, **243**, 1.
Spitzer, L., and Hart, M. 1971*a*, *Ap. J.*, **164**, 399.
_____. 1971*b*, *Ap. J.*, **166**, 483.
Thuan, T. X., and Seitzer, P. O. 1979, *Ap. J.*, **231**, 680.
Tinsley, B. M. 1981, *M.N.R.A.S.*, **194**, 63.
Tremaine, S., and Gunn, J. E. 1978, *Phys. Rev. Letters*, **42**, 407.
Tully, R. B., Bottinelli, L., Fisher, J. R., Gouguenheim, L., Sancisi, R., and van Woerden, H. 1978, *Astr. Ap.*, **63**, 37.
Zinn, R., and Searle, L. 1976, *Ap. J.*, **209**, 734.

S. M. FABER and D. N. C. LIN: Lick Observatory and Board of Studies in Astronomy and Astrophysics, University of California, Santa Cruz, CA 95064

THE ASTROPHYSICAL JOURNAL, 274:L1–L5, 1983 November 1
© 1983. The American Astronomical Society. All rights reserved. Printed in U.S.A.

CLUSTERING IN A NEUTRINO-DOMINATED UNIVERSE

SIMON D. M. WHITE,[1,2] CARLOS S. FRENK,[1] AND MARC DAVIS[1,3]
University of California, Berkeley
Received 1983 June 17; accepted 1983 July 1

ABSTRACT

We have simulated the nonlinear growth of structure in a universe dominated by massive neutrinos using initial conditions derived from detailed linear calculations of earlier evolution. Codes based on a direct N-body integrator and on a fast Fourier transform Poisson solver produce very similar results. The coherence length of the neutrino distribution at early times is directly related to the mass of the neutrino and thence to the present density of the universe. We find this length to be too large to be consistent with the observed clustering scale of galaxies if other cosmological parameters are to remain within their accepted ranges. The conventional neutrino-dominated picture appears to be ruled out.

Subject headings: cosmology — galaxies: clustering — neutrinos

I. INTRODUCTION

In an earlier paper (Frenk, White, and Davis 1983, hereafter Paper I) we presented simulations of the growth of structure in a universe where the density distribution at early times possessed a large coherence length. In agreement with earlier work (Klypin and Shandarin 1981) we found that the nonlinear structure which grows from such conditions retains considerable large-scale coherence. It has a filamentary appearance reminiscent of the observed galaxy distribution (e.g., Einasto, Joeveer, and Saar 1980; Davis *et al.* 1982); this contrasts sharply with the structure in simulations started from white-noise initial conditions (Aarseth, Gott, and Turner 1979; Efstathiou and Eastwood 1981). A population of massive neutrinos could dominate the mass budget of the present universe and would possess just such a large initial coherence length (Doroshkevich *et al.* 1980; Bond, Efstathiou, and Silk 1980). Detailed linear calculations of the evolution of density fluctuations in a neutrino gas have been carried out by Peebles (1982) and Bond and Szalay (1983). Scaling our simulations to their results, we found a unique time at which the particle autocorrelation function matched that observed for galaxies. Three difficulties marred this apparent success. Our scaling required a Hubble parameter of 80 km s^{-1} Mpc^{-1} which, in a flat universe, implies an uncomfortably low age. Our model had the right autocorrelation function after expansion by a factor of 2.5 from the first collapse of structure; this is too early to be consistent with observations of quasars at redshifts approaching 4. Finally, the model autocorrelation function should be compared with that of neutrinos rather than galaxies in the real universe. Galaxies form only in dense, collapsed regions in the neutrino-dominated picture; as we shall see, this makes its large coherence length more difficult to reconcile with the relatively small scale of observed galaxy clustering.

This *Letter* presents extensions of our previous work. We consider open universe models, models with an initial fluctuation distribution derived from detailed linear calculations, and models which use an entirely different simulation method. These experiments allow us to check our numerical techniques and to investigate the three difficulties listed above.

II. SIMULATION METHODS

Our earlier calculations followed the motion of 1000 representative particles with a direct N-body integrator. To obtain a large initial coherence length, we placed our particles on a uniformly expanding cubic lattice with a spherical outer boundary; we then perturbed their positions and velocities with a superposition of random waves. Following previous work (Klypin and Shandarin 1981) we required that the power density of the imposed waves should be flat between upper and lower wavenumber cutoffs. We chose $k_{\min} = 0.6\,k_{\max}$. This fluctuation spectrum is a poor representation of that expected after recombination in a neutrino-dominated universe. The latter is quite well fitted by

$$|\delta_k|^2 \propto k^n 10^{-2(k/k_\nu)^{1.5}}, \tag{1}$$

where

$$\begin{aligned}\lambda_\nu &= 2\pi/k_\nu \\ &= 41 m_{30}^{-1}\theta^{-1}\ \text{Mpc} \\ &= 13\Omega^{-1}h^{-2}\theta^2\ \text{Mpc},\end{aligned} \tag{2}$$

[1]Department of Astronomy.
[2]Space Sciences Laboratory.
[3]Department of Physics.

m_{30} is the neutrino mass in units of 30 eV, Ω is the cosmological density parameter, h is the Hubble parameter in units of 100 km s^{-1} Mpc^{-1}, and θ is the microwave background temperature in units of 2.7 K (Bond and Szalay 1983). The wavelengths here are comoving and referred to the present day. The exponent n in equation (1) is set by whatever process originally produced the fluctuations. The value $n = 1$ is often favored because the resulting curvature fluctuations all enter the particle horizon with equal amplitude; this is the spectrum predicted by inflationary models for the formation of irregularities (Barrow and Turner 1982). We have now performed simulations with the same direct N-body integrator using initial conditions in which the perturbing waves satisfy equation (1) between upper and lower cutoffs; these were chosen as $k_\nu = 0.5\ k_{max} = 4\ k_{min} = 1.6\ 2\pi/R$, where R is the radius of the region simulated. We have run models in which $n = 0$, 1, and 4. These new calculations can be used to check the scaling of our earlier models. As before, in order to estimate statistical uncertainties, we formed ensembles from four models run from different realizations of each set of initial conditions.

The experiments described above all model the evolution of an Einstein–de Sitter universe. In order to investigate how clustering depends on cosmological density, we multiplied the particle mass by a factor of 0.8 in the initial conditions of the eight pancake simulations of Paper I and then reran their evolution. This produces two ensembles of open universes which can be compared directly with the original models.

The results of these experiments may be uncertain because of the relatively small number of particles we used and the way we distributed them on a grid. In collaboration with G. Efstathiou we have now developed a number of simulation codes in which the interparticle force is calculated using fast Fourier transform methods. Details of these codes and of numerous tests of their performance will be published elsewhere. They allow us to study the sensitivity of our results to numerical technique. We have performed a number of simulations which follow the motion of 32,768 particles within a cubic cell of a periodic universe. Forces are calculated using a momentum-conserving cloud-in-cell scheme on a 64^3 density and potential grid (Hockney and Eastwood 1981). We set up initial conditions by placing the particles either on a grid or at random and then perturbing their positions and velocities with a superposition of growing mode fluctuations. For these experiments we chose λ_ν equal to 20% of the side of the cube and set the amplitude in equation (1) to correspond to an rms density fluctuation of 22%. These experiments are very similar to that of Klypin and Shandarin (1981), but our force calculation has higher resolution than theirs, and our initial conditions are a much closer approximation to those of a neutrino-dominated universe. We have run ensembles with $n = 1$ and both grid and random initial conditions, and a grid ensemble with $n = 4$. Each ensemble contains four models. A model takes 1.5–3 hr to execute on a VAX 11/780.

III. RESULTS

In an open universe, linear fluctuations grow more slowly than in a flat universe. We compare our ensembles of open models with their Einstein–de Sitter counterparts at times chosen to match the linear growth factor from the two sets of initial conditions; linear theory then predicts that the particle distributions should be identical. This expectation is remarkably well obeyed on large scales. Even after close inspection, equidensity plots for the open models are indistinguishable from those of Paper I. In Figure 1 we compare the correlation functions of an open ensemble and of the corresponding ensemble from Paper I. The agreement is excellent for values of $\xi < 10$; as expected, clustering on smaller scales is stronger in the open ensemble. Velocity correlations in the open models have very similar form to those shown in Paper I, but their amplitude is smaller. When spatial correlations are matched as above, peculiar velocities scale by the expected factor of $\sqrt{\Omega}$.

Figure 1 also compares a correlation function from our earlier work with results from new Einstein–de Sitter models with initial conditions based on equation (1). Length scales were matched by equating the minimum wavelength present in the earlier calculations to $1.19\ \lambda_\nu$, as is appropriate for $n = 1$ (see Paper I and Peebles 1982). The correlation functions agree well in shape, but there is a slight scale discrepancy which reflects the uncertainty arising from the substantial difference between the initial fluctuation spectra. Our two integration schemes give results which agree very well; both methods seem entirely adequate for the problem at hand. In tests to be published elsewhere, we have found the correlation function of a Fourier model to be affected by potential softening only on scales somewhat smaller than the mesh size of the potential grid. The correlation functions of the two Fourier ensembles in Figure 1 are almost identical. This demonstrates the insensitivity of our results to the grid initial conditions of most of our models. It also shows that a long-wavelength feature with a peak power an order of magnitude above the white-noise level can imprint coherent large-scale structure on a universe which had white-noise initial conditions on small scales.

In a "standard" neutrino-dominated universe, galaxies, and presumably quasars, form once large-scale structures begin to collapse. We identify the onset of galaxy formation in our models as the time when 1% of the elemental volumes of our initial grid have undergone collapse; most galaxies are likely to form well after this. We parameterize later times by the expansion factor,

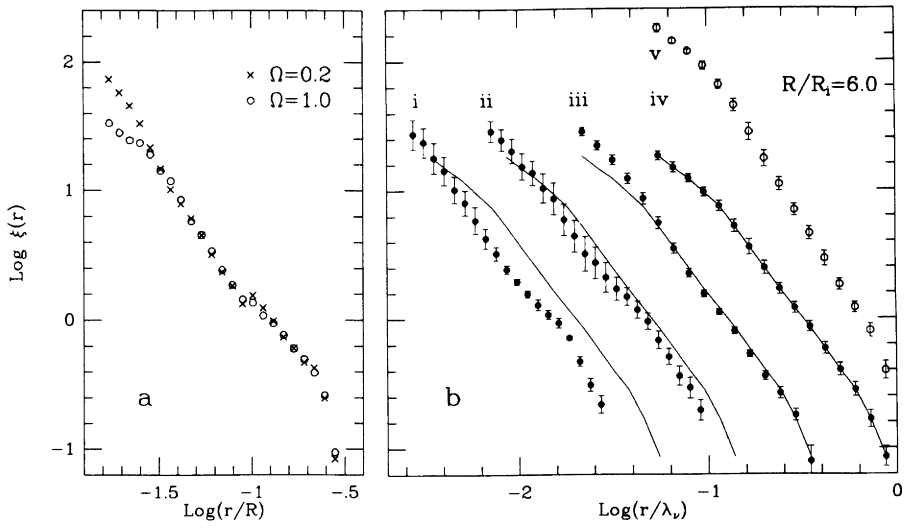


FIG. 1.—(*a*) The correlation function of ensemble A of Paper I (*open circles*) is compared with the correlation function of an open ensemble which initially had the same particle positions and velocities but smaller particle masses (*crosses*). The data for ensemble A are shown at an expansion factor of 6, while the data for the open ensemble are shown at an expansion factor of 15.6 when $\Omega = 0.2$ and the linear growth factor from the initial conditions is 6. (*b*) Correlation functions for four Einstein–de Sitter ensembles are compared at an expansion factor of 6. Plot (i) is ensemble B of Paper I; (ii) is an ensemble of 1000 particle direct *N*-body simulations with initial conditions derived from eq. (1) with $n = 1$; (iii) and (iv) are ensembles of 32,768 particle Fourier simulations with this same perturbation spectrum imposed on random and grid underlying particle distributions respectively; (v) is the "galaxy" function for the Fourier grid ensemble. The abscissa is correct for (iv) and (v), but for clarity the other three functions have been shifted to the left by multiples of 0.4. The solid lines show similarly shifted versions of (iv) superposed on each of the "neutrino" correlation functions. Error bars give the standard deviation in the plotted points derived using the four simulations in each ensemble.

$1 + z_{GF}$, since this first collapse, and we measure the scale, r_c, of nonlinear clustering using the integral relation

$$r_c^{-3} \int_0^{r_c} \xi(r) r^2 \, dr = 1. \tag{3}$$

In our models, this length can be expressed in units of the initial coherence length, λ_ν. The observed value of r_c for the galaxy distribution is $5.0 \pm 0.7\ h^{-1}$ Mpc (Davis and Peebles 1983*a*), and the value of λ_ν for a neutrino-dominated universe is related to the observable parameters Ω, h, and θ through equation (2). Equating the model value of r_c/λ_ν to the observational prediction thus gives a relation between the cosmological parameters

$$(r_c/\lambda_\nu)_{\text{model}} = 0.39 \pm 0.06 \Omega h \theta^{-2}. \tag{4}$$

For any ensemble of models, Ω and r_c/λ_ν are functions of time and thus of z_{GF}. For any particular value of z_{GF}, equation (4) then gives the value of $h\theta^{-2}$ required for consistency with observation.

In Figure 2 we plot this relation for three of our ensembles of simulations. For Ω significantly smaller than 1, the length scale of a neutrino-dominated universe is much too large to be compatible with the relatively small scale of galaxy clustering; this is reflected in the very large values of $h\theta^{-2}$ required for "consistency" with observation. In a flat universe the conservative requirement that $z_{GF} > 3$ requires $h\theta^{-2} > 1$ unless $n > 3$. This forces a choice between an uncomfortably young universe and an initial power spectrum with substantially more power on small scales than the constant curvature spectrum which is now generally favored; if such a spectrum were indeed present at early times, the universe would have been chaotic on the horizon scale well after the epoch of nucleosynthesis.

These comparisons are somewhat misleading because they identify the model correlation function with the observed galaxy function. In fact, galaxies form only in dense "pancake" shocks in the neutrino picture, and the low-density regions between pancakes should be devoid of observable systems. In such a situation the correlation length of the galaxy distribution is expected to be greater than that of the neutrinos, thus exacerbating scaling problems. To illustrate this we have calculated correlation functions for our models assuming that particles are "visible" only after their cell of the initial particle grid has collapsed. In Figure 1 we plot such a "galaxy"

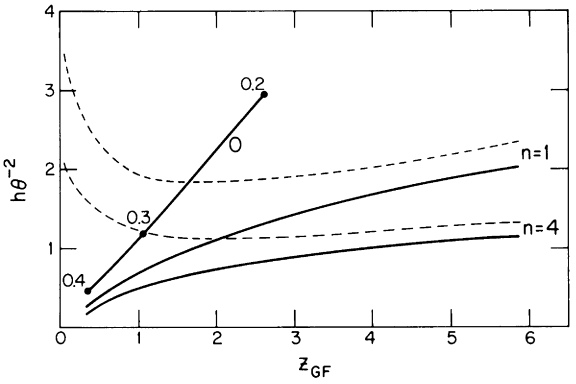


FIG. 2.—The value of $h\theta^{-2}$ required for the coherence length of a neutrino-dominated universe to be consistent with the observed scale of galaxy clustering is plotted as a function of redshift since the onset of galaxy formation for three kinds of initial condition. The line marked O refers to the open version of ensemble A of Paper I scaled assuming $n = 1$. The value of Ω is shown at three points of its evolution. The solid lines refer to two Fourier ensembles with initial conditions given by eq. (1), while the dashed lines show the required parameters if the "galaxy" correlation function of these same ensembles (rather than their "neutrino" function) is matched to the observations.

correlation function for comparison with the corresponding total correlation function. The "galaxy" function has a higher amplitude than the "neutrino" function and has a large comoving clustering length which actually decreases until the time when half the particles have become visible. Using equation (4) we can derive the scaling to observation implied by this length; its evolution is plotted in Figure 2 for Fourier ensembles with $n = 1$ and 4. If $\Omega \leq 1$, agreement with observation is not possible for any acceptable values of h, θ, and n.

IV. CONCLUSIONS

We have used two quite different numerical techniques to carry out simulations of the nonlinear growth of structure in a neutrino-dominated universe. Our initial conditions reproduce the perturbation spectra obtained from detailed linear calculations. We find that the predicted neutrino clustering scale in such a universe substantially exceeds the observed scale of galaxy clustering. Agreement can be obtained only by going well outside the range of currently fashionable cosmological parameters. If $\Omega > 2$, acceptable values of the Hubble constant, the redshift of galaxy formation, and the age of the universe are obtained for a constant curvature initial power spectrum ($n = 1$). If $\Omega \approx 1$, acceptable values are obtained only for $n > 3$, implying a highly inhomogeneous universe at quite recent epochs. If the universe is open by a significant margin, no conventional neutrino-dominated cosmology is possible. This scale discrepancy is exacerbated by the requirement that galaxies form in "pancakes" and thus avoid low-density regions; it could be alleviated only if galaxy formation were strongly suppressed in regions of high neutrino density. Figure 2 shows that scaling difficulties cannot be avoided by assuming galaxy formation to switch off after the collapse of the first pancakes.

Unconventional cosmological assumptions might rescue the neutrino-dominated picture. Increasing the number of massive neutrino species makes things worse, since the neutrino coherence length is then larger for given values of Ω and h. Allowing a highly inhomogeneous universe at relatively recent epochs might help by providing extra entropy generation and so increasing the photon-to-neutrino ratio; in addition it might produce a steep fluctuation spectrum with $n \approx 4$. (Other ways to increase the photon-to-neutrino ratio or otherwise salvage the neutrino picture are discussed by Davis *et al.* 1981). A positive cosmological constant might make a high-density universe consistent with observational constraints on the deceleration parameter and the age of globular clusters; however, it does not affect upper bounds on Ω from the dynamics of the Local Supercluster (cf. Davis and Peebles 1983*b*). Finally, large-amplitude isothermal perturbations in the baryon component of the universe might allow galaxy formation to precede pancake formation; the latter might then occur recently enough to be consistent with the observed clustering. None of these alternatives will seem attractive unless direct measurements of a neutrino mass force us to consider them. Until then, the discrepancy between the large coherence length of neutrino-dominated universes and the small scale of observed galaxy clustering makes it appear unlikely that neutrinos provide the missing mass.

The Fourier simulations reported above would not have been possible without the collaboration of George Efstathiou, who developed the first version of our FFT code. We wish to thank Alex Szalay for useful discussions. This work was supported by DOE contract DE-AT03-82ER40069 and NSF grant AST-8118557.

REFERENCES

Aarseth, S. J., Gott, J. R. and Turner, E. L. 1979, *Ap. J.*, **228**, 664.
Barrow, J. D., and Turner, M. S. 1982, *Nature*, **298**, 801.
Bond, J. R., Efstathiou, G., and Silk, J. I. 1980, *Phys. Rev. Letters*, **45**, 1980.
Bond, J. R., and Szalay, A. S. 1983, *Ap. J.*, **272**, in press.
Davis, M., Huchra, J., Latham, D., and Tonry, J. 1982, *Ap. J.*, **253**, 423.
Davis, M., Lecar, M., Pryor, C., and Witten, E. 1981 *Ap. J.*, **250**, 423.
Davis, M., and Peebles, P. J. E. 1983*a*, *Ap. J.*, **267**, 465.
_____. 1983*b*, *Ann. Rev. Astr. Ap.*, **21**, in press.
Doroshkevich, A. G., Khlopov, M. Yu., Sunyaev, R. A., Szalay, A. S., and Zel'dovich, Ya. B. 1980, *Ann. NY Acad. Sci.*, **375**, 32.
Efstathiou, G., and Eastwood, J. W. 1981, *M.N.R.A.S.*, **194**, 503.
Einasto, J., Joeveer, M., and Saar, E. 1980, *M.N.R.A.S.*, **193**, 353.
Frenk, C. S., White, S. D. M., and Davis, M. 1983, *Ap. J.*, **271**, 417 (Paper I).
Hockney, R. W., and Eastwood, J. W. 1981, *Computer Simulation Using Particles* (New York: McGraw-Hill).
Klypin, A. A., and Shandarin, S. F. 1981, preprint.
Peebles, P. J. E. 1982, *Ap. J.*, **258**, 415.

MARC DAVIS, CARLOS S. FRENK, and SIMON D. M. WHITE: Department of Astronomy, University of California, Berkeley, CA 94720

THE ASTROPHYSICAL JOURNAL, 263:L1–L5, 1982 December 1
© 1982. The American Astronomical Society. All rights reserved. Printed in U.S.A.

LARGE-SCALE BACKGROUND TEMPERATURE AND MASS FLUCTUATIONS DUE TO SCALE-INVARIANT PRIMEVAL PERTURBATIONS

P. J. E. PEEBLES
Joseph Henry Laboratories, Physics Department, Princeton University
Received 1982 July 2; accepted 1982 August 13

ABSTRACT

The large-scale anisotropy of the microwave background and the large-scale fluctuations in the mass distribution are discussed under the assumptions that the universe is dominated by very massive, weakly interacting particles and that the primeval density fluctuations were adiabatic with the scale-invariant spectrum $P \propto$ wavenumber. This model yields a characteristic mass comparable to that of a large galaxy independent of the particle mass, m_x, if $m_x \gtrsim 1$ keV. The expected background temperature fluctuations are well below present observational limits.

Subject headings: cosmic background radiation — cosmology — galaxies: formation

I. INTRODUCTION

It is useful to consider which scenarios for the nature and evolution of the mass distribution in the universe can fit the observations without undue contrivance. We may hope that as the observations improve the list of candidates will narrow, and that this process may in time help guide us to a fundamental theory of the origin of structure in the universe. The picture discussed here is motivated by the argument that, if the initial conditions for conventional classical cosmology were set at some exceedingly high redshift (perhaps the Planck time or the grand unified theory epoch), and if the initial conditions did not involve exceedingly large or small numbers, then the cosmological density parameter ought to be $\Omega = 1$, and, assuming adiabatic density perturbations, the power spectrum P ought to be proportional to the wavenumber k (e.g., Hawking 1982). Density fluctuations on scales greater than the horizon are defined as in Peebles (1980, § 91, hereafter LSS). If $P \propto k$, perturbations to the geometry diverge only as $\log k$ so the cutoffs can be at very large and small k and the spectrum can be truly scale invariant (Harrison 1970; Peebles and Yu 1970, § VI*a*; Zel'dovich 1972). Density fluctuations appearing on the horizon have a fixed value, $\delta M/M \sim 10^{-4}$, and, as this number is not greatly different from unity, we might imagine it is fixed by fundamental physics.

Observational constraints include the measurements of background temperature fluctuations, $\delta T/T \lesssim 1 \times 10^{-4}$ (Boughn, Cheng, and Wilkinson 1981; Melchiorri *et al.* 1981) and the estimate of the coherence length of the galaxy distribution. They tell us $P \propto k$ is unacceptable in the usual cosmology with baryons, electrons, radiation, and massless neutrinos. The observations of $\delta T/T$ imply $\epsilon \lesssim 1 \times 10^{-4}$. The wanted growth factor $\sim 10^4$ to make nonlinear density fluctuations form by the present epoch is obtained if $\Omega \sim 1$ (Silk and Wilson 1981) because density fluctuations on scales greater than the matter-radiation Jeans length, λ_x, grow before decoupling. However, λ_x is large so it makes the mass autocorrelation function unacceptably broad (Silk and Wilson 1981; Peebles 1981*a*). Press and Vishniac (1980) emphasized that there is no hope for the development of appreciable density fluctuations on scales smaller than λ_x.

The problem is relieved if the universe is dominated by massive, weakly interacting particles because density fluctuations on small scales can grow before decoupling. This effect has been widely discussed in the case that the weakly interacting particle mass, m_x, is some tens of electron volts (e.g., Doroshkevich *et al.* 1981 and references therein). If $\Omega \sim 1$, the mass coherence length is broad but perhaps not unacceptable (Peebles 1982*a*). The case $m_x \sim 1$ keV is discussed by Bond, Szalay, and Turner (1982) and Blumenthal, Pagels, and Primack (1982). I discuss here a particularly simple and perhaps important limiting case, $m_x \gtrsim 1$ keV. The main results are the spectrum of mass fluctuations, which seems quite reasonable for the production of galaxies and clusters of galaxies, the statistical character of the background temperature fluctuations, and the expected size of the mass density anticorrelation at large separations.

II. CALCULATION

I assume zero cosmological constant and $\Omega = 1$, the mass being mainly in weakly interacting particles, mass m_x. Following Davis *et al.* (1981), I take the particle distribution in phase space at $v \ll c$ in the absence of

perturbations to be

$$\mathfrak{N} = (\exp m_x v c/kT_x \pm 1)^{-1}, \quad (1)$$

meaning the particles thermally decoupled from the radiation when the particles were relativistic. The parameter T_x is adjusted to make $\Omega = 1$. Hubble's constant and the present radiation temperature are written as

$$H = 100\,h\ \mathrm{km\ s^{-1}\ Mpc^{-1}}, \quad T = 2.7\,\tau\ \mathrm{K}. \quad (2)$$

a) Free Streaming

The smoothing of the mass distribution by the motions of the free particles has been widely discussed (e.g., Wasserman 1981). For either sign in equation (1), one finds that, to good accuracy, the rms peculiar velocity is

$$v = (12)^{1/2}\pi^{2/3}\hbar\rho_c^{1/3}m_x^{-4/3}(1+z), \quad (3)$$

where $v \ll c$, ρ_c is the Einstein–de Sitter density, and z is the redshift. The rms coordinate displacement is

$$r = \int_0^{t_0} dt\, v(t)(1+z), \quad (4)$$

expressed in units of proper displacement at the present epoch. The dominant part of this integral comes between the epochs z_1, where v first drops appreciably below c, and z_{eq}, where $\rho_r \sim \rho_x$ (the universe becomes matter dominated). Taking account only of this interval, one gets (for $h = \tau = 1$)

$$r \sim (0.8 + 0.3\log m_x)m_x^{-4/3}\ \mathrm{Mpc}, \quad (5)$$

where m_x is expressed in units of kilo-electron volts. Thus, if $m_x \gtrsim 1$ keV, thermal motions are unimportant on the characteristic scale, $r \sim 3\ h^{-1}$ Mpc, of large galaxies. This is seen also in Figure 2 of Peebles (1982*a*), Figure 1 of Bond, Szalay, and Turner (1982), and Figure 1 of Blumenthal, Pagels, and Primack (1982).

b) Spectrum of Mass Distribution

When thermal motions are negligible, the free particles can be described as a zero pressure ideal fluid. At high redshift, the density fluctuations are supposed to be adiabatic with spectrum $P \propto k$. The part that appears within the horizon while $\rho_r > \rho_x$ stops growing until $\rho_r \sim \rho_x$ (because the radiation distribution is oscillating as an acoustic wave). This tilts the mass spectrum to $P \propto k^{-3}$ at large k (LSS § 92D). At small k, the spectrum keeps its primeval shape, $P \propto k$. The results of a numerical integration of the shape of the spectrum between these limiting cases (Peebles 1982*a*) are approximated by the formulae

$$P = Ak\left(1 + \alpha k + \beta k^2\right)^{-2},$$

$$\alpha = 6(\tau/h)^2\ \mathrm{Mpc}, \quad \beta = 2.65(\tau/h)^4\ \mathrm{Mpc}^2. \quad (6)$$

Here and below, k is expressed in units of radians per megaparsec at the present epoch.

c) Large-Scale Background Temperature Fluctuations

If the primeval density fluctuations are adiabatic, the angular distribution of the background temperature is (LSS eq. [93.35])

$$T(\theta,\phi)/T_b - 1 = -\tfrac{1}{2}H^2\sum k^{-2}\delta_{\boldsymbol{k}}\exp i\boldsymbol{k}\cdot\boldsymbol{x}. \quad (7)$$

This equation assumes that the density fluctuations on scales of interest appear on the horizon when $\rho_r < \rho_x$ and that the universe subsequently has behaved like the Einstein–de Sitter model. The term $\propto Y_1^m$ due to our peculiar motion is not included. The vector $\boldsymbol{x}$ with length $2cH^{-1}$ points in the direction of observation. In the linear perturbation approximation, the present mass density is

$$\rho = \rho_b\left(1 + \sum\delta_{\boldsymbol{k}}\exp i\boldsymbol{k}\cdot\boldsymbol{r}\right). \quad (8)$$

The expansion of the background temperature in spherical harmonics is

$$T(\theta,\phi) = T_b\left(1 + \sum a_l^m Y_l^m\right),$$

$$a_l^m = -2\pi i^l H^2\sum_{\boldsymbol{k}} k^{-2}\delta_{\boldsymbol{k}} j_l(kx)Y_l^m(\Omega_{\boldsymbol{k}}). \quad (9)$$

With the normalization

$$\sum_{\boldsymbol{k}} \rightarrow \int k^2\,dk\,d\Omega_{\boldsymbol{k}}/4\pi, \quad (10)$$

we get

$$a_l^2 \equiv \langle|a_l^m|^2\rangle = \pi H^4\int_0^\infty dk\,k^{-2}|\delta_k|^2 j_l(kx)^2. \quad (11)$$

At large angular scales, $|\delta_k|^2 \propto k$, and the integral is

$$\int_0^\infty \frac{dy}{y} j_l(y)^2 = [2l(l+1)]^{-1}. \quad (12)$$

It is convenient to express the normalization in terms of the quadrupole moment a_2. Then (eq. [6]),

$$A = 12(a_2)^2/(\pi H^4), \quad (13)$$

and the angular power spectrum of the background is

$$(a_l)^2 = 6(a_2)^2/[l(l+1)]. \quad (14)$$

The s-wave part a_0 diverges, corresponding to the logarithmic divergence in the primeval spectrum $P \propto k$. The p-wave does not include the effect of our peculiar motion.

The autocorrelation function of the background temperature after the p-wave has been eliminated is (LSS § 46)

$$w(\theta_{12}) = \langle T_1 T_2 \rangle / \langle T \rangle^2 - 1$$

$$= \sum_{l>1} (a_l)^2 (2l+1) P_l(\cos\theta_{12})/4\pi. \quad (15)$$

At $\theta_{12} \ll 1$ radian, this with equation (14) gives

$$w(\theta) = (3/\pi) a_2^2 \log(\Theta/\theta), \quad (16)$$

where Θ is on the order of the size of the field within which $w(\theta)$ is measured.

d) Mass Fluctuations

A convenient measure is the rms fluctuation in the mass found within a randomly placed sphere of radius R. We have from equation (8)

$$\delta M/M = 3\sum \delta_k (\sin kR - kR\cos kR)/(kR)^3. \quad (17)$$

With equations (6) and (10), this gives the rms value

$$\frac{\delta M}{M} = \left(\frac{108}{\pi}\right)^{1/2} a_2 \left(\frac{c}{H}\right)^2 \left(\frac{h}{\tau}\right)^4 I(Rh^2/\tau^2),$$

$$[I(R)]^2 = \int_0^\infty \frac{k^3\, dk}{(1 + 6k + 2.65\, k^2)^2} \times \frac{(\sin kR - kR\cos kR)^2}{(kR)^6}. \quad (18)$$

Figure 1 shows $\delta M/M$ for $h = \tau = 1$ and the normalization

$$\frac{\delta M}{M}(R = 8\ \text{Mpc}) = 1, \quad (19)$$

which agrees with the rms fluctuation, $\delta N/N$, in the counts of bright galaxies at $R = 8\ h^{-1}$ Mpc (LSS § 59; Davis and Peebles 1983). At $R \lesssim 0.1$ Mpc, $\delta M/M$ varies only slowly, as $|\log R|^{1/2}$, fluctuations with fixed variance per octave of R having been stored when

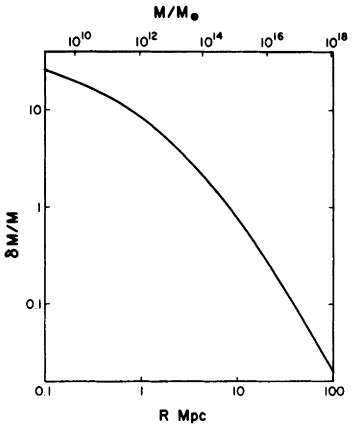


FIG. 1.—The rms fluctuation in the mass found within a sphere of radius R. The top scale is the mean mass within the sphere. The vertical normalization is adjusted to make $\delta M/M$ agree with the observed fluctuation, $\delta N/N$, in the count of bright galaxies at $R = 8\ h^{-1}$ Mpc. This curve is based on the assumptions that the universe is dominated by very massive, weakly interacting particles and that the initial density fluctuations were adiabatic with power spectrum $P \propto k$.

$\rho_r \gtrsim \rho_x$. Shortward of the scale r fixed by m_x (eq. [5]), the power spectrum is truncated by thermal motions so $\delta M/M$ is independent of R. At large R, $\delta M/M \propto R^{-2}$, which is the primeval spectrum. At $1 \lesssim R \lesssim 30$ Mpc, $\delta M/M$ varies roughly as $R^{-1.25}$.

Figure 1 can be compared to the curve $n = 1$ in Figure 2 of Bond, Szalay, and Turner (1982). The short-wavelength cutoff in the latter curve results from the assumption $m_x \sim 1$ keV. The Bond *et al.* curve peaks at $M \sim 10^{12}\ M_\odot$ as does the Jeans mass found by Blumenthal, Pagels, and Primack (1982), both of which agree with the position of the break in Figure 1. This is consistent with the fact that $m_x = 1$ keV is roughly equivalent to the limit of very large m_x.

By equations (18) and (19), the expected quadrupole moment of the microwave background is

$$a_2 = 3.5 \times 10^{-6}. \quad (20)$$

For the case of baryon-dominated matter, Silk and Wilson (1981) found $a_2 = 4 \times 10^{-5}$. The larger value is the result of the much broader mass coherence length, which increases the integral $J_3 = \int r^2\, dr\, \xi$ (Peebles 1981*b*).

The Boughn, Cheng, and Wilkinson (1981) measurements imply $a_2 \sim 3 \times 10^{-4}$, but the more recent measurements of Lubin (1982) and Fixsen (1982) suggest the extragalactic anisotropy may be appreciably less than that and hence perhaps not inconsistent with equation (20).

The expected temperature anisotropy at intermediate angular scales is given by equation (16). The rms fluctuation in T smoothed over $\theta = 10°$ in a sample of size $\Theta = 100°$ is

$$\delta T/T = w^{1/2} \sim 5 \times 10^{-6}. \quad (21)$$

The mass autocorrelation function is

$$\xi(r) = \langle \rho(r)\rho(0)\rangle/\langle\rho\rangle^2 - 1 = \int_0^\infty k^2\, dk\, P \sin kr/kr. \quad (22)$$

At large r, this is dominated by the primeval spectrum, $P = Ak$, so that (LSS § 42)

$$\xi = -(24/\pi)a_2^2(c/Hr)^4 = -(9.4\ \mathrm{Mpc}/r)^4. \quad (23)$$

III. DISCUSSION

Figure 1 is based on the primeval spectrum $P \propto k$ that has some theoretically attractive and perhaps important properties. It is normalized so $\delta M/M$ averaged over a sphere of radius 8 Mpc agrees with the rms fluctuation, $\delta N/N$, in the counts of bright galaxies. This radius is small enough that $\delta N/N$ is fairly well known yet large enough that the time evolution of $\delta M/M$ in the absence of nongravitational forces is accurately given by the linear perturbation calculation (Peebles and Groth 1976). If galaxy formation were aided by the Ostriker-Cowie (1981) process, the primeval amplitude could be lower. However, $\delta M/M$ is almost flat at $R \lesssim 0.1$ Mpc, so if we wanted the first generation of objects to form at redshift $z \gtrsim 10$, we could not decrease a_2 by a factor of more than about 3 unless we were willing to go to exceedingly small values of R.

This scenario yields a characteristic length on the order of 1 Mpc, which certainly is observationally interesting. It is fixed by the horizon size when $\rho_r \sim \rho_x$, independent of the free particle mass m_x if $m_x \gtrsim 1$ keV. At R between 3 and 30 Mpc, $\delta M/M$ scales roughly as $M^{-1/2}$ so the similarity argument (LSS §§ 26 and 73) suggests this could develop into the observed mass clustering hierarchy (Davis and Peebles 1983).

All of this discussion has dealt with the second moment of the mass distribution. It remains to be seen whether the rather rapidly decreasing large-scale fluctuations, $\delta M/M \propto R^{-2}$, could produce large-scale features such as the cluster-cluster coherence length, $r_c \sim 25\ h^{-1}$ Mpc (where $\xi_{cc} = 1$; Hauser and Peebles 1973; Bahcall and Soneira 1982).

In the spirit of the scenario, we expect $\Omega = 1$. The relative velocity data at $r \lesssim 1\ h^{-1}$ Mpc in the Center for Astrophysics (CfA) redshift sample yield $\Omega \sim 0.3$, but there is an indication of a more broadly distributed mass component, so $\Omega = 1$ certainly is not improbable. The p-wave part of the background would be almost entirely due to our peculiar velocity, v. The expected rms velocity is ~ 1000 km s^{-1} if $\Omega \sim 1$ (LSS § 74), which is larger than the observed background p-wave but not to be ruled out: as we are not in a particularly strongly clustered spot, we might expect our velocity is no greater than the median, and as the degree of clustering is highly variable, we might expect the distribution of v is broad and hence that the median is well below the rms value.

Another interesting possibility is that the mass autocorrelation function, $\xi(r)$, vanishes at $r \gtrsim 20\ h^{-1}$ Mpc. If so, the mass fluctuations at smaller r, where $\xi > 0$, produce long-range gravitational potential fluctuations that lead to a value of a_2 that agrees with the present observations (Peebles 1981*b*, 1982*b*). This is an attractive coincidence. Of course, if a_2 were substantially lowered, this picture would be much less attractive. By going from a flat spectrum to $P \propto k$ in the present scenario, we have introduced anticorrelated mass fluctuations, which strongly reduces a_2 (Peebles 1981*b*, § V; Silk and Wilson 1981). If this model is correct, a measurement of the background temperature autocorrelation function, $w(\theta)$, at large θ (eqs. [16] and [21]) will require a considerable advance in the observations. The computed mass anticorrelation amounts to $\xi(20\ \mathrm{Mpc}) \sim -0.05$. As this number is based on the linear approximation, it is not highly accurate, but certainly $\xi(r)$ must be negative at large r to reduce a_2. There is a tentative indication of anticorrelation at $hr \sim 30$ Mpc in the CfA redshift catalog (Davis and Peebles 1983), but a firm test awaits deeper samples.

If the extragalactic part of a_2 were found to be appreciable, say $a_2 \gtrsim 3 \times 10^{-5}$, several interpretations could be considered. It is doubtful that the discrepancy with equation (20) could be due to an underestimate of ξ (by a factor ~ 100 at $r \sim 5\ h^{-1}$ Mpc) for that would conflict with the relative velocity data (Davis and Peebles 1983). The anisotropy could be inserted by sources along the line of sight. If the source density were proportional to the mass density, the angular spectrum in the present scenario would be $a_l \propto l^{1/2}$ (compared to $a_l \propto l^{-1}$ for the intrinsic part in eq. [14]). Hogan (1982) has pointed out that one could assume that the luminosity per unit mass averaged over the mass coherence length $\sim 5\ h^{-1}$ Mpc is a random variable with broad dispersion, so that the background temperature fluctuations approximate white noise, $a_l \sim$ constant. In these two cases, the rms fluctuation in the background temperature smoothed over the angle θ scales as $\delta T/T \sim$

$\theta^{-3/2}$ and θ^{-1} respectively. Thus, if a_2 can be detected, it ought to be possible to check the spectrum. If these two models can be ruled out, likely interpretations of a large a_2 will be that the growth of clustering at $r \lesssim 20$ h^{-1} Mpc was hindered, perhaps by thermal motions ($m_x \ll 1$ keV), or else that the primeval spectrum was not adiabatic with $P \propto k$.

This research was supported in part by the National Science Foundation.

REFERENCES

Bahcall, N. A., and Soneira, R. M. 1982, preprint.
Blumenthal, G. R., Pagels, H., and Primack, J. R. 1982, *Nature*, in press.
Bond, J. R., Szalay, A. S., and Turner, M. S. 1982, *Phys. Rev. Letters*, **48**, 1636.
Boughn, S. P., Cheng, E. S., and Wilkinson, D. T. 1981, *Ap. J. (Letters)*, **243**, L113.
Davis, M., Lecar, M., Pryor, C., and Witten, E. 1981, *Ap. J.*, **250**, 423.
Davis, M., and Peebles, P. J. E. 1983, *Ap. J.*, submitted.
Doroshkevich, A. G., Khlopov, M. Yu., Sunyaev, R. A., Szalay, A. S., and Zel'dovich, Ya. B. 1981, in *Proceedings of the Tenth Texas Symposium, Ann. NY Acad. Sci.*, **375**, 32.
Fixsen, D. J. 1982, Ph.D. thesis, Princeton University.
Harrison, E. R. 1970, *Phys. Rev.*, **D1**, 2726.
Hauser, M. G., and Peebles, P. J. E. 1973, *Ap. J.*, **185**, 757.
Hawking, S. W. 1982, preprint.
Hogan, C. J. 1982, *Ap. J. (Letters)*, **256**, L33.
Lubin, P. M. 1982, paper presented at the 86th course of the International School of Physics, Varenna.
Melchiorri, F., Melchiorri, B. O., Ceccarelli, C., and Pietranera, L. 1981, *Ap. J. (Letters)*, **250**, L1.
Ostriker, J. P., and Cowie, L. L. 1981, *Ap. J. (Letters)*, **243**, L127.
Peebles, P. J. E. 1980, *The Large-Scale Structure of the Universe* (Princeton: Princeton University Press) (LSS).
_____. 1981*a*, *Ap. J.*, **248**, 885.
_____. 1981*b*, *Ap. J. (Letters)*, **243**, L119.
_____. 1982*a*, *Ap. J.*, **258**, 415.
_____. 1982*b*, *Ap. J.*, **259**, 442.
Peebles, P. J. E., and Groth, E. J. 1976, *Astr. Ap.*, **53**, 131.
Peebles, P. J. E., and Yu, J. T. 1970, *Ap. J.*, **162**, 815.
Press, W. H., and Vishniac, E. T. 1980, *Ap. J.*, **236**, 323.
Silk, J., and Wilson, M. L. 1981, *Ap. J. (Letters)*, **244**, L37.
Wasserman, I. 1981, *Ap. J.*, **248**, 1.
Zel'dovich, Ya. B. 1972, *M.N.R.A.S.*, **160**, 1P.

P. J. E. PEEBLES: Joseph Henry Laboratories, Physics Department, Princeton University, Princeton, NJ 08544

GROWTH OF PERTURBATIONS BETWEEN HORIZON CROSSING AND MATTER DOMINANCE: IMPLICATIONS FOR GALAXY FORMATION

Joel R. Primack
Board of Studies in Physics and Santa Cruz Institute for Particle Physics, University of California, Santa Cruz, CA 95064

George R. Blumenthal
Board of Studies in Astronomy and Astrophysics, Lick Observatory, University of California, Santa Cruz, CA 95064

The dark matter (DM) that appears to be gravitationally dominant on all astronomical scales larger than the cores of galaxies can be classified, on the basis of its characteristic free-streaming damping mass M_D, as hot ($M_D \sim 10^{15} M_\odot$), warm ($M_D \sim 10^{11} M_\odot$), or cold ($M_D < 10^8 M_\odot$) (Bond and Szalay, 1983; Primack and Blumenthal, 1983). For the case of cold DM, the shape of the DM fluctuation spectrum is determined by (a) the primordial spectrum (on scales larger than the horizon), which is usually assumed to have a power spectrum of the form $|\delta_k|^2 \propto k^n$ (inflationary models predict the "Zeldovich spectrum" $n = 1$); and (b) "stagspansion", the stagnation of growth of DM fluctuations which enter the horizon while the universe is still radiation-dominated. Stagspansion flattens the fluctuation spectrum for $M \lesssim 10^{15} M_\odot$. We report here the results of a numerical evaluation of the fluctuation spectrum in which all relevant physical effects have been included (Blumenthal and Primack, in preparation), together with implications for galaxy formation (Blumenthal, Faber, Primack, and Rees, in preparation; hereafter BFPR).

1. WHY COLD DM IS INTERESTING

If a species of neutrino is the gravitationally dominant component of the universe, its mass $m_\nu = 100\,\Omega h^2\,\mathrm{eV}$ (where $h = H_o/100\,\mathrm{km\,s^{-1}Mpc^{-1}}$ lies in the range $1/2 \leq h \leq 1$) implies a free-streaming damping mass $M_D \sim 10^{15} M_\odot$ corresponding to hot DM. Probably because this type of DM has been the most intensively studied, a number of potential problems have been identified - for example the late formation of supercluster "pancakes", at $z_p \lesssim 2$ (see, e.g., Frenk, White, and Davis, 1983), which later fragment into galaxies, in contradiction to the observed abundance of QSOs with $z > 2$, which implies earlier galaxy formation.

F. Mardirossian et al. (eds.), Clusters and Groups of Galaxies, 435–440.
© 1984 by D. Reidel Publishing Company.

We would like to call attention to another discrepancy between the hot DM hypothesis and observations, including data presented at this conference: namely the absence of evidence for growth of the DM/baryon ratio from the scale of ordinary galaxies to rich clusters. If the DM is composed of neutrinos, which first collapse gravitationally on supercluster scales with corresponding velocities $\sim 10^3$km/s, only a small fraction of the neutrinos can be captured to form the DM halos of galaxies having escape velocities typically $\sim 10^2$km/s, while a much larger fraction will be bound in rich clusters. One-dimensional simulations (Bond, Szalay, and White, 1983) confirm this expectation. The most relevant observational parameter is not the oft-discussed M_{tot}/L but rather $M_{tot}/M_{\ell um}$, where M_{tot} is the total mass determined dynamically and $M_{\ell um}$ is the mass in stars plus that in x-ray emitting gas. The available data, including that on our galaxy (which imply that its massive halo extends at least to ~50kpc), on small spiral groups, on small cD groups, and on rich clusters, is all consistent with $M_{tot}/M_{\ell um} \approx 10$ (BFPR).

Yet another problem for the hypothesis that the DM is neutrinos is the recent prediction (Faber and Lin, 1983) that the dwarf spheroidal galaxies near the Milky Way have heavy halos; this is supported by recent observational determinations of the velocity dispersion of Draco (Aronson, 1983; Lin and Faber, 1983) and Carina (Aronson, private communication to Faber). The phase space constraint (Tremaine and Gunn, 1979) implies that $m_\nu \gtrsim 500$ eV. This is of course inconsistent with the cosmological density bound on m_ν, implying that the DM in these dwarf spheroidal halos is not hot. It is probably not warm DM either. Warm DM first collapses on a scale $\sim 10^{12} M_\odot$ with $\sigma \sim 10^2$ km/s (Blumenthal, Pagels, and Primack, 1982; Bond, Szalay, and Turner, 1982), and too little could be captured by dwarf spheroidals, having $\sigma \sim 10$ km/s, to form the heavy halos indicated by the observations.

Besides the evidence just summarized against hot and warm DM, a further reason to consider cold DM is the existence of several plausible physical candidates, including <u>axions</u> of mass $\sim 10^{-5}$ eV (Preskill, Wise, and Wilczek, 1983; Abbott and Sikivie, 1983; Dine and Fischler, 1983; Ipser and Sikivie, 1983); heavy stable particles, such as the <u>photino</u>, with a mass ~10 GeV and very weak interactions; and <u>primordial black holes</u> with $10^{17}\,g \lesssim m_{PBH} \lesssim M_\odot$.

2. CALCULATION AND IMPLICATIONS OF THE FLUCTUATION SPECTRUM

We will follow the current conventional wisdom and assume that the primordial fluctuations were adiabatic. In the standard formulation (e.g. Peebles, 1980), fluctuations $\delta \equiv \delta\rho/\rho$ grow as $\delta \sim a^2$ on scales larger than the horizon, where a = scale factor = $(1+z)^{-1}$. When a fluctuation enters the horizon in the radiation-dominated era, the photons (together with the charged particles) oscillate as an acoustic wave, and the neutrinos (assumed to have negligible mass) free stream away. As a result, the main driving terms for the growth of δ_{DM}

disappear, and the growth of δ_{DM} stagnates ("stagspansion") until matter dominates (Guyot and Zeldovich, 1970; Meszaros, 1974; Groth and Peebles, 1975); see Fig. 1. Matter domination first occurs at $z = z_{eq}$, where

$$z_{eq} = 4.2 \times 10^4 h^2 \Omega (1 + 0.68\, n_\nu)^{-1} = 2.5 \times 10^4 h^2 \Omega \text{ for } n_\nu = 3 \quad (1)$$

The first study of the growth of cold DM fluctuations was the numerical calculation of Peebles (1982), who for simplicity ignored neutrinos: $n_\nu = 0$ in (1). We have done numerical calculations including the effects of the known neutrino species ($n_\nu = 3$, $m_\nu \approx 0$) both outside and inside the horizon. Numerically, the largest effect of including neutrinos is the change in z_{eq}.

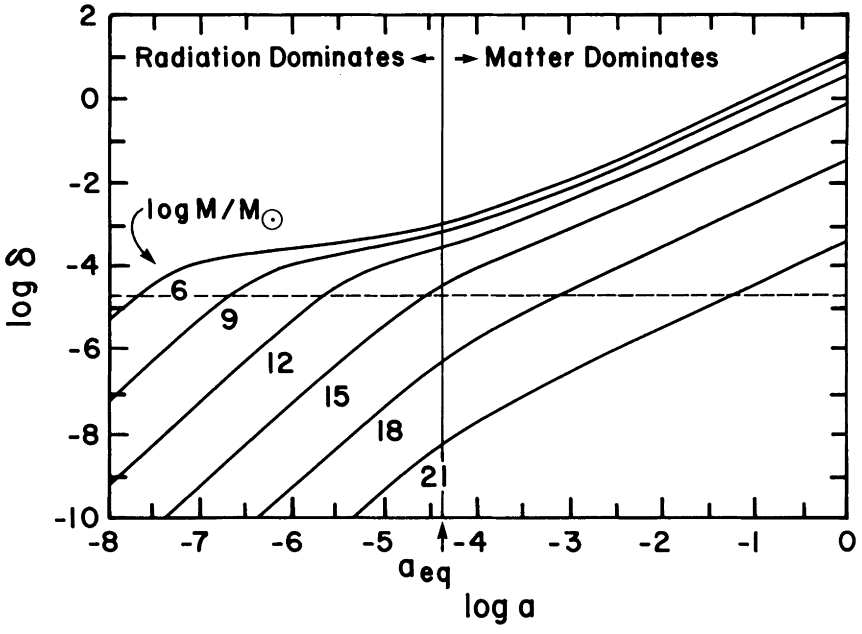


Figure 1. Numerical results for the growth of $\delta = k^{3/2}\delta_k$ versus scale factor a for fluctuations of various masses $M = 4\pi^4 k^{-3}\rho_c/3$. The curves are drawn for $n = 1$, $\Omega = h = 1$, and a baryonic to total mass ratio of 0.1. The vertical line represents the value of a when the universe becomes matter dominated, and the dashed line shows the (constant, for $n = 1$) value of δ when each mass scale crosses the horizon. These curves illustrate the stagnation of perturbation growth after small mass scales cross the horizon and show why at late times $\delta(k)$ is nearly flat for large k (small M).

It is instructive to make the further approximation of setting $\delta_{\gamma+b} = \delta_\nu = 0$ once a fluctuation is inside the horizon. Then one can analytically match the solution for $a > a_{horizon}$ (Peebles, 1980, p.58)

$$\delta_{DM}(a) = A_1 D_1(a) + A_2 D_2(a) = \delta_1 + \delta_2, \tag{2}$$

$$D_1 = 1 + 1.5y, \quad \text{where } y = a/a_{eq}, \tag{3a}$$

$$D_2 = D_1 \ln\left[\frac{(1+y)^{\frac{1}{2}} + 1}{(1+y)^{\frac{1}{2}} - 1}\right] - 3(1+y)^{\frac{1}{2}}. \tag{3b}$$

to the growing mode $\delta_{DM} \sim a^2$ for $a < a_{horizon}$. Matching the derivatives requires δ_2 comparable to δ_1 but opposite in sign. For $a >> a_{horizon}$ only the growing mode D_1 survives, which explains the moderate growth in δ_{DM} between horizon crossing and matter dominance. In the limit of large k, one finds $\delta_k \propto k^{n/2-2} \ln k$. Correspondingly, for $M << M_{eq} \approx 10^{16} M_\odot$, the rms fluctuation in the mass within a random sphere containing average mass M is $\delta M/M \propto |\ln M|^{3/2}$. Turner, Wilczek, and Zee (1983) considered only the Meszaros solution (3a) and erroneously inferred that the fluctuation spectrum would be essentially flat for $M < M_{eq}$ for a Zeldovich primordial spectrum, which would then be inconsistent with observations.

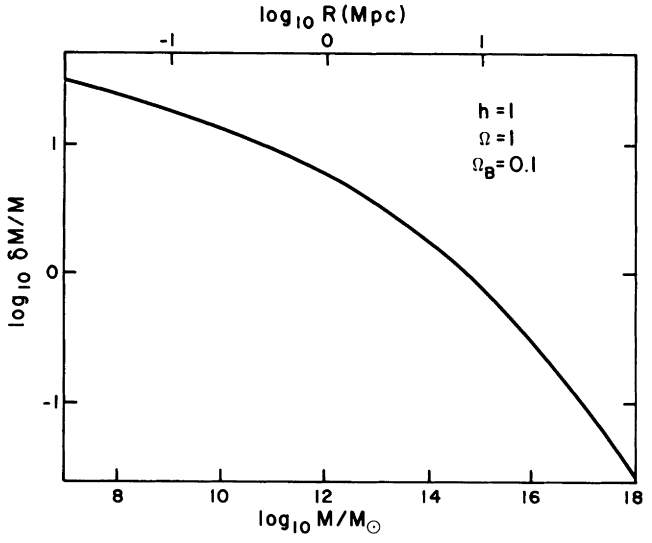


Figure 2. The r.m.s mass fluctuations within a randomly placed sphere of radius R in a cold DM universe. The curve is normalized at 8 Mpc and assumes an initial Zeldovich (n = 1) fluctuation spectrum, and $h = \Omega = 1$.

Our numerical results for $\delta M/M$ are shown in Fig. 2 for $\Omega = h = 1$, assuming a Zeldovich ($n = 1$) spectrum (reflected in $\delta M/M \propto M^{-2/3}$ for $M > M_{eq}$). We have followed Peebles (1982) in normalizing $\delta M/M = 1$ at $8h^{-1}$ Mpc. For either h or Ω less than unity, $\delta M/M$ is somewhat flatter.

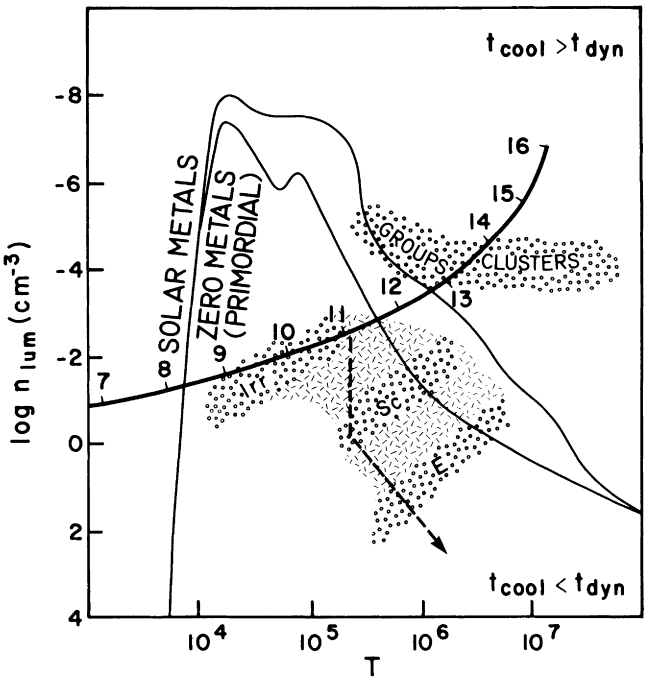


Figure 3. The baryonic density versus temperature as average perturbations having total mass M become nonlinear and virialize. The numbers on the tick marks are the logarithm of M in solar units. This curve assumes $n = 1$, $\Omega = h = 1$, and a baryonic to total mass ratio of 0.07. The region where baryons can cool within a dynamical time lies below the cooling curves. Also shown are the positions of observed galaxies, groups, and clusters of galaxies. The dashed line represents a possible evolutionary path for dissipating baryons.

Fig. 3 (Primack and Blumenthal, 1983) shows the density of ordinary (baryonic) matter vs. internal kinetic energy (temperature) of typical fluctuations of various sizes, just after virialization, calculated from $\delta M/M$ of Fig. 2. This is superimposed upon the Rees-Ostriker (1977) cooling curves (for which cooling time equals gravitational free fall time) and data on galaxies (with kinetic energy determined from rotation velocity for spirals and velocity dispersion for ellipticals). Fluctuations that start with greater amplitude than average will turn

around earlier, at higher density, and thus lie below the virilization curve on Fig. 3. As the baryons in a virialized fluctuation dissipate, their density will initially increase at constant T within the surrounding isothermal halo of dissipationless material (DM), and then T will increase as well when the baryon density exceeds the DM density, as suggested by the dashed line in the figure. Note that the galaxy data lies below the virialization curve (the curve is somewhat higher for $h = \frac{1}{2}, \Omega = 1$). In this respect, the Zeldovich primordial spectrum is more consistent with the data than an $n = 2$ (or $n = 0$) primordial spectrum, which lies too low (too high) on the figure compared to the galaxies (see BFPR for more details). With the Zeldovich spectrum, the important conclusion is that one should observe dissipated systems with large halos having total mass $10^8 M_\odot \lesssim M \lesssim 10^{12} M_\odot$. This is essentially the range of observed galaxy masses.

REFERENCES

Abbott, L., and Sikivie, P.: 1983, Phys. Lett. 120B,pp.133-136.

Aronson, M.: 1983, Astrophys. J. (Lett.) 266,pp.L11-L15.

Blumenthal, G.R., Pagels, H., and Primack, J.R.: 1982, Nature 299,pp.37-38.

Bond, J.R., and Szalay, A.S.: 1983, Astrophys. J. (in press).

Bond, J.R., Szalay, A.S., and Turner, M.S.: 1982, Phys. Rev. Lett. 48, pp.1636-1639.

Bond, J.R., Szalay, A.S., and White, S.D.M.: 1983, Nature 301,pp.584-585.

Dine, M., and Fischler, W.: 1983, Phys. Lett. 120B,pp.137-141.

Faber, S.M., and Lin, D.N.C.: 1983, Astrophys. J. (Lett.) 266,pp.L17-L20.

Frenk, C., White, S.D.M., and Davis, M.: 1983, Astrophys. J. (in press).

Groth, E.J., and Peebles, P.J.E.: 1975, Astron. Astrophys. 41,pp.143-145.

Guyot, M., and Zeldovich, Ya. B.: 1970, Astron. Astrophys. 9,pp.227-231.

Ipser, J., and Sikivie, P.: 1983, Phys. Rev. Lett. 50,pp.925-927.

Lin, D.N.C., and Faber, S.M.: 1983, Astrophys. J. (Lett.) 266,pp.L21-L25.

Meszaros, P.: 1974, Astron. Astrophys. 37,pp.225-228.

Peebles, P.J.E.: 1980, *The Large Scale Structure of the Universe* (Princeton University Press, Princeton, NJ).

Peebles, P.J.E.: 1982, Astrophys. J. (Lett.) 263,pp.L1-L5.

Preskill, J., Wise, M., and Wilczek, F.: 1983, Phys. Lett. 120B,pp.127-132.

Primack, J.R., and Blumenthal, G.R.: 1983, *Formation and Evolution of Galaxies and Large Structures in the Universe*, ed. J. Audouze and J. Tran Thanh Van (Reidel, Dordrecht, Holland); *Proceedings of the Fourth Workshop on Grand Unification* (Birkhäuser, Boston).

Rees, M.J., and Ostriker, J.: 1977, M.N.R.A.S. 179,pp.541-559.

Tremaine, S.D., and Gunn, J.E.: 1979, Phys. Rev. Lett. 42,pp.407-410.

Turner, M.S., Wilczek, F., and Zee, A.: 1983, Phys. Lett. 125B,pp.35-40 and (E)p.519.

Formation of galaxies and large-scale structure with cold dark matter

George R. Blumenthal* & S. M. Faber*

* Lick Observatory, Board of Studies in Astronomy and Astrophysics, University of California, Santa Cruz, California 95064, USA

Joel R. Primack†§ & Martin J. Rees‡§

† Stanford Linear Accelerator Center, Stanford University, Stanford, California 94305, USA
‡ Institute of Theoretical Physics, University of California, Santa Barbara, California 93106, USA

The dark matter that appears to be gravitationally dominant on all scales larger than galactic cores may consist of axions, stable photinos, or other collisionless particles whose velocity dispersion in the early Universe is so small that fluctuations of galactic size or larger are not damped by free streaming. An attractive feature of this cold dark matter hypothesis is its considerable predictive power: the post-recombination fluctuation spectrum is calculable, and it in turn governs the formation of galaxies and clusters. Good agreement with the data is obtained for a Zeldovich ($|\delta_k|^2 \propto k$) spectrum of primordial fluctuations.

WHY are there galaxies, and why do they have the sizes and shapes that we observe? Why are galaxies clustered hierarchically in clusters and superclusters, separated by enormous voids in which bright galaxies are almost entirely absent? And what is the nature of the invisible mass, or dark matter, that we detect gravitationally roundabout galaxies and clusters but cannot see directly in any wavelength of electromagnetic radiation? Of the great mysteries of modern cosmology, these three may now be among the ripest for solution.

Because there is evidence that the mass of dark matter in the Universe exceeds that of the visible matter by at least an order of magnitude[1], the third question may hold the key to the first two. We now consider the hypothesis that the dark matter is cold[2-7], that is, that its thermal velocity is cosmologically negligible in the early Universe.

Most modern theories of the origin of structure in the Universe assume that the extreme inhomogeneity of the present Universe grew gravitationally from initially very small density fluctuations[8,9]. One hypothesis extensively explored[10-12] is that the dark matter (DM) consists of neutrinos of mass ~30 eV. Because these particles are noninteracting for $t \gtrsim 1$ s but are still relativistic at $t \sim 1$ yr when galaxy-size masses first come within the horizon ($R_{\rm horizon} = ct$), they would freely stream away, smoothing out fluctuations of horizon size or smaller. We refer to such particles as hot DM, and to the phenomenon just mentioned as damping by free streaming. The mass of the neutrinos inside the horizon when they first become nonrelativistic is roughly $10^{15}\,M_\odot$, about the mass of a supercluster ($M_\odot$ is the solar mass). This is consequently the mass of the first structures to collapse gravitationally in a neutrino-dominated Universe.

It is also possible that the DM consists of elementary particles that decoupled thermally from the big bang much earlier than neutrinos. It can be shown that such particles would have lower number density today, and thus could be more massive (~1 keV) and become nonrelativistic sooner than neutrinos[13,14]. We refer to this as warm DM. The first structures to form in such a Universe would weigh about $10^{11}\,M_\odot$, about the mass of a typical galaxy[15-17].

In contrast, in a Universe dominated by cold DM, the free streaming damping mass is, by definition, smaller than galaxy masses. Below, we will discuss possible particle physics candidates for cold DM and explain why we believe it is more plausible that the Universe is dominated gravitationally by cold DM than by ordinary matter (baryons) or hot or warm DM. We next consider galaxy formation in the cold DM picture, and show that galaxy and cluster data compare favourably with the predictions of this model. We then discuss the extent to which the data constrain the cosmological density in the cold DM picture, and the evolution of superclusters and voids. The cold DM hypothesis seems to lead to a remarkably attractive theory for galaxy formation and to account for large-scale structure at least as well as any competing theory.

Evidence that dark matter is cold

Cold DM consists of particles having negligible thermal velocity with respect to the Hubble flow and having nongravitational interactions that are much weaker than the weak interactions[3,4]. A popular cold DM candidate is the axion[18], a pseudoscalar field proposed originally to avoid large CP (constant parity) violation in the strong interactions (which would imply, for example, much too large a value for the neutron electric dipole moment)[19-21]. Instanton effects generate a nonzero axion mass at the quark deconfinement temperature ($T \sim 10^2$ MeV), below which the axions act as a nonrelativistic, massive, pressureless fluid. The requirement that the axion density be less than the critical density implies that the axion mass $m_a \gtrsim 10^{-5}$ eV (refs 22-24), while the longevity of helium burning stars implies that $m_a < 10^{-1}$ eV (ref. 25). Thus, if axions exist, they may be cosmologically important, and, for $m_a \approx 10^{-5}$ eV, they would be gravitationally dominant. If axions comprise the dark halo of our Galaxy, laboratory experiments have recently been proposed that could detect them[26].

Another cold DM candidate particle is the photino, the spin $\frac{1}{2}$ supersymmetric partner of the photon. Photinos are thought to be the lightest supersymmetric particle, with $m_\gamma \gtrsim 0.5$ GeV (the lower limit corresponding to cosmological critical density)[27]. As photino annihilation at high temperatures is incomplete, the remnant photinos can, because of their large mass, contribute a critical density today.

A third cold DM candidate is black holes[28,29] of mass $10^{-16}\,M_\odot \lesssim M_{\rm BH} \lesssim 10^6\,M_\odot$, the lower limit implied by the non-observation of γ rays from black hole decay by Hawking radiation, and the upper limit required to avoid disruption of galactic disks and star clusters[30,31]. Stronger but more controversial upper limits are $M_{\rm BH} < 10^2\,M_\odot$ from dwarf spheroidal haloes[32] and $M_{\rm BH} < 10^{-2}\,M_\odot$ from the non-observation of focusing of quasar cores[33].

§ Permanent addresses: Santa Cruz Institute of Particle Physics, University of California, Santa Cruz, California 95064, USA (J.R.P.); Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK (M.J.R.).

Another exotic cold DM candidate, recently proposed by Witten, is "nuggets" of u-s-d symmetric quark matter[34]. There is thus no shortage of cold DM candidate particles—although there is no direct evidence that any of them actually exists. Our motivation for considering the hypothesis that the Universe is dominated by cold DM is twofold. First, there is more than a little evidence against baryonic, hot, and warm DM. Moreover, a cold DM Universe correctly predicts many of the observed properties of galaxies, including their range of masses, irrespective of the identity of the cold particle.

The most conventional assumption might be that DM is baryonic, since we know that baryons pervade the Universe. There are, however, three arguments that make a nonbaryonic form of DM more plausible. First, if DM consists of baryons, it must be clumped into objects with mass $\ll M_\odot$ to avoid nuclear burning and the emission of too much radiation[35], and there is no compelling theory for the formation of such a large density of objects of planetary mass or smaller. The difficulties with other possible forms of baryonic dark matter have been reviewed elsewhere[36]. The second argument involves the observed deuterium abundance, $D/H = (1\text{-}4) \times 10^{-5}$ (by number), which provides a lower limit on the primordial D abundance because deuterium is readily consumed but not produced in stars. This then corresponds to an upper limit for the baryonic density parameter (ratio of baryon density to critical density) of $\Omega_b \leq 0.035h^{-2}(T_0/2.7)^3$ (ref. 37), where T_0 is the present temperature of the microwave background radiation and h is Hubble's constant expressed in units of 100 km s^{-1} Mpc. (Observationally, $1/2 \leq h \leq 1$.) However, there is also strong observational evidence that the total density parameter $\Omega \geq 0.1$. Therefore a baryon dominated Universe ($\Omega = \Omega_b$) is consistent with the deuterium limit only for Ω and h very near their observational lower limits. Finally, the existence of galaxies and clusters today requires that perturbations in the density must have become nonlinear before the present epoch. In a baryonic Universe, for adiabatic perturbations at recombination, this implies present-day fluctuations in the microwave background much larger than the present observational upper limits[38,39]. For baryonic DM, this problem can be avoided if the primordial fluctuation spectrum is isothermal rather than adiabatic, contrary to grand unified models of baryosynthesis[40,41], or if there is significant reheating of the intergalactic medium after recombination[42]. For nonbaryonic DM, on the other hand, the predicted fluctuations in the microwave background can be consistent with the observations as the fluctuations in the baryon-photon fluid are small at recombination and only later grow to the same size as fluctuations in the DM.

The neutrino dominated picture of galaxy formation also has serious weaknesses. Studies of nonlinear clustering (on scales $\lambda < 10$ Mpc) show that supercluster collapse must have occurred quite recently, at redshift $z_{sc} < 2$ (refs 43, 44). This is also consistent with a study of streaming velocity in the linear regime ($\lambda > 10$ Mpc), which indicates $z_{sc} < 0.5$ (ref. 45). However, the best limits on galaxy ages coming from globular clusters and other stellar populations, plus the possible association of QSOs with galactic nuclei, indicate that galaxy formation took place before $z = 3$ (ref. 46). This is inconsistent with the neutrino 'top-down' theory, in which superclusters form before galaxies rather than after them.

Another problem with the neutrino picture is that large clusters of galaxies can accrete neutrinos more efficiently than ordinary galactic haloes, which have lower escape velocities. One-dimensional numerical simulations predict that the ratio of total-to-baryonic mass M/M_b should be ~5 times larger for clusters ($\sim 10^{14} M_\odot$) than for ordinary galaxies ($M \sim 10^{12} M_\odot$)[47]. While there is evidence that mass-to-light ratio M/L does increase with scale, there is also considerable evidence that the more physically meaningful ratio of total-to-luminous mass M/M_{lum} remains constant from large clusters through groups of galaxies, binary galaxies, and ordinary spirals. (M_{lum}, which is the mass visible in galactic stars and gas plus hot, X-ray emitting gas, is $\leq M_b$, as an unknown fraction of the baryons is

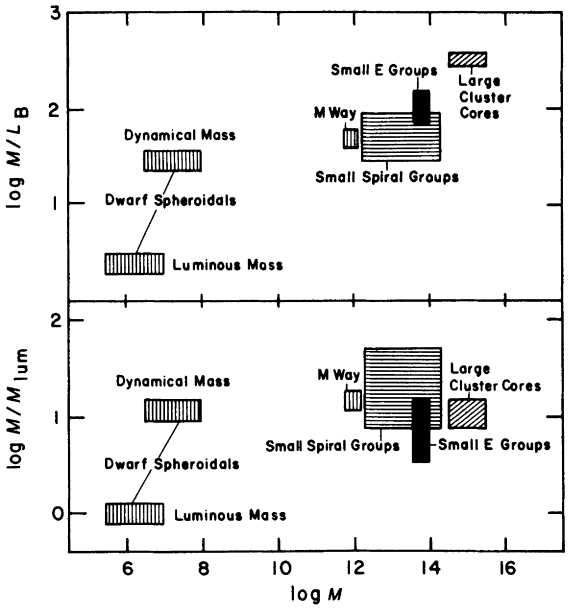


Fig. 1 Mass-to-light ratio, M/L_B, and total-to-luminous mass, M/M_{lum}, for structures of various size in the Universe. The data come from Table 1. Although M/L_B increases systematically with mass, the more physically meaningful ratio M/M_{lum} appears to be constant on all scales within the errors. If the velocity dispersion data for the dwarf spheroidal galaxies are interpreted to imply heavy haloes, the upper estimates result. The lower estimates follow from assuming that all the mass is visible. We believe the former estimate to be more realistic.

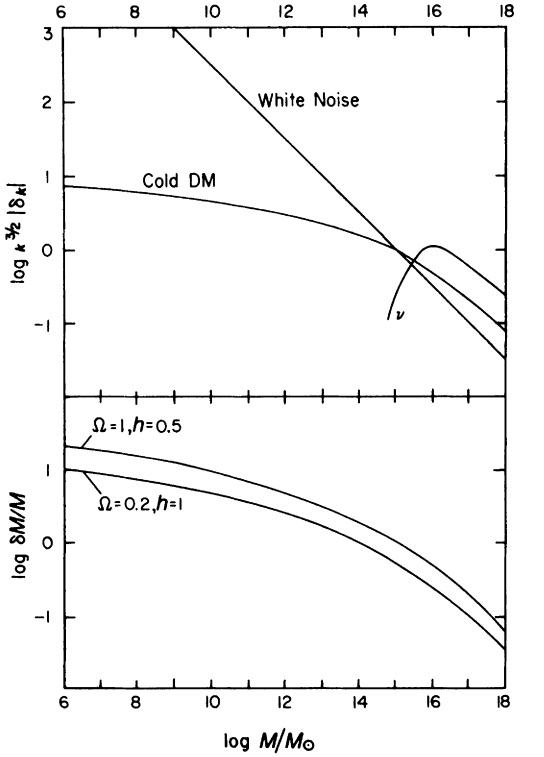


Fig. 2 Density fluctuations as a function of mass. *a*, $k^{3/2}|\delta_k| = \delta\rho/\rho(M)$, where $M = 4\pi^4\rho_0/3k^3$, for isothermal white noise ($n = 0$), and adiabatic Zeldovich ($n = 1$) neutrino[16] and cold DM spectra. *b*, r.m.s. mass fluctuation within a randomly placed sphere containing mass M for cold DM, $n = 1$, and ($\Omega = 1$, $h = 0.5$), ($\Omega = 0.2$, $h = 1$).

invisible—for example, in the form of diffuse ionized intergalactic gas at $T \sim 10^4$ K.) This is illustrated in Table 1 and Fig. 1, where we have plotted observed data for M/L and $M/M_{\rm lum}$. The fact that the total-to-luminous mass of rich clusters is similar to that of galaxies including their massive haloes, even though the clusters' mass-to-light ratio is larger, is due mainly to the different stellar population in the E and S0 galaxies in rich clusters plus the large contribution of X-ray emitting gas to $M_{\rm lum}$. In very rich clusters such as Coma, there is perhaps ~2–5 times as much mass in hot gas as there is in stars.

Finally, preliminary velocity dispersion data for Draco, Carina, and Ursa Minor[48–50] as well as theoretical arguments[51] suggest that a significant amount of DM may reside in dwarf spheroidal galaxies. Because of the low velocity dispersion of dwarf galaxies, phase space constraints give a lower limit of $m > 500$ eV for the mass of particles comprising this DM[32]. The present velocity dispersion estimates are uncertain owing to possible stellar oscillations, mass outflow, and binary motions, but these effects can be discovered and eliminated with careful monitoring. Taken at face value, however, the existing mass limit of 500 eV would rule out neutrinos completely.

Although warm DM provides a natural (free streaming) scale for ordinary galaxies, it cannot account for massive haloes in dwarf spheroidals even though the warm DM mass is actually, if barely, consistent with the phase space constraint. This is because free streaming damps out fluctuations with mass $<10^{11}\ M_\odot$, so dwarf galaxies with mass $\sim 10^7\ M_\odot$ can form in this picture only by the fragmentation of much larger scale galactic masses. Because dwarf galaxies with escape velocity ~ 10 km s^{-1} would capture only a small fraction of the warm DM particles, whose velocity dispersion would be ~ 100 km s^{-1}, typical of ordinary spirals, one would expect $M/M_{\rm lum}$ to be much smaller in dwarf galaxies than in spirals. However, the dynamical mass data on dwarf spheroidals shown in Fig. 1 suggest a similar value for the ratio of total-to-luminous mass. In view of the great importance of dwarf spheroidal haloes to constraints on warm as well as hot DM, it is clearly urgent to continue the monitoring programme on velocity dispersions mentioned above.

Recently proposed models having two or more kinds of possibly unstable DM[52–58] may be easier to reconcile with observations than models with one stable DM species, in part because of their additional adjustable parameters. Such models are, however, beyond the scope of this paper.

Galaxy formation with cold dark matter

The key features of galaxy formation in the cold DM picture are these: the DM fluctuation spectrum at recombination is determined by a small number of physical parameters; after recombination (at $z_{\rm rec} \approx 1{,}300$) the amplitude of the baryonic fluctuations rapidly grows to match that of the DM fluctuations; smaller-mass fluctuations grow to nonlinearity and virialize, and then are hierarchically clustered within successively larger bound systems; and finally the ordinary matter in bound systems of total mass $\sim 10^{8-12}\ M_\odot$ cools rapidly enough within the DM haloes to form galaxies, while larger mass fluctuations form clusters.

To calculate the growth of initially small perturbations and their ultimate gravitational collapse into condensed systems, one must first determine an initial spectrum for density perturbations in the very early Universe. We assume here that the initial fluctuations are adiabatic, which is consistent with fashionable particle theories in which a small excess of baryons over antibaryons is generated by the decay of supermassive grand unified theory particles[40,41]. Then the mass (or energy) density at any point can be written as $\rho(r, t) = \rho_0(1+\delta)$, where $\rho_0(t)$ is the average density in the Universe and δ represents the fractional density perturbation in the synchronous gauge. If the fluctuation spectrum is characterized by a power law distribution, then the r.m.s. fluctuation on mass scale M can be written $\delta \propto M^{(-1/2)(-n/6)}$, which corresponds to a Fourier power spectrum $|\delta_k|^2 \propto k^n$. Alternatively, we can characterize the perturbations on mass scale M when that mass scale crosses (first comes within) the horizon as $\delta_{\rm H} = \varepsilon(M/M_0)^{-\gamma}$, where M_0 is the present horizon mass and $n = 6\gamma + 1$. To avoid having too much power on either large or small scales requires $\gamma \approx 0$, which corresponds to $n \approx 1$. Limits on the large-scale variation of the microwave background further imply $\varepsilon \lesssim 10^{-4}$. The case $n = 1$ is commonly referred to as the constant curvature or (Harrison–Peebles–) Zeldovich[59–61] spectrum, and is predicted in inflationary models. It has recently been shown that not only inflation but also $\varepsilon \approx 10^{-4}$ can be arranged in suitably fine-tuned grand unified theories[62–64]. The Zeldovich spectrum also arises automatically if the fluctuations are due to cosmic strings[65,66]. We show below that $n = 1$ is, furthermore, in best agreement with galaxy and cluster data.

Table 1 $M/L_{\rm B}$ and $M/M_{\rm lum}$ on various mass scales

Unit	M*	$M/L_{\rm B}$†	$M_{\rm gas}/M_{\rm lum}$	$M/M_{\rm lum}$
Large clusters	$10^{15}\ M_\odot$	316 ± 40‡	$0.84^{+0.0}_{-0.1}$§	$8.4^{+7.0}_{-1.0}$
Small E-dominated groups	5×10^{13}	83^{+80}_{-10}‖	$0.61^{+0.1}_{-0.1}$‖	$5.4^{+10.0}_{-2.0}$‖
Small spiral-dominated groups	2×10^{13}	40^{+50}_{-10}¶	0(?)	14.2^{+36}_{-6}#
Whole Milky Way	10^{12}**	50††	0(?)	14‡‡
Dwarf spheroidals§§				
Stellar masses	10^{5-7}	2.5	0	1
Dynamical masses	10^{6-8}	30	0	12

* Total mass including dark matter. $H = 50$ km s^{-1} Mpc.

† Mass-to-light ratio on B(0) magnitude system as described in ref. 1.

‡ From ref. 120.

§ $M_{\rm gas}/M_{\rm lum}$ is calculated as the product of two factors: $(M_{\rm gas}/M) \times (M_{\rm lum}/M)^{-1}$. $M_{\rm gas}/M = 0.10^{+0.02}_{-0.05}$ from ref. 121 for 14 X-ray clusters. Formal clusters have been increased to allow for the unknown gas distribution outside the core and possible settling of gas relative to galaxies or vice versa. $M_{\rm lum}/M$ is $(M_* + M_{\rm gas})/M$, where M_*/M is the fractional mass in stars. It may be calculated as $(M/L_{\rm B})_*(M/L_{\rm B})^{-1}$, where $(M/L_{\rm B})_*$ is the stellar mass-to-light ratio. For E and S0 galaxies in large clusters, $(M/L_{\rm B})_*$ appears to be about 6 (ref. 1).

‖ $L_{\rm V}$, $M_{\rm gas}$, and M from ref. 122 based on five groups. $L_{\rm V}$ corrected to B(0) system based on data provided by Kriss (personal communication). Comparison with Beers *et al.*[123] suggests that M may be underestimated by a factor of two. Quoted errors take this into account.

¶ From ref. 1.

From ref. 72; includes baryonic mass in stars and neutral gas.

** Mass assumed to be 1/3 total mass of Local Group[1].

†† B(0) luminosity from ref. 1.

‡‡ $M_{\rm lum}$ from ref. 124.

§§ Basic data from ref. 51. Upper line assumes that only stars are present. Lower line includes excess dark matter, based on mean dynamical masses of Faber and Lin[51] and Aaronson *et al.*[48–50].

The first study of the growth of cold DM fluctuations was the numerical calculation of Peebles[2], who included cold DM, photons, and ordinary matter. Two of us (G.R.B. and J.R.P.)[3–5,7] extended these numerical calculations to include three massless neutrino species. Fluctuations having mass $M < M_{\rm eq} = 2\times10^{15}\ (\Omega h^2)^{-2}\ M_\odot$ cross the horizon when the Universe is still radiation dominated, that is, when $z > z_{\rm eq} = 2.5\times10^4\ \Omega h^2$. After such fluctuations cross the horizon, their neutrino components dissipate by free streaming, and the remaining photon and charged particle perturbations oscillate as an acoustic wave (whose amplitude is ultimately damped by photon diffusion for[67] $M < M_{\rm Silk} \approx 3\times10^{13}\ \Omega_{\rm b}^{-1/2}\Omega^{-3/4}h^{-5/2}M_\odot$). As a result, the main driving terms for the growth of cold dark matter fluctuations $\delta_{\rm DM}$ decrease, and consequently $\delta_{\rm DM}$ begins to grow only very slowly until the Universe becomes (dark) matter dominated at $z_{\rm eq}$, after which $\delta_{\rm DM} \propto a = 1/(1+z)$ until $z \approx \Omega^{-1}$. This stagnation of the growth of DM fluctuations between the epochs of horizon crossing and matter domination is called 'stagspansion'[3–7].

Because fluctuations with $M < M_{\rm eq}$ grow very little during the stagspansion era and as fluctuations on all scales grow at essentially the same rate after the Universe becomes matter dominated, an initial Zeldovich spectrum, $\delta_{\rm DM} \propto M^{-2/3}$, evolves to a much flatter spectrum for $M < M_{\rm eq}$ by the time of recombination. In Fig. 2 we plot the resulting cold DM fluctuation spectrum at present (ignoring the nonlinear evolution which is important for $\delta \gtrsim 1$, as discussed below; thus the plotted spectrum δ_k is larger than the spectrum at recombination by a constant factor). The quantity plotted in Fig. 2*a*, $k^{3/2}|\delta_k|$, roughly

represents the spatial density fluctuation δ as a function of total mass $M = 4\pi^4\rho_0/3k^3$. Also sketched are the fluctuation spectra in the hot DM scenario and in an isothermal scenario with white noise fluctuations. Note that the hot DM model requires more power on large scales today to form superclusters by $z \approx 2$. Figure 2*b* shows cold DM fluctuation spectra plotted in another form, $\delta M/M$, which represents the r.m.s. mass fluctuation within a randomly-placed sphere of radius R containing mass M. Following Peebles[2], we have normalized the curves so that, at the present epoch, $\delta M/M = 1$ at $R = 8h^{-1}$ Mpc.

Even though baryonic fluctuations do not grow (and are damped for $M < M_{\rm Silk}$) before recombination, after recombination the baryons 'fall into' the DM perturbations so that quickly $\delta_{\rm b} = \delta_{\rm DM}$ (refs 68, 69). This will occur if $\Omega_{\rm DM}\delta_{\rm DM} \gg \Omega_{\rm b}\delta_{\rm b}$, as we expect, and if the baryonic fluctuation mass exceeds the baryonic Jeans mass[6] $M_{\rm J,b} \sim 10^6\,\Omega_{\rm b}\Omega^{-3/2}h^{-1}(T_{\rm b}/T)^{3/2}M_\odot$, where $T_{\rm b}$ is the temperature of the baryonic gas and T is the photon temperature. On scales smaller than this, the pressure of the baryonic gas prevents it from developing the same density contrast δ as the cold DM. The value of $T_{\rm b}/T$ is kept close to unity for $z > 100$ by the coupling of the residual free electrons with the radiation, but for smaller z it falls off approximately as $(1+z)$. This means that at $z \approx 10$, $M_{\rm J,b}$ may be as small as $\sim 10^3\,M_\odot$.

At any mass scale M, when the fluctuation $\delta M/M$ approaches unity, nonlinear gravitational effects become important. The fluctuation then separates from the Hubble expansion, reaches a maximum radius, and begins to contract. Spherically symmetric fluctuations, for example, contract to about half their maximum radii. During this contraction, violent relaxation due to the rapidly varying gravitational field converts enough potential energy (PE) into kinetic energy (KE) for the virial theorem, $\langle PE\rangle = -2\langle KE\rangle$, to be satisfied. After virialization, the mean density within a fluctuation is roughly eight times the density corresponding to the maximum radius of expansion[8].

As the cold-DM fluctuation spectrum $\delta M/M$ is a decreasing function of M, smaller mass fluctuations will, on the average, become nonlinear and begin to collapse at earlier times than larger mass fluctuations. Smaller mass fluctuations are themselves typically clustered within larger mass perturbations, which go nonlinear later. This hierarchical clustering of smaller systems into larger and yet larger gravitationally bound systems begins at the baryon Jeans mass, $M_{\rm J,b}$, and continues until the present. The baryonic substructures within larger mass clusters will then be disrupted by virialization of the clusters unless significant mass segregation between baryons and DM has occurred before cluster virialization. Hence, to maintain their existence as a separate substructure, the baryons must cool and gravitationally condense within their massive DM haloes before virialization occurs on larger scales[70].

Figure 3 shows the baryonic number density $n_{\rm b}$ plotted against temperature T (or halo velocity dispersion) for virialized spherically symmetric protocondensations resulting from an initial Zeldovich spectrum of cold DM fluctuations. The simplifying assumption of sphericity is not expected to be an unreasonable approximation for hierarchical clustering. Two cases are considered: $\Omega = 1$ and $h = 0.5$ (solid lines in Fig. 3), and $\Omega = 0.2$ and $h = 1$ (dashed lines). The curves assume that the protocondensations have virialized, but that the baryons have not yet cooled and condensed. The curves labelled 1σ assume that for each mass scale, $\delta M/M$ has the r.m.s. value; those labelled 2σ correspond to fluctuations $\delta M/M$ twice as great; and so on. For given total mass M, the distribution of values of $\delta M/M$ spreads out along lines of constant M (light diagonal lines in Fig. 3). We have also shown in Fig. 3 the present positions of clusters and groups of galaxies and of individual galaxies, including dwarf spheroidals. Note that different types of galaxies, the Hubble sequence, are spread out in this diagram.

Baryonic cooling moves structures downward in Fig. 3, below the model curves. Baryons can radiatively cool through collisional excitation of atoms and molecules, and by Compton cooling off the cosmic background radiation; Fig. 3 correspondingly includes the medium solid curves labelled 'No metals' and

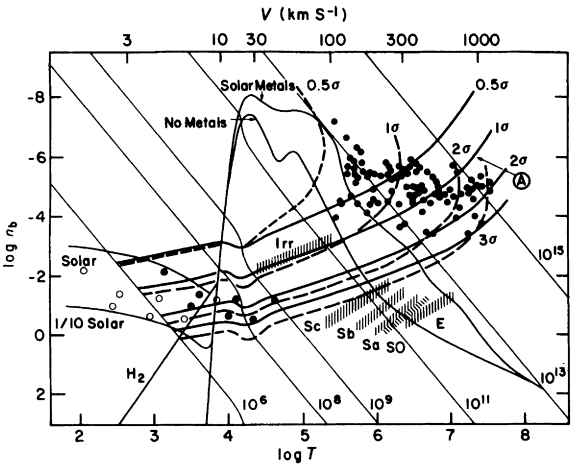


Fig. 3 Baryon density $n_{\rm b}$ versus three-dimensional, r.m.s. velocity dispersion V and virial temperature T for structures of various size in the Universe. The quantity T is $\mu V^2/3k$, where μ is mean molecular weight (≈0.6 for ionized, primordial H + He) and k is Boltzmann's constant. Dots at upper right represent nearly all groups and clusters within 5,000 km s^{-1} in the CfA catalogue[82]. Baryon density is obtained from total density assuming $\rho_{\rm b}/\rho = 0.1$ (see Table 1), where ρ is defined as $M/(4\pi R_{\rm vir}^3/3)$, M is total mass, and $R_{\rm vir}$ is GM/V^2. Catalogue groups are chosen to exceed a minimum threshold luminosity density. Their minimum baryon density should thus fall on a horizontal line, slightly modified by the higher M/L of rich clusters, as is observed. Point A represents Abell clusters of richness 2 and 3 studied by Dressler[120]. Here $R_{\rm vir} = 3.0\,R_{\rm eff}$, valid for a de Vaucouleurs cluster profile[125]. The quantity T for clusters corresponds to central velocity dispersion, with the arrow indicating the effect of the falloff in velocity dispersion at large radii observed in the Coma cluster[126,127]. For galaxies, $M_{\rm lum}$ is set equal to $(M_{\rm lum}/L_{\rm B}) \times L_{\rm B}$, where $L_{\rm B}$ is blue luminosity. Assumed values of $(M_{\rm lum}/L_{\rm B})$: E = 8, S0 = 6, Sa = 4, Sb = 2.5, Sc = 1.5, Dw. Irr = 1.00, Dw. Spheroidal = 2.5 (all but the last are based on $h = 0.5$). Galaxy radii are approximate virial radii, assuming that the baryonic components of galaxies are self-gravitating (to obtain a radius consistent with the definition used for groups and clusters above). To achieve this, isophotal radii from various sources have been appropriately scaled, as follows: E-Sb, $R_{\rm vir} = R_{25}$ (refs 128-131); Sc, $R_{\rm vir} = 1.24\,R_{25}$ (ref. 132); Dw-Irr, $R_{\rm vir} = 1.24\,R_{\rm H}$ (ref. 133); Dw-Sph. $R_{\rm vir} = 0.70\,R_{\rm tid}$ (ref. 51). Velocities: E-S0, $V = \sqrt{3}\times\sigma$ (nucleus) (refs 129, 134); Sa-Sc, $V = v_{\rm rot}^{\rm max}$ (refs 130, 131, 132); Dw-Irr, $V = 1/2$ FWHM (Δv) (ref. 133), corrected for inclination. Dw Spheroidals are plotted twice; open circles, $M = 2.5\,L_{\rm B}$ (stars only) and $V = (GM/R_{\rm vir})^{1/2}$; large dots, $M = 30\,L_{\rm B}$ (stars plus dark matter[48-51]), with V as before. Light diagonal lines represent the masses of self-gravitating bodies with the indicated values of $n_{\rm b}$ and T and assuming $\rho_{\rm b} = 0.1\rho$. The discontinuities near 10^4 K are due to the effects of H and He ionization. Cooling curves[78,135,136] (medium lines) separate regions where cooling is efficient ($\tau_{\rm cool} < \tau_{\rm dyn}$, lower region) from regions where it is inefficient ($\tau_{\rm cool} > \tau_{\rm dyn}$, upper region). All curves assume a residual electron fraction after recombination of 10^{-4}, consistent with $\Omega_{\rm b}/\Omega = 0.1$. Model curves represent the equilibria of structures that collapse dissipationlessly from the cold dark matter initial fluctuation spectra of Fig. 2 with $n = 1$. DM haloes and groups and clusters of galaxies should lie on these curves, whereas the baryon components of galaxies should lie below due to baryonic dissipation. The curves labelled $v\sigma$ refer to fluctuations with $\delta M/M$ equal to v times the r.m.s. value shown in Fig. 2. Heavy curves: $\Omega = 1$, $h = 0.5$; dashed curves: $\Omega = 0.2$, $h = 1$. Model curves appear to pass through groups and clusters in about the right place, and the horizontal spread in T is roughly as expected for a gaussian distribution in $\delta M/M$. Baryon components of galaxies lie below the loci for dissipationless collapse and generally within the region where strong baryonic cooling is expected. Dwarf spheroidal galaxies lie near the beginning of the clustering hierarchy and may be typical of the earliest structures to collapse (see ref. 73). Hubble types are spread out along different loci, perhaps due in part to different baryon collapse factors (Es larger, Irrs smaller) and in part to intrinsic differences in initial $\delta M/M$ (see also Fig. 4).

'Solar metals', 'H_2', and 'Compton'. Below these curves the baryonic cooling time is shorter than the dynamical time and above them the reverse is true[71]. Figure 3 immediately shows that, while the Hubble sequence of galaxies shows strong evidence for baryonic cooling and dissipation (core condensation in heavy haloes), dwarf spheroidals are only marginally able to cool, and groups and clusters of galaxies have too long a cooling time to have dissipated much energy on their scale[72,73].

On average, for the cold DM fluctuation spectrum, what range of total masses yields baryon condensation within a massive halo? Let us first consider large galaxies. Naively, Fig. 3 suggests

that there is an upper bound for galactic masses of $M \lesssim 10^{12}\,M_\odot$, where the baryonic cooling time for gas of primordial composition begins to exceed the dynamical time. More massive galaxies could form in dense regions, where perhaps an early generation of Population III stars enriched the remaining baryonic gas with metals. Actually, the situation is likely to be more complicated, with many large galaxies formed by the merger of smaller ones. The resulting galaxy mass distribution arises from competition between hierarchical clustering and decreasing galaxy collision cross-sections due to dissipation, and it is thus difficult to calculate reliably. Regarding the smallest galaxies, collisional excitation of atomic hydrogen provides a lower limit of $M \gtrsim 10^8\,M_\odot$, corresponding to virialized baryonic temperature $T_b \gtrsim 10^4$ K. This range of protogalactic total masses, $10^8\,M_\odot \lesssim M \lesssim 10^{12}\,M_\odot$, encompasses virtually all the mass that is observed to comprise galaxies. For protogalaxies in this mass range, the velocity dispersion of the baryons will initially remain nearly constant ($T \approx$ constant) as they condense within the gravitational potential of the virialized (and presumably roughly isothermal) DM halo. When the baryon density increases enough that their gravitational potential dominates that of the halo, the baryons' velocity dispersion will rise as they continue to dissipate energy and condense. Baryonic contraction is finally halted by rotation and perhaps, in some protogalaxies, by star formation (see below).

The collapse of fluctuations with mass $M > 10^{13}\,M_\odot$ leads to clusters of galaxies in this picture. In clusters, only the outer parts of member galactic haloes are stripped off by collisions—the inner baryonic cores are able to contract to form the observed stellar systems. More of the baryons in the richest clusters are observed to be in the form of hot gas than in galaxies, as we have already mentioned in connection with Fig. 1. Perhaps this is because rich clusters tend to contain high-density cores, which collapse early, simultaneously with many of the galaxies they contain.

Can the cold DM picture account for the wide range of morphologies displayed by clusters of galaxies in X-ray[74] and optical-band[75] observations, ranging from regular, apparently relaxed configurations to complex, multicomponent structures? Preliminary results are encouraging. In particular, simulations show that large central condensations form quickly and can grow by subsequent mergers to form cD galaxies if most of the DM is in haloes around the baryonic substructures, as expected for cold DM, but not if the DM is distributed diffusely[76,77].

What happens to small clouds whose baryonic mass M_b lies in the range $M_{J,b} < M_b < 10^{7-8}\,M_\odot$ and for which $T_b < 10^4$ K after virialization? For a primordial element abundance, the molecular cooling time (primarily due to H_2) is less than the dynamical time for fluctuations that satisfy

$$M_b > M_{C,b} \approx 4.1\times10^6(\Omega h^2)^{-0.917}\left(\frac{Y_e}{10^{-4}}\right)^{-0.625}\left(\frac{\Omega_b}{0.1\Omega}\right)^{-1.04} \times\left(\frac{1+z_t}{10}\right)^{-2.75} M_\odot$$

where z_t is the redshift of maximum expansion of the fluctuation. (This result uses the cooling rates given by Yoneyama[78] and assumes the simple 'top hat' model of collapsing fluctuations. It corrects an equation given in ref. 17.) The quantity Y_e, the fraction of free electrons, is essentially the fraction that escapes recombination as the universe cools below ~1,000 K, and is roughly $10^{-4}(\Omega_b/0.1)^{-1}\Omega^{1/2}h^{-1}$. Systems with $M_b < M_{C,b}$ will persist as pressure supported clouds until they are disrupted by the virialization of larger scale clusters. Clouds having $M_b > M_{C,b}$ can dissipate energy and collapse, although the influence of rotation and the efficiency of fragmentation are poorly understood. The end product may be an irregular or dwarf spheroidal galaxy. Other possibilities such as a protoglobular cluster or one or more very massive objects (VMOs) would require greater contraction and are likely to be inhibited by angular momentum.

Moreover, it is unclear what fraction of the original baryonic mass can be retained by such small clouds rather than expelled. There is—despite the uncertainties—a strong probability that energy output from fragmented subsystems can influence the gas that remains uncondensed, thereby exerting a feedback on further condensation. In particular, UV emission from massive or supermassive stars can photoionize the remaining diffuse baryons, raising T_b up to 10^4 K. If such stars radiate a fraction q_{UV} of their Eddington luminosity in Lyman continuum photons, then the entire baryonic medium can be ionized even by a mass fraction of stars as low as $\sim 3\times10^{-5}/q_{UV}$. If the baryonic gas is re-ionized in systems with total mass $M \lesssim 10^8\,M_\odot$, the baryons become so hot that they flow out of the cloud, whose gravitational field is not strong enough to bind them.

This may explain the origin of dwarf spheroidal galaxies, with $10^6\,M_\odot < M < 10^8\,M_\odot$, which show little evidence of dissipation in Fig. 3. As the baryons in these systems begin to condense through molecular cooling, if fragmentation and star formation are very efficient the baryons quickly turn into stars and condensation ceases. On the other hand, if a fraction of the baryons fragment into stars having strong UV luminosity, the remaining baryons are heated until they leave the system, and dissipation will again cease. The apparent absence of dissipation in dwarf spheroidals could also be due to stripping of their baryons in encounters with more massive systems[32], which would move the dwarfs upward in Fig. 3.

In the cold DM scenario, globular clusters are probably not primordial objects. For one thing, there is no evidence that they have massive haloes (although, as Peebles[79] points out, there is not much direct contrary evidence either). Furthermore, if truly primordial, they should be distributed in the Universe like DM whereas, at least within galaxies, they seem to have dissipated and condensed like the other baryonic matter. However, the existence of 'standard objects' inside galaxies with mass $\sim 10^6\,M_\odot$ demands some explanation. In the model discussed here, there is a natural mass scale of order $M_{g,b} \sim M_{Gal,b}(T_{virial}/10^4\,K)^{-2}$, where $M_{Gal,b}$ is the baryonic mass within a galaxy; $M_{g,b}$ is the Jeans mass of a cloud at 10^4 K in pressure balance with protogalactic gas at the virial temperature[80]. During the dissipation phase of galaxy formation, the gas might be likely to have a two-phase structure with a hot phase at T_{virial} and a cool phase at $\sim 10^4$ K, in which case subcondensations of mass $M_{g,b}$ would be expected with density contrast $\approx T_{virial}/10^4$ K. We identify these with protoglobular clusters (noting that the considerations of Fall and Rees[81] could further limit the mass range).

Galaxies and clusters

While the n_b–T plot (Fig. 3) is useful for comparing data and predictions with the cooling curves, it is also useful to consider total mass M versus T, as in Fig. 4. This avoids having to take into account the differing amounts of baryonic dissipation suffered by various galaxies. The heavy solid and dashed curves again correspond to the $n = 1$ cold DM spectrum, for ($\Omega = 1$, $h = 0.5$) and ($\Omega = 0.2$, $h = 1$) respectively. It is striking that galaxies in the M–T diagram lie along lines of roughly the same slope as these curves. This occurs because the effective slope of the $n = 1$ cold DM fluctuation spectrum in the galaxy mass range is $n_{eff} \approx -2$, which corresponds to the empirical Tully–Fisher and Faber–Jackson laws: $M \propto v^4$. The light dashed lines in Fig. 4 are the postvirialization curves for primordial fluctuation spectra with $n = 0$ (white noise) and $n = 2$. The $n = 1$ (Zeldovich) spectrum is evidently the one that is most consistent with the data.

The points in Fig. 4 represent essentially all of the clusters identified by Geller and Huchra[82] in the CfA catalogue within 5,000 km s^{-1}. The cluster data lie about where they should on the diagram, and even the statistics of the distribution seem roughly to correspond to the expectations represented by the 0.5, 1, 2, and 3σ curves. Note that the galaxies and clusters are neatly separated by the Compton cooling line of Fig. 3, suggesting that the era of galaxy formation ceased in the Universe when Compton cooling off microwave background photons became inefficient (J. E. Gunn, personal communication).

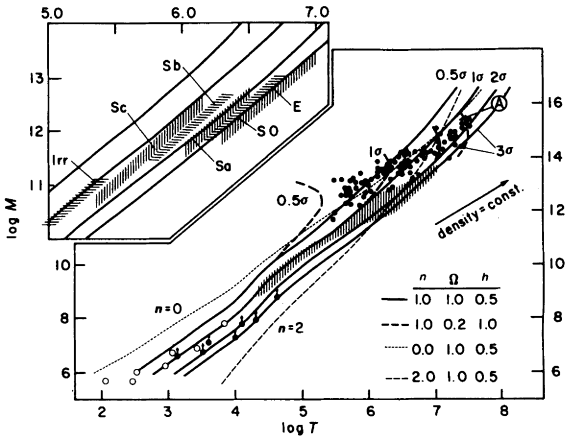


Fig. 4 Total mass M versus virial temperature T. Data sources and symbols are as in Fig. 3. M for groups and clusters is total dynamical mass. For galaxies, M is assumed to be 10 M_{lum} (see Fig. 3 legend). If Dw spheroidals actually have $M/L_B = 30$, they may have suffered baryon stripping[32], in which case M is a lower limit (arrows). Details of the region occupied by massive galaxies are shown in the inset, upper left. In addition to the $n = 1$ models from Fig. 3, two 1σ curves for $n = 0$ and $n = 2$ are also shown (light dashes). Either set of curves for an $n = 1$ (Zeldovich) spectrum provides a good fit to the observations over 9 orders of magnitude in mass. Curves with $n = 0$ and $n = 2$ do not fit as well. The apparent gap between galaxies and groups and clusters in Fig. 3 (which stems from baryonic dissipation) vanishes in this figure, and the clustering hierarchy is smooth and unbroken from the smallest structures to the largest ones. The Fisher-Tully and Faber-Jackson laws for galaxies ($M \propto V^4$ or T^2) arise naturally as a consequence of the slope of the cold DM fluctuation spectrum in the mass region of galaxies. Groups and clusters are distributed around the $n = 1$ loci about as expected. The apparent upward trend among the groups is not physically meaningful but arises from their selection as density enhancements above a minimum threshold (see caption, Fig. 3, and constant-density arrow, this figure). The exact locations of galaxies are uncertain. In particular, the temperatures of Es and S0s may be overestimated owing to the use of nuclear rather than global velocity dispersions. The masses of Dw. Irrs may also be too low owing to neglect of mass in H_2. Taken at face value, however, the data suggest that early-type galaxies (Es and S0s) arise from high-$\delta M/M$ fluctuations, whereas late-type galaxies (Scs and Irrs) arise from low-$\delta M/M$ fluctuations. Groups and clusters appear to fill a wider band than galaxies. If real, this difference may indicate that low-$\delta M/M$ fluctuations on the mass scale of galaxies once existed but did not give rise to visible galaxies. This suggests further that galaxy formation, at least in some regions of the Universe, may not have been fully complete and that galaxies are therefore not a reliable tracer of total mass. The sharp demarcation line between galaxies and clusters corresponds to the Compton cooling limit in Fig. 3 and suggests that galaxy formation ceased when Compton cooling off the microwave background became inefficient.

Note that spiral galaxies lie roughly along the 1σ curve while elliptical galaxies lie along the 2σ curve. Although this displacement is not large compared with the uncertainties, it is consistent with the fact that more than half of all galaxies are spirals, while only ~15% are ellipticals. We have suggested elsewhere[83] that, in hierarchical clustering scenarios, the higher σ fluctuations will develop rather smaller angular momenta, as measured by the dimensionless parameter $\lambda (= JE^{1/2}G^{-1}M^{-5/2})$. This difference seems to exist with either white noise or a flatter spectrum, but to be larger in the latter case. If high σ fluctuations have little angular momentum, their baryons can collapse by a large factor in radius, forming high-density ellipticals and spheroidal bulges, as shown in Fig. 3. Because, with a flat spectrum, higher σ fluctuations occur preferentially in denser regions destined to become rich clusters (the statistics of such correlations can be treated[84] by the methods of Peebles[6]), one expects to find more ellipticals there—as is observed. Indeed, the rich clusters lie along the same 2 and 3σ curves in Fig. 4 as do the elliptical galaxies.

Presumably the collapse of the low-λ protoelliptical galaxies is halted by star formation well before a flattened disk can form, yielding a stellar system of spheroidal shape. The mechanism governing the onset of star formation in these systems is unfortunately not yet understood, but may involve a threshold effect which sets in when the baryon density exceeds the DM halo density by a sufficient factor[70,72]. Disks (spirals and irregulars) form from more typical higher-λ protogalaxies, which, for a given mass, are larger and more diffuse than their protoelliptical counterparts. The collapse of disks thus occurs through a relatively slow infall of baryons from $\sim 10^2$ kpc, and is halted by angular momentum. Infall from such distances is consistent both with the extent of dark haloes inferred from observations and with the high angular momenta of present-day disks ($\lambda \sim 0.4$)[85,86]. The location of galaxies in Fig. 3 is consistent with these ideas if the baryons in all galaxies collapsed in radius by roughly the same factor, about an order of magnitude, but less for late-type irregulars and more for early-type Es and spheroidal bulges.

It has been theorized that the Hubble sequence originates in the distribution of either the initial angular momenta[87,88] or else the initial densities[89] of protogalaxies. However, if overdensity and angular momentum are linked, with the high-σ fluctuations having lower λ, then these two apparently competitive theories become opposite sides of the same coin.

Consider finally the difference in Fig. 4 between the solid and dashed lines. The dashed lines, representing a lower-density Universe ($\Omega = 0.2$), curve backward at the largest masses and lie far away from the circle representing the cores of the richest clusters, Abell classes 2 and 3. Because these regions of very high galaxy density contain at least several per cent of the mass in the Universe, the circle should lie between the 2 and 3σ lines (assuming gaussian statistics). It does so for the solid ($\Omega = 1$) lines, but not for the dashed lines. At face value, this is evidence favouring an Einstein-de Sitter Universe for cold DM. However, there are at least two reasons why this argument should not be taken too seriously. First, the velocity dispersions represented by the Abell cluster circle in Fig. 4 correspond to the cluster cores. The model curves on the other hand refer to the entire virialized cluster, over which the velocity dispersion is considerably lower (as indicated schematically by the arrow attached to the circle in Fig. 4). Second, the assumption of spherical symmetry used in obtaining both sets of curves is only an approximation. The initial collapse is probably often quite anisotropic—more like a Zeldovich pancake than a sphere. It is, therefore, preferable to compare these data with N-body simulations than with the simple model represented by the curves in Fig. 4. Until this becomes possible we do not believe that the data in Fig. 4 allow a clear-cut discrimination between the $\Omega = 0.2$ and $\Omega = 1$ cases, especially if the Hubble parameter h is allowed to vary simultaneously within the observationally allowed range, as we have assumed.

Cosmological density

The most straightforward interpretation of the approximate constancy of $M/M_{lum} \approx 10$ from galaxy through rich cluster scales (Fig. 1) is that dark matter clusters with galaxies. This is precisely what cold DM is expected to do. One then expects the density parameter $\Omega = (M/M_{lum})\Omega_{galaxies} \approx 10 \times 0.02 = 0.2$, in agreement with $\Omega = 0.2 \times 1.5^{\pm 1}$ from the cosmic energy equation and the stability of clustering[90].

The value $\Omega \approx 0.2$ is consistent with all reliable measurements and furthermore gives an age for the Universe, t_0, consistent with globular cluster age estimates[91] (≥ 15 Gyr) if $h \leq 0.6$. The observed abundance of deuterium plus helium-3 implies a lower limit[37] $\Omega_b h^2 \geq 0.01$, which is also consistent with $\Omega = 0.2$ and $\Omega_b \approx 0.02$ if $h \geq 0.7$.

Another argument favouring $\Omega = 0.2$ for cold DM is based on preliminary N-body results[92], which indicate that superclusters and voids form on the observed scale for $\Omega = 0.2$, but on too small a scale for $\Omega = 1$ unless the Hubble constant is unrealistically small. However, this conclusion follows from comparing the N-body mass autocorrelation function $\xi_m(r)$ with the observed galaxy autocorrelation function $\xi_g(r)$. This is justified only if the galaxies are a good tracer of mass—if $M/L \approx$ constant. This is certainly not true for rich clusters, as illustrated in the top part of Fig. 1 and in Table 1, where M/L for very rich clusters is roughly six times larger than it is for small groups

and for the Milky Way (including its DM halo). This means that some of the actual small-distance-mass autocorrelation is not included in the galaxy autocorrelation function; that $\xi_m(r)$ is steeper for $r \leq 1$ Mpc than $\xi_g(r)$. This effect is in the right direction to bring the $\Omega = 1$ N-body simulations into consistency with observations. Note that galaxy velocity dispersion averages also underestimate the actual mass-weighted velocity dispersion, because the high-velocity, dense regions of rich clusters are under-represented in the counts. This effect is also compatible with higher Ω.

The absence of measurable fluctuations in the cosmic microwave background also restricts Ω, because in a low-Ω Universe there is less growth of fluctuations both before and after recombination. The latest limits[38,39] on $\Delta T/T$ require $\Omega \geq 0.2h^{-4/3}$ (refs 93, 94), unless the Universe was reionized after recombination.

Both prejudice and the inflationary hypothesis favour the Einstein-de Sitter value $\Omega = 1$. (Actually, inflation implies more generally that $\Omega = 1 - \Lambda/3H^2$, where H is Hubble's constant and Λ is the cosmological constant[95], but we assume here that $\Lambda = 0$.) For $\Omega = 1$, $t_0 = 6.5h^{-1}$ Gyr, so consistency with globular cluster age estimates requires the Hubble constant to be perhaps unrealistically small. As we have discussed, the Zeldovich ($n = 1$) spectrum of primordial adiabatic fluctuations, which also follows from inflation, is compatible with models of galaxy formation in the cold DM scenario. Of course, the Zeldovich spectrum does not necessarily entail inflation. Moreover, inflation does not necessarily imply $\Omega = 1$: if there is a great deal of inflation, then, of course, Ω is very close to unity (assuming vanishing cosmological constant); but if we speculate that greater amounts of inflation are increasingly unlikely, then our horizon might happen to lie in a patch with $\Omega = 0.2$.

Could $\Omega = 1$ in the cold DM model? Based on Fig. 1, this could happen only if M/M_{lum} increases substantially on scales larger than rich clusters. Perhaps galaxy formation is suppressed in the voids, and the resulting luminosity contrast is so strong, for reasons we do not yet understand, that $M/L > 1{,}500$ there (see below). Alternatively, if the virial mass of clusters of galaxies is significantly underestimated, as suggested recently by Kaiser, and if galactic haloes extend much further than ~70 kpc, then the ratio M/M_{lum} may be substantially underestimated in Fig. 1, leading to a larger value for Ω. If galaxy formation is inefficient so that only large-overdensity perturbations can form galaxies—a possibility suggested by Fig. 4—the $\Omega = 1$, cold DM N-body simulations[92] can be brought into agreement with observations[96,97]. We conclude that a straightforward interpretation of the evidence summarized above favours $\Omega \approx 0.2$ in the cold DM picture, but that $\Omega = 1$ is not implausible.

Note that the cold DM model may require some means of suppressing galaxy formation in voids for $\Omega \approx 0.2$ as well as for $\Omega = 1$. This is required in the latter case to hide most of the mass. In a low-Ω Universe, on the other hand, large, very underdense regions cannot form by gravitation alone[98,99]; and galaxy formation must be suppressed somehow in regions of moderately low density if the density of galaxies in voids is less than one-quarter of the average density, the quoted upper limit for the Boötes void[100,101]. Suppression of galaxy formation may occur partly because of the substantial difference at $z < \Omega^{-1}$ between $\langle\rho\rangle$ and ρ_c in a low-density Universe.

Superclusters and voids

Recent, accurate redshift measurements of several thousand galaxies have revealed the presence of voids with linear dimensions ranging up to $\sim 50h^{-1}$ Mpc that are almost completely empty of bright galaxies. Most galaxies are concentrated in irregularly shaped, flattened, or elongated superclusters (see ref. 102 for fuller discussion and references). Two related observations are particularly interesting for their possible bearing on the origin of this large scale structure: a correlation of galaxy type with galaxy number density[103,104] and, possibly, a correlation between the orientation of cluster major axes and the direction to neighbouring clusters within $\sim 15h^{-1}$ Mpc (ref. 105).

In the hot DM picture, the fluctuation spectrum is rather sharply peaked at masses $\sim 4 \times 10^{15}\,(30\text{ eV}/m)^2$. The suppression of smaller scale structure results in an essentially pressure-free collapse on the supercluster scale, with the formation of sharp caustics—Zeldovich pancakes[106,107]. If galaxies form only in the densest regions, where the shocked baryons have cooled sufficiently[108,109], most of the Universe would be essentially devoid of galaxies. Simulations[110] also show large-scale cluster orientation correlations resembling those observed. On the other hand, the rapid evolution of the autocorrelation function requires that the pancakes form at $z \leq 2$, which is uncomfortably recent if smaller structures including galaxies must form subsequently.

The cold DM scenario avoids this latter problem because galaxies and clusters would already have formed by the time of supercluster collapse. Would superclusters and voids arise in a cold DM Universe? There are good reasons to believe that pancakes, filaments, and voids would indeed form, and preliminary indications from N-body simulations suggest that they do[92,111]. The cold and hot DM fluctuation spectra are identical on the very largest scales, differing only in that the latter is cut off below $\sim 10^{15}\,M_\odot$ by neutrino free streaming and normalized larger above this scale (see Fig. 2a). The presence of already-formed and partially-virialized substructure in the cold DM case causes supercluster collapse to be less dissipative than in the neutrino picture because the caustics are thickened and because a smaller fraction of the baryons remains as cold, uncondensed gas. Nevertheless, as Dekel[112] has shown, persistent flattened or elongated structures can form even in the absence of dissipation as a result of continued expansion in directions orthogonal to the collapse. Indeed, the sharp caustics and highly dissipative shocks in the neutrino picture may produce superclusters that are too flat compared with the observations, even when gravitational interactions with neighbouring superclusters are taken into account.

Another related difference between hot and cold DM is that the hot DM Universe is predicted to have a rather simple cellular structure while the cold DM Universe probably has a considerably richer structure, perhaps more like that observed. In particular, the sizes of superclusters and voids in the cold case should span a fairly broad range. This will be an important test of the models when enough galaxy redshift data become available to indicate the statistics of the void distribution.

What should one expect to find in the voids? In the hot DM picture, all galaxies form from the dense gas along the caustics. Hence, the centres of voids should be entirely empty of galaxies and should contain only low-density DM and hot primordial gas (heated by radiation from pancake shocks and too dilute to have cooled)[113]. In the cold DM picture, one might at first suppose that galaxies form more or less uniformly in space, with their density subsequently enhanced by gravitational clustering and the pancake distortion of the Hubble flow. One would then expect to find galaxies in the voids, although with lower density than in superclusters.

In a low-density Universe, both analytical calculations[98,99] and N-body simulations[92] suggest that large regions having $\rho/\langle\rho\rangle \leq 0.2$ cannot form by gravitational clustering alone. Nonetheless, large regions with few bright galaxies may actually be more likely in a low density cold DM Universe, for the following reason. On average, large galaxies form late when the average density has dropped well below the critical density. Larger initial fluctuations are thus required to form large galaxies in low-density regions. But larger fluctuations are statistically less likely in low-density regions with the flat, cold DM spectrum. This suppresses formation of large galaxies in moderate-size voids, and formation of clusters in larger voids[84].

In addition, feedback from nongravitational processes could amplify the number density contrast of bright galaxies compared with the underlying density of dark and baryonic matter. For example, at $z \sim 10$ the average density is highest where pancakes will later occur. This is where most of the earliest galaxies and rich clusters form, and these will be the regions earliest enriched

in metals. Radiation from these early galaxies (and quasars, and possibly Population III stars or VMOs) could heat the gas in lower density regions, raising the Jeans mass, as discussed above, and suppressing small galaxy formation. Indeed, if this early radiation has a hard enough spectrum and heats the gas sufficiently rapidly, it could raise the gas temperature throughout the Universe to 10^6 K and essentially halt galaxy formation outside pancakes. One of us (M.J.R.) has recently discussed[114] processes for suppressing galaxy formation in protovoids. Finally, explosive shocks could also enhance galaxy formation in denser regions[115]. Although the efficacy of all these processes is uncertain and needs further investigation, we conclude that the existence of large regions in which the density of bright galaxies is low is probably not a serious problem for the cold DM picture.

Now let us briefly discuss the two types of correlations mentioned above. The observed correlation of galaxy type with galaxy number density[103,104] can, at least partly, be understood as a consequence of the greater statistical likelihood of large-σ fluctuations in regions of greater density (protoclusters), together with the effect discussed above that results in higher-σ fluctuations acquiring lower angular momenta on average and becoming elliptical galaxies or spheroidal bulges. There is also an environmental effect: lower-σ fluctuations yield dark haloes that are physically larger and more diffuse than higher-σ fluctuations. Disks, which form from such fluctuations, thus form slowly, by infall of gas from large radii within these extended haloes. Because large haloes have correspondingly large collision cross-sections, few disks can form in regions of high galaxy number density. In dense regions, the halo gas is stripped by collisions and is mixed with enriched gas from galactic winds to become the hot intergalactic medium observed in X rays. It will be interesting to investigate these effects with N-body simulations and a more detailed theory of galaxy formation.

Regarding the second type of correlation, Binggeli[105] has found that the position angles of nearby, elongated Abell clusters are within 45° of the direction to the nearest cluster, provided the two clusters are separated by $\lesssim 15h^{-1}$ Mpc. He found a similar, though less significant, correlation between the position angle of the brightest cluster galaxy's major axis and the direction to the nearest cluster. In the hot DM picture, where superclusters form before clusters and galaxies, such correlations arise naturally. Indeed an N-body simulation[110] has shown that such correlations also occur, but more weakly, even with a fluctuation spectrum having an adiabatic peak at $\sim 10^{15}\ M_\odot$ superimposed on an $n=0$ spectrum, so that rich clusters and superclusters form almost simultaneously. However, a preliminary analysis[116] of cold DM N-body simulations finds Bingelli-type correlations only if $\Omega h \leq 0.2$.

It is important to find other observationally accessible information that can discriminate between cosmological models. The large-scale velocity fields of superclusters should be rather different in the hot and cold DM schemes because of the much greater dissipation in the former. With new instruments it will also be possible to study the $z<2$ evolution of superclustering of quasars and of Lyα absorbing clouds, and perhaps the density and composition of the gas in voids.

It was once hoped that percolation analysis of the large scale galaxy distribution could help to distinguish between different cosmological models.[11,117]. However, when this analysis is applied to the CfA galaxy data, the results are found to be a sensitive function of the depth of the survey[118]. Moreover, realistic N-body simulations of isothermal and adiabatic scenarios (that is, with accurate treatment of gravitational interactions on small as well as large scales) have nearly identical percolation properties[110,119]. Better statistical tests are needed to compare objectively the large-scale distribution of matter in models and observations.

Conclusions

We have shown that a Universe with ~10 times as much cold dark matter as baryonic matter provides a remarkably good fit to the observed Universe. This model predicts roughly the observed mass range of galaxies, the dissipational nature of galaxy collapse, and the observed Faber-Jackson and Tully-Fisher relations. It also gives dissipationless galactic haloes and clusters. In addition, it may also provide natural explanations for galaxy-environment correlations and for the differences in angular momenta between ellipticals and spiral galaxies. Finally, the cold DM picture seems reasonably consistent with the observed large-scale clustering, including superclusters and voids. In short, it seems to be the best model available and merits close scrutiny and testing.

We thank Dick Bond, Jim Bardeen, Marc Davis, Avishai Dekel, George Efstathiou, Mike Fall, Carlos Frenk, Margaret Geller, Jim Gunn, Jim Peebles, Nicola Vittorio, and Simon White for valuable conversations. In addition, J.R.P. acknowledges the hospitality of the Max Planck Institute for Physics and Astrophysics, Munich during the summer 1983. This work was supported by the Department of Energy under contract no. DE-AC03-76SF00515 and by NSF grants AST-82-11551 and PHY-81-15541. This is Lick Observatory Contribution No. 436.

1. Faber, S. M. & Gallagher, J. S. *A. Rev. Ast. Astrophys.* **17**, 135-187 (1979).
2. Peebles, P. J. E. *Astrophys. J. Lett.* **263**, L1-L5 (1982).
3. Primack, J. R. & Blumenthal, G. R. *Formation and Evolution of Galaxies and Large Structures in the Universe* (eds Audouze, J. & Tran Thanh Van, J. 163-183 (Reidel, Dordrecht, 1983).
4. Primack, J. R. & Blumenthal, G. R. *4th Workshop on Grand Unification* (eds Weldon, H. A., Langacker, P. & Steinhardt, P. J.) 256-288 (Birkhauser, Boston, 1983).
5. Primack, J. R. & Blumenthal, G. R. *Clusters and Groups of Galaxies* (eds Mardirossian, F., Giuricin, G. & Mezzetti, M.)(Reidel, Dordrecht, in the press).
6. Peebles, P. J. E. *Astrophys. J.* **277**, 470-477 (1984).
7. Blumenthal, G. R. & Primack, J. R. Preprint (SLAC-PUB-3388, 1984).
8. Peebles, P. J. E. *The Large Scale Structure of the Universe* (Princeton University Press, 1980).
9. Efstathiou, G. & Silk, J. *Fundm. Cosmic Phys.* **9**, 1-138 (1983).
10. Lyubimov, V. A. *et al. Phys. Lett.* **B94**, 266-268 (1980).
11. Zeldovich, Ya. B., Einasto, J. & Shandarin, S. F. *Nature* **300**, 407-413 (1982).
12. Shandarin, S. F., Doroshkevich, A. G. & Zeldovich, Ya. B. *Sov. Phys. Usp.* **26**, 46-76 (1983).
13. Pagels, H. R. & Primack, J. R. *Phys. Rev. Lett.* **48**, 223-226 (1982).
14. Olive, K. A. & Turner, M. S. *Phys. Rev.* **D25**, 213-216 (1982).
15. Blumenthal, G. R., Pagels, H. & Primack, J. R. *Nature* **299**, 37-38 (1982).
16. Bond, J. R., Szalay, A. S. & Turner, M. S. *Phys. Rev. Lett.* **48**, 1636-1639 (1982).
17. Bond, J. R. & Szalay, A. S. *Astrophys. J.* **274**, 443-468 (1984).
18. Ipser, J. & Sikivie, P. *Phys. Rev. Lett.* **50**, 925-927 (1983).
19. Peccei, R. & Quinn, H. *Phys. Rev.* **D16**, 1791-1797 (1977).
20. Weinberg, S. *Phys. Rev. Lett.* **40**, 223-226 (1978).
21. Wilczek, F. *Phys. Rev. Lett.* **40**, 279-282 (1978).
22. Preskill, J., Wise, M. & Wilczek, F. *Phys. Lett.* **120B**, 127-132 (1983).
23. Abbott, L. & Sikivie, P. *Phys. Lett.* **120B**, 133-136 (1983).
24. Dine, M. & Fischler, W. *Phys. Lett.* **120B**, 137-141 (1983).
25. Fukugita, M., Watamura, S. & Yoshimura, M. *Phys. Rev. Lett.* **48**, 1522-1525 (1982).
26. Sikivie, P. *Phys. Rev. Lett.* **51**, 1415-1417 (1983); erratum **52**, 695 (1984).
27. Ellis, J., Hagelin, J. S., Nanopoulos, D. V., Olive, K. & Srednicki, M. *Nucl. Phys. B* **238**, 453-476 (1984).
28. Stecker, F. W. & Shafi, Q. *Phys. Rev. Lett.* **50**, 928-931 (1983).
29. Freese, K., Price, R. & Schramm, D. N. *Astrophys. J.* **275**, 405-412 (1983).
30. Carr, B. J. *Commts Astrophys.* **7**, 161-173 (1978).
31. Lacey, C. G. *Formation and Evolution of Galaxies and Large Scale Structures in the Universe* (eds Audouze, J. & Tran Thanh Van, J.) 351-360 (Reidel, Dordrecht, 1984).
32. Lin, D. N. C. & Faber, S. M. *Astrophys. J. Lett.* **266**, L21-L25 (1983).
33. Canizares, C. R. *Astrophys. J.* **263**, 508-517 (1982).
34. Witten, E. *Phys. Rev.* **D30**, 272-285 (1984).
35. Carr, B. J., Bond, J. R. & Arnett, W. D. *Astrophys. J.* **277**, 445-469 (1984).
36. Hegyi, D. J. & Olive, K. A. *Phys. Lett.* **126B**, 28-32 (1983).
37. Yang, J., Turner, M. S., Steigman, G., Schramm, D. N. & Olive, K. A. *Astrophys. J.* (in the press).
38. Uson, J. M. & Wilkinson, D. T. *Astrophys. J. Lett.* **277**, L1-L4 (1984).
39. Uson, J. M. & Wilkinson, D. T. *Inner Space/Outer Space* (eds Kolb, E. W. *et al.*) (University of Chicago Press, in the press).
40. Weinberg, S. *Astrophysical Cosmology: Proc. Study Week on Cosmology and Fundamental Physics* (eds Brück, H. A., Coyne, G. V. & Longair, M. S.) 503-528 (Pontifical Scientific Academy, Vatican, 1982).
41. Kolb, E. W. & Turner, M. S. *A. Rev. Nucl. Part. Sci.* **33**, 645-696 (1983).
42. Rees, M. J. *The Very Early Universe* (eds Gibbons, G., Hawking, S. & Siklos, S.) 29-58 (Cambridge University Press, 1983).
43. Frenk, C., White, S. D. M. & Davis, M. *Astrophys. J.* **271**, 417-430 (1983).
44. Dekel, A. & Aarseth, S. *Astrophys. J.* **283** (in the press).
45. Kaiser, N. *Astrophys. J. Lett.* **273**, L17-L20 (1983).
46. Faber, S. M. *Proc. 1st ESO-CERN Symp., Large Scale Structure of the Universe, Cosmology, and Fundamental Physics* (in the press).
47. Bond, J. R., Szalay, A. S. & White, S. D. M. *Nature* **301**, 584-585 (1983).
48. Aaronson, M. *Astrophys. J. Lett.* **266**, L11-L15 (1983).
49. Aaronson, M. & Cook, K. *Bull. Am. astr. Soc.* **15**, 907 (1983).
50. Cook, K., Schechter, P. & Aaronson, M. *Bull. Am. astr. Soc.* **15**, 907 (1983).
51. Faber, S. M. & Lin, D. N. C. *Astrophys. J. Lett.* **266**, L17-L20 (1983).
52. Davis, M., Lecar, M., Pryor, C. & Witten, E. *Astrophys. J.* **250**, 423 (1981).
53. Hut, P. & White, S. D. M. *Nature* **310**, 637-644 (1984).
54. Doroshkevich, A. G. & Khlopov, M. Yu. *Mon. Not. R. astr. Soc.* (in the press).
55. Turner, M. S., Steigman, G. & Krauss, L. M. Preprint (Fermilab, 1984).
56. Primack, J. R. & Roncadelli, M. Preprint (SLAC-PUB-3304, 1984).
57. Shafi, Q. & Stecker, F. W. Preprint (Bartol Research Foundation, 1984).
58. Gelmini, G., Schramm, D. N. & Valle, J. W. F. Preprint (CERN-TH-3865, 1984).
59. Harrison, E. R. *Phys. Rev.* **D1**, 2726-2730 (1970).
60. Peebles, P. J. E. & Yu, J. T. *Astrophys. J.* **162**, 815-836 (1970).
61. Zeldovich, Ya. B. *Mon. Not. R. astr. Soc.* **160**, 1P-3P (1972).

62. Shafi, Q. & Vilenkin, A. *Phys. Rev. Lett.* **52**, 691–694 (1984).
63. Steinhardt, P. J. & Turner, M. S. Preprint (Fermilab, 1984).
64. Quinn, H. R. & Gupta, S. *Phys. Rev.* **D29**, 2791–2797 (1984).
65. Zeldovich, Ya. B. *Mon. Not. R. astr. Soc.* **192**, 663–667 (1980).
66. Vilenkin, A. *Phys. Rev.* **D24**, 2082–2089 (1981).
67. Bond, J. R., Efstathiou, G. & Silk, J. *Phys. Rev. Lett.* **45**, 1980–1984 (1980).
68. Doroshkevich, A. G., Zeldovich, Ya. B., Sunyaev, R. A. & Khlopov, M. Yu. *Soviet Astr. Lett.* **6**, 252–256 (1980).
69. Chernin, A. D. *Soviet Astr.* **25**, 14–16 (1981).
70. White, S. D. M. & Rees, M. J. *Mon. Not. R. astr. Soc.* **183**, 341–358 (1978).
71. Rees, M. J. & Ostriker, J. P. *Mon. Not. R. astr. Soc.* **179**, 541–559 (1977).
72. Faber, S. M. *Astrophysical Cosmology: Proc. Study Week on Cosmology and Fundamental Physics* (eds Brück, H. A., Coyne, G. V. & Longair, M. S.) 191–234 (Pontifical Scientific Academy, Vatican, 1982).
73. Silk, J. *Nature* **301**, 574–578 (1983).
74. Forman, W. & Jones, C. *Ann. Rev. Astr. Astrophys.* **20**, 547–585 (1982).
75. Geller, M. J. & Beers, T. C. *Proc. Astr. Soc. Paci.* **94**, 421–439 (1982).
76. Thuan, T. X. & Romanishin, W. *Astrophys. J.* **248**, 439–459 (1981).
77. Cavaliere, A., Santangelo, P., Tarquini, G. & Vittorio, N. *Clusters and Groups of Galaxies* (eds Mardirossian, F., Giuricin, G. & Mezzetti, M.) (Reidel, Dordrecht, in the press).
78. Yoneyama, T. *Publ. Astr. Soc. Jap.* **24**, 87–98 (1972).
79. Peebles, P. J. E. Preprint (Princeton Univ., 1984).
80. Fall, S. M. & Rees, M. J. (in preparation).
81. Fall, S. M. & Rees, M. J. *Mon. Not. R. astr. Soc.* **181**, 37P–42P (1977).
82. Geller, M. & Huchra, J. *Astrophys. J. Suppl.* **52**, 61–87 (1983).
83. Faber, S. M., Blumenthal, G. R. & Primack, J. R. *Astrophys. J. Lett.* (submitted).
84. Blumenthal, G. R., Faber, S. M. & Primack, J. R. (in preparation).
85. Efstathiou, G. & Jones, B. J. T. *Mon. Not. R. astr. Soc.* **186**, 133–144 (1979).
86. Fall, S. M. & Efstathiou, G. *Mon. Not. R. astr. Soc.* **193**, 189–206 (1980).
87. Sandage, A., Freeman, K. C. & Stokes, N. R. *Astrophys. J.* **160**, 831–844 (1970).
88. Efstathiou, G. & Barnes, J. *Formation and Evolution of Galaxies and Large Scale Structures in the Universe* (eds Audouze, J. & Tran Thanh Van, J.) 361–377 (Reidel, Dordrecht, 1984).
89. Gott, J. R. & Thuan, T. X. *Astrophys. J.* **204**, 649–667 (1976).
90. Davis, M. & Peebles, P. J. E. *Astrophys. J.* **267**, 465–482 (1983).
91. Sandage, A. *Proc. 1st ESO–CERN Symp., Large Scale Structure of the Universe, Cosmology, and Fundamental Physics* (in the press).
92. Davis, M., Efstathiou, G., Frenk, C. & White, S. D. M. (in preparation).
93. Bond, J. R. & Efstathiou, G. Preprint (NSF-ITP-84-79, 1984).
94. Vittorio, N. & Silk, J. Preprint (U. Cal. Berkeley, 1984).
95. Peebles, P. J. E. Preprint (Princeton Univ., 1984).
96. Kaiser, N. *Inner Space/Outer Space* (eds Kolb, E. W. *et al.*) (University of Chicago Press, in the press).
97. Bardeen, J. *Inner Space/Outer Space* (eds Kolb, E. W. *et al.*) (University of Chicago Press, in the press).
98. Hoffman, G. L., Salpeter, E. E. & Wasserman, I. *Astrophys. J.* **268**, 527–539 (1983).
99. Hoffman, Y. & Shaham, J. *Astrophys. J. Lett.* **262**, L23–L26 (1982).
100. Kirshner, R. F., Oemler, A., Schechter, P. L. & Schectman, S. A. *Astrophys. J. Lett.* **248**, L57–L60 (1981).
101. Kirshner, R. F., Oemler, A., Schechter, P. L. & Schectman, S. A. *Early Evolution of the Universe and its Present Structure, IAU Symp.* No. 104 197–202 (1983).
102. Oort, J. H. *A. Rev. Astr. Astrophys.* **21**, 373–428 (1983).
103. Dressler, A. *Astrophys. J.* **236**, 351–365 (1980).
104. Postman, M. & Geller, M. J. *Astrophys. J.* **281**, 95–99 (1984).
105. Binggeli, B. *Astr. Astrophys.* **107**, 338–349 (1982).
106. Zeldovich, Ya. B. *Astr. Astrophys.* **5**, 84–89 (1970).
107. Arnold, V. I., Shandarin, S. F. & Zeldovich, Ya. B. *Geophys. Astrophys. Fluid Dyn.* **20**, 111 (1982).
108. Bond, J. R., Centrella, J., Szalay, A. S. & Wilson, J. R. *Formation and Evolution of Galaxies and Large Structures in the Universe* (eds Audouze, J. & Tran Thanh Van, J.) 87–99 (Reidel, Dordrecht, 1984).
109. Shapiro, P. R., Struck-Marcell, C. & Melott, A. L. *Astrophys. J.* **275**, 413–429 (1983).
110. Dekel, A., West, M. J. & Aarseth, S. J. *Astrophys. J.* **279**, 1 (1984).
111. Melott, A., Einasto, J., Saar, E., Suisalu, I., Klypin, A. A. & Shandarin, S. F. *Phys. Rev. Lett.* **51**, 935–938 (1983).
112. Dekel, A. *Astrophys. J.* **264**, 373–391 (1983).
113. Doroshkevich, A. G., Shandarin, S. F. & Zeldovich, Ya. B. *Commt Astrophys.* **9**, 265–273 (1982).
114. Rees, M. J. *Clusters and Groups of Galaxies* (eds Mardirossian, F., Giuricin, G. & Mezzetti, M.) (Reidel, Dordrecht, in the press).
115. Ostriker, J. P. *Astrophysical Cosmology: Proc. Study Week on Cosmology and Fundamental Physics* (eds Brück, H. A., Coyne, G. V. & Longair, M. S.) 473–493 (Pontifical Scientific Academy, Vatican, 1982).
116. Dekel, A. *8th Johns Hopkins Workshop on Current Problems in Particle Theory* (eds Domokos, G. & Kovesi-Domokos, S.) (World Scientific Publishing, Singapore, in the press).
117. Shandarin, S. F. *Soviet Astr. Lett.* **9**, 104 (1983).
118. Barrow, J. D. & Bhavsar, S. P. *Mon. Not. R. astr. Soc.* **205**, 66 (1983).
119. Dekel, A. & West, M. J. Preprint (NSF-ITP-84-44, 1984).
120. Dressler, A. *Astrophys. J.* **226**, 55–69 (1978).
121. Jones, C. & Forman, W. *Astrophys. J.* **276**, 38–55 (1983).
122. Kriss, G. A., Cioffi, D. F. & Canizares, C. R. *Astrophys. J.* **272**, 439–448 (1983).
123. Beers, T. C., Geller, M. J., Huchra, J. P., Latham, D. W. & Davis, R. J. *Astrophys. J.* (in the press).
124. Gunn, J. E. *Astrophysical Cosmology* (eds Brück, H. A., Coyne, G. V. & Longair, M. S.) 233–259 (Pontifical Scientific Academy, Vatican, 1982).
125. Young, P. J. *Astr. J.* **81**, 807–816 (1976).
126. Kent, S. M. & Gunn, J. E. *Astr. J.* **87**, 945–971 (1982).
127. Kent, S. M. & Sargent, W. L. W. *Astr. J.* **88**, 697–708 (1983).
128. de Vaucouleurs, G., de Vaucouleurs, A. & Corwin, H. R. *2nd Reference Catalogue of Bright Galaxies* (University of Texas, Austin, 1976).
129. Dressler, A. & Sandage, A. R. *Astrophys. J.* **265**, 664–680 (1983).
130. Rubin, V. C. *Internal Kinematics and Dynamics of Galaxies* (ed. Athanassoula, E.) 3–6 (Reidel, Dordrecht, 1983).
131. Rubin, V. C., Ford, W. K., Thonnard, N. & Burstein, D. *Astrophys. J.* **261**, 439–456 (1982).
132. Burstein, D., Rubin, V. C., Thonnard, N. & Ford, W. K. *Astrophys. J.* **253**, 70–85 (1982).
133. Thuan, T. X. & Seitzer, P. O. *Astrophys. J.* **231**, 680–687 (1979).
134. Tonry, J. & Davis, M. *Astrophys. J.* **246**, 680–695 (1981).
135. Raymond, J. C., Cox, D. P. & Smith, B. W. *Astrophys. J.* **204**, 290–292 (1976).
136. Dalgarno, A. & McCray, R. A. *A. Rev. Astr. Astrophys.* **10**, 375–426 (1972).

THE ASTROPHYSICAL JOURNAL, 292:371–394, 1985 May 15
© 1985. The American Astronomical Society. All rights reserved. Printed in U.S.A.

THE EVOLUTION OF LARGE-SCALE STRUCTURE IN A UNIVERSE DOMINATED BY COLD DARK MATTER

MARC DAVIS,[1,2] GEORGE EFSTATHIOU,[1,3] CARLOS S. FRENK,[1,4] AND SIMON D. M. WHITE[1,5]
Received 1984 August 20; accepted 1984 November 30

ABSTRACT

We present the results of numerical simulations of nonlinear gravitational clustering in universes dominated by weakly interacting, "cold" dark matter (e.g., axions or photinos). These studies employ a high resolution N-body code with periodic boundary conditions and 32,768 particles; they can accurately represent the theoretical initial conditions over a factor of 16 in length scale. We have followed the evolution of ensembles of models with $\Omega = 1$ and $\Omega < 1$ from the initial conditions predicted for a "constant curvature" primordial fluctuation spectrum. We also ran one model of a flat universe with a positive cosmological constant. Large filamentary structures, superclusters of clumps, and large low-density regions appear at certain times in all our simulations; however, we do not find large regions as extreme as the apparent void in Boötes. The evolution of the two-point correlation function, $\xi(r)$, is not self-similar; its effective power-law index becomes more negative with time. Models with $\Omega = 1$ are inconsistent with observation if galaxies are assumed to be unbiased tracers of the underlying mass distribution. The peculiar velocities of galaxies are predicted to be much too large. In addition, at times when the shape of $\xi(r)$ matches that observed, the amplitude of clustering is inferred to be too small for any acceptable value of the Hubble constant. Better agreement is obtained for $\Omega = 0.2$, but in both cases the rms relative peculiar velocity of particle pairs decreases markedly with pair separation, whereas the corresponding quantity for galaxies is observed to increase slowly. In all models the three-point correlation function ζ is found to fit the observed form, $\zeta \propto Q\xi^2$, but with Q depending weakly on scale. On small scales Q substantially exceeds its observed value. Consistent with this, the mass distribution of clusters is very broad, showing the presence of clumps with a very wide range in mass at any given time. The model with a positive cosmological constant closely resembles an open model with the same value of Ω. If galaxies are a random sampling of the mass distribution, none of our models is fully consistent with observation. An alternative hypothesis is that galaxies formed only at high peaks of the initial density field. The clustering properties of such "galaxies" are biased; they appear preferentially in high-density regions and so are more correlated than the overall mass distribution. Their two- and three-point correlation functions and their relative peculiar velocity distribution may be consistent with observation even in a universe with $\Omega = 1$. If this is an appropriate model for galaxy formation, it may be possible to reconcile a flat universe with most aspects of the observed galaxy distribution.

Subject headings: galaxies: clustering — galaxies: formation — numerical methods

I. INTRODUCTION

The fundamental problem of understanding the structure of our universe can be approached by phenomenological study of its present configuration or by analytical study of its origins. In recent years considerable progress has been made on both these fronts, but it has become clear that a major difficulty lies in bridging the gap between them. Large-scale surveys of galaxy redshifts have revealed many aspects of the three-dimensional morphology of the galaxy distribution. (See the recent reviews Oort 1983; Davis 1984.) At the same time there has been widespread enthusiasm in the cosmological community over new ideas injected both by Grand Unified Theories and by the possibility of an early inflationary epoch. The present ratio of photons to baryons may be explained by non-equilibrium particle interactions at grand unification energies (Weinberg 1979), while inflation can account for the homogeneity and flatness of the universe and for the low abundance of magnetic monopoles (Guth 1981). Thus some of the most fundamental parameters of our universe may be explained in a natural fashion. Inflationary models make a specific prediction for the form of the density perturbation spectrum from which observed structure must grow (Guth and Pi 1982; Hawking 1982; Starobinskii 1982; Bardeen, Steinhardt, and Turner 1983). Perturbations in curvature arise from quantum fluctuations during inflation. The time invariance of de Sitter space ensures that their amplitude as they are swept across the event horizon during the inflationary epoch, and consequently as they return across it in the present Friedmann universe, is independent of scale. Their distribution thus has the Harrison-Zel'dovich "constant curvature" form. Perturbations generated in this way are adiabatic, with no fluctuation in photon-to-baryon ratio. So far inflationary models have been able to predict the correct fluctuation amplitude only for very specific and finely tuned theories for the underlying particle physics. The prediction of an adiabatic scale-invariant spectrum is not, however, restricted to such models (see, e.g., Harrison 1970; Zel'dovich 1972; Kibble 1976; Turner and Schramm 1978; Press 1980).

If any inflationary model is valid, the present value of the cosmological density parameter Ω is expected to be very close to unity. Constraints from Big Bang nucleosynthesis suggest, however, that the density of baryonic material is an order of

[1] Institute for Theoretical Physics, University of California, Santa Barbara.
[2] Astronomy and Physics Departments, University of California, Berkeley.
[3] Institute of Astronomy, University of Cambridge.
[4] Astronomy Centre, University of Sussex.
[5] Steward Observatory, University of Arizona.

magnitude lower (Yang *et al.* 1984). Furthermore, in a purely baryonic universe an adiabatic constant-curvature spectrum leads to small-scale anisotropies in the microwave background in excess of the reported upper limits (Wilson and Silk 1981; Uson and Wilkinson 1984). These apparent inconsistencies are avoided if the dark matter that appears to dominate the dynamics of galaxy clusters and of the universe as a whole is made up of weakly interacting, nonbaryonic elementary particles. Candidate particles can be grouped into three categories on the basis of their effect on the fluctuation spectrum (Bond *et al.* 1983). If the dark matter is composed of abundant light particles which remain relativistic until shortly before recombination, then it may be termed "hot." The best candidate for hot dark matter is a neutrino with a mass in the range $10 \text{ eV} < m_\nu < 100$ eV. The phase mixing of freely streaming relativistic particles erases density fluctuations of scale smaller than the horizon, and for massive neutrinos detailed calculations show that fluctuations are damped out if their present wavelength is less than the critical damping scale:

$$\lambda_c = 41(m_\nu/30 \text{ eV})^{-1} \text{ Mpc} \quad (1)$$

(Peebles 1982; Bond and Szalay 1983). A second possibility is for the dark matter particles to interact more weakly than neutrinos, to be less abundant, and to have a mass of order 1 keV. Such particles are termed "warm" dark matter, because they have lower thermal velocities than massive neutrinos and are able to erase fluctuations only on wavelengths smaller than about 1 Mpc, corresponding roughly to the mass of a bright galaxy. There are, at present, few candidate particles which fit this description. Gravitinos and photinos have been suggested (Pagels and Primack 1982; Bond, Szalay, and Turner 1982), but currently it seems more likely that supersymmetric particles have a mass in excess of 1 GeV. Any particles which became nonrelativistic very early, and so were able to diffuse a negligible distance, are termed "cold" dark matter (CDM). There are many candidates for CDM, including supersymmetric particles (with masses above 1 Gev) and axions with a mass of 10^{-5} eV (Preskill, Wise, and Wilczek 1983).

A priori the most plausible elementary-particle candidate for the dark matter would seem to be the neutrino. Extensive studies of a neutrino-dominated universe have been carried out using analytic and N-body techniques, while related studies have focused on the hydrodynamics of pancake formation and cooling (Klypin and Shandarin 1983; Frenk, White, and Davis 1983; Centrella and Melott 1983; White, Frenk, and Davis 1983; Shapiro, Struck-Marcell, and Melott 1983; Bond *et al.* 1983). The transient filamentary structure that forms in these models is reminiscent of the long supercluster chains apparent in the galaxy distribution. However, it is not possible to reconcile the relatively small length-scale of galaxy clustering with the large length-scale of equation (1), unless galaxies formed at unacceptably recent epochs and in a manner very different from that predicted in the standard pancake picture. For these and other reasons, it is very unlikely that stable massive neutrinos dominate the universe (White, Frenk, and Davis 1983; White, Davis, and Frenk 1984).

Detailed calculations of the linear evolution of the CDM density-perturbation spectrum have been given by several authors (Peebles 1982, 1984; Blumenthal and Primack 1983; Bond and Efstathiou 1984). During the radiation-dominated epoch, fluctuations grow significantly only while their wavelength exceeds the horizon scale. However, once the universe is matter-dominated, all linear modes grow at the same rate. As a result the fluctuation spectrum develops a bend at a scale corresponding to the horizon size at the transition between the two epochs. For a constant-curvature initial spectrum, the power density at late times at comoving wavenumber k tilts gradually from the Harrison-Zel'dovich form, $|\delta_k|^2 \propto k$, to $|\delta_k|^2 \propto k^{-3}$ over more than two decades in scale. It is well approximated by the following expression:

$$|\delta_k|^2 = Ak/(1 + \alpha k + \beta k^{3/2} + \gamma k^2)^2 , \quad (2)$$

where A is a normalization constant. If k is expressed in terms of present length units, the parameters in equation (2) take the values $\alpha = 1.7l$, $\beta = 9.0l^{3/2}$, $\gamma = 1.0l^2$, where $l = (\Omega h^2 \theta^{-2})^{-1}$ Mpc, $h = H_0/100$ km s^{-1} Mpc^{-1}, Ω is the present cosmological density parameter, and θ is the microwave background temperature in units of 2.7 K. This fit is based on the numerical calculations of Blumenthal and Primack (1983) and of Bond and Efstathiou (1984), which have $\Omega = 1$ and $\theta = 1$ and include three massless neutrino flavors. If the baryon density is small compared to that of the CDM, Silk damping has a negligible effect on the CDM fluctuation spectrum and the scaling with l becomes exact; otherwise it is approximate. The scaling of l with $\Omega h^2 \theta^{-2}$ simply reflects that of the horizon diameter when matter and radiation had equal energy density; it is independent of the nature of the CDM. As a result, it may be possible to constrain this combination of cosmological parameters by requiring the nonlinear structure that forms at late times to agree with observation. Alternatively, it may be possible to reject the CDM model, if we find no match to the observations for acceptable values of the cosmological parameters. Since the overall amplitude of fluctuations cannot, at present, be calculated *a priori*, we are free to adjust it to give the correct level of clustering today. This is the only free parameter of the model.

In this paper we study the nonlinear evolution of clustering from the linear initial conditions characterized by equation (2). Analytic treatments of this problem can be carried through only for simple, highly symmetric models of clusters or voids. Peebles (1984) and Blumenthal *et al.* (1984) have investigated the cosmogony of CDM universes by applying scaling arguments based on such models to the predictions of linear theory. While these studies give considerable insight into galaxy formation, clustering patterns in the real universe show no special symmetry, and direct simulations are required to study their morphology and their statistical properties. We have studied the full nonlinear problem by running ensembles of N-body simulations of Einstein–de Sitter, and of open Friedmann models, together with one simulation with zero curvature but with a nonzero cosmological constant, $\Lambda > 0$. Our models have an initial resolution limit corresponding to a present scale in excess of 2 Mpc and so are insensitive to details of the initial spectrum below this wavelength. As a result, the nonlinear clustering that they develop differs very little from that expected in a universe dominated by warm dark matter.

Results from a single simulation of an "axion/gravitino/photino–dominated universe" have been published previously by Melott *et al.* (1983). Unfortunately, these authors chose an initial power spectrum (power-law segments with $n = 1$ and $n = -3$ separated by a sharp break) which is an extremely poor approximation to that expected in a universe dominated by cold dark matter (eq. [2]). In addition they used a particle-mesh code on a 32^3 grid, and so had too little resolution to be able to follow the hierarchical growth of structure from small scales (see Efstathiou *et al.* 1985, hereafter EDFW). As a result, the relevance of their work to CDM universes is unclear.

In § II we describe our numerical methods and the way in which we generate initial conditions, and we catalog the simulations we have performed. The evolution of the fundamental statistical properties of the models (power spectra, two- and three-point correlation functions, velocity correlations, multiplicity functions, cluster properties, and cell-count distributions) is described in § III, where we also discuss their compatibility with observation. In § IV we present graphical comparisons of our open models with the observed galaxy distribution in a large redshift survey. We conclude from the data in these two sections that an $\Omega = 1$ model is quite unacceptable if galaxies trace the mass distribution and that models with $\Omega \approx 0.2$, while better, still do not provide a fully acceptable match to observation. Finally, in § V we investigate the consequences of relaxing the assumption that the galaxy distribution follows the mass. Instead we identify "galaxies" with high peaks in the initial density distribution; such points form a biased subset of the overall mass distribution and have enhanced correlation properties. This procedure models a situation in which galaxy formation was suppressed except in sufficiently dense regions. It leads to models which can agree with observation quite well even for $\Omega = 1$. Conclusions and general remarks are given in § VI.

II. THE SIMULATIONS

All the simulations discussed here were done on VAX computers using a particle-particle/particle-mesh code (P^3M; Efstathiou and Eastwood 1981; Hockney and Eastwood 1981). This code was designed to have high force resolution and uses a comoving formulation of the equations of motion to follow evolution within the fundamental cube of a triply periodic universe; a detailed examination of its performance and a comparison with simpler particle-mesh codes are given by EDFW. In our models the long-range force field was calculated by Fourier methods on a grid of 64^3 cells. Inclusion of a short-range force correction resulted in point-mass interactions between particles separated by more than a softening length, $\eta = L/213$, where L is the comoving length of the side of our computational volume. At separations smaller than η our force law corresponded to the interaction of two interpenetrating spheres with radii $\eta/2$ and linear density profiles. We applied the short-range correction only at separations below 9η; the discontinuity in the total force at this separation was less than 2%, and the force field was isotropic on all scales to much better accuracy than this. All our simulations integrated the motion of 32,768 particles using a leapfrog scheme with a, the expansion parameter, as the time variable and a timestep of $\Delta a = 0.02$. Their accuracy can be judged using the Layzer-Irvine cosmic energy equation; typically the integration constant in the integral form of this equation had changed by less than 0.1% of the current potential energy at any stage of a run (see EDFW). The longest simulations were integrated until $a = 8.4$ and took about 70 hours of CPU time on a VAX 11/750.

We generated initial conditions using the Zel'dovich (1970) approximation, following the algorithm described by EDFW. Starting with the particles on a cubic lattice, we assigned a displacement and peculiar velocity to each in proportion to the gradient of the potential derived from a random realization of the linear fluctuation distribution. The latter corresponded to a superposition of 64^3 waves with random phases and with amplitudes distributed normally with variance given by equation (2). In all cases velocities were set so that only the growing mode was present. Since we used 32^3 particles, waves with a projected wavelength along any coordinate axis less than $L/16$ were undersampled; initial perturbations cannot be generated at such high frequencies. We therefore chose the amplitude of the initial perturbations so that the theoretically predicted power would be equal to the white-noise level at a wavelength of $L/16$. Tests show that this choice leads to the best possible random-phase representation of the theoretical spectrum and that the subsequent evolution of all structures made up of more than a few particles is unaffected by residual artificialities of the method (see EDFW § V, and Figs. 2 and 5 below). For our Einstein–de Sitter models we used equation (2) with the parameters as given; for our open models we used parameters obtained by fitting this equation to Blumenthal and Primack's (1983) results for $\Omega = 0.2$. The resulting spectra agree quite well with the expected scaling; small residual differences are reflected in a reduction of power in the open models by about 10% at long wavelengths (see Fig. 2*b*).

Because the linear CDM fluctuation spectrum has considerable power on large scales, statistically equivalent initial conditions can lead to substantially different levels of clustering even after expansion by quite a modest factor. Thus several simulations from different realizations of the same initial spectrum must be compared to assess the statistical uncertainties in the results. Furthermore, although the spectrum behaves asymptotically as $|\delta_k|^2 \propto k^n$, with $n = 1$ on large scale and $n = -3$ on small scale, the transition between these two regimes is quite slow; the length scales where $n = 0.5$ and $n = -2.5$ differ by a factor of 100 (cf. eq. [2]). Since we can reproduce this spectrum only over a factor of 16 in scale, simulations in different-size "boxes" are useful to extend the dynamic range of our results and to study the effect of their limited resolution.

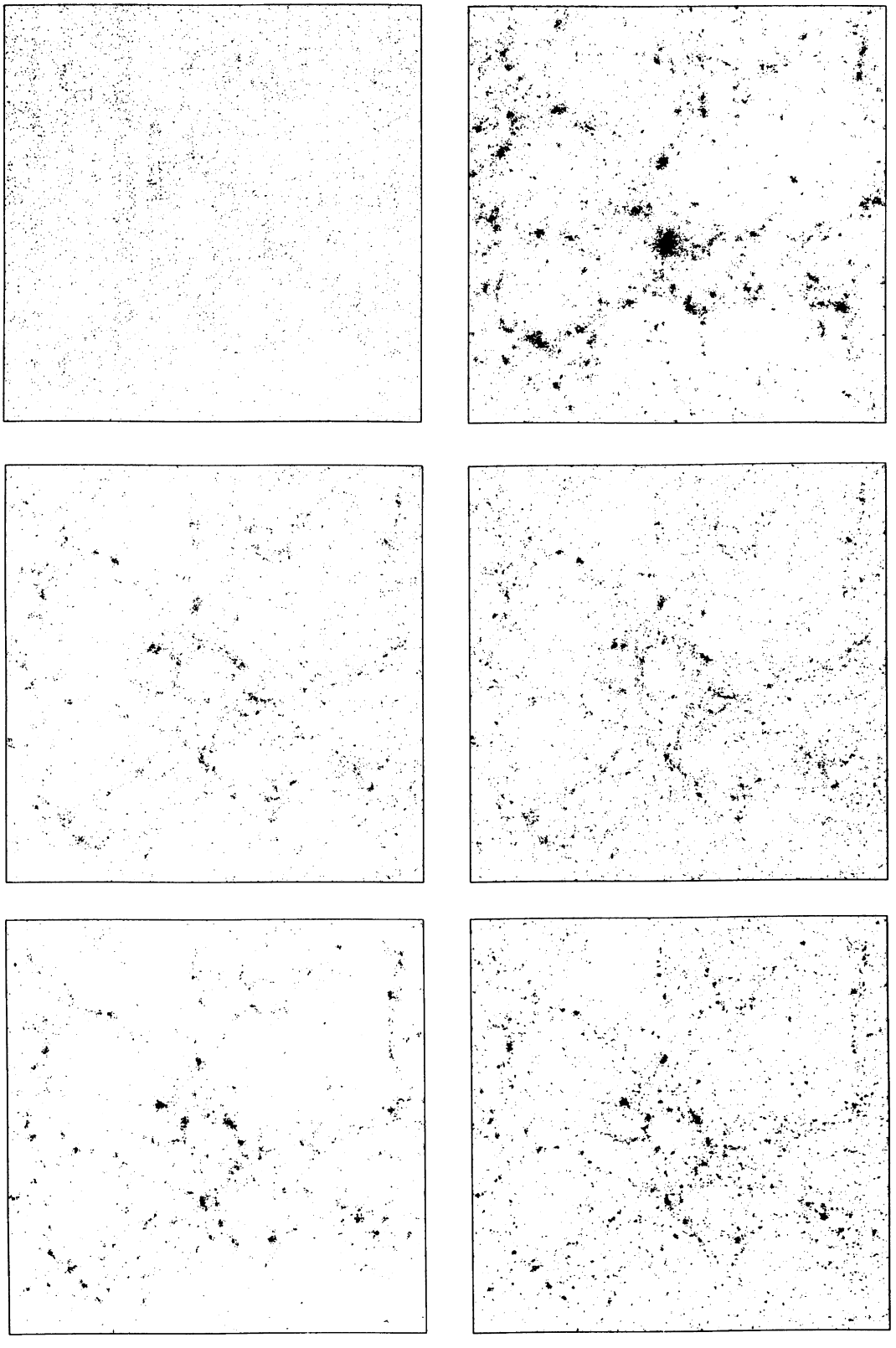
We ran three ensembles of simulations with the physical parameters given in Table 1. While h and θ affect the simulations only through their relation to the length scale L, the value of Ω influences the dynamical evolution and so must be specified separately. Henceforth we assume $\theta = 1$. Models EdS1–5 represent an $\Omega = 1$ universe with $L = 32.5h^{-2}$ Mpc in present units. Models O1–4 were evolved from similar initial conditions, but the mass of each particle was reduced so that $\Omega = 0.2$ after an expansion of 3.2. If this time is identified with the present, then $L = 162.5h^{-2}$ Mpc. Output times were chosen so that the linear growth factors in the two ensembles would match. Models EdS6–8 also had $\Omega = 1$, but the present size of the box was taken to be $L = 65h^{-2}$ Mpc. Again the output times were chosen so that one could compare these results to EdS1–5 at the same linear fluctuation amplitude. Finally, we ran a single model (L1) of a universe with zero curvature and a

TABLE 1

MODEL PARAMETERS

Model	L (l)	η (l)	M_T (m)	m_p ($10^{-4}m$)	Ω_i
EdS1–EdS5	32.5	0.15	9.4	2.9	1
EdS6–EdS8	65.0	0.30	75	23	1
L1	32.5	0.15	9.4	2.9	0.796
O1–O4	32.5	0.15	9.4	2.9	0.446

NOTE.—L is the side of the computational box, M_T the total mass it contains, η the softening length of the force law, and m_p the mass of an individual particle. The length and mass units are defined by $l = (\Omega h^2\theta^{-2})^{-1}$ Mpc and $m = 10^{15}(\Omega h^2)^{-2}\theta^6\ M_\odot$, where Ω, h, and θ denote the *present* value of the cosmological parameters. Ω_i is the value of Ω at the start of the simulations.



positive cosmological constant. Its initial condition matched those of EdS1 and O1, and the initial values of Ω and Λ were chosen so that when $\Omega = 0.2$ the linear fluctuation growth factors in O1 and L1 would agree. Because of our adopted normalization there are substantial initial fluctuations on small scales, so that many small objects collapse immediately after the start of a simulation. The steep initial spectrum then results in the rapid formation of larger structures as evolution proceeds. For models EdS1–5 the clustering scale approaches the box size for $a \geq 4$ and the experiments must be terminated.

III. THE EVOLUTION OF CLUSTERING

a) Morphology of Clustering

We begin by examining pictures of the models to get a qualitative feel for the results. Figure 1 shows four projections of model EdS1 and two of O1; the times plotted for O1 were chosen so that its linear-fluctuation amplitude corresponded to that of EdS1 at the two intermediate times. The initial conditions of the two models are almost indistinguishable, so only those of EdS1 are shown. Note that they have substantial large-scale fluctuations, and that careful examination reveals a trace of the underlying grid from which the particles were perturbed. As the simulation evolves, structure appears very rapidly, and filaments, superclusters of clumps, and large low-density regions develop. The range of structure present is much richer than in similar models evolved from white-noise initial conditions (Efstathiou and Eastwood 1981; Efstathiou and Barnes 1983). Because of the strong subclustering which is present, the filaments and voids in these models are less conspicuous than in simulations of neutrino-dominated universes. At times $a = 1.8$ in EdS1 and $a = 3.2$ in O1, the two-point correlation function is close to a power law, $\xi \propto r^{-\gamma}$, with $\gamma = 1.8$; under an appropriate scaling, these pictures might therefore be considered a reasonable statistical match to the observed galaxy distribution (however, see our more detailed analysis below). At later times much of the substructure is erased as clumps merge into larger and larger systems. By the end of the integration of EdS1, at $a = 4.5$, the distribution is dominated by a few large clumps with little apparent substructure. Note how closely the structure matches in the two models with different values of Ω. Nevertheless, it is clear that clusters are more concentrated in the open model. The simulation with a positive cosmological constant matches O1 very closely at the time when $\Omega = 0.2$ and the linear growth factor is 1.8 in both models.

b) Power Spectra and Two-Point Correlation Functions

In this section we consider the power spectrum $|\delta_k|^2$ of our simulations and its Fourier transform, the two-point correlation function $\xi(r)$. To obtain power spectra we derived a density field from the point distribution using a nearest-grid-point mass assignment on a 64^3 grid; we then Fourier transformed the result and binned the power density to the nearest integral wavenumber (see EDFW). The evolution of the power spectrum for EdS1–5 is shown in Figure 2*a*. At the initial time the spectrum flattens to the white noise level for $k > 16$, the Nyquist frequency corresponding to the initial particle grid, but is otherwise very close to the input theoretical power spectrum; remaining differences at small k are due to sampling noise. (Here and below we quote values of k in units such that $k = 1$ corresponds to the fundamental wave of our box.) The effective spectral index for this initial condition ranges from $n \approx -1$ for $1 < k < 6$ to $n \approx -2$ for $6 < k < 12$. At later times and at high wave number ($k > 16$), the power in excess of white noise is reduced somewhat as a result of the smoothing introduced by the mass assignment; it is also affected slightly by aliasing from wavenumbers with $k > 32$. In linear theory, all Fourier modes grow at the same rate, and $|\delta_k|^2 \propto a^2$ for $\Omega = 1$. We can check that the models remain linear on the largest scales by measuring the growth rate of the $k = 1$ mode and checking that it follows this scaling. In Figure 2*a* the growth expected is 12.2, slightly greater than the factor 10.5 found from the data; the discrepancy appears only at the end of the simulations. Once the fundamental mode becomes nonlinear, its growth rate is expected to decrease in this way (cf. EDFW). In the present models the fundamental mode saturates and grows very slowly for $a > 4$. At this stage, the periodic boundary conditions are becoming unrealistic, and the integration must be stopped. This reflects the fact that there is only a relatively small time window between the collapse of the smallest and largest structures in our models.

Figure 2*b* shows the evolution of the power spectrum of our ensemble of open models. The corresponding results from Figure 2*a* are superposed in order to facilitate comparison and have been multiplied by 0.9 to account for the slight difference between the theoretical input spectra (see § II). The growth of large-scale perturbations ($k < 8$) matches extremely well in the two sets of models, even though nonlinearities appear to have affected the evolution significantly for $k \geq 4$. At higher frequencies ($k > 8$) and at later times, there is an excess of power in the open models. This excess reflects the difference in cluster concentration visible in Figure 1 and is a consequence of the larger expansion of the open models after cluster collapse. When times with the same linear growth factor are compared, the power spectrum of the model with a positive cosmological constant is very similar to that of the corresponding open model but with slightly less power at high frequency.

An important test of the simulations comes from comparison of the power spectra for calculations in boxes of different size. Figure 3 shows the evolution of the power spectrum of the "large box" ensemble superposed on the appropriately shifted version of the data from Figure 2*a*. There are no free-fitting parameters in this comparison and, in the absence of spurious effects due to the finite particle number and the finite box size, the two power spectra should coincide. At large k there are slight differences due to the smoothing inherent in our method for calculating the power spectrum, but in the range $2 < k < 8$ the agreement is very good. This shows that the evolution of the Fourier components accessible to the "small box" calculation is not affected by its finite size. In principle, we could thus use linear theory to specify the evolution of longer wavelength components and so embed our models in an "infinite" universe. We have not bothered to do this because the neglected large-scale power is quite small at the times that interest us. The evolution of the power spectra in Figures 2 and

FIG. 1.—Two-dimensional projections of the particle positions in an Einstein–de Sitter and an open CDM simulation. The initial conditions for EdS1 are shown at top left. Subsequent plots show the model after expansion by factors, $a = 1.8$ (*center left*), $a = 2.4$ (*bottom left*), and $a = 4.5$ (*top right*). The remaining two plots show O1 at $a = 3.2$ when the linear growth factor from its initial conditions is 1.8 and $\Omega = 0.2$ (*center right*), and at $a = 8.4$ when the growth factor is 2.4 and $\Omega = 0.09$. Note that the side of the box corresponds to $32.5(\Omega h^2)^{-1}$ Mpc in these models, where Ω and h are the *present* values of the cosmological parameters.

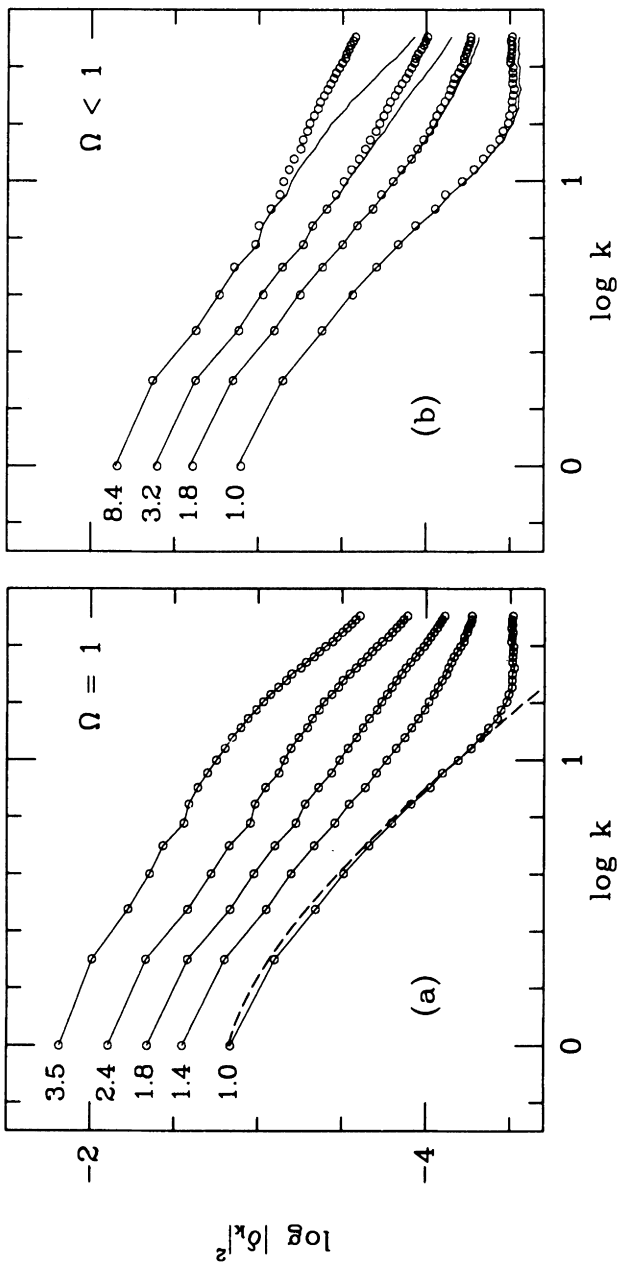


FIG. 2.—The evolution of the power spectrum for two ensembles of models, (*a*) EdS1–5 and (*b*) O1–4. The value of the expansion parameter *a* corresponding to each spectrum is given. The dashed line in (*a*) shows the theoretical spectrum (eq. [2]) used to generate the initial conditions. The solid lines in (*b*) repeat the results of (*a*) after multiplication by 0.9.

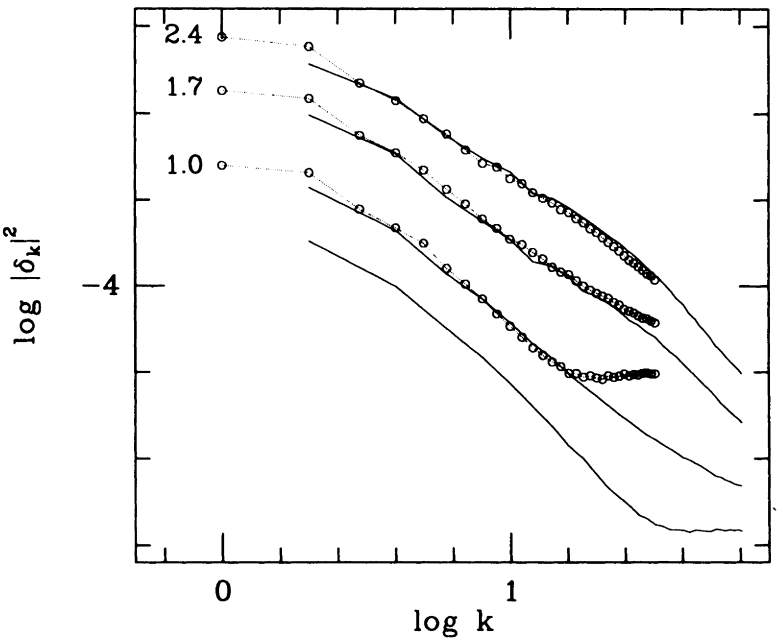


FIG. 3.—The evolution of the power spectrum of the "big box" models, EdS6–8, is compared with appropriately shifted results for our "standard" models, EdS1–5 (shown as solid lines). The agreement at small wavenumber shows that the evolution of large-scale structure in the standard models has not been adversely affected by the absence of wavelengths larger than the box size.

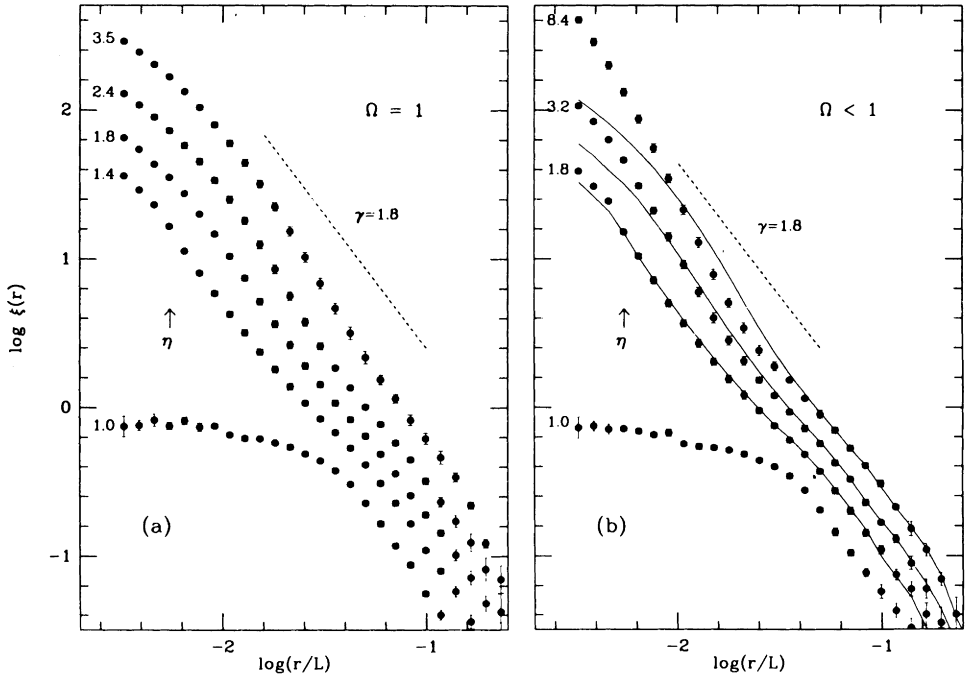


FIG. 4.—Two-point correlation functions $\xi(r)$ are shown for models (*a*) EdS1–5 and (*b*) O1–4, for various values of the expansion parameter *a*. The separation *r* is given in units of the side of the computational volume. The error bars give errors in the mean derived from the scatter within each ensemble. The dotted line is a power law with the index $\gamma = 1.8$, which fits the galaxy distribution. The solid lines in (*b*) repeat the results in (*a*) after multiplication by 0.9. An arrow in each panel marks the softening length of our simulations.

3 is such that as short-wavelength modes go nonlinear, they grow more rapidly than the linear mode, resulting in a progressive flattening of the power spectra. The power spectra do not evolve in a self-similar fashion.

A corresponding lack of self-similarity is shown by the evolution of the two-point correlation function $\xi(r)$, which we plot in Figure 4*a* for models EdS1–5. In order to achieve the highest possible resolution, correlation functions were computed by pair counts rather than by a Fourier transform of the power spectrum; all pairs were used for $\log(r/L) < -1$, and about 10% of pairs were used for larger separations. The effects of force softening are apparent as a slight flattening of $\xi(r)$ on the smallest scales shown [recall that $\log(\eta/L) = -2.3$], but generally the correlations have near power-law form on large scales. Note that $\xi(r)$ steepens with time and has a power-law index $\gamma = 1.8$ at $a = 1.8$. By the end of the run $\gamma = 2.1$, which is much steeper than the observed galaxy function. Power law fits to ξ over the range $30 > \xi > 0.3$ are given in Table 2 for all three of our ensembles. In Figure 4*a* the growth of $\xi(r)$ on large scale matches the linear expectation, $\xi \propto a^2$, while on small scale [$\log(r/L) = -2.5$], ξ is growing as $a^{2.2}$. This reflects the accelerated growth seen also in the power spectrum. As we discuss in the next section, the nonlinear clustering in these models does not match the predictions of earlier similarity arguments (Peebles 1974, 1980; Davis and Peebles 1977).

Two-point correlation functions for the open models are shown in Figure 4*b* and are compared with those of EdS1–5 at

TABLE 2

POWER-LAW FITS TO $\xi(r)$ OVER THE RANGE $0.3 \leq \xi \leq 30$

A. EdS1–EdS5

a	Δ	γ	$\log(r_0/L)$
1.4.....	1.4	1.70 ± 0.01	-1.55 ± 0.01
1.8.....	1.8	1.84 ± 0.01	-1.41 ± 0.01
2.4.....	2.4	1.96 ± 0.02	-1.29 ± 0.02
3.0.....	3.0	2.09 ± 0.02	-1.19 ± 0.02
4.0.....	4.0	2.14 ± 0.04	-1.06 ± 0.03

B. EdS6–EdS8

a	Δ	γ	$\log(r_0/L)$
1.3.....	1.8	1.65 ± 0.01	-1.76 ± 0.01
1.7.....	2.4	2.02 ± 0.01	-1.60 ± 0.02
2.1.....	3.0	2.14 ± 0.02	-1.50 ± 0.02
2.4.....	3.5	2.17 ± 0.03	-1.43 ± 0.03
2.8.....	4.0	2.21 ± 0.03	-1.37 ± 0.03

C. "Galaxies"

a	Δ	γ	$\log(r_0/L)$
1.4.....	1.4	1.96 ± 0.04	-1.17 ± 0.04
1.8.....	1.8	1.99 ± 0.05	-1.13 ± 0.04
2.4.....	2.4	1.92 ± 0.05	-1.07 ± 0.04
3.0.....	3.0	2.09 ± 0.07	-1.01 ± 0.05

D. O1–O4

a	Δ	γ	$\log(r_0/L)$	Ω
1.4.....	1.2	1.52 ± 0.01	-1.72 ± 0.02	0.37
1.8.....	1.4	1.73 ± 0.01	-1.59 ± 0.01	0.31
3.2.....	1.8	1.82 ± 0.02	-1.45 ± 0.02	0.20
5.1.....	2.1	1.75 ± 0.02	-1.39 ± 0.03	0.14
8.4.....	2.4	1.75 ± 0.02	-1.33 ± 0.02	0.09

NOTE.—Δ is the amplitude of mass fluctuations in the linear regime relative to that in the initial conditions of EdS1–5. The quoted errors come from least-squares fits to $\log[\xi(r)]$.

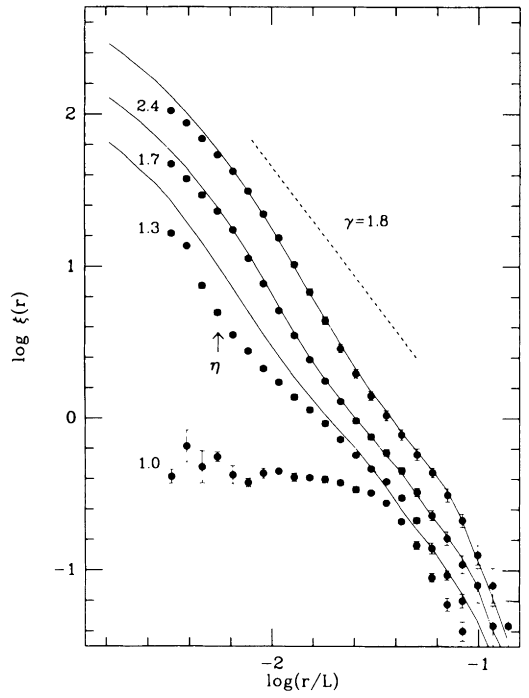


FIG. 5.—A test of boundary and discreteness effects. The points show ξ averaged over the ensemble EdS6–8 at various values of the expansion variable. The box length in these models is $L = 65h^{-2}$ Mpc. The correlation functions for the ensemble EdS1–5 at corresponding linear fluctuation amplitudes are shown as solid lines. These have been shifted by a factor of 2 to the left to enable a direct comparison with the points. Note the excellent agreement at late times and the size of the error bars (these are often smaller than the symbols plotted).

times chosen to have the same linear growth from the initial conditions. The correlations in the EdS models have been multiplied by 0.9 to account in the mean for the slight difference in the initial power spectra discussed in § II. As we found in earlier work (Frenk, White, and Davis 1983), the correlation functions agree well when times are matched in this way. For the open and positive Λ models, ξ is somewhat larger on small scales, and smaller on intermediate scales, than in the flat models, consistent with the excess power seen at high frequencies in Figure 2*b*. As Ω becomes small this disagreement becomes worse; $\xi(r)$ steepens on small scales but stops growing at large scale. At late times its shape is not well fit by a single power law, so that the exponents given in Table 2D must be treated with some caution.

Figure 5 compares the two-point functions of our two ensembles with $\Omega = 1$ and boxes of differing size. The lines representing our "standard" models (EdS1–5) have been shifted to the left by a factor of 2 to account for the difference in box size, and the times plotted were chosen so that the linear fluctuation amplitudes on large scale should match. Once again we have no freedom to adjust the fit in this diagram; the curves should coincide, as indeed they do once initial transients have had time to subside. Slight differences on very small scale show the effect of the larger softening in the "big box" models. This test confirms the similar test carried out above for the power spectra and demonstrates that the rapid steepening of

$\xi(r)$ in the smaller simulations is an intrinsic property of the nonlinear clustering process rather than a consequence of their finite size. (This behavior is expected because of the non–self-similarity of the initial conditions.) This test also shows that the statistical properties of the small-scale structure are very similar in models where the particle masses differ by a factor of 8 and the softening lengths by a factor of 2. Thus, discreteness and two-body relaxation effects do not seem to be important on the scales which concern us (cf. Fall 1979). Figure 5 suggests that we can use periodic boxes of different size to extend our dynamic range; however, this technique is of limited use in practice, because we can never evolve a simulation beyond the time when the $k = 1$ modes begin to go nonlinear.

Under the hypothesis that the galaxies are an unbiased tracer of the mass distribution in the universe, we must require $\xi(r)$ in our models to have the form of the observed galaxy function. In addition, we must scale the models so that its amplitude is correct. The correlation length, r_0, defined by $\xi(r_0) = 1$, is measured to be $5 \pm 0.7h^{-1}$ Mpc for the galaxy distribution (Davis and Peebles 1983). At the time when the slope of the correlation function in our models has the observed power-law index, $\gamma = 1.8$, we find $r_0 = 1.27h^{-2}$ Mpc for the $\Omega = 1$ ensemble (at $a = 1.8$) and $r_0 = 5.8h^{-2}$ Mpc for the open ensemble (at $a = 3.2$, when $\Omega = 0.2$). Thus, for consistency with the observations, $h = 0.25$ is required if $\Omega = 1$ and $h = 1.1$ if $\Omega = 0.2$; both these values are outside the range covered by most modern determinations. For acceptable values of h, our models do not simultaneously reproduce the correct shape and the correct amplitude for $\xi(r)$. For example, when $a = 4$ in the EdS1–5 ensemble, we find $h = 0.57$ and $\gamma = 2.14$ (see Table 2A). The uncertainty as to which model time should be identified with the present is probably sufficient to allow large, but acceptable, values of h for the open models. Matching the correlation length of the positive Λ model to observation when $\Omega = 0.2$ also leads to $h = 1.1$, but in this case the age of the universe is 2.3 times the usual Einstein–de Sitter value, $\frac{2}{3}H_0^{-1}$.

A more direct test of whether the simulations match the observed shape of ξ is illustrated in Figure 6. Here we show the projected correlation function,

$$w(r_p) = \int_{-\infty}^{\infty} dy \xi[(r_p^2 + y^2)^{1/2}] , \quad (3)$$

for the $\Omega = 1$ models at $a = 1.8$ and for the open models at $a = 3.2$. In evaluating the integral in equation (3), we used the mean values of $\log \xi$ shown in Figures 4*a* and 4*b*. Note that the results are significantly affected by our force softening to the left of the vertical arrow. The filled circles in Figure 6 show observational estimates of $w(r_p)$ from the CfA redshift survey (Davis and Peebles 1983, Fig. 2). Our $\Omega = 1$ results fit the data very well, provided $h = 0.22$. The open arrows in the figure show how the model function shifts if it is scaled using $h = 0.3$. Clearly the value of h required for agreement at any given model time is determined to about 10%. The major uncertainty lies in deciding which model times could be consistent with the observational data. At $a = 3.5$ the amplitude of $w(r_p)$ matches the data if $h = 0.47$, but its dependence on r_p is much too steep to be acceptable. The amplitude of $w(r_p)$ shown for the $\Omega = 0.2$ ensemble in Figure 6 corresponds to the choice $h = 1.1$. In this case its shape deviates significantly from a power law, but the model still represents the CfA data fairly well. A value of Ω in the range $0.2 < \Omega < 1$ would clearly produce rough agreement with both the shape and the amplitude of the observed correlation function for any of the currently popular values of h. However, as we now discuss, such models would predict peculiar velocities for galaxies which are considerably larger than those observed.

c) Velocity Moments

The relative peculiar velocity distribution of pairs provides another important diagnostic for the simulations. Consider the moments of this distribution as a function of pair separation. The first moment points along the separation vector of the pairs and must cancel the Hubble flow if the pairs are to

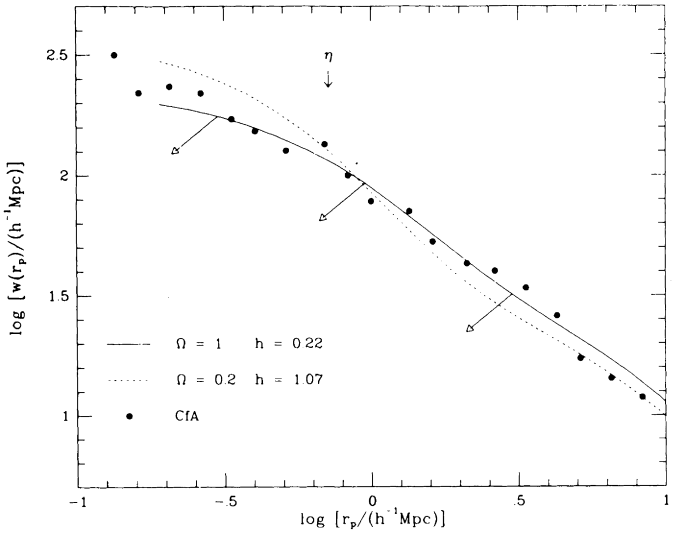


FIG. 6.—The projected correlation function $w(r_p)$ (eq. [3]). The filled circles show the results from the CfA redshift survey (Davis and Peebles 1983, Fig. 2), while the curves show results from our open and Einstein–de Sitter ensembles at $a = 3.2$ and $a = 1.8$ respectively. The vertical arrow marks the softening length of our simulations, and the diagonal arrows show how the results for a flat universe would change if they were scaled using $h = 0.3$ rather than $h = 0.22$.

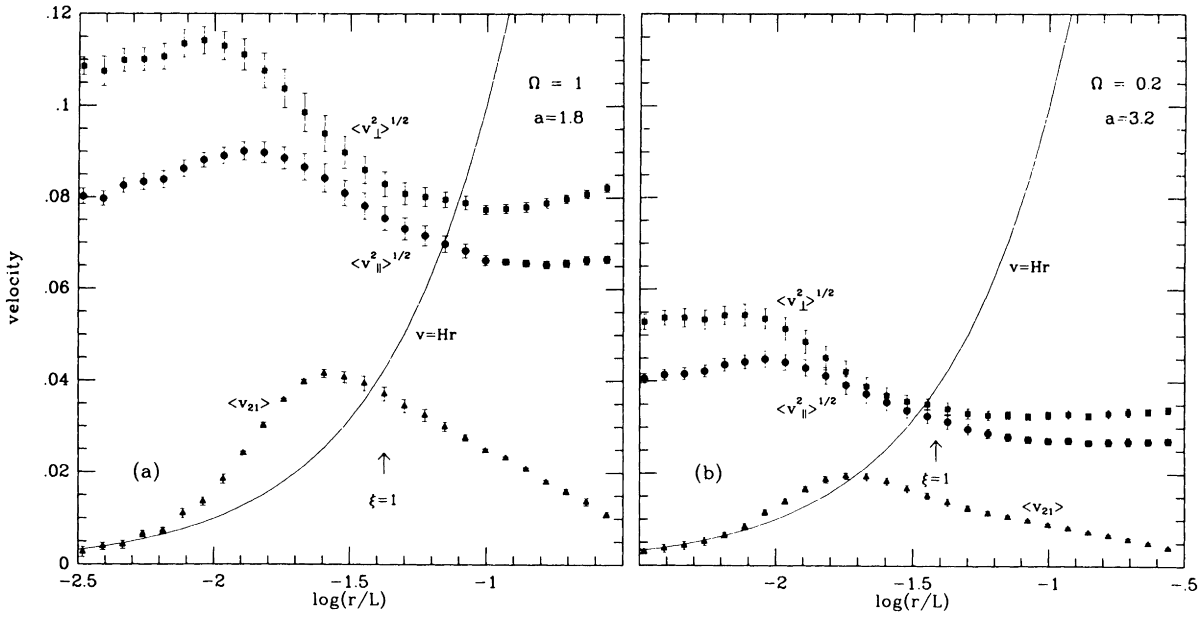


FIG. 7.—Velocity moments are shown in (*a*) for the $\Omega = 1$ ensemble EdS1–5 at $a = 1.8$, and in (*b*) for the open ensemble O1–4 at $a = 3.2$. The velocities have been scaled to units in which the Hubble velocity across the computational box is unity. A velocity of one thus corresponds to $3250(\Omega h)^{-1}$ km s^{-1}. The solid line shows the Hubble flow $v = Hr$. $\langle v_{\parallel}{}^2\rangle^{1/2}$ and $\langle v_{\perp}{}^2\rangle^{1/2}$ are the rms peculiar relative velocities of pairs in the radial and tangential directions respectively. A vertical arrow in each panel marks the point where $\xi = 1$.

remain, on average, at fixed physical separation. The second moments form a symmetric second-rank tensor; its two independent eigenvalues correspond to the radial and transverse velocity dispersions. In Figures 7*a* and 7*b* we plot these moments as a function of separation for models EdS1–5 (at $a = 1.8$) and for models O1–4 (at $a = 3.2$) respectively; the solid line is the Hubble line $v = Hr$, and arrows mark the correlation length r_0. As discussed above, $\xi(r)$ for both ensembles is approximately described by a power law of index $\gamma = 1.8$ at these times. Plots similar to Figure 7 have been given by Davis and Peebles (1977) for BBGKY models, and by Efstathiou and Eastwood (1981) for N-body simulations with white noise and subrandom initial conditions.

Our results are very similar to those of Efstathiou and Eastwood (1981) but differ substantially from the predictions of the BBGKY similarity solutions. The first moment, $\langle v_{21}(r)\rangle$, rises above the Hubble line for $\xi(r) > 1$ in both ensembles, and lies well above it for $\Omega = 1$, indicating that, on average, pairs on these scales are physically moving together. On the smallest scales $\langle v_{21}(r)\rangle = -Hr$, as expected for stable bound structures. The peculiar velocity dispersion tensor is slightly radially biased on all scales. (In accord with previous practice the radial dispersion plotted is not a central moment, and so should have the mean motion subtracted in quadrature when computing the anisotropy.) In our simulations, as in those of Efstathiou and Eastwood, the dispersions fall substantially with increasing separation, whereas the BBGKY solutions rise slowly. This discrepancy is probably the consequence of some of the simplifying assumptions which Davis and Peebles adopted in order to make the BBGKY equations tractable. The rms velocity difference drops slightly more rapidly with separation in the open models than in the closed models. This is yet another reflection of the fact that clusters are relatively more compact in the open models. The decrease of the dispersions at the smallest scales [$\log(r/L) < -2.3$] can be at least partly ascribed to the softening of the particles.

In order to test the effects of softening on the velocity distribution, we ran one simulation with $\Omega = 1$ and a severely softened force law. The potential was taken to vary logarithmically for separations less than $L/32$ and was further softened to approach a constant at separations smaller than $L/430$. It thus corresponded approximately to the interaction of truncated isothermal spheres. This model developed a correlation function with a pronounced bend at the outer softening length, and it had considerably reduced peculiar velocities on small scales. However, the qualitative nature of the velocity field remained unchanged; the dispersions decreased with separation for $\log(r/L) > -1.7$, and $\langle v_{21}(r)\rangle$ exceeded the Hubble flow in the region where $\xi \approx 1$. The velocity field at these separations is thus insensitive to the behavior of the interaction law on smaller scales. This model produced weakly concentrated clusters and a correlation function that is quite inconsistent with observations of galaxies.

A useful test that the velocities shown in Figure 7 are not unduly influenced by the discreteness of our particles (their individual mass exceeds that of an axion by a factor of order 10^{80}) is to compare them with the velocities found for the "big box" models EdS6–8. If this is done at times later than that plotted in Figure 7*a*, good agreement is found over the range $0.01 < r/L < 0.3$. However, $a = 1.8$ in EdS1–5 corresponds to $a = 1.27$ in EdS6–8, and at such an early time transients from the initial conditions are still visible in the velocity correlations of the "big box" models; corresponding effects may be seen in the two-point functions plotted in Figure 5. The good agreement at later times confirms that discreteness effects are indeed negligible in our models. This might also be inferred from the

fact that the dispersion on all scales in Figure 7 substantially exceed the Keplerian velocity of a binary. The dispersions must therefore result from virial motions within more massive aggregates.

To scale our models to physical units we set the Hubble velocity across the correlation length to its observed value of 500 km s^{-1}. (As discussed above, this corresponds to an unrealistically small value of h in EdS1–5.) For $\Omega = 1$, particles at small separation are then inferred to be moving with one-dimensional rms relative velocities up to 1100 km s^{-1}, while velocities up to 550 km s^{-1} are found in the open ensemble. As we discuss in § IV, values of 200–300 km s^{-1} are required to fit observations of galaxies, and the trend of the observed dispersions with separation disagrees with that seen in our models. Smaller velocities would be obtained if we considered models with lower Ω, but the scale discrepancy discussed in the previous section would then be worse. If one associates the observed galaxy distribution with the mass distribution of the simulations, then velocities in the Einstein–de Sitter models are clearly much too high to be consistent with observation. This disagreement is exacerbated if the present is identified with a larger expansion factor in EdS1–5 (the correlation amplitude of galaxies could then be matched for a larger value of h). It just reflects the well-known fact that the mass-to-light ratios obtained from dynamical analyses of groups and clusters of galaxies are much too small to close the universe. The open models give acceptable velocities at large separations ($\sim 5h^{-1}$ Mpc) but velocities which are still too large on smaller scales. The positive Λ model has very similar velocity correlations to O1 if the two models are compared when $\Omega = 0.2$, and so agrees no better with observation.

d) Three-Point Correlation Function

We have estimated the three-point correlation function $\zeta(r_1, r_2, r_3)$ for our models. This may be parameterized in terms of the usual "size" and "shape" parameters defined by

$$r = r_1 , \quad u = r_2/r_1 , \quad v = (r_3 - r_2)/r_1 , \quad (r_1 < r_2 < r_3)$$

(cf. Peebles and Groth 1975). Observational estimates from the galaxy distribution show that ζ is closely approximated by

$$\zeta(r_1, r_2, r_3) = Q[\xi(r_1)\xi(r_2) + \xi(r_2)\xi(r_3) + \xi(r_3)\xi(r_1)] , \quad (4)$$

where Q is a constant. Groth and Peebles (1977) find $Q = 1.3 \pm 0.2$ from an analysis of the angular correlation functions measured from the Zwicky and Lick catalogues. Efstathiou and Jedredjewski (1985) have estimated Q directly from the three-dimensional distribution using four independent redshift surveys; they find $Q = 0.8 \pm 0.2$, consistent with the Groth and Peebles' result. In our models ζ obeys the scaling with u and v implied by equation (4) very well. In Figure 8, we therefore plot Q, averaged over u and v, against the size parameter r for EdS1–5 and O1–4. In each case, the value of Q on small scales is significantly higher than that observed, and declines to $Q \approx 1$ where $\zeta \approx 1$. Thus the scale dependence of ζ in the N-body models disagrees slightly with observation. In the most extreme cases we find $Q > 2$ on small scales, well outside the observational limits. Note that as the models evolve, the small-scale values of Q decline. This behavior, and

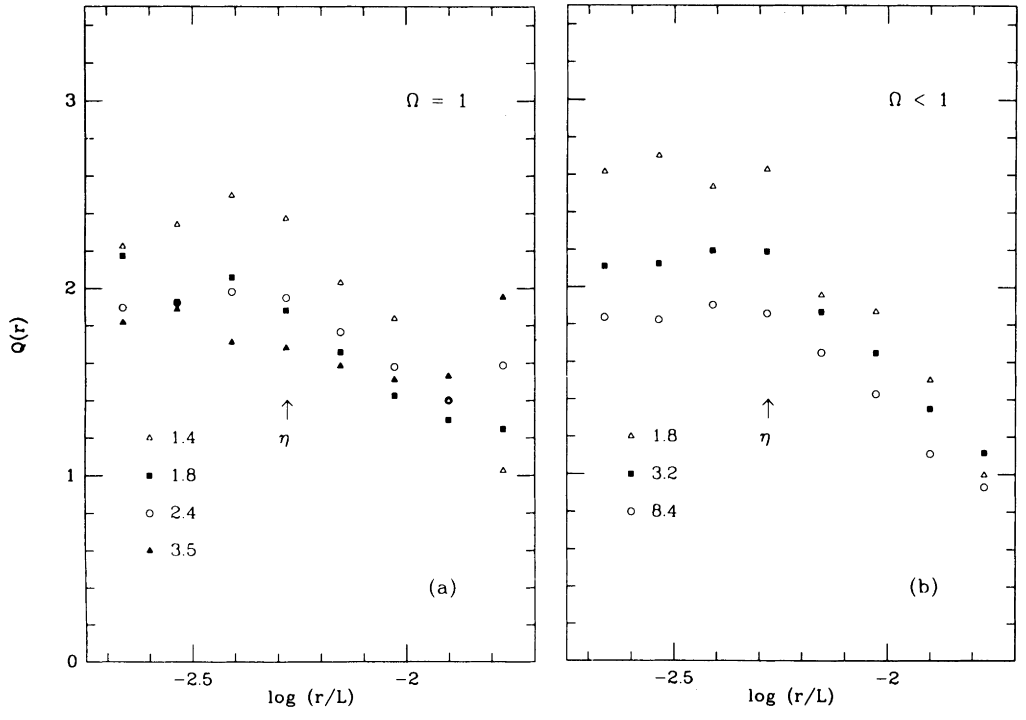


FIG. 8.—The dimensionless amplitude of the three-point correlation function Q (eq. [4]) plotted against the size parameter r. We have averaged Q over u and v because we find it to be almost independent of the triangle shape parameters. Results for the ensemble EdS1–5 are given in (*a*), and for O1–4 in (*b*). Values of Q for different expansion factors are plotted using different symbols as labeled in each panel. Arrows mark the softening length of our simulations.

the high initial values of Q, reflect the shape of the initial fluctuation spectrum. In the hierarchical model of equation (4), Q provides a measure of the dispersion in the number of neighbors found within a sphere of radius r centered on a randomly chosen particle. This dispersion is large compared to the mean number of neighbors if the initial fluctuation spectrum has considerable power on large scales (cf. the discussion of the multiplicity function below). Thus we expect Q to be larger in these models than the value $Q = 1.2 \pm 0.1$ found by Efstathiou and Eastwood for simulations with white noise initial conditions. Similarly, the gradual decrease in the value of Q with time is probably a result of the non–power-law initial conditions in our models. It is worth mentioning that Fry (1984) finds a dependence of Q on primordial spectral index n that is qualitatively similar to the results found here. Fry's calculation is based on second-order perturbation theory, and is not, however, directly applicable to the regime $\zeta > 1$.

e) Multiplicity Functions

Correlation studies of structure in our models may be complemented by an analysis of the distribution of cluster masses. This distribution, often known as the multiplicity function, is heavily influenced by correlations of very high order which cannot be estimated directly. We have found groups in our simulations by linking all particle pairs separated by less than a fraction b of the mean interparticle separation; each distinct subset of connected particles is then defined as a group. This method of cluster analysis has several advantages. In the first place it produces a unique group catalog for any chosen value of b. Second, these catalogs obey a nesting condition; all the members of any group defined for one particular b are also members of the same group in any catalog defined by a larger b (cf. Efstathiou, Fall, and Hogan 1979). Finally, the method makes no *a priori* assumption about the shapes of groups. We have found that if relatively large values of b are used ($b \approx 0.5$), the resulting groups are often quite irregular with several separate centers of concentration. Smaller values of b ($b \leq 0.25$) tend to pick out objects with a well-defined center and relatively regular structure. If the points had a fractal distribution, the shape of groups would be independent of b, so this behavior demonstrates a departure from spatial self-similarity in our models; further work will be required to determine whether this is real or is a consequence of our force softening on small scales ($b = 0.25$ corresponds to a separation of 1.7η in our models). Nevertheless it is interesting that the distribution of nearby galaxies appears to show a similar behavior (Einasto *et al.* 1984).

In Figure 9 we plot the multiplicity functions for our ensemble of Einstein–de Sitter models at two times and for $b = 0.5$ and 0.25. We have followed White and Negroponte (1982) and Einasto *et al.* (1984) in plotting the number of particles as a function of the multiplicity of the group to which they belong; we have also chosen logarithmic bins such that the first contains singles, the second pairs, the third triplets and quadruples, the fourth quintuples to octuples, etc. The most striking aspect of the distributions in Figure 9 is that as the simulations evolve they become flat over the entire multiplicity range 2–1000; a particle is thus equally likely to find itself in a group in each logarithmic mass interval over this range. Multiplicity functions for our ensemble of open models resemble those shown in Figure 9; these seem to be characteristic of CDM models. Evolution from initial conditions with a large coherence length leads to U-shaped distributions in which particles are likely either to be singles or members of very large clusters. This may be seen clearly in the diagrams shown by Einasto *et al.* (1984) for models of a neutrino-dominated universe, and we have found similar results for the neutrino models of White, Frenk, and Davis (1983). Scale-free initial conditions with less power on large scales (e.g., white-noise models) produce configurations in which most particles tend to be in systems of intermediate multiplicity; they look rather like the non-dynamical hierarchical models illustrated by Einasto *et al.* Thus, as emphasized by Bhavsar, Gott, and Aarseth (1981), multiplicity functions seem to discriminate clearly between different kinds of linear initial condition. The flat multiplicity functions of Figure 9 are very similar to the histograms

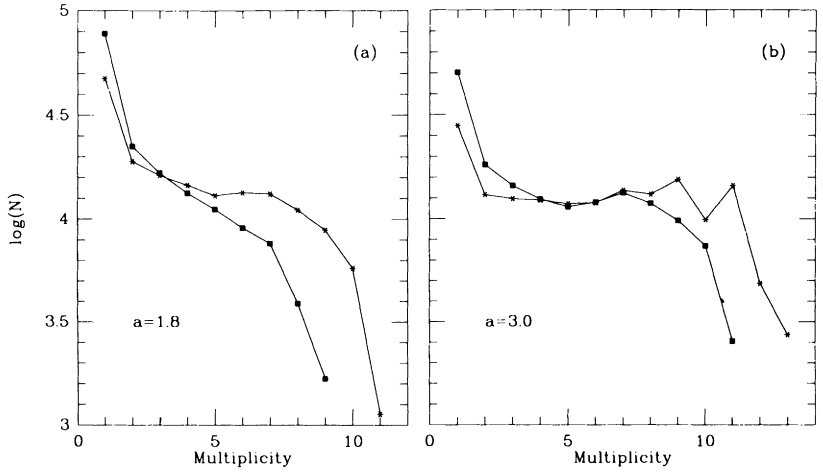


FIG. 9.—The number of particles in ensemble EdS1–5 in groups of a given multiplicity is plotted as a function of that multiplicity. The multiplicity parameter is defined to be the integral part of $\log_2 (N_g) + 1$, where N_g is the number of particles in a group. It is thus equal to 1 for singles, to 2 for pairs, to 3 for triplets and quadruples, and so on. Results are shown for $a = 1.8$ in (*a*) and for $a = 3.0$ in (*b*). In both plots filled squares give results for $b = 0.25$ and stars give results for $b = 0.5$, where b is the maximum separation at which particles are linked into a group; b is expressed in units of the mean interparticle separation.

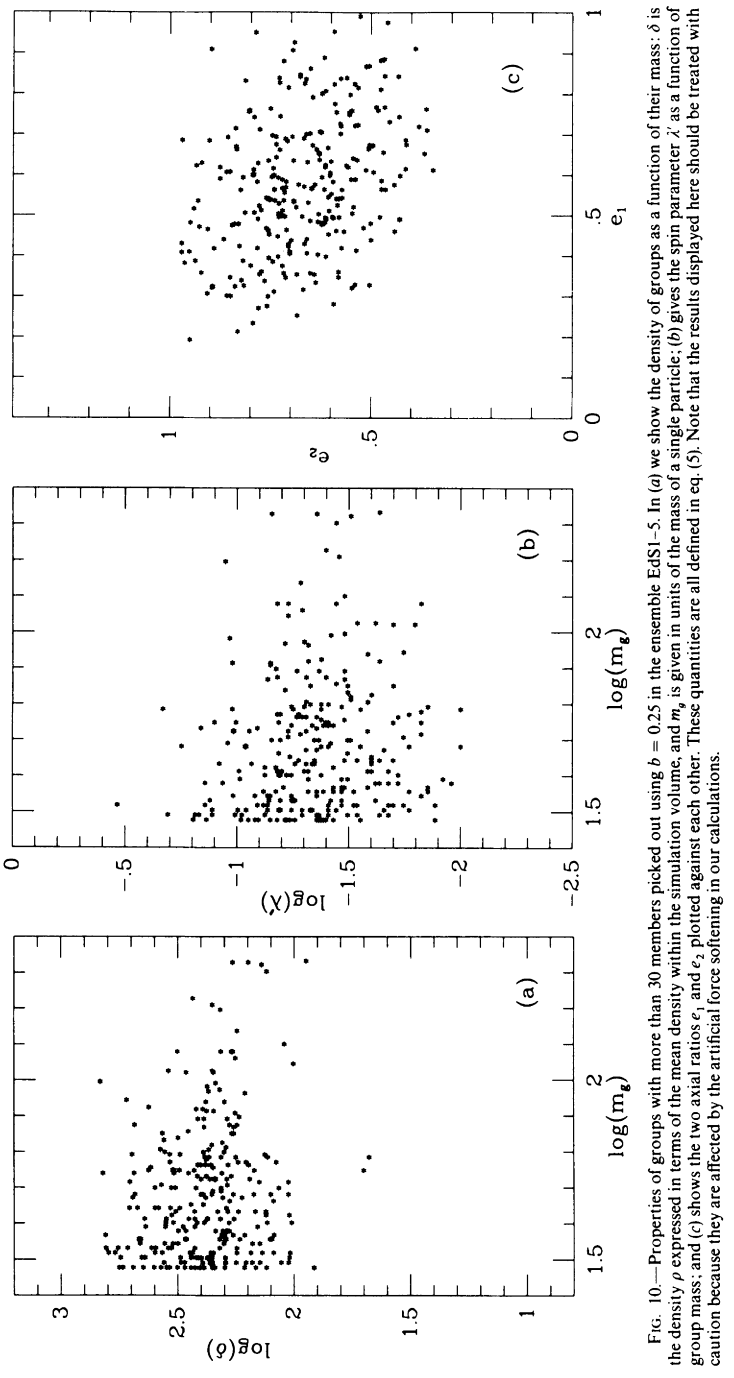


FIG. 10.—Properties of groups with more than 30 members picked out using $b = 0.25$ in the ensemble EdS1–5. In (*a*) we show the density of groups as a function of their mass: δ is the density ρ expressed in terms of the mean density within the simulation volume, and m_g is given in units of the mass of a single particle; (*b*) gives the spin parameter λ' as a function of group mass; and (*c*) shows the two axial ratios e_1 and e_2 plotted against each other. These quantities are all defined in eq. (5). Note that the results displayed here should be treated with caution because they are affected by the artificial force softening in our calculations.

obtained for the nearby galaxy distribution by Einasto *et al.* for corresponding values of b. We hope to come back to this question and make a detailed comparison with a larger sample of observational data in a later paper.

The groups in our simulations correspond to the bound clumps of cold dark matter which must make up the halos of individual galaxies and of groups and clusters of galaxies in the real universe. For small values of b our group-finding algorithm picks out single, well-defined clumps that can be identified with such structures. Although we do not have space here to present a thorough discussion of clump properties, we show some results in Figure 10 to give an idea of what is possible. For each group found by our algorithm we calculated a mass m_g, an inertia tensor with principal components $m_g a_1^2 > m_g a_2^2 > m_g a_3^2$, an internal velocity dispersion $\langle v^2 \rangle$, and an angular momentum vector $\boldsymbol{J}$. From these quantities we derive a mean density, two shape parameters, and a spin parameter using the formulae:

$$\rho = 3m_g/(4\pi 5^{3/2} a_1 a_2 a_3) ,$$
$$e_1 = a_2/a_1 ,$$
$$e_2 = a_3/a_2 ,$$
$$\lambda' = J(\langle v^2 \rangle/2)^{1/2}/Gm_g^2 . \quad (5)$$

In these equations ρ is the density of a uniform ellipsoid with the same mass and inertia tensor as the group, and λ' differs from the usual spin parameter in that it is defined using the kinetic energy of the group rather than its total energy. For an oblate system $e_1 = 1$, while for a prolate system $e_2 = 1$. Figure 10 shows the distribution of some of these quantities for the 273 groups with 30 or more members in models EdS1–5 at $a = 1.8$. We give ρ in units of the mean density of the calculation and m_g in units of the particle mass. For the value $b = 0.25$ used here, the algorithm clearly picks out groups at a fairly well-defined density contrast of about 200. There is no correlation of density with mass except for a slight tendency for the very largest groups to have low density. The spin parameter also shows at most a very slight anticorrelation with mass, and has a median value of about 0.045. The lack of trend agrees with the results of Efstathiou and Barnes (1983) for simulations from white noise initial conditions, but our characteristic value is 30% lower. This difference would be interesting if correct, but it is almost certainly due to the effects of softening in the present calculations; the mean separation of particles in our groups is approximately equal to η. The problem merits further careful investigation using higher resolution simulations. The groups tend to be quite strongly aspherical; axial ratios greater than 2 are quite common. Their generic shape is neither oblate nor prolate, although a clear majority of groups have $e_2 > e_1$, showing some preference for a prolate figure. Again this merits further study to investigate the influence of tides, of internal substructure, and of force-softening on the measured quantities. We have found very similar distributions for $a = 3$ in models EdS1–5, the main difference being that the mass distribution then extends to higher masses (see Fig. 9). The properties of groups in our open models are very similar to those shown in Figure 10, but since these groups are rather more compact, they are even more strongly affected by force softening.

A final problem which we have addressed with our group-finding algorithm is the derivation of the value of b for which linked structures extend across our computational volume and percolation occurs. Einasto *et al.* (1984) suggest this to be a useful measure of structure both in N-body models and in data catalogs; Dekel and West (1985), on the other hand, point out a number of difficulties in interpreting percolation properties. Models EdS1–5 percolate for values of b in the neighborhood of 0.7, both at $a = 1.8$ and at $a = 3$. Percolation is slightly harder at the later time and is also slightly harder in the open ensemble. This value of b is smaller than the value $b = 0.86$ required for percolation in a three-dimensional Poisson distribution. Thus by the criteria of Einasto *et al.* the present models appear filamentary rather than hierarchical, even though Figure 1 shows that clusters do grow in a hierarchical fashion. For the nearby galaxy distribution Einasto *et al.* found percolation to occur for $b = 0.65$, quite close to the value we obtain for our models.

f) *Counts in Cells*

Further information about galaxy clustering is contained in the distribution of counts in cells, and, in particular, in its departure from a Gaussian curve. For the N-body simulations this statistic is very easy to generate, but for the data it is essential to use volume-limited catalogs; this reduces the size of available surveys and makes comparison quite difficult. Figure 11 shows the distribution of counts in cubes of side $L/8$ and $L/4$ for EdS1–5 at $a = 1.8$. At this epoch $r_0/L = 1/26$, so these cubes should be compared with regions of the real universe which are $16h^{-1}$ and $32h^{-1}$ Mpc on a side. For the smaller cube the distribution is extremely non-Gaussian; there are an extended tail at high density and an excess of low density regions. For the larger cube the distribution is much closer to a Gaussian, but even here the departures are quite substantial. The expected rms fluctuation in the number in a cube is $(\delta\rho/\rho)_{\rm rms} = \bar{\xi}^{1/2}$, where $\bar{\xi}$ is the average of the correlation function over the cube. From our cell counts we find rms fluctuations of 70% and 35% for the small and large cubes respectively.

The dashed histogram in Figure 11 corresponds to the distribution of counts in cells of *redshift* space for cubes of side $L/4$. This curve was generated by displacing each particle in the z-direction by its z-component of peculiar velocity divided by the model Hubble constant and then using the same algorithm as before; this procedure approximates the construction of a volume-limited redshift catalog, the z-axis being the redshift direction. Observational catalogs of large-scale clusters and holes are, of course, measured in redshift space, not real space, and should be compared to this histogram. The difference between the redshift-space and the real-space histograms is seen mainly in the low-density cells. Because of the coherent outflowing of material from low-density regions, cells tend to appear emptier in redshift space than in real space. This effect can be substantial, particularly in a high-density universe.

From the cumulative distribution corresponding to Figure 11, we find that 4% of cubes of side $L/4$ have $n/\bar{n} < 0.5$ in real space, versus 8% in redshift space. However, even in redshift space, the fraction of very empty cubes is very small; only 0.6% of cubes of side $L/4$ have $n/\bar{n} < 0.3$. A very similar result is found in models EdS6–8, if we analyze them at $a = 1.3$ when the correlations are similar to those of EdS1–5 at $a = 1.8$. In these models, 10% of $32h^{-1}$ Mpc cubes in redshift space have density less than half the mean. In the open models, the distribution of counts in large cells is virtually identical to that for the $\Omega = 1$ models, provided times with the same linear growth factor are compared. At late times, linear fluctuations cease to

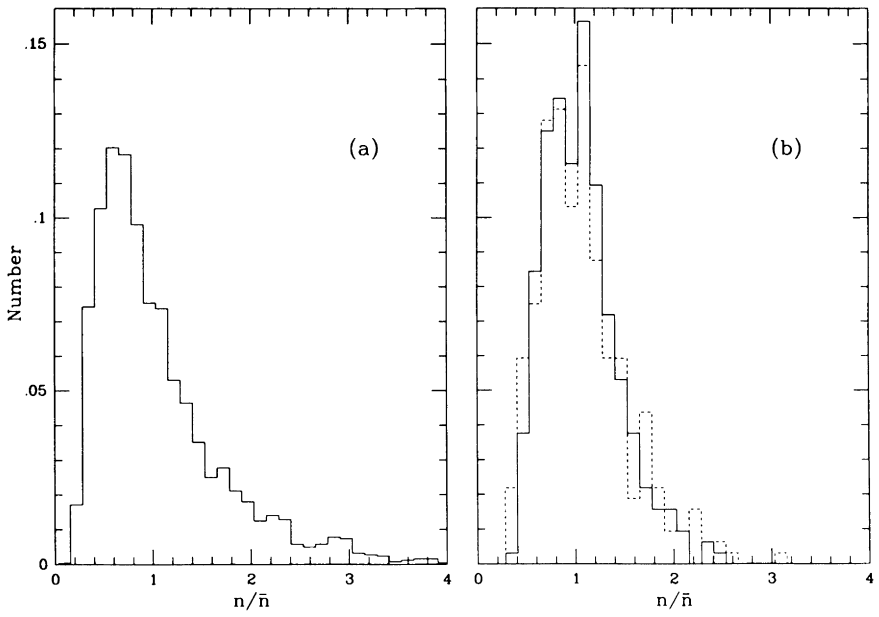


FIG. 11.—Histograms of counts in cells for the ensemble EdS1–5. The distribution of counts in cubes of side one-eighth that of the simulation volume is shown in (*a*), while (*b*) shows the corresponding distribution for cubes of side one-quarter that of the box. The dashed histogram in (*b*) shows the result of making the counts in "redshift" space rather than in real space.

grow in an open universe, and large holes would no longer expand in comoving coordinates.

It is difficult to find observational data to compare with these results. In a subset of the CfA redshift survey (Davis *et al.* 1982; Huchra *et al.* 1983) volume-limited to 6000 km s^{-1} and magnitude-limited to 7500 km s^{-1}, an analysis of counts in randomly placed spheres suggests that 20% of spheres of radius $20h^{-1}$ Mpc (i.e., with the same volume as a $32h^{-1}$ Mpc cube) have density less than half the mean. There is thus a discrepancy with the CDM prediction by about a factor of 2. Note, however, that the total volume of the survey is less than that of eight such spheres, so that the statistics are very poor. Nevertheless the number and sizes of voids within the CfA data set appears to be quite representative (cf. Oort 1983). If really large voids in the galaxy distribution such as the $30h^{-1}$ Mpc radius void in Boötes reported by Kirshner *et al.* (1981, 1983) turn out to be common, they may be difficult to explain within a CDM model. The Boötes region has not been fully surveyed, but Kirshner *et al.* claim the density of bright galaxies there to be less than 25% of the mean. This is particularly startling because the volume of the void is about 10% of all space that has been fully surveyed to date! Within our simulations we see no low-density regions of this size despite the fact that the total simulation "volume" (in units of r_0^3) is greater than the volume of "known" space.

IV. COMPARISON WITH A REDSHIFT CATALOG

In the previous section, we showed that CDM models with $\Omega = 1$ give excessively high velocities and two- and three-point correlation functions incompatible with observation. The amplitudes of the relative peculiar velocities, and of ξ, are better reproduced in models with $\Omega \approx 0.2$, but the discrepancy in the value of Q is worse. Since these open models (at $a \approx 3.2$) are the best approximation to the observed galaxy distribution we have found so far, we now compare some of them with the CfA redshift survey by generating "redshift" catalogs following the procedure of Davis *et al.* (1982). We situate the observer within the computational box and choose points at random in such a way as to mimic the selection function appropriate to the CfA northern catalog, when it is volume-limited to a distance of 4000 km s^{-1}, and magnitude-limited at $m_z = 14.5$ from 4000 km s^{-1} to 10,000 km s^{-1}. Further details are given by Davis and Huchra (1982). We recall that $\gamma = 1.8$ at $a = 3.2$ in O1–4, and that setting the model correlation length to its observed value then requires $h = 1.07$ and a physical box size of 142 Mpc.

Redshift catalogs constructed in this way from runs O2 and O3 (at $a = 3.2$) and projected onto the sky are shown in Figures 12*a* and 12*b*. The semi-volume-limited CfA northern catalog, which these samples model, is shown in Figure 12*c*. To illustrate the kind of fluctuation which may be expected, we have situated the origin near a prominent cluster in model O3, in much the same way as we happen to be near the Virgo supercluster. The selection criteria were unchanged between models O2 and O3, so this procedure results in the number of selected "galaxies" being considerably higher in O3. Wedge diagrams of these three catalogs are shown in Figures 13*a–c*. The simulation data have been plotted in redshift space, which includes the peculiar velocity distortion in the radial direction; the rms value of this velocity is 256 km s^{-1} for the catalog from O2 and 357 km s^{-1} for that from O3. A comparison of these three catalogs is quite instructive. "Fingers of God" can be seen in all three samples, as can long filamentary structures. The cluster we chose in O3 is clearly richer than Virgo, and there is as much difference between the two simulated catalogs as between them and the real data. The large-scale power in the perturbation spectrum produces significant fluctuations in volumes the size of the CfA catalog, so that it is difficult to

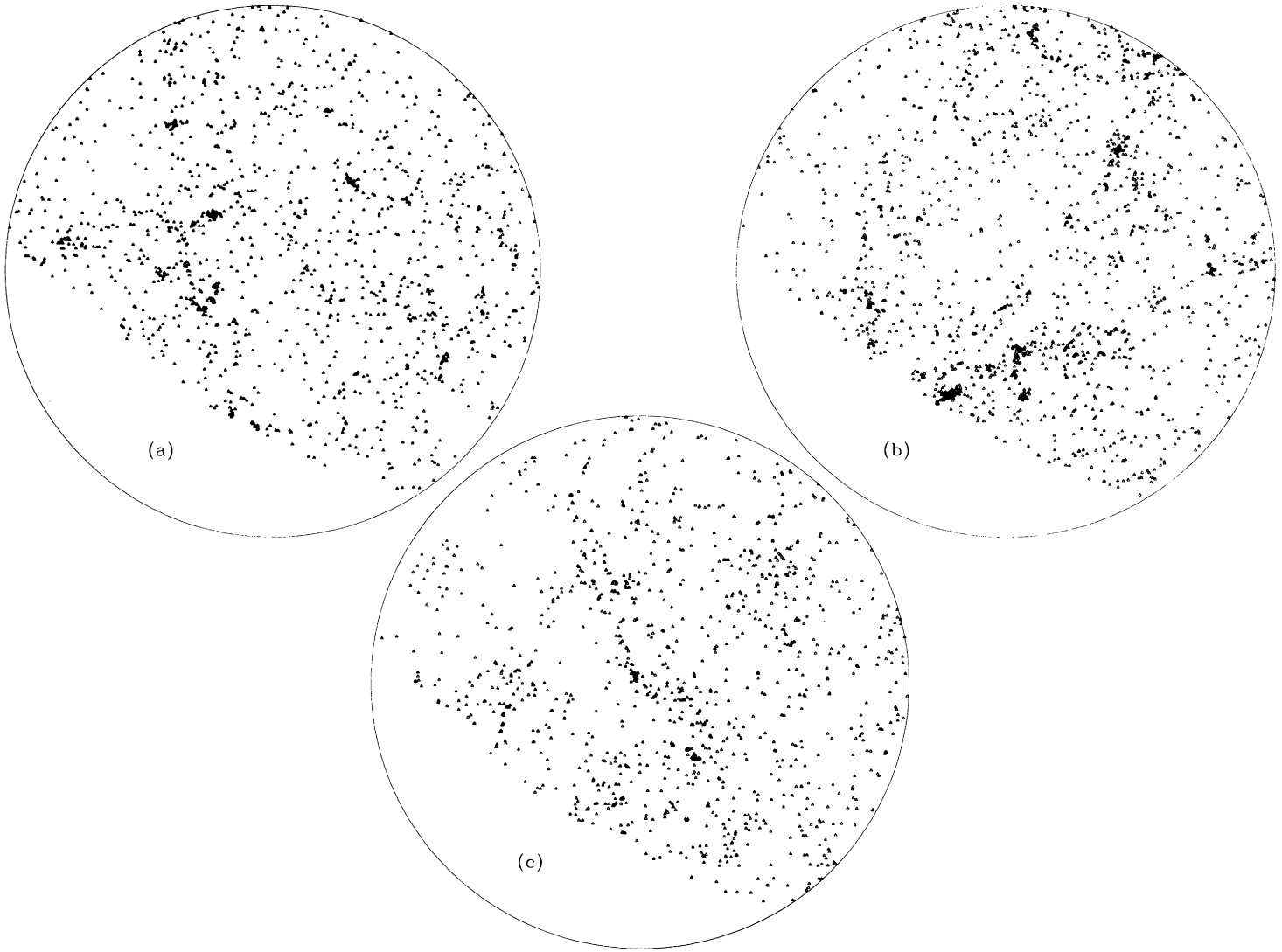


FIG. 12.—Redshift catalogs constructed from two open models (O2 and O3) are shown in (*a*) and (*b*) as projections onto the "sky." Particles were selected for inclusion in these catalogs in such a way as to mimic the northern CfA survey. The real data are shown in the same format in (*c*). These are equal area plots of the sky; the outer circle corresponds to Galactic latitude +40°, while the empty regions correspond to declinations below 0°. In constructing the catalog from O3 shown in (*b*), the "observer" was purposely sited near a prominent cluster.

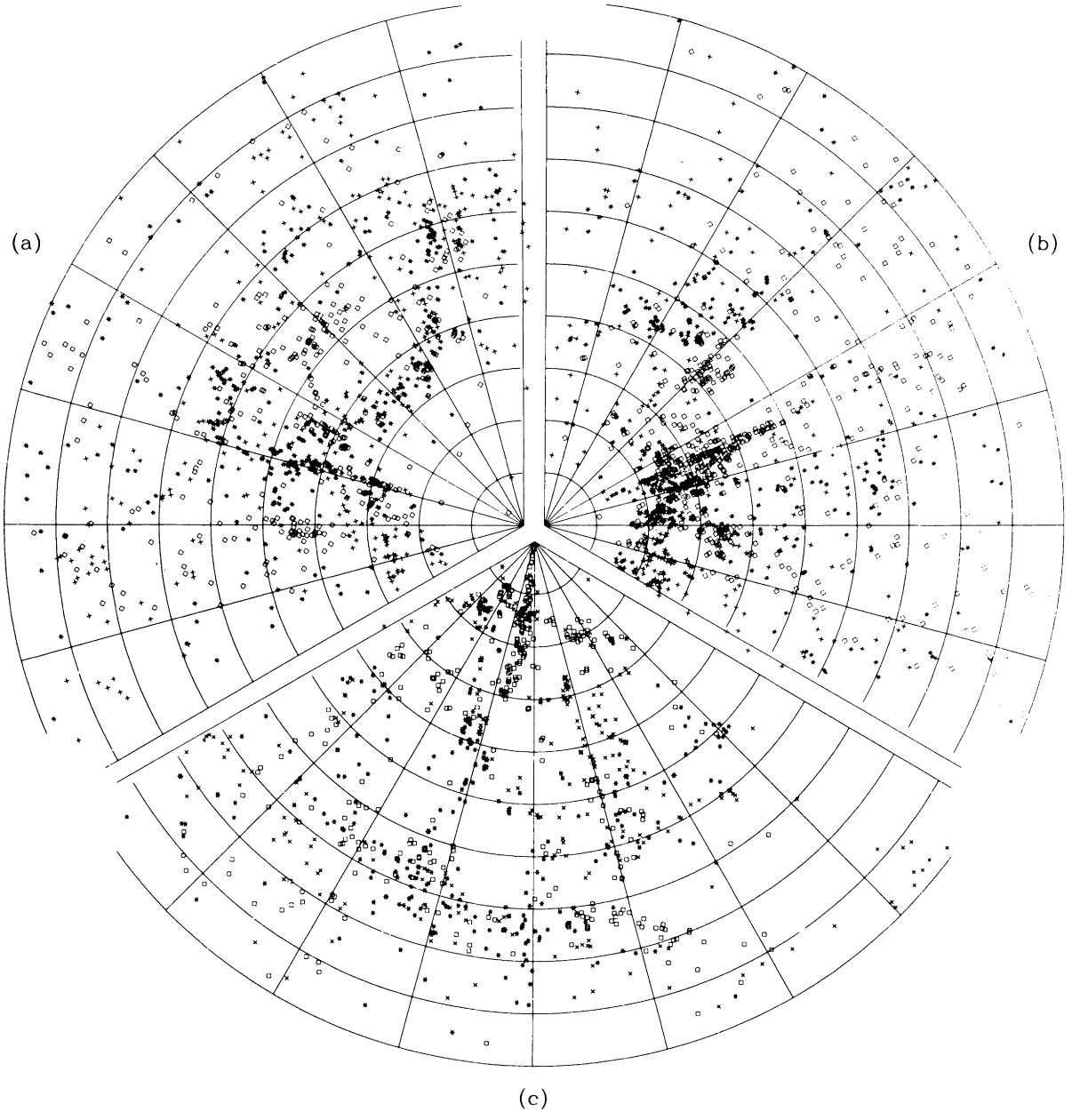


FIG. 13.—Wedge diagrams for the three catalogs illustrated in Fig. 12. The radial coordinate in these plots is recession velocity with successive circles corresponding to increments of 1000 km s^{-1}. The angular coordinate is right ascension with a line plotted for each hour. All "galaxies" in the declination range $0° < \delta < 45°$ are shown, with squares corresponding to $0° < \delta < 15°$, stars to $15° < \delta < 30°$, and crosses to $30° < \delta < 45°$. The three slices of this pie are labeled to correspond to Fig. 12. Note the large cluster in O3 and the Virgo Cluster in the CfA data.

judge the acceptability of a model on the basis of one representation. Such fluctuations also affect the observations, as may be seen by comparing galaxy distributions in the northern and southern hemispheres out to a distance of $50h^{-1}$ Mpc. Figures 12 and 13 give the overall impression that our open models match the observed galaxy distribution very well.

We produced the redshift catalogs for O2 and O3 by scaling the simulations so that the two-point correlation function of the open ensemble matches that of the observations. However, since the redshift catalogs contain only 3%–4% of the original particles, we can expect substantial fluctuations in estimates of $\xi(r)$ derived from them. In Figure 14 we plot $\xi(r)$ computed from the three redshift catalogs in a fashion that attempts to give equal weight to each volume element; this was the method used by Davis and Peebles (1983) for the CfA data. We give ξ as a function of the apparent spatial separation in redshift space; this is a biased estimator of the true $\xi(r)$. The catalog from O3 has correlations quite similar to the CfA data, but that from O2 has a smaller correlation length. Since the models have larger peculiar velocities on small scale than the real data (Fig. 7*b*), their correlation function in redshift space is shallower than that observed. This effect is evident in the O3 catalog for $r < 5h^{-1}$ Mpc. Note also that the CfA data become negative for $r > 28h^{-1}$ Mpc, that the same is true for the O3 catalog at $r > 20h^{-1}$, but that the O2 catalog is positive on almost all scales. The true $\xi(r)$ is actually positive over all these scales in models O2 and O3. The negative correlation seen in the O3 redshift catalog, and perhaps also that in the CfA data, results from the substantial inhomogeneity of these data sets and from the method of estimation. In the analysis of redshift catalogs, one solves for the mean density $\bar{n}$, as well as for $\xi(r)$. Despite the use of a density estimator that is minimally biased by inhomogeneity effects (see Davis and Huchra 1982 for details), the derived $\bar{n}$ for the O3 catalog is slightly higher than the true mean; this reduces the correlations on large scale. In addition, the foreground cluster in this catalog "beats" against a large low-density region just behind it (cf. Fig. 13*b*); this beating produces most of the anticorrelation seen on intermediate scales. One should clearly be cautious when interpreting reports of anticlustering in redshift surveys (Davis and Peebles 1983; Shanks *et al.* 1983).

The distribution of peculiar velocities provides the most stringent observational test of our models. In redshift surveys one can study $\xi(r_p, \Delta v)$, the two-point correlation of pairs as a function of their projected separation r_p and their line-of-sight velocity difference Δv. The rms distortion of ξ in the v direction as a function of r_p is termed $\sigma(r_p)$; its square is a linear combination of the radial and transverse components of $\langle v_{21}{}^2 \rangle$, suitably averaged over projection factors. There is excellent agreement between the estimates of $\sigma(r_p)$ from the CfA and AAT redshift surveys (Davis and Peebles 1983; Bean *et al.* 1983). Both groups report the velocity field to be smooth; the CfA data are best fit by a slowly rising $\sigma(r_p)$, while the AAT data are consistent with this, but are better fit by a constant $\sigma(r_p)$. We have applied the procedure of Davis and Peebles (1983) to our artificial "redshift" catalogs, and the results are plotted in Figure 15. The error bars are formal errors from fits to an exponential broadening function, which represents both the CfA data and the models quite well. The f parameter used by Davis and Peebles is here set to unity, which agrees with $\langle v_{21}(r) \rangle$ as plotted in Figure 7*b*. It was to be expected from the behavior of the full peculiar velocity-dispersion tensor that $\sigma(r_p)$ would fall with radius; Figure 15 demonstrates that it is indeed possible to see this trend in a redshift catalog, and that it is inconsistent with the CfA and AAT data. We showed in § IIIc that the falling dispersions of our models are not a consequence of their experimental limitations; they must therefore be considered a serious disagreement with observation. The substantial difference between $\sigma(r_p)$ in the two simulations is caused by the large cluster in the foreground of the O3 catalog.

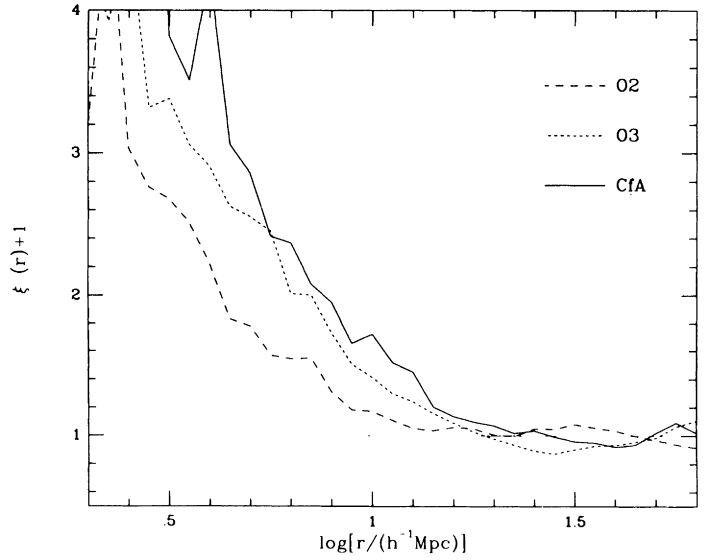


FIG. 14.—Correlation functions calculated in redshift space for the three catalogs of Fig. 12. Note that r is not a true physical distance in this plot since it includes distortions caused by peculiar velocities. The scaling of the simulations assumes $h = 1.1$, the value required to get $\xi(r)$ for the ensemble of open models to match observation (see Fig. 6).

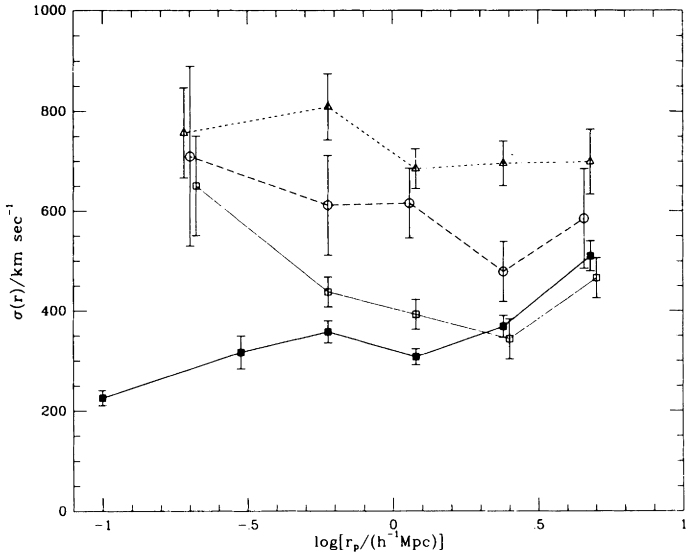


FIG. 15.—The rms line-of-sight peculiar relative velocity as a function of projected separation for catalogs made from O2 (*open squares*), from O3 (*open triangles*), from the CfA redshift survey (*filled squares*; from Davis and Peebles 1983), and from the "galaxies" found in EdS1–5 by the algorithm described in § V. The two open models have been scaled using $h = 1.1$, while the biased formation ensemble has been scaled using $h = 0.44$.

The velocity moments are computed on a pair-weighted basis, and the presence of this cluster skews the results. This effect is undoubtedly larger in the O3 catalog than in the real data, because of the richness of the cluster we chose to sit next to.

We have found that no CDM model is in complete agreement with observation if galaxies are distributed in the same way as the mass; the open models are a better fit than models with $\Omega = 1$, but are still not fully acceptable. To try and circumvent these problems, we now explore the effect of relaxing the assumption that galaxies trace the mass. This is clearly necessary if we wish to find a viable model with $\Omega = 1$.

V. BIASED GALAXY FORMATION

In a universe dominated by CDM, galaxy formation presumably occurs at peaks of the matter-density field. Now suppose that galaxy formation is a process that can go to completion only under favorable circumstances and can be suppressed if the environment changes. This might occur, for example, if the first generation of stars photoionizes, or shock heats to high temperature, those regions that were somewhat slower to collapse, cool, and fragment. It is then reasonable to expect that the highest peaks (which have the shortest cooling and collapse times) will be the most likely to complete the process of galaxy formation successfully. A similar biasing of the bright-galaxy distribution might occur if the stellar initial mass function is sensitive to local environment, or perhaps simply if galaxy formation is sufficiently sensitive to protogalactic cooling time.

To simulate this kind of situation, we applied a low-pass filter to the initial random-density field of our models and tabulated the result on a 64^3 grid. We then identified all peaks of amplitude at least ν times the rms density fluctuation and tagged the particle whose Lagrangian position lay nearest each one as a "galaxy." Only one "galaxy" was allowed for each connected region above the threshold. The tagged particles were followed over the course of the simulation. This identification procedure involves two parameters, the dispersion r_s of the Gaussian filter $\exp(-r^2/2r_s^2)$ used to smooth the density field, and the threshold ν. The scale parameter defines the size of the regions from which a galaxy forms, while the threshold is some measure of the "difficulty" of galaxy formation. Because there is so much power on large scales in the CDM models, it is significantly more probable for a local peak to rise above a fixed threshold if it lies in an incipient protocluster than if it lies in an incipient void. As a result our "galaxies" are a biased subset and are more strongly clustered than the simulation as a whole. We find the strength of this bias to increase strongly with ν, but only rather weakly with r_s; the number of particles tagged as galaxies depends strongly on both parameters. A similar effect has recently been explored by Kaiser (1984*a*) as a possible explanation for the large correlation length of the distribution of rich galaxy clusters. Analytic discussions of biasing in the present context are given by Bardeen (1984) and Kaiser (1984*b*). With $r_s = L/192$ and $\nu = 2.5$, we obtain about 1000 tagged points, or about 3% of the original. After adopting a suitable scaling, this turns out to be comparable to, or somewhat less than, the expected number of bright galaxies in a volume the size of our box (i.e., to a density $\sim 0.003h^3$ Mpc^{-3}).

The result of applying this procedure to run EdS1 is illustrated in Figure 16, which compares the distribution of "galaxies" with that of the dark matter at $a = 1.4$. The "galaxies" follow the clustering structure of the underlying matter field but with greatly enhanced amplitude. The effect of this bias on $\xi(r)$ is shown in Figure 17*a*, where we plot the evolution of the "galaxy" correlation function in models EdS1–5 and compare it with the unbiased results obtained when all particles are used; power-law fits to these correlation functions are given in Table 2C. The "galaxy" correlations substantially exceed the mass correlations on all scales, but the effect is most marked at small separations and produces an increase in the effective power-law index γ. Much of the clustering of the "galaxies" is due to the pattern imposed by the

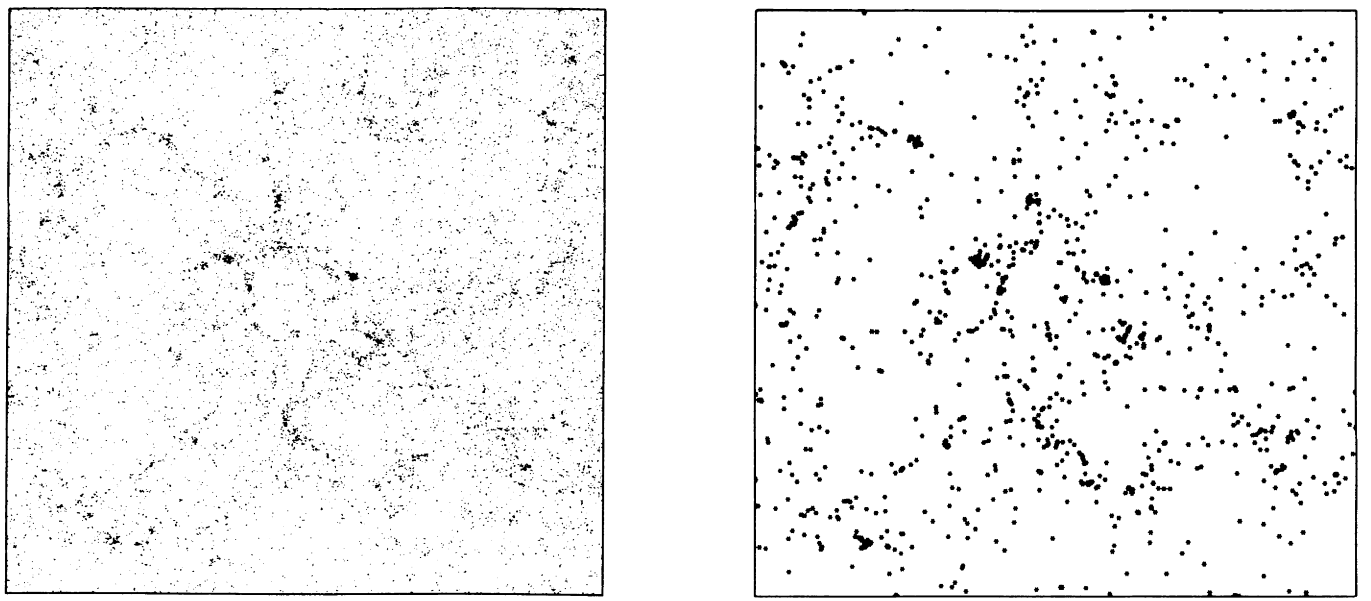


FIG. 16.—The projected distribution of all particles (*left*) and of the "galaxies" (*right*) in EdS1 at $a = 1.4$. The side of the box is $32.5h^{-1}$ Mpc. "Galaxies" are assumed to form only at the 2.5 σ peaks of the linear density distribution.

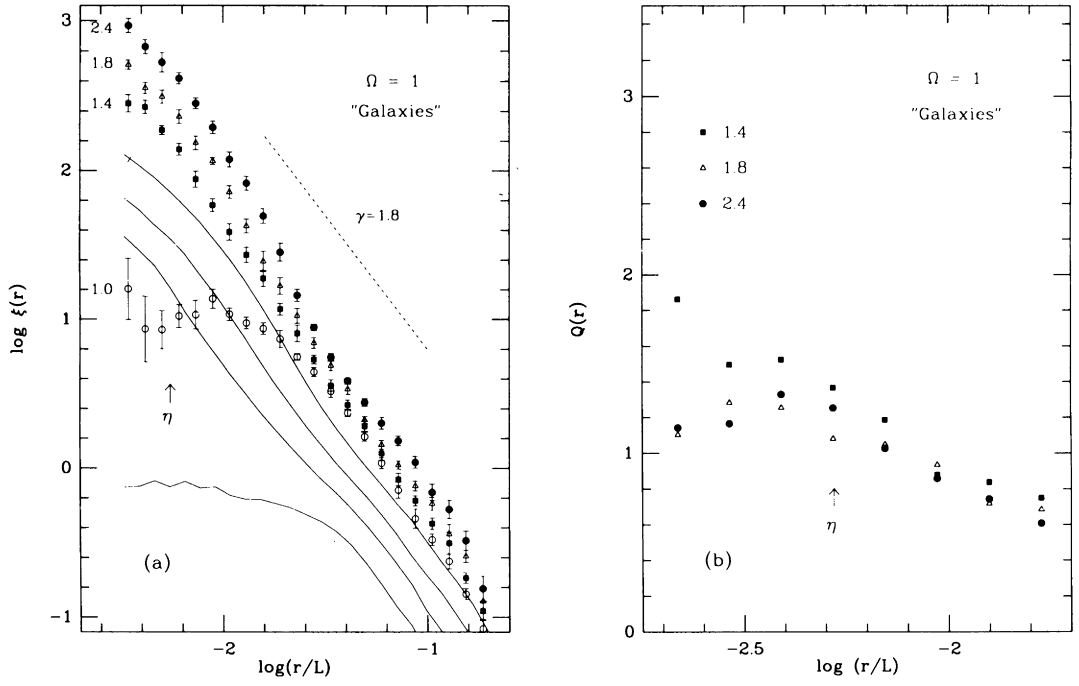


FIG. 17.—Correlation properties of "galaxies" in EdS1–5 on the biased galaxy formation model. (*a*) compares the evolution of $\xi(r)$ for the "galaxies" (*symbols with error bars*) with that of the mass distribution as a whole (*solid lines*). (*b*) shows Q values for the "galaxies" in the same format as Fig. 8.

statistical properties of the linear density field rather than to the effects of gravity. This is particularly evident on large scales, where correlations grow very little until the underlying matter distribution has become significantly clustered. As a result, $\xi(r)$ for the "galaxies" steepens even more rapidly than for the simulation as a whole. By $a = 1.4$, our models have an effective γ which exceeds 1.8, and by $a = 1.8$, the galaxy correlation function is clearly too steep to match observation. The correlation length r_0 for the "galaxies" exceeds that of the underlying mass distribution by a factor of 2.4 at $a = 1.4$. The two- and three-point correlations of the "galaxies" obey the empirical formula of equation (4) at least as well as those of the mass, and, as shown in Figure 17*b*, they imply Q values which agree better with observation. The relative peculiar velocities of the "galaxies" behave in a similar way to those of the simulations as a whole; the bias is evident only as a slight increase in the dispersions. This may be seen in Figure 18, as may the fact that the dispersion profiles for the "galaxies" are somewhat flatter than those of Figure 7*a*.

To scale these "galaxy" catalogs to the real world, we again set the correlation length r_0 to $5h^{-1}$ Mpc. Unfortunately, while the correlation exponent γ is about 1.8 at $a = 1.2$, the peculiar velocity distribution shows the effects of initial transients until about $a = 1.8$. As a compromise we identify the present with $a = 1.4$ when ξ is only slightly steeper than the observations and velocity transients have already subsided to a relatively low level. At this time $r_0 = 0.07L$, leading to a value of 0.44 for Ω and to a box size $L = 168$ Mpc. With this scaling the velocity unit in Figure 18 is 7400 km s^{-1}, and the one-dimensional dispersion has a maximum of 575 km s^{-1}, from which it drops to values of order 400 km s^{-1}. When redshift catalogs are made from the "galaxy" distributions, these dispersions are found to lead to the values of $\sigma(r_p)$ shown in Figure 15, which are in as good agreement with observation as those for unbiased models with $\Omega = 0.2$. The redshift catalogs for our "galaxies" are too sparse for visual comparison with the CfA survey (as in Figs. 12 and 13), but we can get some information about the probability of large low-density regions from the distribution of counts in cells. In EdS1–5, we find only one cube out of 40 with side $7r_0$ ($35h^{-1}$ Mpc) to have a density less than half the mean. Holes appear to be relatively less abundant in these "galaxy" samples than in the overall mass distribution (cf. § III*f*). Thus voids as large as the one in Boötes may be hard to accommodate in our model. Simulations of larger scale are required to obtain a definitive answer on this point.

Although our prescription for galaxy formation is very crude, it is clear that the biases which arise if there is an effective threshold for galaxy formation can be very strong. Furthermore, their effects are of the kind required to bring an Einstein–de Sitter CDM universe into agreement with observation. The particular models which we discuss above have two-point correlations and peculiar velocity distributions which are quite similar to those that we obtained earlier for our open models. They are more successful than the open models in reproducing the observed three-point correlations, and they imply an age for the universe which is not uncomfortably short. Furthermore, they satisfy the theoretical desire for a flat universe without contradicting dynamical studies of galaxy clustering. Finally, they lead to predicted fluctuations in the microwave background which are far below the current observational upper limits (Bond and Efstathiou 1984; Vittorio and Silk 1984). On the negative side, the biased models seem less able to produce large-scale voids than the open models, and they predict that galaxies formed at quite recent epochs—this

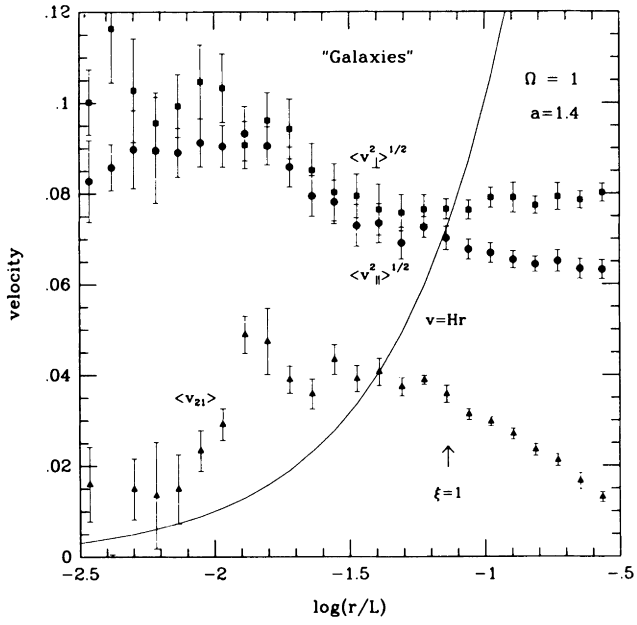


FIG. 18.—Moments of the relative peculiar velocity distribution of pairs of "galaxies" in EdS1–5 at $a = 1.4$ in the same format as Fig. 7.

may conflict with observation. Clearly, what is now required is a proper physical model for galaxy formation which can be grafted onto simulations to see if the distribution of our "galaxies" is indeed realistic. These biased galaxy-formation models are in many ways the closest we have come to matching the observed galaxy distribution, and they involve the minimum gravitational interaction!

VI. CONCLUSIONS

The numerical simulations discussed in this paper were designed to study nonlinear gravitational clustering in a universe dominated by cold dark matter. We assumed adiabatic primordial density fluctuations with a constant-curvature spectrum. We followed the evolution of two ensembles of simulations with $\Omega = 1$, another ensemble with $\Omega < 1$, and a single simulation of a flat universe with a nonzero cosmological constant. Our main conclusions are as follows:

1. Initial conditions of this kind lead to a very rapid growth in the characteristic scale of clustering. As a result, nonlinear structures on a wide range of scales are present at any given moment. This is reflected in the filamentary structures, the superclusters of clumps, and the large low-density regions which are visible in pictures of the particle distribution. This large-scale morphology is clearly a very general consequence of nonlinear evolution from such initial conditions. While its scale and its evolution rate depend on Ω, Λ, and h, its form is almost independent of them.

2. The particle two-point correlation function $\xi(r)$ steepens with time in all the models, showing that their evolution is not self-similar. The good agreement we find between ensembles of different scale argues strongly that this result is not a consequence of the limited spatial and temporal resolution of our experiments, but rather reflects the lack of spatial self-similarity in the theoretical initial conditions. In each ensemble of simulations, a time interval can be chosen so that ξ provides an acceptable match to the shape of the observed galaxy two-point function over a limited range of scales. The amplitude of the observed function can then be matched by choosing h appropriately. For Einstein–de Sitter models we require $h \approx 0.22$, which is unacceptably low. On the other hand, for models in which ξ has the right shape when $\Omega = 0.2$, we need $h \approx 1.1$; this seems rather high. Our model with zero spatial curvature and positive cosmological constant has very similar properties to open models with the same value of Ω; however, it implies a considerably greater age for the universe.

3. The peculiar velocity distribution in our simulations does not give a good match to observations of galaxies. For $\Omega = 1$ we find rms one-dimensional relative peculiar velocities of $\sigma \approx 1000$ km s^{-1} at separations $r \approx 1h^{-1}$ Mpc, much in excess of the observational estimates, $\sigma \approx 200$–300 km s^{-1}; this is a reflection of the well-known fact that the mass-to-light ratios of small groups and clusters appear much smaller than that needed to close the universe. For $\Omega = 0.2$ we find $\sigma \approx 500$ km s^{-1} on small scales, which is still rather too big. At $r \approx 5h^{-1}$ Mpc, however, peculiar velocities in the open models are quite similar to those observed. In the simulation with a nonzero cosmological constant, the peculiar velocity distribution resembles that in an open model with the same value of Ω. In all the models σ is found to decline markedly with increasing pair separation, whereas the observations show it to be constant, or to increase slightly with scale.

4. If we write the irreducible three-point correlation function as $\zeta = Q\xi^2$, we find that, although Q is independent of triangle shape, it decreases slowly with triangle size in all our models. On scales $r \leq 5h^{-1}$ Mpc, we find values of Q for the mass distribution which are considerably greater than those estimated for galaxies. This discrepancy gets smaller as our models evolve, but it always remains significant.

5. The distribution of cluster masses is very broad in our

models and is quite different from that found either in simulations with white-noise initial conditions or in models where the initial density field has a large coherence length. Despite the fact that structure grows hierarchically in the present models, percolation across them is easier than across a Poisson distribution of points; in this sense our models show filamentary structure. These properties correspond quite closely to the results of Einasto *et al.* (1984) for the galaxy distribution in the neighborhood of the Local Supercluster.

6. The distribution of particle counts in cells is found to have significantly non-Gaussian tails, even for cubic cells with side exceeding $6r_0$ (where $\xi(r_0) = 1$). 4% of such cubes are underdense by a factor of 2. This frequency doubles if the distribution is analyzed in "redshift space." The observed frequency of such large low-density regions is very poorly known but appears to be about a factor of 2 larger than in our models. None of our simulations contains an underdense region as extreme as the apparent void in Boötes.

7. Artificial redshift catalogs of points, chosen from our simulations in a way which mimics the selection of galaxies for the CfA redshift survey, show considerable resemblance to the real data. These samples also show that substantial fluctuations are to be expected in estimates of clustering statistics derived from samples of this size. In particular it is quite likely that the anticorrelation seen in the CfA survey at separations of order $20h^{-1}$ Mpc is an artifact of the relatively small volume surveyed.

8. In all the above comparisons with observation, we assume that the distribution of galaxies may be identified with the mass distribution in our models. Observed galaxies may, however, be a biased reflection of the real underlying mass distribution. We have investigated the kind of bias to be expected if bright galaxies form only at relatively high peaks of the linear density distribution. If this is the case, their large-scale clustering follows that of the dark matter, but with strongly enhanced amplitude. For plausible parameters we find that the correlation length of galaxies in the present universe may exceed that of the dark matter by a factor of about 2.5. The peculiar velocities of galaxies are then smaller than those expected for an unbiased distribution by a similar factor. Applying a simple model of biased galaxy formation to our Einstein–de Sitter simulations, we find a "galaxy" distribution which fits the observed two- and three-point correlations and the observed peculiar velocity distribution quite well. This kind of model may therefore reconcile the theoretical imperative of a flat universe with the low density values implied by observations of galaxy motions.

The evolution of a universe dominated by cold dark matter is difficult to simulate because of the speed with which the characteristic scale of clustering grows. Even with a very large increase in computing resources, it would be possible to follow clustering at times before the start of our models over expansion by at most a factor of 2. Nevertheless, models with higher spatial resolution than those of the present paper are desirable in order to study the formation of galactic halos in greater detail. This is clearly a prerequisite for any more detailed theory of galaxy formation, and it would also be useful for a more thorough investigation of the consequences of our own very simple prescription. Models with a larger dynamic range are necessary to study the large-scale morphology of the "galaxy" distribution because small-scale resolution must be retained to locate the initial density peaks on which "galaxies" form. Until we have made such models, we cannot be sure that we can reproduce not only the correlation functions and other small-scale statistics of galaxies but also the filaments and voids which characterize their large-scale distribution. Higher resolution models would also help us to pin down the epoch of galaxy formation. This is important because it appears to be quite recent, particularly in the biased formation models, and so may be in conflict with observation.

Structure forms in a universe dominated by CDM in a well-specified way, and it is worth mentioning several aspects of its evolution which may be amenable to direct observational testing. The rapid increase in the characteristic scale of clustering might be observable on Mpc scales if $\Omega = 1$ and galaxies trace the mass distribution; however, such models are already ruled out by the kinematics of nearby galaxies. If the universe is open, or if our biased galaxy formation model is appropriate, the correlation function is a much weaker function of lookback time. Existing data on angular correlations of galaxies at $z \approx 0.3$ (Koo and Szalay 1984) may put significant constraints on these possibilities. The biased formation model implies that "failed" galaxies should be visible in our local region of space. The voids should contain some baryons either in low-density gas clouds or in low-surface-brightness galaxies. Space Telescope will be an excellent tool to search for such clouds. One would look for low-redshift Lyα absorption clouds in the spectra of distant QSOs. It is known that the Lyα clouds observed at $z = 2$ are more weakly clustered than bright galaxies (Sargent *et al.* 1980); this may already be evidence for biased galaxy formation. At low redshift the distribution of clouds found by Space Telescope could be correlated with the galaxy distribution measured by redshift surveys; if the clouds clustered only weakly with galaxies, and lay preferentially in "voids," it would be clear evidence that galaxy formation is biased on large scales.

The distribution of rich clusters of galaxies and of large voids presumably gives us information about initial fluctuations on large scale. The correlation function of Abell clusters can be understood as reflecting the enhanced clustering of rare peaks in the matter distribution if the underlying mass distribution is positively correlated on the relevant scales (Kaiser 1984*a*; Barnes *et al.* 1985). The linear correlation function corresponding to equation (2) is negative for $r > 18(\Omega h^2)^{-1}$ Mpc, so that if Abell clusters are indeed positively correlated at $r = 100h^{-1}$ Mpc (Bahcall and Soneira 1983), then we must conclude that $\Omega h < 0.18$. The Boötes void appears incompatible with CDM models in which galaxies are assumed to follow the mass distribution, as well as with the recipe for biased galaxy formation which we have considered. It is not clear whether an alternative recipe could explain the low density of galaxies in such a large region of space.

Further constraints on the CDM model come from observational limits on the anisotropy of the microwave background radiation. Recent calculations have shown that the stringent limit of Uson and Wilkinson (1984) is able to rule out CDM models with $\Omega h^{1.3} < 0.2$ at the 95% confidence level unless the universe was reionized at $z \geq 50$ (Bond and Efstathiou 1984; Vittorio and Silk 1984). The energy requirements for reionization are quite extreme, and if this loophole is discounted, the microwave background limits exclude all unbiased models for which we find even rough agreement with the observed peculiar velocity distribution of galaxies. The microwave limits are, however, very far from constraining the kind of biased formation model we have considered.

A great virtue of the general theoretical framework investi-

gated in this article is the fact that it makes very specific predictions for the initial conditions from which structure must form. It may therefore be possible to disprove it, in the same way as we found the standard neutrino-dominated universe to be untenable once its observational consequences were analyzed sufficiently carefully (White, Frenk, and Davis 1983; White, Davis, and Frenk 1984). Conversely, if the properties of a universe filled with CDM are found to agree with observation, this must be considered a significant success, since there is very little freedom to adjust the theoretical predictions. The direct detection of most of the elementary-particle candidates for CDM remains a remote possibility, so their cosmological implications may be the strongest indication of their existence. The major uncertainty in these implications comes from the fact that while we can predict the distribution of mass, what we see is the distribution of galaxies. As we have shown, the latter may be strongly influenced by the details of galaxy formation, a process which we are unlikely to understand fully in the foreseeable future. Nevertheless, it is remarkable how many aspects of the observed galaxy distribution are reflected quite faithfully by the distribution of CDM, and our crude attempts to model galaxy formation suggest that $\Omega = 1$ might even be *required* to get good agreement with observation. This seems too good to be true, but perhaps it hints that we are at last approaching a correct resolution of the missing mass problem.

This research has been carried out over a period of more than two years and has benefited greatly from discussions with many people. We would like to extend particular thanks to all participants in the 1984 program of the Evolution of Structure in the Universe at the Institute for Theoretical Physics. We are also very grateful to the director and to his staff for their warm hospitality during our stay. Various stages of this work were supported by the UK SERC, by the NSF under grants nos. AST 81-14715, AST 81-18557, PHY 77-27084 (supplemented by funds from the National Aeronautics and Space Administration), and by Department of Energy contract AT03-82ER1240069. Considerable amounts of computer time were provided by the ITP, by the Royal Greenwich Observatory, and by the Institute of Astronomy, Cambridge.

REFERENCES

Bahcall, N. A., and Soneira, R. M. 1983, *Ap. J.*, **270**, 20.
Bardeen, J. M. 1984, in *Inner Space/Outer Space* (Chicago: University of Chicago Press), in press.
Bardeen, J. M., Steinhardt, P. J., and Turner, M. S. 1983, *Phys. Rev. D*, **28**, 679.
Barnes, J., Dekel, A., Efstathiou, G., and Frenk, C. S. 1985, *Ap. J.*, submitted.
Bean, A. J., Efstathiou, G., Ellis, R. S., Peterson, B. A., and Shanks, T. 1983, *M.N.R.A.S.*, **205**, 605.
Bhavsar, S. P., Gott, J. R., and Aarseth, S. J. 1981, *Ap. J.*, **246**, 65.
Blumenthal, G. R., Faber, S. M., Primack, J. R., and Rees, M. J. 1984, *Nature*, **311**, 517.
Blumenthal, G. R., and Primack, J. R. 1983, in *Fourth Workshop on Grand Unification*, ed. H. A. Weldon, P. Langacker, and P. J. Steinhardt (Boston: Birkhauser), p. 256.
Bond, J. R., Centrella, J., Szalay, A. S., and Wilson, J. R. 1983, in *Formation and Evolution of Galaxies and Large Structures in the Universe*, ed. J. Audouze and J. Tran Thanh Van (Dordrecht: Reidel), p. 87.
Bond, J. R., and Efstathiou, G. 1984, *Ap. J. (Letters)*, **285**, L45.
Bond, J. R., and Szalay, A. S. 1983, *Ap. J.*, **274**, 443.
Bond, J. R., Szalay, A. S., and Turner, M. S. 1982, *Phys. Rev. Letters*, **48**, 1636.
Centrella, J., and Melott, A. L. 1983, *Nature*, **305**, 196.
Davis, M. 1984, in *Inner Space/Outer Space* (Chicago: University of Chicago Press), in press.
Davis, M., and Huchra, J. 1982, *Ap. J.*, **254**, 425.
Davis, H., Huchra, J., Latham, D. W., and Tonry, J. 1982, *Ap. J.*, **253**, 423.
Davis, M., and Peebles, P. J. E. 1977, *Ap. J. Suppl.*, **34**, 425.
———. 1983, *Ap. J.*, **267**, 465.
Dekel, A., and West, M. J. 1985, *Ap. J.*, **288**, 411.
Efstathiou, G., and Barnes, J. 1983, in *Formation and Evolution of Galaxies and Large Structures in the Universe*, ed. J. Audouze and J. Tran Thanh Van (Dordrecht: Reidel), p. 361.
Efstathiou, G., Davis, M., Frenk, C. S., and White, S. D. M. 1985, *Ap. J. Suppl.*, **57**, 241 (EDFW).
Efstathiou, G., and Eastwood, J. W. 1981, *M.N.R.A.S.*, **194**, 503.
Efstathiou, G., and Jedredjewski, R. 1985, in preparation.
Efstathiou, G., Fall, S. M., and Hogan, C. 1979, *M.N.R.A.S.*, **189**, 203.
Einasto, J., Klypin, A. A., Saar, E., and Shandarin, S. F. 1984, *M.N.R.A.S.*, **206**, 529.
Fall, S. M. 1979, *Rev. Mod. Phys.*, **51**, 21.
Frenk, C. S., White, S. D. M., and Davis, M. 1983, *Ap. J.*, **271**, 417.
Fry, J. 1984, *Ap. J.*, **279**, 499.
Groth, E. J., and Peebles, P. J. E. 1977, *Ap. J.*, **217**, 385.
Guth, A. 1981, *Phys. Rev. D*, **23**, 347.
Guth, A., and Pi, S.-Y. 1982, *Phys. Rev. Letters*, **49**, 1110.
Harrison, E. R. 1970, *Phys. Rev. D*, **1**, 2726.
Hawking, S. 1982, *Phys. Letters*, **115B**, 295.
Hockney, R. W., and Eastwood, J. W. 1981, *Computer Simulations using Particles* (New York: McGraw-Hill).
Huchra, J., Davis, M., Latham, D. W., and Tonry, J. 1983, *Ap. J. Suppl.*, **53**, 89.
Kaiser, N. 1984*a*, *Ap. J. (Letters)*, **284**, L9.
———. 1984*b*, in *Inner Space/Outer Space* (Chicago: University of Chicago Press), in press.
Kibble, T. W. B. 1976, *J. Phys. A*, **9**, 1387.
Kirshner, R. P., Oemler, A., Schechter, P. L., and Shectman, S. A. 1981, *Ap. J. (Letters)*, **248**, L57.
———. 1983, in *IAU Symposium 104, Early Evolution of the Universe and its Present Structure*, ed. G. O. Abell and G. Chincarini (Dordrecht: Reidel), p. 197.
Klypin, A. A., and Shandarin, S. F. 1983, *M.N.R.A.S.*, **204**, 891.
Koo, D., and Szalay, A. S. 1984, *Ap. J.*, **282**, 390.
Melott, A. L., Einasto, J., Saar, E., Suisalu, I., Klypin, A. A., and Shandarin, S. F. 1983, *Phys. Rev. Letters*, **51**, 935.
Oort, J. 1983, *Ann. Rev. Astr. Ap.*, **21**, 373.
Pagels, H. R., and Primack, J. R. 1982, *Phys. Rev. Letters*, **48**, 223.
Peebles, P. J. E. 1974, *Ap. J. (Letters)*, **189**, L51.
———. 1980, *The Large Scale Structure of the Universe* (Princeton: Princeton University Press).
———. 1982, *Ap. J.*, **258**, 415.
———. 1984, *Ap. J.*, **277**, 470.
Peebles, P. J. E., and Groth, E. J. 1975, *Ap. J.*, **196**, 1.
Preskill, J., Wise, M. B., and Wilczek, F. 1983, *Phys. Letters*, **120B**, 127.
Press, W. H. 1980, *Phys. Scripta*, **21**, 702.
Sargent, W. L. W., Young, P. J., Boksenberg, A., and Tytler, D. 1980, *Ap. J. Suppl.*, **42**, 41.
Shanks, T., Bean, A. J., Efstathiou, G., Ellis, R. S., and Peterson, B. A. 1983, *Ap. J.*, **274**, 529.
Shapiro, P. R., Struck-Marcell, C., and Melott, A. L. 1983, *Ap. J.*, **275**, 473.
Starobinskii, A. 1982, *Phys. Letters*, **117B**, 175.
Turner, M., and Schramm, D. 1978, *Nature*, **279**, 303.
Uson, J. M., and Wilkinson, D. T. 1984, *Ap. J. (Letters)*, **277**, L1.
Vittorio, N., and Silk, J. 1984, *Ap. J. (Letters)*, **285**, L39.
White, S. D. M., Davis, M., and Frenk, C. S. 1984, *M.N.R.A.S.*, **209**, 27P.
White, S. D. M., Frenk, C. S., and Davis, M. 1983, *Ap. J. (Letters)*, **274**, L1.
White, S. D. M., and Negroponte, J. 1982, *M.N.R.A.S.*, **201**, 401.
Weinberg, S. 1979, *Phys. Rev. Letters*, **42**, 850.
Wilson, M. L., and Silk, J. 1981, *Ap. J.*, **243**, 14.
Yang, J., Turner, M. S., Steigman, G., Schramm, D. N., and Olive, K. A. 1984, *Ap. J.*, **281**, 493.
Zel'dovich, Ya. B. 1970, *Astr. Ap.*, **5**, 84.
———. 1972, *M.N.R.A.S.*, **160**, 1P.

MARC DAVIS: Department of Astronomy, University of California, Berkeley, CA 94720

GEORGE EFSTATHIOU: Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, England

CARLOS FRENK: Astronomy Centre, Physics Building, University of Sussex, Falmer, Brighton BN1 9QH, England

SIMON WHITE: Steward Observatory, University of Arizona, Tucson, AZ 85721

THE ASTROPHYSICAL JOURNAL, 313:505–516, 1987 February 15
© 1987. The American Astronomical Society. All rights reserved. Printed in U.S.A.

CLUSTERS, FILAMENTS, AND VOIDS IN A UNIVERSE DOMINATED BY COLD DARK MATTER

SIMON D. M. WHITE
Steward Observatory, University of Arizona

CARLOS S. FRENK
Physics Department, University of Durham

MARC DAVIS
Astronomy Department, University of California, Berkeley

AND

GEORGE EFSTATHIOU
Institute of Astronomy, Cambridge
Received 1986 July 1; accepted 1986 August 6

ABSTRACT

We present new N-body results concerning the behavior of a flat cold dark matter universe on scales larger than 10 Mpc. We adopt the biasing model which we previously found to provide a good match to observations of galaxy clustering on small and intermediate scales. This model assumes that galaxy formation occurs only near high peaks of the filtered linear density field. There are no further free parameters in the present investigation which focuses on the properties of rich clusters, voids, filaments and other large-scale features. Our model produces the observed abundance of Abell clusters of richness class greater than 0 and is consistent with estimates of the abundance of richer systems. The apparent mass-to-light ratios of these clusters are similar to those found for real clusters. However, their correlation length, while 5 times that of the mass distribution, is only twice that of the galaxies and so seems to be significantly smaller than that inferred for Abell clusters. The three-dimensional rms peculiar velocities of clusters in the model are ~ 150 km s^{-1}, but the peculiar velocities of galaxies are comparable to our velocity relative to the microwave background. The effective bias in regions the size of the Local Supercluster is sufficient to make our observed Virgocentric infall consistent with $\Omega = 1$. In our model voids approaching the size of the one in Bootes are readily formed, and deep pencil-beam surveys often show large empty redshift intervals. Projections of artificial magnitude-limited catalogs are quite similar to the Zwicky catalog, and wedge diagrams in redshift space resemble equivalent representations of real data. Long, curving filaments with sharp edges are often seen in such diagrams. Taken with previous work, these results not only provide strong support for the cold dark matter model, they also demonstrate that present observations of the galaxy distribution do not require an open universe, initial conditions with nonrandom phases, pregalactic explosions, or other large-scale nongravitational processes.

Subject headings: cosmology — early universe — galaxies: clustering

I. INTRODUCTION

A universe dominated by cold dark matter (CDM) has attracted considerable attention in recent years. The dynamics of such a universe are determined by collisionless elementary particles which interact only through gravity and which had a low peculiar velocity dispersion at early times. Candidates for such dark matter include photinos, gravitinos, axions, and leptons with GeV masses. During the radiation-dominated era the growth of fluctuations in the dark matter is inhibited once they enter the horizon. As a result a characteristic scale is imposed on the fluctuation spectrum, associated with the horizon size when dark matter became dominant. This scale is proportional to $(\Omega H^2)^{-1}$, where Ω is the cosmological density parameter and H is the Hubble constant. (Hereafter, $H = 100h$ km s^{-1} Mpc^{-1}.) In this model primordial fluctuations are normally assumed to be adiabatic with a constant curvature power spectrum. The post-(re)combination spectrum then preserves its original shape on large scales but has relatively less power on small scales. The limiting behavior corresponds to a power-law form for the spectral density, $\delta_k^2 \propto k^n$, with $n = 1$ and $n = -3$ on large and small scales, respectively. In a cold dark matter universe, nonlinear structures build up from small to large scales in a roughly hierarchical way. This model is attractive because the growth of structure in the mass distribution is driven by gravity alone, and proceeds from initial conditions which are known *a priori*. Reviews may be found in Blumenthal *et al.* (1984) and White (1986).

In a previous paper we used N-body simulations to follow the nonlinear evolution of both open and flat dark matter universes on scales $1–10h^{-1}$ Mpc (Davis *et al.* 1985, hereafter Paper I). We found that structures form very rapidly with a wide range of scales collapsing almost simultaneously. The mass correlations acquire a power-law form which steepens with time. Identifying the simulated mass distribution with the observed galaxy distribution, we found models with $\Omega = 0.2$ to agree reasonably well with observation, as judged by their two-point correlation functions, by the peculiar relative velocities of galaxy pairs, and by their overall morphology. However, these models produced overly strong three-point correlations and are marginally ruled out by current limits on the anisotropy of the microwave background (Bond and Efstathiou 1984; Vittorio and Silk 1984). A similar identification for the flat ($\Omega = 1$)

model led to mass and velocity correlations in serious disagreement with observation. The identification of the mass and galaxy distributions has, at best, only the weak justification of simplicity. The alternative assumption, that galaxies form at high peaks of the smoothed linear density field, leads to a bias in the galaxy distribution favoring regions of high dark matter density. We found that such a biased cold dark matter model could agree well with the known properties of the galaxy distribution on megaparsec scales even for $\Omega = 1$. Our biasing scheme was parametrised by the width of the low-pass filter that modulates the background matter distribution and the threshold height for galaxy formation. This recipe is similar to that invoked by Kaiser (1984) to explain the large correlations of Abell clusters. It has also been discussed in connection with galaxy formation by Kaiser (1986), Bardeen (1986), Bardeen *et al.* (1986), and Schaeffer and Silk (1985) among others. Several feed back mechanisms which might provide a bias of this kind have been discussed in the context of the cold dark matter cosmogony (Rees 1985; Silk 1985), but no compelling evidence in favor of any specific scheme has been found. Ongoing research in our own program suggests that such mechanisms may not be necessary, and that a bias of the type required may arise automatically from the nonlinear evolution of the dark matter distribution.

A particularly attractive feature of our flat cold dark matter model is that it is specified by only four parameters. These are the two biasing parameters just discussed, and the amplitude and characteristic length scale of the fluctuation spectrum. The latter is determined uniquely by the value of H and so cannot be varied freely. The fluctuation amplitude is, in principle, set by the physics of the early Universe while the bias parameters are likely to depend on complex astrophysical phenomena. All four parameters can be determined empirically by matching the observed mean space density of galaxies and their position and velocity correlations. The success of the models of Paper I consisted in showing that a set of parameters which was plausible *a priori* could produce sufficient bias to match observations successfully. Once parameter values have been fixed to fit the galaxy distribution on Mpc scales, there is no further freedom in the model. It is then natural to ask whether it can also successfully account for observations on other scales. In a subsequent paper we completed a preliminary study of the behavior of a cold dark matter universe on much smaller scales, comparable to those of luminous galaxies and their halos (Frenk *et al.* 1985). We found that our standard model produces the observed abundance of clumps with the binding energies and structure inferred for galaxy halos. Galaxy formation is predicted to be a recent but protracted process, starting at a redshift $z \lesssim 3$. Recent reviews of our work are given by Frenk (1986) and by White (1986).

In the present paper we investigate the large-scale properties of our standard model. In particular we focus on the abundance and properties of rich clusters and on the overall topology of the galaxy distribution. Observational data on these large scales are less abundant and systematic than those on small and intermediate scales. The amplitude and even the sign of the galaxy correlation function is uncertain beyond $10h^{-1}$ Mpc (Davis and Peebles 1983; Bean *et al.* 1983; Stevenson *et al.* 1985). The abundance, luminosities, and masses of rich Abell clusters are fairly well determined, and their spatial correlation function is reported to have an amplitude ~ 18 times that of the galaxies (Bahcall and Soneira 1983; Klypin and Kopylov 1983). This result is based on a sample of only ~ 100 clusters, and its uncertainty may be large (Ling, Frenk, and Barrow 1986). Evidence for large voids and filaments in the galaxy distribution is still somewhat anecdotal. Long quasi-linear superclusters of galaxies and large underdense regions are clearly visible in projection in maps constructed from the Lick and Zwicky catalogs (Seldner *et al.* 1977; Fall 1979), but existing redshift samples survey only relatively small volumes of space. The largest known filamentary structure is the Perseus-Pisces supercluster which extends at least $50h^{-1}$ Mpc (Gregory, Thompson, and Tifft 1981; Oort 1983; Haynes and Giovanelli 1986). The largest void discovered so far is a sphere of radius $31h^{-1}$ Mpc in the constellation of Bootes which is reported to contain less than 25% of the mean density of bright galaxies (Kirshner *et al.* 1981, 1983). Large gaps appear in the velocity distribution in deep redshift surveys of small areas of the sky (Koo, Kron, and Szalay 1986; Broadhurst, Ellis, and Shanks 1986). On the basis of an extension of the CfA redshift survey (Davis *et al.* 1982) to $m_{\text{Zwicky}} = 15.5$, de Lapparent, Geller, and Huchra (1986) claim to have detected "striking... surfaces of bubble-like structures" of typical diameter $25h^{-1}$ Mpc.

The remainder of this paper is organized as follows. In § II we describe a new set of large-scale N-body simulations and a new algorithm to impose the bias required to define the galaxy distribution. We use these techniques to find "Abell clusters" in the models and to construct both magnitude-limited and volume-limited "galaxy" catalogs. The properties of the predicted clusters, voids and filaments are discussed in §§ III–V, respectively. We conclude in § VI with a summary of our results. Because our standard model requires $h = 0.5$ to fit the data, we shall adopt this value throughout the present paper.

II. METHODS

a) *The Simulations*

We have run simulations of comoving cubic boxes of present size $L = 280$ Mpc and $L = 360$ Mpc. The initial conditions were generated by perturbing particle positions and velocities from a comoving cubic lattice as described by Efstathiou *et al.* (1985, hereafter EDFW). For the smaller box size we ran a set of 25 calculations each containing 32,768 particles; we refer to this ensemble as E28. Such a large number of realizations was required to obtain good enough statistics for correlation studies of clusters. For the larger box size we ran a single simulation with 216,000 particles; we refer to this as C36. The larger particle density of this model was necessary for us to obtain a high enough sampling frequency to be able to generate the magnitude limited "galaxy" catalogs used in §§ IV and V. All simulations were performed with the P^3M code described by EDFW using a 64^3 mesh and softening parameter $\eta = 0.2$. In all cases we assume $\Omega = 1$. Expansion factors were adjusted to correspond to the epoch which we identified as best matching the present day in the biased models of Paper I ($a = 1.4$ in ensemble EdS1-5). The initial amplitude of the input spectrum was reduced by 0.5 in E28 and by 0.57 in C36 in order for the initial time to be early enough that initial transients do not affect the results. The expansion factors required to reach the present day then turn out to be 1.86 for E28 and 2.0 for C36. The mass represented by each particle is 4.6×10^{13} $M_\odot$ for E28 and 1.5×10^{13} $M_\odot$ for C36. Figure 1*a* shows power spectra of the initial conditions in E28 and in C36; the solid line represents the input theoretical power spectrum. For frequencies below the Nyquist frequency of the initial particle

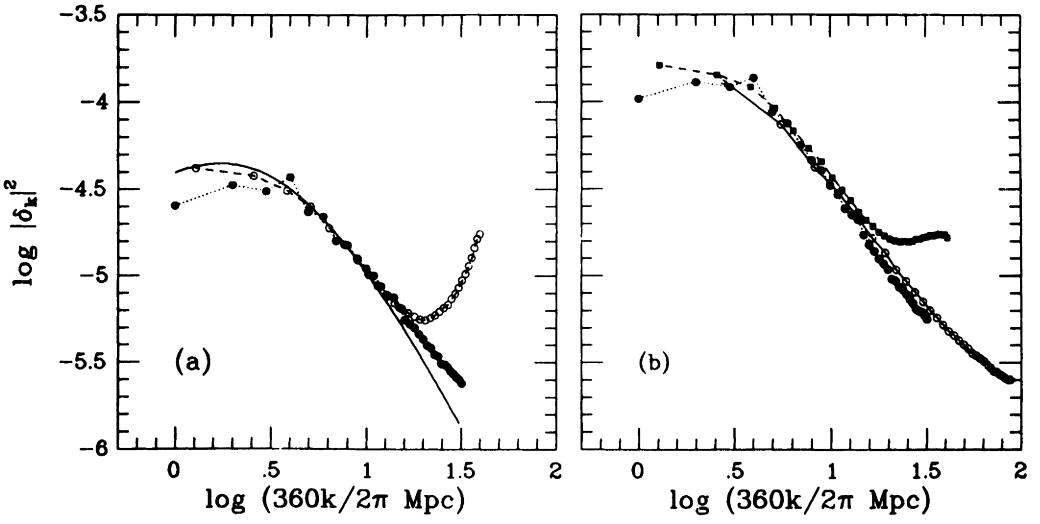


FIG. 1.—Power spectra of the initial conditions in the ensemble E28 (*open circles*) and in the single model C36 (*filled circles*) are compared to the theoretical linear power spectrum from Bardeen *et al.* (1986) (*solid curve*) in (*a*). In (*b*) the spectra of the evolved models (squares for E28, filled circles for C36) are compared with those of ensemble EdS1-5 of Paper I (*open circles*). These power spectra were calculated as described in Paper I.

grid, the computed power spectra are very close to the expected values, in agreement with the results of Paper I. The deviations at low wavenumber for C36 are a result of random statistical fluctuations. Figure 1*b* shows power spectra for the same simulations at the final time overlaid on those of the "standard" models of Paper I. Once again there is good agreement below the Nyquist frequency. The mass distributions in these new calculations thus extend those of our earlier work correctly.

b) The Prescription for Biased Galaxy Formation

In Paper I we used a simple recipe which identified the high peaks of the smoothed linear density field as the sites of galaxy formation. We applied a Gaussian low-pass filter to the initial density field of the models, and found all peaks rising above 2.5 times the rms density fluctuation. We then tagged the particle whose Lagrangian position was closest to each peak as a "galaxy," ensuring that only one such "galaxy" was associated with each connected region above the threshold. This recipe cannot be implemented in the present simulations because we are unable to resolve the mass distribution on the scales required. We circumvent this difficulty by using the analytic machinery which Bardeen *et al.* (1986, hereafter BBKS) developed to study the statistics of peaks in Gaussian random fields. Their results allow us to calculate the expected number of high peaks (and thus "galaxies") associated with each simulation particle. This number is a nonlinear function of the density field smoothed on a large scale which we can resolve.

The procedure works as follows. Let F be a Gaussian random field which describes the linear density fluctuations in the cold dark matter. When F is smoothed on the galactic scale, r_s, it produces a new field F_s with rms value σ_s. Peaks of F_s higher than 2.5 σ_s are to be identified as the sites of galaxy formation, but F_s is not resolved by our numerical models. Now let F be smoothed on the larger scale, r_b, to produce a field F_b (with rms value σ_b) which we can resolve. In their Appendix E BBKS derive an exact formula for the quantity $\mathcal{N}_{\rm pk}(\nu_s|\nu_b)d\nu_s$, the number density of peaks of F_s with height between $\nu_s\sigma_s$ and $(\nu_s + d\nu_s)\sigma_s$ in the vicinity of a point where F_b has the value $\nu_b\sigma_b$. If we imagine representing F_b as a contour plot, this formula allows us to populate the volume between neighboring contours with a uniform density of "peaks." By construction, the relative number of "peaks" at different contour levels must then be correct. However, as BBKS point out, the correlation properties of this "peak" distribution may not correspond to those of the peaks of F_s, even on scales much larger than r_b. This is because the difference field $F_s - F_b$ in general has substantial correlations with F_b. This unpleasant property can be avoided if the smoothing on scale r_b is chosen to correspond to a step function in k-space; the Gaussian field $F_s - F_b$ is then uncorrelated with F_b, and the correlation properties of the "peak" distribution do agree asymptotically with those of the real peaks of F_s.

To implement this machinery, we constructed smoothed linear density fields for each of our simulations. The smoothing was accomplished by removing all power at frequencies above $k_{\rm crit}$ from the density field used to generate the initial conditions. We adopted $k_{\rm crit} = 16\ 2\pi/280$ Mpc, the Nyquist frequency of the particle grid in E28; we have checked that, as expected, our final results are very insensitive to this choice. After tabulating the smoothed density on a 64^3 mesh, we used the formulae of BBKS to find the expected number of "galaxy" peaks associated with each mesh cell. "Galaxy" peaks were assumed to be 2.5 σ peaks of the density field smoothed with a Gaussian of width $r_s = 1.08$ Mpc, corresponding to a smoothing mass of $1.4 \times 10^{12}\ M_\odot$. This threshold agrees with that adopted in Paper I. The value of r_s was chosen to give the same density of peaks as we found before. (It is larger than the value of r_s quoted in Paper I because the 64^3 density grids used by our original peak-finding technique were too coarse to resolve F_s fully; as a result we missed a significant number of peaks.) With our new procedure we find an average of 10,400 peaks in each 280 Mpc model and about 22,000 peaks in C36. These numbers agree very well with those predicted by the exact formulae of BBKS. In what follows we do not try to identify these peaks as individual bright galaxies.

Rather we find the expected number of peaks in the initial volume occupied by each particle. We then take this peak number to be proportional to the total luminosity associated with that particle. The proportionality constant is set by requiring the mean luminosity density of the model to equal the observed value, $0.0019L_*/\mathrm{Mpc}^3$ with $L_* = 3.9 \times 10^{10}\ L_\odot$. The mean luminosity associated with each "peak" is then $4.0L_*$, and the mean luminosity associated with each particle is $1.3L_*$ in E28 and $0.41L_*$ in C36. Finally we assume that the probability of finding a galaxy with luminosity L associated with a particle is given by the universal luminosity function,

$$\phi(L) \propto L^{-1.25} \exp(-L/L_*),$$

normalized to produce the correct total luminosity. These luminosity function and luminosity density parameters are taken from the recent review of Felten (1985); L_* has been corrected to the value expected on the magnitude system of the Zwicky catalog when high latitude extinction is assumed negligible. Our motivation for this procedure is as follows. Galaxies of different brightness undoubtedly correspond to peaks of different mass in the linear dark matter distribution. However, while peak mass and peak number density depend strongly on r_s, the effective bias depends only weakly on it (see BBKS). Thus it seems reasonable to calculate the bias using peak parameters appropriate to the halo of a bright galaxy but to set the mean galaxy density to its observed value.

Figure 2 shows "galaxy" correlation functions for several models. The filled squares are the results presented for our biased models in Paper I. The open triangles were produced by applying our new procedure to these same models; these correlations were calculated using all particles and weighting each pair by the product of the associated peak numbers. The agreement of the two curves shows that our new biasing algorithm gives very similar results to the simpler one used in Paper I. The new scheme gives slightly stronger correlations where $100 > \xi > 1$, but the difference never exceeds 0.1 in $\log(\xi)$. The filled circles in Figure 2 give the ensemble mean "galaxy" correlation function for E28. For separations larger than our softening length, these data agree well with the correlations found for the small box models. This again indicates that our new simulations are correctly extending our earlier results. It is amusing, but probably not significant, that our new biasing scheme leads to a "galaxy" correlation function in E28 which is an even better approximation to a $r^{-1.8}$ power law than that given in Paper I.

III. RICH CLUSTERS

We identify rich clusters in our simulations using a criterion patterned after the one which Abell (1958) used to find clusters

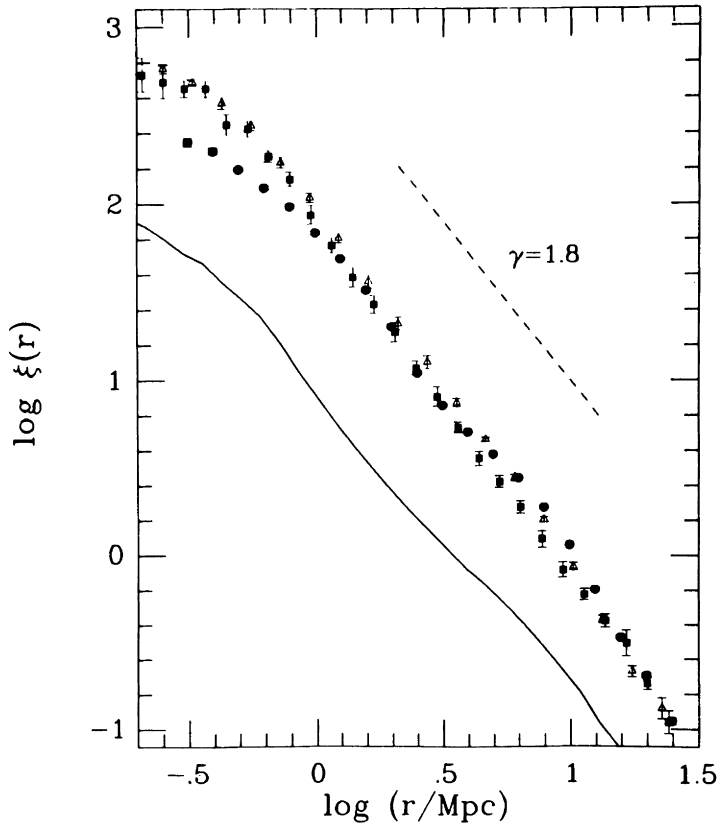


FIG. 2.—"Galaxy" correlations for a variety of biased models. Filled squares are taken from Fig. 17 of Paper I (ensemble EdS1-5). Open triangles shown the results of applying the bias algorithm of this paper to the same set of models from Paper I. Filled circles show the results found here for the final time of ensemble E28. In each case error bars give the error in the mean derived from the scatter within the ensemble. The solid line gives the *mass* correlations in ensemble EdS1-5.

in the Palomar Sky Survey. For inclusion in his catalog, Abell required a cluster to have at least n_a galaxies within projected radius $r_a = 3$ Mpc, and brighter than $m_3 + 2$, where m_3 is the apparent magnitude of the third brightest galaxy. For richness class 1, $n_a = 50$ and for richness class 2, $n_a = 80$. If we assume a Schechter luminosity function for the galaxies within r_a,

$$\phi(L) = N_* L^{-1.25} \exp(-L)/\Gamma(0.75),$$

where L is in units of L_* and $N_* L_*$ is the total luminosity projected within r_a, we find Abell's criterion to require $N_* > 55$ for richness class 1 and $N_* > 112$ for richness class 2. For typical cluster luminosity profiles (Seldner and Peebles 1977), these values correspond to luminosities of $42L_*$ and $86L_*$, respectively, within spheres of radius r_a centered on the clusters. For comparison, Oemler's (1974) photometry of the Coma cluster gives a total luminosity of $235L_*$, a luminosity projected within r_a of about $160L_*$, and an inferred luminosity of about $120L_*$ within a 3 Mpc sphere.

To find "Abell clusters" in the simulations, we located high-density clumps using the group-finding algorithm of Paper I with linking parameter $b = 0.2$. We then found the total luminosity within 3 Mpc of the barycenter of each clump and identified the clump as a rich cluster if this luminosity exceeded $42L_*$. With this definition we found a mean of 17.6 rich clusters in each simulation of E28. The observed abundance of Abell clusters with richness class greater than 0 is 7.5×10^{-7} Mpc^{-3} (Bahcall and Soneira 1983); thus the expected number in a volume the size of these calculations is 16.5. In C36 we found 27 clusters whereas 35 were expected. This agreement is remarkable since our model has no adjustable parameters. It appears that a flat cold dark matter model can reproduce the observed abundance of virialized structures spanning the mass range between galaxy halos and rich galaxy clusters.

The luminosity function of the rich clusters in E28 is displayed in Figure 3. It is steep but has a long tail toward high luminosities; the brightest cluster that we find has $200L_*$ within 3 Mpc of its center. No comparable data set exists for the real universe at present, but it may become available in the near future when current photometric and spectroscopic

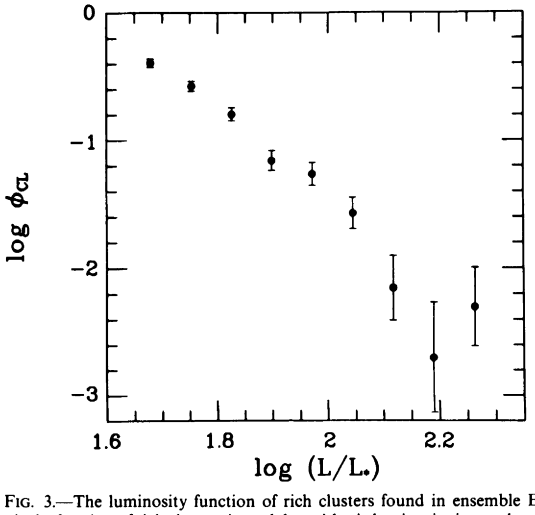


FIG. 3.—The luminosity function of rich clusters found in ensemble E28. Φ_{CL} is the fraction of rich clusters in each logarithmic luminosity interval.

galaxy surveys are completed. In the interim we can compare our results with Abell's richness data. In Bahcall and Soneira's (1983) sample the space density of clusters with richness class greater than 1 is 16% of that of the sample as a whole. In our simulations 10% of all clusters have luminosities putting them in this class. Since these estimates are based on 22 clusters in the observational sample and 42 clusters in E28, the discrepancy is not statistically significant. The mean mass, luminosity, and one-dimensional velocity dispersion of the inner 3 Mpc of our clusters are $9.6 \times 10^{14}\ M_\odot$, $60.5L_*$, and 550 km s^{-1}; the mean mass and luminosity within 1.5 Mpc are $6.0 \times 10^{14}\ M_\odot$ and $39.7L_*$. These masses and luminosities are typical of rich clusters in the real universe; in particular, the mean luminosity overdensity of 280 within r_a which we find for our clusters is in good agreement with the value of 289 inferred from the cluster-galaxy cross-correlation of Seldner and Peebles (1977). However, the velocity dispersion we find appears to be about 30% too low. This is probably a result of the rather poor mass resolution of our calculations which prevents them from modeling adequately the centrally concentrated mass distribution of real clusters. In addition, our Poisson solver produces softened forces on scales $\lesssim 0.5$ Mpc, leading to an artificial reduction in measured velocity dispersions. Within our clusters we find the mass-to-light ratio to be almost independent of radius despite the overall segregation of mass and light imposed by the bias. The mass-to-light ratios and velocity dispersions of the clusters are plotted against luminosity in Figure 4, where we use the same luminosity bins as in Figure 3. Note that M/L decreases slowly with L, and as a result σ increases more slowly than $L^{1/2}$. The mean M/L for the clusters is only 20% of the value for the calculation as a whole; this accords well with the observation that the mean M/L of Abell clusters is about 20% of the value required for closure.

The mean M/L of our clusters does, of course, increase as we average their properties out to larger and larger radii. The distance of the Virgo cluster corresponds to a Hubble velocity of about 1350 km s^{-1}. The mean mass and luminosity within spheres of radius 27 Mpc centered on our clusters are $8 \times 10^{15}\ M_\odot$ and $340L_*$. These correspond to mean relative overdensities of 0.43 and 1.21 in mass and light, respectively, and to a mean M/L which is 1.55 smaller than that of the calculations as a whole. The Local Supercluster is thought to have a relative luminosity overdensity of about 2 (Davis and Huchra 1982); with a bias factor of 1.55 this corresponds to a relative mass overdensity of 0.94. In the presence of such a bias the standard analysis assuming that mass follows light would lead to an underestimate of Ω by a factor of 3.5. Thus the values, $\Omega = 0.3$, inferred from our infall toward Virgo are exactly what would be expected in a flat cold dark matter universe with a bias of this kind.

The mean overdensities we find in large regions surrounding clusters seem somewhat low. The cross-correlation between galaxies and Abell clusters obtained by Seldner and Peebles (1977) predicts a mean luminosity overdensity of 3.1 within 27 Mpc of a cluster, a clustering amplitude 2.5 times that we have found. This problem is related to the most serious discrepancy we find between our models and existing data. In Figure 5 we plot autocorrelation functions both for the galaxies and for the rich clusters in E28. The correlation length of the clusters (i.e., the separation at which ξ is 1) is $r_{0,cc} = 22$ Mpc, about twice that of the galaxies ($r_{0,gg} = 10.6$ Mpc). However, it is a factor of ~ 2.2 smaller than that obtained by Bahcall and Soneira (1983) and Klypin and Kopylov (1983) from a sample of about ~ 100

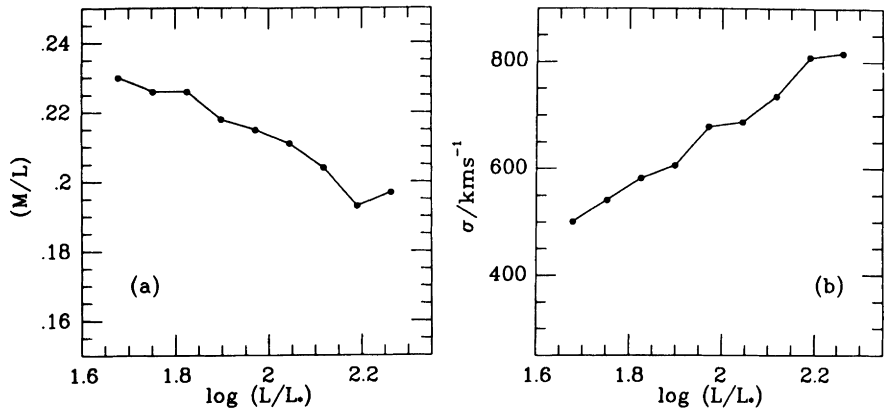


FIG. 4.—The mean mass-to-light ratio and the mean velocity dispersion of rich clusters are plotted against their luminosity in (*a*) and (*b*), respectively. M/L is given in units of the value for the universe as a whole.

Abell clusters with measured redshifts. For such a small sample, statistical uncertainties are expected to be large; a bootstrap analysis of these data and of the CfA survey by Ling, Frenk, and Barrow (1986) gives a probable error ± 5.5 Mpc for $r_{0,cc}$ and a best estimate of 4.3 ± 1.1 for the ratio $r_{0,cc}/r_{0,gg}$. This should be compared with the value 2.1 found in our simulations. We show the data of Bahcall and Soneira in Figure 5 for comparison with the models. Although there is actually moderately good agreement on small scales, there appears to be a significant deficit of pair counts in the models on scales of 15–50 Mpc. From Monte Carlo simulations of cluster samples with the selection characteristics of the Bahcall and Soneira data and the mean correlation properties of our models, we estimate that sampling fluctuations will lead to an apparent clustering strength exceeding that of the real data with a probability of roughly 10%. Thus the cluster correlations in the data appear inconsistent with our model. Both the cluster-galaxy cross-correlation and the cluster-cluster correlation are based on the Abell catalog which its author did not intend to be used for such statistical studies; they may perhaps be affected by systematic trends in the catalog. Larger and more homogeneous cluster samples are clearly very desirable in order to confirm these observational results.

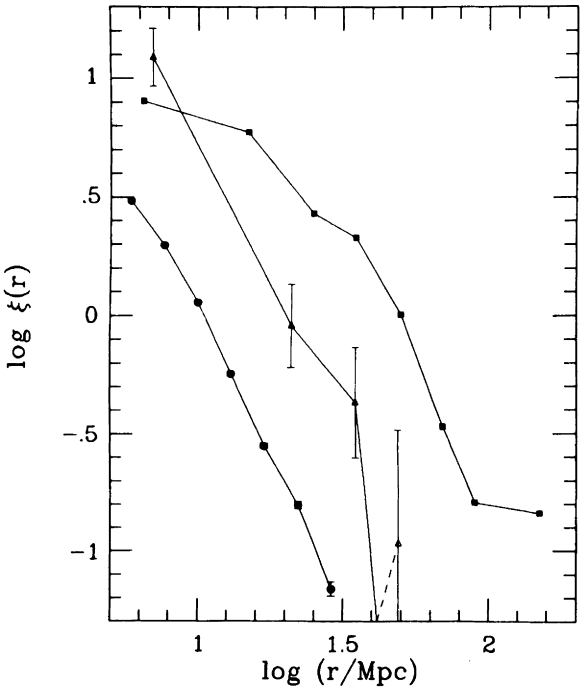


FIG. 5.—The two-point correlation function for the 440 rich clusters in ensemble E28 (*triangles*). For the last data point ξ is negative. Error bars are calculated from the scatter in pair counts among the 25 members of the ensemble. The "galaxy" correlation function from Fig. 2 is also shown (*circles*) as is the cluster-cluster correlation estimate of Bahcall and Soneira (1983) (*squares*).

We have also calculated the rms peculiar velocity of clusters and the rms relative peculiar velocity of cluster pairs in our simulations; these are 145 km s^{-1} and 511 km s^{-1} respectively (in three dimensions). The first value is in agreement with the upper limits derived by Aaronson *et al.* (1986). However, the second, which refers to all pairs closer than 56 Mpc, is far too small to be consistent with the velocity broadening of ~ 2000 km s^{-1} which Bahcall and Soneira (1983) claim to detect in their study of correlations. The low peculiar velocities that we find both for galaxies and for clusters are inconsistent with the large-scale streaming motions reported recently by Collins, Joseph, and Robertson (1986) and by Burstein *et al.* (1986). However, the rms peculiar velocity of "galaxies" in E28 is 490 km s^{-1}, quite comparable to our observed velocity through the microwave background.

IV. VOIDS

In this section we first consider the distribution of voids in volume-limited catalogs constructed from our simulations. We fix the galaxy density in these catalogs to be the density expected at the median distance in a magnitude-limited survey. This has the advantage that the clustering distribution is independent of position within our catalogs but will be similar to that in the part of real galaxy surveys where most of the data lie. As a descriptive statistic for the voids we use the distribution of the size of the largest empty sphere surrounding a randomly chosen point. In addition we implement a sampling procedure similar to that followed by Kirshner *et al.* (1981,

1983, hereafter together KOSS) in their investigation of the large void in Bootes, and we look for similar voids in our models. Finally we study voids in deep pencil-beam redshift surveys.

For a luminosity function of the form $\phi = \phi_* L^{-1.25} e^{-L}$, the expected galaxy density at the median distance in a magnitude-limited survey is $0.59\phi_*$. Thus, using the value $\phi_* = 0.0015$ Mpc^{-3} recommended by Felten (1985), we require 19,430 galaxies in each calculation in E28. We constructed volume-limited catalogs at this density by assigning a "galaxy number" to each particle which is 1.87 times the peak number defined in § Ib; 19,430 "galaxies" are then predicted in each model. We interpreted this new galaxy number as the expected number of galaxies associated with each particle. By comparing it to a random number drawn uniformly from the interval [0, 1], we decided whether the particle should be included in the catalog. This procedure fails for a number of particles in dense regions which have galaxy numbers greater than one; these should, in principle, appear in the catalog more than once on average. Since such multiple inclusion would clearly not change the distribution of empty regions, we ignored this problem. As a result of it our catalogs on average contain 14,592 rather than 19,430 galaxies.

To investigate the sizes of spherical voids in these catalogs we calculated the distribution of nearest neighbors for randomly chosen points, $N_s(r)dr$. This distribution is shown in Figure 6*a*; it cuts off quite sharply at large radii indicating that the biggest voids in each simulation have very similar sizes. The largest void in the whole ensemble has a diameter of 61 Mpc and almost all the simulations have voids over 56 Mpc across. These values are significantly greater than those we inferred in Paper I by counting particles in fixed cells; as has been noted by Politzer and Preskill (1986) such an increase in maximum void size is to be expected if the centers of the cells can be chosen freely. The integral

$$p(r) = \int_r^\infty N_s(r)dr$$

gives the probability that a randomly chosen sphere of radius r is empty. For a Poisson distribution of points with mean density n this probability is $p(r) = \exp[-(4\pi/3)r^3 n]$. If the distribution is clumpy, we can define an effective number of particles per clump, $N_{\rm eff}$, through $p(r) = \exp[(4\pi/3)r^3 n/N_{\rm eff}]$. This quantity is a measure of the deviation of the frequency distribution of voids from that expected in the Poisson case. In Figure 6*b* we plot the values of $N_{\rm eff}$ obtained from the cumulative distribution of $N_s(r)$ from Figure 6*a*. The effective clumping of particles becomes greater as we consider rarer and rarer voids of ever increasing size.

In order to test whether voids as large as that discovered by KOSS can be found in our simulations we attempted to follow their observational procedure quite closely. We took the catalogs described above and inserted imaginary parallel skewers through them on a square mesh with a separation of 8 Mpc, the mean separation of the skewers used by KOSS at the center of the Bootes void (A. Oemler 1986, private communication). We then constructed new catalogs containing all "galaxies" within distance r of a skewer; r was chosen so that these new catalogs would, on average, contain 31 "galaxies" in a sphere 124 Mpc in diameter. This is the number of galaxies expected by KOSS in the region of the Bootes void. We then calculated $N_s(r)dr$ as above. This distribution and the corresponding values of $N_{\rm eff}$ are shown with dashed lines in Figure 6. The largest void in E28 has a diameter of 113 Mpc and three voids in this ensemble have diameters larger than 112 Mpc. Thus, voids very close in extent to the one in Bootes are readily produced in the flat cold dark matter model. The values of $N_{\rm eff}$ in Figure 6 show that even with such sparse sampling, the distribution of large voids deviates significantly from that expected in an uncorrelated distribution.

Striking voids have also been seen in deep pencil-beam redshift surveys (Kirshner *et al.* 1981; Koo, Kron, and Szalay 1986; Broadhurst, Ellis, and Shanks 1986). In the latter studies redshifts were obtained for galaxies in several fields about 0°.5 in diameter. The samples are complete to about $B = 21$ and become progressively more incomplete at fainter magnitudes. Each field typically has redshifts with a median of about 0.2. Histograms of the redshift distribution vary considerably; some show little structure, whereas others show large peaks separated by low density regions up to 20,000 km s^{-1} across. We have made such magnitude-limited pencil beam surveys of our large model, C36. We pick a vantage point within the simulation volume and look out at a particular area of the "sky." For each particle whose position, either in the funda-

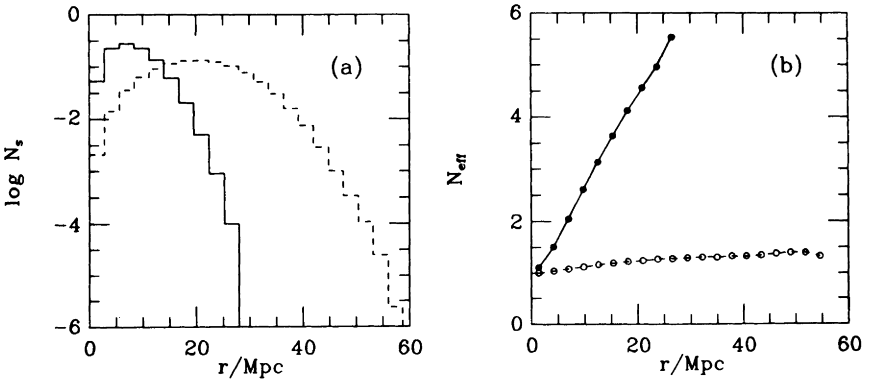


FIG. 6.—The distribution of the distance to the nearest "galaxy" for randomly chosen points within ensemble E28 is shown in (*a*) by the solid histogram. The dashed histogram gives results for catalogs in which "galaxies" are sampled only along skewers in the manner adopted by KOSS. Panel (*b*) gives values of the effective number of "galaxies" per clump derived from these counts; filled and open circles correspond to the solid and dashed histograms in (*a*).

mental cube or in one of its periodic replications, is projected on this area, we use the local luminosity function discussed in § II*b* to determine the expected number of "galaxies" above the apparent magnitude limit. This number is compared with a random number drawn uniformly from [0, 1] to decide if the particle should be included in the survey. For a few particles in nearby clumps, this procedure breaks down because the expected galaxy number exceeds unity. When compiling redshift histograms, we increment the relevant bins for such particles (which are always included in the survey) by the expected galaxy number rather than by unity.

Figure 7 shows redshift histograms for four of these artificial surveys, each taken to $B = 20.2$ over a 15′ radius circular field; these parameters lead to an average count and median redshift similar to those found observationally (the discrepancy in limiting magnitude is probably a result of our neglect of cosmological, evolutionary, and K-corrections). The directions were chosen arbitrarily and the redshift range plotted embraces five or six replications of the fundamental cube. However, as a result of the narrowness of the beam, individual structures appear only once in each plot. There is a lot of variety in these histograms. The total number of galaxies found ranges between 84 and 189. Occasional massive structures are encountered in most directions with velocity widths of 1000–2000 km s^{-1}. They can appear singly or at apparently regular separations of a few tens of thousands of kilometers per second. Some directions show few galaxies and rather little structure. In others, large redshift intervals containing very few galaxies can be found. This diversity of structure is a result of the small transverse dimension of the surveys and of the strong small-scale clustering of "galaxies"; it includes the range of structure seen in the real data.

V. BUBBLES, FILAMENTS, AND OTHER FROTH

Much recent observational work on large-scale structure has focused on the qualitative morphology of the galaxy distribution and its possible implications for theories of cosmogony. Thus by 1980 Einasto, Joeveer, and Saar (1980), appraising a large but heterogeneous data base, felt able to conclude: "By far the most important property of the galaxy distribution is the presence of large holes in space . . . . The vast majority of galaxies belong to cluster chains or lie in their vicinity . . . . The system of cluster chains resembles a string polyhedron . . . . It seems reasonable to associate superclusters with polyhedron walls . . . . This distribution resembles cells . . . . The structural data available prefer the adiabatic scenario for galaxy formation." Two years later Davis *et al.* (1982) commented on structure in the newly completed CfA redshift survey: "The distribution of galaxies is frothy, characterized by large filamentary superclusters and large holes devoid of galaxies . . . . Our results present a severe challenge to all theories of galaxy and cluster formation." Most recently, de Lapparent, Geller, and Huchra (1986) describe the distribution they find in a deeper survey of a narrow declination band in the northern hemisphere as follows: "The distribution of galaxies looks like a slice through the suds in the kitchen sink; it appears that the galaxies are on the surface of bubble-like structures . . . . The best available model for generating these structures is the explosive galaxy formation theory." We feel that it is dangerous to use such subjective appreciations of the morphology of

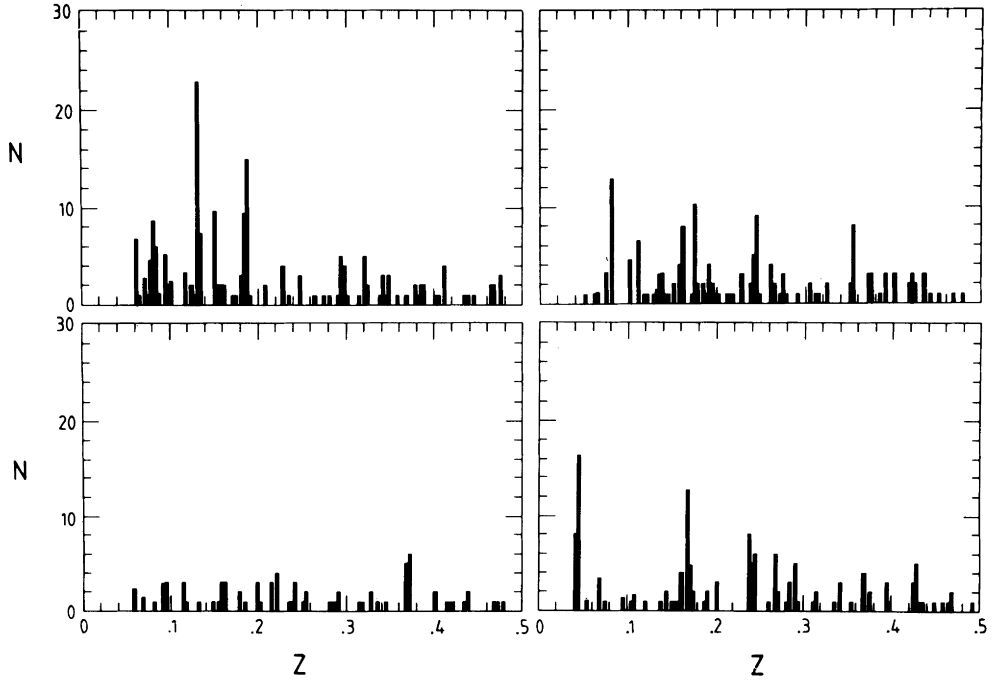


FIG. 7.—Histograms of the galaxy distribution in redshift for four magnitude-limited "pencil-beam" surveys of model C36. Number of galaxies is plotted against redshift. The redshift bins are 1000 km s^{-1} wide.

clustering to draw far-reaching conclusions about cosmogony, particularly when the theories of galaxy formation considered are too complex for the morphology in a galaxy catalog to be predicted with any confidence. In Paper I we showed by constructing artificial "CfA" catalogs that the morphology of cold dark matter models matches the frothiness of the real data quite well. In the present section we use our large C36 model to construct "magnitude-limited" catalogs to $m_{\text{Zwicky}} = 15.5$. We can then investigate how filamentary or "sudsy" our models appear when viewed in the same manner as the real data.

We construct these catalogs in a similar way to the pencil beam surveys discussed in the last section. However, because they are much shallower we do not need to consider replications outside the fundamental cube of our model. As before, we have problems in a few nearby, dense clumps; we end up with fewer galaxies than we should in such clumps because there are not enough particles to represent all the galaxies they are predicted to contain. In Figure 8 we show an equal area plot of the distribution of the 9785 galaxies with $m_{\text{Zwicky}} < 15.5$ in the North Polar Cap, $b > 40°$, $\delta > -2°.5$. We compare this with a simulated catalog which turned out to contain 8800 "galaxies" (after correction for 1200 "galaxies" missed in dense clumps). The surface density of this catalog is 25% lower than that of the Zwicky catalog; this appears to be a result of the presence of the Local Supercluster in the foreground of the real data. The morphology of the two plots appears quite similar to our prejudiced eyes. In the model we chose our vantage point so that the richest cluster in the simulation was 140 Mpc away towards the North Pole. This cluster is about a factor of 2 less rich than Coma so that its effect is not as impressive. Nevertheless, filamentary structures like the Coma supercluster can be seen in the artificial plot.

To compare our models with the data of de Lapparent, Geller, and Huchra (1986) we took "Zwicky" catalogs similar to that plotted in Figure 8 and extracted a strip between declinations 26°.5 and 32°.5. We then plotted right ascension against both true and apparent distance in order to get wedge diagrams analogous to the one they present. In each case we again ensured that the richest cluster in the model was in the center of the strip at a distance of 140 Mpc. The results are shown in Figure 9 for three different catalogs. In all of our maps it is easy to find large low-density regions with diameters in the range $\sim$50–100 Mpc. The filaments bounding these regions often have quite sharp edges. The main body of our dominant cluster is never as impressive as that of the real Coma cluster. This is partly because it is only half as rich and has a velocity dispersion of only 655 km s^{-1}, and partly because we miss 20–30 of the galaxies which "should" have been associated with it as a result of the saturation effects referred to above. It is interesting to notice how some filamentary features are enhanced in the "apparent" distribution as compared to the "true" distribution while others are degraded. The first effect is a result of systematic inflow, and the second of the masking of such inflow by "fingers of God" corresponding to the motions within virialized clumps.

It is also of interest to try to follow filaments through adjacent strips to see if they appear to be sections of curved sheets. In Figure 10 we show the apparent distribution for a fourth strip together with the apparent distribution in the 6° strips immediately above and below it. Some of the filaments can be clearly traced from one strip to the next suggesting that they are, in fact, sections through sheets. Note that several of the filaments in these diagrams are of order 100 Mpc long and so are comparable in size to the Perseus-Pisces structure. Our Figures 9 and 10 can also be compared with the galaxy distributions shown by Haynes and Giovanelli (1986) although the geometries of the diagrams do not correspond exactly. The observations show a rather wide range of morphologies. Some regions appear to show a "sudsy" structure like that emphasized by de Lapparent, Geller, and Huchra, whereas others

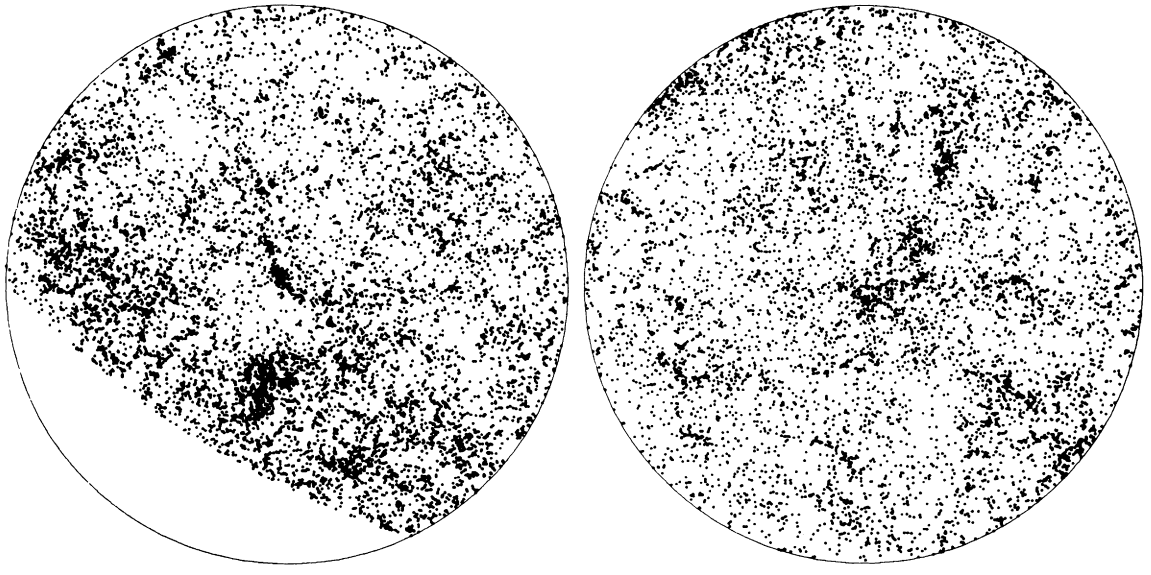


FIG. 8.—Equal area plots of the distribution of galaxies with $m_{\text{Zwicky}} < 15.5$ in the North Galactic Cap (*left*), and in a magnitude-limited survey of model C36 (*right*).

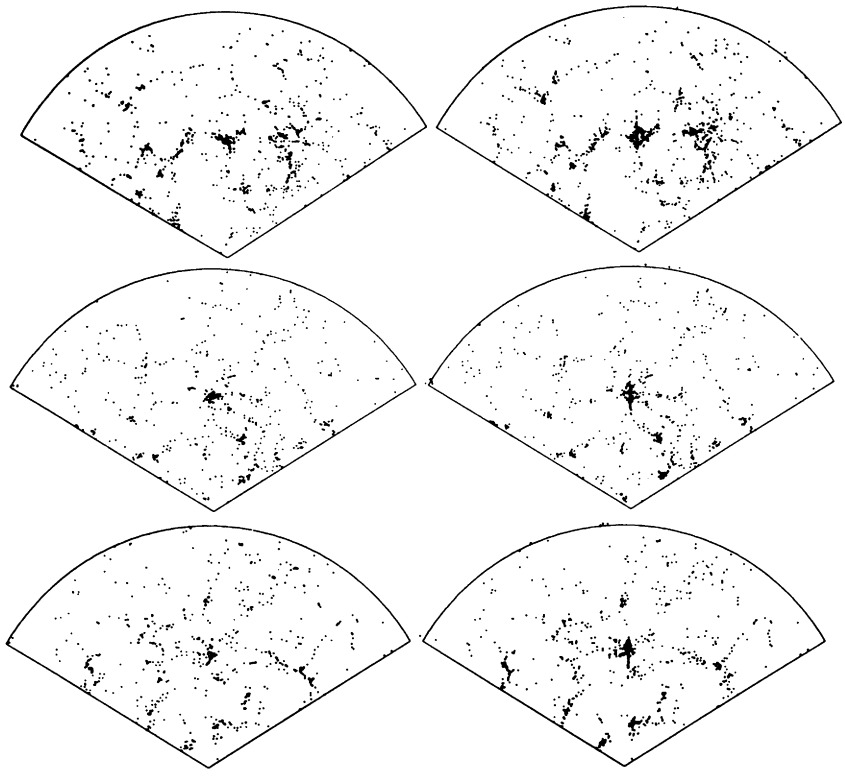


FIG. 9.—Distributions of "galaxies" for three redshift surveys of model C36 with the geometry chosen by de Lapparent, Geller, and Huchra (1966). For each survey the wedge diagram to the left uses true distance as radial coordinate, while the one to the right uses apparent distance inferred from redshift. The angular coordinate corresponds to right ascension and the maximum radial velocity plotted is 15,000 km s^{-1}.

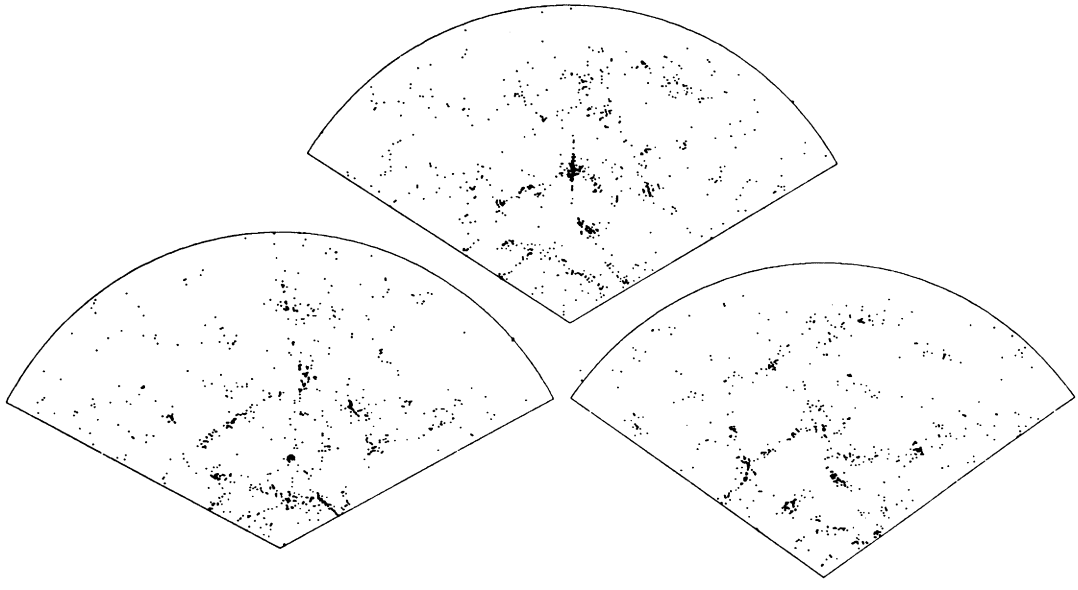


FIG. 10.—An apparent distance wedge diagram similar to those of Fig. 9 (*top*) is compared with similar surveys of the adjacent 6° declination strips (lower declination to the left, higher declination to the right).

appear less well organized. A similar range of apparent morphologies can be seen in our numerical models. The galaxy distribution in our models appears to us to have a spongy topology. Voids do not seem to be surrounded by continuous, unbroken sheets of galaxies, but rather to percolate through the computational volume. The discovery of continuous sheets surrounding voids could be evidence in favor of an explosive model. However, a purely qualitative appreciation of structure is not to be trusted, and quantitative comparison of the topology of our models with the data and with explosive models would be valuable (Gott, Melott, and Dickinson 1986). It is our impression that the structure in our models is sufficiently like that found by de Lapparent, Geller, and Huchra and by Haynes and Giovanelli that it would be premature to conclude that suds require explosions.

VI. SUMMARY AND CONCLUSIONS

We have investigated the behavior of a flat cold dark matter (CDM) universe on scales, $r \geq 10$ Mpc. On such scales the predictions of our model are uniquely determined, since all the free parameters were fixed in Paper I by comparison with observed galaxy clustering on smaller scales. For this investigation we used N-body techniques complemented by the statistical analysis of peaks in Gaussian random fields developed by BBKS. We ran an ensemble of 25 simulations of comoving regions of present size 280 Mpc and a single simulation of a region of present size 360 Mpc. In our flat cold dark matter model, galaxies are required to form only near high peaks of the smoothed linear density field. We located "galaxies" in the mass distributions of our simulations by assuming their abundance to be proportional to the local number density of peaks given by the BBKS formalism. We were than able to select clusters in a similar way to Abell (1958), to construct volume-limited and magnitude-limited "galaxy" catalogs, to look for voids like that found by Kirshner *et al.* (1981), and to study the morphology of the galaxy distribution using the methods adopted in a variety of recent surveys. Our main conclusions are as follows.

1. Our biased CDM model produces the observed abundance of Abell clusters of richness class greater than 0; the predicted abundance of clusters of richness class greater than 1 is lower, but not significantly lower, than in the sample of Bahcall and Soneira (1983). The model clusters have masses and luminosities typical of clusters in the real universe. The predicted mass-to-light ratios inside the "Abell radius" are essentially independent of cluster radius and decrease slowly with cluster luminosity. The mean mass-to-light ratio of the clusters is about 20% of the value required to close the universe.

2. The bias in mean mass-to-light ratio remains large over regions the size of the Local Supercluster. If the real universe is biased in the same way as our model, the observed luminosity overdensity and Virgocentric infall velocity are compatible with $\Omega = 1$.

3. The correlation function of rich clusters in our models falls to unity at a separation of about 22 Mpc. This correlation length is 5 times larger than that of the mass distribution but only twice that of the galaxies. This result is in disagreement with existing estimates of the cluster correlation length. For the available redshift sample of ~ 100 clusters, the correlation length is 4–5 times that observed for galaxies. In addition the amplitude of the galaxy-cluster cross-correlation function in our models is a factor of 2.5 lower than the estimate of Seldner and Peebles (1977) on scales of order 30 Mpc.

4. The three-dimensional rms peculiar velocity of clusters in our model is only 145 km s^{-1}, in agreement with the upper limits derived by Aaronson *et al.* (1986). The rms relative peculiar velocity of cluster pairs is 511 km s^{-1}, much smaller than the values quoted by Bahcall and Soneira (1983). The rms peculiar velocity of individual "galaxies" in our simulations is 490 km s^{-1}, comparable to our observed velocity through the microwave background. Large-scale streaming motions in our models are very much smaller than those reported recently by Collins, Joseph, and Robertson (1986) and by Burstein *et al.* (1986).

5. Voids approaching the size of the one in Bootes are readily produced in our CDM model. The Bootes void is 124 Mpc in diameter; in our ensemble of 280 Mpc simulations we found three voids with diameters exceeding 112 Mpc. These numbers are larger than those quoted in Paper I simply because our previous work used the distribution of counts in fixed cells to measure the frequency of voids, whereas in the present paper we followed the same procedure as Kirshner *et al.* (1981, 1983).

6. Simulations of deep pencil-beam surveys show considerable large-scale structure and rather large variations from field to field. Large clumps are often separated by low-density regions extending up to 30,000 km s^{-1} in depth. This behavior is in good agreement with observational studies by Koo, Kron, and Szalay (1986) and by Broadhurst, Ellis, and Shanks (1986).

7. Artificial magnitude-limited catalogs with an apparent magnitude limit $m_{\text{Zwicky}} < 15.5$ show similar morphology to the real Zwicky catalog when projected on the sky. Filaments, large low-density regions, and apparent superclusters are seen in the artificial data as well as in the real map.

8. Wedge diagrams in redshift space similar to those presented recently by de Lapparent, Geller, and Huchra (1986) have a strong resemblance to the real data. Long curving filaments, often with sharp edges, are seen in our artificial catalogs, and objects similar to the Perseus-Pisces filament can be found. The range of morphology in such catalogs is similar to that in the different regions surveyed by Haynes and Giovanelli (1986).

Our results demonstrate that observed large-scale structures can be produced by the gravitational amplification of random-phase density fluctuations in a flat universe. Other possibilities such as open models, nonrandom phase initial conditions, or explosive generation of structure are not required by present observations, and in the latter two cases it still remains to be seen if they are, in fact, consistent with observation. On the other hand, if further work confirms the existence of large cluster-cluster correlations and of substantial streaming velocities on large scales, the particular flat CDM model we have analyzed will probably be excluded; however, an extension of it may well be able to accomodate these data. At the moment, the biasing we have invoked is qualitatively plausible but lacks a quantitative physical justification. On the basis of further work which is almost completed, we are hopeful that biasing of the type required will turn out to be an automatic consequence of the nonlinear evolution of the CDM distribution on small scales. While continuing to seek such a justification, we are impressed that a single model with $\Omega = 1$ can explain such a

wealth of structural information on objects ranging from the halos of individual galaxies to the largest known clusters, filaments, and voids.

This work was supported by NSF Presidential Young Investigator award AST-83352062, by NSF grant AST-8642105, and by NATO travel grant 689/84.

REFERENCES

Aaronson, M., Bothun, G., Mould, J., Huchra, J. P., Schommer, R. A., and Cornell, M. E. 1986, *Ap. J.*, **302**, 206.
Abell, G. O. 1958, *Ap. J. Suppl.*, **3**, 211.
Bahcall, N., and Soneira, R. 1983, *Ap. J.*, **270**, 20.
Bardeen, J. M. 1986, in *Inner Space/Outer Space*, ed. E. W. Kolb, M. S. Turner, D. Lindley, K. Olive, and D. Seckel (Chicago: University of Chicago Press), p. 212.
Bardeen, J. M., Bond., J. R., Kaiser, N., and Szalay, A. S. 1986, *Ap. J.*, **304**, 15 (BBKS).
Bean, A. J., Efstathiou, G., Ellis, R. S., Peterson, B. A., and Shanks, T. 1983, *M.N.R.A.S.*, **205**, 605.
Blumenthal, G. R., Faber, S. M., Primack, J. R., and Rees, M. J. 1984, *Nature*, **311**, 517.
Bond, J. R., and Efstathiou, G. 1984, *Ap. J. (Letters)*, **285**, L45.
Broadhurst, T. J., Ellis, R. S., and Shanks, T., in preparation.
Burstein, D., Davies, R. L., Dressler, A., Faber, S. M., Lynden-Bell, D., Terlevich, R., and Wegner, G. 1986, preprint.
Collins, A., Joseph, R. D., and Robertson, N. A. 1986, *Nature*, in press.
Davis, M., Efstathiou, G., Frenk, C. S., and White, S. D. M. 1985, *Ap. J.*, **292**, 371 (Paper I).
Davis, M., and Huchra, J. P. 1982, *Ap. J.*, **254**, 437.
Davis, M., Huchra, J. P., Latham, D., and Tonry, J. 1982, *Ap. J.*, **253**, 423.
Davis, M., and Peebles, P. J. E. 1983, *Ap. J.*, **267**, 465.
de Lapparent, V., Geller, M. J., and Huchra, J. P. 1986, *Ap. J. (Letters)*, **302**, L1.
Efstathiou, G., Davis, M., Frenk, C. S., and White, S. D. M. 1985, *Ap. J. Suppl.*, **57**, 241 (EDFW).
Einasto, J., Joeveer, M., and Saar, E. 1980, *M.N.R.A.S.*, **193**, 353.
Fall, S. M. 1979, *Rev. Mod. Phys.*, **51**, 21.
Felten, J. E. 1985, *Comm. Ap. Sp. Sci.*, **11**, 53.
Frenk, C. S. 1986, *Phil. Trans. Roy. Soc.*, in press.
Frenk, C. S., White, S. D. M., Efstathiou, G., and Davis, M. 1985, *Nature*, **317**, 595.
Gott, J. R., Melott, A., and Dickinson, M. 1986, *Ap. J.*, **306**, 341.
Gregory, S. A., Thompson, L. A., and Tifft, W. G. 1981, *Ap. J.*, **243**, 411.
Haynes, M. P., and Giovanelli, R. 1986, *Ap. J. (Letters)*, **306**, L55.
Kaiser, N. 1984, *Ap. J. (Letters)*, **284**, L9.
———. 1986, in *Inner Space/Outer Space*, ed. E. W. Kolb, M. S. Turner, D. Lindley, K. Olive, and D. Seckel (Chicago: University of Chicago Press), p. 258.
Kirshner, R. P., Oemler, A., Schechter, P. L., and Shectman, S. A. 1981, *Ap. J. (Letters)*, **248**, L57 (KOSS).
———. 1983, in *IAU Symposium 104, Early Evolution of the Universe and its Present Structure*, ed. G. O. Abell and G. Chincarini (Dordrecht: Reidel), p. 197 (KOSS).
Klypin, A. A., and Kopylov, A. A. 1983, *Soviet Astr. Letters*, **9**, 41.
Koo, D., Kron, R., and Szalay, A. S. 1986, in preparation.
Ling, N., Frenk, C. S., and Barrow, J. 1986, *M.N.R.A.S.*, in press.
Oemler, A. 1974, *Ap. J.*, **194**, 1.
Oort, J. H. 1983, *Ann. Rev. Astr. Ap.*, **21**, 373.
Politzer, H., and Preskill, J. 1986, *Phys. Rev. Letters*, **56**, 99.
Rees, M. J. 1985, *M.N.R.A.S.*, **213**, 75P.
Schaeffer, R., and Silk, J. I. 1985, *Ap. J.*, **292**, 319.
Seldner, M., and Peebles, P. J. E. 1977, *Ap. J.*, **215**, 703.
Seldner, M., Siebers, B., Groth, E. J., and Peebles, P. J. E. 1977, *A.J.*, **82**, 49.
Silk, J. I. 1985, *Ap. J.*, **297**, 1.
Stevenson, P. R. F., Shanks, T., Fong, R., and McGillivray, H. T. 1985, *M.N.R.A.S.*, **213**, 953.
Vittorio, N., and Silk, J. I. 1984, *Ap. J. (Letters)*, **285**, L39.
White, S. D. M. 1986, in *Inner Space/Outer Space*, ed. E. W. Kolb, M. S. Turner, D. Lindley, K. Olive, and D. Seckel (Chicago: University of Chicago Press), p. 228.

MARC DAVIS: Astronomy Department, University of California, Berkeley, CA 94720

GEORGE EFSTATHIOU: Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

CARLOS S. FRENK: Physics Department, University of Durham, Durham DH1 3LE, UK

SIMON D. M. WHITE: Steward Observatory, University of Arizona, Tucson, AZ 87521

THE ASTROPHYSICAL JOURNAL, 282:L47–L50, 1984 July 15
© 1984. The American Astronomical Society. All rights reserved. Printed in U.S.A.

LARGE-SCALE ANISOTROPY OF THE MICROWAVE BACKGROUND AND THE AMPLITUDE OF ENERGY DENSITY FLUCTUATIONS IN THE EARLY UNIVERSE[1]

L. F. ABBOTT[2]
Physics Department, Brandeis University
AND
MARK B. WISE
California Institute of Technology
Received 1984 January 12; accepted 1984 April 10

ABSTRACT

The spectrum of energy density fluctuations predicted by the inflationary cosmology is used to compute multipole moments of the microwave background temperature and the probability distributions for these moments. Comparison with the observed dipole moment gives a 90% confidence level upper bound of 3.4×10^{-6} on the amplitude of these fluctuations when they cross the horizon.

Subject headings: cosmic background radiation — cosmology

The large-scale anisotropy of the microwave background can provide important information about the amplitude and spectrum of long-wavelength energy density fluctuations in the early universe. In the inflationary cosmology (Guth 1981; Linde 1982; Albrecht and Steinhardt 1982), the spectrum of energy density perturbations is predicted (Guth and Pi 1982; Bardeen, Steinhardt, and Turner 1983; Starobinskii 1982; Hawking 1982) to be scale invariant (Harrison 1970; Zeldovich 1972), i.e., the amplitude of a fluctuation when it crosses the horizon is equal to a constant ε_H, independent of the wavelength of that fluctuation. The value of ε_H is determined by parameters of the scalar potential used to produce the phase transition that gives rise to the inflationary epoch. Thus, a knowledge of ε_H is of interest in both astrophysics and particle physics. Here we discuss what measurements of the microwave background imply about the value of ε_H.

In the inflationary cosmology, the universe is very near critical density so, in first approximation, it is described by a spatially flat Robertson-Walker metric with scale factor R and Hubble constant $H \equiv \dot{R}/R$. We use Bardeen's (1980) gauge-invariant variable, ε, to characterize energy density fluctuations. For a fluctuation of coordinate wavenumber k,

$$\varepsilon(\boldsymbol{k},t) = \frac{\delta\rho}{\rho} + 3\left(\frac{\rho+p}{\rho}\right)\left(\frac{RH}{k}\right)(v - B), \qquad (1)$$

where ρ and p are the unperturbed energy density and pressure and $\delta\rho$, v, and B are the amplitudes of fluctuations in the energy density, velocity, and off-diagonal metric respectively (see Bardeen 1980 for details). The variable ε reduces to $\delta\rho/\rho$ when the physical wavelength of the fluctuation is much less than the horizon size, $1/H$.

[1]Work supported by US Department of Energy Contract no. DE-AC03-76-ER03230 and DE-A103-81-ER40050.
[2]Supported in part by an Alfred P. Sloan Foundation Fellowship.

The prediction of the inflationary cosmology for wavelengths which enter the horizon during the matter-dominated era is

$$\varepsilon(\boldsymbol{k},t) = \tfrac{9}{4}\hat{a}(\boldsymbol{k})\varepsilon_H k^2 t_0^{4/3} t^{2/3}, \qquad (2)$$

where t_0 is the present time and $\hat{a}(\boldsymbol{k})$ is a random variable satisfying

$$\langle \hat{a}^*(\boldsymbol{k})\hat{a}(\boldsymbol{q})\rangle = \frac{1}{k^3}\delta^3(\boldsymbol{k} - \boldsymbol{q}). \qquad (3)$$

Equation (2) can also be written in terms of the conformal time variable $\tau = 3t_0^{2/3}t^{1/3}$ as

$$\varepsilon(\boldsymbol{k},\tau) = \tfrac{1}{4}\hat{a}(\boldsymbol{k})\varepsilon_H k^2 \tau^2. \qquad (4)$$

The time of horizon crossing is $\tau = 2/k$, so the average value of ε^2, defined as

$$k^3 \int d^3q \langle \varepsilon^*(\boldsymbol{k},t)\varepsilon(\boldsymbol{q},t)\rangle, \qquad (5)$$

is equal to ε_H^2 at the time of horizon crossing. The random nature of the prediction for ε reflects the underlying quantum mechanical origin of the fluctuations in the inflationary cosmology.

The fluctuation in the observed cosmic blackbody temperature (Sachs and Wolfe 1967; Abbott and Wise 1984*c*) viewed along the direction unit vector $\boldsymbol{e}$, produced by an energy density fluctuation of coordinate wave vector $\boldsymbol{k}$, is

$$\frac{\delta T_0}{T_0}(\boldsymbol{e},\boldsymbol{k}) = \frac{1}{k^2}\int_0^{\tau_0 - \tau_E} e^{i\boldsymbol{k}\cdot\boldsymbol{e}y}(\boldsymbol{e}\cdot\boldsymbol{k})^2 \dot{\varepsilon}(\boldsymbol{k}, \tau_0 - y). \qquad (6)$$

Here the dot represents a τ derivative, $\tau_0 = 3t_0$ and τ_E is the conformal time of recombination, $\tau_E = 3t_0^{2/3}t_E^{1/3}$. The micro-

wave background temperature can be expanded in multipoles

$$\frac{\delta T_0}{T_0} = \sum_{l,m} a_{lm} Y_{lm}(e). \tag{7}$$

The quantities

$$a_l^2 = \sum_{m=-l}^{l} |a_{lm}|^2 \tag{8}$$

are rotationally invariant. Predictions for the multipole moments of the background temperature have been obtained from equations (3), (4), and (6) by projecting out a given multipole in equation (6), integrating over wave vectors, k, squaring, and taking an expectation value. The result, for $l \geq 2$ is given to a good approximation by (Peebles 1982; Abbott and Wise 1984*a*)

$$\langle a_l^2 \rangle = \frac{2\pi^2 \varepsilon_H^2 (2l+1)}{l(l+1)}. \tag{9}$$

For $l \leq 20$, this prediction depends only on very long wavelength fluctuations and so is insensitive to fluctuations which have gone nonlinear and to the details of the process of recombination.

Equation (9) is the prediction for an expectation value averaged over many universes. Measurements of the microwave background are of course restricted to a single universe. Fortunately, the inflationary cosmology (in the weak-coupling approximation for the scalar field in the inflationary epoch) also predicts the higher correlation functions of $\hat{a}(\boldsymbol{k})$, and these can be used to construct a probability distribution for $a_l \equiv (a_l^2)^{1/2}$. In the inflationary cosmology, the multipole coefficients a_{lm} are Gaussian random variables, so a_l has the probability distribution

$$P_l(a_l) = \sqrt{\frac{2}{\pi}} \frac{1}{1 \cdot 3 \cdot 5 \cdots (2l-1)} \frac{(a_l)^{2l}}{\sigma_l^{(2l+1)}} e^{-[1/2(a_l/\sigma_l)^2]} \tag{10}$$

with σ_l given by

$$\sigma_l = \frac{2\pi^2 \varepsilon_H^2}{l(l+1)}. \tag{11}$$

This determines the statistical uncertainties in the predictions of equation (9). The results for the first 10 moments are shown in Figure 1 where 68% confidence level error bars are given. Even with the relatively large statistical uncertainties, these predictions for a_l^2 are significantly different from those of some non-scale-invariant spectra (Abbott and Wise 1984*c*).

A limit on the value of ε_H can be obtained by comparing the prediction for $l = 2$ with the present limit on the quadrupole anisotropy (Lubin, Epstein, and Smoot 1983; Fixen, Cheng, and Wilkinson 1983)

$$a_2^2 < 5.2 \times 10^{-8}. \tag{12}$$

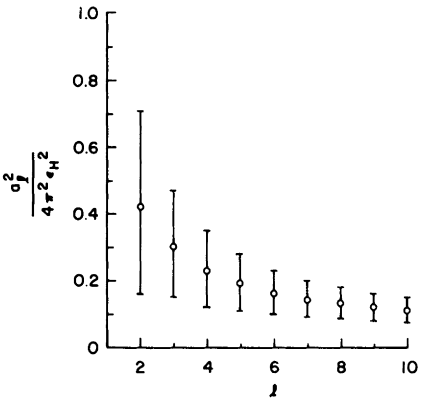


FIG. 1.—Predictions for moments 2 through 10 with error bars showing 68% confidence level limits.

At 90% confidence level, the probability distribution (10) gives

$$a_2^2 > 0.3 \frac{5\pi^2}{3} \varepsilon_H^2, \tag{13}$$

so we find that

$$\varepsilon_H < 1.0 \times 10^{-4}. \tag{14}$$

A more stringent bound on ε_H can be obtained from the observed dipole anisotropy. The dipole moment has been considered in a similar way by Silk and Wilson (1981*a*, *b*) and by Kaiser (1983). The prediction for $l = 1$ is obtained by combining equations (3), (4), and (6). We find that wavelengths with $k < k_{max}$, which entered the horizon during the matter-dominated era, give a contribution to the dipole moment of

$$\langle a_1^2 \rangle = 6\pi^2 \varepsilon_H^2 k_{max}^2 t_0^2. \tag{15}$$

Unlike the higher multipoles, the dipole moment is sensitive to short-wavelength fluctuations. This is because the dipole moment is dominated by the peculiar velocity of the observer. However, the observed dipole moment can still be used to obtain a bound on ε_H without having to consider modes which have gone nonlinear. If we assume that the short-wavelength modes that have gone nonlinear are uncorrelated with the long-wavelength modes that are evolving linearly, then the contributions of both of these sets of modes to $\langle a_1^2 \rangle$ is positive. The full probability distribution for the dipole moment is a convolution of probability distributions for the linear and nonlinear contributions. With the distribution for the linear modes given by equation (10), it is straightforward to show that the nonlinear modes can only make the full probability distribution broader than that of the linear modes alone. We thus obtain a bound on ε_H by demanding that the computable long-wavelength contribution to the dipole anisotropy is no larger than the observed dipole anisotropy (Lubin, Epstein, and Smoot 1983; Fixen, Cheng, and Wilkinson 1983)

$$a_1^2 = 5.5 \times 10^{-6}, \tag{16}$$

at 90% confidence level. (The statistical error in our prediction dominates over the experimental error in the measurement of the dipole moment.)

To derive the best possible bound, we must consider wavelengths which entered the horizon before the time of matter domination, but which are still evolving linearly today. In keeping with the prediction of inflation that we are at critical density, we assume that the universe is dominated by an uncoupled form of dark matter such as neutrinos (Cowsik and McClelland 1972; DeRujula and Glashow 1980), axions (Preskill, Wise, and Wilczek 1983; Abbott and Sikivie 1983; Dine and Fischler 1983), photinos (Cabibbo, Farrar, and Maiani 1983), or gravitinos (Pagels and Primack 1982), and that the universe became matter dominated by this dark matter at about 1,000 years. Before the time of matter domination, the inflationary cosmology predicts that fluctuations in the radiation density satisfy

$$\varepsilon_{\rm rad}(\boldsymbol{k},t) = \hat{a}(\boldsymbol{k}) A_{\rm rad}\left[\frac{\sin x - x\cos x}{x}\right], \tag{17}$$

where

$$x = \frac{k}{\sqrt{3}\,RH}, \tag{18}$$

$A_{\rm rad}$ is a constant, and $\hat{a}$ is a random variable satisfying equation (3). To analyze how these fluctuations are transferred to the dark matter, we use an extension of Bardeen's (1980) gauge-invariant formalism to the case of multiple, uncoupled fluids (Abbott and Wise 1984*b*).

The amplitude of fluctuations which entered the horizon before the matter-dominated era is determined by numerically integrating the coupled dark matter–radiation evolution equations for $\varepsilon_{\rm rad}$ and $\varepsilon_{\rm dm}$. We find that the amplitude of the dark matter fluctuations during the matter-dominated era is

$$\varepsilon_{\rm dm}(\boldsymbol{k},t) = \tfrac{9}{4}\hat{a}(\boldsymbol{k}) A_{\rm dm}(k) k^2 t_0^{4/3} t^{2/3}, \tag{19}$$

where $A_{\rm dm}$ is given approximately by

$$A_{\rm dm} \approx \frac{\varepsilon_{\rm H}}{1 + 0.04x_m^2 + 0.0005x_m^4} \tag{20}$$

with

$$x_m = \frac{2}{\sqrt{3}}(3/4)^{2/3} k t_0^{2/3} t_m^{1/3}. \tag{21}$$

Here t_m is the time the universe becomes matter dominated. Equation (20) is a good approximation to our numerical results for $x_m \leq 8$. For wavelengths which enter the horizon after the time of matter domination, x_m is small so equations (19) and (20) reproduce equation (2).

The bound we derive from the dipole anisotropy will be expressed in terms of $\varepsilon_{\rm H}$. However, $\varepsilon_{\rm H}$ and $A_{\rm rad}$ are proportional to each other, so our bound also applies to $A_{\rm rad}$. In the radiation-dominated era, for small x, the dark matter fluctuations satisfy (Abbott and Wise 1984*b*)

$$\varepsilon_{\rm dm} = \tfrac{9}{16}\varepsilon_{\rm rad} \tag{22}$$

and for large x, $\varepsilon_{\rm dm}$ grows logarithmically, eventually growing by a factor of 4–5. To relate $\varepsilon_{\rm H}$ and $A_{\rm rad}$ we take equation (22) and match it onto the growing and shrinking matter-dominated modes of $\varepsilon_{\rm dm}$ at the time when matter first dominates. Then comparing the coefficient of the growing mode with the prediction (2), we find[3]

$$A_{\rm rad} = 8\left(\tfrac{4}{3}\right)^{4/3}\varepsilon_{\rm H}. \tag{23}$$

To determine the contribution of modes that are evolving linearly to the dipole moment, we use equations (19) and (20) in equation (6), integrate over $\boldsymbol{k}$, project out the dipole term, square, and take the expectation value. The k integration must be cut off at the point where linear perturbation theory breaks down, i.e., when the average value of ε^2, as defined in equation (5), is of order unity. With a Hubble constant, $H_0 = 100$ km s^{-1} Mpc^{-1}, the average value of ε^2 for a 60 Mpc wavelength is at most 10^{-1}, so we feel confident that linear perturbation theory is valid for wavelengths longer than this. We find that the contribution of modes with a wavelength longer than 60 Mpc to a_1^2 is greater than or equal to $4.8 \times 10^5 \varepsilon_{\rm H}^2$ at the 90% confidence level. Comparison of this result with the observed dipole moment (16) gives the bound

$$\varepsilon_{\rm H} < 3.4 \times 10^{-6} \tag{24}$$

at the 90% confidence level. If we are less conservative and use a 30 Mpc cutoff, then the bound improves to $\varepsilon_{\rm H} < 1.9 \times 10^{-6}$. The bound is not very sensitive to the present value of the Hubble constant. If a Hubble constant $H_0 = 50$ km s^{-1} Mpc^{-1} is used, then we find $\varepsilon_{\rm H} < 3.8 \times 10^{-6}$ for a 120 Mpc cutoff and $\varepsilon_{\rm H} < 2.7 \times 10^{-6}$ for a 60 Mpc cutoff.

The bound of equation (24) is valid for any form of cold dark matter. For neutrinos (hot dark matter), our bound is modified somewhat due to free streaming. Using the results of Bond and Szalay (1983), we find that

$$\varepsilon_{\rm H} < 7 \times 10^{-6} \tag{25}$$

for neutrinos using a 60 Mpc cutoff.

The bound (24) implies that the contribution of energy density fluctuations to the quadrupole moment satisfies $a_2^2 <$

[3] To obtain this result, we must determine the value of the scale factor R in the radiation-dominated period and use this result in equation (18). If we write

$$R = \left[\frac{t + (1/3)t_m}{t_0}\right]^{2/3}$$

in the matter-dominated era, then by continuity of R and $\dot{R}$,

$$R = \left(\frac{4}{3}\right)^{2/3}\left(\frac{t_m}{t_0}\right)^{2/3}\left(\frac{t}{t_m}\right)^{1/2}$$

in the radiation-dominated era.

3.2×10^{-10} at the 90% confidence level. If the scale of the symmetry breaking responsible for inflation is of order 10^{17} GeV, then long-wavelength gravitational waves generated during the inflationary era will contribute significantly to the quadrupole and higher moments (Starobinskii 1979; Rubakov, Sazhin, and Veryaskin 1982; Fabbri and Pollock 1983). Thus, in this case, a_2^2 could be bigger than 3.2×10^{-10} without violating our dipole bound (gravitational waves do not contribute to the dipole moment). If the scale of inflation is below 10^{17} GeV, then no appreciable gravitational contribution exists. The bound on ε_H in equation (24) is close to the value of ε_H determined by Peebles from an analysis of the galaxy-galaxy correlation function (Peebles 1982). This analysis does not involve any large theoretical statistical uncertainties; however, the correlation function is not well measured at large distances. Also, the correlation function approach must assume that galaxies are a good indicator of the distribution of mass in the universe, even though the total mass is dominated by dark matter.

Throughout this *Letter* we have assumed an inflationary cosmology. However, our results are valid in any cosmology with $\Omega = 1$ and a scale-invariant spectrum of energy density fluctuations that has a Gaussian probability distribution.

We wish to thank Alan Guth, Nathan Isgur, and John Preskill for helpful conversations.

REFERENCES

Abbott, L., and Sikivie, P. 1983, *Phys. Letters*, **120B**, 133.
Abbott, L., and Wise, M. 1984*a*, *Phys. Letters*, **135B**, 279.
_____. 1984*b*, *Nucl. Phys. B*, **237**, 226.
_____. 1984*c*, in preparation.
Albrecht, A., and Steinhardt, P. 1982, *Phys. Rev. Letters*, **48**, 1220.
Bardeen, J. 1980, *Phys. Rev. D*, **22**, 1882.
Bardeen, J., Steinhardt, P., and Turner, M. 1983, *Phys. Rev. D*, **28**, 679.
Bond, J., and Szalay, S. 1983, *Ap. J.*, **274**, 443.
Cabibbo, N., Farrar, G., and Maiani, L. 1983, *Phys. Letters*, **105B**, 155.
Cowsik, R., and McClelland, J. 1972, *Phys. Rev. Letters*, **29**, 669.
DeRujula, A., and Glashow, S. 1980, *Phys. Rev. Letters*, **45**, 942.
Dine, M., and Fischler, W. 1983, *Phys. Letters*, **120B**, 137.
Fabbri, R., and Pollock, M. 1983, *Phys. Letters*, **125B**, 445.
Fixen, D., Cheng, E., and Wilkinson, D. 1983, *Phys. Rev. Letters*, **50**, 620.
Guth, A. 1981, *Phys. Rev. D.*, **23**, 347.
Guth, A., and Pi, S.-Y. 1982, *Phys. Rev. Letters*, **49**, 1110.
Harrison, E. 1970, *Phys. Rev. D.*, **1**, 2726.
Hawking, S. 1982, *Phys. Letters*, **115B**, 295.
Kaiser, N. 1983, *Ap. J. (Letters)*, **273**, L17.
Linde, A. 1982, *Phys. Letters*, **108B**, 389.
Lubin, P., Epstein, G., and Smoot, G. 1983, *Phys. Rev. Letters*, **50**, 616.
Pagels, H., and Primack, J. 1982, *Phys. Rev. Letters*, **48**, 223.
Peebles, P. J. E. 1982, *Ap. J. (Letters)*, **263**, L1.
Preskill, J., Wise M., and Wilczek, F. 1983, *Phys. Letters*, **120B**, 127.
Rubakov, V., Sazhin, M., and Veryaskin, A. 1982, *Phys. Letters*, **115B**, 189.
Sachs, R., and Wolfe, A. 1967, *Ap. J.*, **147**, 73.
Silk, J., and Wilson, M. 1981*a*, *Ap. J.*, **243**, 14.
_____. 1981*b*, *Ap. J. (Letters)*, **244**, L37.
Starobinskii, A. 1979, *JETP Letters*, **30**, 683.
_____. 1982, *Phys. Letters*, **117B**, 175.
Zeldovich, Y. 1972, *M.N.R.A.S.*, **160**, 1P.

L. F. ABBOTT: Physics Department, Brandeis University, Waltham, MA 02254

MARK B. WISE: Physics Department, California Institute of Technology, Pasadena, CA 91125

THE ASTROPHYSICAL JOURNAL, 285:L39–L43, 1984 October 15
© 1984. The American Astronomical Society. All rights reserved. Printed in U.S.A.

FINE-SCALE ANISOTROPY OF THE COSMIC MICROWAVE BACKGROUND IN A UNIVERSE DOMINATED BY COLD DARK MATTER

NICOLA VITTORIO
Department of Astronomy, University of California, Berkeley; and Istituto Astronomico, Universita di Roma "La Sapienza," Roma
AND
JOSEPH SILK
Department of Astronomy, University of California, Berkeley
Received 1984 May 30; accepted 1984 July 10

ABSTRACT

The fine-scale anisotropy of the cosmic microwave background radiation has been studied in cosmological models with a scale-invariant primordial adiabatic density fluctuation spectrum that are dominated by cold, weakly interacting particles such as axions or photinos. Normalization of the present fluctuation spectrum to the observed galaxy distribution, equivalent to the assumption that mass and light are correlated on large scales, results in excessive temperature anisotropy when compared to a recent upper limit on 4'.5 unless the density parameter Ω_0 exceeds 0.4 (50 km s^{-1} Mpc$^{-1}/H_0$). Combining this result with the requirement that the universe be at least 13 billion years old, we conclude that if the cosmological constant is zero, $0.4 \lesssim \Omega_0 \lesssim 1$ and 60 km s^{-1} Mpc^{-1} $\gtrsim H_0 \gtrsim$ 50 km s^{-1} Mpc^{-1}.

Subject headings: cosmic background radiation — cosmology

Primordial nucleosynthesis and grand unification are generally considered to be desirable aspects of the evolution of the early universe, leading to the usual Friedmann-Lemaître cosmological model in which the baryon density parameter satisfies $0.03 < \Omega_b < 0.1$ (Yang *et al.* 1984), and galaxies form from primordial adiabatic density fluctuations. The search for temperature anisotropy in the cosmic background radiation induced by density fluctuations on the last scattering surface has proved to be one of the most important tests of a baryon-dominated universe. Suppression of small-scale structure by radiative damping guarantees that the first nonlinear structures must have developed recently. Consequently reionization cannot plausibly occur early enough to modify the last scattering surface. Fine-scale anisotropy limits on angular scales in excess of several arc minutes have unambiguously ruled out baryon-dominated universes for any power-law initial adiabatic fluctuation power spectrum, including the plausible and natural hypothesis of scale-invariant fluctuations (Wilson and Silk 1981; Wilson 1983).

Hence attention has focused on a cosmological model dominated by dark matter in the form of a weakly interacting nonbaryonic species (Bond and Szalay 1983). Two candidate particle species have emerged, a neutrino of mass[1] $\sim 100\ \Omega\ h^2$eV which first becomes nonrelativistic at $kT \lesssim m_\nu c^2$ when the horizon scale contains $M_\nu \approx 10^{15}\ (m_\nu/100\ \text{eV})^{-2}\ M_\odot$, and cold relics, such as photinos or axions, which are nonrelativistic (and thereby suppress any free-streaming) at epochs when the horizon scale contains masses of interest for galaxy formation. Free-streaming erases all substructure for the neutrino fluctuations below M_ν, but substructure survives in the cold relic spectrum down to scales $\lesssim 10^6\ M_\odot$. This leads to very different scenarios for galaxy formation. In the top-down scenario, fragmentation of large-scale pancakes occurs at a late epoch ($z \lesssim 5$) in a neutrino-dominated universe (Zel'dovich and Sunyaev 1972), whereas in the bottom-up picture, hierarchical clustering develops at an earlier epoch ($z \gtrsim 10$) in the cold relic–dominated cosmology (Blumenthal *et al.* 1984). Recently, *N*-body simulations of galaxy clustering (White, Frenk, and Davis 1983) and predictions of excessive dipole anisotropy (Kaiser 1983) have cast considerable doubt on the neutrino–dominated model but appear to lead to acceptable correlations in a cold relic–dominated universe either if $\Omega_0 h \approx 0.2$ or if the dark matter is more weakly correlated than the galaxy distribution (Davis *et al.* 1984).

The purpose of this *Letter* is to study the fine-scale anisotropy of the cosmic background radiation in a universe dominated by weakly interacting cold particles. Reducing Ω_0 below unity increases the amplitude of the fluctuations at the epoch of last scattering, for a given normalization to large-scale structure. Hence we may anticipate that observational limits on radiation anisotropy should impose a lower bound on Ω_0. The cold particle scenario is of especial interest since, precisely because of the persistence of small-scale power, it yields the smallest large-scale fluctuations and consequently $\delta T/T$, for any model involving primordial scale-invariant adiabatic fluctuations. Throughout, we assume that the cosmological constant is zero. We show that a recent observational upper limit on $\delta T/T$ imposes the significant constraint $\Omega_0 h \gtrsim 0.2$. This result is encouraging for believers in inflation, which predicts that $\Omega_0 = 1$, but it cannot be straightforwardly

[1] We adopt the usual notation $h \equiv H_0/100$ km s^{-1} Mpc^{-1}.

reconciled with astronomical determinations of Ω_0, which generally require $\Omega_0 \approx 0.15$ (± 0.1) (e.g., Sandage and Tammann 1984).

We consider the growth of perturbations in a universe dominated by cold particles and also containing radiation, baryons, and three species of massless neutrinos. Our analysis follows that of Peebles and Yu (1970) and Wilson and Silk (1981). The synchronous gauge is used:

$$ds^2 = c^2\,dt^2 - a^2(t)\left[\delta_{\alpha\beta} - h_{\alpha\beta}(x,t)\right] dx^\alpha\, dx^\beta, \quad (1)$$

where $a(t)$ is the expansion factor and the perturbed metric $h_{\alpha\beta} \ll 1$. Radiation and massless neutrinos are described by the Boltzmann equation. We describe the coupling of radiation with the matter by means of isotropic Thomson scattering. The angular dependence is treated by an expansion in Legendre polynomials. The evolution of the relevant quantities is evaluated solving the appropriate set of differential equations.

In the subhorizon limit, for a $p = 0$ universe, the perturbed Boltzmann equation has the following analytical solution for the fractional brightness at the present epoch t_0:

$$\delta(u,k,t_0) = |\delta_{ss}(\mu,k)|e^{-ik\mu c\tau} - \frac{8i\mu}{\pi}\frac{\lambda(t_0)}{\lambda_H(t_0)}\delta_c(k,t_0) + \frac{8}{\pi^2}\left[\frac{\lambda(t_0)}{\lambda_H(t_0)}\right]^2\delta_c(k,t_0), \quad (2)$$

where k is the comoving wavenumber; $\mu = \cos(\theta) = \boldsymbol{k}\cdot\boldsymbol{p}/|\boldsymbol{k}||\boldsymbol{p}|$ with $\boldsymbol{p}$ being the photon momentum; λ_H is the dimension of the horizon at t_0 and $\tau = \int_0^{t_0} dt'/a(t')$. For the small angular scale anisotropy, only the first term is important, and it is evaluated by numerical solution of the Boltzmann equation. An observer at present measures temperature anisotropy with fractional brightness $|\delta_{ss}(\mu,k)|$. A standard recombination history has been computed for each set of Ω_0, Ω_b, and h, and the smearing due to the finite thickness of the last scattering surface is incorporated into the numerical solution for $|\delta_{ss}(\mu,k)|$. Provided the last scattering surface is at a redshift $z^* \gg \Omega_0^{-1}$, the angle subtended by a given comoving length scale l_0 is independent of z^*:

$$\theta(l_0) = \frac{\Omega_0}{2c}l_0 H_0 \equiv \frac{l_0}{r_0} = 34''\Omega_0 h l_0(\mathrm{Mpc}). \quad (3)$$

To study the small-scale anisotropy of the cosmic microwave background radiation, we use the angular correlation function of two measurements spaced an angle α apart and made with an antenna of beam width σ (Doroshkevich, Zel'dovich, and Syunyaev 1978)

$$C(\alpha,\sigma) \propto \int_0^\infty k^{2+n}dk\int_{-1}^{1} d\mu|\delta_{ss}(\mu,k)|^2\exp(-\zeta^2\sigma^2)\cdot J_0(\zeta\alpha), \quad (4)$$

where

$$\zeta = k(1-\mu^2)^{1/2}r_0. \quad (5)$$

The index n refers to the spectral index of the initial power spectrum $|\delta_k|^2 \propto k^n$ adopted for the primordial density fluctuations. The exponential factor takes into account the smearing at very high k due to the finite beam of the antenna, and J_0 is a Bessel function of order zero.

Once Ω_0 and Ω_b, referring to the normalized (in units of the critical closure density, $3H_0^2/8\pi G$) total and baryonic densities are specified, the only uncertainty in the computations of $\delta T/T$ arises with the normalization to the density fluctuation spectrum. The variance $\delta M/M$ for the mass within a randomly placed sphere of radius r is defined by (Peebles 1980)

$$\left[\frac{\delta M}{M}(r)\right]^2 \propto \int_0^\infty k^{2+n}dk|\delta_c(k,t_0)|^2\frac{(\sin kr - kr\cos kr)^2}{(kr)^6}. \quad (6)$$

We normalize this to $\delta N/N$ for luminous galaxies, equal to unity at a scale $r_0 = 8h^{-1}$ Mpc (Davis and Peebles 1983). If the dark matter were more weakly correlated than the baryonic matter in luminous galaxies with a clustering scale smaller than r_0, the predicted fluctuations $\delta M/M$ would be correspondingly reduced.

The models to be discussed here include inflationary models ($\Omega_0 = 1$) and low-density, cold dark matter–dominated models ($0.1 \lesssim \Omega_0 \lesssim 1$). Use of the variance $\delta M/M$ at r_0 for normalization introduces a dependence of the predicted $\delta T/T$ fluctuations on Ω_0 and h. Consider first the simplest case, $\Omega_0 = 1$. The Hubble constant is not well determined, and we consider the range $0.5 \le h \le 1$. For $\Omega_0 \approx 1$, it is most unlikely that h exceeds 0.5 (if $\Lambda = 0$); otherwise, the universe would be too young. Reducing h has the effect of increasing the comoving horizon scale at the epoch of matter radiation equality,

$$L_{eq} \approx 12\left(\Omega_0 h^2\right)^{-1}\ \mathrm{Mpc}. \quad (7)$$

This scale demarcates the transition in the fluctuation spectrum between the Zel'dovich scale-invariant limit on large scales, which first enter the horizon at matter-dominated epochs, and the flattening in the regime of suppressed growth on the smaller scales which enter the horizon during the radiation-dominated era (Peebles 1982; Bond, Szalay, and Turner 1983; Blumenthal and Primack 1984).

We can distinguish several effects that contribute to the dependence of the predicted $\delta T/T$ on Ω_0 and h. Effect (a), increasing L_{eq}, flattens the fluctuation spectrum and increases the large-scale amplitude for the same normalization. Moreover, effect (b), the reduction in h means that the scale of normalization increases, further enhancing the large-scale power. However, effect (c), the prediction of $\delta T/T$ at some specified angular scale implies that larger comoving scales are contributing as $\Omega_0 h$ is reduced: this has the effect of tending to reduce the predicted amplitude of $\delta T/T$. Effects (a) and (c) approximately cancel each other, and effect (b) results in a

scaling that is approximately proportional to h^{-1} since the fluctuation spectrum is normalized at the transition regime where the approximate slope is $\delta M/M \propto M^{-1/3}$. Finally, effect (d), reducing Ω_0 cuts the growth period since recombination by a factor $1/\Omega_0$ in redshift, and the required amplitude of $\delta M/M$ at recombination must increase by this factor for a given normalization at the present epoch.

We infer that the predicted $\delta T/T$ should scale, to a first approximation, as $(\Omega_0^l h^m)^{-1}$, with $l \approx m \approx 1$. Varying Ω_b over the range allowed by primordial nucleosynthesis ($0.03 \lesssim \Omega_b \lesssim 0.2$) only has a small effect: if Ω_0 is fixed, and Ω_b is increased, some modest suppression of growth occurs prior to recombination and slightly raises the predicted $\delta T/T$. Hence we set conservative limits on $\Omega_0 h$ by taking the low value of Ω_b (equal to 0.03). A more extensive set of models, including a discussion of the large-scale anisotropy, different values of Ω_b, the predictions of $\delta T/T$ in a massive neutrino-dominated universe, a nonzero cosmological constant, and nonstandard ionization histories of the universe, will be described elsewhere (Vittorio and Silk 1984). The computed temperature fluctuation angular correlation function for a cold relic–dominated universe is shown in Figure 1*a*. The effects of observing with a finite beam width reduce the power substantially for a steep ($n = 4$) initial spectrum, but a small beam (1.5) has little effect on the scale-invariant ($n = 1$) spectrum predictions: smearing due to the finite depth of the last scattering surface cuts the power off at large k. The limiting value $n = 4$ corresponds to the minimal fluctuations expected on large scales due to any casual reshuffling of matter on small scales (Peebles 1980).

The most sensitive observational limit on anisotropy to date has been performed over an angular scale of 4.5, where the finite thickness of the last scattering surface does not strongly suppress the power contributing to $\delta T/T$ (Uson and Wilkinson 1984*a*, *b*). In order of magnitude, if recombination occurs over $\Delta z/z \approx 0.1$, one expects the thickness of the surface of last scattering to subtend an angle of order 10 $\Omega_0^{-1/2}$ arcmin. Power due to fluctuations of comoving scale $< 15\ \Omega_0^{-1/2} h^{-1}$ Mpc, corresponding to the thickness of the last scattering surface, will be smeared out. The observational technique in this experiment involved beam switching between three positions in the sky, where subscript 1 denotes the central beam,

$$\delta T_{\rm rms} = \left\langle |T_1 - \tfrac{1}{2}(T_2 + T_3)|^2 \right\rangle^{1/2}. \tag{8}$$

In terms of the correlation function $C(\alpha, \sigma)$ defined in equation (4), we have

$$\left\langle \left| \frac{\delta T}{T}(\alpha, \sigma) \right|^2 \right\rangle = \tfrac{3}{2} C(0, \sigma) - 2C(\alpha, \sigma) + \tfrac{1}{2} C(2\alpha, \sigma). \tag{9}$$

The Uson-Wilkinson experiment yields a 95% confidence upper limit

$$(\delta T/T)_{\rm rms} < 3 \times 10^{-5}.$$

In Figure 1*b*, we compare our theoretical predictions of $\delta T/T(\alpha, \sigma)$ with the experimental limit. We infer that the scale-invariant–fluctuation spectrum leads to excessive $\delta T/T$ in a low-density ($\Omega_0 h \lesssim 0.2$) universe.

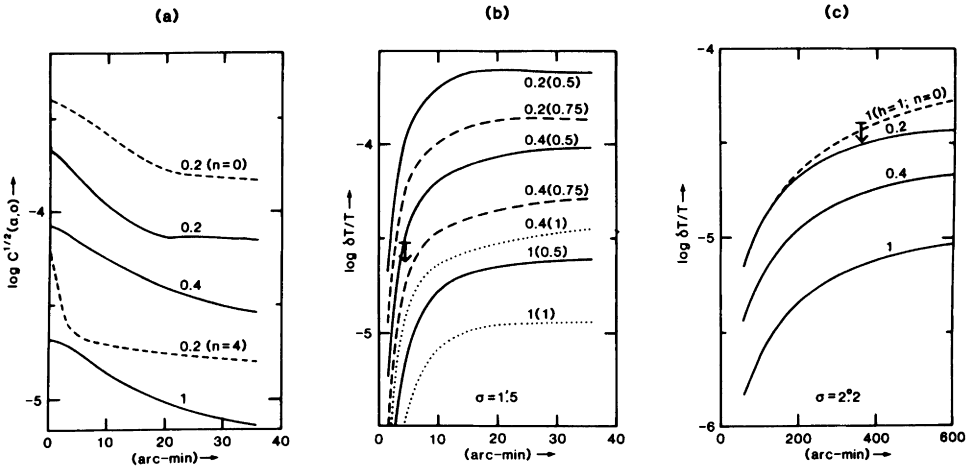


FIG. 1.—Predictions of $\delta T/T$ for a universe dominated by cold particles. (*a*) The angular correlation function measured by an antenna with perfect resolution is shown for models with $\Omega_0 = 0.2$, 0.4, and 1 and a scale-invariant spectrum ($n = 1$). Also shown are predictions for white noise ($n = 0$) and minimal fluctuations ($n = 4$) in the $\Omega_0 = 0.2$ model (*dashed lines*): all models have Hubble constant $h = 0.5$. (*b*) The rms temperature fluctuations measured in a triple beam switching experiment with a beam of 1.5 are shown for various combinations of Ω_0, with the corresponding value of h in parentheses. All predictions are for $n = 1$. The experimental upper limit (95% confidence) is also indicated (Uson and Wilkinson 1984*b*). (*c*) The rms temperature fluctuations measured in a double beam switching experiment with a beam of 2.2 are shown for $\Omega_0 = 0.2$, 0.4, and 1, with h taken to be 0.5 and a scale-invariant ($n = 1$) primordial spectrum . Also shown is the prediction for white noise fluctuations ($n = 0$), and $\Omega_0 = 1$ and $h = 1$, and the upper limit (95% confidence) from the experiment of Melchiorri *et al.* (1981).

This conclusion can be avoided if the universe becomes reionized sufficiently early so as to smear out the fine-scale anisotropy. Two arguments may be marshaled to demonstrate that this is unlikely. First, following Peebles (1984), we have calculated the earliest epoch at which nonlinearity occurs. Lowering Ω_0 flattens the spectrum, and we find, for example, for $\Omega_0 = 0.2$, $h = 0.5$, that $\langle(\delta M/M)^2\rangle^{1/2}$ is enhanced by a factor of 2 at the low-mass end of the spectrum relative to the normalization scale. Hence even 3 σ fluctuations could have achieved nonlinearity and collapsed no earlier than a redshift of about 30. It seems most unlikely that ionizing pregalactic sources, developing self-consistently from the same fluctuation spectrum designed to produce large-scale structure, could have yielded significant scattering optical depth subsequent to the epoch of hydrogen neutralization and matter-radiation decoupling. Unit optical depth to Thomson scattering is achieved in a fully ionized universe at redshift

$$1 + z = 70(\Omega_0/0.2)^{1/3}(0.02/\Omega_b h)^{2/3}$$

Second, any such smearing is inoperative over angular scales larger than a few degrees, the angle subtended by the horizon at last scattering. An experiment performed by Melchiorri *et al.* (1981) sets a 95% confidence limit of $\delta T/T < 4 \times 10^{-5}$ on an angular scale of 6° for an experiment utilizing a double beam switching technique with a 2° beamwidth. Since this experiment measures $\delta T_{\rm rms} = \langle|T_1 - T_2|^2\rangle^{1/2}$, we recover the usual expression $|\delta T/T(\alpha, \sigma)|^2 = 2[C(0, \sigma) - C(\alpha, \sigma)]$. In Figure 1*c*, we show a comparison of the upper limit over 6° with our theoretical predictions.

A primordial white noise spectrum ($n = 0$) is found to be inconsistent even if $\Omega = 1$, but $h < 1$. The 6° two-beam experiment, despite the large beam which smooths out much of the primordial power, still yields the useful constraint $\Omega_0 \gtrsim 0.14$, $h \lesssim 0.7$. In the limit of small α, $\delta T/T \propto \alpha^2$ for double subtraction (three beams) but $\delta T/T \propto \alpha$ for single subtraction (two beams). Moreover, at fixed angle α, expansion of the Bessel function in expression (4) for $C(\alpha, \sigma)$ shows that the contribution from the large wavelength perturbations is weighted by a factor proportional to k^2 in the single subtraction experiment and k^4 in the double subtraction experiment. This means that the latter experiment probes a narrower region of the primordial fluctuation spectrum. Since lowering h independently of Ω_0 also flattens the spectrum and enhances the predicted anisotropy, we summarize our results in Figure 2, where we show 95% confidence contour limits on excluded regions in the Ω_0-h plane. Also shown are constraints attainable by a future three-beam experiment with sensitivity $\delta T/T = 1 \times 10^{-5}$ at 4'.5.

Age dating of globular cluster stars provides an independent constraint on the age of the universe and on Ω_0 and h. Recent studies (e.g., Iben and Renzini 1984) conclude that the globular cluster ages are 17 (± 2) billion years. We show in Figure 2 the inferred constraint if the age of the universe is at least 13 billion years (and $\Lambda = 0$). We conclude that, combining our result on $\delta T/T$ with a conservative age constraint requires $0.4 \lesssim \Omega_0 \lesssim 1$ and $0.6 \gtrsim h \gtrsim 0.5$.

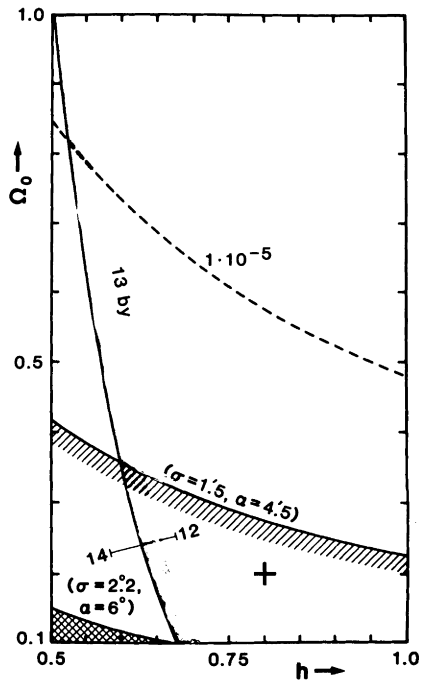


FIG. 2.—Constraints on Ω_0 and h from $\delta T/T$. The upper limits on $\delta T/T$ at 4'.5 due to Uson and Wilkinson (1984*b*) and 6° (Melchiorri *et al.* 1981) are expressed as constraint contours in the h-Ω_0 plane. The excluded region (below the hatched line) results in excessive $\delta T/T$ in a cold particle–dominated universe provided the large-scale mass distribution follows that of the galaxies, assumed to have formed from primordial scale-invariant density fluctuations. The dashed line indicates the improvement potentially available were the current upper limits obtained in the Uson and Wilkinson (1984*a*, *b*) experiment lowered to $\delta T/T < 1 \times 10^{-5}$. The light hatched line shows the constraints on Ω_0 and h if the universe is older than the oldest globular cluster stars, taken to have formed 13 billion years ago. The bar shows the Ω_0-h limit for ages of 12 and 14 billion years (with zero cosmological constant). Also, the cross is inferred from N-body simulations (see text).

Recent N-body simulations of large-scale structure in a cold particle–dominated universe have only succeeded in obtaining the observed slope for the galaxy correlation function if $\Omega_0 \approx 0.2$, in a run with $h = 0.8$, under the assumption that the mass and light distributions are identical: too steep a slope is obtained for larger values of $\Omega_0 h$ (Davis *et al.* 1984). We tentatively conclude that our $\delta T/T$ limit for an axion or photino-dominated universe in which galaxies form from scale-invariant adiabatic fluctuations can be reconciled with the N-body simulations and with an age for the universe in excess of 13 billion years only if the mass distribution is more weakly clustered than the observed galaxy distribution. In fact, fluctuations in the Gaussian tail of the adopted fluctuation spectrum do not go nonlinear sufficiently early to reionize the universe and reduce the anisotropy significantly. If mass and light are correlated on large scales, the microwave limits

alone imply $\Omega_0 \gtrsim 0.4$ if the universe is at least 13 billion years old and are both compatible with, and may even provide the first unambiguous astrophysical evidence in support of, the inflationary prediction $\Omega_0 = 1$ for a scale-invariant primordial fluctuation spectrum in a universe dominated by cold weakly interacting particles.

This research has been supported in part by DOE contract AT03-82ER40069, by NASA grant NGR 05-003-578, and by NATO. We acknowledge helpful discussions with J. R. Bond, M. Davis, C. Frenk, N. Kaiser, A. Lasenby, J. Uson, and S. White. Similar results have been independently obtained by J. R. Bond and G. Efstathiou.

REFERENCES

Blumenthal, G. R., and Primack, J. R. 1984, in preparation.
Blumenthal, G., Faber, S., Primack, J., and Rees, M. 1984, preprint.
Bond, J. R., and Szalay, A. S. 1983, *Ap. J.*, **276**, 443.
Bond, J. R., Szalay, A. S., and Turner, M. S. 1982, *Phys. Rev. Letters*, **48**, 1636.
Davis, M., and Peebles, P. J. E. 1983, *Ap. J.*, **267**, 465.
Davis, M., Efstathiou, G., Frenk, C. S., and White, S. D. M. 1984, preprint.
Doroshkevich, A. G., Zel'dovich, Ya. B., and Syunyaev, R. A. 1978 *Sov. Astr.–AJ*, **22**, 523.
Iben, I., and Renzini, A. 1984, *Rept. Prog. Phys.*, in press.
Kaiser, N. 1983, *Ap. J. (Letters)*, **273**, L17.
Melchiorri, F., Melchiorri, B., Ceccarelli, C., and Pietranera, L. 1981, *Ap. J.*, **250**, L1.
Peebles, P. J. E. 1980, *The Large Scale Structure of the Universe* (Princeton: Princeton University Press).
Peebles, P. J. E. 1982, *Ap. J. (Letters)*, **263**, L1.
_____. 1984, *Ap. J.*, **277**, 470.
Peebles, P. J. E., and Yu, J. T. 1970, *Ap. J.*, **162**, 815.
Sandage, A., and Tammann, G. 1984, in *Proc. First ESO-CERN Conference on Elementary Particles and Large-Scale Structure of the Universe*, in press.
Uson, J. M., and Wilkinson, D. T. 1984*a*, *Ap. J. (Letters)*, **277**, L1.
_____. 1984*b*, in *Proc. Inner Space–Outer Space Symposium*, in press.
Vittorio, N., and Silk, J. 1984, in preparation.
Wilson, M. 1983, *Ap. J.*, **273**, 2.
Wilson, M. L., and Silk, J. 1981, *Ap. J.*, **243**, 14.
White, S. D. M., Frenk, C. S., and Davis, M. 1983, *Ap. J. (Letters)*, **274**, L1.
Yang, S., Turner, M. S., Steigman, G., Schramm, D. N., and Olive, K. A. 1984, *Ap. J.*, **281**, 493.
Zel'dovich, Ya. B., and Sunyaev, R. A. 1972, *Astr. Ap.*, **20**, 189.

JOSEPH SILK: Department of Astronomy, University of California, Berkeley, CA 94720

NICOLA VITTORIO: Istituto Astronomico, Universita di Roma "La Sapienza," Via Lancisi 29, 00100 Roma, Italy

THE ASTROPHYSICAL JOURNAL, 285:L45–L48, 1984 October 15
© 1984. The American Astronomical Society. All rights reserved. Printed in U.S.A.

COSMIC BACKGROUND RADIATION ANISOTROPIES IN UNIVERSES DOMINATED BY NONBARYONIC DARK MATTER

J. R. BOND[1,2] AND G. EFSTATHIOU[2,3]
Received 1984 June 4; accepted 1984 July 17

ABSTRACT

We present detailed calculations of the temperature fluctuations in the cosmic background radiation for universes dominated by massive collisionless relics of the big bang. We assume an initially adiabatic constant curvature perturbation spectrum. In models with cold dark matter, the simplest hypothesis—that galaxies follow the mass distribution—leads to small-scale anisotropies which exceed current observational limits if $\Omega < 0.2$ $h^{-4/3}$. Since low values of Ω are indicated by dynamical studies of galaxy clustering, cold particle models in which light traces mass are probably incorrect. Reheating of the pregalactic medium is unlikely to modify this conclusion. In cold particle or neutrino-dominated universes with $\Omega = 1$, our predictions for small-scale and quadrupole anisotropies are below current limits. In all cases, the small-scale fluctuations are predicted to be $\sim 10\%$ linearly polarized.

Subject headings: cosmic background radiation — cosmology — galaxy formation

I. INTRODUCTION

Current observational constraints on anisotropies in the cosmic background radiation (CBR) and on the clustering of galaxies have considerably narrowed the range of acceptable models for galaxy formation. The recent limits of Uson and Wilkinson[4] (1984*a*, *b*) of $(\Delta T/T) < 2.9 \times 10^{-5}$ at angular scales of 4'.5 essentially rule out all models with an adiabatic primordial fluctuation spectrum in which the present mean mass density of the universe is composed entirely of baryonic matter (Wilson and Silk 1981; Wilson 1983).

In neutrino-dominated universes this difficulty may be avoided (Bond, Efstathiou, and Silk 1980; Doroshkevich *et al.* 1980). However, detailed computations of the coherence length in the mass distribution (Bond and Szalay 1981, 1983; Peebles 1982) combined with *N*-body simulations of galaxy clustering (White, Frenk, and Davis 1983) show that the neutrino picture conflicts with observations of the galaxy distribution unless galaxies formed at uncomfortably recent epochs. Similar conclusions follow from considerations of large-scale streaming motions (Kaiser 1983*a*).

Models in which the dark matter is cold (e.g., axions, photinos, etc.) preserve many of the salient features of earlier hierarchical clustering theories and offer a promising way of overcoming some of the difficulties associated with massive neutrinos (e.g., Blumenthal *et al.* 1984). In the simplest cold particle schemes, galaxies are assumed to be good tracers of the underlying mass distribution. If this is the case, then observations of the peculiar velocities between galaxy pairs imply a low-density cosmological model with $\Omega = 0.14 \times 2^{\pm 1}$ (Davis and Peebles 1983; Bean *et al.* 1983). *N*-body simulations (Davis *et al.* 1984) do indeed show that a low-density model with $\Omega = 0.2$ can match many features of the observed clustering pattern, whereas numerical simulations with $\Omega = 1$ lead to excessive peculiar velocities and only reproduce the observed shape of the galaxy correlation function $\xi(r)$ if Hubble's constant is unreasonably small ($h \approx 0.2$, where h is Hubble's constant in units of 100 km s^{-1} Mpc^{-1}).

In this *Letter*, we describe detailed computations of the small angular scale temperature anisotropies expected in cold particle and neutrino-dominated universes on the assumption that the primordial fluctuations were adiabatic with a constant curvature spectrum $|\delta_k|^2 \propto k$. As we will show, the recent upper limits of Uson and Wilkinson (1984*a*, *b*) provide powerful constraints on low-density cold particle models.

II. CALCULATIONS

In this section we summarize the main features of our calculations. A detailed discussion will be given in a subsequent paper. We assume a metric $ds^2 = a^2(\tau)\, dx^\mu\, dx^\nu(\eta_{\mu\nu} + h_{\mu\nu})$, where $a(\tau)$ is the scale factor, $\tau = \int dt/a$ is conformal time, and $\eta_{\mu\nu} = (1, -1, -1, -1)$ is the background metric. We adopt the synchronous gauge, $h_{0\mu} = 0$, and consider only scalar perturbations. The flat metric gives an accurate description of the background in open models at early times provided one considers temperature anisotropies on angular scales $\theta < (\Omega/2)(1 - \Omega)^{-1/2}$. On larger scales, curvature of the spacelike hypersurfaces cannot be neglected (e.g., Wilson 1983). Our methods are similar to those used by Peebles and Yu (1970) and Wilson and Silk (1981). The main new aspects of our calculations follow.

a) Radiation

We follow the transfer of the linear polarization of the photons as well as of total radiation intensity since the two are

[1] Physics Department, Stanford University.
[2] Institute for Theoretical Physics, University of California, Santa Barbara.
[3] Institute of Astronomy, Cambridge University.
[4] This is their 95% "field" limit which can be directly compared with our calculations and is $\sqrt{1.5}$ times their quoted "sky" limit (see § II*c*).

coupled by Thomson scattering (Kaiser 1983*b*). The evolution of the radiation anisotropy is described by the Boltzmann transport equation for the perturbation to the photon distribution function, $\delta f_T(\boldsymbol{q}, \boldsymbol{x}, \tau) = (T\partial\bar{f}/\partial T)\Delta_T/4$, where $\bar{f}$ is the unperturbed photon distribution function, T is the photon temperature, and $\boldsymbol{q}$ is the comoving momentum. For Thomson scattering, there is only one other nonzero Stokes parameter (Chandrasekhar 1960), the net linear polarization, $\delta f_P = (T\partial\bar{f}/\partial T)\Delta_P/4$. Both Δ_P and $\Delta T/T = \Delta_T(\hat{\boldsymbol{q}}, \boldsymbol{x}, \tau)/4$ are independent of photon energy, depending only on the direction of the photon momentum $\hat{\boldsymbol{q}}$. The spatial Fourier transforms of Δ_T and Δ_P, defined by $\Delta(\hat{\boldsymbol{q}}, \boldsymbol{k}, \tau) = \int d^3\boldsymbol{x}/V\Delta(\hat{\boldsymbol{q}}, \boldsymbol{x}, \tau)e^{-i\boldsymbol{k}\cdot\boldsymbol{x}}$ (where V is a box-normalizing volume), obey

$$\dot{\Delta}_T + ik\mu\Delta_T = \dot{h}(1-\mu^2) + \dot{h}_{33}(3\mu^2 - 1) - n_e\sigma_T a\left[\Delta_T - \Delta_{T0} - 4\mu v - \tfrac{1}{2}P_2(\mu)(\Delta_{T2} + \Delta_{P2} - \Delta_{P0})\right] \quad (1a)$$

$$\dot{\Delta}_P + ik\mu\Delta_P = -n_e\sigma_T a\left\{\Delta_P - \tfrac{1}{2}\left[1 - P_2(\mu)\right]\times(\Delta_{P0} - \Delta_{T2} - \Delta_{P2})\right\}. \quad (1b)$$

Here, dots denote differentiation with respect to τ, $h = -h_i^i$, $\mu = \hat{\boldsymbol{k}}\cdot\hat{\boldsymbol{q}}$ with $\boldsymbol{k}$ in the z-direction, v is the Fourier transform of the baryon velocity, and σ_T is the Thomson cross section. The background electron number density, n_e, is determined by solving the ionization equations of Peebles (1968) allowing for a 25% mass fraction of neutral helium and the presence of dark matter. We solve equations (1) by expanding Δ_T and Δ_P in angular moments, Δ_{Tl} and Δ_{Pl}, with respect to Legendre polynomials $\Delta(\hat{\boldsymbol{q}}, \boldsymbol{k}, \tau) = \Sigma(2l+1)\Delta_l(k, \tau)P_l(\mu)$. The expansions are truncated at a sufficiently large l to ensure high accuracy (e.g., $l \sim 300$ for $k \sim 1$ Mpc^{-1}). To check our code, we ran the adiabatic baryon-dominated model considered by Wilson and Silk and found excellent agreement.

b) Neutrinos and Cold Dark Matter

We include massless neutrinos and, when appropriate, massive neutrinos. The Boltzmann equations for both massless and massive neutrinos were solved by direct integration, following the methods described by Bond and Szalay (1983). For low wave numbers ($k < 0.15$ Mpc^{-1}), we expanded the massless neutrino distribution function in Legendre polynomials. For cold dark matter, the density perturbation obeys $\delta_X = \dot{h}/2$.

c) Observed Temperature Fluctuations

The correlation function for the temperature fluctuations at the present epoch τ_0 is

$$C(\theta, \tau_0) = \langle\Delta_T(\hat{\boldsymbol{q}}, \boldsymbol{x}, \tau_0)\Delta_T(\hat{\boldsymbol{q}}', \boldsymbol{x}, \tau_0)\rangle/16, \quad \cos\theta = \hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{q}}', \quad (2a)$$

$$\approx \int \frac{Vk^2\,dk}{2\pi^2}\frac{1}{16}\sum_{l>2}(2l+1)|\Delta_{Tl}(k, \tau_f)|^2 \times \sin(kR\theta)/kR\theta, \quad (2b)$$

$$R = \Delta\tau, \qquad \Delta\tau = \tau_0 - \tau_f, \qquad \Omega = 1,$$

$$= L_c\sinh(\Delta\tau/L_c), \quad L_c = H_0^{-1}(1-\Omega)^{-1/2}a_0^{-1}, \quad \Omega < 1,$$

(see Wilson 1983). Equation (2b) is an excellent approximation for small angle fluctuations ($kR\theta \ll 1$), if τ_f is chosen so that the optical depth to the observer at τ_0 is small.

In principle, the average in equation (2a) is taken over a large spatial volume; in practice, the initial fluctuations are assumed to have random phases for each Fourier mode and the average is over the ensemble: the spatial ergodic hypothesis is assumed to hold. The theoretical prediction for the results of a typical observation, $(\Delta T/T)_{\mathscr{R}} = \int\mathscr{R}(\hat{\boldsymbol{q}})\,d\Omega_{\hat{q}}\Delta_T(\hat{\boldsymbol{q}}, \boldsymbol{x}, \tau_0)/4$, where $\mathscr{R}(\hat{\boldsymbol{q}})$ is the angular response of the experiment, is a Gaussian probability distribution of $(\Delta T/T)_{\mathscr{R}}$ with variance

$$\langle(\Delta T/T)^2_{\mathscr{R}}\rangle = \int\mathscr{R}(\hat{\boldsymbol{q}})\,d\Omega_{\hat{q}}\mathscr{R}(\hat{\boldsymbol{q}}')\,d\Omega_{\hat{q}'}C(\hat{\boldsymbol{q}}\cdot\hat{\boldsymbol{q}}', \tau_0). \quad (3)$$

A beam-switching experiment with beam pattern F has $\mathscr{R} = F(\hat{\boldsymbol{q}} - \hat{\boldsymbol{q}}_1) - F(\hat{\boldsymbol{q}} - \hat{\boldsymbol{q}}_2)$, hence $\langle(\Delta T/T)^2_{\mathscr{R}}\rangle = 2[C_F(0) - C_F(\theta)]$, where $\cos\theta = \hat{\boldsymbol{q}}_1\cdot\hat{\boldsymbol{q}}_2$ and C_F is defined by equation (3). Uson and Wilkinson (1984*a*, *b*) measure the temperature difference between the direction $\hat{\boldsymbol{q}}_1$ and the average of the temperatures at $\hat{\boldsymbol{q}}_2$ and $\hat{\boldsymbol{q}}_3$, 4.5 on either side of $\hat{\boldsymbol{q}}_1$: $\mathscr{R} = F(\hat{\boldsymbol{q}} - \hat{\boldsymbol{q}}_1) - [F(\hat{\boldsymbol{q}} - \hat{\boldsymbol{q}}_2) + F(\hat{\boldsymbol{q}} - \hat{\boldsymbol{q}}_3)]/2$. Hence, we must compare their experimental results with our predictions of

$$\langle(\Delta T/T)^2_{\mathscr{R}}\rangle = 2\left[C_F(0) - C_F(\theta)\right] - \tfrac{1}{2}\left[C_F(0) - C_F(2\theta)\right]. \quad (4)$$

d) Normalization

The cold particle models are normalized by matching $J_3(x_0) = \int_0^{x_0}\xi(x)x^2\,dx$ to observational estimates from the CfA redshift survey (Davis and Peebles 1983) at $x_0 = 10\ h^{-1}$ Mpc. Here, ξ is the two-point correlation function of the cold particles, $\langle\delta_X(\boldsymbol{x} + \boldsymbol{x}')\delta_X(\boldsymbol{x}')\rangle$. J_3 is insensitive to the details of nonlinear clustering if $\xi(x_0) \ll 1$ (Peebles and Groth 1976; Efstathiou 1979). Accordingly, x_0 has been chosen so that observed fluctuations on this scale are close to linear. In § III, we show how sensitive our results are to this choice.

Implicit in the above procedure is the assumption that the observed galaxy distribution follows that of the dark matter. This is not expected to hold in neutrino-dominated universes (Bond *et al.* 1984). We therefore normalize to the redshift $z_{\rm nl}$, where $\xi(0, z_{\rm nl}) = 1$. Using the statistical distributions given by Doroshkevich and Shandarin (1978), we find that 11% of the pancakes have begun to form caustics and 1% of the gas has shocked by $1 + z = 1.44(1 + z_{\rm nl})$; by $z_{\rm nl}$, 50% of the pan-

TABLE 1

COLD DARK MATTER FLUCTUATION PARAMETERS

Ω (1)	Ω_B (2)	h_0 (3)	a(Mpc) (4)	b(Mpc) (5)	c(Mpc) (6)	ν (7)	θ_c (arc min) (8)	β (9)	$C(0)^{1/2}$ $\times 10^5$ (10)	$\Delta T/T$ $\times 10^5$ (11)	$J_3(x_0)$ (h^{-3} Mpc^3) (12)	(13)	z_{nl} (14)	z_{reh} (15)	Ω_{BC} (16)
1.0	0.03	0.75	11.3	5.29	3.10	1.13	8.4	0.94	1.2	0.28	210	340	30	117	1.9×10^{-6}
1.0	0.03	0.50	23.1	11.4	6.48	1.25	8.8	1.25	2.8	0.42	190	380	20	151	6.9×10^{-15}
0.4	0.03	0.75	23.5	17.3	7.15	1.07	5.2	0.90	3.0	1.1	180	450	25	86	1.2×10^{-5}
0.3	0.03	0.75	37.1	21.1	10.8	1.12	4.5	0.88	4.4	1.8	170	470	20	79	4.2×10^{-6}
0.2	0.03	0.75	48.2	31.6	15.6	1.30	3.7	0.94	8.0	3.6	160	520	19	70	1.9×10^{-5}
0.2	0.10	0.75	76.2	71.1	27.7	2.15	5.2	1.82	22.0	7.9	150	580	15	32	2.6×10^{-3}
0.2	0.03	1.00	37.9	18.4	10.1	1.21	3.9	0.75	5.4	2.3	170	490	23	58	2.7×10^{-4}
0.2	0.03	0.50	207.0	47.9	38.5	1.73	3.6	1.22	15.6	7.2	160	530	24	91	1.1×10^{-5}
CfA Survey											180	550			
$x_0 h$ Mpc											7.5	15			

NOTES ON SELECTED COLUMNS.—Parameters for the transfer functions (eq. [6]) and the radiation correlation functions (eq. [5]) are given in columns (4)–(10). If Silk damping is negligible, a, b, and c should scale as $(\Omega h^2)^{-1}$, and ν should be constant. Column (11) gives our prediction of $\langle(\Delta T/T)^2\rangle^{1/2}$ at 4'.5 for the experimental setup used by Uson and Wilkinson (eq. [4]) assuming a Gaussian beam response with half-width 1'.5. Columns (10) and (11) have been normalized using $J_3(10\ h^{-1}\ \text{Mpc}) = 270\ h^{-3}\ \text{Mpc}^3$ estimated from the CfA redshift survey. Columns (12) and (13) give J_3 at two additional values of x_0 which may be compared with the CfA estimates.

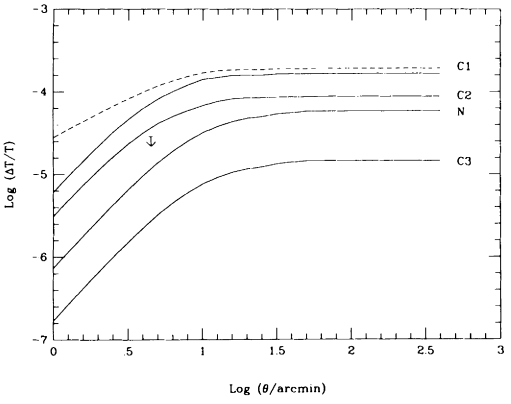


FIG. 1.—Temperature fluctuation as a function of angular scale. Curves C1–C3 show results for cold particle models with the following parameters: (C1) $\Omega = 0.2$, $h = 0.5$; (C2) $\Omega = 0.2$, $h = 0.75$; (C3) $\Omega = 1.0$, $h = 0.75$. The curve labeled (N) shows results for a massive neutrino model with $\Omega = 1.0$, $h = 0.75$, normalized so that $\xi(0, z_{nl}) = 1$ at $z_{nl} = 3$. In all cases $\Omega_B = 0.03$. The solid lines show our predictions for the experimental setup used by Uson and Wilkinson (see eq. [4]). In case (C1), we compare their procedure with the results expected in a standard beam-switching experiment shown as the dashed line. The 95% confidence upper limit of Uson and Wilkinson (1984*b*) is marked by the top of the arrow. The theoretical curves have been calculated assuming a Gaussian beam response of half-power width 1'.5.

cakes have begun to form caustics and 10% of the gas has shocked. These estimates agree with the N-body results of White, Frenk, and Davis (1983).

III. RESULTS AND DISCUSSION

The fitting formula

$$C(\theta)/C(0) = \left[1 + (\theta/\theta_c)^2/(2\beta)\right]^{-\beta} \qquad (5)$$

accurately reproduces our computed values of $C(\theta)$. The parameters θ_c and β are given in Table 1 for various cases. The coherence angle, θ_c, characterizes the loss of correlation with distance on the last scattering surface. If we define a correlation function $C_P(\theta)$ for the polarization amplitude Δ_P in an analogous way to equation (2a), then we find $|C_P(0)/C(0)|^{1/2} \approx 0.08$ in all our models. It is only in determining $C(0)$ that we require the overall normalization amplitude. Our X-particle transfer functions may be accurately represented by the formula

$$T_X(k) = \left\{1 + \left[ak + (bk)^{3/2} + (ck)^2\right]^\nu\right\}^{-1/\nu}. \qquad (6)$$

In Table 1, we list the values for the parameters. We also illustrate the sensitivity to variations in the J_3 normalization radius, x_0. To some extent, this is a measure of how well the model reproduces the observed large-scale clustering properties. The low-density models agree quite well with the CfA results and lend confidence that our anisotropy predictions are not sensitive to the choice of x_0. We have also normalized our results by equating the rms overdensity in a sphere of radius x_s, $\langle[(\Delta M/M)(x_s)]^2\rangle^{1/2}$, to the observed rms fractional excess of galaxy counts in such a sphere, which is unity at $x_s = 8\ h^{-1}$ Mpc (Peebles 1984). We find good agreement with our J_3 normalization.

Our final results are given in Figure 1 and Table 1. By comparing with the Uson and Wilkinson (1984*b*) limit, we find that the models with $\Omega \leq 0.2\ h^{-4/3}$ for $\Omega_B = 0.03$ can be ruled out at the 95% confidence level; notice from Table 1 that the h scaling in this formula applies only to low Ω models.

One way of avoiding these constraints is to invoke reheating. Small-scale anisotropies can be erased by rescattering of the CBR photons off gas that has been ionized after recombination. Erasure is significant if reionization occurs earlier than the minimum redshift required to make the optical depth to us unity, z_{reh}. If full ionization is maintained since reheating,

then $z_{reh} \approx 100$ for low Ω_B models. (See Table 1.) An early population of massive stars is one of the few ways to reionize at such early epochs. To assess whether this is plausible, we now estimate the amount of gas that would have collapsed by z_{reh}. Although for this spectrum the X-particle fluctuations diverge logarithmically as $r \to 0$, the baryons cannot cluster below their Jeans length. If $\Omega_B \ll \Omega$, and the gas and radiation temperatures are approximately the same (valid for $z \sim z_{reh}$), then the baryon transfer function is $T_B(k) \approx T_X(k)[1 + (k/k_B)^2]^{-1}$ (Peebles 1984), where $k_B = 0.9(\Omega h^2)^{1/2}$ kpc^{-1} is the gas Jeans wavenumber just after recombination. With this approximate spectrum, $\xi_B(0)$ is finite and we can define the redshift at which the rms density fluctuations in the gas reach unity by $\xi_B(0, z_{nl}) = 1$. The fraction of the gas which has $\delta_B > 1$ at z_{reh} is given by $\Omega_{BC}/\Omega_B = \frac{1}{2}\mathrm{erfc}\{[2\xi_B(0, z_{reh})]^{-1/2}\}$. Values for z_{nl} and Ω_{BC} are listed in Table 1. Carr, Bond, and Arnett (1984) have estimated the fraction of the closure density in massive stars of various types required for reionization to occur via the overlapping of their H II regions. Even in the optimal case of massive objects radiating into unclumped gas, they find $\Omega_* > 10^{-5}$ to 10^{-6} is necessary. For such a density to arise only from the collapsed gas at z_{reh} would imply extremely high efficiencies of massive star formation for low Ω_B models. This illustrates that only exotic models of pregalactic star formation could lead to significant damping of the anisotropy.

The $\Omega = 1$ massive neutrino model shown in Figure 1 lies below the observational limits unless $z_{nl} > 8.7$ if $h = 0.75$. The limit is more severe if $h = 0.5$: $z_{nl} > 3.5$ if $\Omega_B = 0.03$. These constraints are less stringent than those from N-body studies (White, Frenk, and Davis 1983) but are not as sensitive to the details of galaxy formation.

Current observational limits on large angular scales do not lead to strong constraints. For example, Lubin, Epstein, and Smoot (1983) and Fixsen, Cheng, and Wilkinson (1983) give 90% confidence limits for the quadrupole of $Q < 0.38$ mK and $Q < 0.18$ mK, respectively, where $Q^2 = (Q_2^2 + Q_3^2 + Q_4^2 + Q_5^2)/4$ and the Q_l are defined in the Berkeley convention. (We do not include the Q_1 component.) If $\Omega = 1$ and $\Omega_B \ll 1$, we obtain $Q = 5.5 \times 10^{-3}\ h^{-1}$ mK for the cold particle case and $Q = 2.8 \times 10^{-3}\ h^{-2}(1 + z_{nl})$ mK for the massive neutrino case. We do not quote values for the low Ω models as there is no natural generalization of the Zel'dovich spectrum on scales exceeding that of the spatial curvature.

We have shown that strong limits on low-density universes dominated by dark matter follow from the predicted small-angle CBR fluctuations, and that these are unlikely to be affected by reheating. Cold particle and massive neutrino models with $\Omega = 1$ are not strongly constrained at present. Low-density cold particle models may be reconciled with observations if mass is less strongly clustered than galaxies. In this case, the anisotropy limits only require that the epoch of nonlinearity be less than $\sim 17(\Omega h/0.2)^2$. However, if mass does not follow light, many of the difficulties that the high-density cold particle models face when compared with the observed clustering properties and peculiar velocities might also be surmounted.

We thank Alex Szalay for useful discussions and for his assistance in the early stages of this project. We thank J. Uson and D. Wilkinson for discussions regarding their experimental results. N. Vittorio and J. Silk have done similar calculations to those reported here and their results are in substantial agreement with ours. This research was supported in part by the UK SERC and by the NSF under grant PHY77-27084, supplemented by funds from NASA.

REFERENCES

Bean, A. J., Efstathiou, G., Ellis, R. S., Peterson, B. A., and Shanks, T. 1983, *M.N.R.A.S.*, **205**, 605.
Blumenthal, G. R., Faber, S. M., Primack, J. R., and Rees, M. J. 1984, *Nature*, in press.
Bond, J. R., Centrella, J., Szalay, A. S., and Wilson, J. 1984, *M.N.R.A.S.*, in press.
Bond, J. R., Efstathiou, G., and Silk, J. 1980, *Phys. Rev. Letters*, **45**, 1980.
Bond, J. R. and Szalay, A. S. 1981, in *Neutrino 81*, ed. R. J. Cence, E. Ma, and A. Roberts (Honolulu: University of Hawaii, Department of Physics and Astronomy), Vol. **1**, p. 59.
_____. 1983, *Ap. J.*, **277**, 443.
Carr, B. J., Bond, J. R., and Arnett, W. D. 1984, *Ap. J.*, **277**, 445.
Chandrasekhar, S. 1960, *Radiative Transfer* (New York: Dover).
Davis, M., Efstathiou, G., Frenk, C. S., and White, S. D. M. 1984, preprint.
Davis, M., and Peebles, P. J. E. 1983, *Ap. J.*, **267**, 465.
Doroshkevich, A. G., Khlopov, M., Sunyaev, R. A., Szalay, A. S., and Zel'dovich, Ya. B. 1980, *Ann. NY Acad. Sci.*, **375**, 32.
Doroshkevich, A. G., and Shandarin, S. F. 1978, *M.N.R.A.S.*, **182**, 27.
Efstathiou, G. 1979, *M.N.R.A.S.*, **187**, 117.
Fixsen, D. J., Cheng, E. S., and Wilkinson, D. T. 1983, *Phys. Rev. Letters*, **50**, 620.
Kaiser, N. 1983*a*, *Ap. J.* (*Letters*), **273**, L17.
_____. 1983*b*, *M.N.R.A.S.*, **202**, 1169.
Lubin, P. M., Epstein, G. L., and Smoot, G. F. 1983, *Phys. Rev. Letters*, **50**, 616.
Peebles, P. J. E. 1968, *Ap. J.*, **153**, 1.
_____. 1982, *Ap. J.*, **258**, 415.
_____. 1984, *Ap. J.*, **277**, 470.
Peebles, P. J. E., and Groth, E. J. 1976, *Astr. Ap.*, **53**, 131.
Peebles, P. J. E., and Yu, J. T. 1970, *Ap. J.*, **162**, 815.
Uson, J. M., and Wilkinson, D. T. 1984*a*, *Ap. J. Letters*, 277, L1.
_____. 1984*b*, Proc. of the Inner Space/Outer Space Workshop, Fermilab, in press.
Wilson, M. L. 1983, *Ap. J.*, **273**, 2.
Wilson, M. L., and Silk, J. 1981, *Ap. J.*, **243**, 14.
White, S. D. M., Frenk, C. S., and Davis, M. 1983, *Ap. J.* (*Letters*), **274**, 1.

J. R. BOND: Department of Physics, Stanford University, Stanford, CA 94305

G. EFSTATHIOU: Institute of Astronomy, Cambridge University, Cambridge CB3 OHA, England, UK

Improved limits on small-scale anisotropy in cosmic microwave background

Juan M. Uson & David T. Wilkinson

Physics Department, Princeton University, Princeton, New Jersey 08544, USA

As a remnant of the early Universe, the cosmic microwave background provides unique information on the initial conditions from which matter has evolved to form the structures we see today. All efforts to detect small-scale structure in this radiation have so far been unsuccessful (ref. 1 and refs therein)[1-4]. Nevertheless, upper limits set on possible underlying fluctuations restrict the range of physical models for perturbations of the density in the early Universe. Our search for small-scale anisotropy in the background radiation has now resulted in a lowering of the upper limit on root-mean-square fluctuations ($\Delta T_{r.m.s.}$) observed at an angular scale of ~4 arc min to $\Delta T_{r.m.s.}/T < 2.1 \times 10^{-5}$ at the 95% confidence level (where $T = 2.7$ K, the temperature of the background radiation). The actual limits deduced from our experiment depend on the model assumed for the unseen fluctuations. Several possibilities are discussed as well as the implications this new measurement has for various cosmological models.

The 140-foot telescope of the US National Radio Astronomy Observatory (NRAO), with the K-band maser receiver mounted at the Cassegrain focus, is an excellent instrument for high-sensitivity continuum observations[2]. This system routinely achieves a temperature of 50 K with 370 MHz bandwidth, centred at 19.5 GHz. Moreover, it is exceedingly stable for long integrations and consistently reaches a sensitivity of $\Delta T_{r.m.s.} = 10^{-4}$ K (antenna temperature) for 2 h of observation in the best weather conditions (clear sky and ambient temperature below 0 °C). At this frequency, the half-power beamwidth is about 1.5 arc min.

We have used this system to search for small-scale anisotropy in the microwave background radiation. Although our first observations were made in December 1980, we shall be concerned here only with data collected since December 1982, which are of much better quality than previous data. Good data were obtained from 27-31 December 1982 (82 h) and 16-20 May 1983 (night-time only, 30 h): these have been discussed in ref. 2. The new observations reported here were made between 25 February and 5 March 1984 (88 h) with additional data taken between 3 and 7 January 1984 (4 h) in a filler programme scheduled with other observations. Thus, a total of 174 h of high quality data were obtained from these observing runs out of a total of 445 h scheduled initially.

Fluctuations in the gain of the receiver, as well as variable contributions of radiation from the atmosphere and the ground, are important sources of systematic errors in this kind of measurement. To overcome these problems, the telescope beam is switched at a rate of 3.33 Hz between two positions in the sky, separated by 4.5 arc min, by mechanically oscillating the secondary reflector. The receiver output, δT, which is detected synchronously with beam position, is thus a measure of the difference between the antenna temperatures of the radiation

collected from the two positions in the sky with the addition of a variable offset, due to the differences of atmospheric and ground radiation in the two beams. To eliminate this variable offset, the telescope is periodically moved by 4.5 arc min to place a given sky position (called the field) alternately in the two beams. This gives two sets of data for each field

$$\delta T_{on} = T_{Field} - T_{ref2} + T_{offset\text{-}on}$$

and (1)

$$\delta T_{off} = T_{ref1} - T_{Field} + T_{offset\text{-}off}$$

where T_{ref1} and T_{ref2} correspond respectively to the westerly and easterly reference positions, 4.5 arc min away from the corresponding field. The two offset signals in equation (1) cannot be measured with sufficient accuracy to subtract them; furthermore, they are slowly varying functions of time and telescope position. By accurately timing the off-on observing sequence, one hopes that both offsets will be equal and will cancel in the expression

$$\Delta T_{Field} = (\delta T_{on} - \delta T_{off})/2$$
$$= T_{Field} - (T_{ref1} + T_{ref2})/2 \quad (2)$$

which is the quantity measured for each field. The atmospheric and ground effects are too complex to predict at the 10^{-4} K level, so one must try and see what happens.

The 12 candidate fields were chosen as close to the north celestial pole as safe operation of the 140-foot telescope allows: declination (1950.0) 86°51′ with right ascensions: 1 h 30 min, 3 h 30 min, . . . , 23 h 30 min. This allows 2 h of observation of each field on each day, while minimizing the motion of the telescope with respect to the ground and the atmosphere. At this declination and with a 4.5 arc min beam separation, the telescope goes through the same range of positions with respect to the ground for an off-on pair if the on-scan is started 5 min 3 s after the off-scan. As changing the position of the telescope takes about 15 s, the best cancellation of T_{offset} is achieved with an integration time of 4 min 48 s for each scan. We emphasize that keeping the residual systematic errors below 10^{-4} K requires the atmospheric and ground contributions to be repeatable to better than one part in 10^5 over times longer than the 10 min duration of the off-on sequence. It is remarkable that, with good weather conditions, this requirement is indeed satisfied with this particular telescope.

After correcting the data for the beam efficiency of the antenna (55%) and converting to units of thermodynamic temperature, we compute the mean values ΔT^i_{Field} for the observations of each field on each day (i). The associated standard deviations, σ^i_{Field}, are calculated from the scatter among the different off-on measurements (typically 12 on each day), and are an estimate of the statistical uncertainty in the mean, after 2 h of observation. A typical value for data taken with clear sky is $\sigma^i_{Field} = 0.25 \times 10^{-3}$ K, whereas one expects 0.14×10^{-3} K from system noise alone. The excess noise at this stage is due to a $1/f$ component in the noise power spectrum of the receiver and to the variable components of the atmospheric and ground contributions to δT_{on} and δT_{off}.

The weighted average of the values of ΔT^i_{Field} from the different days in which each field was observed yield the final values $\overline{\Delta T}_{Field}$ plotted in Fig. 1. The errors associated with each point correspond to the standard deviations, σ_{Field}, estimated from the measurement statistics in the usual way. It is crucial to estimate accurately the errors in Fig. 1 because the data will be used to decide whether a significant sky signal has been detected or, alternatively, at what level a sky signal is excluded. A major concern is whether ΔT^i_{Field} changes from day to day owing to slow changes in $T_{offset\text{-}on} - T_{offset\text{-}off}$. This can be checked by comparing the measured day-to-day scatter in ΔT^i_{Field} with the corresponding values of σ^i_{Field}. Figure 2 shows a histogram of the weighted residuals $(\Delta T^i_{Field} - \overline{\Delta T}_{Field})/\sigma^i_{Field}$ for the 87 2-h observations. The histogram matches the gaussian distribution expected if all the day-to-day scatter were due to fluctuations

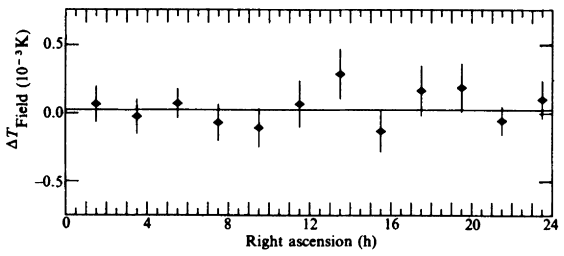


Fig. 1 Final averages for the observed fields. The solid line indicates the overall weighted average, which is $\Delta T_{ave} = (26 \pm 39) \times 10^{-6}$ K. $\overline{\Delta T}_{Field}$ is the difference in thermodynamic temperature between the field and the average of two reference positions, 4.5 arc min away on opposite sides. (See equation (2).)

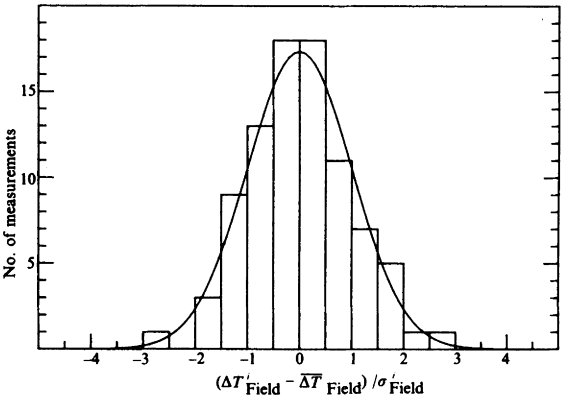


Fig. 2 The day-to-day scatter in measurements averaged over ~2 h. As the residuals from the field averages ΔT_{Field} are used, the possible sky signals have been removed. If there were no extraneous day-to-day effects, the histogram would fit the gaussian curve shown; it does so reasonably well.

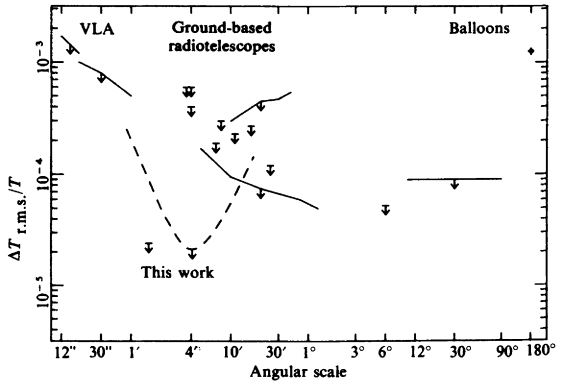


Fig. 3 ↧, Upper limits (95% confidence level) on anisotropy of the 2.7 K radiation at various angular scales. The Very Large Array (VLA) results[3,4] are new and will probably be improved. Some of the points between 1′ and 1° have been revised[1] from the values given in the original references. Anisotropy at large angular scales[8] must be measured from balloons or satellites because of atmospheric noise.

contained in $\sigma^{\perp}_{\text{Field}}$. Indeed, the corresponding χ^2 is 78 with 75 degrees of freedom. Therefore the values of σ_{Field} displayed in Fig. 1 are not contaminated by day-to-day effects and are accurate estimates of the measurement errors.

The weighted average of the values in Fig. 1 is $\Delta T_{\text{ave}} = (26 \pm 39) \times 10^{-6}$ K. As none of the 12 points differs significantly from this value, no anisotropy has been detected at a level of $\sim 2\bar{\sigma}_{\text{Field}}$. A more stringent limit can be set by using all the data in a statistical argument, although this requires adopting a model for the anisotropy. The underlying sky signal is expected to be gaussian, distributed between the 12 field positions (separation $>1.6°$), and can be characterized by its standard deviation, σ_{sky}. Assuming there is no correlation between the signals arriving from a given field and associated reference positions (4.5 arc min scale), the r.m.s. residual in $\overline{\Delta T}_{\text{Field}}$ due to a true sky signal is (from equation 2)

$$\sigma^2_{\text{r.m.s.}} = (4\,\sigma^2_{\text{sky}} + \sigma^2_{\text{sky}} + \sigma^2_{\text{sky}})/4 = 1.5\,\sigma^2_{\text{sky}} \qquad (3)$$

In this case, the off-on procedure is a probe of σ_{sky} with $\sqrt{1.5}$ efficiency. This efficiency factor depends on the angular autocorrelation that is assumed for the possible sky signal and must be evaluated separately for each different model.

The Neyman–Pearson lemma prescribes the optimal statistic for testing a hypothetical value of σ_{sky}. In our case the best statistic is essentially a weighted χ^2 (ref. 2), which yields

$$\sigma_{\text{r.m.s.}} = (0.1 \pm 4.7) \times 10^{-5}\ \text{K} \qquad (4)$$

and an upper limit

$$\sigma_{\text{r.m.s.}} < 7.8 \times 10^{-5}\ \text{K (95\% confidence level)} \qquad (5)$$

Using equation (3) and $T = 2.7$ K gives

$$\sigma_{\text{sky}}/T < 2.4 \times 10^{-5}\ \text{(95\% confidence level)} \qquad (6)$$

Limits on anisotropy from different experiments are difficult to compare. Because various scanning patterns and statistical methods have been used, careful statements of limits on a particular model should examine each measurement separately. Figure 3 shows the present experimental limits, most of which assume that the underlying anisotropy is uncorrelated gaussian noise. For this assumption the result of equation (6) applies at an angular scale of 1.5 arc min.

A more general way to interpret our result is to take the Fourier transform of the scan pattern (three gaussian beams 1.5 arc min across, separated by 4.5 arc min) and then to average it over the orientation angle. Thus, the limit shown in Fig. 3 (dashed line) assumes that all of the anisotropy is the result of gaussian sinusoidal fluctuations of a certain scale. The strongest limit corresponds to a scale (half wavelength) of 4 arc min and is

$$\Delta T_{\text{r.m.s.}}/T < 2.1 \times 10^{-5}\ \text{(95\% confidence level)} \qquad (7)$$

which is more stringent than the one given in equation (6) because the sinusoidal fluctuations are anticorrelated on the scale of our beam-throw, increasing the efficiency of the observing procedure for this kind of fluctuation.

A physical interpretation of the results requires a model for the nature of the underlying fluctuations. Most authors assume fluctuations with a spectrum of the form $(\delta\rho/\rho) \propto k^n$, for the dominant contribution to the mass of the Universe at the time of decoupling of matter and radiation (a redshift of $\sim$1,000). The amplitudes of the fluctuations are fixed by matching the evolved autocorrelation function of the matter to the data available from studies of galaxy clustering. The details of the interaction between the mass components and the background radiation determine the anisotropy required by each particular model. Comparing our results with recent calculations[5] rules out those models in which the initial fluctuations are adiabatic and dominated by baryonic matter. Baryon-dominated universes in which the initial perturbations are isothermal are currently out of fashion but are not seriously constrained by our limit: for $n < 0$ one needs to satisfy $0.1 < \Omega < 0.3$ (Ω is the ratio of the density of the Universe to that of an Einstein–de Sitter universe), whereas no restriction is imposed if $n \geq 0$. Models dominated by massive neutrinos require $\Omega \sim 1$, although they suffer other problems[6,7]. More exotic possibilities are models in which the dominant component is in the form of yet undiscovered cold, dark matter, such as axions or photinos. These models require that $\Omega \cdot H > 0.2$ (H is Hubble's constant in units of 100 km s^{-1} Mpc^{-1}).

Finally, we must caution that the analysis of our data rests on the assumption that the underlying fluctuations have a gaussian distribution: if untrue, the theoretical predictions could be uncertain by factors of 2. Although there is a conflict between the theoretical models and the experimental data, the fate of the various models is still far from settled.

We thank Drs C. Brockway and R. Fisher for their continuous improvement of the telescope, receiver and data acquisition system, and all the NRAO–Green Bank staff for their help and constant vigilance. This research was supported in part by the NSF. The National Radio Astronomy Observatory of the USA is operated by Associated Universities, Inc. under contract with the NSF.

Received 20 August; accepted 18 September 1984.

1. Lasenby, A. N. & Davies, R. D. *Mon. Not. R. astr. Soc.* **203**, 1127–1169 (1983).
2. Uson, J. M. & Wilkinson, D. T. *Astrophys. J.* **283**, 471–478 (1984).
3. Fomalont, E. B., Kellermann, K. I. & Wall, J. V. *Astrophys J. Lett.* **277**, 23–26 (1984).
4. Knocke, J. E., Partridge, R. B., Ratner, M. I. & Shapiro, I. I. *Astrophys J.* (in the press).
5. Silk, J. *Proc. Inner Space/Outer Space Conf.*, Fermilab (University of Chicago Press, in the press).
6. Peebles, P. J. E. *Astrophys J.* **258**, 415–424 (1982).
7. White, S. D. M., Frenk, C. S. & Davis, M. *Astrophys J. Lett.* **274**, 1–5 (1983).
8. Wilkinson, D. T. *Phil. Trans. R. Soc.* **A307**, 55–65 (1982).

THE ASTROPHYSICAL JOURNAL, 298:L1–L5, 1985 November 1
© 1985. The American Astronomical Society. All rights reserved. Printed in U.S.A.

A MAP OF THE COSMIC BACKGROUND RADIATION AT 3 MILLIMETERS

PHILIP LUBIN, THYRSO VILLELA,[1] GERALD EPSTEIN, AND GEORGE SMOOT
Space Sciences Laboratory and Lawrence Berkeley Laboratory, University of California, Berkeley
Received 1985 June 17; accepted 1985 August 13

ABSTRACT

We present data from a series of balloon flights covering both the northern and southern hemispheres measuring the large angular scale anisotropy in the cosmic background radiation at 3.3 mm wavelength. The data cover 85% of the sky to a limiting sensitivity of 0.7 mK per 7° field of view. The data show a 50 σ (statistical error only) dipole anisotropy with an amplitude of 3.44 ± 0.17 mK and a direction of $\alpha = 11^{h}.2 \pm 0^{h}.1$ and $\delta = -6^{\circ}.0 \pm 1^{\circ}.5$. A 90% confidence level upper limit of 7×10^{-5} is obtained for the rms quadrupole amplitude. Flights separated by 6 months show the motion of Earth around the Sun. Galactic contamination is very small with less than 0.1 mK contribution to the dipole and quadrupole terms. A map of the sky has been generated from the data.

Subject heading: cosmic background radiation

The isotropy of the cosmic background radiation is one of the fundamental tests of modern cosmological theories. The angular distribution has been measured on scales from arc seconds to 180°. The only variations seen are the dipole at large angular scales, presumably due to our motion, and perhaps an effect due to the scattering of the radiation by free electrons in galactic clusters at small angular scales (Sunyaev-Zel'dovich effect). Neither of these effects is due to an intrinsic cosmological perturbation. The lack of any observed intrinsic anisotropy on any angular scale is disturbing and has contributed to recent modifications of the standard model such as the inflationary universe models, which are more consistent with observations (Bond and Efstathiou 1984; Vittorio and Silk 1984).

We have made a series of observations on the large-scale ($> 7°$) anisotropy in the background radiation at 3.3 mm (90 GHz) wavelength. The instrument consists of a liquid-helium–cooled Schottky diode receiver which measures the temperature difference between two 7° fields of view 90° apart in the sky. The apparatus is flown from a balloon at an altitude of 30 km. The instrument is described in detail in Lubin (1982), Lubin, Epstein, and Smoot (1983), and Epstein (1983).

The system has a sensitivity of about 13 mK $Hz^{-1/2}$, a noise temperature of 130 K, and a bandwidth of 600 MHz with a beam width of 7° FWHM. The gondola rotates with a period of about 100 s. The rotation of the gondola and Earth yields a sky coverage of about 30% in one 12 hr flight. An inherent instrument asymmetry is caused by the beam chopper, which consists of a sector of highly polished aluminum. The emissivity of a piece of metal with our polarization orientation is $\varepsilon = (4\pi\delta/\lambda_0)\cos\theta$, where θ is the angle of incidence, 45° for us, and δ is the skin depth. The conductivity of aluminum at the ambient flight air temperature of 240 K gives an emissivity of 6×10^{-4}. The observed inflight asymmetry (offset) was 200 mK giving an emissivity of about 8×10^{-4}, in relatively good agreement. The offset was stable to about 1 mK hr^{-1} during the flight.

Calibration was performed inflight every 27 minutes. In the first two data flights, a small blackbody target was used. Though it provided good relative calibration, it was difficult to use as an absolute calibrator to better than about 5%. On the third data flight (Brazil), it was replaced by a full beam calibrator which appeared to be capable of a 1%–2% calibration. Unfortunately, the gondola was lost in this last flight. Though the data were recovered, the full calibration accuracy has not been achieved due to the loss of the flight tape from the gondola tape recorder. The Moon has also been used as an inflight calibration source. In particular, during the last flight, the beam passed directly over the Moon, permitting a good comparison between the full beam calibrator and the Moon. The lunar emission has been analyzed using a model of the Moon due to Keihm (1984). The calibration using the Moon agrees with the calibration obtained using the target calibrator within 3%.

The preflight and postflight calibrations were the same within 2% for all flights except the last flight where no postflight calibration was possible. The inflight calibration stability was ±1% over all flights. The overall calibration uncertainty was taken as 5%. Because of the high sensitivity of the experiment, this uncertainty dominates the error in the dipole amplitude, though it has negligible effect on the higher order moments (quadrupole, etc).

Galactic emission is a problem for large-scale anisotropy measurements because of the extended size of the Galaxy and uncertainty in modeling its emission. At long wavelengths ($\lambda > 1$ cm) synchrotron and bremsstrahlung radiation dominate, while at short wavelengths ($\lambda < 1$ mm) dust emission dominates. Typical Galactic emission as a function of wavelength is shown in Figure 1. This estimate is based on low-frequency ground-based surveys for the synchrotron and

[1]Also Instituto de Pesquisas Espaciais/CNPq, Departamento de Astrofísica, São José dos Campos, SP, Brazil.

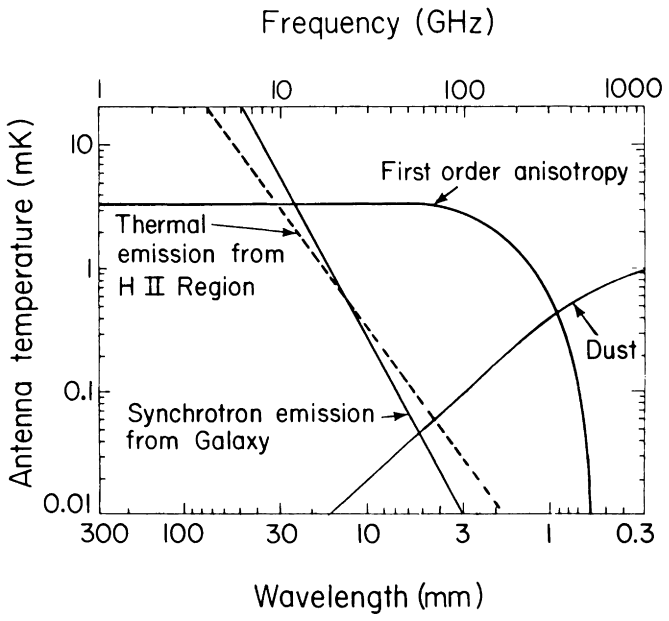


FIG. 1.—Estimated galactic emission. See Lubin and Villela (1984) for more details and comparison of recent measurements.

bremsstrahlung components and from submillimeter balloon measurements (Owens, Muehlner, and Weiss 1979) combined with our 3.3 mm data for the dust. A natural minimum, or window, occurs around 3 mm where our measurement was made. More recent measurements (Halpern 1983) suggest a somewhat flatter dust spectrum but with a minimum galactic contribution still near 3 mm (Lubin and Villela 1984).

At 30 km, the residual atmospheric pressure is 10 mb. For our frequency and bandwidth, the vertical atmospheric emission is calculated to be 6 mK (Lubin and Villela 1984) based on the line strengths of oxygen, water, and ozone and a model of their atmospheric distribution and temperature profile. To first order, the atmospheric contribution cancels out because the beam is symmetrical with respect to the zenith. Gondola wobble is measured with the magnetometers to be less than $\pm 0.^{\circ}5$, which combined with our offset gives a systematic error of less than 0.1 mK. The low sidelobe antenna and dual ground shields reduced the contribution of radiation from Earth to less than 0.1 mK.

The instrument has been flown four times. The first flight in 1981 July was on the gondola of D. Wilkinson of Princeton. Due to interference, this flight did not return useful data. The second flight, in 1981 November, was on the gondola of R. Weiss of MIT. The third flight, in 1982 April, and the fourth flight, in 1982 November, used our own gondola. The first three flights were from Palestine, Texas, at the National Scientific Balloon Facility. The fourth flight was from Cachoeira Paulista, SP, Brazil, at the Instituto de Pesquisas Espaciais. The sky coverage from the three data flights is shown in Figure 2 on a Hammer-Aitoff projection. Sky coverage is now 85% with coverage from -68° to $+77^{\circ}$ declination. The data from the Brazilian flight were taken from the telemetry tapes. The tapes were digitized and the data were calibrated using both the inflight observation of the Moon and the inflight target calibrator. The calibration was constant to $\pm 1\%$ during the flight.

The dipole obtained from the Brazilian flight only is $\alpha = 11^{h}.14 \pm 0^{h}.20$, $\delta = -8.^{\circ}2 \pm 3.^{\circ}0$ and $\Delta T_A = 2.86 \pm 0.14$ mK, in good agreement with the northern hemisphere data (Epstein 1983) of $\alpha = 11^{h}.25 \pm 0^{h}.10$, $\delta = -5.^{\circ}7 \pm 1.^{\circ}8$ and $\Delta T_A = 2.83 \pm 0.14$ mK, where ΔT_A is the "antenna temperature." The fit to the dipole and quadrupole spherical harmonics is summarized in Table 1. As can be seen, no significant quadrupole is present. The quadrupole fits imply a 90% confidence level upper limit of 7×10^{-5} for the rms quadrupole component. The latest Princeton 25 GHz maser data (Fixsen, Cheng, and Wilkinson 1983) and these and our previous data (Lubin, Epstein, and Smoot 1983) all are consistent with quadrupole upper limits at the 7×10^{-5} level and appear to rule out the previous quadrupole reports. A fit to a galactic dust model with a cosec b (b is galactic latitude) distribution flattened at $b = 5^{\circ}$ yields 44 ± 11 μK at the poles. While this appears to be significant, it should be interpreted cautiously. Excluding data within 5° of the Galactic plane fit yields an insignificant 14 ± 19 μK, indicating that the previous fit is dominated by emission near the Galactic equator as would be expected. Galactic emission appears to be a very small effect in these data as can be seen by the very small change in the multipole amplitudes when the Galactic plane is omitted.

The inferred velocity of the Sun is 382 ± 25 km s^{-1} toward $\alpha = 11^{h}.2 \pm 0^{h}.1$ and $\delta = -6.^{\circ}0 \pm 1.^{\circ}5$ for a background radiation temperature of 2.7 K. Using a Galactic solar velocity of 220 km s^{-1} toward $l^{II} = 90^{\circ}$, $b^{II} = 0^{\circ}$ gives a Galactic (Local Group) velocity of $V_{LG} = 550 \pm 40$ km s^{-1} toward $\alpha = 10^{h}.7$

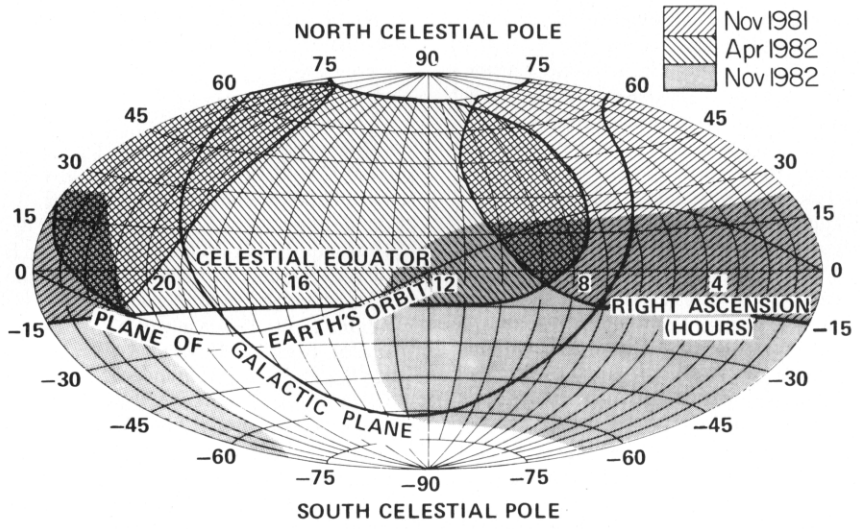


FIG. 2.—Sky coverage of three data flights by date

TABLE 1

SPHERICAL HARMONICS FITS

Fit	Coefficients[a]	All Data	Galactic Cut 5°	Statistical Error	Total Error
Dipole	T_x	−3.36	−3.36	0.07	0.17
	T_y	0.62	0.71	0.07	0.10
	T_z	−0.37	−0.36	0.06	0.08
Dipole and Quadrupole ...	T_x	−3.37	−3.37	0.09	0.17
	T_y	0.63	0.69	0.09	0.09
	T_z	−0.49	−0.42	0.09	0.09
	Q_1	0.21	0.10	0.09	0.09
	Q_2	0.27	0.15	0.09	0.10
	Q_3	0.16	0.12	0.10	0.11
	Q_4	−0.10	−0.09	0.07	0.09
	Q_5	0.05	0.09	0.06	0.08

NOTE.—Values are mK thermodynamic temperature for $T = 2.7$ K.

[a] $T = T_x \cos\delta\cos\alpha + T_y \cos\delta \sin\alpha + T_z \sin\delta + Q_1(3\sin^2\delta - 1)/2 + Q_2 \sin 2\delta \cos\alpha + Q_3 \sin 2\delta \sin\alpha + Q_4 \cos^2\delta \cos 2\alpha + Q_5 \cos^2\delta \sin 2\alpha$.

$\pm 0^h3$, $\delta = -22° \pm 5°$. This gives an angle of 44° between the center of the Virgo Cluster and V_{LG}.

The orbital velocity of Earth around the Sun is roughly 30 km s^{-1} ($\beta \approx 10^{-4}$) or 10% of Earth's velocity relative to the cosmic background radiation ($\beta \approx 10^{-3}$). Flights separated by 6 months should be able to detect Earth's orbital motion as a 10% change in the dipole. Our 1982 April and November 1982 flights clearly show this effect. The dipole difference vector between the flights is:

$$(\Delta T_x, \Delta T_y, \Delta T_z) = (-0.30 \pm 0.14, 0.40 \pm 0.14, 0.08 \pm 0.14),$$

where the numbers are antenna temperature in mK. The predicted difference based on Earth's orbital motion and assuming a 2.7 K background radiation temperature is $(-0.32, 0.28, 0.12)$ which is in good agreement.

The data set consists of approximately 10^5 measurements of temperature differences distributed over 85% of the celestial sphere. To convert these different measurements into a map of the sky, we use a technique suggested by Gulkis and Janssen (1981). In essence, the idea is to find a sky distribution which minimizes

$$\chi^2 = \sum \frac{1}{\sigma_{ij}^2}(T_i - T_j - \Delta T_{ij})^2,$$

where T_i and T_j are the map temperatures desired at pixels i and j and T_{ij} is the measured temperature difference between pixels i and j. In this procedure, the T_i are free variables. Minimizing χ^2 with respect to T_i yields a set of N linear equations, where N is the number of pixel elements. These equations take the form:

$$\sum_1^N \frac{1}{\sigma_{ij}^2}(T_i - T_j - \Delta T_{ij}) = 0.$$

Our integrated beam pattern yields a solid angle of about 70 deg^2. This would give about 600 pixels for the full sky. We split each pixel into four parts giving about 2400 pixels for the full sky or 2000 pixels for our 85% sky coverage. Solving these 2000 linear equations is tantamount to inverting a 2000 by 2000 matrix. Instead of a direct matrix inversion, we use an iterative Gauss-Sidel inversion method which is efficient at iteratively inverting a large matrix. In addition, it converges rapidly, typically within 100 iterations. A subsidiary condition is also needed since we do not measure the absolute temperature. We chose to set the weighted average to 0. Because the data are essentially taken at constant latitude, and each pixel is connected by observations to a few other pixels (the matrix

is sparse and nearly singular), certain mathematical instabilities can occur in the map. For this reason, we have run numerous Monte Carlo simulations to test the stability of the maps. These tests show that on large scales the maps are a faithful representation of the data. A map made from the data is shown in Figure 3 (Plate L1) as a $\sin\delta$ projection which is an equal area projection. The map clearly shows the dipole anisotropy. The residual map with the dipole removed is shown in Figure 4 (Plate L2). No significant structure appears to be left. No corrections for Galactic emission have been made in the maps.

A summary of recent dipole measurements is shown in Figure 5. Recent measurements have been limited by about a 5% calibration error which is included in Figure 5. Combining the dipole measurements of the Princeton group at 25 GHz and these data at 90 GHz gives a spectrum which is consistent with a 2.7 K blackbody (Lubin and Villela 1984). Further analysis of our recently recovered Brazilian package may improve the calibration from 5% to 2% which would give a better idea of the spectrum (Villela 1985).

As large-scale anisotropy measurements are being fundamentally limited by systematic errors and particularly by Galactic emission, it is of interest to consider how much better the next generation of experiments will be. New technologies, such as superconducting receivers and ^{3}He-cooled bolometers, may give an order of magnitude increase in system sensitivity. To take advantage of this, though, a much better understanding of Galactic emission, particularly at the millimeter wavelengths, will be necessary. With the recent launch of the *Prognoz 9* satellite, space-based measurements are now possible. These have the advantage of no atmosphere and a longer observing time but generally do not use state-of-art techniques. Later in this decade the COBE satellite will be launched and should give us a much better understanding of Galactic emission and about a factor of 5 increase in sensitivity over the data presented here.

This work was supported by the California Space Institute under CS 48-81, by the National Aeronautics and Space Administration under NAGW-66, by the National Science Foundation under SPI 8166057, and by the US Department of Energy under DE-AC03-76SF00098. G. E. acknowledges the generous support of the Fannie and John Hertz Foundation. T. V. acknowledges the support and assistance of Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazilian Government research agency, Brazil, under 202372/83-FA, and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), State of São Paulo research agency, Brazil, under 83/9834-1-Astronomia, during his research stay. The northern hemisphere balloon launches were conducted by NASA at the National Scientific Balloon Facility, Palestine, Texas. The southern hemisphere flight was supported and conducted by INPE/CNPq at its balloon launching facility at Cachoeira Paulista, SP, Brazil. This project would not have been fruitful without the fabrication assistance of Hal Dougherty and John Gibson and the encouragement and assistance of D. Wilkinson on whose gondola we flew for the first time. We appreciate the perseverance of the MIT group and especially of R. Weiss in allowing us to fly our "MIT/Bezerkeley Lunar Lander." Special thanks go to P. Price, N. Boggess, R. Kubara, and also to the staff of the National Scientific Balloon Facility for their assistance and hospitality. Agradecemos os esforços e assistência do Departamento de Astrofísica do INPE durante o vôo de novembro de 1982 e o desenvolvimento subsequente deste trabalho.

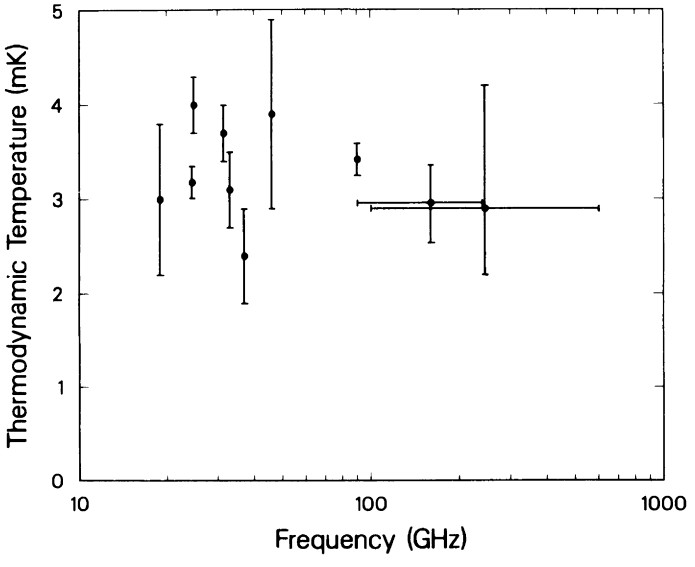


FIG. 5.—Recent dipole measurements vs. frequency. This work at 90 GHz.

REFERENCES

Bond, J. R., and Efstathiou, G. 1984, *Ap. J. (Letters)*, **285**, L44.
Epstein, G. L. 1983, Ph.D. thesis, University of California, Berkeley.
Fixsen, D. J., Cheng, E. S., and Wilkinson, D. T., 1983, *Phys. Rev. Letters*, **50**, 620.
Gulkis, S., and Janssen, M. 1981, NASA COBE Report 4004.
Halpern, M. 1983, Ph.D. thesis, Massachussetts Institute of Technology.
Keihm, S. 1984, private communication.
Lubin, P. 1982, in *Proc. of the International School of Physics "Enrico Fermi," Course on Gamow Cosmology*, Varenna, Italy, in press.
Lubin, P., Epstein, G., and Smoot, G. 1983, *Phys. Rev. Letters*, **50**, 616.
Lubin, P., and Villela, T. 1984, in *Proc. of the Third Rome Meeting on Astrophysics*, Rome, Italy, in press.
Owens, D. K., Muehlner, D. J., and Weiss, R. 1979, *Ap. J.*, **231**, 702.
Villela, T. 1985, Ph.D. thesis, in preparation.
Vittorio, N., and Silk, J. 1984, *Ap. J.*, **285**, L39.

GERALD EPSTEIN: Office of Technology Assessment, US Congress, Washington, DC 20510

PHILIP LUBIN, GEORGE SMOOT, and THYRSO VILLELA: Space Sciences Laboratory and Lawrence Berkeley Laboratory, University of California, Berkeley, CA 94720

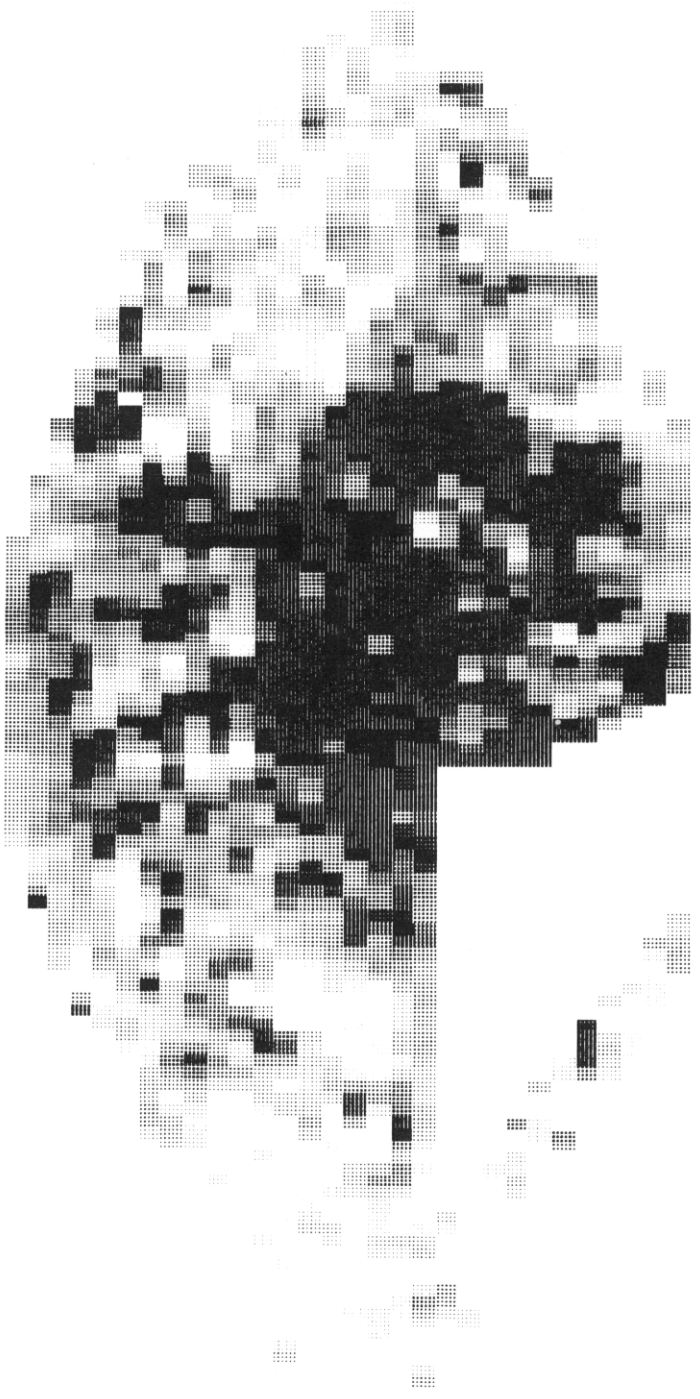


FIG. 3.—Map at 3 mm in celestial coordinates (α, δ). $\alpha = +24^h$ at left and 0^h at right. $\delta = +77°$ at the top and $-68°$ at the bottom. Projection is $\sin\delta$. Pixels are nominally $3° \times 5°$ $(\alpha \times \delta)$.

LUBIN, VILLELA, EPSTEIN, AND SMOOT (*see* page L4)

PLATE L2

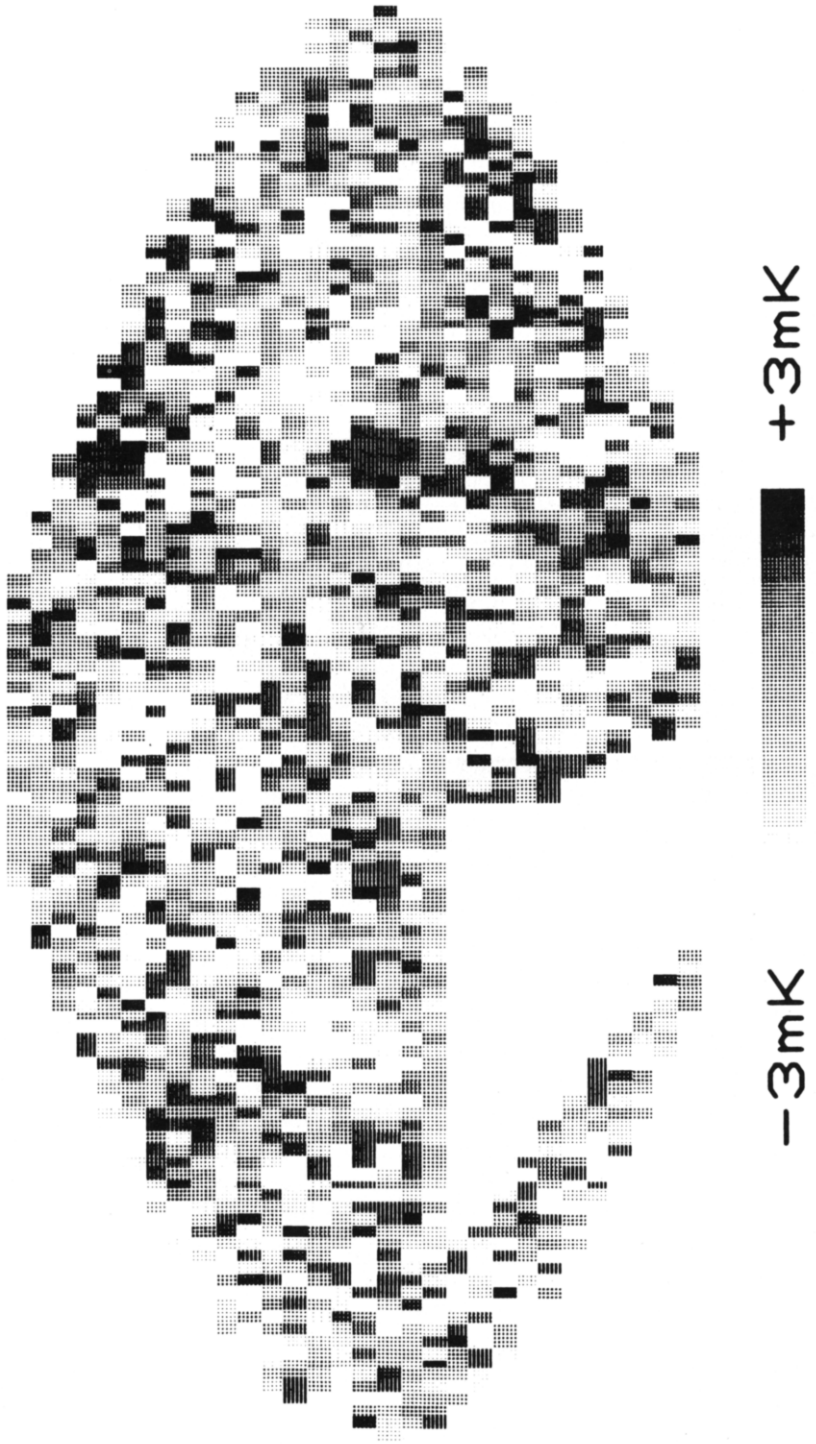


FIG. 4.—Residual map after subtraction of the dipole. Other parameters are same as Fig. 3.

LUBIN, VILLELA, EPSTEIN, AND SMOOT (*see* page L4)

COSMOLOGY OF THE INVISIBLE AXION

John PRESKILL [1], Mark B. WISE [2]
Lyman Laboratory of Physics, Harvard University, Cambridge, MA 02138, USA

and

Frank WILCZEK
Institute for Theoretical Physics, University of California, Santa Barbara, CA 93106, USA

Received 10 September 1982

We identify a new cosmological problem for models which solve the strong *CP* puzzle with an invisible axion, unrelated to the domain wall problem. Because the axion is very weakly coupled, the energy density stored in the oscillations of the classical axion field does not dissipate rapidly; it exceeds the critical density needed to close the universe unless $f_a \leqslant 10^{12}$ GeV, where f_a is the axion decay constant. If this bound is saturated, axions may comprise the dark matter of the universe.

Ever since the discovery [1] of the *CP*-violating parameter $\bar{\theta}$ of the strong interactions, the small value of $\bar{\theta}$ has posed a serious puzzle. Current experimental limits [2] on the electric dipole moment of the neutron require [3] $\lesssim 10^{-9}$. The small value of $\bar{\theta}$ is naturally explained in models with a spontaneously broken Peccei–Quinn symmetry [4], but the resulting axion [5] has not been seen. In a recently proposed variant of the Peccei–Quinn scheme [6,7], the axion mass and its couplings to ordinary matter are inversely proportional to the large vacuum expectation value f_a of an $[SU(2) \times U(1)]_W$-singlet scalar field. For $f_a \gtrsim 10$ TeV, this axion is consistent with all known laboratory experiments, but a more stringent constraint on f_a comes from astrophysical considerations [8]. To suppress the power radiated in axions by the helium core of a red giant star to an acceptable level, it is required that $f_a \gtrsim 10^9$ GeV. Since f_a must be very large anyway, one is tempted to suppose [7] that $f_a \sim 10^{15}$ GeV, the mass scale of grand unification [9].

In this paper, we point out a new cosmological constraint on the "invisible" axion satisfying $f_a \gtrsim 10^9$ GeV. Because the couplings of this axion are so weak, the energy density stored in the oscillations of the classical axion field does not dissipate rapidly. To prevent this energy density from exceeding the critical density we must require $f_a \lesssim 10^{12}$ GeV; the invisible axion must not be too visible. The most interesting possibility is $f_a \sim 10^{12}$ GeV, for in that case a nondissipative, pressureless axion gas could constitute the dark matter of the universe.

The axion is a pseudo-Goldstone boson; all of its couplings are suppressed by inverse powers of the axion decay constant f_a. Its only nonderivative coupling, to $\mathrm{tr}(F\tilde{F})$, arises from the color anomaly of the Peccei–Quinn current. In the effective field theory describing physics at energies below f_a, the action of the nonlinearly realized $U(1)_{PQ}$ Peccei–Quinn symmetry on the axion field a is

$$a/f_a \to a/f_a + \theta. \tag{1}$$

We have chosen to normalize f_a so that a global rotation of a/f_a is equivalent to a rotation of the QCD angle θ; a/f_a must therefore be regarded as a periodic variable with period $2\pi N$, where N is the number of elements in the discrete subgroup of the $U(1)_{PQ}$ symmetry group which is not explicitly broken by nonperturbative QCD effects.

[1] Alfred P. Sloan Fellow.
[2] Junior Fellow, Harvard Society of Fellows. On leave from California Institute of Technology, Pasadena, CA 91125, USA.

0 031-9163/83/0000–0000/$ 03.00 © 1983 North-Holland

In the standard big-bang cosmology, the axion is in thermal equilibrium at temperatures $T > f_a$. At temperatures satisfying $f_a \gg T$ and $H \gg m_a$, where H is the Hubble parameter and m_a is the axion mass, the axion is decoupled, and behaves like a free massless scalar field. Fluctuations in the angle a/f_a averaged over a volume λ^3 are expected to be of order one if λ exceeds the horizon size t. The fluctuations begin to red-shift away when the wavelength enters the horizon [10]; thereafter, they decay like $\delta a/f_a \sim \lambda/t$. Hence a/f_a is approximately constant over length scales which are smaller than the horizon size, and assumes an arbitrarily chosen value which is typically of order one. In the new inflationary cosmology [11], the size of a causally connected region is enormous, and the background value of a/f_a can be effectively constant throughout the universe, provided that the inflation occurs after the spontaneous breakdown of the Peccei–Quinn symmetry. (There *are* fluctuations in the background value, due to the Hawking temperature.) The fluctuations about this background value, generated by reheating, are small. ‡1

The nonderivative couplings of the axion can be described by an effective lagrangian

$$\mathcal{L} = \tfrac{1}{2}(\partial_\mu a)^2 - f_a^2 m_a^2 F(a/f_a), \tag{2}$$

where F is a dimensionless function with period 2π and the Taylor expansion $F(x) = \frac{1}{2}x^2 + \ldots$. ($F$ is also a function of the temperature T.) Because a/f_a is a periodic variable with period $2\pi N$, the potential in eq. (2) has N distinct minima, and if the axion field is at any one of these minima, $\bar{\theta} \simeq 0$ and the *CP* violation of the strong interactions is acceptably small. However, as noted by Sikivie [12], if $N > 1$ we expect the axion field to choose different minima in causally disconnected regions of space, giving rise to axion domains separated by walls, in conflict with standard cosmology [13].

‡1 If the axion fluctuations on all scales are in thermal equilibrium at the reheating temperature T_{RH}, so that $\delta a \sim (T_{RH}/\lambda)^{1/2}$ initially, then fluctuations entering the "horizon" H^{-1} at a lower temperature T are larger than thermal fluctuations by the factor T_{RH}/T; thermal equilibrium is not maintained by the expansion for $\lambda > H^{-1}$, because the axion is not conformally coupled. Averaging over a "horizon" volume, we have $\delta a^2/f_a^2 \sim (T_{RH}/f_a)^2 \times (H/T)$. These fluctuations are very small, if $T_{RG} \lesssim f_a$ and H^{-1} is much larger than a thermal wavelength.

The problem pointed out by Sikivie can be avoided in various ways. Several authors have shown that it is possible to construct models in which N is effectively 1, so that there are no domains [14]. Furthermore, in the new inflationary cosmology, it is expected that a/f_a is nearly constant throughout the visible universe, if the inflation occurs after the Peccei–Quinn symmetry is spontaneously broken; it is therefore possible for the entire visible universe to lie within a single domain.

But there is another possible cosmological problem with the axion which has not been discussed previously. Whether or not the axion potential in eq. (2) has multiple vacua, and whether or not the universe has undergone a period of exponential inflation, the expectation value $\langle a \rangle$ of the axion field has no reason to be at the minimum of the potential when the universe cools to a temperature such that $H \sim m_a$. Typically, we expect $\langle a \rangle \sim f_a$. Thus, the axion field has a potential energy density of order $f_a^2 m_a^2$, and begins to oscillate coherently.

There are two mechanisms by which the energy stored in these oscillations can be dissipated – Hubble expansion and particle production. We will show that, for $f_a > 10^{13}$ GeV, neither mechanism can prevent the axion energy density from exceeding the critical density needed to close the universe, assuming that $\langle a \rangle \sim f_a$ initially.

We first neglect particle production and analyze the effect of the Hubble expansion; it will be shown below that particle production really is negligible. Because the other modes red-shift away rapidly, we consider only the zero-momentum mode of the axion field. For values of $\langle a \rangle / f_a$ which are somewhat smaller than one, it is reasonable to approximate the potential in eq. (2) by the quadratic term, and the equation satisfied by the zero-momentum mode is

$$\mathrm{d}^2\langle a \rangle/\mathrm{d}t^2 + 3H(t)\mathrm{d}\langle a \rangle/\mathrm{d}t + m_a^2(t)\langle a \rangle = 0. \tag{3}$$

As will be described in more detail below, the axion mass m_a varies as the temperature changes, so we have allowed it to be time-dependent in eq. (3). If H and $\dot{m}_a/m_a$ are small compared to m_a, eq. (3) has an approximate solution of the form

$$\langle a \rangle = A(t)\cos(m_a t), \tag{4}$$

where $A(t)$ satisfies

$$\mathrm{d}(m_a A^2)/\mathrm{d}t = -3H(m_a A^2). \tag{5}$$

Eq. (5) can be integrated, giving

$$(m_a A^2)_f/(m_a A^2)_i = (R_f/R_i)^{-3}, \quad (6)$$

where R is the Robertson–Walker scale factor. This result has a simple interpretation. The zero-momentum mode of the axion field has energy density $\rho_a = \frac{1}{2} m_a^2 A^2$, and may be regarded as a coherent state of axion particles at rest with number density $n_a = \frac{1}{2} m_a A^2$. Eq. (6) says that the number of zero-momentum axions per comoving volume in the expanding universe is conserved. This statement is true because we have neglected axion interactions and have assumed that $m_a(t)$ is so slowly varying that no additional axions can be produced. Eq. (6) begins to apply when the axion Compton wavelength comes within the "horizon" H^{-1}.

Using eq. (6), we can derive an expression for the axion energy density ρ_a as a function of the photon temperature T_f,

$$\rho_a = \tfrac{1}{2}(m_a)_i (m_a)_f f_a^2 (N_f/N_i)^{3/4} (T_f/T_i)^3 (A_i/f_a)^2. \quad (7)$$

Here N is the effective number of spin degrees of freedom, and the factor $(N_f/N_i)^{3/4}$ quantifies the effect of photon heating as the degrees of freedom freeze out. The amplitude A begins to decay according to eq. (6) when $m_a(t) \sim 3H(t)$. Thus we expect that A_i/f_a is of order one when T_i is chosen to satisfy

$$m_a(T_i) \sim 3H = (24\pi G\rho)^{1/2} = 39 T_i^2/m_p, \quad (8)$$

where $m_p = 1.2 \times 10^{19}$ GeV is the Planck mass. [The radiation species contributing to ρ in eq. (8) are assumed to be photons, gluons, three neutrinos, two charged leptons, and three quark flavors.] If the temperature T_f is much smaller than the electron mass, we have $(N_f/N_i)^{3/4} \sim 0.13$, recalling that neutrinos decouple before electron–positron annihilation. Therefore, combining (7) and (8) yields

$$\rho_a \simeq 2.5 (f_a/m_p)(m_a f_a/T_i) T_f^3 (A_i/f_a)^2, \quad (9)$$

where m_a is now the zero temperature value of the axion mass, and A_i/f_a is expected to be of order one.

To proceed with the evaluation of ρ_a using (9), we must solve eq. (8) for T_i; first we need to derive an expression for $m_a(T)$. A global rotation of a/f_a is equivalent to a rotation of the QCD angle θ. Therefore, the axion mass can be calculated in terms of the θ-dependent free energy density as

$$m_a^2(T) = (1/f_a^2) d^2F(T,\theta)/d\theta^2|_{\bar\theta=0}. \quad (10)$$

At high temperature, the θ-dependent part of the free energy density can be computed in the dilute-instanton-gas approximation. Using the methods of ref. [15], we find, for QCD with three quark flavors,

$$m_a(T) \sim (2 \times 10^{-2})(\Lambda^2/f_a)(m_u m_d m_s/\Lambda^3)^{1/2} \times (\Lambda/\pi T)^4 [9 \ln(\pi T/\Lambda)]^3, \quad (11)$$

where $m_{u,d,s}$ are the quark masses, and $\Lambda \sim 200$ MeV is the QCD scale factor.

At low temperature, the chiral symmetries of QCD are spontaneously broken, and the axion mass must be computed using the methods of current algebra; rotating θ causes the quark condensate to rotate also. Generalizing the calculation of ref. [16], the result is

$$m_a = (1/f_a)[m_u m_d)^{1/2}/(m_u + m_d)] f_\pi m_\pi, \quad (12)$$

assuming $m_{u,d} \ll m_s$ and choosing $f_\pi = 93$ MeV. To find the axion energy density today, we combine (9) and (12), making the substitutions $T_f = 2.7$ K and $(m_u m_d)^{1/2}/(m_u + m_d) \sim 0.5$. After dividing by the critical density, $\rho_c \sim 2 \times 10^{-29} h^2$ g cm^{-3}, we obtain

$$\Omega_a = \rho_a/\rho_c = 10^7 (f_a/m_p)(200 \text{ MeV}/T_i)(A_i/f_a)^2 h^{-2}, \quad (13)$$

where h is the Hubble constant in units of 100 km sec^{-1} Mpc^{-1}. By solving (8) and (11), we find that T_i is a slowly decreasing function of f_a/m_p which takes the value $T_i \sim 4\Lambda \sim 800$ MeV for $f_a/m_p \sim 4 \times 10^{-7}$. Therefore, the axion energy density exceeds the critical density unless

$$f_a \lesssim 4 \times 10^{12} \text{ GeV}, \quad (14)$$

if A_i/f_a is of order one. This bound is our main result.

It has been implicit in the above discussion that the change in m_a during the chiral phase transition can be regarded as adiabatic. The assumption is valid if the chiral phase transition is second-order in the chiral symmetry limit, for it is then very weakly first-order in the presence of small non-zero quark masses. If the phase transition actually occurs suddenly, however, Ω_a is larger by about a factor of four, and the limit on f_a is correspondingly more stringent, $f_a \lesssim 10^{12}$ GeV. Other sources of uncertainty are the Hubble constant h and the QCD parameter Λ = (200

MeV)λ. Roughly, our result is $f_a \lesssim (4 \times 10^{12}\ \text{GeV}) \times h^2\lambda^{3/4}$.

It should also be noted that (14) was derived by assuming the initial value of $\langle a\rangle/f_a$ to be of order one. In a conventional cosmological scenario, there are large fluctuations in a/f_a over distances exceeding the horizon size. The axion energy density begins to behave like nonrelativistic matter when the axion Compton wavelength comes within the horizon, and it is unavoidably true that at this time the energy density stored in the nonrelativistic axion modes is of order $f_a^2 m_a^2$ [‡2]. In an inflationary cosmology in which the inflation occurs after the Peccei–Quinn symmetry is spontaneously broken, however, a/f_a can be approximately constant throughout the visible universe. If its arbitrarily chosen value, which we expect to be of order one, happens to be small, the bound on f_a can be relaxed. For example, if $f_a \sim 10^{15}$ GeV, the axion energy density is acceptably small for $(A_i/f_a)^2 \lesssim 10^{-3}$ [‡3]. Thus, for the value of f_a suggested by grand unified theories, we are faced with a naturalness problem which is not quite so severe as the strong *CP* problem that the invisible axion is intended to solve. [‡4]

Let us now examine whether the axion energy density can be dissipated by particle production [18]. Naively, particle production is insignificant, because the invisible axion with $f_a \gtrsim 10^9$ GeV has a lifetime which far exceeds the age of the universe. However, the oscillating axion field represents a coherent space consisting of a large density of zero-momentum axions, and one wonders whether coherence effects can greatly amplify the rate at which this state decays. We will show that in fact the cosmological red shift prevents such coherence effects from substantially reducing the axion energy density.

The two-photon decay mode of the axion, which is the dominant one at zero temperature, is forbidden for an axion at rest when $T \gtrsim 6\ \text{keV}(10^9\ \text{GeV}/f_a)^{2/3}$, because the electron plasma frequency exceeds the photon energy $\frac{1}{2}m_a$. At higher temperatures, the dominant process that destroys zero-momentum axions is the $4a \rightarrow 2a$ process, in which the energy of four zero-momentum axions is transferred to two axions with $E = 2m_a$; this energy can subsequently red-shift away harmlessly. The $4a \rightarrow 2a$ process causes the axion zero-momentum mode to decay at the net rate

$$\Gamma_{4a\rightarrow 2a} \sim (m_a^3/f_a^2)(A/f_a)^6 n_{2m_a}, \tag{15}$$

where n_{2m_a} is the occupation number for the $E = 2m_a$ modes. (This is the difference of the $4a \rightarrow 2a$ rate and the rate for the inverse process.) In standard cosmology we expect $n_{2m_a} \sim f_a^2/3m_a^2$ while $H \gtrsim 2m_a$. (In an inflationary scenario, n_{2m_a} is much smaller.) Subsequently, n_{2m_a} and A begin to decrease rapidly because of the Hubble expansion. Therefore, $\Gamma_{4a\rightarrow 2a}$ never becomes as large as H, and the $4a \rightarrow 2a$ process cannot significantly deplete the axion zero-momentum mode.

Eventually, the universe cools to the point where the two-photon decay mode of the axion is allowed. At this point, were it not for the cosmological red shift, induced emission would cause the $E = \frac{1}{2}m_a$ photon modes to grow exponentially like $e^{\Gamma t}$ where [19]

$$\Gamma \sim \alpha(A/f_a)m_a. \tag{16}$$

But this exponential growth is blocked by the red-shift if $\Gamma\tau \lesssim 1$, where $\tau \sim \Gamma(HM_a)^{-1}$ is the time required for the band of width Γ of growing modes to be red-shifted by an amount of order Γ. The condition

$$\Gamma\tau \sim \alpha^2(A/f_a)^2 m_a H^{-1} \lesssim 1 \tag{17}$$

is well satisfied at all times. Thus, the photon modes do not grow exponentially, and the net $a \rightarrow \gamma\gamma$ decay rate of the axion is

[‡2] Another means of dissipating the axion energy density has been discussed by Vilenkin and Everett [17]. They propose that nearly all of the energy density is carried by oscillating pieces of domain wall bounded by strings, and that these oscillations are damped by the emission of gravitational radiation. We expect that their mechanism can actually dissipate only a small fraction of the axion energy density, and have ignored it in our discussion.

[‡3] Reheating does not much affect A_i/f_a. If T_i is the temperature at which the axion Compton wavelength becomes comparable to the "horizon" size H^{-1}, the fluctuations averaged over a "horizon" volume are $\delta a^2/f_a^2 \sim (T_{RH}/f_a)^2 \times (40T_i/m_p)$, which is quite small for $T_{RH} \lesssim f_a$.

[‡4] The grand unified axion may be made acceptable in yet another way, if inflation is used to avoid axion domain walls. In a model in which the PQ symmetry is spontaneously broken at the GUT scale, we actually expect $f_a \sim (10^{15}\ \text{GeV})/N$, where Z_N is the subgroup of $U(1)_{PQ}$ which is not explicitly broken by nonperturbative QCD effects. We can make N large by incorporating into the model many superheavy fermions with PQ charges; for N sufficiently large, the constraint (14) can be satisfied.

$$\Gamma_{a\to\gamma\gamma} \sim \alpha^2(m_a^3/f_a^2)n_\gamma$$

$$\sim 10^{-24}\ \mathrm{yr}^{-1}(10^9\ \mathrm{GeV}/f_a)^5 n_\gamma, \quad (18)$$

where $n_\gamma = [\exp(m_a/2T) - 1]^{-1}$ is a thermal occupation number. The axions produced in the early universe are still around.

The axion field is still oscillating coherently today, but with a very small amplitude. From eq. (9) we can deduce that, when the photon temperature is 2.7 K, the amplitude is

$$A/f_a \sim 10^{-22}(f_a/10^9\ \mathrm{GeV})^{1/2}(100\ \mathrm{MeV}/T_i)^{1/2}(A_i/f_a). \quad (19)$$

Therefore, although $\bar{\theta}$ is still oscillating today with frequency m_a, the *CP* violation in the strong interactions is extremely small; the Peccei–Quinn mechanism works. (Actually, the *CP* violation in the weak interactions shifts the minimum of the axion potential to $\bar{\theta} \sim 10^{-14}$, in the Kobayashi–Maskawa [20] model.)

The astrophysical and cosmological constraints on the axion together require $10^9\ \mathrm{GeV} \lesssim f_a \lesssim 10^{12}\ \mathrm{GeV}$; if we wish to solve the strong *CP* problem by the Peccei–Quinn mechanism, there must be a symmetry-breaking scale in that range. Could the new physics associated with this symmetry-breaking scale be experimentally accessible? One possibility is that there are other flavor symmetries, besides the Peccei–Quinn U(1) symmetry, which are spontaneously broken at the scale f_a. Thus there may be other very weakly coupled Goldstone bosons besides the axion, perhaps including ones with generation nondiagonal couplings. For example, if $f_a \sim 10^{11}$ GeV, the branching ratios for $\mu \to e\phi$ and $K^+ \to \pi^+\phi$, where ϕ is a Goldstone boson, may not be far below current experimental limits.

If f_a is of order 10^{12} GeV, we are led to entertain the possibility that axions make up a significant part of the dark matter of the universe. An axion gas with $\Omega_a \sim 1$ behaves like dust; it is nondissipative and pressureless. [‡5] Fluctuations in the axion energy density can begin to grow when the universe becomes matter dominated by axions at $T \sim 10$ eV, and these fluctuations can act as a seed for galaxy formation when ordinary matter decouples from radiation at $T \sim 0.3$ eV.

[‡5] The energy–momentum tensor of a *coherent* axion gas includes an oscillating pressure term, proportional to $\cos 2m_a t$, which *averages* to zero. But it can be shown by explicit calculation that small fluctuations in the axion energy density grow like $t^{2/3}$ when the universe is dominated by axions; this is the same behavior as for nonrelativistic matter.

Results very similar to ours have been obtained by Abbott and Sikivie [2]. We thank them, S. Hawking and A. Vilenkin for very helpful conservations. This research has been supported in part by the National Science Foundation under Grants. No. PHY77-22864 and PHY77-27084, and by the Department of Energy under Contract No. DE-ACO3-81ER40050.

Notes added. The astrophysical bound on f_a derived in ref. [8] has been reexamined by Fukugita et al. [22] who claim that the actual limit is $f_a \gtrsim 4 \times 10^7$ GeV.

The result eq. (12) was previously derived by Dine et al. in ref. [6].

References

[1] G. 't Hooft, Phys. Rev. Lett. 37 (1976) 8; Phys. Rev. D14 (1976) 3432;
C.G. Callan, R.F. Dashen and D.J. Gross, Phys. Lett. 63B (1976) 334;
R. Jackiw and C. Rebbi, Phys. Rev. Lett. 37 (1976) 177.
[2] W.B. Dress et al., Phys. Rev. D15 (1977) 9.
[3] V. Baluni, Phys. Rev. D19 (1979) 2227;
R.J. Crewther et al., Phys. Lett. 88B 1979) 123.
[4] R. Peccei and H. Quinn, Phys. Rev. Lett. 38 (1977) 1440.
[5] S. Weinberg, Phys. Rev. Lett. 40 (1978) 223;
F. Wilczek, Phys. Rev. Lett. 46 (1978) 279.
[6] J.E. Kim, Phys. Rev. Lett. 43 (1979) 103;
M. Dine, W. Fischler and M. Srednicki, Phys. Lett. 104B (1981) 199.
[7] M.B. Wise, H. Georgi and S.L. Glashow, Phys. Rev. Lett. 47 (1981) 402.
[8] D. Dicus et al., Phys. Rev. D18 (1978) 1829.
[9] H. Georgi and S.L. Glashow, Phys. Rev. Lett. 32 (1974) 438;
H. Georgi, H. Quinn and S. Weinberg, Phys. Rev. Lett. 33 (1974) 451.
[10] A. Vilenkin, Phys. Rev. Lett. 48 (1982) 59.
[11] A. Guth, Phys. Rev. D23 (1981) 347;
A.D. Linde, Phys. Lett. 106B (1982) 389;
A. Albrecht and P.J. Steinhardt, Phys. Rev. Lett. 48 (1982) 1220.
[12] P. Sikivie, Phys. Rev. Lett. 48 (1982) 1156.
[13] Ya.B. Zeldovich, I.Yu. Kobzarev and L.B. Okun, Zh. Eksp. Teor. Fiz. 67 (1974) 3.

[14] G. Lazarides and Q. Shafi, Phys. Lett. 115B (1982) 21; H. Georgi and M.B. Wise, Phys. Lett. 116B (1982) 123; S. Dimopoulos et al., Phys. Lett., 117B (1982) 185.
[15] D.J. Gross, R.D. Pisarski and L.G. Yaffe, Rev. Mod. Phys. 53 (1981) 43.
[16] M. Peskin, unpublished; S. Weinberg, Phys. Rev. Lett. 40 (1978) 223; W. Bardeen and S.-H.H. Tye, Phys. Lett. 74B (1978) 229.
[17] A. Vilenkin and A.E. Everett, Phys. Rev. Lett. 48 (1982) 1867.
[18] L.F. Abbott, E. Farhi and M.B. Wise, Phys. Lett. 117B (1982) 29.
[19] N.W. McLachlan, Theory and Application of Mathieu Functions (Oxford, U.P. London, 1947); B.R. Mollow and R.J. Glauber, Phys. Rev. 160 (1967) 1076.
[20] M. Kobayashi and K. Maskawa, Prog. Theor. Phys. 49 (1973) 652.
[21] L. Abbott and P. Sikivie, to appear.
[22] M. Fukugita et al., Phys. Rev. Lett. 48 (1982) 1522.

A COSMOLOGICAL BOUND ON THE INVISIBLE AXION

L.F. ABBOTT [1]
Physics Department, Brandeis University, Waltham, MA 02254, USA

and

P. SIKIVIE [2]
Particle Theory Group, University of Florida, Gainesville, FL 32611, USA

Received 14 September 1982

The production of axions in the early universe is studied. Axion models which break the $U(1)_{PQ}$ symmetry above 10^{12} GeV are found to produce an unacceptably large axion energy density.

The absence of *CP* violation in the strong interactions can be explained naturally by incorporating the $U(1)_{PQ}$ symmetry of Peccei and Quinn [1] into the standard model. A consequence of this symmetry is the axion [2], a spin 0 boson of mass

$$m_A \approx m_\pi f_\pi / v, \tag{1}$$

where v is the magnitude of the vacuum expectation value which breaks the $U(1)_{PQ}$ symmetry. An axion of mass (1) with v of order the weak interaction scale has not been observed. However, it has been pointed out [3] that v may be much larger, for example of order the grand unification mass. The axion is then extremely light and weakly coupled and is essentially invisible to laboratory experiments. It has recently been noted [4] that most axion models have a spontaneously broken $Z(N)$ symmetry and that the resulting discrete, degenerate vacua give rise to domain walls which are incompatible with standard cosmology. Several solutions to this problem have been proposed [4,5]. One approach is to adopt an inflationary cosmology [6]. This solves the domain wall problem provided that the reheating temperature, after the inflationary period, is not high enough to restore the $U(1)_{PQ}$ symmetry.

In this letter, we derive a new constraint which standard cosmology imposes on axion models. We consider the production of axions in the early universe and find that axions are copiously produced when the temperature, T, gets of order Λ, where Λ is the QCD scale parameter. At this time, QCD instanton effects align the vacuum producing a coherent state of axions at rest. After they are produced, the axions decouple. They have a large energy density which only goes down like T^3 since they are nonrelativistic. Thus, we expect a large energy density of axions in the present universe. We find that unless v is less than 10^{12} GeV, the present axion energy density will exceed the closure density of the universe by a factor of ten. This bound holds whether or not the vacuum of the axion model is unique. It is also independent of the history of the universe before the temperature reaches about 100 GeV.

To obtain our bound we must consider how the axion field behaves as the universe cools. At a high temperature of order v, a scalar field develops an expectation value

$$\langle \Phi \rangle = v e^{i\theta}, \tag{2}$$

breaking the $U(1)_{PQ}$ symmetry. The axion field is

[1] Supported in part by DOE contract DE-ACO3-76ERO3230-A005.
[2] Supported in part by DOE contract DSR80136je7.

0 031-9163/83/0000–0000/$ 03.00 © 1983 North-Holland

$\phi_A \equiv \theta v$. Initially, the value of θ in eq. (2) is chosen randomly, but after the quarks develop a mass at a temperature of about 100 GeV, QCD instantons introduce a potential for the axion field with minima at values of θ for which there is no *CP* violation. This is the Peccei–Quinn mechanism [1]. We will choose our phases so that the vacuum is described by $\theta = \phi_A = 0$. The initial value of θ in eq. (2) is presumably some number of order one. Thus, initially ϕ_A is of order v and as T approaches Λ the instanton induced potential will cause the axion field to oscillate,

$$\phi_A \approx Av \cos m_A t. \tag{3}$$

We will parameterize the amplitude of these oscillations by the dimensionless variable A. Initially, $A = \theta$ and is of order one. The oscillating ϕ_A field is equivalent to a coherent state of axions at rest. Since we are dealing with large numbers of axions in a coherent state we can use classical field theory as a good approximation. The present value of A in eq. (3) must be extremely small if the axion energy density which is about $m_A^2 v^2 A^2 \approx m_\pi^2 f_\pi^2 A^2$ is to be compatible with cosmology. If we require this energy density to be less than ten times the critical energy density of the universe we find that

$$A_{\text{present}} < 10^{-21}. \tag{4}$$

We will now compute the present value of the amplitude of the axion field oscillations, A, to see if the bound of eq. (4) is satisfied. If we ignore (for the moment) axion interactions, the axion field is governed by the equation of motion

$$\mathrm{d}^2\phi_A/\mathrm{d}t^2 + 3H(t)\mathrm{d}\phi_A/\mathrm{d}t + m_A^2(T)\phi_A = 0, \tag{5}$$

where $H(t)$ is the Hubble "constant" and $m_A(T)$ is the temperature-dependent axion mass. We use the Hubble constant ($H = 1/2t$) and temperature–time relation ($T \propto t^{-1/2}$) of a spacially flat, radiation-dominated cosmology. The axion mass, $m_A(T)$, can be determined from finite-temperature instanton calculations [7]. For three quark flavors (which is the relevant number for the region of interest) we find

$$m_A(T) = 0.04\, \alpha_s^{-3}(\pi^2 T^2)(m_u m_d m_s \Lambda^9)^{1/2} v^{-1} T^{-4}. \tag{6}$$

Eq. (5) is the equation of a damped harmonic oscillator with time-dependent parameters. Until the age of the universe, t, is of order $1/m_A(t)$ the axion field will not have had enough time to oscillate. At a time t_1, where $t_1 = 1/m_A(t_1)$, when the temperature is T_1, the axion field begins to oscillate. The drag term in eq. (5) will cause the amplitude of these oscillations to decrease like $T^{3/2}$. The time variation of the axion mass also causes the oscillation amplitude to decrease. After the time t_1, the relative change in the axion mass is slower than the oscillation frequency, that is $(1/m_A)\mathrm{d}m_A/\mathrm{d}t < m_A$. This means that the mass variation is adiabatic and the quantity $m_A(T)A^2$ will remain a constant since it is an adiabatic invariant [8]. Thus, the amplitude is reduced by the square root of the ratio of the axion mass at temperature T_1 to the final axion mass, m_A of eq. (1). Putting all this together we find that at a temperature T (where $T < \Lambda$)

$$A = [m_A(T_1)/m_A]^{1/2}(T/T_1)^{3/2}, \tag{7}$$

assuming initially $A = \theta = 1$. T_1 and $m_A(T_1)$ can be determined from eq. (6). We find that $T_1 \approx 1$ GeV $(10^{11}\ \text{GeV}/v)^{1/6}$ and

$$A_{\text{present}} \approx 10^{-21}\ (v/10^{12}\ \text{GeV})^{7/12}. \tag{8}$$

For eq. (4) to be satisfied, we must therefore require that

$$v < 10^{12}\ \text{GeV}. \tag{9}$$

This bound is relatively insensitive to the normalization and even to the power of T used in the formula for $m_A(T)$. Uncertainties about the value of Λ and about the effects of chiral symmetry breaking introduce at most a factor of 2–3 uncertainty in the bound.

Up to now we have been ignoring axion couplings. The invisible axion, of course, decouples at temperatures much below v. However, when the axion field is oscillating, we have a huge density of axions in a coherent state and coherence can enhance the effects of axion interactions. Axion couplings to fermions do not lead to coherent effects and so are irrelevant. Likewise, axion couplings to gravitons are too weak. However, axion self-couplings and photon couplings are of interest.

Since the coherent state axions are non-relativistic, their energy density goes down like T^3 as indicated by eq. (7). However, if through self-couplings these axions could become relativistic, their energy density would start going down like T^4 and the present value of A would be much smaller than the result (8). To see if significant amounts of energy can be transferred

from the initial zero momentum mode to non-zero momentum modes let us consider the effect of a self-coupling term $\lambda\phi_A^4/4!$ and write the axion field as

$$\phi_A = Av\cos m_A t + \delta A_k(t)\exp(i\boldsymbol{k}\cdot\boldsymbol{x}). \quad (10)$$

Then, for small δA_k, the non-zero momentum mode is governed by the equation (ignoring for the moment the Hubble term)

$$d^2\delta A_k/dt^2 + (k^2 + m_A^2 + \tfrac{1}{2}\lambda A^2 v^2\cos^2 m_A t)\delta A_k = 0. \quad (11)$$

This is the Mathieu equation and it has the following properties [9]. There are bands of k values for which eq. (11) is unstable and δA_k grows exponentially. The most rapidly growing modes occur for a band of width

$$\Delta k \approx (m_A/4\sqrt{3})(\tfrac{1}{8}\lambda A^2 v^2/m_A^2)^2, \quad (12)$$

which grow at a rate

$$\Gamma \approx \tfrac{1}{16}\, m_A(\tfrac{1}{8}\lambda A^2 v^2/m_A^2)^2. \quad (13)$$

A given unstable mode will grow as long as it lies in the band of instability, but eventually it will be red-shifted out of this band. The time it takes to be red-shifted out of the unstable band is

$$\tau \approx \Delta k/Hm_A. \quad (14)$$

For the axion field $\lambda \approx m_A^2/v^2$, so,

$$\Gamma\tau \approx (1/64\sqrt{3})(\tfrac{1}{8}A^2)^4(m_\Lambda/H). \quad (15)$$

For $v \geqslant 10^{12}$, we find that $\Gamma\tau$ is always much less than one so no appreciable growth takes place in the non-zero momentum modes. Thus, axion self-couplings have no effect on the bound of eq. (9).

The axion lifetime for decays into two photons is longer than the age of the universe for an invisible axion. However, for our coherent state of axions, stimulated emission can lead to an enhancement of this decay mode. The axion photon system can be evaluated exactly like the non-zero momentum modes considered above. In this case, the oscillating axion field couples two polarization modes of the electromagnetic field and the system acts like a parametric amplifier [10]. However, the axion decay into photons is blocked until the temperature reaches about 1 MeV because the plasma frequency $[(4\pi e^2/m_e)n_e]^{1/2}$ is greater than the axion mass. After this, the axion density is too low for significant coherent effects to occur. Thus, our bound is also unaffected by photon couplings of the axion.

In conclusion, axion models which break the $U(1)_{PQ}$ symmetry above 10^{12} GeV result in a present energy density of non-relativistic axions more than ten times greater than the critical energy density. In the introduction, we suggested that the domain wall problem could be avoided in an inflationary cosmology provided that $T_{\text{reheat}} < v$. However, our bound would then require that $T_{\text{reheat}} < 10^{12}$ GeV. Since the baryon asymmetry must be generated after the inflationary period, it may be difficult to have inflation take place at such a low temperature.

Astrophysical arguments [11] place a lower bound on v of roughly 10^9 GeV [‡1]. Thus, invisible axion models with v between about 10^9 GeV and 10^{12} GeV are still acceptable. For such models, axions are a major component of the present energy density of the universe. (For example, if $v \approx 1.4 \times 10^{11}$ GeV axions would give the universe a critical energy density.) The presence of these axions would clearly have profound implications for astrophysics and cosmology.

Work similar to ours has been carried out by J. Preskill, M. Wise and F. Wilczek. We thank J. Preskill, M. Wise and S. Coleman for helpful discussions. This work was done at SLAC and by P.S. also at the Aspen Center for Physics. We thank both institutions.

Note added. We have recently learned that M. Dine and W. Fischler have also studied this problem.

[‡1] This is the value from ref. [11]. Fukugita et al. [12] find a lower bound of 4×10^7 GeV.

References

[1] R.D. Peccei and H. Quinn, Phys. Rev. Lett. 38 (1977) 1440.
[2] S. Weinberg, Phys. Rev. Lett. 40 (1978) 223;
F. Wilczek, Phys. Rev. Lett. 40 (1978) 279.
[3] M. Dine, W. Fischler and M. Srednicki, Phys. Lett. 104B (1981) 199;
J. Kim, Phys. Rev. Lett. 43 (1979) 103;
M. Wise, H. Georgi and S.L. Glashow, Phys. Rev. Lett. 47 (1981) 402;
H.P. Nilles and S. Raby, SLAC-PhB-2743 (1981).
[4] P. Sikivie, Phys. Rev. Lett. 48 (1982) 1156.

[5] G. Lazarides and Q. Shafi, Phys. Lett. 115B (1982) 21;
B. Holdom, Stanford preprints ITP-712 (1982); ITP-713 (1982);
H. Georgi and M. Wise, Phys. Lett. 116B (1982) 123;
S. Dimopoulos, P. Frampton, H. Georgi and M. Wise, Phys. Lett. 117B (1982) 185;
S. Barr, D. Reiss and A. Zee, Phys. Lett. 116B (1982) 227;
S. Barr, X.C. Gao and D. Reiss, University of Washington preprint 40048-22P2 (1982);
Y. Fujimoto, K. Shigemoto and Zhao Zhi-yong, ICTP preprint IC-82-140 (1982).

[6] A. Guth, Phys. Rev. D23 (1981) 347.

[7] D.J. Gross, R.D. Pisarski and L.G. Yaffe, Rev. Mod. Phys. 53 (1981) 43.

[8] L.D. Landau and E.M. Lifshitz, Mechanics (Addison-Wesley, Reading, MA, 1960) pp. 154–157.

[9] N.W. McLachlan, Theory and applications of Mathieu functions (Oxford U.P., London, 1947).

[10] B.R. Mollow and R.J. Glauber, Phys. Rev. 160 (1967) 1076, 1097.

[11] D.A. Dicas, E.W. Kolb, V.L. Teplitz and R.V. Wagoner, Phys. Rev. D18 (1978) 1829.

[12] M. Fukugita, S. Watamura and M. Yoshimura, Phys. Rev. Lett. 48 (1982) 1522.

THE NOT-SO-HARMLESS AXION

Michael DINE
The Institute for Advanced Study, Princeton, NJ 08540, USA

and

Willy FISCHLER
Department of Physics, University of Pennsylvania, Philadelphia, PA 19104, USA

Received 17 September 1982
Received manuscript received 14 October 1982

Cosmological aspects of a very weakly interacting axion are discussed. A solution to the problem of domain walls discussed by Sikivie is mentioned. Demanding that axions do not dominate the present energy density of the universe is shown to give an upper bound on the axion decay constant of at most 10^{12} GeV.

It has been suggested that the strong *CP* problem may be solved by extending the Peccei–Quinn idea [1] to grand unified theories [2–4]. In other words, one should require of a grand unified theory that it possesses a U(1) symmetry broken explicitly only by anomalies. Moreover, this symmetry should be broken spontaneously at the unification scale. In such a theory, the *CP* violating angle, θ, becomes dynamical; it is the would-be Goldstone boson (axion) of this spontaneously broken symmetry. QCD gives rise to a potential for this angle with a minimum very near the origin. The axion itself receives a tiny mass as a result of this symmetry-violating potential. Its interactions are extremely weak, having a strength inversely proportional to f_A, the axion decay constant. In particular, this axion would not appear in experiments nor would it play any role in astrophysical environments such as stars.

However, this lack of interaction raises concerns of a cosmological nature. For temperatures much higher than QCD scales, we expect θ-dependent effects to fall as some large power of T [5]. Thus, at the grand-unified scale, the Peceei–Quinn symmetry will be an essentially exact symmetry. At temperatures above f_A, this symmetry will be unbroken; as the temperature is lowered, spontaneous symmetry breaking will occur, and θ will take some random value on the interval $[0, 2\pi]$ (actually, on a somewhat smaller interval; see below).

At temperatures well below the QCD scale, we can calculate the axion potential using current algebra arguments. For definiteness, we focus on the model of ref. [2], in which the Peccei–Quinn symmetry is broken by the expectation value of a singlet field (with Peccei–Quinn charge 1)

$$\langle\varphi\rangle = 2^{-1/2} f_A \, e^{i\theta} .$$

Following ref. [6], it is natural to work with the anomaly-free, partially conserved current (specializing for simplicity to the case of two light quarks; inclusion of the strange quark is straightforward)

$$\tilde{j}_\mu = j_\mu^{PQ} - [N/(1+Z)]\,(\bar{u}\gamma_\mu\gamma_5 u + Z\bar{d}\gamma_\mu\gamma_5 d) ,$$

where $Z = m_u/m_d$. We must also include the current of axial isospin,

$$j^3_{\mu A} = \tfrac{1}{2}(\bar{u}\gamma_\mu\gamma_5 u - \bar{d}\gamma_\mu\gamma_5 d)$$

Following a method discussed by Sikivie, we write the expectation values of fermion bilinears as

$$\langle\bar{u}_L u_R\rangle = |\langle\bar{u}_L u_R\rangle| \exp\{i\{\alpha + [N/(1+Z)]\,\theta\}\} ,$$

$$\langle\bar{d}_L d_R\rangle = |\langle\bar{d}_L d_R\rangle| \exp\{i\{-\alpha + [NZ/(1+Z)]\,\theta\}\} .$$

0 031-9163/83/0000–0000/$ 03.00 © 1983 North-Holland

Then the effective potential for α and θ is obtained by replacing fields by their expectation values (using $\langle\bar{u}u\rangle \approx \langle\bar{d}d\rangle$)

$$V = |\langle\bar{u}u\rangle|\,[m_u \cos(\alpha + [N/(1+Z)]\,\theta) + m_d \cos(ZN\theta/(1+Z) - \alpha)]\,. \quad (1)$$

Here αf_π is the pion field and θf_A is the axion field. Expanding to second order in the π^0 and axion fields, the cross terms cancel (this is the reason for the choice of Z) we find the usual results:

$$f_\pi^2 m_\pi^2 = (m_u + m_d)\langle\bar{u}u\rangle\,,$$

$$f_A^2 m_A^2 = [Z/(1+Z)^2]\,N^2 m_\pi^2 f_\pi^2\,.$$

As Sikivie has emphasized, this potential has a discrete Z_N symmetry. It corresponds to the transformation

$$\theta \to \theta + 2\pi/N\,, \quad \alpha \to \alpha + 2\pi - 2\pi/(1+Z)$$

Modulo this discrete transformation, the minimum of the potential lies at $\theta = 0$ (it will, of course, acquire a tiny non-zero value once we include *CP* violating effects). However, the question arises, how quickly will θ approach its minimal value? In particular, since the axion is so weakly coupled to matter, there is virtually no damping term in the equation of motion for θ. If we neglect, for a moment, the expansion of the universe, θ obeys the equation (keeping, for simplicity, only the quadratic term in the potential),

$$\ddot{\theta} + \Gamma\dot{\theta} + m^2\theta = 0 \quad (2)$$

The only damping comes from the decay of the axion to two photons. The axion lifetime in vacuum is 10^{58} yr. In the early universe (essentially for any temperature greater than about 10 eV) in fact, it is not clear that the axion can decay at all! The reason is that for these temperatures the Debye frequency of the plasma is larger than the axion mass. In any case, Γ is *extremely* tiny. Thus if we neglect the expansion of the universe, θ simply rolls back and forth in the potential, eq. (1), essentially forever.

Of course, the universe does expand, and this will induce a significant decrease in θ with time. At temperatures small compared to QCD scales, the generally covariant version of eq. (2), in a Robertson–Walker metric, is

$$\ddot{\theta} + 3(\dot{R}/R)\dot{\theta} + m^2\theta = 0\,. \quad (3)$$

This equation has the solution, for a radiation-dominated universe ($\dot{R}/R = 1/2t$)

$$\theta = t^{-1/4}[aJ_{1/4}(mt) + bJ_{-1/4}(mt)] \underset{t \gg m^{-1}}{\sim} \theta_0 (mt)^{-3/4} \sin mt \quad (4)$$

For a matter-dominated universe ($\dot{R}/R = 2/3t$)

$$\theta = t^{-1/2}[a'J_{1/2}(mt) + b'J_{-1/2}(mt)] \underset{t \gg m^{-1}}{\sim} \theta_0 (mt)^{-1} \sin mt \quad (5)$$

The coefficients of the Bessel functions in these equations depend on the initial conditions which we impose. The asymptotic behavior, however, is not particularly sensitive to these conditions. The question of appropriate initial conditions will be described below, when we describe the axion potential at high temperatures.

The axion mass, of course, does not simply turn on at the phase transition, but presumably falls off, at high temperatures, as some power of the temperature. The θ-dependence of the free energy arising from instantons can, in fact, be calculated at high temperature and will fall as a high power of T [5]. For example, for three light quarks, instanton calculations give a result for the free energy

$$F \approx -C(\pi T)^{-8}\Lambda_{PV}^9 m_u m_d m_s\,[4\pi^2/g^2(T)]^6 \cos N\theta\,, \quad (6)$$

Here the constant, C, is obtained from a numerical integration of an expression for the instanton density due to Gross et al. [5]. We have made the estimate

$$C = 2.4 \times 10^{-2}\,. \quad (7)$$

Λ_{PV} is the QCD scale factor appropriate to the Pauli–Villars scheme. From ref. [7], we know $\Lambda_{PV} \approx \Lambda_{\overline{MS}}$. Also, in obtaining the expression (6), it should be noted that we have neglected the scale dependence of the $(1/g^2)^6$ factor in the integration over scale sizes.

Of course, at low temperatures, many effects (apart from instantons) presumably conspire to give the θ-dependence of the vacuum energy, and these presumably also fall as some power of T. It is possible that even at large temperatures these effects are the dominant ones. Thus to obtain an idea of what θ may do in the early universe not too deeply tied to our prejudices from in-

stanton physics, we take

$$m_A^2(T) = \tilde{m}^2(T/\tilde{T})^{-a} , \tag{8}$$

and consider a range of plausible values for $\tilde{m}$, $\tilde{t}$ and a, including the values obtained from instanton computations. Note that eq. (8) is written in this form for purposes of motivation; obviously only the product $\tilde{m}^2\tilde{T}^a$ is physically meaningful. We expect $\tilde{T}$ to be some characteristic QCD scale, and m to be some fraction of the zero-temperature axion mass.

The equation of motion for θ follows from conservation of energy at constant entropy. In a radiation-dominated universe one obtains

$$\ddot{\theta} + (3/2t)\dot{\theta} + \tilde{m}^2[T(t)/\tilde{T}]^{-a} = 0 . \tag{9}$$

Our previous remarks about the Peccei–Quinn phase transition suggest that at some very small time, t_0, corresponding to a temperature well above QCD scales (the precise meaning of "well-above" will become clear shortly) we should set $\theta = \theta_0$, $\dot{\theta}_0 = 0$. The temperature in eq. (9) depends upon the time through the relation

$$T = (M_P/2\gamma t)^{1/2} , \tag{10}$$

where above QCD scales, for three light quarks,

$$\gamma \approx 13 .$$

This equation has the solution, with these initial conditions

$$\begin{aligned}\theta &= \theta_0(t/\tilde{t})^{-1/4}J_\nu[(t/\tilde{t})^\beta]\,\Gamma(\nu+1)2^\nu \\ &\underset{t\gg\tilde{t}}{\simeq} 2^\nu\Gamma(\nu+1)(2/\pi)^{1/2}\theta_0(t/\tilde{t})^{-(\beta+1/2)/2} \\ &\quad\times\cos[(t/\tilde{t})^\beta - \tfrac{1}{2}\pi\nu - \tfrac{1}{4}\pi] ,\end{aligned} \tag{11}$$

where

$$\beta = \tfrac{1}{4}a + 1 , \quad 2\beta\nu = \tfrac{1}{2}$$

$$\tilde{t} = [\beta^{-1}(\tilde{m}^2\tilde{T}^a)^{1/2}(M_P/2\gamma)^{-a/4}]^{-1/\beta} . \tag{12}$$

Note that this time corresponds to a temperature which can be computed from eq. (10). $\tilde{t}$ can be thought of as a characteristic time for the motion of θ. In particular, we should choose $t_0 \ll \tilde{t}$.

We are now in a position to discuss the cosmological development of θ. First, we can see that θ today is extremely small, no matter what we assume for a, $\tilde{T}$, $\tilde{m}$, the location and nature of the phase transition, etc. If $\theta \sim 1$ at $T = 100$ MeV, then $\theta \approx 10^{-18}$ today. Certainly, then, we do not need to worry about the neutron electric dipole moment.

Also, in this picture, we can discuss a serious concern raised by Sikivie [8]. He points out that, since the Z_N symmetry respected by QCD is spontaneously broken by $\langle\phi\rangle$, one expects the formation of domain walls. This leads to conflicts with standard cosmology. One way out of this problem is to choose the fermion content so that the left-over discrete symmetry is trivial[‡1]. This has been discussed by Sikivie and by Georgi and Wise [9]. An alternative possibility is suggested by the picture above of the time evolution of θ. In the early universe scenario discussed by Guth [10] and refined by Linde [11] and Albrecht and Steinhardt [12], the universe goes through a period of exponential expansion at temperatures below the grand unified mass. We live, essentially, in one of the causal regions of this early universe and hence in one domain. In this region, θ will take a single value, which will evolve in time according to the equations of motion we have discussed earlier. [One can easily check, from eq. (3), that θ is virtually constant during the period of exponential expansion.] Thus a period of exponential expansion can solve the domain wall problem, provided that the Peccei–Quinn transition occurs at high enough temperatures, and that one does not reheat the universe above this temperature.

A more serious concern is the energy density associated with the axion field, at the time of helium synthesis and today (this point has been emphasized to us by L. Abbott and P. Sikivie). The pressure and energy density of the axion field obey an "equation of state" appropriate to non-relativistic matter, for $t \gg m^{-1}$. This can be seen directly by substituting our solutions, eqs. (4) and (5), into the expressions for the energy density, ρ. and pressure, p, and averaging over a period, m^{-1}

$$\begin{aligned}\rho &= \tfrac{1}{2}f_A^2(\dot{\theta}^2 + m_A^2\theta^2) \\ &\underset{mt\gg1}{\approx} (mt)^{-3/2} , \quad \text{radiation domination,} \\ &\underset{mt\gg1}{\approx} (mt)^{-2} , \quad \text{matter domination} ,\end{aligned} \tag{13}$$

‡1 This can also be achieved by choosing the Peccei–Quinn charge of ϕ to be N.

$$p = \tfrac{1}{2} f_A^2 (\dot{\theta}^2 - m_A^2 \theta^2)$$

$$\underset{mt \gg 1}{\approx} 0 . \qquad (13 \text{ cont'd})$$

Thus, if at some time ρ is large compared to the energy density of radiation, the universe will expand as if matter dominated. Note that, in either case $\rho \sim R^{-3}$ for large times.

Suppose that the potential is given by the zero temperature expression, eq. (1), below some temperature T_c (presumably near the chiral symmetry violating phase transition). If the universe has been expanding adiabatically since that time, then the ratio of ρ to the entropy density is a constant. If we calculate the entropy at T_c appropriate to a quark–gluon plasma[‡2], tje requirement that the universe be radiation dominated at the time of helium synthesis gives

$$\theta^2(T_c)/T_c^3 \lesssim 30 \text{ GeV}^{-3} \qquad (14)$$

From the limits on the present energy density of the universe, and the present temperature of the cosmic radiation background we obtain the stronger limit

$$\theta^2(T_c)/T_c^3 \lesssim 10^{-4} \text{ GeV}^{-3} . \qquad (15)$$

If we take $T_c = 0.1$ GeV, we obtain, from the latter equation

$$\theta^2(T_c) \lesssim 10^{-7} , \qquad (16)$$

whereas if we take the much more optimistic value, $T_c = 1$ GeV, the limit is

$$\theta^2(T_c) \lesssim 10^{-4} . \qquad (17)$$

We can calculate $\theta^2(T_c)$ from our solution of eq. (9). As a pessimistic estimate, we can use the formula for the axion mass implied by instantons. We will see later that more optimistic assumptions for $\tilde{m}$, $\tilde{T}$, T_c, and a do not drastically alter the situation.We consider the case of three light quarks (this will be seen to be self-consistent), and we will, as stated earlier, neglect the logarithmic variation of the $g^{-2}(T)$ terms, taking these to be simply g^{-2} (1 GeV). Of course, the instanton calculation cannot be trusted down to arbitrarily low temperatures, since the free energy diverges. We will make the ad hoc assumption that the temperature, T_c, is just the temperature where formula (6) gives the zero temperature result, eq. (1). One can check that the final results are not too sensitive to this choice. In any case, these assumptions yield (taking $\Lambda_{\overline{MS}} = 0.150$ GeV)

$$a = 8, \qquad \tilde{m}^2 \tilde{T}^8 f_A^2 = 5.8 \times 10^{-12} \text{ (GeV)}^{12}$$

$$T_c = 0.089 \text{ GeV}, \quad \tilde{t} = 5 \times 10^{18} \text{GeV}^{-1} (f_A/10^{15} \text{GeV})^{1/3} .$$

Plugging this into our solution, we find that we need (assuming $\theta_0 \sim 1$)

$$f_A \lesssim 10^{11} \text{ GeV} \qquad (18)$$

in order to satisfy the bound, eq. (15).

As a more optimistic extreme, we can take

$$\tilde{m}^2 = 0.1 \, m_A^2 , \quad a = 2 ,$$

$$\tilde{T} = 1 \text{ GeV} , \qquad T_c = 1 \text{ GeV} . \qquad (19)$$

This yields the bound

$$f_A \lesssim 10^{12} \text{ GeV} . \qquad (20)$$

So it appears that the estimate from instantons is not likely to be off by more than a factor of 10.

So we see that the current mass density and photon temperature severely constrain the value which θ can take at the grand unified phase transition. The invisible axion idea thus appears to be in serious trouble. In particular, in conventional grand unified theories satisfying the bounds on f_A described above requires a fine-tuning of one part in 10^6–10^{10}. Thus the strong *CP* problem, or the problem of the fine-tuning of θ, has simply been replaced by a fine-tuning of other parameters.

A few solutions to this dilemma may be imagined but none, at present, look particularly attractive. The most obvious is a process occurring after T_c which significantly increases the entropy (by a factor of order 10^4). An example is the decay of the long-lived gravitinos in certain supersymmetric theories [12]. However, no particularly satisfactory model of this type yet exists. Moreover, such a process would also dilute the baryons, which will be difficult to reconcile with current ideas concerning baryon number generation.

[‡2] This is an optimistic estimate. A pessimist would calculate the entropy relevant to a gas of pions and nucleons. The truth is probably somewhere in between, and depends on the details of the true QCD equation of state at and around the chiral symmetry breaking transition. These considerations, however, can change our results only by small factors.

Finally, apart from fine-tuning, there is another disturbing feature of an axion with decay constant less than 10^{12} GeV. As mentioned earlier, the domain wall problem might be solved by a period of exponential expansion. However, in order to generate baryons, this period must be followed by significant reheating of the universe, almost certainly to temperatures above 10^{12} GeV, i.e. above the temperature of the new Peccei–Quinn transition. So the domain wall problem discussed by Sikivie will still be with us.

All of this adds up to a troubling picture. For the moment, at least, the strong *CP* puzzle remains a puzzle.

We thank L. Abbott, P. Sikivie, L. Susskind and E. Witten for instructive conversations. This work was supported by the U.S. Department of Energy under Contract Nos. EY-76-C-02-3071 (W.F.) and DE-ACO2-76ERO2220 (M.D.).

References

[1] R.D. Peccei and H. Quinn, Phys. Rev. Lett. Lett. 38 (1977) 1440.
[2] M. Dine, W. Fischler and M. Srednicki, Phys. Lett. 104B (1981) 199.
[3] J. Kim, Phys. Rev. Lett. 43 (1979) 103.
[4] M. Wise, H. Georgi and S.L. Glashow, Phys. Rev. Lett. 47 (1981) 402.
[5] D.J. Gross, R.D. Pisarski and L.G. Yaffe, Rev. Mod. Phys. 53 (1981) 43.
[6] W.A. Bardeen and S.-H.H. Tye, Phys. Lett. 74B (1978) 229.
[7] R. Dashen and D.J. Gross, Phys. Rev. D23 (1981) 2340.
[8] P. Sikivie, Phys. Rev. Lett. 48 (1982) 1156.
[9] H. Georgi and M. Wise, Phys. Lett. 116B (1982) 123;
G. Lazarides and Q. Shafi, Rockefeller preprint RU89-B027 B027 (1982);
B. Holdom, Stanford preprints ITP-712 (1982), ITP-713 (1982);
S. Dimopoulos, P. Frampton, H. Georgi and M. Wise, Harvard preprint HUTP-82-A030 (1982);
S. Barr, D. Reiss and A. Zee, Phys. Lett. 116B (1982) 227;
S. Bar, X.C. Gao and D. Reiss, University of Washington preprint 40048-22P2 (1982);
Y. Fujimoto, K. Shigemoto and Zhao Zhiyong, ICTP preprint IC-82-410 (1982).
[10] A. Guth, Phys. Rev. D23 (1981) 347.
[11] A. Linde, Phys. Lett. 108B (1982) 389;
A. Albrecht and P. Steinhardt, Phys. Rev. Lett. 48 (1982) 1220.
[12] S. Weinberg, Phys. Rev. Lett. 48 (1982) 1303.

LIMITS ON MASSES AND NUMBER OF NEUTRAL WEAKLY INTERACTING PARTICLES

P. HUT
Institute for Theoretical Physics, University of Utrecht, Utrecht, Netherlands

Received 25 April 1977

Limits on the masses and number of neutral weakly interacting particles are derived using cosmological arguments. No such particles with a mass between 120 eV and 3 GeV can exist within the usual big band model. Similar, but much more severe, restrictions follow for particles that interact only gravitationally. This seems of importance with respect to supersymmetric theories.

Following an idea, put forward by Shvartsman [1], Steigman et al. [2] presented arguments leading to an upper limit to the number of different types of massless neutrinos, which may be summarized as follows.

According to the hot big bang model all forms of matter in the universe, even neutrinos, are initially in thermal equilibrium. The total energy density of relativistic particles is then given at a temperature T by

$$\rho = \kappa a T^4. \tag{1}$$

a is the radiation density constant, appearing in the black-body radiation law, and κ is given by

$$\kappa = \tfrac{1}{2}(n_b + \tfrac{7}{8} n_f). \tag{2}$$

The quantities n_b and n_f are the total number of internal degrees of freedom of the different types of bosons and fermions respectively. For a photon gas $\kappa = 1$, while for a mixture of photons, electrons, electron and muon neutrinos, together with their antiparticles, $\kappa = 9/2$.

A second expression for the total energy density ρ is given as a function of the expansion time t by solving the Einstein equations in a radiation dominated homogeneous and isotropic universe,

$$\rho = 3/32\,\pi G t^2, \tag{3}$$

where G is the gravitational coupling constant, $G = 6.7 \times 10^{-45}$ MeV^{-2}‡. Combining (1) and (3) we get

$$T = (3/32\,\pi G a)^{1/4}\, \kappa^{-1/4}\, t^{-1/2}. \tag{4}$$

‡ We use units such that $\hbar = c = k = 1$, and the temperature is expressed in MeV.

Adding more types of neutrinos relative to the standard big bang model increases the value of κ. This would have the following observable effect.

The neutron/proton ratio is given by the equilibrium value $n/p = \exp\{-(m_n - m_p)/T\}$ as long as the rate of weak interactions, like e.g. $n + e^+ \leftrightharpoons p + \bar{\nu}_e$, is high enough. But this ratio freezes in soon after the time between successive collisions grows bigger than, say, the expansion time. The mean free time is $\tau = (\sigma N)^{-1}$ as long as the electrons are relativistic. The cross section $\sigma \sim T^2$ and the number density of protons and neutrons $N \sim R^{-3}$, where R is the scale factor of the expanding universe. At these early times the number of nucleons is far smaller than the number of photons, electrons, positrons and neutrinos, so the cooling proceeds adiabatically like $T \sim R^{-1}$. Therefore $N \sim T^3$ and thus

$$\tau = \text{const.} \times T^{-5}. \tag{5}$$

Putting $t = \tau$ in (4), from (5) we get an effective temperature T_f at which the neutron/proton ratio freezes in, given by

$$T_f = \text{const.} \times \kappa^{1/6}. \tag{6}$$

When the temperature falls off further nearly all neutrons are captured to form deuterium and subsequently helium. In the standard model $T_f \approx 1$ MeV $\approx 10^{10}$ K and the abundance by weight of helium produced in this way is $Y \approx 0.23$ to 0.27, depending on the present density of nucleons in the universe. An observational upper limit [4] $Y \lesssim 0.29$ agrees well with the standard model.

Increasing now the number of neutrino types would

increase κ and via (6), T_f. Therefore the neutron/proton ratio would freeze in at a higher value, and more helium would be formed. Since there is not much room left between the observational upper limit to Y and the predictions of the standard model, rather strong restrictions can be posed on κ. Steigman et al. [2], using numerical calculations by Wagoner [3], find $\kappa \lesssim 9$, twice the standard value $\kappa = 9/2$ at temperatures around $T_f \approx 1$ MeV ($m_\mu > T_f > m_e$). Assuming a doublet structure for heavy leptons they concluded from (2) that the total number of heavy leptons should be less than or equal to 5.

This argument can be extended and combined with arguments concerning the present matter density of the universe to apply to massive neutral weakly interacting particles in general, to be called massive neutrinos for short. No detailed assumption about the multiplet structure of which such neutrinos would be part is needed in this context. Also their spin is not essential, only the number of degrees of freedom is relevant.

Suppose now the existence of new neutrinos with a mass less than 1 MeV. Then these particles would be copiously present till the temperature dropped to $T \approx 1$ MeV. This is also the temperature at which the neutron/proton ratio freezes in, more specifically the temperature at which the universe becomes transparent for weak processes. Therefore neutrinos having a mass less than 1 MeV are as effective in the previous argument concerning the helium abundance as are massless neutrinos.

Neutrinos more massive than 1 MeV disappear by annihilation before the time t_f when the neutron/proton ratio freezes in. Therefore no observational consequences arise from such neutrinos.

The argument concerning the observed helium abundance applies also to unstable neutrinos provided their lifetime exceeds 1 second, which is roughly the mean free time at $T \approx 1$ MeV. Otherwise they do not play a significant role in the equilibrium situation.

Further restrictions on the number of stable neutrino types are imposed by the present lifetime of the universe, being 10^{10} years or more. Together with the observed expansion rate of the universe this puts an upper limit on the mean mass density of about 10^{-29} g/cm^3. Since for highly relativistic particles the energy density varies like $\rho \sim T^4$, as opposed to nonrelativistic particles for which $\rho \sim T^3$, the universe is initially dominated by radiation until the temperature drops to $T \approx 10^{-7} - 10^{-5}$ MeV (depending on the present mass density). Hereafter the energy density of the different forms of radiation left drops below the density of massive particles. If there would exist a stable kind of neutrino having a mass in the keV region, these particles would be present copiously at the time when the universe became transparent for weak interactions. Now they cool by expansion and they would soon become nonrelativistic. Therefore at present they would have a mass density far greater than the upper limit posed by the lifetime of the universe. Several authors have used this argument to find an upper limit to the rest mass of electron and muon neutrinos [5].

For light stable neutrinos an upper limit to the mass is computed in this way by Szalay and Marx [5]. The result is 30 eV in the case of electron or muon neutrinos, assuming two degrees of freedom per particle. Allowing a particle with only one degree of freedom which is its own antiparticle leads thus to an upper limit of 120 eV. At first sight this seems to imply for a particle with four degrees of freedom, accompanied by an antiparticle, a mass limit of 15 eV, but actually this value is about 1.5 eV. The reason is that the observed helium abundance is only compatible with eight extra degrees of freedom if the universe contains a mass density of at most 1/10 of the critical mass density [2]. These results are drawn in fig. 1 (full line).

For heavy stable neutrinos a lower limit to the mass m can be obtained by the following argument. When the temperature drops to $T \approx m$, most of the heavier particles have vanished already. Therefore the neutrinos can disappear only by annihilation with their antiparticles, and we have to estimate the temperature $T_f(m)$ at which annihilation effectively stops.

Suppose for simplicity that the neutrinos are in thermal equilibrium till the temperature drops to $T_f(m)$ and that they are completely free afterwards. This temperature can again be estimated by equating the mean free time τ between reactions involving only these heavy neutrinos and their antiparticles with the expansion time of the universe. For massless neutrinos it follows from (4) and (5) that

$$t/\tau = [T/1\ \text{MeV}]^3, \qquad (7)$$

where the constant of proportionality is determined by the temperature $T \approx 1$ MeV when the universe becomes transparent for weak interactions. For heavy

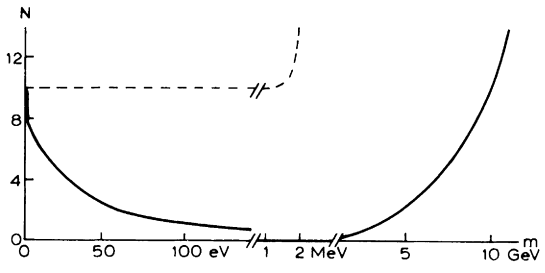


Fig. 1. Upper limits to the sum of the number of the internal degrees of freedom of the different types of neutral weakly interacting particles, apart from the electron and muon neutrinos. The full line applies to stable particles; the dashed line to unstable particles if their halflife is more than 1 second. m denotes the mass of the particles, while N gives the number of degrees of freedom for fermions. For bosons, multiply N by a factor 7/8.

neutrinos, however, the energy density $\rho_L(m)$ compared to that of electron neutrinos at $T = T_f$ is approximately

$$\rho_L(m)/\rho_\nu = \tfrac{1}{2} \exp(-m/T_f(m)). \quad (8)$$

We assume here that the heavy neutrinos have only one degree of freedom and are their own antiparticle. Therefore the reaction rate τ^{-1} gets this extra dilution factor, and (7) goes over into

$$t/\tau = \tfrac{1}{2}[T_f(m)/1\ \text{MeV}]^3 \exp(-m/T_f(m)) \equiv 1, \quad (9)$$

or

$$m = 3T_f(m)\{\ln T_f(m) - 0.23\}. \quad (10)$$

Now we must estimate the influence of the left-over neutrinos on the present mass density, which must obey

$$\rho_{L0} < 10^{-29}\ \text{g/cm}^3 \approx 5 \times 10^{-3}\ \text{MeV/cm}^3. \quad (11)$$

According to the standard model the present temperature $T_{\nu 0}$ of electron neutrinos is related to that of the photons by $T_{\nu 0} = (4/11)^{1/3} T_{\gamma 0} = 1.7 \times 10^{-10}$ MeV, where the observed background radiation temperature $T_{\gamma 0} = 2.7\ \text{K} = 2.3 \times 10^{-10}$ MeV. The present energy density of electron neutrinos is thus given by

$$\rho_{\nu 0} = 1.1 \times 10^{-34}\ \text{g/cm}^3 = 6.2 \times 10^{-8}\ \text{MeV/cm}^3, \quad (12)$$

resulting in the inequality

$$\rho_{L0}/\rho_{\nu 0} < 10^5. \quad (13)$$

The energy density of the heavy neutrinos falls off like $\rho_L \sim T^3$, since they are already nonrelativistic at $T = T_f$, while $\rho_\nu \sim T^4$. From (13) it follows that at $T = T_f$

$$\frac{\rho_L(m)}{\rho_\nu} = \frac{T_{\nu 0}}{T_f(m)} \cdot \frac{\rho_{L0}}{\rho_{\nu 0}} < 2[T_f(m)/10^{-5}\ \text{MeV}]^{-1}. \quad (14)$$

Comparing this with (8) and (10) gives

$$m > 3000\ \text{MeV} \quad (15)$$

for the mass of a heavy neutrino with one degree of freedom. Similar calculations for more degrees of freedom are easily done and result in the full line in fig. 1.

In summary, stable massive neutral leptons cannot exist in the mass range 120 eV – 3000 MeV. Below 60 eV at most 10 degrees of freedom are available for fermions. For instance, only two types of neutrinos are allowed if they have two spin states per particle and antiparticle. Bosons have a weight factor of 8/7 with respect to fermions. Above 3000 MeV the allowed number rises quickly (see fig. 1).

Less stringent upper limits can be put on the number of unstable massive neutrino types. If their lifetime is more than 1 second the dashed line in fig. 1 applies, but for much shorter lifetimes no interesting limits can be obtained.

Finally the same arguments can be applied to particles which interact only gravitationally. For them the universe becomes transparent already at a much earlier time, estimated by Matzner [6] to be 10^{-16} sec. at a temperature of 10^{20} K $\approx 10^{10}$ MeV. These particles thus have a forbidden mass range of 60 eV – 10^{19} MeV. The upper limit of this range is still well below the Planck mass, being about 10^{22} MeV. Of course this limit is less reliable than the upper limit found for weakly interacting particles since not much is known for such early times after the big bang. But at least it gives a strong indication that the upper limit is very much more than, say, 1 TeV. Assuming the standard big bang model to be of relevance in the con-

text of supergravity theories, one can make the following remark. If there exist light massive spin 3/2 particles interacting only gravitationally, having four spin degrees of freedom, their mass must be less than 15 eV if they are their own antiparticles, otherwise their mass is less than 1.5 eV. Also they may exist with masses very much larger than 1 TeV.

The author is indebted to Prof. M. Veltman for suggesting this investigation, and for helpful comments and criticism.

References

[1] V.F. Shvartsman, JETP Lett. 9 (1969) 184.
[2] G. Steigman et al., Phys. Lett. 66B (1977) 202.
[3] R.V. Wagoner, Astrophys. J. 179 (1973) 343.
[4] M. Peimbert, Ann. Rev. Astron. and Astrophys. 13 (1976) 113.
[5] Y.B. Zel'dovich and I.D. Novikov, Relativitskaya Astrofizika [Nauka, Moscow] (1969);
R. Cowsik and J. McClelland, Phys. Rev. Lett. 29 (1972) 669.
A.S. Szalay and G. Marx, Astron. and Astrophys. 49 (1976) 437.
[6] R.A. Matzner, Astrophys. J. 154 (1968) 1123.

PHYSICAL REVIEW
LETTERS

VOLUME 39 | 25 JULY 1977 | NUMBER 4

Cosmological Lower Bound on Heavy-Neutrino Masses

Benjamin W. Lee[a]
Fermi National Accelerator Laboratory,[b] *Batavia, Illinois 60510*

and

Steven Weinberg[c]
Stanford University, Physics Department, Stanford, California 94305
(Received 13 May 1977)

The present cosmic mass density of possible stable neutral heavy leptons is calculated in a standard cosmological model. In order for this density not to exceed the upper limit of 2×10^{-29} g/cm^3, the lepton mass would have to be *greater* than a lower bound of the order of 2 GeV.

There is a well-known cosmological argument[1] against the existence of neutrino masses greater than about 40 eV. In the "standard" big-bang cosmology,[2] the present number density of each kind of neutrino is expected[3] to be $\frac{6}{11}$ the number density of photons in the 3°K black-body background radiation, or about 300 cm^{-3}; hence if the neutrino mass were above 40 eV, their mass density would be greater than 2×10^{-29} g/cm^3, which is roughly the upper limit allowed by present estimates[4] of the Hubble constant and the deceleration parameter.

However, this argument would not apply if the neutrino mass were much larger than 1 MeV. Neutrinos are generally expected[2] to go out of thermal equilibrium when the temperature drops to about 10^{10} °K, the temperature at which neutrino collision rates become comparable to the expansion rate of the universe. If neutrinos were much heavier than 1 MeV, then they would already be much rarer than photons at the time when they go out of thermal equilibrium, and hence their number density would now be much less than 300 cm^{-3}.

Of course, the familiar electronic and muonic neutrinos are known to be lighter than 1 MeV. However, heavier stable neutral leptons could easily have escaped detection, and are even required in some gauge models.[5] In this Letter, we suppose that there exists a neutral lepton L^0 (the "heavy neutrino") with mass well above 1 MeV, and we assume that L^0 carries some additive or multiplicative quantum number which keeps it absolutely stable. We will present arguments based on the standard big-bang cosmology to show that the mass of such a particle must be above a lower bound of order 2 GeV.

At first glance, it might be thought that the present number density of heavy neutrinos would simply be less than the above estimate of 300 cm^{-3} by the value $\exp[-m_L/(1\text{ MeV})]$ of the Boltzmann factor at the time the heavy neutrinos go out of thermal equilibrium. If this were the case, then an upper limit of 2×10^{-29} g/cm^{-3} on the present cosmic mass density would require that $m_L\exp[-m_L/(1\text{ MeV})]$ should be less than 40 eV, and hence that m_L should either be less than 40 eV or greater than 13 MeV.

However, the true lower bound on the heavy-neutrino mass is considerably more stringent.

The heavy neutrinos are assumed to carry a conserved quantum number, so their number density can relax to its equilibrium value n_0 only by annihilation of a heavy neutrino with a heavy antineutrino, in processes such as

$$L^0\bar{L}^0 \to \nu\bar{\nu}, \quad e^-e^+, \quad \mu^-\mu^+, \quad \pi^-\pi^+, \quad \text{etc.} \tag{1}$$

But although the energy distribution of the heavy neutrinos is kept thermalized by collisions with ν, e, etc., down to a temperature of 10^{10} °K, at that temperature they are so rare that their annihilation rate is already much less than the cosmic expansion rate. Thus the heavy neutrinos go out of *chemical* equilibrium (in the sense that their number density begins to exceed its equilibrium value) at a "freezing" temperature T_f which is much higher than 1 MeV. The condition on m_L is then that $m_L \exp(-m_L/kT_f)$ should be less than 40 eV, and the resulting lower bound on m_L must therefore be greater than 13 MeV.

To make this quantitative, we use the rate equation

$$\frac{dn}{dt} = -\frac{3\dot{R}}{R}n - \langle\sigma v\rangle n^2 + \langle\sigma v\rangle n_0^2. \tag{2}$$

Here n is the actual number density of heavy neutrinos at time t; R is the cosmic scale factor; $\langle\sigma v\rangle$ is the average value of the $L^0\bar{L}^0$ annihilation cross section times the relative velocity; and n_0 is the number density of heavy neutrinos in thermal (and chemical) equilibrium[6]:

$$n_0(T) = \frac{2}{(2\pi)^3}\int_0^\infty 4\pi p^2 dp\left[\exp\left(\frac{(m_L^2+p^2)^{1/2}}{kT}\right)+1\right]^{-1}. \tag{3}$$

(We use units with $\hbar = c = 1$ throughout.)

At the temperatures we are considering here, the energy density and the entropy are dominated by highly relativistic particles, including γ, ν_e, ν_μ, and e. It follows that R, T, and t are related by

$$\dot{R}/R = -\dot{T}/T = (8\pi\rho G/3)^{1/2}, \tag{4}$$

where ρ is the energy density

$$\rho = N_F a T^4 = N_F \pi^2 (kT)^4/15, \tag{5}$$

with N_F an effective number of degrees of freedom, counting $\frac{1}{2}$ and $\frac{7}{16}$, respectively, for each boson or fermion species and spin state. For temperatures in the range of 10–100 MeV (which most concern us here) we must include just γ, ν_e, $\bar{\nu}_e$, ν_μ, $\bar{\nu}_\mu$, e^-, and e^+, so $N_F = 4.5$, a value we will adopt for most purposes. However, if current ideas about the strong interactions are correct, then N_F rises steeply at a temperature equilibrium because, as we have seen, the upper limit on the present cosmic mass density requires the Boltzmann factor $\exp(-m_L/kT_f)$ to be very small. For nonrelativistic velocities, the cross-section $\sigma(v)$ for the exothermic processes (1) behaves like $1/v$, so we can take $\langle\sigma v\rangle$ as a temperature-independent constant. If there were only a single annihilation channel open (say $L^0\bar{L}^0 \to \nu\bar{\nu}$) and if the annihilation process were due to an ordinary "V-minus-A" charged-current Fermi interaction, then $\langle\sigma v\rangle$ would be just $G_F^2 m_L^2/2\pi$, where G_F is the Fermi coupling constant 1.15×10^{-5} GeV^{-2}. To take account of more general possibilities, we shall write

$$\langle\sigma v\rangle = G_F^2 m_L^2 N_A/2\pi, \tag{6}$$

where N_A is a dimensionless fudge factor which depends both on the number of annihilation channels open and on the details of the weak $L^0\bar{L}^0$ annihilation interaction. (In general, the annihilation proceeds both through exchange of charged and of neutral intermediate vector bosons.) Where a numerical value is needed, we will take $N_A = 14$. [This corresponds to the annihilation channels $L^0\bar{L}^0 \to \nu_e\bar{\nu}_e$, $\nu_\mu\bar{\nu}_\mu$, $\nu_\tau\bar{\nu}_\tau$, e^-e^+, $\mu^-\mu^+$, $u\bar{u}$, $d\bar{d}$, $s\bar{s}$, with all fermions (antifermions) of helicity $-\frac{1}{2}$ ($+\frac{1}{2}$), and three colors for each of the quarks u, d, and s.] New channels open for m_L above about 5 GeV, leading to a moderate rise in N_A.

Equations (3)–(6) allow us to rewrite the rate equation (2) in the convenient form

$$df/dx = C\mu^3(f^2 - f_0^2), \tag{7}$$

where

$$\begin{aligned} &x = kT/m_L, \quad \mu^3 = m_L^3 N_A/\sqrt{N_F}, \quad f(x;\mu) = n/T^3, \\ &f_0(x) = n_0/T^3 \\ &\quad = \pi^{-2}k^3\int_0^\infty du\, u^2[\exp(x^{-2}+u^2)^{1/2}+1]^{-1}, \end{aligned} \tag{8}$$

and

$$\begin{aligned} C &= \frac{1}{k^3}G_F^2\left(\frac{45}{32\pi^5 G}\right)^{1/2} \\ &= 1.28\times 10^6\left(\frac{\text{cm °K}}{\text{GeV}}\right)^3. \end{aligned} \tag{9}$$

Equation (7) is to be solved subject to the initial condition that, as $x \to \infty$, $f(x)$ approaches the equilibrium value $f_0(x)$. For $x \ll 1$, $f(x)$ approaches an asymptotic value, which evidently depends only on

the single unknown parameter μ:

$$\lim_{x\to 0} f(x;\mu) = F(\mu). \tag{10}$$

The subsequent annihilation of electron-positron pairs increases RT by a factor $(\frac{11}{4})^{1/3}$, and hence decreases n/T^3 by a factor $\frac{4}{11}$, so the present mass density of heavy neutrinos (and antineutrinos) will be given by

$$\rho_L = 2(\tfrac{4}{11}) m_L T_\gamma{}^3 F(\mu), \tag{11}$$

with $T_\gamma = 3°\mathrm{K}$ the present radiation temperature. (For freezing temperature above 100 MeV, the factor $\frac{4}{11}$ must be decreased to take account of the subsequent heating of photons by annihilation of $\mu^-\mu^+$, $\pi^-\pi^+$, etc.)

Computer solutions of Eq. (7) are shown for a variety of special cases in Fig. 1. We find that over the range of parameters considered here, the function $F(\mu)$ of Eq. (10) can be very well represented by

$$F(\mu) \simeq 1.20\times 10^{-5}\ (\mathrm{cm}\ °\mathrm{K})^{-3}[\mu(\mathrm{GeV})]^{-2.85}. \tag{12}$$

The present mass density of heavy neutrinos and antinuetrinos is then given by Eq. (11) as

$$\rho_L = (4.2\times 10^{-28}\ \mathrm{g/cm^3}) \times [m_L(\mathrm{GeV})]^{-1.85}(N_A/\sqrt{N_F})^{-0.95}. \tag{13}$$

If we require[4] that $\rho_L < 2\times 10^{-29}\ \mathrm{g/cm^3}$, then m_L must be subject to the lower bound

$$m_L(N_A/\sqrt{N_F})^{0.51} \gtrsim 5.2\ \mathrm{GeV}. \tag{14}$$

For $N_A = 14$ and $N_F = 4.5$, this gives $m_L \gtrsim 2$ GeV. Allowing for a factor-of-4 uncertainty in the $L^0\bar{L}^0$ annihilation rate, the lower bound on m_L would lie in the range 1–4 GeV.

It is enlightening to see how these results can be obtained by an analytic approximation to the asymptotic value of n/T^3. We expect that $f(x)$ remains approximately equal to $f_0(x)$ until the temperature drops to a freezing value T_f, at which the annihilation rate per unit volume equals the rate of change of n_0, or

$$df_0/dx = C\mu^3 f_0^2, \text{ at } x = x_f = kT_f/m_L, \tag{15}$$

and that thereafter f obeys approximately the equation

$$df/dx = C\mu^3 f^2, \quad x < x_f, \tag{16}$$

with the initial condition $f(x_f) = f_0(x_f)$. Since $kT_f \ll m_L$, we can approximate f_0 in Eq. (8) by the nonrelativistic formula

$$f_0(x) = n_0/T^3 \simeq 2k^3(2\pi x)^{-3/2}e^{-1/x} \tag{17}$$

and take only the exponential into account in evaluating df_0/dx. The freezing temperature is then obtained from

$$\begin{aligned} x_f{}^{-1/2}\exp(1/x_f) &= Ck^3 2(2\pi)^{-3/2}\mu^3 \\ &= 1.35\times 10^7 N_A N_F{}^{-1/2}[m_L(\mathrm{GeV})]^3. \end{aligned} \tag{18}$$

For μ in the range 1 to 10 GeV, we find $x_f{}^{-1}$ between 15 and 22. Thus, from Eqs. (17) and (18), we find that

$$f(x_f) \simeq f_0(x_f) = (C\mu^3 x_f{}^2)^{-1}. \tag{19}$$

Equation (16) can now be solved:

$$f(x) = [f_0(x_f)^{-1} + C\mu^3(x_f - x)]^{-1}, \tag{20}$$

so that, with $x_f{}^{-1} \simeq 20$,

$$\begin{aligned} F = f(0) &= (C\mu^3 x_f{}^2)^{-1}(1 + x_f{}^{-1})^{-1} \\ &\simeq 1.5\times 10^{-5}[\mu\,(\mathrm{GeV})]^{-3}(\mathrm{cm}\ °\mathrm{K})^{-3} \end{aligned} \tag{21}$$

in reasonable agreement with the more accurate result (12). This analysis shows that the freezing temperature is likely to be about $\frac{1}{20}$ the heavy-neutrino mass, or 100 MeV for $m_L = 2$ GeV.

Of course, if a stable heavy neutral lepton were discovered with a mass of order 1–15 GeV, the gravitational field of these heavy neutrinos

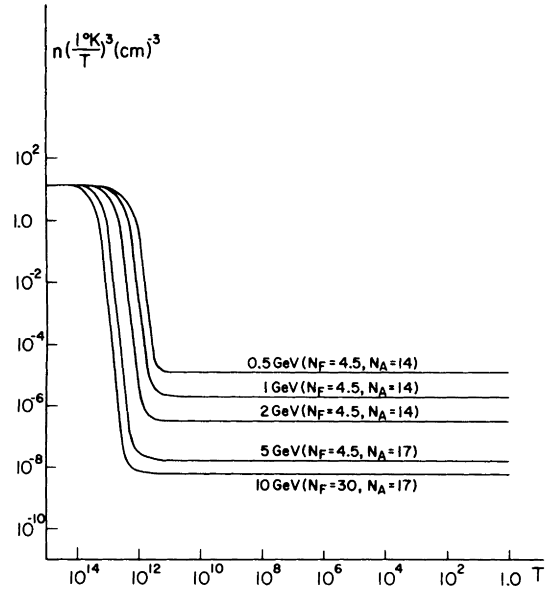


FIG. 1. n/T^3 vs T for a variety of special cases of m_L, N_F, and N_A.

would provide a plausible mechanism for closing the universe.

One of us (S.W.) would like to thank Robert Wagoner for helpful conversations, and the other (B.W.L.) wishes to thank Robert Pearson for his assistance in computer calculations.

---

[a]Deceased.

[b]Operated by Universities Research Association Inc. under contract with the U. S. Energy Research and Development Administration.

[c]On leave 1976–1977 from Harvard University.

[1]See e.g., G. Marx and A. S. Szalay, in *Proceedings of the Neutrino-1972 Conference*, edited by A. Frankel and G. Marx (OMDK-TECHNIOFORM, Budapest, 1972), Vol. I, p. 191; R. Cowsik and J. McClelland, Phys. Rev. Lett. 29, 669 (1972), and also Astrophys. J. 180, 7 (1973).

[2]For a review with references to the original literature, see S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, New York, 1972), Chap. 15.

[3]This includes a factor $\frac{3}{4}$ for the effects of Fermi statistics; a factor $\frac{4}{11}$ because electron-positron annihilation has increased the photon temperature by an extra factor of $(\frac{11}{4})^{1/3}$; and a factor of 2 because both neutrinos and antineutrinos are included. The massive neutrinos are, like the photons, supposed to exist in two helicity states. (The 40-eV upper limit given here differs from the 8-eV bound given by Cowsik and McClelland in Ref. 1, because we take account of the heating of photons by electron-positron annihilation, and because we do not assume that the muon and electron-type neutrinos have equal mass.)

[4]To take account of the uncertainty in galactic evolution, we assume that $q_0 < 2$. A value $H_0 = 50.3 \pm 4.3$ km/sec/Mpc is given by A. Sandage and G. A. Tammann, Astrophys. J. 210, 7 (1976). The desntiy 2×10^{-29} g/cm^3 corresponds (for zero cosmological constant) to $q_0 = 2$ and $H_0 = 50$ km/sec/Mpc. [These quantities are defined in terms of the cosmic scale factor $R(t)$ as $(dR/dt)_0 \equiv R_0 H_0$ and $(d^2R/dt^2)_0 \equiv -R_0 H_0^2 q_0$, a subscript 0 denoting the present instant.]

[5]See, e.g., B. W. Lee and S. Weinberg, Phys. Rev. Lett. 38, 1237 (1977). It is possible that the heavy neutral leptons in this model actually have lifetimes which are shorter than the age of the universe, in which case the lower bound on m_L derived here need not be satisfied. [For calculations of neutral-lepton decay rates, see J. N. Bahcall, N. Cabibbo, and A. Yahil, Phys. Rev. Lett. 28, 318 (1972); W. J. Marciano and A. I. Sanda, Phys. Lett. 67B, 303 (1977); T. Goldman and G. J. Stephenson, Jr., Los Alamos Laboratory Report No. LA-UR-77-941 (to be published); and H. Fritsch and P. Minkowski, Phys. Lett. 62B, 72 (1976).]

[6]We suppose here that the chemical potentials associated with any conserved quantum numbers carried by L^0 are vanishingly small. If there were any appreciable degeneracy, the mass density of the heavy neutrinos would greatly exceed reasonable cosmological limits.

[7]This includes contributions of $\gamma, \nu_e, \bar{\nu}_e, \nu_\mu, \bar{\nu}_\mu, e^-, e^+, \mu^-, \mu^+$; plus three color triplets u, d, s, of quarks and the corresponding antiquarks; plus eight massless spin-one gluons.

Cosmological limits on the masses of neutral leptons

M. I. Vysotskiĭ and A. D. Dolgov

Institute of Theoretical and Experimental Physics

Ya. B. Zel'dovich

Institute of Applied Mathematics, USSR Academy of Sciences
(Submitted June 30, 1977)
Pis'ma Zh. Eksp. Teor. Fiz. **26**, No. 3 200–202 (5 August 1977)

Cosmological arguments are presented which forbid the existence of stable weakly interacting particles in the mass interval 30 eV $< m <$ 2.5 GeV. Limits are also imposed on the masses of new neutral leptons if the latter are unstable.

PACS numbers: 95.30.Cq, 14.60.Gh

Recently experimenters at SLAC[1] announced the discovery of a new charged lepton $\tau^{\pm}$ with mass ~ 1.8 GeV and with a suggested decay channel $\tau \to \mu(e) + \nu_{\mu(e)} + \nu_{\tau}$. A study of the decay spectrum imposes a rather weak limitation on the mass of the new neutrino $m(\nu_{\tau}) < 600$ MeV (90% confidence level).[2] We shall show that cosmological arguments lead to a much stronger limitation: $m_{\nu_{\tau}} < 30$ eV. For a small mass [$m(\nu_{\tau}) < 3$ meV] we can reason in the same manner as in the case of the muon neutrino. For the latter, Gershteĭn and Zel'dovich[3] using cosmological arguments, obtained $m(\nu_{\mu}) < 400$ eV. It is stated in[4] that if it is assumed that $m(\nu_e) = m(\nu_{\mu})$ then the limit is $m < 8$ eV. Finally, arguments are advanced in[5] favor $m(\nu_{\mu}) = 15$ eV. The idea in the cited paper is the following. When the temperature of the expanding universe is high, i.e., $T \gtrsim m_{\nu}$, the density of the massive neutrinos is the same as that of the photons. With further expansion and cooling ($T < m_{\nu}$) the annihilation $\nu + \tilde{\nu} \to$ anything does not have time to take place (if $m_{\nu} \lesssim 3$ MeV). At the present time the temperature and the density of the electromagnetic relict radiation are respectively 2.7 K and 400 cm^{-3}. Taking into account the arguments advanced in[6], we obtain $n_{\nu} + n_{\tilde{\nu}} \approx 200$ cm^{-3}, and the corresponding mass density is $\rho_{\nu} = 200\ m_{\nu}$ g/cm^3.

At the contemporary value of the Hubble constant, the critical density is $\rho_c = 5 \times 10^{-30}$ g/cm^3. The large density leads to a closed model of the universe, to a large slowing-down parameter $q_0 = \Omega/2 = \rho/2\rho_c > 0.5$, and to a small age of the universe, $T < 12 \times 10^9$ years. Astrophysical observations contradict this picture. With reasonable caution we can conclude that $\rho_{\nu} < 2\rho_c$. This corresponds to $m_{\nu} < 30$ eV.

For ν_{τ} it is necessary to consider also the case $m_{\nu_{\tau}} > 3$ MeV. With increasing mass, the annihilation cross section increases, so that the assumption $n(\nu_{\tau}) \approx n(\gamma)$ now turns out to be incorrect. We estimate the residual (quenched) concentration of ν_{τ}, following,[7] where the concentration of the residual quarks was considered.

For the rate of the $\nu_{\tau}\bar{\nu}_{\tau}$ annihilation at rest in the Weinberg-Salam model,

0021-3640/77/2603-0188$00.60 © 1978 American Institute of Physics

with allowance for the four types of light leptons, we obtain $\sigma v = G^2 m^2/\pi$, where $G = 10^{-5}/m_p^2$, $m \equiv m_{\nu_\tau}$.

The time dependence of the relative concentration $r = n(\nu_\tau)/n(\gamma)$ is determined by the equation

$$\frac{dr}{dt} = -\sigma v n(\gamma)(r^2 - r_{eq}^2), \tag{1}$$

where r_{eq} is the thermodynamic equilibrium concentration. For small values of $\Theta = T/m$ $r_{eq} \approx \Theta^{-3/2}\exp(-1/\Theta)$. $\Theta = t^{-1/2}$ (sec)m^{-1} (MeV) so that r_{eq} decreases rapidly with increasing t. The instant of quenching is determined by the condition $4\sigma v n_{eq}(\nu_\tau) t\,\Theta = 1$. After this instant, r_{eq} becomes small in comparison with the real concentration. The residual (as $t \to \infty$) relative concentration is determined by integrating Eq. (1) with $r_{eq} = 0$. The result is

$$r_\infty \approx 3 \times 10^{-7}(m_p/m)^3. \tag{2}$$

Numerical integration of Eq. (1) gives the same answer for $m_\nu \approx 2.5\, m_p$. The increase of the neutrino density in comparison with quarks[7] is due to the smallness of the annihilation cross section. The present-day energy density of new neutrinos is

$$\rho(\nu_\tau) = \frac{4}{11} n(\gamma) m 3 \times 10^{-7}(m_p/m)^3 = 7 \times 10^{-29}(m_p/m)^2 \ \text{g/cm}^3, \tag{3}$$

where the factor $\frac{4}{11}$ is connected with the increase of the photon density as a result of e^+e^- annihilation.[6] From the upper bound of ρ we obtain $m(\nu_\tau) > 2.5$ GeV. Thus, the only remaining possibility is $m(\nu_\tau) < 30$ eV.

It should be noted that the condensation of matter into stars can lead to an increase of the density ν_τ and to the appearance of secondary annihilation. If a fraction β of the heavy neutrinos annihilates at a certain red shift z, the energy density of the decay products will at the present time be smaller by a factor $(1+z)$. The resultant energy density, including the remaining heavy neutrinos and products of their decay, is equal to

$$\rho = \rho_0(1-\beta) + \frac{\rho_0 \beta}{1+z}, \tag{4}$$

where ρ_0 is defined by Eq. (3).

At first glance it appears that at $(1-\beta) \ll 1$ and $z \gg 1$ this makes the limits on the mass of ν_τ much worse. Actually, about half of entire energy release in annihilation goes over ultimately into electromagnetic radiation. The density of the relict radiation is 4×10^{-34}, that of the optical radiation is 4×10^{-36}, and that of the x rays is 10^{-37} g/cm^3. Even the weakest limit yields

$$\frac{\rho\beta}{1+z} < 4 \times 10^{-34} \ \text{g/cm}^3,$$

and for example, for $\beta = 0.1$ and $z = 100$ we obtain $\rho < 4 \times 10^{-31}$ and $m > 12$ GeV. We see thus that the possibility of secondary annihilation improves the lower limit for $m(\nu_\tau)$.

The obtained limits are valid for a stable neutral lepton, i.e., for a lepton having a new conserved quantum number. If the new neutral lepton is of the electronic or muonic type[1)] it can decay into $\nu\bar{\nu}\nu_e$ or $\nu_e\gamma$. The lifetime of this particle will be of the order of $\tau = 2 \times 10^{-6}(m_\mu/m)^5$ sec. This time exceeds the

cosmological ($t_c = 6 \times 10^{17}$ sec) if $m < 2 \times 10^{-5}\, m_\mu = 2$ keV. This condition, however, is not necessary. If the particle has decayed at a certain instant $t = t_c(1+z)^{-3/2}$, the energy density at the present time will be smaller by a factor $(1+z)$. Even this can be too large. In fact

$$\rho_{\text{today}} = \frac{200m}{1+z} = 200m(t/t_c)^{2/3}$$

and this will be less than ρ_c if $m > 10^{-4} m_\mu = 10$ keV.

A model for the violation of the muon charge was proposed in[8]. In this model $W(\nu \to \nu_e + \gamma) = 10^{-2}\,(m_\nu/m_\mu)^5$ sec^{-1}. In this case the cosmological considerations exclude the mass interval 30 eV $< m(\nu_\tau) <$ 2 MeV.

It is most probable that the masses of all the neutrinos (ν_e, ν_μ, ν_τ) are equal to zero. It is puzzling why the electric charge in the (l, ν_l) pairs is connected with the mass. Another interesting problem connected with the neutrinos is whether there exists a right-hand neutrino having no weak interaction, but with undisputed gravitational interaction.[9] Using the ideas of[10], a limitation on the number of types of neutrinos was obtained in[11].

The work was reported at the Session of the Division of Nuclear Physics of the USSR Academy of Sciences on May 26, 1977. On June 16 we received a preprint[12] containing analogous results for the case of stable neutrinos.

We are grateful to L. B. Okun' for support and interest in the work.

[1)]It should be remembered that if τ is of the electronic type, then the existence of the new neutrino is not obligatory.

[1]M. L. Perl, Phys. Rev. Lett. 35, 1489 (1975).
[2]M. L. Perl, Phys. Lett. B 63, 466 (1976).
[3]S. S. Gershteĭn and Ya. B. Zel'dovich, Pis'ma Zh. Eksp. Teor. Fiz. 4, 174 (1966) [JETP Lett. 4, 120 (1966)].
[4]R. Cowsik and J. McClelland, Phys. Rev. Lett. 29, 669 (1972).
[5]A. S. Szaloy and G. Marx. Astronomy and Astrophysics 49, 437 (1976).
[6]P. I. E. Peebles, Phys. Rev. Lett. 16, 410 (1966).
[7]Ya. B. Zel'dovich, L. B. Okun', and S. B. Pikel'ner. Usp. Fiz. Nauk 87, 113 (1965) [Sov. Phys. Usp. 8, 702 (1966)].
[8]J. D. Bjorken and S. Weinberg, Phys. Rev. Lett. **38**, 622 (1977).
[9]I. Yu. Kobzarev and L. B. Okun', Zh. Eksp. Teor. Fiz. **43**, 1904 (1962) [Sov. Phys. JETP **16**, 1343 (1963)].
[10]V. F. Shvartsman, Pis'ma Zh. Eksp. Teor. Fiz. **9**, 315 (1969) [JETP Lett. **9**, 184 (1969)].
[11]G. Steigman, D. M. Schramm, and J. E. Cunn, Phys. Lett. B 66, 202 (1977).
[12]B. W. Lee and S. Weinberg, FERMILAB-PUB-77/41-THY, May 1977.

THE ASTROPHYSICAL JOURNAL, 223:1015–1031, 1978 August 1
© 1978. The American Astronomical Society. All rights reserved. Printed in U.S.A.

SOME ASTROPHYSICAL CONSEQUENCES OF THE EXISTENCE OF A HEAVY STABLE NEUTRAL LEPTON

J. E. GUNN*
California Institute of Technology; and Institute of Astronomy, Cambridge, England
B. W. LEE†
Fermi National Accelerator Laboratory;‡ and Enrico Fermi Institute, University of Chicago
I. LERCHE
Enrico Fermi Institute and Department of Physics, University of Chicago
D. N. SCHRAMM
Enrico Fermi Institute and Departments of Astronomy and Astrophysics and Physics, University of Chicago
AND
G. STEIGMAN
Astronomy Department, Yale University
Received 1977 December 1; accepted 1978 February 14

ABSTRACT

Recently, high-energy particle theorists have constructed new extended gauge theories which may fit experiment somewhat better than previous already very successful theories. One of the predictions which is often discussed is the possible existence of a stable neutral lepton, probably with a mass of a few GeV/c^2. Following this motivation we here investigate some cosmological consequences of the existence of *any* stable, massive, neutral lepton, and show that it could well dominate the present mass density in the universe. The contribution to the mass density depends on the mass of the lepton, which should eventually be determined with high-energy accelerators. It is interesting that the more massive the lepton, the smaller its contribution to the present mass density. It is unlikely that these leptons affect big bang nucleosynthesis or condense into stellar size objects. However, such a lepton is an excellent candidate for the material in galactic halos and for the mass required to bind the great clusters of galaxies. Annihilation radiation from these structures should be detectable. At the end of the paper a brief mention is made of the astrophysical constraints on the mass-lifetime relationship if the neutral lepton is unstable.

Subject headings: cosmology — elementary particles

I. INTRODUCTION

The spectrum of leptons has always been, and continues to be, an intriguing puzzle in elementary particle physics. Until recently, the known species were the electron, the muon, their associated neutrinos ν_e and ν_μ, and their antiparticles. These neutrinos are assumed to be massless, since this naturally renders the chiral structure of the charged weak leptonic currents; experimentally there are reasonably stringent upper bounds on their masses: $m_{\nu_e} < 60$ eV and $m_{\nu_\mu} < 0.65$ MeV. Although in the past there was no convincing reason why the muon should exist in addition to the electron (and similarly for the muon neutrino), modern renormalizable unified gauge theories of the weak and electromagnetic interactions (Weinberg 1974; t'Hooft and Veltman 1973) have provided motivations for the existence of heavy leptons, both charged and neutral. These include quark-lepton symmetry, good high-energy behavior of scattering amplitudes, and the cancellation of certain anomalous divergences which would otherwise spoil the renormalizability of the theory. Typically, heavy leptons are unstable and decay into three lighter leptons via the exchange of a virtual gauge boson which gives rise to an effective current-current interaction. If the masses of the three decay leptons are negligible relative to that of the parent lepton, then the decay rate scales as the fifth power of the mass of this parent lepton. Thus, in general, as the mass of a lepton increases, its lifetime decreases quite rapidly.

In this paper we shall study the astrophysical consequences of the most extreme exception to this rule, namely the case of an absolutely stable heavy lepton. The initial motivation for this study was provided by a recently developed model of weak and electromagnetic interactions (Lee and Weinberg 1977*a*; Lee and Shrock 1977*a*, *b*)

* Alfred P. Sloan Foundation Fellow.

† The authors regret the tragic death of Dr. Lee in an auto accident 1977 June 16.

‡ Operated by the University Research Association, Inc., under contract with the Energy Research and Development Administration.

based upon the enlarged gauge group SU(3) × U(1). This model was initially constructed in order to account for the trimuon events observed by the Fermilab-Harvard-Pennsylvania-Rutgers-Wisconsin neutrino experiment (Benvenuti *et al.* 1977*a*, *b*). The most likely explanation of these unusual events seems to be the direct production and sequential decay of a charged heavy lepton M^-. In order to account for the observed rate it is most natural to invoke a full-strength weak coupling of ν_μ to M^-, via a new gauge boson, and thus necessarily to enlarge the gauge group beyond the standard Weinberg-Salam SU(2) × U(1) model (Weinberg 1967, 1971, 1972*a*, *b*; Salam 1967). The SU(3) × U(1) model successfully explains both the total rate and the kinematic characteristics of the trimuon events. Furthermore, it naturally ensures quark-lepton universality and $e = \nu$ universality, absence of right-handed currents in β and μ decay, and suppression of strangeness-changing neutral currents to order G_F^2 in the $K_L - K_S$ mass difference and $K_L \to \mu\bar{\mu}$ decay rate. Various discrete symmetries play a crucial role in maintaining these properties.

The heavy lepton production and cascade decay mechanism for trimuon production involves the decay of M^- into a neutral heavy lepton, M^0. An analysis of the trimuon data yields the approximate mass values $m_M = 7.0(+3.0, -1.0)$ GeV and $m_{M^0} = 3.5(+1.5, -0.4)$ GeV (private communication Shrock 1977). Thus, it is reasonable to propose that already two heavy leptons have manifested their presence in the trimuon data. It might be noted in passing, to bolster the reader's confidence in the seriousness and reality of leptons heavier than the muon, that experiments at the e^+-e^- colliding beam machine SPEAR have accumulated substantial evidence for the existence of a charged heavy lepton, called τ^-, with a mass of 1.9 GeV (Perl *et al.* 1975, 1976). This heavy lepton is also incorporated in the SU(3) × U(1) model. The M^-, M^0, and τ^- are all unstable.

For our purposes, one remarkable property of the model is of prime importance: as a result of an exact discrete symmetry in the theory (a kind of internal parity) the lightest of the massive leptons with odd internal parity is forbidden from decaying, i.e., it is absolutely stable. In the model this lepton is the electron-type partner to the M^0 and is accordingly labeled E^0.

Whatever the success or failure of this particular model, symmetries of this sort are likely to be needed in any attempt to expand the gauge group of weak and electromagnetic interactions since in order to preserve the universality of weak interaction strengths it is necessary to prevent the mixing of old and new leptons or quarks. The lightest particle which carries some new quantum number of this sort will of course be stable.

Although we were initially motivated by the SU(3) × U(1) model, *any stable neutral massive lepton would produce the astrophysical consequences we discuss in this paper.* Therefore we use the somewhat more general notation L^0 here to denote any absolutely stable neutral heavy lepton. Although the most recent theoretical ideas might identify L^0 with E^0, it may well be that future developments may give alternative motivation for a stable neutral heavy lepton. In any case the question should eventually be resolved with high-energy particle accelerators.

Lee and Weinberg (1977*b*, hereafter referred as LW2) showed that such neutral leptons must have a mass m_L greater than about ~ 2 GeV or less than ~ 50 eV (see also Cowsik and McClelland 1972);[1] otherwise they would be in conflict with the present observational constraints on the deceleration parameter q_0, and the lower limit to the age of the universe and the corresponding conservative upper limit on the density of the universe of $\sim 2 \times 10^{-29}$ g cm^{-3}. The argument is straightforward; first consider low-mass neutrinos ($M_L \ll 1$ MeV). At early epochs the universe was almost certainly hotter than $kT = 1$ MeV, so electron-positron pairs were as abundant as photons (see, for example, Weinberg 1972*c*). At such epochs ($kT > 1$ MeV), neutrino pairs would be copiously created and would be in thermodynamic equilibrium. At temperatures lower than $kT \approx 1$ MeV the weak interaction rate drops below the expansion rate and the neutrinos decouple from the ordinary matter. The present ratio of neutrinos of type i to photons is: $n_{\nu_i}/n_\gamma = \frac{3}{8} g_{\nu_i} (T_\nu/T_\gamma)^3$, where g_{ν_i} is the statistical weight of the type i neutrinos (e.g., for two helicity states and antineutrinos as well as neutrinos, $g_{\nu_i} = 4$). At present, $(T_\gamma/T_\nu)^3 \approx 11/4$ and $n_\gamma \approx 400$ cm^{-3} so that $n_{\nu_i} \approx 55\, g_{\nu_i}$. The present mass density contributed by such neutrinos would be $\rho_{\nu_i} \approx 9 \times 10^{-31}\, g_{\nu_i} m_i$ (eV) g cm^{-3}. For $\rho_0 \lesssim 2 \times 10^{-29}$ g cm^{-3} (equivalent to $\Omega = 1$, $H_0 = 100$ km s^{-1} Mpc^{-1}) we obtain: $\sum_i g_{\nu_i} m_i$ (eV) $\lesssim 220$ (e.g., for $g_{\nu_i} = 4$, $m_i \lesssim 55$ eV).

The situation is somewhat different for high-mass neutrinos ($m_L \gg 1$ MeV). Such neutrinos become nonrelativistic when they are still in thermodynamic equilibrium. As a result their abundance relative to photons is low due to the presence of a Boltzmann factor: $n_\nu/n_\gamma \sim \exp(-m_L c^2/kT) \ll 1$. These neutrinos "freeze out" when their annihilation rate drops below the expansion rate. LW2 shows that the freezing temperature is of the order of 1/20 the neutrino mass for masses of a few to a few tens of GeV. The residual mass density of L^0 is shown by LW2 to be

$$\rho_L = 4.2 \times 10^{-28}\ \text{g cm}^{-3} [m_L(\text{GeV})]^{-1.85} \left(\frac{N_A}{N_F^{1/2}} \right)^{-0.95}, \tag{1}$$

where N_A is the number of annihilation channels open and depends on the details of the $L^0\bar{L}^0$ interaction and N_F is the effective number of degrees of freedom allowed, counting 1/2 for each boson and 7/16 for each fermion species and spin state. Similarly the number density $n_L \propto [m_L(\text{GeV})]^{-2.85}$. Clearly there is some uncertainty in N_A depending on the model, but m_L only goes as approximately $N_A^{1/2}$. Taking $N_A \approx 14$ and $N_F = 4.5$ gives, for $\rho < 2 \times 10^{-29}$, $m_L \gtrsim 2$ GeV. Figure 1 summarizes the relationship between m_L and ρ_L.

[1] Throughout this paper the subscript L will refer to the neutral lepton L^0 and not to any of its charged relatives.

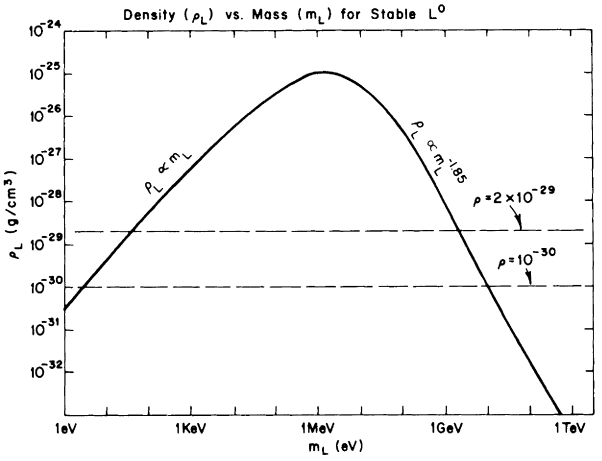


FIG. 1.—The cosmological relationship between density ρ_L and the mass of the stable L^0, m_L, making the assumptions given in the text.

The cosmological production of particles with masses this high and with concomitant freezing temperatures in the range around 1 GeV are plausible only if the multiplicity of hadron states is not exponential in the mass, as has been suggested (see, e.g., Hagedorn 1970; Huang and Weinberg 1970). If quarks are infinitely tightly bound, such a model may still be viable. In such a case the number of particles with masses larger than 2–3 GeV would be much smaller than the calculations of LW2 since the maximum temperature would then be the order of 170 MeV. Thus densities much smaller than 2×10^{-29} g cm^{-3} need not imply very high L^0 masses. Most of the subsequent discussion centers around L^0 masses of the order several GeV, but the reader must keep in mind that the uncertainties are large. It should be remembered that this mass question should eventually be resolvable with accelerator experiments. Notice that if massive neutrinos exist and if the universe reached equilibrium at temperatures $\gtrsim 10^{12}$ K, then such neutrinos *must* contribute to the mass density of the universe. Whether such a contribution is interesting depends on m_L. If m_L is very large (> 10 GeV), then the contribution to the density would be small. However, for m_L between ~ 2 and ~ 10 GeV such neutrinos would be the dominant constituent of the present universe. It should also be noted that if the neutrinos cluster with the baryons in galaxies and clusters of galaxies so that the cosmological density limits from galaxies apply a more restrictive limit on ρ is possible (Gott *et al.* 1974, hereafter GGST). In such a case m_L would be $\gtrsim 10$ GeV and the neutrinos could provide the "missing mass" for galaxies and clusters but would not close the universe. If for some reason the neutrinos do not all cluster with the baryons and m_L is found to be the order of 2–5 GeV, then the neutrinos could close the universe. It will be shown that this latter possibility is unlikely (but, given the uncertainties about the physics of the early universe perhaps not impossible).

The only interactions that are important at the present epoch are large-scale gravitational ones and annihilation. For the latter we use the rate $\langle \sigma v \rangle_a = G_F^2 m_L^2 N_A / 2\pi$, with the Fermi coupling constant $G_L = 1.15 \times 10^{-5}$ GeV^{-2}, and N_A set equal to 14. Numerically, $\langle \sigma v \rangle_a \approx 1.4 \times 10^{-26} (m_L/2 \text{ GeV})^2$ cm^3 s^{-1}. In the simple picture presented in LW2, 9/14 of the decays go into hadron channels. We will use a picture in which half the decay energy emerges in photons, which for definiteness we will assume come out in an energy band E_0 centered at some fraction f of the mass m_L ($E_0 = f m_L c^2$) and of about that width. If roughly half the energy comes out in photons, this parametrization gives the average number of photons per annihilation, g_γ, as $\sim 1/2f$. Note also that a stable heavy neutral lepton (L^0) will interact with other matter through simple weak exchange of a neutral vector boson. Thus the $L^0 + p \rightarrow L^0 + p$, low-energy cross section is the order of a typical weak cross section. Then, for a "typical" galaxy of radius ~ 20 kpc and average density of roughly 1 H atom per cm^3, the collision probability is negligibly small—galaxies are transparent to such neutral leptons.

The bulk of the paper is arranged more or less chronologically in the cosmological setting. We consider the thermal history of the early universe, big bang nucleosynthesis, and the development of density perturbations and their growth. After the primeval hydrogen plasma combines, the density perturbations in the L^0 fluid and in the baryon fluid, which need not be correlated, develop into galaxies and clusters, and the extent to which the different species can segregate is investigated and shown to be probably very small. The effects of annihilation in the universe as presently seen are investigated, and we show that critical tests of the theory can be formulated if the "missing mass" is indeed heavy neutrinos with roughly the adopted annihilation rate and roughly the adopted decay scheme. For completeness we also show that other mechanisms, such as spin-flip radiation and Compton scattering, are far too weak to yield detection possibilities. A brief discussion is also given on the unlikely possibility of forming stellar

size aggregates of heavy neutrinos. The paper concludes with a description of restrictions which might be placed on neutrino lifetimes if the symmetry were broken and the massive neutrinos were not absolutely stable.

II. THE EARLY UNIVERSE: THERMAL HISTORY

Since at temperatures much above the rest mass of L^0, the L^0's are simply one of a very large number of relativistic species present, the effects on the total mass density of the universe is very small. One can reasonably ask whether once the rest mass of the L^0 is important, the L^0 might dominate the mass density. It is clear from inspection of the expression for the mass density of such particles in equilibrium,

$$\rho = \text{const. } T^4 \int u^2 du \left[u^2 + \frac{m^2c^4}{K^2T^2}\right]^{1/2} \left[\exp\left(u^2 + \frac{m^2c^4}{K^2T^2}\right)^{1/2} + 1\right]^{-1} \tag{2}$$

that though ρ is in fact an increasing function of mc^2/kT, it never exceeds the value it has for $m = 0$ by very much and, of course, falls exponentially for large m. Since, again, there are many low-mass species present at $kT \sim m_Lc^2$, the L^0, $\bar{L}^0$ mass density is always a small fraction of the total until, long after freezing, the radiation density drops once and for all below the cold rest-mass density of the residual L^0's. Stronger arguments can be made for the pressure: the pressure of the L^0's is *never* important. Since the number density is roughly the zero-mass thermal density of states, which is proportional to T^3 at $kT \gg m_Lc^2$ and falls rapidly below that value for kT less than m_Lc^2 (see the curves plotted in LW2), the particles never constitute a degenerate fluid (that is to say, neither the L^0's nor the $\bar{L}^0$'s constitute a degenerate fluid).

In fact, since $\langle \sigma v \rangle_a$ is independent of velocity, the distribution function would not change shape as a result of annihilation if there were not ongoing thermalization. We shall see, however, that the heavy neutrinos are in effective thermal contact with the rest of the universe until the temperature is somewhat lower than the freezing temperature, at which the L^0's are very nonrelativistic. The result is that a roughly Maxwellian distribution of velocities is set up in what for all practical purposes is a completely classical fluid of noninteracting particles.

We can estimate how long the L^0's are in thermal contact with the rest of the universe as follows. The momentum exchange in a collision with a relativistic thermal particle with $kT \ll m_Lc^2$ is

$$\frac{\Delta p}{p} \approx kT(m_LkT)^{-1/2}. \tag{3}$$

The collision rate is

$$\lambda \approx n\sigma_t c\,,$$

where σ_t is the total collision cross section and n is the density of the light thermal particles (the only heavy particles besides the L^0's themselves at the relevant epoch are nucleons, whose number densities are down by about 10^8 compared to everything else [cf. Weinberg 1972*c*], as are the number densities of the L^0's). The thermalization time is the time required for the momentum to change of order itself by random encounters, or

$$\lambda \Delta t \left(\frac{\Delta p}{p}\right)^2 \sim 1\,, \tag{4}$$

which yields

$$t_{\text{th}} \approx \frac{m_Lc}{NaT^4}\,, \tag{5}$$

where N is the number of species of relevant thermal particles. As mentioned above, the interaction depends on the simple exchange of vector boson. Thus the cross section for weakly interacting fermions with $mc^2 < kT < m_Lc^2$ is

$$\sigma \sim G_F{}^2(kT)^2 \sim 3.8 \times 10^{-44}\left(\frac{T}{10^{10}\text{ K}}\right)^2 \text{cm}^2\,. \tag{6}$$

The cross sections for interaction with photons is much smaller (see § VII*c*) since photons do not couple directly to the neutral L^0 but must couple through a charged virtual intermediary, either a lepton or a charged boson. Thus interactions with photons are not important. The relevant interactions at $T \lesssim 10^{12}$ K are with electrons and the various massless neutrinos and their antiparticles; N is about 3. Thus

$$t_{\text{th}} \sim 4 \times 10^3 \left(\frac{10^{10}\text{ K}}{T}\right)^6 \left(\frac{m_L}{2\text{ GeV}}\right) \text{seconds}\,. \tag{7}$$

The expansion rate is $\sim 1/2t$ for a radiation-dominated universe, where t is the age, so

$$t_{\text{exp}} \sim 2t \sim 2.2(10^{10}\text{ K}/T) \text{ seconds} \tag{8}$$

for $5 \times 10^9 \lesssim T \lesssim 10^{12}$ K. The L^0's are in thermal kinetic equilibrium as long as $t_{\rm th} \lesssim t_{\rm exp}$, or for

$$T \gtrsim 10^{11}(m_L/2\ \text{GeV})^{1/4}\ \text{K}\,, \tag{9}$$

which is somewhat lower than the freezing temperature. At later times, the L^0's "cool" with the expansion (they do not interact), but the distribution function remains Maxwellian with a "temperature"

$$T_L \approx \left(\frac{T_\nu^2}{1.4 \times 10^{11}\ \text{K}}\right)\text{K}\,, \tag{10}$$

where T_ν is the neutrino temperature, currently $(4/11)^{1/3}$ times the blackbody photon temperature, or about 2 K. The L^0's cool faster than radiation because they are nonrelativistic. We shall see that this temperature is meaningful only up until about the epoch of decoupling ($T_{\rm ph} \sim 4000$ K, $1 + z \sim 1500$), at which time it has fallen to about 5×10^{-5} K. Thus the L^0 fluid in the early universe becomes very cold.

III. BIG BANG NUCLEOSYNTHESIS

It has been shown that big bang nucleosynthesis places severe constraints on the baryon density of the universe (cf. review by Schramm and Wagoner 1977). In particular the amount of deuterium produced in big bang nucleosynthesis is very sensitive to the baryon density, with present indications being that the baryon density is more than an order of magnitude below the critical density necessary to close the universe. In addition, the ^{4}He yield, while weakly dependent on the baryon density, is very sensitive to the competition between the cosmological expansion rate and the weak interaction rate for the neutron decay processes $n + e^+ \leftrightarrow p + \bar{\nu}_e$ and $p + e^- \leftrightarrow n + \nu_e$. This latter fact was utilized by Steigman, Schramm, and Gunn (1977) in showing that the helium abundance limits the number of types of massless particles. ("Massless" here means $\ll 100$ keV.) This limit occurred because during and preceding big bang nucleosynthesis the density ρ of the universe is radiation dominated with $\rho \propto T^4$, the proportionality constant being dependent on the number of relativistic particle types. Therefore each new massless or low-mass ($\lesssim 1$ MeV) particle type would add to the density ρ, which would increase the expansion rate and thus increase the ^{4}He yield.

For massive neutrinos, the above scenario no longer occurs. At the temperatures relevant for big bang nucleosynthesis, 10^{11} to 10^9 K, the massive neutrinos do not significantly contribute to the density of the universe. Thus they have no effect on the expansion rate and cause *no significant effect* in big bang nucleosynthesis. Thus big bang nucleosynthesis proceeds in the same manner whether or not the massive neutrinos exist; and, if deuterium is cosmological in origin, its abundance still specifies the *baryon* density.

What *is* important is that the relation between the *total* matter density and the nucleosynthesis can be altered, since the L^0's can contribute to the mass density at present. The relations between the current baryon density and the nucleosynthesis are unaltered. Thus the present abundances of helium and deuterium seem to require a low baryon density now, of order 5×10^{-31} g cm^{-3} *or less* (see, for example, GGST for review and discussion). This requires a density parameter $\Omega \sim 0.08$ ($\Omega = \rho/\rho_c$, where ρ_c is the critical density of the universe), if all the mass is in baryons; but if a substantial density is in L^0's, Ω can be much larger, *argued from the deuterium and helium alone*; indeed, this argument cannot preclude closure ($\Omega > 1$), with most of the matter in the universe in L^0's. We will later fall back on dynamical arguments to show that, even with L^0's, Ω is probably less than 1, but it is important to note that, if they exist, a prime argument for a low-density universe no longer holds.

With regard to nucleosynthesis, it should be noted that if the L^0's provide the bulk of the dynamical mass density ($\Omega \sim 0.08$), then the baryon density is even less than 5×10^{-31} g cm^{-3}. This would imply that much more deuterium is synthesized in the big bang than is currently observed and consequently more astration is required to reach its observed value.

IV. DENSITY PERTURBATIONS IN THE EARLY UNIVERSE

Once the annihilations become rare, the L^0's only interact via gravity. It is very difficult for the L^0's to transfer their internal energy to the ordinary matter, and they will tend to get left behind in any collapse. For an interaction with a constant gravitational potential, it is clear that an L^0 which is initially not bound to gravitational system will remain unbound. It will pass a mass point in some hyperbolic orbit, thus creating a density enhancement $\rho^*/\rho \propto (r/r^*)^{3/2}$ near the mass, but each neutrino would have no effective way of dissipating energy to go from an unbound to a bound orbit. However, in a time-varying potential such as that produced in collapse, energy is not conserved and it is possible to capture initially unbound L^0's and, for initially bound L^0's, to partially follow the collapse.

In addition to the effect of the time-varying potential, it may be possible for a density perturbation of L^0's to be sufficiently large so as to be bound. It is known that in order to form galaxies in the universe, bound perturbations seem to be required. It is known that even for matter which only interacts gravitationally, there can be some collapse (the order of a factor 2) due to the time-changing effect of the collective gravitational potential. This type of collapse has come to be known as a violent relaxation (Lynden-Bell 1967). Regardless of whether one has a self-made condensate or a primordial density fluctuation, it can be seen that without some way to dissipate energy, the

L^0's will always be less compact than the collapsing normal matter and will tend to form an extended halo. Although the L^0's will not gravitationally collapse as much as ordinary matter, they can condense sufficiently that annihilation might again become important.

Since the existence of galaxies seems to imply the existence of primordial density perturbations, one question we need to address at the outset is whether small-scale but large-amplitude density perturbations can significantly alter the conclusions of LW2 about the annihilation and survival of the L^0 fluid. The L^0's are thermally produced, and the universe is thermal-radiation dominated in the era of their production. Therefore, temperature perturbations are not physical, since they merely represent a spatial variation of the zero of time (Rees, private communication). What may be important are curvature fluctuations. Imagine a small portion of the universe which locally has positive curvature, so that if it were big enough in the course of time it would collapse. In a matter-dominated universe, if the region has not begun its recollapse before the horizon grows to include neighboring regions of negative curvature (and hence lower density), the radiation can flow out, saving the perturbations from collapse but leaving behind any noninteracting nonrelativistic matter (the L^0's) with a large local density excess. Such perturbations do not grow so long as the universe is radiation-dominated, but they can have interesting effects if their magnitude is large.

First, it is clear that their relative amplitude cannot be much greater than unity, since a positive curvature region at turnaround has a density only a few times the mean (critical) density. It is indeed doubtful that they could be near unity, since if there are fluctuations in the amplitude of the curvature fluctuations, many would be large enough to result in an early collapse, with the perturbation being sufficiently large that the annihilation radiation could not escape. Such perturbations would result in the production of large numbers of black holes of mass

$$M \sim \rho(ct)^3 \sim \left(\frac{30aT^4}{c^2}\right)\left[\left(\frac{3}{32\pi G}\right)^{1/2}\left(\frac{30aT^4}{c^2}\right)\right]^3 \sim 6 \times 10^{30}\left(\frac{kT}{1\ \mathrm{GeV}}\right)^{-2}\ \mathrm{g} \tag{11}$$

for $T \gtrsim 0.5\ \mathrm{GeV}/k \sim 5 \times 10^{12}$ K, where the factor 30 in the density is the effective number of species present (see LW2; it is assumed again that there is in fact no maximum temperature). Thus a large fraction of the mass-energy of the universe would be dumped into black holes of masses large enough that they do not evaporate by the Hawking (1973) mechanism. They will accrete, but that is really irrelevant. What is important is that very soon the universe becomes matter-dominated, and the present ratio of matter density to radiation density is

$$\left(\frac{\rho_m}{\rho_{\mathrm{rad}}}\right)_0 \gtrsim \frac{T_{\mathrm{BH}}}{T_0} \approx 3 \times 10^{12}\left(\frac{kT_{\mathrm{BH}}}{1\ \mathrm{GeV}}\right), \tag{12}$$

where T_{BH} is the temperature at black-hole formation. The ratio has an upper limit observationally of about 10^4.

Thus if the curvature fluctuations lead to collapse, they must do so in a vanishingly small fraction of the dense regions, and we can conclude that the rms amplitude is very small. This means that regions are not formed in which the annihilation proceeds much further than in other regions.

Let us now see what happens to perturbations in the L^0 density. In particular we need to examine the question of whether or not L^0 diffusion will damp out perturbations. The question comes down to how far the L^0's might travel in the time scale of perturbation growth. We will see that because the thermal velocities of the L^0's are low, they do not travel far and so damp out only small-scale fluctuations.

In the era of radiation dominance $R \propto t^{1/2}$, and one obtains

$$\Delta u = \frac{v_0 t_0}{R_0} \ln \frac{t}{t_0} \tag{13}$$

for the radial coordinate traversed at time t for an initial velocity v_0. As shown above, at thermal decoupling of the L^0's from the rest of the universe, the temperature was of order $10^{11}\ \mathrm{K}(m_L/2\ \mathrm{GeV})^{1/4}$. Thus

$$\left.\frac{v}{c}\right|_0 \approx \left(\frac{kT_0}{mc^2}\right)^{1/2} \sim 2.7 \times 10^{-2}\left(\frac{m_L}{2\ \mathrm{GeV}}\right)^{-1/4}, \tag{14}$$

$$t_0 \approx 1.1\left(\frac{T_0}{10^{10}\ \mathrm{K}}\right)^{-2} \sim 2.7 \times 10^{-3}\left(\frac{m_L}{2\ \mathrm{GeV}}\right)^{-1/2}\ \mathrm{seconds}. \tag{15}$$

At t_{eq}, the time at which the radiation density and mass densities are equal, we shall see that we get rapid growth of fluctuations. That epoch occurs at a redshift

$$(1 + z)_{\mathrm{eq}} \approx \left.\frac{\rho_m}{\rho_{\mathrm{rad}}}\right|_{\mathrm{present}} = 10^4\Omega\left(\frac{H_0}{60\ \mathrm{km\ s^{-1}\ Mpc^{-1}}}\right)^2, \tag{16}$$

or a temperature (for $T_0 = 2.9$ K, and a radiation field at present consisting of photons, ν_e, ν_μ, ν_τ, and their antiparticles)

$$T_{eq} = (1 + z)_{eq} T_0 3 \times 10^4 \Omega (H_0/60)^2 \text{ K} \tag{17}$$

at an age

$$t_{eq} \approx \left(\frac{3}{32\pi G \rho_{eq}}\right)^{1/2} \approx 2 \times 10^{11} \Omega^{-2} \left(\frac{60}{H_0}\right)^4 \text{ seconds} . \tag{18}$$

Thus free diffusion occurs for a time t_{eq}, and

$$r_0 \approx R_0 \Delta u \sim c \frac{v}{c}\bigg|_0 t_0 \ln \frac{t_{eq}}{t_0} \sim 7 \times 10^7 \text{ cm} \left(\frac{m_L}{2 \text{ GeV}}\right)^{-3/4} . \tag{19}$$

The corresponding mass (assuming the mass *now* is dominated by L^0's) is

$$m_{diff} \approx 4 \rho_L|_{t_0} r_0^3 \approx 2.4 \times 10^{28} \Omega (H_0/60)^2 \text{ g} , \tag{20}$$

a small fraction of a solar mass. Therefore, perturbations of this size or larger will survive without diffusing away.

Let us assume there are primordial perturbations in the L^0 density of the type that seem to be necessary in the baryon density in order to form galaxies. We will now examine how these L^0 perturbations evolve and develop. Let us look at perturbations in which the radiation density is uniform, but the L^0 density is not: we shall see that this is the relevant picture for a large range of perturbations sizes (all but those corresponding to the very largest observed structures). Let $\rho_L = \rho_L(1 + \delta)$, with $\delta \ll 1$. The peculiar velocity of the L^0 fluid responds only to the gravitational potential induced by fluctuations in the L^0 fluid itself, since we assume the radiation to be uniform. The perturbation equations can be developed in the standard way (see, e.g., Weinberg 1972*c*, p. 578, for a treatment of this problem, and Mészáros 1974 for perturbations of this type in radiation-dominated universes) and yield

$$\ddot{\delta} + 2H\dot{\delta} = 4\pi G \bar{\rho}_L \delta , \tag{21}$$

where $H = \dot{R}(t)/R(t)$. Let $R_{eq} = R(t_{eq})$, and let

$$\tau_{eq}^2 = \left(\frac{3}{8\pi G \rho_{rad}}\right)_{eq} = \left(\frac{3}{8\pi G \rho_m}\right)_{eq} . \tag{22}$$

The quantity τ_{eq} is of the same order as t_{eq}. The field equations yield the equation for R,

$$\dot{R}^2 = \frac{8\pi G R^2}{3} (\rho_m + \rho_r) = \frac{R_{eq}^2}{\tau_{eq}^2} \xi^{-1} + \xi^{-2} , \tag{23}$$

where $\xi = R/R_{eq}$. After some manipulation we can cast equation (21) in the form

$$\frac{d^2\delta}{dt^2} (\xi + \xi^2) + \frac{d}{dt} (1 + 3\xi/2) = 3\delta/2 . \tag{24}$$

One solution of this equation is $\delta = \frac{2}{3} + \xi$. The other solution (for a matter-dominated universe) is proportional to $\xi^{-3/2}$. This other solution is in fact monotonically decreasing in amplitude, as may be seen by the following simple argument. At large ξ, δ may be taken to be positive and has a negative first derivative and a positive second derivative. If for smaller ξ it is to have a positive first derivative, there must be a ξ such that $d\delta/d\xi = 0$. Let ξ_t be the largest such ξ. Clearly δ is positive there, since $d\delta/d\xi < 0$ for $\xi > \xi_t$. Clearly also $d^2\delta/d\xi^2$ must be *negative* there, since $d\delta/d\xi = 0$ for $\xi = \xi_t$. But the coefficients in equation (24) are all positive, so no such ξ_t can exist. For small ξ, the other solution goes like $\ln \xi^{-1}$, and must fit smoothly and monotonically on to const. $\times \xi^{-3/2}$. Thus the two modes are monotonically decaying and growing, respectively, but neither does very much for $\xi < 1$. Physically the reason is that the time scale for expansion is set by the radiation density; and since the matter responds only to matter fluctuations, there is simply not time for the perturbations to grow. All this development assumes the absence of random particle motions which affect only perturbations of order m_{diff} or smaller, so the approximation is good for all perturbations of interest. Thus, perturbations in the L^0 fluid cannot grow appreciably prior to t_{eq}.

The baryon perturbations are locked to the radiation by electron-scattering viscosity until the recombination epoch, or at $T \sim 4000$ K, which may be earlier or later than t_{eq} according as whether Ω is smaller or larger than $\Omega = 0.2(60/H_0)^2$, respectively. For $t > t_{eq}$, $\delta \propto \xi \propto t^{2/3}$ for the growing mode.

The baryon perturbations can in general be decomposed into an *adiabatic* component ($\delta\rho_B = \text{const.}\ \delta\rho_{rad}$) and an *isothermal* one ($\delta\rho_{rad} = 0$). Since the sound speed in the medium before t_{eq} is $c/\sqrt{3}$, perturbations in the radiation density begin behaving like acoustic waves almost as soon as the horizon is bigger than the length scale of the perturbation, and the adiabatic perturbations are damped by photon viscosity, a well-known result (Silk 1967). The

horizon at decoupling contains about $1.5 \times 10^{18}\Omega(H_0/60)^2\, M_\odot$, so adiabatic perturbations on the scale of galaxies and even great clusters of galaxies had been propagating like waves for a long time prior to decoupling. Thus on average the cold L^0 fluid saw a uniform background, which motivated the development of a model for perturbations in the L^0 fluid alone. (Of course, the *isothermal* perturbations entail no radiation perturbations.) All but the largest of the adiabatic baryon perturbations are damped by photon viscosity (Silk 1967), but the L^0 perturbations do not couple to the photons and thus survive on all scales.

The effective damping of the adiabatic baryon perturbations is changed by the L^0's only if Ω is large, so that there is a period in which the universe is dominated by the L^0 fluid before recombination. The time scale is speeded thereby, so that smaller perturbations can survive for a given Ω_B, the fraction of the critical density presently in baryons. For smaller Ω_B the "classical" results hold (cf. Weinberg 1972*c*). In any case, it seems that only isothermal perturbations on the scale of galaxies and clusters survive. Remember, however, that L^0's only damp out their own perturbations on scales smaller than $M_{\rm diff} \sim 10^{-5}\, M_\odot$.

It is important to note that the perturbations are static in the *comoving* frame prior to $t_{\rm eq}$; their metric sizes grow with the expansion, and their density drops like $1/T^3$. Since one cannot by curvature perturbations produce density fluctuations in the L^0 fluid of large amplitude, one does not have to worry about perturbations in which the density of L^0's and $\bar{L}^0$'s is high enough to produce appreciable annihilation. It is instructive to note that the annihilation time is given by

$$\tau_a = \frac{1}{n_L \langle \sigma v \rangle_a} \approx \frac{7.3 \times 10^{24}\ \text{s}}{n_L} \left(\frac{2\ \text{GeV}}{m_L}\right)^2 \tag{25}$$

and is about the present Hubble time for a density of about $n_{\rm crit} = 10^8(2/m_L)^2\ \text{cm}^{-3}$. At $t_{\rm eq}$, the number density of L^0's is

$$n_{\rm eq} = 6 \times 10^5 \left[\Omega\left(\frac{H_0}{60}\right)^2\right]^4 \left(\frac{2}{m_L}\right)\ \text{cm}^{-3}, \tag{26}$$

much smaller than the value $n_{\rm crit}$ above for $\Omega \sim 0.1$, $H_0 \sim 60$, but note the very strong dependence on H_0.

V. THE DEVELOPMENT OF STRUCTURE AND THE VALUE OF Ω

In this section we will investigate the growth of perturbations to form the structure we see in the universe today. We will see that regardless of the initial correlation between L^0 and baryon perturbations, in the end they become correlated, and thus the L^0's will be associated with galaxies. To examine the development of structure, let us go to the era, $t_{\rm eq}$, when the radiation and matter densities are equal.

In the period between $t_{\rm eq}$ and the recombination time $t_{\rm rec}$ [recall that this period exists only for $\Omega \gtrsim 0.2(60/H)^2$], the perturbations in the L^0 density grow freely, and their amplitude increases by a factor $(1 + z)_{\rm eq}/(1 + z)_{\rm rec} \approx 5\Omega(H/60)^2$ (recall that the pressure effects are completely negligible, the velocity dispersion in the L^0's corresponding to a temperature of order 10^{-4} K at these epochs). Thus for large Ω, if there is any scale on which $\delta\rho_L/\rho_L$ is of order unity, bound structures will form with densities of order $\rho_{L,\rm eq}$. If these survive, they will be copious sources of annihilation radiation, but we will argue that they do not. There is some question, in fact, about whether they would exist—we have argued that it is difficult to make perturbations with large $\delta\rho_L/\rho_L$, and an attractive scheme has been forwarded in which all the density perturbations are in the *baryons* (Gott and Rees 1975) and arise associated with the same sort of curvature fluctuations we considered for the origin of the L^0 perturbations, with the amplitudes $\sim 10^{-4}$ when each new scale enters the horizon. In this picture the L^0 density perturbations on interesting scales would be very small, and nothing would happen until recombination.

The situation is sufficiently uncertain that one would like to find a general argument that indicates how the L^0 and baryon perturbations develop, given the distinct possibility that they are completely uncorrelated at recombination, and with (possibly also uncorrelated) velocity perturbations which are of the same order as the density perturbations.

We do know that the perturbation amplitudes are small for scales like galactic masses and larger; the density amplitude is related to the collapse time for a bound perturbation by (Gunn and Gott 1972)

$$t_c = \frac{\pi}{H_{\rm init}} (\delta_+)^{-3/2}, \qquad H_{\rm init} \sim H_0 \Omega^{1/2} (1 + z_{\rm init})^{3/2}, \tag{27}$$

where $H_{\rm init}$ is the value of R/R at $t_{\rm init}$, the larger of $t_{\rm eq}$ and $t_{\rm rec}$, and δ_+ is the relative density excess then above the critical density $\rho_c = 3H_{\rm init}^2/8\pi G$. The average density at these epochs is nearly the critical density;

$$\delta_e = \frac{\rho - \rho_c}{\rho_c} = \frac{\Omega - 1}{\Omega(1 + z_{\rm init})}. \tag{28}$$

For a cluster like Coma [$t_c \sim 6 \times 10^9(60/H_0)$ yr], $\delta_+ \sim 5 \times 10^{-3}\Omega^{-1/2}$ while $\delta_e \sim 10^{-3}(\Omega - 1)/\Omega$. Thus even for $\Omega = 0.1$, the cluster excess above critical is twice as large as the excess of the critical density above the mean

density. For a galaxy with a collapse time for 10^9 years, the perturbation is $\delta^+ \sim 1.8 \times 10^{-2}\Omega^{-1/2}$, about 6% for a low-density ($\Omega = 0.1$) model.

Thus initially the baryons and L^0's can each be considered to fair accuracy a uniform fluid, and the question is whether by virtue of their initial conditions they can separate.

For any reasonable value of $\Omega(>0.02)$ now, $\Omega_{\rm init} = 1 + \delta_e$ is near unity, so $R \propto t^{2/3}$ during the initial development of the perturbation. Let us suppose that the density is dominated by L^0's; this assumption is really irrelevant but simplifies the discussion. Suppose we have a more or less spherical perturbation, which for the sake of definiteness has $\delta_+ > 0$, so that it is bound, and suppose the baryons are moving radially in the perturbation with a velocity different from the L^0's by an amount Δv_0 initially at an initial comoving radial coordinate u. We assume for definiteness that the perturbation has constant density inside some boundary r_m, so that it expands uniformly; clearly $r_m \propto t^{2/3}$ initially also, if the initial amplitude of the perturbation is small. Then the perturbation behaves like a small piece of a somewhat denser universe, and we can apply the same argument as we made to determine the diffusion size of the L^0 perturbations in the very early universe. Scale the coordinate u such that $u = r/r_m$ initially; thus $u = 1$ corresponds to the edge of the perturbation. Again,

$$r_m \dot{u} = \Delta v_0 (r_m)_0 / r_m \,. \tag{29}$$

This integrates to yield

$$u = u(t) - H_0 = 3 \frac{v_0 t_0}{(r_m)_0} \left[1 - \left(\frac{t_0}{t} \right)^{1/2} \right] \tag{30}$$

But $v_0 = (\dot{r}_m)_0 u_0 = [2(r_m)_0/3t_0]u_0$, so

$$\frac{\Delta u}{u} = 2 \frac{\Delta v_0}{v_0} \left[1 - \left(\frac{t_0}{t} \right)^{1.2} \right], \tag{31}$$

and a velocity perturbation of a few percent results in a *total* radial separation of only a few percent, and converges to a final value very rapidly. The above situation can be generalized without difficulty to the exact motion for a bound perturbation, with similar results. A crucial (well known) point is that gravitational perturbations never get much smaller than their maximum size. Coming to equilibrium merely involves generating random kinetic energy comparable to the gravitational energy, so the final gravitational energy is twice the total energy, whereas at maximum expansion it *is* the total energy. Thus, crudely, $r_{\rm final} = \frac{1}{2} r_{\rm max}$. Hence we conclude that the baryons and L^0's cannot separate by gravitational processes alone, and any structure which collapsed as a result of gravitational instability must contain matter which has a ratio of baryons to L^0's which is representative of the cosmic average ratio, provided it developed from small perturbations. The subsequent development of the perturbation is largely through the violent relaxation process described by Lynden-Bell (1967) which does not segregate by mass.

They *can* separate subsequently by virtue of the fact that if gas pressure is ever important and the densities are such that the gas (*baryon* gas, or course) can cool by radiative processes, the baryons can sink deeper into the potential well. Thus the L^0's can be left behind in more extended structures than *galaxies*, for which these conditions are satisfied, but cannot in *clusters* of galaxies. This implies that the L^0 to baryon ratio in clusters and perhaps in extended elliptical galaxies will have the cosmic ratio, but the disks of spirals will not and stars will have essentially no L^0's.

Galaxies in groups and clusters will experience dynamical friction against the L^0 background in these structures, but the effects are the same as they are for stars or gas, since the friction does not depend on the mass of the particles in the background fluid if these masses are much smaller than the galaxy mass. Some aspects of violent relaxation and dynamical friction have also been considered by Faulkner, Sarazin, and Steigman (1978) with regard to L^0's.

Gunn (1977) has noted that the mass-to-light ratio of structures large enough not to have suffered the dissipative separation discussed above have a mass-to-light ratio of about $120(H_0/60)$ in solar units (Zwicky photographic magnitude system) normalized to the stellar population of spiral galaxies. The luminosity density with the same normalization is $6.3 \times 10^7 (H_0/60) L_\odot \,{\rm Mpc}^{-3}$ (Gott and Turner 1976), leading to a mass density of $7.5 \times 10^9 (H_0/60)^2 M_\odot \,{\rm Mpc}^{-3}$, $\rho = 5.1 \times 10^{-31} (H_0/60)^2$ g cm^{-3}, or an H_0-independent value of Ω of 0.08. The mass-to-light ratios of stellar material seem to only be in the range of 5–10, and this mass can be readily accounted for. Therefore, between 90 and 95% of the matter, even for this small value of Ω, is "missing." This mass, if it is assumed to be nondissipative and otherwise inert, was shown to form the roughly $1/r^2$ halos required for galaxies as the galaxy perturbation developed; these halos would be tidally stripped in the great clusters, and most of the halo mass would move with random velocities of the same order as the galaxy velocities in the cluster potential. Heavy noninteracting neutral particles present since the early universe, constrained by the arguments above to cluster with the matter, could not be better as stuff to constitute the dynamical missing mass.

Admittedly there are some uncertainties in the exact determination of limits on Ω from the dynamical mass associated with galaxies. Davis and Peebles (1977), using the BBGKY formalism, obtain a value for Ω closer to unity, although Aarseth, Gott, and Turner (1977) argue from n-body calculations that the observed clustering is also compatible with $\Omega \sim 0.1$. It is probably reasonable to say that Ω from the dynamical mass associated with galaxies is uncertain by about a factor of ~ 3, but seems to be below its critical value of unity. Therefore, if L^0's do

provide the "missing" mass and cluster with galaxies, then the *L^0's would not close the universe.* The only way around this would be if, for some as yet unknown reason, the bulk of the L^0's were able to avoid the perturbations which produced the galaxies and clusters and remain unbound. It should be noted that if the L^0's do provide the dynamical missing mass and $\Omega \sim 0.1$, then $m_L \approx 10$ GeV rather than the 2 GeV lower limit mentioned before. Assuming our model for the annihilation cross section and that no Hagedorn (1970) limiting temperature occurs, then the critical test for L^0's being the missing mass is for m_L to be determined to be ~ 10 GeV. In fact, 10 GeV *may be a more realistic lower limit to m_L than* 2 GeV; however, there are sufficiently large uncertainties that for safety we will continue to write our expressions using the conservative 2 GeV normalization.

One other comment should be made before we leave the subject of perturbations. Baryon perturbations smaller than $10^6\ M_\odot$ or so, the Jeans mass at decoupling, cannot collapse because of ordinary gas pressure; the L^0 fluid suffers no such restriction, because of its much lower temperature. If there are large-amplitude L^0 perturbations on this scale (or smaller), they will collapse. It can be argued (Fall and Rees 1977)—and n-body calculations and observations of galaxies and clusters support the arguments—that very little such small-scale structure persists as larger aggregates form. Galaxies themselves are the interesting exception to this rule of large-scale structure destroying smaller-scale structure, and it seems almost certain that dissipative processes play a vital role in their formation. Thus we expect very few of these high-density, low-mass structures to survive; but if any do, they will be very interesting objects indeed, as we will discuss later.

VI. THE RESIDUAL ANNIHILATION RADIATION AND γ-RAY OBSERVATIONS

The observed γ-ray background has an isotropic component whose spectral flux may be approximated by a power law with (Fichtel, Simpson, and Thompson 1977)

$$\frac{d\mathscr{F}_\gamma}{dE} \approx 8.3 \times 10^{-2} E_{\rm mev}^{-2.85} \text{ photons (cm}^2 \text{ s sr MeV)}^{-1}. \tag{32a}$$

The corresponding integral flux is

$$\mathscr{F}_\gamma(E) \approx 0.045 E_{\rm mev}^{-1.85} \text{ photons (cm}^2 \text{ s sr)}^{-1}. \tag{32b}$$

We will investigate the extent to which residual L^0–$\bar{L}^0$ annihilation might account for that background; we will also consider whether discrete sources of annihilation radiation (e.g., galactic halos) can be seen against the background.

The following calculations are intended to provide estimates; in keeping with the spirit of these estimates, several simplifying assumptions will be made. For each L^0–$\bar{L}^0$ annihilation, we assume that g_γ γ-rays are produced at energy E_0; in analogy with p-$\bar{p}$ annihilation we assume that $2g_\gamma E_0 \approx m_L$. It is expected that the average photon would have energies like those of the bulk of a hadron shower ($E_0 \sim 200$ MeV), which implies for $M_L \sim Z$ GeV, $g_\gamma \approx (2f)^{-1} \approx 5$. In principle this point could eventually be checked experimentally. The annihilation radiation emissivity then is

$$\delta(E) = g_\gamma (\sigma v)_a n_L{}^2 \delta(E - E_0) [\text{photons (cm}^3 \text{ s MeV)}^{-1}]. \tag{33}$$

Consider the cosmological contribution to the background. The number density at any time in the past (corresponding to a redshift z) is

$$n_L = n_{\bar{L}} \approx 2 \times 10^{-5} M_L{}^{-3} (1 + z)^3 \text{ cm}^{-3}. \tag{34}$$

In equation (34) and subsequently, M_L is in GeV. For $(\sigma v)_a \approx 2.5 \times 10^{-28} N_A M_L{}^2$ (cm^3 s^{-1}), we obtain

$$j(E, z) = j_0(E)(1 + z)^6, \tag{35a}$$

where

$$j_0(E) \approx 10^{-37} g_\gamma N_A M_L{}^{-4} (E - E_0) [\text{photons (cm}^3 \text{ MeV)}^{-1}]. \tag{35b}$$

The present γ-ray flux at energy E is due to photons emitted in the past (at redshifts z) at energies $E(1 + z)$. To convert the appropriate time integral to an integral over redshift, we use (for a matter-dominated cosmology),

$$dt/dz = H_0{}^{-1} (1 + z)^{-2} (1 + \Omega_0 z)^{-1/2}. \tag{36}$$

It is convenient to replace the redshift integral by an integral over energies; this is accomplished by noting that $E' = E(1 + z)$. In this manner we obtain

$$\frac{d\mathscr{F}_\gamma}{dE} = \left(\frac{1}{4\pi}\right)\left(\frac{C}{H_0}\right)\left(\frac{1}{E^3}\right) \int_E^{E_m} j_0(E') E'^2 \left[(1 - \Omega_0) + \frac{\Omega_0 E'}{E}\right]^{-1/2} dE'. \tag{37}$$

In (37), $E_M = E(1 + z_M)$; z_M is the redshift at which the universe is opaque to γ-rays. Substituting (35) in (37), we easily obtain for $E > E_0 > E_M$,

$$\frac{d\mathscr{F}_\gamma}{dE} \approx 1.2 \times 10^{-10}\left(\frac{60}{H_0}\right) g_\gamma \frac{N_A}{M_L^4}\frac{E_0^2}{E^3}\left[(1 - \Omega_0) + \Omega_0 \frac{E_0}{E}\right]^{-1/2} \text{photons (cm}^2\text{ s sr MeV)}^{-1}. \tag{38}$$

For any production spectrum which is reasonably peaked around some energy E_0, equation (38) should provide a satisfactory order-of-magnitude estimate. For $E \gtrsim \Omega_0 E_0$, the integral flux we predict is

$$\mathscr{F}_\gamma(E) \approx 6 \times 10^{-11}\left(\frac{60}{H_0}\right) g_\gamma N_A M_L^{-4}(E_0/E)^2 \text{ photons (cm}^2\text{ s sr)}^{-1}. \tag{39}$$

Note that the spectral index of the predicted integral flux (~ 2) is in excellent agreement with that of the observed flux (1.85[+0.50, −0.35], Fichtel, Simpson, and Thompson 1977). (It is interesting to note that Stecker 1977 also showed that annihilation radiation could fit the γ-ray background; however, Stecker assumed baryon annihilation which has other problems [cf. Steigman 1976].) What of the magnitude of the predicted flux? For $H_0 = 60$, $N_A = 14$, E in MeV, and recalling that $2g_\gamma E_0 \approx M_L$ (with M_L in GeV), we obtain

$$\mathscr{F}_\gamma(E) \approx 2.2 \times 10^{-4}(g_\gamma M_L^2)^{-1} E_{\text{MeV}}^{-2} \text{ photons (cm}^2\text{ s sr)}^{-1}. \tag{40}$$

For $E_{\text{MeV}} \approx 100$, the observed flux is $F_\gamma^{\text{obs}} \approx 9 \times 10^{-6}$, whereas for $M_L \gtrsim 2$ the predicted flux is $F_\gamma^{\text{pred}} \approx 5.4 \times 10^{-9}\text{ g}^{-1}$. Thus, γ-ray production in cosmological annihilation fails to account for the observed background by roughly three orders of magnitude.

We have seen that the average density of uniformly distributed L^0's is too low to contribute significantly to the observed, isotropic component of the γ-ray background. If L^0's form halos around galaxies, the density would be greatly enhanced. Let us now consider whether such discrete sources could be seen against the isotropic background. Consider the Galaxy itself. If the Galaxy is typical of other big spirals, it has a circular velocity at large radii of about 200 km s^{-1}. For a heavy halo of L^0's the corresponding density distribution is (Gunn 1977)

$$n_L = n_{\bar{L}} \approx v_c^2[8\pi G M_L r^2]^{-1}. \tag{41}$$

Since the central regions of galaxies are *not* dominated by high mass-to-light ratio stuff, this distribution does not extend to the center; the distribution must turn over at some core radius, a, which must be of the order of 10 kpc. We thus adopt for the L^0 distribution in galactic halos,

$$n_L = n_{\bar{L}} \approx n_0 a^2(a^2 + r^2)^{-1}, \tag{42a}$$

$$n^0 = \frac{v_c^2}{8\pi G M_L a^2} \approx 0.15 v_{200}^2 M_L^{-1} a_{10}^{-2}\text{ cm}^{-3}. \tag{42b}$$

The γ-ray emissivity is given in (33) where, now, we use (42) for n_L. The predicted flux will depend on the direction of observation and on the ratio of the core radius, a, to the distance, r_0, of the Sun from the galactic center. As a first estimate of the magnitude of the flux we compute F_0 defined by

$$\mathscr{F}_0(E) \equiv \frac{1}{4\pi}\int_E^\infty dE' \int_0^R dr j(E', r)\,, \tag{43a}$$

$$\mathscr{F}_0(E) \approx \frac{g_\gamma (\sigma v)_a n_0^2 \mathscr{A}}{4\pi}\int_0^{R/a} (1 + x^2)^{-2} dx\,. \tag{43b}$$

In (43), R is the maximum extent of the halo; for $R/a \gg 1$, the integral in (43b) is $\approx \pi/4$ so that

$$\mathscr{F}_0 \approx \tfrac{1}{16} \times g_\gamma (\sigma v)_a n_0^2 \mathscr{A} \approx 1.1 \times 10^{-8} N_A v_{200}^4 a_{10}^{-3} g_\gamma \text{ photons cm}^{-2}\text{ s}^{-1}\text{ sr}^{-1}. \tag{44}$$

The flux expected in various directions may be related to $\mathscr{F}_0$; for $a_{10} \approx r_0 \approx 1$ (10 kpc), $v_{200} \approx 1$(200 km s^{-1}), and $N_A \approx 14$ we obtain

$$\mathscr{F}_\gamma(\text{center}) = \left(\frac{3}{2} + \frac{1}{\pi}\right)\mathscr{F}_0 \approx 2.8 \times 10^{-7} g_\gamma\,, \tag{45a}$$

$$\mathscr{F}_\gamma(\text{pole}) = \frac{1}{2\sqrt{2}}\mathscr{F}_0 \approx 5.4 \times 10^{-8} g_\gamma\,, \tag{45b}$$

$$\mathscr{F}_\gamma(\text{anticenter}) = \left(\frac{1}{2} - \frac{1}{\pi}\right)\mathscr{F}_0 \approx 2.8 \times 10^{-8} g_\gamma\,. \tag{45c}$$

Unless g_γ is very large ($\gtrsim 100$, in which case E_0 will be very small), the predicted flux is not in conflict with observations (for $E \approx 100$ MeV; see eq. [32b]). However, the unique distribution on the sky of halo annihilation γ-rays, as well as their spectral distribution (flat compared to the observed isotropic background) suggest the possibility that such a component may be discovered in future γ-ray observations.

The halos of other galaxies might be discrete sources of annihilation γ-rays; the sum of such discrete sources would contribute to the overall isotropic background. To estimate these effects, first compute the luminosity of a galactic halo in annihilation γ-rays:

$$L_\gamma = g_\gamma \int_{\text{halo}} d^3r\,[(\sigma v)_a n_L^2(r)]\,, \tag{46a}$$

$$L_\gamma = g_\gamma[\pi^2(\sigma v)_a n_0^2 a^3] \approx 2.3 \times 10^{40} g_\gamma \text{ photons s}^{-1}\,. \tag{46b}$$

In (46), the usual assumptions have been adopted: $a \approx 10$ kpc, $v \approx 200$ km s^{-1}, $N_A \approx 14$ (note that these results, as well as those in [44], are independent of M_L).

In our picture, all galaxies will be γ-ray sources; for example, M31 should be comparable to the Galaxy in luminosity and should be a source with a diameter somewhat larger than 1°.5. The total flux from M31 would be

$$\mathscr{F}_\gamma(\text{M31}) \approx 3.8 \times 10^{-10} g_\gamma \text{ photons (cm}^2\text{ s)}^{-1}\,. \tag{47}$$

This is well below the upper limit (Fichtel, Simpson, and Thompson 1977) of 10^{-6}. It is, thus, not surprising that the total contribution to the isotropic background from all galaxies is small:

$$\mathscr{F}_\gamma(\text{Total}) \approx \frac{L_\gamma}{4\pi} ng\left(\frac{c}{H_0}\right) \approx 4.4 \times 10^{-9} g_\gamma \left(\frac{H_0}{60}\right)^2 \text{ photons cm}^{-2}\text{ s}^{-1}\text{ sr}^{-1}\,. \tag{48}$$

Before passing on to other topics, it is worth investigating the expected fluxes from L^0's associated with other mass aggregations. Because of their low densities, the great clusters are inefficient radiators. Coma, for example, with $V_c \approx 1200$ km s^{-1}, $a \approx 100$ kpc, will have a total luminosity only ~ 100 times that of an individual galaxy. Even though the cluster galaxies almost certainly contain only a small fraction of the cluster mass, the galaxy contribution to the γ-rays probably exceeds that from the cluster as a whole.

It is interesting to note that if $L^0\bar{L}^0$ aggregates of globular cluster size were to survive subsequent dynamical evolution, then their γ-ray luminosity would be $\sim 10^{37}$ photons s^{-1}. It is not clear what their optical counterparts might be.

Although our estimates lead to somewhat disappointing conclusions, the situation is not entirely discouraging. Our crude calculations suggest that with a better model for γ-ray production in $L^0\bar{L}^0$ annihilation and more sensitive data (with respect to distribution on the sky as well as spectral distribution) a definitive test of the hypothesis of an L^0 halo around our Galaxy might be possible. In this regard it is worth noting that Stecker (1978) has carried out some model calculations which are essentially consistent with our general estimates.

VII. OTHER ASTROPHYSICAL ASPECTS OF L^0's

In this section several interesting but probably undetectable aspects of the L^0's will be mentioned.

a) Magnetic Moment

Although at the present time we see no direct way to take advantage of the fact, it is nevertheless interesting to note that the L^0's have a magnetic moment. Because the L^0 is neutral, there is obviously no lowest order moment $\sim e/2m_L$ (where $c = \hbar = 1$) but from several single loop Feynman diagrams one can calculate following Fujikawa, Lee, and Sanda (1972) and Lee and Shrock (1977*b*) a moment

$$\mu_{L^0} = -\frac{3}{2}\frac{G_F m_L^2}{\pi^2\sqrt{2}}\left(\frac{e}{2m_L}\right) \equiv -a\left(\frac{e}{2m_L}\right)\,, \tag{49}$$

where

$$a = 5.0 \times 10^{-6}\left(\frac{m_L}{2\text{ GeV}}\right)^2\,.$$

This moment is accurate to lowest order in $(m_L/m_w)^2$, where m_w is the mass of the intermediate vector boson [$m_w \sim 86$ GeV in the minimal SU(3) × U(1) with favored parameter values].

b) Spin-Flip Radiation

Placed in a magnetic field of strength which extends over a volume V, magnetic moments can spin-flip, producing radiation at about angular frequency $\omega = 2u_L B$. Let the spin-flip transition time be τ. Then the amount of radiation is

$$\tau \sim 2|\mu_L| B n_L V/\tau \text{ ergs s}^{-1} \tag{50}$$

at angular frequency ω. The transition time τ can be shown to be given by ($\alpha = e^2/4\pi \sim 1/137$):

$$\tau \sim \pi(2m_L)^2\alpha^{-1}(\mu_L B)^{-3} \sim 2.6 \times 10^{19}\left(\frac{2}{m_L}\right)\left(\frac{10^{13}\text{ gauss}}{B}\right) \text{ s}.$$

Thus

$$I \sim 3 \times 10^{-26}\left(\frac{m_L}{2}\right)\left(\frac{B}{10^{13}\text{ gauss}}\right)^4 n_L V \text{ ergs s}^{-1}.$$

Clearly, even for galactic densities $n_L \sim 1$ and for magnetic fields near neutron stars or magnetic white dwarfs, the intensity is still negligible.

c) Compton Scattering

Because the L^0 can look like a virtual electron ($\dot{e}$) or heavy "electron" (E^-) and an intermediate vector boson (W or U), it is possible for it to Compton scatter with the universal blackbody photons. In a typical gauge model such as the SU(3) × U(1) model of Lee and Weinberg the cross section for such a process can be estimated roughly to be

$$\sigma \sim \frac{\alpha^2}{\pi^3} G_F^2 m_E^{-2} \frac{m_L^4}{m_u^4},$$

$$\sim 10^{-47}\left(\frac{m_E}{10\text{ GeV}}\right)^2\left(\frac{m_L}{2\text{ GeV}}\right)^2\left(\frac{80\text{ GeV}}{m_U}\right)^4 \text{ cm}^2,$$

where E^- is a charged unstable heavy lepton and U is a gauge boson, degenerate with the W which appear in this SU(3) × U(1) model. The mass of the U is determined in this model to be the same as $m_W \sim 86$ GeV, and $m_{E^-} \sim 10$ GeV. There would also be terms with $m_{E^-}{}^2$ replaced by $\sim m_e{}^2$ or $\sim m_U{}^2$, but those terms are obviously small. The above cross section can be compared with that for normal Compton scattering off electrons $\sigma_{\text{comp}} \sim \pi\alpha^2/m_e 2 \sim \sim 2.5 \times 10^{-25}$ cm^2. Notice that the present L^0–photon collisional lifetime is about

$$\tau \sim (\sigma c n_{\text{BB}})^{-1} \text{ s};$$

with $n_{\text{BB}} \approx 4 \times 10^2$ photons cm^{-3} we have $\tau \approx 8 \times 10^{33}$ s, which is clearly negligible compared to the age of the universe ($\lesssim 10^{18}$ s). Even for the early universe it is clear that photon scattering was not occurring on a rapid time scale unless $T \gg 10^{10}$ K, and for those temperatures we have already seen that direct weak interactions dominate.

d) L^0 Stars

We have already seen that L^0's will probably not form condensed bound systems of stellar size since we know of no mechanism to dissipate the energy. We have also seen that, even if one could form such a bound system, the $L^0\bar{L}^0$ annihilation would prevent a significant density buildup and would prevent the L^0's from becoming degenerate. However, for the sake of speculation let us presume tht some condensation mechanism took place and some $L^0\bar{L}^0$ separation mechanism took place to prevent annihilation. (It is possible, for instance, that the net L^0 number is nonzero so there is an excess of L^0's over $\bar{L}^0$ or vice versa.) In such a situation the L^0's would then be the idealized fermion of the Oppenheimer-Volkoff calculation. This would mean that degeneracy would at most support a 0.7 $M_\odot$ L^0 star with GeV mass L^0's. It is also intriguing that because of the magnetic moment such an L^0 star might have an intrinsic magnetic field.

It is also interesting that if a regular (nondegenerate) star contains a nonsymmetric L^0 and/or $\bar{L}^0$ component, then a sufficient admixture of such L^0's can change both radiative transport and luminosity estimates of stars, and also influence the nucleosynthetic stellar evolution. The points here are that part of the radiation pressure necessary to hold a star up against its gravity could then arise in part from the neutrino annihilation photons, thereby relieving the burden on the core of the star to produce, by convential nucleosynthesis, the major part of the photon budget. This, in turn, means that the core can be cooler than under conventional scenarios of element production, which in turn means that reaction rates proceed more slowly. A star may then not undergo nuclear evolution as rapidly as normally presumed. Precise estimates clearly depend on the heavy neutrino admixture assumed *a priori*. This point is also discussed by Faulkner, Sarazin, and Steigman (1977). It should be remembered that at present we have no reason to believe in such asymmetries, nor do we see a plausible way to get bound L^0's into stellar size objects.

VIII. ASTROPHYSICAL LIMITS ON POSSIBLE HEAVY NEUTRINO DECAY TIME SCALES

While the SU(3) × U(1) formalism requires that the heavy neutrino (L^0) be stable, it is possible that this symmetry group is only approximate (it could, for instance, be an embedded part of a much larger group with a small, but nonzero, coupling to the whole group). It is then possible that the heavy neutrino is not absolutely stable. Supposing it can decay, what astrophysical consequences result? These depend on its decay products. As has been pointed out by Zel'dovich (1977) and Dicus, Kolb, and Teplitz (1977), it may be possible to use the astrophysical constraints to bracket the possible lifetime and decay products; in fact, much more stringent constraints than those discussed there can be placed. There are strong arguments against present radiation dominance in the universe (GGST, Mészáros 1974), but the γ-ray background limits are, as we shall see, much more restrictive for the current density of decay radiation.

The constraints we can impose fall into several different regimes depending on the decay time, let us consider them in turn.

a) Long Decay Times ($t \gtrsim 10^5$ years)

The redshift of the dominant decay period is less than a few thousand, and the universe is transparent to the decay photons. The intensity of the decay photons is roughly

$$6 \times 10^{-9} \frac{N}{4\pi} \frac{2}{m_L} [1 - \exp(-t/t_L)] = 16N[1 - \exp(-t/E_d)] \frac{2}{m_L} \text{ photons cm}^{-2}\,\text{s}^{-1}\,\text{sr}^{-1}. \quad (51)$$

If t/t_L is long, i.e., if the decay redshift is large, this intensity must be compared with the observed number above $E = E_0/(1 + z_d)$, where again E_0 is the average photon energy and z_d is the redshift corresponding to $t \sim t_L$. One obtains from the γ-ray background

$$16N\left(\frac{2}{m_L}\right) \lesssim 2 \times 10^{-5}\left[\frac{E_0}{100\text{ MeV}(1 + z)}\right]^{-1.2}. \quad (52)$$

If $NE_0 \sim \phi m_L$, we can write

$$\dot{\phi} \lesssim 5 \times 10^{-9}\left(\frac{E_0}{100\text{ MeV}}\right)^{-0.2}(1 + z)^{1.2}, \quad (53)$$

which requires that the fraction of energy ϕ going into γ's is less than about 10^{-4}, which seems unlikely. For very low masses ($m \ll 1$ MeV) the decay will not be limited by the γ-ray background but can instead be restricted, for all but extremely small masses, by other wavelength backgrounds or by arguments that there cannot be a radiation-dominated universe during the epoch of galaxy formation. Thus we conclude that the range of decay times between about 10^5 years and the present age of the universe is ruled out for essentially any but extremely small masses for the L^0. For even larger decay times we can neglect the redshift of the decay photons, and the condition becomes

$$\frac{16Nt_0}{\tau_d(2/m_L)} \leq 2 \times 10^5\left(\frac{E_0}{100\text{ MeV}}\right)^{-1.2}. \quad (54)$$

The particles are for all intents and purposes stable as far as dynamics is concerned, and the arguments for the mass of the L^0 remain unchanged—they must be more massive than a few GeV. We can write

$$\phi \leq 5 \times 10^{-9} \frac{\tau_d}{t_0}\left(\frac{E_0}{100\text{ MeV}}\right)^{-0.2}. \quad (55)$$

If $\phi \sim 1$, τ must be longer than $10^8 t_0 \sim 10^{18}$ years! Thus if the decay time is longer than about 10^5 years, it must be very long indeed.

b) Intermediate Decay Times (2000 s $\lesssim t_L \lesssim 10^5$ years)

In this range we are in the curious situation that the universe is very opaque to the decay photons, and they come to thermal kinetic equilibrium with the charged particles present very quickly, but the photon production time is too long for the production of a blackbody photon distribution. In this temperature range ($10^4 \leq T \leq 10^8$) the dominant photon production is through bremsstrahlung, for which

$$\frac{d}{dt}\left(\frac{n_{\text{ph}}}{T^3}\right) \approx 1.6 \times 10^{-27} k^{-1} T^{-3.5} n_{e0}^2, \quad (56)$$

where n_{e0} is the present electron density. Since $t \sim 2 \times 10^{19}(1 + z)^{-2}$ s for the same temperature range, we have

$$\frac{d}{dt}\left(\frac{n_{\rm ph}}{T^3}\right) \approx 4 \times 10^{-5}\Omega^2(1 + z)^{-1/2} \tag{57}$$

or

$$(n_{\rm ph}/T^3)|_{\rm brems} \approx 8 \times 10^{-5}\Omega^2(1 + z)^{1/2}\,. \tag{58}$$

Now the value of this ratio for blackbody radiation is about 50 ($\sim a/k$), which says that bremsstrahlung cannot seriously change the blackbody photon number for $z \lesssim 4 \times 10^{11}\Omega^{-4}$. This is in fact not quite true, since for $1 + z \geq 10^8$, pair production begins to increase the electron density markedly, and other more efficient processes for making photons come into play. We know, however, that the present radiation spectrum is accurately blackbody with a temperature near 2.9 K. The distortion of the microwave background due to decay is going to be affected by the temperature at decay. The longer the lifetime, the lower the temperature at decay and thus the easier it is to create a distortion. The total mass-energy density, ρ_L, in the form of such particles prior to decay will follow the mass dependence in Figure 1. Therefore higher masses, which can contribute less energy density, will be restricted more for longer decay times. This is quantitatively shown in Figure 2 where it was assumed that the Zel'dovich distortion parameter for the microwave background was less than 0.02 (Jones and Steigman 1977).

For very low masses where $am_Lc^2 < kT_*$ (T_* is the temperature at the time of decay), it should be remembered that the number of density of L's prior to decay will be comparable to the number density of photons with the normal statistical factors ($n_L/n_\gamma = 6/11$). Thus, unless the branching ratio to radiative decay is extremely small (< 0.03), the effect of any decay will be to change the number density of photons substantially and thus cause an unacceptable distortion. In particular there would be too many photons for a given kT.

c) *Short Decay Times* ($t_L \lesssim 2000$ seconds)

For decay times shorter than about 200 seconds, whatever the mass range, there are no observable consequences that the L^0 existed except a probable distortion of the blackbody neutrino background; it can be confidently stated that that will remain a secret for quite some time.

For decay times between about 200 and 2000 seconds we need to look at the possible effect on big bang nucleosynthesis. It is useful to think of nucleosynthesis occurring in two distinct stages. Although this is clearly an oversimplification, short of doing a numerical computation, it is quite accurate.

Stage I: $T \gtrsim 10^{10}$ K.—During this stage, the weak interaction is in equilibrium ($T_\nu \approx T_\gamma$). The crucial n/p ratio is primarily determined during this stage. Of importance here is the competition between the expansion rate

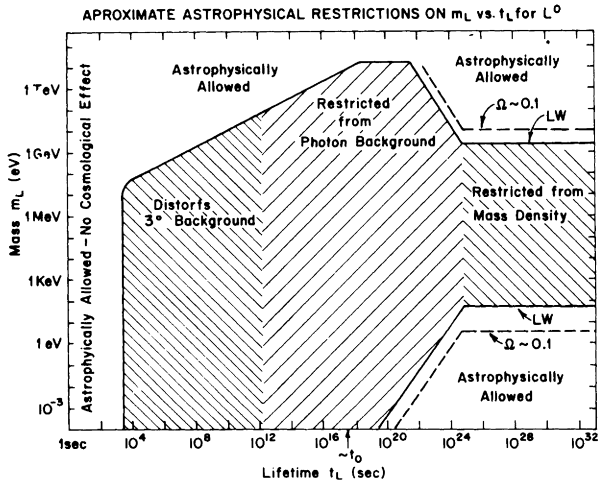


FIG. 2.—The restrictions on the relationship between the mass m_L and lifetime t_L for L^0's assuming the decay mode produces at least one photon. For long lifetimes the restrictions come from Fig. 1. For intermediate lifetimes (note $t_0 \approx$ age of the universe) the restrictions come from the various limits on the photon background (in particular the γ-ray background) if t_L is less than the radiation decoupling time ($\sim 10^5$ yr); but if t_L is greater than the time of big bang nucleosynthesis, then the restrictions come from the distortions in the present microwave background. No strong astrophysical constraints come for decays at the epoch of nucleosynthesis or earlier.

($\propto \rho^{-1/2}$) and the weak interaction rate (cf. the review by Schramm and Wagoner 1977). The entropy per baryon is not important during this stage.

Stage II: $T_\gamma \lesssim 2 \times 10^9$ K.—During this stage, $T_\gamma \approx (11/4)^{1/3} T_\nu$ and the light nuclei are built up via two-body reactions. Of importance here is the reaction rate (which depends on nucleon density) versus the expansion rate. A change in the expansion rate can be compensated, during this stage, by a change in the nucleon density (or, entropy per baryon).

We will now see that the heavy neutrinos can play no major role during Stage I. In particularly, for $T \gtrsim 10^{10}$, we will see that $\rho_L \ll \rho_{\rm rad}$; thus the expansion rate is virtually uncharged during this stage.

From Figure 1 it follows that the present density of L's, ρ_L, is $\lesssim 10^{-26} T_\nu^3$; and we know that the present photon density $\rho_\gamma \sim 10^{-35} T_\gamma^4$. Thus for the time interval in question the density in L's is small ($\lesssim 0.036$) compared to the density of radiation and therefore the change in expansion rate will be small. In the limit of zero mass this will go to the results of Steigman, Schramm, and Gunn (1977), where it was shown that an additional zero-mass neutrino pair will increase the helium mass fraction by less than 0.02. When it is remembered that similar variations are possible by the uncertainty on the entropy per baryon alone, then it is clear that strict restrictions on L^0 lifetime cannot be placed in Stage I.

In Stage II ($T < 2 \times 10^9$) the expansion rate is competing with the nuclear reaction rates. However, any speedup in the expansion rate during Stage II has the same effect as lowering the baryon density (or raising the entropy per baryon). Since the baryon density is not well known, the deuterium and helium abundances cannot be used to isolate any such speedup.

The conclusion, therefore, is that if the heavy neutrino decays during nucleosynthesis, there is no unambiguous effect on primordial nucleosynthesis. If one does truly know the baryon density, then one can use the deuterium abundance to specify the entropy per baryon (cf. Schramm and Wagoner 1977) and thus restrict decays which affect this parameter. Recently Dicus *et al.* (1977) have used the above argument using what they believe to be the baryon density from galactic dynamics. However, we feel such restrictions may be premature. In any case, we are able to obtain comparable restrictions without precise assumptions on the baryon density.

Figure 2 summarizes what we believe to be the safest astrophysical limits on m_L and t_L.

IX. DISCUSSION AND CONCLUSION

We have investigated the consequences of the suggestion by LW2 that massive (few GeV) stable neutrinos L^0 and their antiparticles $\bar{L}^0$ may have been produced in the early universe and dominate the mass density in the universe at present. The L^0's remain in thermal kinetic equilibrium down to temperatures of about 10^{11} K, after which they decouple from the ambient radiation and cool as a nonrelativistic gas. Since annihilation has reduced their number density far below the thermal one, they constitute classical, albeit completely noninteracting Maxwellian gas.

Since they do not interact either with each other or with the baryonic matter present, and since at that time they make a negligible contribution to the mass density, they do not sensibly affect primordial nucleosynthesis, with the result that the relation between the helium and deuterium abundances and the baryon density at present is unchanged.

We showed that the only plausible mechanism for producing perturbations in the $\rho_{L^0}/\rho_{\rm rad}$ is through the presence of curvature perturbations, which must have small amplitude in order that the universe now not be dominated by black holes by a factor large compared to the limits on matter-to-radiation ratios. Hence annihilation rates calculated for a smooth distribution are almost certainly accurate.

Perturbations present in either ρ_L or the baryon density grow either after the universe becomes matter-dominated or after recombination, and we showed that even though the perturbations need not be correlated initially, both forms of matter respond to the total perturbation potential, and are effectively locked together gravitationally for all but very large-amplitude (and hence small-scale) perturbations. They can separate subsequently only if the baryonic matter can dissipate energy by radiation or other means. This can happen on galactic scales or smaller, but cannot happen on scales like the great clusters of galaxies or in any structure which is still expanding, like the Local Supercluster. Thus mass measurements in structures like these should measure the total mass of a mixture of baryons and L^0's in which their ratio is the primordial one—they are "fair samples" of the matter in the universe. Assuming such a case, standard dynamical arguments can be made which indicate that the density parameter Ω is less than 1, and the universe is not closed by the L^0's. The L^0's still serve well to account for the missing 90% of the mass required to make the extended halos of spiral galaxies and to bind the great clusters.

The production rate of γ-rays via annihilation in structures like galactic halos now is interestingly high, even though the annihilation times are very long. Our picture of the annihilation processes is very crude, however; and until a better model is forthcoming, we regard the predictions as tantalizing: the prospect of detecting the "missing" matter through its fundamental properties is a very exciting one.

We also pointed out some academically intriguing though probably unobservable aspects of L^0's such as their magnetic moment and spin flip radiation.

Finally, we investigated some of the consequences of the decay of the L^0 if it should not be absolutely stable. The basic conclusion is that, if the L^0 ever dominated the mass density in the universe, its decay time must be shorter than about 3000 seconds or longer than about 10^{18} years in order that it not produce a γ-ray background higher than

observed, unacceptable distortions of the blackbody background, or do violence to helium production. We noted that for decay times less than about 100 seconds, the L^0 never can dominate the mass density, so it can be responsible for matter dominance and concomitant growth of perturbations in the early universe for only a very small interval, clearly of no importance dynamically.

If massive neutrinos are shown to exist in accelerator experiments, they will be present in the standard big bang model with no special ad hoc assumptions. Although other neutrino species can also be used to add to the density, special ad hoc assumptions (e.g., degeneracy) are needed to increase the density of such unobserved species much beyond that which occurs in the standard model, and this directly affects big bang nucleosynthesis. For the massive neutral stable leptons, equilibrium at $T \gtrsim 10^{12}$ K is all that is needed.

In other words: the question of the nature of the missing mass might well be decided by Earth-based particle accelerators rather than optical telescopes. Measurements of a stable heavy neutrino of about a few GeV mass will then have the astrophysical consequences pointed out above.

We wish to acknowledge most gratefully the aid of Steve Weinberg, who with B. L. initiated this paper and offered advice as a coauthor throughout most of the development, but commitments to axions and related things unfortunately took their toll on his time.

Special thanks also go to R. Shrock of Fermilab and Princeton who made significant contributions to the weak interaction physics in this paper, particularly with regard to the L^0 magnetic moment and Compton cross section. The γ-ray section of this paper profited greatly from discussions with Floyd Stecker who was extremely helpful in accurately interpreting the γ-ray observations for us. Floyd has gone on to do his own detailed modeling of γ's from $L^0\bar{L}^0$ annihilation which he is publishing as a separate paper elsewhere in this issue. One of us (D. N. S.) acknowledges a conversation with Ya. Zel'dovich who had independently thought about several of the points mentioned. Also I. L. is grateful to Jim Ipser for beneficial discussion. D. N. S. and G. S. also acknowledge conversations with R. Wagoner, V. Teplitz, and E. Kolb. J. E. G. and D. N. S. thank the Institute of Astronomy in Cambridge, UK, for hospitality while some of this work was carried out. The work reported here was supported by the National Science Foundation (J. E. G., D. N. S.), by the Energy Research and Development Agency (B. L.) and by the National Aeronautics and Space Administration (I. L., D. N. S.).

REFERENCES

Abers, A., and Lee, B. W. 1973, *Phys. Rept.*, **9**, 1.
Aarseth, S., Gott, J. R., and Turner, E. 1977, private communication.
Benvenuti, A., *et al.*, 1977*a*, *Phys. Rev. Letters*, **38**, 1110.
Benvenuti, A., *et al.*, 1977*b*, *Phys. Rev. Letters*, **38**, 1183.
Cowsik, R., and McClelland, J. 1972, *Phys. Rev. Letters*, **29**, 669.
Davis, M., and Peebles, J. 1977, *Ap. J. Suppl.*, **34**, 425.
Dicus, D. A., Kolb, E. N., and Teplitz, V. L. 1977, *Phys. Rev. Letters*, **39**, 168.
Dicus, D. A., Kolb, E. N., Teplitz, V. L., and Wagoner, R. V. 1977, preprint.
Fall, M., and Rees, M. 1977, preprint.
Faulkner, J., Sarazin, C., and Steigman, G. 1978, in preparation.
Fichtel, C. E., Simpson, G., and Thompson, R. 1977, NASA-Goddard preprint.
Fujikawa, K., Lee, B. W., and Sanda, A. I. 1972, *Phys. Rev.*, **D6**, 2923.
Gott, J. R., Gunn, J. E., Schramm, D. N., and Tinsley, B. M. 1974, *Ap. J.*, **194**, 543 (GGST).
Gott, J. R., and Rees, M. 1975, *Astr. Ap.*, **45**, 365.
Gott, J. R., and Turner, E. 1976, *Ap. J.*, **209**, 1.
Gunn, J. 1977, Caltech preprint.
Gunn, J., and Gott, G. R. 1972, *Ap. J.*, **176**, 1.
Hagedorn, R. 1970, *Astr. Ap.*, **5**, 184.
Hawking, S. 1973, *Nature*, **248**, 30.
Huang, K., and Weinberg, S. 1970, *Phys. Rev. Letters*, **25**, 895.
Jones, F., and Steigman, G. 1977, preprint.
Landau, L. D., and Lifshitz, E. M. 1958, *Statistical Physics* (New York: Pergamon Press).
Lee, B. W., and Shrock, R. E. 1977*a*, FERMILAB-PUB-77/21-THY (1977 February) to be published in *Phys. Rev. D.*
Lee, B. W., and Shrock, R. E. 1977*b*, FERMILAB-PUB-77/21-THY (1977 June).
Lee, B. W., and Weinberg, S. 1977*a*, *Phys. Rev. Letters*, **38**, 1237.
———. 1977*b*, *Phys. Rev. Letters*, **39**, 165 (LW2).
Lynden-Bell, D. 1967, *M.N.R.A.S.*, **136**, 101.
Mészáros, P. 1974, *Astr. Ap.*, **37**, 225.
Perl, M., *et al.* 1975, *Phys. Rev. Letters*, **35**, 1489.
Perl, M., *et al.* 1976, *Phys. Rev. Letters*, **63B**, 466.
Salam, A. 1967, in *Elementary Particle Physics*, ed. N. Svortholm (Stockholm: Almqvist and Wiksells: 1968), p. 367.
Sato, K. 1977, preprint.
Schramm, D. N., and Wagoner, R. V. 1977, *Ann. Rev. Nucl. Sci.*, **27**, 37.
Silk, J. 1967, *Nature*, **215**, 1155.
Stecker, F. W. 1977, *Ap. J.*, **212**, 60.
———. 1978, *Ap. J.*, **223**, 1032.
Steigman, G. 1976, *Ann. Rev. Astr. Ap.*, **14**, 339.
Steigman, G., Schramm, D. N., and Gunn, J. E. 1977, *Phys. Letters*, **66B**, 202.
t'Hooft, G., and Veltman, M. 1973, Diagrammar, CERN Rept. 73-9.
Weinberg, S. 1967, *Phys. Rev. Letters*, **19**, 1264.
———. 1971, *Phys. Rev. Letters*, **27**, 1688.
———. 1972*a*, *Phys. Rev. Letters*, **D5**, 1412.
———. 1972*b*, *Phys. Rev. Letters*, **D5**, 1962.
———. 1972*c*, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (New York: Wiley).
———. 1974, *Rev. Mod. Phys.*, **46**, 255.
Zel'dovich, Ya. B. 1977, *Proc. Neutrino-77*, Caucausus, USSR.

JAMES E. GUNN: Department of Astronomy, California Institute of Technology, Pasadena, CA 91125

BEN W. LEE: deceased

IAN LERCHE: Division of Radiophysics, CSIRO, Epping, New South Wales, Australia

DAVID N. SCHRAMM: EFI–LASR, University of Chicago, 933 East 56th Street, Chicago, IL 60637

GARY STEIGMAN: Department of Astronomy, Yale University, New Haven, CT 06520

DYNAMICAL INTERACTIONS AND ASTROPHYSICAL EFFECTS OF STABLE HEAVY NEUTRINOS[a)]

GARY STEIGMAN[b)]
Yale University Observatory, New Haven, Connecticut

CRAIG L. SARAZIN[b)]
Department of Astronomy, University of Virginia, Charlottesville, Virginia 22903

H. QUINTANA
National Radio Astronomy Observatory,[c)] Green Bank, West Virginia 22901

JOHN FAULKNER[b)]
Lick Observatory, Board of Studies in Astronomy and Astrophysics, University of California, Santa Cruz, California 93604
Received 9 February 1978; revised 27 April 1978

ABSTRACT

Recent gauge theories for the weak interaction have predicted the existence of massive, stable, neutral leptons ("heavy neutrinos"). For masses in the range of 2–20 GeV (or higher masses if the particles are not matter–antimatter symmetric), these particles would represent a dynamically significant component of the mass density of the Universe. Could such particles participate in the formation of astrophysical structure (stars, galaxies, clusters of galaxies) through gravitational collapse? We show that they would undergo dissipationless collapse and violent relaxation along with ordinary matter, but that they are left behind in dissipative collapse (when the collapse time exceeds the free-fall time). We show that these heavy neutrinos would have all the required properties to form the "missing mass" in clusters of galaxies, and galactic halos. They would not collapse with stars and affect stellar structure, although stars are very sensitive to the presence of heavy neutrinos because the neutrinos transport energy very efficiently.

I. INTRODUCTION

The existence of a massive, stable, neutral lepton ("heavy neutrino") has been predicted by a gauge theory of the weak and electromagnetic interactions recently proposed by Lee and Weinberg (1977a). While this stable electron type neutrino (the E°) has not yet been observed experimentally, its muon type partners (the M^- and M°) may have been detected in trimuon events (Benvenuti *et al.* 1977). In this paper we will discuss the astrophysical consequences of the existence of any massive, stable, neutral lepton; our conclusions apply to any such particle. Although it is appealing at present to identify this particle with the E° predicted by Lee and Weinberg, the results we give apply to any stable heavy neutrino. Because they are relatively inert, massive neutrinos will have important astrophysical consequences only if they represent a significant fraction of the mass density of the present universe. Cowsik and McClelland (1972, hereafter CM) and Lee and Weinberg (1977b, hereafter LW) have shown that the present density of heavy neutrinos can be determined from their rest mass m. In the early universe when $kT \gg mc^2$, neutrino pairs were copiously produced; in equilibrium the production and annihilation rates balance and an equilibrium density of pairs is present (see, for example, Steigman 1973). For $kT \ll mc^2$, the production of pairs effectively ceases, but annihilation may continue until it too is slowed by further expansion.

There are two cases to consider. If $m \ll 1$ MeV, production and annihilation cease when the heavy neutrinos are relativistic; in this case CM showed that their present number density is proportional to the black-body photon number density so that

$$\rho_\nu \approx 4 \times 10^{-31}\, m_{\rm eV}\ {\rm g\ cm^{-3}} \quad (m \ll 1\ {\rm MeV}). \quad (1)$$

Here, $m_{\rm eV}$ is the mass of the heavy neutrino in eV. In contrast, if $m \gg 1$ MeV, the heavy neutrinos are nonrelativistic when they freeze out. The present density of such neutrinos has been evaluated numerically by LW; an adequate analytic fit to their results is:

$$\rho_\nu \approx \frac{7 \times 10^{-29}}{m_{\rm GeV}^3}\ {\rm g\ cm^{-3}} \quad (m \gg 1\ {\rm MeV}). \quad (2)$$

Here, $m_{\rm GeV}$ is the mass in GeV. Interestingly, the mass of the heavy neutrino is a double-valued function of its present mass density.

These relations can be used either to limit the possible range of the mass, or to determine the present density of the particles given their mass. First, a limit on the mass of the heavy neutrino can be derived by requiring that the density of such particles not exceed the observed upper limit on the total mass density in the universe

[a)] Lick Observatory Bulletin No. 803.

[b)] Visiting Scientist at the National Radio Astronomy Observatory.[c)] Mailing address: Bartol Research Foundation, Franklin Institute, University of Delaware, Newark, Delaware 19711.

[c)] Operated by Associated Universities, Inc., under contract with the National Science Foundation.

 0004-6256/78/8309-1050$00.90 © 1978 Am. Astron. Soc.

(based on limits on the value of q_0 and the age of the universe). For the conservative upper limit on the present mass density of $\rho_{\rm tot} \lesssim 2 \times 10^{-29}$ gm cm^{-3}, the mass is limited to

$$m_{\rm eV} \lesssim 50 \quad (\text{if } m \ll 1 \text{ MeV}), \tag{3a}$$

$$m_{\rm GeV} \gtrsim 2 \quad (\text{if } m \gg 1 \text{ MeV}). \tag{3b}$$

Alternatively, Eqs. (1) and (2) indicate that if m is in the range $1 \lesssim m_{\rm eV} \lesssim 50$ or $2 \lesssim m_{\rm GeV} \lesssim 15$, the heavy neutrinos would make an important, possibly dominant, contribution to the total mass in the universe. In fact, the heavy neutrinos may be significant even if $m_{\rm eV} < 1$ or $m_{\rm GeV} > 15$, because the calculation of CM and LW assumed that the universe is symmetric between neutrinos and antineutrinos. In principle, there is no reason why the heavy leptons need be symmetric; it is almost certain, for example, that the baryons are not symmetric (Steigman 1976). If there is an excess of heavy neutrinos over heavy antineutrinos, then Eqs. (1) and (2) give lower limits to the mass density of heavy neutrinos. Thus, these particles could dominate the mass of the universe for any value of m satisfying Eq. (3). Clearly, such neutrinos will play an important role in the dynamical evolution of the universe. Should such a massive, stable, neutral lepton exist, what would be the astrophysical consequences? The question, obviously, is not new but the theory of LW invests it with a new relevance (see also CM, LW, Gunn *et al.* 1978).

Heavy neutrinos interact only via the weak interaction and gravity, and it might reasonably be expected that the weakness of those interactions would render them effectively invisible. If they were uniformly distributed in space, their only gravitational effect would be to make the universe more closed (larger q_0). Moreover, for homogeneously distributed heavy neutrinos, weak interaction processes (such as annihilation) would be unimportant; we have just argued that such processes effectively ceased in the early universe ($T \gtrsim 1$ MeV, $t \lesssim 1$ s). Uniformly distributed heavy neutrinos would be uninteresting, except for their effect on the expansion of the universe.

However, if condensations of heavy neutrinos could form, these regions could affect the dynamics of structures of baryonic matter which form through gravitational collapse, such as clusters of galaxies, galaxies, and stars (see Sec. V). If the density of heavy neutrinos in these condensations exceeded the average density, weak interaction effects could become important. Gunn *et al.* (1978) have argued that observable fluxes of gamma rays might be produced by annihilations of neutrino-antineutrino pairs. For stars, the most condensed astrophysical structures, the effect of even a small contribution of heavy neutrinos would be enormous. These particles would provide alternative sources of gravitation, energy transport, and luminosity, and could drastically modify stellar structure (see Sec. V). As these examples illustrate, most potentially observable effects produced by heavy neutrinos depend crucially on the answer to the following question: To what extent can heavy neutrinos cluster gravitationally?

The heavy neutrinos are, effectively, a collisionless gas. Since exchange of energy between ordinary matter and heavy neutrinos is very difficult, the neutrinos are effectively isolated from energy changing processes such as shock heating, dissipative cooling, etc. Weak though it may be, the gravitational interaction prevents the complete isolation of the heavy neutrinos. The crucial point is simply that in a time dependent gravitational potential, energy is not conserved along particle orbits. The possibility thus exists that, to some extent, the heavy neutrinos may cluster gravitationally. This paper is mainly devoted to an investigation of the extent to which neutrinos may cluster and the astrophysical consequences of that clustering.

II. VIOLENT RELAXATION

Although quite homogeneous on the largest scales, the Universe displays a wealth of structure from stars to gas clouds, galaxies, clusters of galaxies and beyond. It is generally accepted that such large scale structure has as its origin the amplification, via gravitational instability, of small amplitude perturbations. Now, in a classic paper, Lynden-Bell (1967) pointed out that the violently changing gravitational field of a collapsing system will be effective in changing the statistics of particle orbits. Such violent relaxation produces, within a few crossing times (Peebles 1970; Gott 1973), an isothermal distribution in which the temperature is proportional to the mass (i.e., the velocity dispersions are independent of mass). Very quickly, then, phase mixing and violent relaxation lead to a system in virial equilibrium (i.e., $2 \times$ kinetic energy $\approx -$potential energy). Hence, in the absence of dissipation, a perturbation (e.g., galaxy, cluster of galaxies, etc.) will collapse from its maximum scale by a factor ~ 2, increasing its mean density by a factor ~ 8. In such collisionless collapse and relaxation the ordinary matter and the heavy neutrinos would share the same spatial distribution and would have equal velocity dispersions because both react to fluctuations in the overall potential. In any structures formed by dissipationless collapse the ratio of heavy neutrinos to ordinary matter would reflect the universal value of this ratio. Thus, heavy neutrinos might supply the dominant mass on large scales in galaxies and galaxy clusters. Their importance on other scales depends on the extent to which they can transfer energy to ordinary matter which is capable of radiating it away.

III. FURTHER COLLAPSE OF HEAVY NEUTRINOS

In this section we study the collapse of a gas of heavy neutrinos. For our purposes here, the key properties of the heavy neutrinos are that they are nonrelativistic, nondegenerate, inert (only weak and gravitational in-

teractions), and stable. Since the neutrinos cannot radiate away their energy, they can collapse gravitationally only if their energy remains constant or if they transfer their energy to the ordinary matter which can radiate it away.

The first case (constant energy) was discussed in the previous section. The neutrinos can collapse by turning potential energy into kinetic energy. Such a collapse will proceed by about a factor of 2 at which point violent relaxation will virialize the neutrinos (Lynden-Bell 1967). Thus, the neutrinos and the ordinary matter can collapse together for about a factor of 2.

Any further collapse of the neutrinos requires that they transfer their binding energy to the ordinary matter. Since the neutrinos main interaction with the ordinary low density matter is gravitational, they can lose energy only in a changing gravitational potential. Variations of the gravitational potential on many scales are possible. We first consider a region in which the overall potential is changing due to collapse. In Sec. IV, we discuss the situation when potential variations are due to the motions of subaggregates within the region (two-body interactions). These represent extreme limits to the scales of variation.

Consider a collapsing region (of ordinary matter plus neutrinos) of size R and potential ϕ; significant collapse of the neutrinos will be possible only if

$$v_\nu^2 \lesssim -\frac{\partial\phi}{\partial t}\frac{R}{v_\nu}. \tag{4}$$

In Eq. (4), v_ν is the neutrino velocity. If t_c is the collapse time (defined here by $t_c = \phi(\partial\phi/\partial t)^{-1}$ and $-\phi \approx GM/R \approx v_f^2$ (v_f is the free-fall velocity), then,

$$\frac{v_\nu}{v_f} \lesssim \left(\frac{t_f}{t_c}\right)^{1/3}, \tag{5}$$

where $t_f \approx R/v_f$ is the free-fall timescale. Now, since $t_c \gtrsim t_f$, the neutrinos can collapse along with the ordinary matter only if $v_\nu \lesssim v_f$. For an already virialized distribution, $v_f \approx v_\nu$ and thus, as these crude estimates suggest, neutrino collapse is only possible for rapid collapse of the ordinary matter: $t_c \approx t_f$. The collapse then must be dissipationless, or the cooling time must be much less than the free-fall time; the cooling time must here be increased over the usual estimates (e.g., Rees and Ostriker 1977) by the ratio of total mass to baryonic mass.

As the above very crude arguments suggest, the extent to which the neutrinos can collapse along with the ordinary matter depends on the velocity dispersion of the neutrinos and the speed of the collapse. To study this question in greater detail we consider some specific, simplified cases of collapse in which the neutrinos are treated as test particles; the effects due to the self-gravity of the neutrinos is discussed at the end. We consider the idealized case of uniformly contracting, homogeneous bounded spheres. The timescales are defined so that t_f/t_c is the ratio of the collapse velocity to v_f; this ratio is assumed to be constant in space and time. Finally, we consider in turn particles bound to the contracting sphere, particles initially unbound but captured during the collapse, and unbound particles. These highly idealized exercises in classical mechanics are instructive.

a) Slow Contraction ($t_f \ll t_c$)

First let us consider the rate at which initially unbound particles are captured by the decreasing potential. For $t_f \ll t_c$, the potential changes slowly during a particle orbit so that the change in energy can be calculated by perturbation theory. Taking the particle orbits to be those in a steady potential, the energy change is found by integrating the time-varying force along the orbit. The orbit is characterized by the energy per unit mass E and the angular momentum per unit mass L; the change in energy in one passage is found to be

$$\Delta E = -\frac{3\sqrt{2}}{4}\left(\frac{GM}{R}\right)\left(\frac{t_f}{t_c}\right) \times \left\{\left[\frac{ER}{GM}+\frac{1}{2}\right](2\sin^{-1}x - \pi) + 2\left[\frac{2ER}{GM}+2-\frac{L^2}{GMR}\right]^{1/2}\right\}, \tag{6}$$

where $x = (E + GM/2R)[(E + 3GM/2R)^2 - GML^2/R^3]^{-1/2}$. Here, M and R are the mass and radius of the isolated, contracting sphere; E is measured relative to infinity. It is clear from Eq. (6) that, for $t_f \ll t_c$, $|\Delta E| \ll GM/R$. Hence, only those particles with initial energies $E \ll GM/R$ can be captured; also, terms of order ER/GM may be ignored. Now, only those particles with angular momenta less than $L_{\max} = (2R^2E + 2GMR)^{1/2}$ will pass through the contracting sphere; define $\lambda \equiv L/L_{\max}$. Equation (6) reduces to

$$\Delta E = -\frac{3\sqrt{2}}{4}\left(\frac{GM}{R}\right)\left(\frac{t_f}{t_c}\right)\left\{[8(1-\lambda^2)]^{1/2} - \frac{\pi}{2} + \sin^{-1}\left[\left(\frac{1}{9-8\lambda^2}\right)^{1/2}\right]\right\}, \tag{7}$$

Defining $\beta \equiv ER/GM$, the capture condition may be written as

$$\beta < \frac{t_f}{t_c}g(\lambda^2), \tag{8}$$

where

$$g(\lambda^2) = \frac{3\sqrt{2}}{4}\left\{[8(1-\lambda^2)]^{1/2} - \frac{\pi}{2} + \sin^{-1}\left[\left(\frac{1}{9-8\lambda^2}\right)^{1/2}\right]\right\}. \tag{9}$$

Now, $g(\lambda^2)$ is a monotonically decreasing function of λ^2; $g(0) = 1.7$. Equation (8), therefore, can be inverted to solve for the maximum value of $\lambda \equiv \lambda_*(\beta t_c/t_f)$ for

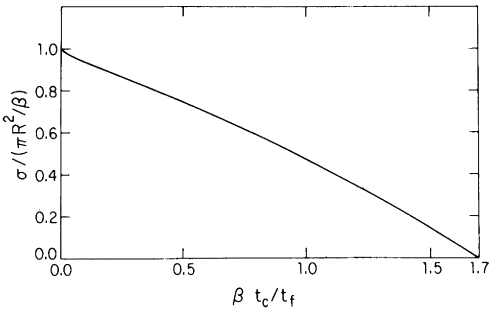


FIG. 1. The function λ_*^2, which is the capture cross section σ normalized to the cross section for passage through the region $\pi R^2/\beta$, is given in terms of the dimensionless parameter $\beta\, t_c/t_f$.

which capture will occur. Then, the capture cross section is

$$\sigma(\beta t_c/t_f) = \frac{\pi R^2}{\beta}\lambda_*^2(\beta t_c/t_f). \tag{10}$$

The function λ_*^2 is shown in Fig. 1; note that only particles with $\beta t_c/t_f < g(0) = 1.7$ can be captured, and that $\pi R^2/\beta$ is the cross section for passage through the contracting sphere.

The capture rate is found by integrating the product $n_\nu(v)\sigma(v)v$ over the velocity distribution of the neutrinos. The nondegenerate neutrinos are assumed to follow a Maxwell–Boltzmann distribution characterized by a line-of-sight velocity dispersion Δv_ν. The capture rate is

$$\langle n_\nu \sigma v\rangle = (8\pi)^{1/2} n_{\nu\infty}\frac{(GM)^2}{(\Delta v_\nu)^3}\frac{t_f}{t_c}\int_0^\infty \lambda_*^2(y) \times \exp(-\alpha y)dy, \tag{11}$$

where $\alpha = (t_f/t_c)(GM/R\Delta v_\nu^2)$ and $n_{\nu\infty}$ is the number density of free neutrinos far from the contracting sphere. For sufficiently slow collapse ($t_f \ll t_c$) and/or for $GM/R \ll \Delta v_\nu^2$, $\alpha \approx 0$ and the integral in Eq. (11) may be done analytically as λ_*^2 is just the inverse of g in Eq. (9) (in any case this will yield an upper limit).

$$\langle n_\nu \sigma v\rangle \approx 0.91(8\pi)^{1/2} n_{\nu\infty}\frac{(GM)^2}{(\Delta v_\nu)^3}\frac{t_f}{t_c}. \tag{12}$$

To find the total number of particles captured in the course of the collapse, the capture rate (12) is integrated over the collapse time

$$\frac{N_{\nu c}}{N_{\nu i}} \approx 0.51\left(\frac{GM}{R_i\Delta v^2}\right)^{3/2}. \tag{13}$$

In Eq. (13), $N_{\nu c}$ is the number of captured neutrinos and $N_{\nu i} = (4\pi/3)n_{\nu\infty}R_i^3$ is the number of neutrinos initially in the precollapse sphere of radius R_i. If the neutrinos and the ordinary matter had originally collapsed together and relaxed violently then

$$\frac{GM}{R_i\Delta v_\nu^2} \approx \left(\frac{M}{M_T}\right)^{2/3}, \tag{14}$$

where M_T is the total mass of the original, relaxed region (e.g., a galaxy) and M is the mass of the region undergoing further collapse (e.g., a gas cloud). Comparing Eqs. (13) and (14) it is clear that for $M \approx M_T$, a significant fraction of the neutrinos initially present in the collapsing sphere will be captured. For most astrophysically interesting cases however, $M \ll M_T$ and virtually none of the original neutrinos is captured.

A further comment is required regarding these "captured" neutrinos. The captured neutrinos are very weakly bound and, therefore, predominantly at large radii. A typical bound neutrino will have a binding energy of order ΔE [see Eq. (6)] and a typical radius of order

$$r_\nu \approx \frac{GM}{\Delta E} \approx \left(\frac{t_c}{t_f}\right) R \gg R_i \gg R_f. \tag{15}$$

Were the collapsing region (final radius R_f) not completely isolated these weakly bound neutrinos would be completely stripped; thus in realistic situations few neutrinos will be captured in slow collapse. It is physically more relevant to consider the number of bound neutrinos within R_f; integrating over the single particle spatial distribution we find

$$\frac{N_{\nu f}}{N_{\nu i}} \approx 0.05\left(\frac{GM}{R_i\Delta v_\nu^2}\right)^{3/2}\left(\frac{t_f}{t_c}\right)^{3/2} \approx 0.05\left(\frac{M}{M_T}\right)\left(\frac{t_f}{t_c}\right)^{3/2}. \tag{16}$$

Thus, even for $M \approx M_T$, only a small fraction of the initial neutrinos remains in the collapsing region for slow collapse $t_f/t_c \ll 1$. Although few neutrinos are captured inside the collapsing volume, their density can increase considerably; we return later to the implications of this result.

Consider, now, the dynamical evolution of heavy neutrinos which are initially bound to the collapsing region. The bound neutrinos may have orbits entirely outside, entirely inside or, partially inside the collapsing region. A bound neutrino whose orbit lies outside the collapsing volume sees a time-independent potential; such neutrinos will not contract. Of more interest are those neutrinos with orbits initially within the collapsing region [of radius $R(t)$; initial radius R_i].

A neutrino with $r_\nu < R$ will feel the potential of a three-dimensional harmonic oscillator with a frequency which varies as

$$\omega^2 \sim \frac{GM}{R^3} \sim R^{-3}. \tag{17}$$

The extrema of the orbit of such a neutrino are

$$r_{\nu\pm}^2 = \left(\frac{1}{\omega}\right)\left(\frac{E_\nu}{\omega}\right)\left\{1 \pm \left[1 - \left(\frac{L}{E_\nu/\omega}\right)^2\right]^{1/2}\right\}. \tag{18}$$

In Eq. (18), E_ν is the neutrino energy per unit mass and L is the neutrino angular momentum per unit mass. For slow collapse, E_ν/ω is conserved (E_ν/ω is an adiabatic

neglecting the self-gravity of the neutrinos. The results can be generalized easily to include the neutrino mass. We define $f = [M_M + M_\nu(t)]/M_o$, where M_ν is the mass of neutrinos within the collapsing region (which will decrease with time if neutrinos are left behind) and $M_o = M_M + M_\nu(0)$ is the initial total mass of the region. Equation (28) generalizes to

$$f^2 \frac{d^2y}{dz'^2} + \frac{1}{2}\frac{d(f^2)}{dz'}\frac{dy}{dz'} = \alpha^2/y^3,$$

where

$$z' = \left(2\int_x^1 \frac{f(x')}{x'^2}\,dx'\right)^{1/2}.$$

This equation may easily be integrated; in fact, the solution is identical to Eq. (29) if we replace z with z'', where

$$z'' = \int_x^1 \frac{dx'}{x'^2}\left(2\int_{x'}^1 \frac{f(x'')dx''}{x''^2}\right)^{-1/2}.$$

The integrand in this equation is positive-definite. Thus, z'' increases monotonically as the collapse proceeds. The number of neutrinos collapsing with the matter is again given by Eq. (30), if we replace z with z''. The neutrinos are left behind just as before. The mass of collapsing neutrinos will therefore decrease as the collapse proceeds; thus, f decreases during the collapse. Then, it is easy to see that $z'' > z$ and the neutrinos are left behind even more quickly if we include their self-gravity. After the collapse has proceeded by a factor of more than 2 or so, the collapsing mass is just the ordinary matter mass M_M. Then, $z'' = z(M_o/M_M)^{1/2}$, and the number of collapsing neutrinos are given by

$$\frac{N_{\nu c}}{N_{\nu i}} = \frac{4}{3\sqrt{\pi}}\left(\frac{\gamma^2 M_M}{z^2 M_o}\right)^{3/2} (\gamma/z \ll 1), \quad (33b)$$

which is the generalized version of Eq. (31).

It is clear from Eqs. (33a) and (33b) that the neutrinos will be left behind even more quickly if they contribute significantly to the mass of a collapsing region. Hence, our neglect of the self-gravity of the neutrinos in Secs. II *a*) and III *b*) leads to an over estimate of the number of collapsing neutrinos. This, then, only strengthens our conclusion that very few neutrinos can collapse along with ordinary matter to form condensed objects such as, for example, stars.

IV. TWO-BODY INTERACTIONS

Neutrinos and ordinary matter can exchange energy through fluctuations in the gravitational potential. In Sec. II and Sec. III, we discussed the effect of fluctuations in the overall potential of a region; here we consider small-scale fluctuations. In a region composed of subunits (i.e., stars in a galaxy or a globular cluster, or galaxies in a cluster of galaxies), the smallest scale fluctuations are due to the motion of these subunits. We use the term "two-body interactions" to describe the cumulative effect of these interactions.

The standard theory is essentially unchanged in the presence of the heavy neutrinos. One can define a Spitzer reference time (Spitzer and Härm 1958) for the subunits (we will call them "stars" for convenience) as

$$t_{\rm ref} = \frac{\sqrt{2}(\Delta v_*)^3}{\pi G^2 M_*^2 n_* \ln (3R\Delta v_*^2/GM_*)}. \quad (34)$$

Here, Δv_*, and M_*, and n_* are the line-of-sight velocity dispersion, individual mass, and number density of the "stars," and R is the radius of the entire region. This expression ignores the presence of the heavy neutrinos. Let us assume the neutrinos and stars have virialized, so that $\Delta v_\nu = \Delta v_*$ (see Sec. II). Then the neutrinos affect only the denominator of Eq. (34). Although the logarithmic term is slowly varying, the mass of the neutrinos is at least 10^{57} times smaller than the "stars", so even this term would be affected if Eq. (34) held rigorously. However, the argument of the logarithm is actually a ratio of all possible impact parameters in the system (Spitzer 1962). Obviously, the minimum impact parameter for a collision between a heavy neutrino and a "star" cannot sensibly be much less than the size of the "star." When this restriction is applied, the logarithm for neutrinos is found to agree with the value for stars (within 40%) for astrophysically interesting cases (globular clusters, galaxies, and clusters of galaxies).

Thus the neutrinos differ from the "stars" principally in having a different mass and density in Eq. (34). They will have two effects on the evolution of such systems. First, the neutrinos represent an additional source of dynamic friction in the cluster. The time scale for dynamic friction of the more massive "stars" will be (Spitzer and Härm 1958)

$$t_{DF} \approx 0.66\, t_{\rm ref}\left(\frac{\rho_*}{\rho_* + \rho_\nu}\right). \quad (35)$$

Here ρ_* and ρ_ν are the mass densities of the "stars" and heavy neutrinos, respectively. This time scale should be compared to timescales that might be derived ignoring the presence of heavy neutrinos. First of all, if only the mass of "stars" is considered, the dynamic friction timescale is $t_{DF}^\circ \approx 0.66\, t_{\rm ref}$. However, for some systems the total mass of the system can be derived from the virial theorem. If this total mass (from "stars" and heavy neutrinos) is assumed to be due only to stars, the dynamical friction timescale is considerably shortened; i.e., $t_{DF}' = 0.66\, t_{\rm ref}\,[\rho_*/(\rho_* + \rho_\nu)]^2$. Thus, $t_{DF}^\circ \geq t_{DF} \geq t_{DF}'$.

This inequality may most easily be compared to the observations of clusters of galaxies. There the virial mass of the cluster is known to exceed by a large factor the mass in luminous material in galaxies (see Sec. V and references therein). If this "missing mass" is physically associated with the visible galaxies, the dynamical friction time scale is relatively short ($t_{DF} = t_{DF}'$). A substantial amount of mass segregation would be expected; this is

certainly not observed (Rood 1965; White 1977). The observed galaxy distribution in the Coma cluster seems to require that the "missing mass" be uniformly distributed with the same *overall* distribution as the galaxies (White 1977); in Sec. V we show that this is the distribution expected for heavy neutrinos which are bound to a cluster.

If dynamic evolution proceeds further, evaporation of the neutrinos will become important. Because of their low mass, heavy neutrinos evaporate much more rapidly than "stars". From Spitzer and Härm (1958) we find $t^{\nu}_{ev} \approx 4.88\, t_{\rm ref}$, while $t^{*}_{ev} = 88\, t_{\rm ref}$, where t^{ν}_{ev} and t^{*}_{ev} are the evaporation times for heavy neutrinos (low mass limit) and the average "stars." Thus, in evolved systems such as open galactic clusters or globular clusters, many of the heavy neutrinos originally bound to the region may have been lost.

If the heavy neutrinos contain a substantial fraction of the mass in the cluster, their evaporation will cause a substantial reduction in the size of the cluster. Because evaporation decreases slightly the energy of the cluster, the radius must decrease to $R' \leq R(M'/M)^2$. Here, M, R and M', R' are the initial and final mass and radius, respectively. At the same time the velocity dispersion increases. The net effect is to shorten the dynamical timescale rather rapidly as evaporation proceeds.

For a typical globular cluster, t^{ν}_{ev} is about a Hubble time. Thus, if heavy neutrinos initially dominated the mass of these systems, their contribution must still be significant. As there is little evidence for any "missing mass" in globular clusters, they apparently did not form with a significant heavy neutrino component.

V. ASTROPHYSICAL CONSEQUENCES

The analysis of the previous sections shows that whenever the heavy neutrinos account for a significant fraction of the total mass (so that $t_f \lesssim t_c$), their collapse cannot keep pace with that of the ordinary matter for more than a factor of two. In nondissipative collapse, potential energy will be turned into kinetic energy; violent relaxation will establish virial equilibrium in which the heavy neutrinos and the ordinary matter have the same velocity dispersions. The ordinary matter may radiate its energy away and collapse further; the heavy neutrinos cannot. Our main conclusion, therefore, is that in any structures formed by dissipationless collapse the ratio of heavy neutrinos to ordinary matter is preserved. In such objects the heavy neutrinos may dominate the total mass. In contrast, any structures formed by collapse involving dissipation will have left the neutrinos behind. We now turn to the implications of these results for galaxy formation, the masses of galaxies and clusters, and stellar structure and evolution. For a discussion of the gamma ray fluxes expected from the annihilation of heavy neutrino pairs see Gunn *et al.* (1978) and Stecker (1978).

a) Galaxy Formation

It is generally accepted that the large-scale structure in the universe has evolved, via gravitational instability, from small-amplitude fluctuations in an otherwise isotropic, homogeneous, expanding cosmology. The origin of the required primordial fluctuations is unknown (but see Rees 1972 and Zeldovich 1972 for intriguing speculations). In the absence of deeper understanding there appears no reason to exclude the heavy neutrinos from such perturbations; in any case, since they respond gravitationally, an initially smooth heavy neutrino background will be perturbed by the necessary curvature fluctuations.

The fluctuations which are amplified gravitationally (curvature perturbations) are of two types. The matter and the radiation can be perturbed together such that the entropy per particle remains constant; such perturbations are called *adiabatic*. In contrast, the matter may be perturbed while the photons remain smoothly distributed; such perturbations are called *isothermal*. For a recent review of the evolution of perturbations in an expanding, hot, Big Bang cosmology, see Jones (1976). There have been several recent comparisons of theory with observation (Gott and Rees 1975; Gott 1977). Here we consider the extent to which the standard results may be modified by the presence of heavy neutrinos.

There are two epochs of importance. Early on the ordinary matter is ionized and therefore strongly coupled to the radiation. This coupling is broken at recombination which occurs at a redshift

$$z_{\rm rec} \approx 10^3. \tag{36}$$

At present, the density in ordinary matter (plus, perhaps, heavy neutrinos), exceeds that in radiation (photons, e-neutrinos, μ-neutrinos). The matter and radiation densities were equal at a redshift

$$z_{\rm eq} \approx 10^{4.5}\Omega_0 h_o^2, \tag{37}$$

where Ω_0 is the usual density parameter and the Hubble parameter is $H_o = 100\, h_o$ (km s^{-1} Mpc^{-1}). At epochs earlier than recombination perturbations enter the horizon and grow until their mass is exceeded by the Jeans mass. For masses less than the Jeans mass, the restoring force due to radiation pressure forces the perturbations to oscillate like acoustic waves. Perturbations on scales of masses larger than the Jeans mass can continue growing. During this prerecombination phase there is an important difference between adiabatic and isothermal perturbations. On small scales, adiabatic fluctuations are damped by viscosity and radiative diffusion (Silk 1968; Weinberg 1971); by recombination, perturbations on mass scales $M \lesssim M_D \approx 3 \times 10^{12} (\Omega_0 h_0^2)^{-5/4}\, M_\odot$ have been severely attenuated (Jones 1976). In contrast, isothermal perturbations are not dissipated. Except on very large scales then, perturbations in the matter cannot grow at least until recombination. Further, density perturbations in the matter

cannot begin growing until the universe is matter dominated ($z < z_{eq}$) (Meszaros 1974). Thus, in a low density universe, for which $z_{eq} < z_{rec}$, growth does not begin at z_{rec} but at z_{eq}.

How is all this changed by the presence of heavy neutrinos? Surprisingly little. During the radiation dominated era, the matter (including now the heavy neutrinos) is dynamically unimportant. It is, therefore, only during the matter dominated epoch that the heavy neutrinos might influence the results. In a high density universe dominated by heavy neutrinos, growth on all scales can begin at $z_{eq} > z_{rec}$ since the heavy neutrinos are not coupled to the radiation. This has two important consequences. Now the damping of adiabatic fluctuations stops at $z_{eq} > z_{rec}$ so the critical scale M_D is smaller than previous estimates (Silk 1968, Weinberg 1971, Jones 1976); adiabatic perturbations can no longer be dismissed as the origin of structure down to scales of galaxies (see, for example, Gott 1977). The second important point is that *all* scales begin growing at z_{eq} in distinction to the "standard" case where only scales $M > M_{Jeans}$ can grow. Thus the "spike" in the covariance function predicted by Gott and Rees (1975) should not occur; it is not observed (Peebles 1974). Thus, the weak arguments from the covariance function data against a high-density universe become weaker still if most of the mass is in the form of heavy neutrinos. For a low-density universe in which $z_{eq} \lesssim z_{rec}$, the standard analysis is unmodified.

b) Masses of Galaxies and Clusters

We have seen that the presence of heavy neutrinos modifies only slightly the standard theory of galaxy formation and clustering. In Sec. II we saw that the heavy neutrinos would collapse with the ordinary matter and virialize with an isothermal distribution; in Sec. III it was shown that the heavy neutrinos are left behind in any subsequent collapse involving dissipation. We thus anticipate that for those objects formed by dissipationless collapse, the distributions of the ordinary matter and the neutrinos are identical. In contrast, the neutrinos would be in an extended halo surrounding any object which has collapsed while radiating away some of its energy; this weakly bound halo would be easily stripped. Let us consider some illustrative examples.

Elliptical galaxies and clusters (or groups) of galaxies are good candidates for objects formed by dissipationless collapse (Gott 1977). After collapse and violent relaxation the velocity dispersion and spatial distribution of all constituents (i.e., stars, heavy neutrinos, etc.) should be the same. For an isothermal distribution the mass increases with radius, in convenient units we have

$$\frac{M}{M_\odot} \approx 10^{11} \left(\frac{\Delta v_{100}}{2}\right)^2 \left(\frac{R_{kpc}}{10}\right). \qquad (38)$$

In Eq. (38), the velocity dispersion is in units of 100 km s^{-1} and the radius is in kiloparsecs. Typically, for galaxies $\Delta v_{100} \approx 2$–3 and $\Delta v_{100} \approx 10$ for clusters. If heavy neutrinos exist and dominate the mass density in the universe ($\rho_\nu \gtrsim 1/2\rho_{TOT}$) then Eq. (38) may well describe the mass distribution in clusters, E-galaxies and the halos of S-galaxies; most of the mass in these objects (and, indeed, in the universe) may be truly invisible. Consider some of the consequences.

In clusters of galaxies the M/L ratios are known to have values around 200 (van den Bergh 1977) while individual galaxies have values between 5 and about 30 (Roberts 1976; Burbidge and Burbidge 1976); here we assume $H_o = 50$ km s^{-1} Mpc^{-1}. A similar quantity of matter is estimated to be in hot gas, as is required by models that account for the x-ray emission by thermal bremsstrahlung. Therefore, of the order of 70%–80% of the cluster mass is in the form of nonluminous material (or very high M/L). As was already noted, Rood (1965) and White (1977) have shown that this mass cannot be associated with the galaxies (or galaxy-mass type objects) because too much mass segregation would occur, which is not observed. The unseen matter has to be diffuse and distributed in the same manner as the galaxies. Studies of the radial distribution of the galaxies in clusters show it to be well approximated by isothermal sphere gas models (Bahcall 1977) although other parametrizations are possible (Rood *et al.* 1972). By the arguments of Sec. II and Sec. III, we predict an isothermal distribution for the heavy neutrinos, which should follow closely the galaxy distribution, since the collapse in clusters is expected to be dissipationless. Gunn (1977) has calculated the detailed radial distribution that inert matter (such as low-mass stars or heavy neutrinos) would have upon collapse in a forming protogalaxy or cluster. He found that under fairly wide conditions the density radial dependence is a power law with an exponent ~ 2, and a mass varying linearly with radius.

Elliptical galaxies are believed to be formed during a nearly dissipationless collapse (although some models require substantial dissipation). For the same reasons as previously mentioned, heavy neutrinos are expected to follow the normal matter distribution. The precise form will depend on the actual dissipation involved. However, more dissipation is expected than in clusters and at some point in the collapse neutrinos will be left behind. The central regions of ellipticals should have a lower ratio of neutrinos to total mass than in comparable regions of the clusters. Thus, the M/L ratios would be an increasing function of radius. M/L values at the central regions of elliptical galaxies have been measured at about 3–15 $M_\odot/L_\odot$ (Williams 1977 and references therein) and are smaller than values deduced for the whole galaxies.

In spiral galaxies a large amount of dissipation is expected from early stages in the collapse of the protogalaxy to form the disks. The slow collapse in this situation means that the heavy neutrinos will be left behind. One expects that the inner regions of these galaxies contain a smaller proportion of heavy neutrinos. The neutrinos

left behind would still be bound to the galaxy and form an extended isothermal halo, with a larger relative scale height than the halos of elliptical galaxies.

If heavy neutrinos make up the missing mass in clusters of galaxies, we would expect that isolated galaxies (i.e., no tidal stripping) would show *total M/L* ratios similar to clusters (i.e., ~200). From Eq. (38) this implies that typical large galaxies have halos of heavy neutrinos extending for several hundred kiloparsecs.

c) Stellar Structure and Evolution

If heavy neutrinos exist and could collapse with ordinary matter to form stars, they would substantially affect stellar structure. They could contribute an extra, invisible source of gravitation, provide an additional source of luminosity through annihilation, and increase the rate of energy transport through elastic collisions with protons.

We have argued above that a substantial fraction of the mass associated with galaxies and clusters of galaxies might be in the form of massive neutrinos. Of course, if the same were true for stars, stellar structure would be drastically modified. For example, the lifetimes of stars would be shortened; in fact, the Sun would not have survived to the present epoch if most of its mass were in the form of massive neutrinos. Fortunately, our analysis (Sec. III) has shown that subsequent to an initial collapse and virialization, the neutrinos cannot significantly follow any further collapse. (The small fraction of neutrinos which do collapse with stars is estimated below.) Thus, heavy neutrinos will not have a significant gravitational effect in stars.

Because of the high densities in stars, weak interaction processes can become significant; of course, the main energy source of solar mass stars, the proton–proton reaction, is just such a process. For heavy neutrinos, the most important of these effects are annihilation and elastic scattering on protons. The rate coefficient for annihilation is estimated by *LW* to be

$$\langle \sigma v \rangle_a \simeq G_F^2 m_\nu^2 N_a/\pi \simeq 3.5 \times 10^{-27} m_{\mathrm{GeV}}^2 \ \mathrm{cm}^{-3}\ \mathrm{s}^{-1}, \tag{39}$$

where G_F is the Fermi coupling constant and N_a is the number of decay channels, assumed to be $N_a = 14$. Of course, annihilation will only occur if both neutrinos and antineutrinos are present in the star. If the neutrinos are completely symmetric (see Sec. I), then the lifetime of the neutrinos against annihilation is

$$t_a \simeq \frac{1}{\langle \sigma v \rangle_a n_\nu} \simeq 10^{19} m_{\mathrm{GeV}}^{-2} n_\nu^{-1}\ \mathrm{yr}. \tag{40}$$

Obviously, symmetric heavy neutrinos could not survive at stellar densities, but we've argued above that neutrinos cannot follow the collapse of ordinary matter to form stars. Some *small* fraction of the neutrinos will be captured during the collapse, and this could result in a relatively high-density for these neutrinos [see Eq. (16)]. Let N_ν/N_p be the ratio of the number of heavy neutrinos captured by a star to the number of protons in the star; thus $N_\nu/N_p \ll 1$. For a star with $\langle \rho \rangle \approx 1$, Eq. (40) gives

$$t_a \simeq 10^{-5} m_{\mathrm{GeV}}^{-2} \left(\frac{N_\nu}{N_p}\right) \mathrm{yr}. \tag{41}$$

As an example, if we consider the Sun we find that the present fraction of heavy neutrinos must be less than $N_\nu/N_p \lesssim 2 \times 10^{-15} m_{\mathrm{GeV}}^{-2}$, if the heavy neutrinos are symmetric and have survived for the 4.6×10^9-yr lifetime of the Sun.

The annihilation of symmetric heavy neutrinos could provide an additional source of luminosity for stars. If neutrinos collapsed to stellar densities, this additional source could easily rival fusion; however, we have argued that heavy neutrinos do not participate significantly in the collapse to form stars. In fact it may easily be shown from Eq. (16) that neutrino annihilation is never a significant source of stellar luminosity, either during or following the collapse to form the star.

Neutrinos present in the Sun would elastically scatter on protons; we estimate the cross section for this process to be $\sigma_{\nu p} \approx 10^{-38}\ \mathrm{cm}^2$ for $m_{\mathrm{GeV}} \gtrsim 1$. This process could have a number of significant effects. First, it produces a distribution of heavy neutrinos quite different than the distribution of matter in the star. Second, it allows neutrinos to exchange energy with ordinary matter; because the path lengths of these neutrinos are very long they can effectively transport energy within a star.

Collisions with protons will tend to bring the neutrinos into equipartition with the gas. However, the neutrinos are much more massive than the mean molecular weight of the gas, and thus are strongly concentrated to the center of the star. For example, heavy neutrinos with the temperature of the gas at the Sun's center would be confined within a radius $r \approx 0.2 m_{\mathrm{GeV}}^{-1/2} R_\odot$. Thus, most of the heavy neutrinos would be within the inner 1% of the Sun's volume. The velocity distribution of the heavy neutrinos will also differ from that of the gas, because high-velocity neutrinos can escape by virtue of their long mean free paths (greater than a stellar radius for the cross section given above). There is a continual evaporation of heavy neutrinos with velocities exceeding the escape speed. Although the distribution of heavy neutrinos in any realistic stellar model is difficult to calculate in detail, it should be reasonably well represented by a Maxwell–Boltzmann distribution cut off at the escape velocity (Gilliland 1978). We have used this simple distribution to estimate the evaporation rate and central density of heavy nuetrinos within the Sun. The evaporation time is shorter than the lifetime of the Sun unless $m_{\mathrm{GeV}} \geq 4$. Because the mass occurs in an exponential, this limit is insensitive to the details of the model, and is a reasonably strong limit (a mass of $m_{\mathrm{GeV}} = 2$ is definitely excluded, for example).

Through these elastic collisions with protons, heavy neutrinos can transport energy very effectively within

invariant); of course, for a central potential, L is conserved too. It then follows from Eq. (18) that

$$r_{\nu\pm} \sim \omega^{-1/2} \sim R^{3/4}. \tag{19}$$

The orbits of these bound neutrinos (those inside the collapsing volume) shrink, but more slowly than the volume as a whole. As long as $r_{\nu+} < R$, the neutrino orbit continues to shrink according to Eq. (19). For orbits initially and/or subsequently outside, then shrinkage is slower and eventually (when $r_{\nu-} > R$) they will get left behind. The number of bound neutrinos remaining inside the collapsing volume is

$$\frac{N_{\nu f}}{N_{\nu b}} \approx \left(\frac{R}{R_i}\right)^{3/4}. \tag{20}$$

Although few neutrinos remain inside the collapsing region ($N_{\nu f} \ll N_{\nu i}$), their density can increase considerably:

$$\frac{n_\nu}{n_{\nu b}} \approx \left(\frac{R_i}{R}\right)^{9/4}. \tag{21}$$

Here, $N_{\nu b}$ and $n_{\nu b}$ are the number and number density of neutrinos which are initially bound within the region.

As a final consideration we turn to those neutrinos not captured but influenced by the potential of a collapsing region. Such particles may be characterized by a distribution function in phase space $f(\vec{r},\vec{v},t)$. Since the heavy neutrinos are collisionless, it follows from Liouville's theorem that f is constant along particle trajectories. For slow contraction (or, if the collapse is complete), the energy is nearly conserved along trajectories; thus, $f = f(E,t)$. Since the neutrinos have a Maxwell-Boltzmann distribution far from the perturbation, they have the same phase-space distribution everywhere. Thus, the spatial distribution of the neutrinos in the vicinity of a potential $\phi(r)$ is

$$n_\nu(r) = \int n(\vec{r},\vec{v})d^3v = \frac{n_{\nu\infty} e^{|\phi(r)|/\Delta v_\nu^2}}{(2\pi\Delta v_\nu^2)^{3/2}} \times \int_{(2|\phi(r)|)^{1/2}}^{\infty} 4\pi v^2 \exp\left(\frac{-v^2}{2\Delta v_\nu^2}\right) dv, \tag{22}$$

$$n_\nu(r) = \left(\frac{4}{\pi}\right)^{1/2} n_{\nu\infty} \exp\left(\frac{|\phi(r)|}{\Delta v_\nu^2}\right) \Gamma\left(\frac{3}{2}, \frac{|\phi(r)|}{\Delta v_\nu^2}\right). \tag{23}$$

The distribution of unbound neutrinos depends on the ratio of the escape velocity $[(-\phi)^{1/2}]$ to the velocity dispersion. For most cases of astrophysical interest this ratio is not large so that $n_\nu(r) \approx n_{\nu\infty}$. For example, neutrinos that collapsed along with our galaxy would have $\Delta v_\nu \approx 200$ km s^{-1} which may be compared with a similar value for the escape velocity from a star. Thus, the increase in density of galactic heavy neutrinos near a star or cluster neutrinos near a galaxy is not very great. Even for a condensed object (black hole or neutron star) for which $|\phi| \approx c^2$, the heavy neutrino density (from unbound particles) is only enhanced by the factor of about $c/\Delta v_\nu$ and is negligible compared to the density of baryonic matter.

Let us summarize the results of this section. During slow collapse ($t_f \ll t_c$) the heavy neutrinos are generally left behind; in objects formed by slow collapse (e.g., gas clouds, stars, possibly spiral galaxies) the ratio of the density of heavy neutrinos to the ordinary matter density should be much less than the average ratio in the universe. Significant numbers of neutrinos can be captured by a collapsing region only if the mass of the region is comparable to the Jeans mass for neutrinos (i.e., $GM/R \approx \Delta v_\nu^2$) and if the collapse is rapid. Even neutrinos initially bound to the collapsing region get left behind ($r_\nu \approx R^{3/4}$). Hence, even if they dominate the mass of the universe, heavy neutrinos should contribute negligibly, either by mass or by numbers, to collapsed objects such as galactic nuclei, dense clouds, and stars. There is, however, one important point worth emphasizing. Despite most neutrinos being left behind during slow collapse, those few remaining neutrinos can have their density considerably enhanced [see Eq. (17) and (21)]. If the neutrino density is large enough, weak interaction processes such as annihilation of $\nu\bar{\nu}$ pairs will become important. It may well be that some neutrinos which survived the "freeze-out" in the early universe are annihilated subsequently. The gamma rays from this annihilation may be detectable (see Gunn *et al.* 1978).

b) Fast Collapse ($t_f \approx t_c$)

Consider a uniform, fully mixed sphere of neutrinos and ordinary matter (but, assume the neutrinos contribute negligibly to the total mass). At $t = 0$, "stop" the ordinary matter and let it collapse in free fall; what happens to the neutrinos? First let us follow the ordinary matter.

For an ordinary matter particle at r_i with $dr/dt = 0$ at $t = 0$, the equation of motion is

$$\frac{d^2r}{dt^2} = -\frac{GM_i}{r^2}. \tag{24}$$

The mass interior to the particle $M_i = 4\pi/3\ \rho(t=0)\ r_i^3$ is a constant. Then, the first integral of the motion is

$$\left(\frac{dr}{dt}\right)^2 = 2GM_i\left(\frac{1}{r} - \frac{1}{r_i}\right). \tag{25}$$

Now consider a heavy neutrino which, at $t = 0$, is located at $r = r_i$; the neutrino will have an angular momentum per unit mass $L = v_\perp r_i$ and a radial velocity v_r (where $v^2 = v_\perp^2 + v_r^2$). The neutrino sees a harmonic oscillator potential with a time dependent oscillator frequency. Thus,

$$\frac{d^2r_\nu}{dt^2} - \frac{L^2}{r_\nu^3} + \omega^2(t) r_\nu = 0, \tag{26}$$

where $\omega^2 = GM/r^3$. Using Eq. (24), Eq. (26) may be rewritten as

$$\frac{d^2r_\nu}{dt^2} - \frac{r_\nu}{r}\frac{d^2r}{dt^2} = \frac{L^2}{r_\nu^3}. \tag{27}$$

If we substitute $y = r_\nu/r$ and $z = [2(1-x)/x]^{1/2}$, where $x = r/r_i$, Eqs. (25) and (26) combine to give

$$\frac{d^2y}{dz^2} = \alpha^2/y^3, \tag{28}$$

where $\alpha = L/\omega_0 r_i^2$, and ω_0 is $\omega(t=0)$. Note that y gives the position of a neutrino relative to the position of an ordinary matter particle which stated at the same position; thus, it is a useful variable for describing the contraction of the neutrinos relative to that of the ordinary matter. The variable z, which goes from 0 to ∞, measures the extent of the collapse in the ordinary matter.

The solution to Eq. (28) subject to our initial conditions is

$$y = \left[1 + \frac{2v_r z}{\omega_0 r_i} + \left(\frac{v_r z}{\omega_0 r_i}\right)^2\right]^{1/2}. \tag{29}$$

Equation (29) can be used to find the position of an individual heavy neutrino given its initial position and velocity. By integrating over the initial velocity distribution (Maxwell–Boltzmann), the initial spatial distribution (a uniform sphere, like the ordinary matter), and the volume of the collapsing region, we can find the number of heavy neutrinos which collapse with the matter $N_{\nu c}$. When compared to the number of neutrinos originally occupying the region prior to collapse $N_{\nu i}$, we find (after considerable algebra)

$$\frac{N_{\nu c}}{N_{\nu i}} = \operatorname{erf}\left(\frac{2\gamma}{z}\right) + \frac{1}{4\sqrt{\pi}}\left[\left(\frac{z}{\gamma}\right)^3 - 6\left(\frac{z}{\gamma}\right)\right] - \frac{1}{4\sqrt{\pi}}\left[\left(\frac{z}{\gamma}\right)^3 - 2\left(\frac{z}{\gamma}\right)\right]\exp\left(-\frac{4\gamma^2}{z^2}\right). \tag{30}$$

Here, $\gamma^2 = (GM/2R_i\Delta v_\nu^2)$, where M is the total mass and R_i the total radius (at $t=0$) of the collapsing region. The degree of collapse in heavy neutrinos thus depends on the extent of the collapse, as measured by z, and the velocity dispersion of the neutrinos compared to the initial free-fall velocity of the region. When $\gamma/z \gg 1$, $N_{\nu c}/N_{\nu i}$ is unity; neutrinos are captured by a free-falling matter distribution only if their velocity dispersion is small and the collapse doesn't proceed too far. When $\gamma/z \ll 1$,

$$\frac{N_{\nu c}}{N_{\nu i}} = \frac{4}{3\sqrt{\pi}}\left(\frac{\gamma}{z}\right)^3 \quad (\gamma/z \ll 1). \tag{31}$$

$N_{\nu c}/N_{\nu i}$ reaches a value of one half for $\gamma/z = 1.6$.

Now, if the neutrinos were initially (for $t \leq 0$) in virial equilibrium within a (possibly larger) region of mass M_t and size R_t, then $\Delta v_\nu^2 \approx GM_t/R_t$. For example, M_t might represent a galaxy in which a gas cloud (of mass M) was collapsing to form stars. Thus $\gamma^2 \approx (M/M_t)^{2/3}$ from Eq. (14). The virialized neutrinos will follow the collapse of a free-falling region (that is $N_\nu/N_{\nu i} \approx 1$) only if $M = M_t$ and the collapse proceeds only by a factor of less than about 2. If the collapse proceeds much further or $M \ll M_t$, then Eq. (31) gives

$$\frac{N_{\nu c}}{N_{\nu i}} \approx \left(\frac{M}{M_t}\right)\left[\frac{R_i}{R_f} - 1\right]^{-1/2} \ll 1, \tag{32}$$

so that the heavy neutrinos are quickly left behind.

c) *Effects Due to the Mass of the Heavy Neutrinos*

In the previous analyses the heavy neutrinos were treated as test particles in that their contribution to the gravitational potential was ignored. Since the interesting astrophysical consequences follow if the universal mass density is dominated by heavy neutrinos, the previous approximations will not always be valid. It is, fortunately, easy to see that our conclusions are generally valid (indeed, even strengthened) when the effects due to the neutrinos are included.

In Sec. III *a*), we considered the distribution and density of "free" neutrinos (not bound to the matter perturbation) and of initially free neutrinos which were captured during slow collapse. Following Eq. (23), we showed that the excess density of free neutrinos near a condensed object is very small; we are therefore justified in ignoring the self-gravity of these neutrinos. Similarly, the number of neutrinos captured during the slow collapse is very small [Eq. (13)], and even those captured occupy a very large region, so that their contribution to the gravitational potential at the collapsing region is insignificant. It is easy to show that the effects of neutrino self-gravity on initially free neutrinos are completely insignificant for slow collapse.

Section III *a*) also considered the fate of neutrinos bound to a slowly collapsing region (Eq. 20). What happens if the bound neutrinos have a significant mass M_ν as compared to the mass of ordinary baryonic matter M_M? The adiabatic invariants of the particle orbits are not changed, and Eq. (19) applies as long as M is taken to be the total mass in the collapsing region $M = M_M + M_\nu$ in Eq. (17). Even if M were constant, Eq. (19) implies that the neutrinos are left behind in the collapse. The mass of neutrinos in the collapsing region M_ν will thus decrease; the neutrinos will be left behind even more quickly according to Eq. (19). In fact, Eq. (19) can easily be solved by including the neutrino mass in ω; one finds that Eq. (20) is an upper limit to the number of neutrinos within the bound region. When the collapse has proceeded by a large factor, the neutrinos are left behind and $M = M_M$. The number of bound neutrinos with the collapsed volume is then generally given by

$$\frac{N_{\nu f}}{N_{\nu i}} \approx \left(\frac{M_M}{M_0}\frac{R}{R_i}\right)^{3/4} \tag{33a}$$

instead of Eq. (20). M_o is the total mass of matter and neutrinos at the start of the collapse.

In Sec. III *b*), we considered fast collapse ($t_c = t_f$)

the core of a star. We have estimated the magnitude of the heavy neutrino energy transport within the core of the Sun, using a standard solar model. As stated previously, the neutrinos are confined to a very small region at the center of the Sun. However, their mean free paths vastly exceed this core size, so successive proton–neutrino collisions are not spatially correlated. This results in a nondiffusive mode of energy transport. The neutrinos relax to a distribution in equilibrium with the *average* gas temperature within the core, not the local temperature at each point. Within the inner parts of the core, elastic collisions transfer energy from gas to neutrinos; in the outer core, the neutrinos lose energy to the gas. Let L_ν be the luminosity transported within the Sun's core by this process; when this exceeds the luminosity carried by photons L_γ (the nuclear luminosity of the core), the core will become isothermal. Using the proton–neutrino cross section given earlier, we have estimated the luminosity L_ν from the collision rate, the volume of the neutrino core, and the temperature variation across the core in the standard solar model. We find

$$\frac{L_\nu}{L_\gamma} \simeq 10^{10}\, m_{\mathrm{GeV}}^{1/2} \left(\frac{N_\nu}{N_p}\right). \tag{42}$$

Thus, even an extremely small population of heavy neutrinos would dominate the thermal balance in the core of the Sun and cause the core to be isothermal. Among other consequences, we note that the predicted solar neutrino flux (normal, massless neutrinos) measureable by current detectors would be substantially reduced in agreement with the observations (Davis, Rogers, and Radeka 1972), if the Sun's core were more isothermal.

In order to determine the importance of this effect, we need to estimate the number of heavy neutrinos captured during the formation of the Sun. Of course if heavy neutrinos are matter–antimatter symmetric, annihilation will reduce the population of neutrinos to a very low level ($N_\nu/N_p \simeq 10^{-15}$), as discussed above. If neutrinos are not symmetric, we can calculate the number of neutrinos captured by the Sun during its formation from Eq. (16). If the neutrinos originally collapsed and virialized with our Galaxy, then the limits on the mass density of neutrinos from Sec. I imply

$$\frac{N_\nu}{N_p} \lesssim 10^{-12} \left(\frac{t_f}{t_c}\right)^{3/2} \tag{43}$$

from Eq. (16). We use this equation because the neutrinos will primarily be bound to the galaxy as a whole, not the presolar nebula, and because the collapse is probably slow. (In fact, rapid collapse also satisfied this limit.) Comparing this with Eq. (42) we see that it is unlikely that heavy neutrinos would have any significant effect on stars like the Sun.

VI. CONCLUSIONS

We have shown that heavy neutrinos are an ideal material from which to form the "missing mass" in clusters of galaxies and galactic halos. The heavy neutrinos would participate fully in the initial dissipationless collapse and virialization of matter perturbations; they are left behind during any subsequent fragmentation or dissipative collapse. In clusters of galaxies, they would give a missing mass component which mimics the galaxy distribution, but is not physically associated with individual galaxies. As such, they could account for the high virial masses and large M/L ratios in clusters, without producing more mass segregation than is observed.

Heavy neutrinos would form extended halos around galaxies; their detailed behavior would depend on the amount of dissipation involved in the galaxy formation. As this is the quantity which is conventionally thought to account for the differences between spiral and elliptical galaxies, we would expect elliptical galaxies to have higher M/L ratios within their luminous regions. However, the total M/L ratios for isolated galaxies ought to be similar.

Heavy neutrinos will be produced in the Big Bang with significant mass densities, as compared to the density of baryonic matter, if their mass lies in the range $1 \lesssim m_{\mathrm{eV}} \lesssim 50$ or $2 \lesssim m_{\mathrm{GeV}} \lesssim 15$ (*CM, LW*). If the neutrinos are not matter-antimatter symmetric, their densities could be even higher. The analytic theory we have given in Sec. II for the collapse of heavy neutrinos would apply to any such particles; we note, however, that the phase-space density of neutrinos in the lighter range (1–30 eV) is probably too low to account for the missing mass in clusters and galactic halos (Gunn 1977).

Remarkably enough, heavy neutrinos could dominate the mass of clusters of galaxies and individual galaxies without seriously affecting the structure of stars. There are two reasons for this; first, if neutrinos are symmetric, annihilation will reduce their population within stars to an insignificantly low value. Second, heavy neutrinos are left well behind during the dissipative collapse and fragmentation of stellar formation. It is perhaps fortunate that the number of neutrinos expected in stars is so low ($N_\nu/N_p \lesssim 10^{-12}$). Because elastic collisions with protons allow the neutrinos to transport energy very efficiently, stellar structure could be altered by even a very small ($N_\nu/N_p \lesssim 10^{-10}$) heavy neutrino component. This leads to an important corollary of our work. If a substantially more efficient way could be found of gathering heavy neutrinos into stars, one could solve the solar neutrino problem without seriously affecting other aspects of stellar evolution.

J. F., C. S., and G. S. are very grateful to the National Radio Astronomy Observatory for their hospitality during the time when this work was begun. J. F. was partially supported by National Science Foundation Grant AST 75-09751.A01. G. S. has profited from discussions with R. V. Wagoner and D. N. Schramm. We thank R. Gilliland for calculating several quantities of relevance to the effect of heavy neutrinos on stellar structure.

REFERENCES

Bahcall, N. A. (1977). Ann. Rev. Astron. Astrophys. **15,** 505.
Benvenuti, A. *et al.* (1977). Phys. Rev. Lett. **38,** 1110.
Burbidge, G. R., and Burbidge, E. M. (1976). In *Galaxies and the Universe,* edited by Sandage, Sandage, and Kristan (Univ. Chicago P., Chicago).
Cowsik, R., and McClelland, J. (1972). Phys. Rev. Lett. **29,** 669.
Davis, R., Jr., Rogers, L. C., and Radeka, V. (1971). Bull. Am. Phys. Soc. **16,** 631.
Gilliland, R. (1978). In preparation.
Gott, J. R. (1973). Astrophys. J. **186,** 481.
Gott, J. R. (1977). Ann. Rev. Astron. Astrophys. **15,** 235.
Gott, J. R., and Rees, M. (1975). Astron. Astrophys. **45,** 365.
Gunn, J. E. (1977). Astrophys. J. **218,** 592.
Gunn, J. E., Lee, B. W., Lerche, I., Schramm, D. N., and Steigman, G. (1978). Astrophys. J. (in press).
Jones, B. J. T. (1976). Rev. Mod. Phys. **48,** 107.
Lee, B. W., and Weinberg, S. (1977a). Phys. Rev. Lett. **38,** 1237.
Lee, B. W., and Weinberg, S. (1977b). Phys. Rev. Lett. **39,** 165.
Lynden-Bell, D. (1967). Mon. Not. R. Astron. Soc. **136,** 101.
Meszaros, P. (1974). Astron. Astrophys. **37,** 225.
Peebles, P. J. E. (1970). Astron. J. **75,** 13.
Peebles, P. J. E. (1974). Astrophys. J. Lett. **189,** L51.
Rees, M. J. (1972). Phys. Rev. Lett. **28,** 1669.
Rees, M. J., and Ostriker, J. (1977). Mon. Not. R. Astron. Soc. **179,** 541.
Roberts, M. S. (1976). In *Galaxies and the Universe,* edited by Sandage, Sandage, and Kristian (Univ. Chicago P., Chicago).
Rood, H. J. (1965). Ph.D. thesis, Univ. Michigan.
Rood, H. J., Page, T. L., Kintner, E. C., and King, I. R. (1972). Astrophys. J. **175,** 627.
Silk, J. (1968). Astrophys. J. **151,** 459.
Spitzer, L. (1962). *Physics of Fully Ionized Gases* (Interscience, New York).
Spitzer, L., and Härm, R. (1958). Astrophys. J. **127,** 544.
Stecker, F. W. (1978). Astrophys. J. (in press).
Steigman, G. (1973). Cargèse Lect. Phys. **6,** 505.
Steigman, G. (1976). Ann. Rev. Astron. Astrophys. **14,** 339.
van den Bergh, S. (1977). Vistas Astron. **21,** 71.
Weinberg, S. (1971). Astrophys. J. **168,** 175.
White, S. D. M. (1977). Mon. Not. R. Astron. Soc. **179,** 33.
Williams, T. B. (1977). Astrophys. J. **214,** 685.
Zeldovich, Ya. B. (1972). Mon. Not. R. Astron. Soc. **160,** 1p.

Constraint on the Photino Mass from Cosmology

H. Goldberg
Department of Physics, Northeastern University, Boston, Massachusetts 02115
(Received 22 February 1983)

A lower bound for the photino mass $m_{\tilde{\gamma}}$ as a function of the spin-0 fermion superpartner mass $m_{\tilde{f}}$ is derived as an extension of the calculation of Lee and Weinberg. The Majorana nature of the photino induces a p-wave threshold for annihilation $\tilde{\gamma}\tilde{\gamma}\to f\bar{f}$ into light fermions, and leads to a rather unexpected form for the bound: for 25 GeV $\lesssim m_{\tilde{f}} \lesssim$ 45 GeV, $(m_{\tilde{\gamma}})_{\min} \simeq m_\tau = 1.8$ GeV; for $m_{\tilde{f}} > 45$ GeV, $(m_{\tilde{\gamma}})_{\min}$ increases approximately *linearly* with $m_{\tilde{f}}$ to a value of 20 GeV when $m_{\tilde{f}} = 100$ GeV.

PACS numbers: 14.80.Pb, 11.30.Pb, 98.80.Bp

The supersymmetric partner of the photon, the photino ($\tilde{\gamma}$), is very possibly the lightest supersymmetric particle. As such, it is stable, and the value of its mass, $m_{\tilde{\gamma}}$, assumes phenomenological importance. Since the photino is massless in a supersymmetric world, one may expect that $m_{\tilde{\gamma}}$ is related to the scale of supersymmetry breaking. For example, in some recently proposed locally supersymmetric grand unified models,[1] the photino is massless at tree level, and $m_{\tilde{\gamma}}$ is radiatively induced[2]:

$$m_{\tilde{\gamma}} \approx C(\alpha/2\pi) m_{\text{gravitino}} \approx \tfrac{1}{100}\, m_{\text{gravitino}} \qquad (1)$$

for $C \approx 10$, where C is an appropriate Casimir number. In another class of models, $m_{\tilde{\gamma}}$ may enter as a parameter in a nonperturbative modification of the gauge kinetic energy, in such a way that both $m_{\text{gravitino}}$ and $m_{\tilde{\gamma}}$ are nonzero in the limit $m_{\text{Pl}} \to \infty$.[3,4] One then obtains relations of the sort[3]

$$m_{\tilde{\gamma}}^2 \propto \sum m_{\text{boson}}^2 - \sum m_{\text{fermion}}^2 - N m_{\text{gravitino}}^2 \qquad (2)$$

(N is a group factor), implying a possibly heavier photino mass. Hence, there is not yet a clear theoretical constraint on $m_{\tilde{\gamma}}$.

As noted by Weinberg,[5] the photino mass should be bounded from below by a cosmological consideration applied previously[6] to neutral lepton masses: namely, the contribution to the present mass density of such particles (those which survive annihilation) is no greater than that needed to close the universe.

It is the purpose of this note to quantify this cosmological result, taking into account the specific nature of photino couplings. The result of interest is displayed in Fig. 1, which shows the dependence of $(m_{\tilde{\gamma}})_{\min}$ on another phenomenologically significant parameter, the (almost) common mass $m_{\tilde{f}}$ of the spin-0 supersymmetric partners of the light fermions. In many models, $m_{\tilde{f}}$ and $m_{\tilde{\gamma}}$ are not independent, and the explication of this relation between them is of interest. The peculiar behavior of the bound can be traced to the presence of a p-wave barrier in the incident channel for the annihilation of cool photinos into the light fermions, while the s-wave annihilation into the heavy fermions (notably τ and c) will lead to sufficient depletion of $\tilde{\gamma}$'s only for $m_{\tilde{f}} < 45$ GeV. I will proceed to the derivation of the result, reserving further comment until later.

The low-energy effective interaction for the annihilations $\tilde{\gamma}\tilde{\gamma}\to f\bar{f}$ (f is a light fermion) is[7] (with λ the Majorana photino field)

$$\mathcal{L}_{\text{eff}} = \sum_f (e^2 Q_f^{\,2}/2m_{\tilde{f}}^2)(\bar{\lambda}\gamma_\mu\gamma_5\lambda)(\bar{f}\gamma^\mu\gamma_5 f), \qquad (3)$$

where $m_{\tilde{f}}$ is the (assumed) common mass of the scalar and pseudoscalar spin-0 $\tilde{f}$ particles and Q_f is the charge of f. The spread in masses among all spin-0 quarks and leptons is severely

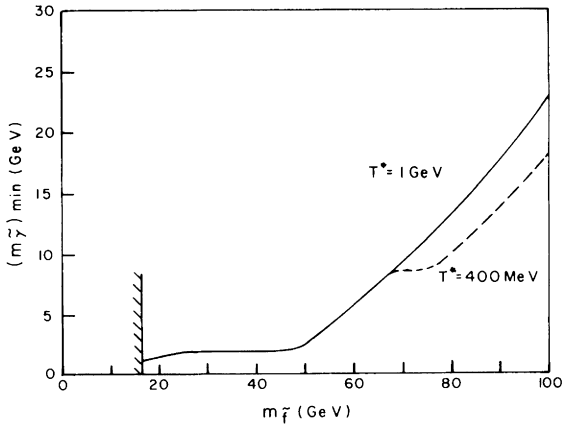


FIG. 1. The behavior of the lower bound on photino mass as a function of $m_{\tilde{f}}$ (spin-0 fermion mass). The region to the left of the crosshatching is disallowed by experiment. The label T^* on the curves denotes the temperature at which the entropy of the quark-gluon plasma attains its massless ideal-gas form.

© 1983 The American Physical Society

constrained by $\Delta Q = 0$ flavor-changing phenomenology,[8] and, in addition, $|m_{\tilde{q}u} - m_{\tilde{l}}|/m_{\tilde{q}u}$ is predicted to be small in supergravity grand unification theories.[1] Hence it is sufficiently accurate to take a common mass $m_{\tilde{f}}$ *independent of* f in Eq. (3).

The annihilation rate at low $\tilde{\gamma}\tilde{\gamma}$ momenta is calculated to be (for Majorana $\tilde{\gamma}$'s)

$$(|\vec{v}_{\rm rel}|\sigma)_{\tilde{\gamma}\tilde{\gamma}} = (e^4/4\pi) m_{\tilde{f}}^{-4} \sum_f Q_f^4 [\tfrac{4}{3}(m_{\tilde{\gamma}}^2 - m_f^2)|\tfrac{1}{2}\vec{v}_{\rm rel}|^2 + m_f^2], \tag{4}$$

where $|\vec{v}_{\rm rel}|$ is the relative velocity of the photinos. The momentum dependence in Eq. (4) can be understood as follows: For massless fermions, only the $J=1$ part of the axial coupling [in Eq. (3)] contributes. The Fermi statistics of the photinos requires that at low energies they be in a spin-0 s wave or spin-1 p wave. For $J=1$, only the latter is possible, giving rise to a p-wave barrier at $\tilde{\gamma}\tilde{\gamma}$ threshold. The factor of $\frac{4}{3}$ comes from averaging over a p-wave $(1+\cos^2\theta)$ factor. The term in m_f^2 is the remnant of the $\tilde{\gamma}\tilde{\gamma}$ s-wave annihilation.

I now replace $v_{\rm rel}^2$ in Eq. (4) by its thermal average $6k_BT/m_{\tilde{\gamma}}$ [$\frac{1}{2}(\frac{1}{2}m_{\tilde{\gamma}})v_{\rm rel}^2 = \frac{3}{2}k_BT$]. At the same time, e^4 is rewritten as $2(4\sin^2\theta_W m_W^2)G_F^2$ in order to bring $v\sigma$ into the standard weak interaction form used in Ref. 6:

$$(v\sigma)_{\tilde{\gamma}\tilde{\gamma}} = (G_F^2/2\pi)(4\sin^2\theta_W)^2 (m_W/m_{\tilde{f}})^4 \sum_f Q_f^4 [2(m_{\tilde{\gamma}}^2 - m_f^2)x + m_f^2], \tag{5}$$

where $x = k_BT/m_{\tilde{\gamma}}$.

Because of the differing x and $m_{\tilde{\gamma}}$ dependence of the cross section from that used in Ref. 6, the results of the exact integration of the rate equation are not directly applicable. However, the analytic approximation obtained in that reference can be used (modified to take into account the $\tilde{\gamma}\tilde{\gamma}$ p-wave factor x), with a result for the relic abundance that is approximately 25% high. As an approximation, I will carry out the analytic approximation and downgrade the resulting abundance by a factor of 0.8.

The reduced abundance $f = n/T^3$ at $T=0$ is found by integrating the rate equation

$$df/dx = \gamma f^2 \tag{6}$$

from the freezeout value of x, $x = x_{\rm fr}$, to $x = 0$. Here

$$\begin{aligned} &\gamma(x) = \tilde{C} m_{\tilde{\gamma}} \mu^2(x), \quad \tilde{C} = N_F^{-1/2} G_F^2 m_{\rm Pl} (45/32\pi^5)^{1/2} (4\sin^2\theta_W)^2 (m_W/m_{\tilde{f}})^4, \\ &\mu^2(x) = \mu_0^2 + 2\mu_1^2 x, \quad \mu_0^2 = \sum_f Q_f^4 m_f^2 \theta(m_{\tilde{\gamma}} - m_f), \quad \mu_1^2 = \sum_f Q_f^4 (m_{\tilde{\gamma}}^2 - m_f^2)\theta(m_{\tilde{\gamma}} - m_f), \end{aligned} \tag{7}$$

and N_F is the effective number of degrees of freedom at freezout.[6] The freezeout temperature is found by setting

$$df_0/dx = \tilde{C} m_{\tilde{\gamma}} \mu^2(x) f_0^2 \tag{8}$$

with f_0 given in Ref. 6. The only difference in procedure is that now the coefficient γ in Eq. (6) is x dependent.

Integrating the rate equation (6) using the linear form (7) for γ, I find

$$f(0) = (\tilde{C} m_{\tilde{\gamma}})^{-1} (\mu_0^2 x_{\rm fr} + \mu_1^2 x_{\rm fr}^2)^{-1}. \tag{9}$$

With the aforementioned reduction of this approximation by 20%, the present matter density due to $\tilde{\gamma}$'s is then (with $m_W = 81$ GeV)

$$\rho_{\tilde{\gamma}} = 0.8 \tfrac{4}{11} T_\gamma^3 m_{\tilde{\gamma}} f(0) \kappa \tag{10a}$$

$$\cong (0.72\times10^{-28}\ {\rm g\ cm^{-3}}) \kappa^{1/2} (\mu_0^2 x_{\rm fr} + \mu_1^2 x_{\rm fr}^2)^{-1} [m_{\tilde{f}}/(100\ {\rm GeV})]^4 \tag{10b}$$

at $T_\gamma = 2.8$ K. (All masses are in gigaelectronvolts.) The factor κ in (10a) adjusts for the inequality $T_{\tilde{\gamma}} \neq T_\gamma$ due to reheating of the photon gas by annihilation of particles when $T < T_{\rm fr}$.[9] $\kappa = 1$ if $m_e \lesssim T_{\rm fr} \lesssim m_{\mu,\pi}$; $\kappa = \frac{43}{69} \cong 0.62$[10] if $m_{\mu,\pi} \lesssim T_{\rm fr} \lesssim T^*$, where T^* is the temperature at which the entropy of the quark-gluon plasma is approximately equal to its massless, ideal-gas value. If $T_{\rm fr} \gtrsim T^*$, then $\kappa = \frac{43}{205} = 0.21$.[10] The value of T^* will be discussed in the conclusion: Some estimates based on finite-temperature QCD will give $T^* \sim 1$ GeV; other heuristic arguments suggest $T^* \sim 400$ MeV. The $\kappa^{1/2}$ in Eq. (10b) is the result of a partial cancellation of κ by $N_F^{1/2}$ ($\propto \kappa^{-1/2}$) as $m_{\tilde{\gamma}}$ crosses thresholds.

The freezeout condition (8) gives $x_{\rm fr} \simeq \frac{1}{17}$ to $\frac{1}{22}$ over a range of $m_{\tilde{\gamma}} = 1.5$ to 25 GeV and $m_{\tilde{f}} = 20$

to 100 GeV (see below for details). Thus to a good approximation

$$\mu_0^2 + \mu_1^2 x_{\rm fr} \simeq \sum_f Q_f^4 (m_f^2 + x_{\rm fr} m_{\tilde\gamma}^2)\,\theta(m_{\tilde\gamma} - m_f). \tag{11}$$

The requirement that $\rho_{\tilde\gamma} \lesssim 10^{-29}$ g cm^{-3} for $T_\gamma = 2.8$ K leads to a condition on the masses:

$$\sum_f Q_f^4 (m_f^2 + x_{\rm fr} m_{\tilde\gamma}^2) \gtrsim 7.2 x_{\rm fr}^{-1} (m_{\tilde f}/100)^4 \kappa^{1/2}. \tag{12}$$

It is convenient to parametrize $x_{\rm fr} = (20\xi)^{-1}$, $0.85 \lesssim \xi \lesssim 1.1$. Then Eq. (12) becomes

$$\left(\sum_f Q_f^4 m_f^2\right) + \tfrac{1}{20}\,\xi^{-1} m_{\tilde\gamma}^2 \left(\sum_f Q_f^4\right) \gtrsim 144\,\xi [m_{\tilde f}/(100\ {\rm GeV})]^4 \kappa^{1/2}\ {\rm GeV}^2. \tag{13}$$

The sum in Eq. (13) is over open channels. As a reasonable approximation, and for clarity of exposition, I take e, μ, u, d, and s fermions as massless, and $m_c \simeq m_\tau = 1.8$ GeV, $m_b \simeq 4.5$ GeV. The assumed degeneracy of c and τ will play a completely negligible role in the results, and the thresholds can be separated if so desired. I then consider the following cases:

(i) $m_{\tilde\gamma} < m_\tau$.—The first term in Eq. (13) is then absent, and we obtain the bound (with $\kappa = 1$)

$$m_{\tilde\gamma} \gtrsim (33\ {\rm GeV})\,\xi [m_{\tilde f}/(100\ {\rm GeV})]^2, \tag{14}$$

which is inconsistent with $m_{\tilde\gamma} < m_\tau$ unless $m_{\tilde f} < 23\,\xi^{-1/2}$ GeV. Since the PETRA lower bound on $m_{\tilde f}$ is about 17 GeV,[11] we find from (14) that $m_{\tilde\gamma} \gtrsim 1.0\,\xi$ GeV. In this mass range ξ turns out to be about 0.86 (see below), so that there is a narrow experimentally allowed range (17 GeV $\lesssim m_{\tilde f} < 25$ GeV) in which $(m_{\tilde\gamma})_{\rm min}$ lies between 0.9 and 1.8 GeV (see Fig. 1).

(ii) $m_{\tilde\gamma} > m_\tau$.—For $m_{\tilde\gamma} < m_b$, Eq. (13) reads

$$\tfrac{43}{27}(1.8)^2 + \tfrac{115}{27}\,\tfrac{1}{20}\,\xi^{-1} m_{\tilde\gamma}^2 > 144\,\xi (m_{\tilde f}/100)^4 \kappa^{1/2}, \tag{15a}$$

and for $m_{\tilde\gamma} > m_b$

$$\tfrac{43}{27}(1.8)^2 + \tfrac{1}{27}(4.5)^2 + \tfrac{116}{27}\,\tfrac{1}{20}\,\xi^{-1} m_{\tilde\gamma}^2 > 144\,\xi (m_{\tilde f}/100)^4 \kappa^{1/2}. \tag{15b}$$

It is easy to see from Eq. (15a) that *if* $m_{\tilde f} \lesssim 44\,\xi^{-1/4}$ *GeV, the lower bound on* $m_{\tilde\gamma}$ *is simply the spectral one* ($m_{\tilde\gamma} \gtrsim m_\tau$). The physical origin of this result is apparent: If $m_{\tilde f} < (44\ {\rm GeV})\,\xi^{-1/4}$, the s-wave $J^P = 0^-$ $\tilde\gamma\tilde\gamma$ annihilation into $\tau\bar\tau + c\bar c$ is large enough to deplete the photino abundance to give a mass density below $\rho_{\rm crit}$. The dependence on m_f^2 (rather than $m_{\tilde\gamma}^2$) of this part of $\langle v\sigma\rangle$ removes any restriction on $m_{\tilde\gamma}$ beyond the spectral one. It is also important to note the insensitivity of this result to ξ, which turns out to be $\simeq 0.92$ in this region of $m_{\tilde f}$.

However, if $m_{\tilde f} \gtrsim 44\,\xi^{-1/4}$ GeV, Eqs. (15a) and (15b) do impose a lower bound on $m_{\tilde\gamma}$ (other than spectral), and this bound grows approximately linearly with $m_{\tilde f}$ until $m_{\tilde f} \simeq 100$ GeV. To obtain the bound on $m_{\tilde\gamma}$ from (15a) and (15b), a value of ξ [$= (20 x_{\rm fr})^{-1}$] as a function of $m_{\tilde f}$ is needed. This is arrived at self-consistently: An initial value of $\xi = 1$ is used in Eq. (15) to generate a set $\{m_{\tilde f}, (m_{\tilde\gamma})_{\rm min}\}$. These are then fed into Eq. (8) to obtain new values of $x_{\rm fr}$ (or ξ) as a function of $m_{\tilde f}$ and $m_{\tilde\gamma}$. The procedure converges very rapidly, and the result is that $x_{\rm fr}$ may be parametrized roughly as

$$x_{\rm fr}^{-1} = 16 + 5.3 m_{\tilde f}/(100\ {\rm GeV}),$$

or

$$\xi = (20 x_{\rm fr})^{-1} = 0.80 + 0.27\, m_{\tilde f}/(100\ {\rm GeV}). \tag{16}$$

The resulting lower bound $(m_{\tilde\gamma})_{\rm min}$ is plotted against $m_{\tilde f}$ in Fig. 1. The behavior outlined in the text is clearly seen. The threshold factor $(1 - m_f^2/m_{\tilde\gamma}^2)^{1/2}$, and the inequality $m_c \neq m_\tau$, both ignored in obtaining Fig. 1, introduce negligible changes in the result. The effect of changing κ from 1 to 0.62 when $x_{\rm fr} m_{\tilde\gamma} > m_\pi$ has been included in both curves. The solid curve assumes that the quark-gluon entropy does not attain its ideal massless value until $T^* = 1$ GeV, whereas the dashed curve assumes that this occurs for $T^* = 400$ MeV.

Several comments are in order:

(1) Clearly the most favorable situation for models in which the photino derives its mass from radiative corrections is that $m_{\tilde f} \lesssim 45$ GeV, in which case $m_{\tilde\gamma}$ can theoretically be as small as ~2 GeV. This situation is also (coincidentally) the most propitious for the detection of double photino production at PETRA.[12]

(2) If $m_{\tilde f} \sim 70$ GeV or larger, then the minimum photino mass is large ($\gtrsim 10$ GeV), and the radiative origin becomes unlikely. Models such as those in Ref. 3 (with all their attendant ambiguity) can accomodate large photino masses.

(3) When $m_{\tilde\gamma} \gtrsim 4$ GeV, the freezeout temperature $T_{\rm fr} \cong \frac{1}{20}\, m_{\tilde\gamma}$ lies above $T_c \cong 200$ MeV, the quark-hadron transition point. It is then appropriate to insert a factor $\kappa = [g(m_e < T < m_{\mu,\pi})/g(T = T_{\rm fr})]^{1/2}$ in Eq. (15b). [Here $g = S/(RT)^3$ is the degeneracy factor for the entropy in a comoving volume R^3 at temperature T.] The question of significance

becomes the following: At what temperature $T^* > T_c$ does g assume its (large) value corresponding to an ideal gas of massless quarks and gluons? An estimate based on the $q\bar{q}$ interaction energy[13] relative to the kinetic energy implies that $T^* \cong 440$ MeV and the dashed curve reflects this number. However, it is crucial to note that there are nonperturbative effects in the quark-gluon plasma which are *much stronger* than the $q\bar{q}$ interaction. Self-mass effects for gluons and quarks are especially important. For example, a massless quark will grow[14] a chirally invariant mass $m_{qu}(T)$ as a result of polarization effects. To lowest order in g, $m_{qu}(T) \cong 1.3gT$.[14] Iteration of this effect leads to the calculation of the entropy of a noninteracting gas of quarks of mass $m_{qu}(T)$; this attains 50% of the massless form only when $m_{qu}(T)/T \lesssim 2$, or $g^2/4\pi \lesssim 0.20$. With $g^2/4\pi = (2\pi/9)\ln(4T/\Lambda_{QCD})$ ($4T \cong$ mean c.m. energy of quark collisions) and $\Lambda_{QCD} \cong 0.2$ GeV, this does not occur until $T \cong 5$ GeV! The thermodynamics is similarly strongly affected by effective mass generation for the gluons. In this context, it has been noted[14,15] that the $O(g^3)$ contribution to the free energy (the Debye-Hückel effect) is comparable to the ideal-gas value for $g^2/4\pi \gtrsim 0.25$, or $T < 0.8$ GeV. The effects of possible magnetic mass growth[14,16] are more speculative. Recent Monte Carlo calculations[17] for SU(2) suggest that the transverse modes of the gluon propagate with an effective mass $m_{mag}(T) \cong 0.25g^2T$. With a modest extrapolation to $m_{mag}(T) = \frac{3}{2}(0.25g^2T)$ for SU(3), the same analysis as above indicates a confluence with massless ideal-gas entropy for $T > 550$ MeV. In view of this discussion one may favor T^* to lie between 0.4 and 1.0 GeV. The two branches in the figure reflect these boundaries.

I would like to thank R. Arnowitt for useful conversations during the course of this work, and L. Krauss for calling to my attention a numerical error in a preliminary version. This research was supported in part by the National Science Foundation.

[1]A. H. Chamseddine, R. Arnowitt, and P. Nath, Phys. Rev. Lett. 49, 970 (1982); L. Ibáñez, Phys. Lett. 118B, 73 (1982).

[2]R. Arnowitt, A. H. Chamseddine, and P. Nath, private communication; L. Alvarez-Gaumé, J. Polchinski, and M. B. Wise, Harvard University Report No. HUTP-82/A063 (to be published).

[3]E. Cremmer, P. Fayet, and L. Girardello, Phys. Lett. 122B, 41 (1983). See also J. Ellis, D. V. Nanopoulos, and K. Tamvakis, Phys. Lett. 121B, 123 (1983), for discussion of this point.

[4]Conditions under which the gauge kinetic energy is expected to be near canonical are considered by S. Weinberg, Phys. Rev. Lett. 50, 387 (1983).

[5]Weinberg, Ref. 4.

[6]B. W. Lee and S. Weinberg, Phys. Rev. Lett. 39, 165 (1977).

[7]P. Fayet, Phys. Lett. 86B, 272 (1979).

[8]J. Ellis and D. Nanopoulos, Phys. Lett. 110B, 44 (1982).

[9]S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1982), p. 536.

[10]G. Steigman, K. A. Olive, and D. N. Schramm, Phys. Rev. Lett. 43, 239 (1979).

[11]R. Brandelik *et al.*, Phys. Lett. 117B, 365 (1982).

[12]P. Fayet, Phys. Lett. 117B, 460 (1982).

[13]R. V. Wagoner and G. Steigman, Phys. Rev. D 20, 825 (1979).

[14]D. J. Gross, R. D. Pisarski, and L. G. Yaffe, Rev. Mod. Phys. 53, 43 (1981).

[15]E. Shuryak, Phys. Rep. 61C, 2 (1980).

[16]A. D. Linde, Rep. Prog. Phys. 42, 389 (1979); T. Appelquist and R. D. Pisarski, Phys. Rev. D 23, 2305 (1981).

[17]A. Billoire, G. Lazarides, and Q. Shafi, Phys. Lett. 103B, 450 (1981); T. A. DeGrand and D. Toussaint, Phys. Rev. D 25, 526 (1982).

Nuclear Physics B238 (1984) 453–476
© North-Holland Publishing Company

SUPERSYMMETRIC RELICS FROM THE BIG BANG*

John ELLIS and J. S. HAGELIN

Stanford Linear Accelerator Center, Stanford University, Stanford, California 94305, USA

D. V. NANOPOULOS, K. OLIVE†, and M. SREDNICKI‡

CERN, CH-1211 Geneva 23, Switzerland

Received 16 September 1983
(Revised 15 December 1983)

We consider the cosmological constraints on supersymmetric theories with a new, stable particle. Circumstantial evidence points to a neutral gauge/Higgs fermion as the best candidate for this particle, and we derive bounds on the parameters in the lagrangian which govern its mass and couplings. One favored possibility is that the lightest neutral supersymmetric particle is predominantly a photino $\tilde{\gamma}$ with mass above $\frac{1}{2}$ GeV, while another is that the lightest neutral supersymmetric particle is a Higgs fermion with mass above 5 GeV or less than O(100) eV. We also point out that a gravitino mass of 10 to 100 GeV implies that the temperature after completion of an inflationary phase cannot be above 10^{14} GeV, and probably not above 3×10^{12} GeV. This imposes constraints on mechanisms for generating the baryon number of the universe.

1. Introduction

In the past few years, supersymmetric extensions of the standard model of particle physics have received a great deal of attention from theorists [1]. All these models predict an exciting variety of new particles. Alas, there are few experimental constraints on their masses and couplings, particularly on those of new neutral particles. To guide future work, both experimental and theoretical, we must learn all we can about the parameters of these theories. Here, as in many other cases, useful information can be obtained from a study of early cosmology. (For a review of this subject, see ref. [2].)

The essential feature of supersymmetric theories which makes them amenable to cosmological study is that, in many models, one of the new particles is absolutely stable. By "new," we mean the superpartners of any ordinary particles: gauge

* Work supported by the Department of Energy, contract DE-AC03-76SF00515.
† Address as of July 1, 1983: Fermilab, PO Box 500, Batavia, IL 60510.
‡ Address as of Sept. 1, 1983: Dept. of Physics, University of California, Santa Barbara, CA 93106.

fermions (gluino, photino, wino, etc.), Higgs fermions (shiggs or higgsino), scalar quarks and leptons (squarks and sleptons), and the gravitino. In the very early universe, all these particles would be present in thermal equilibrium. As the temperature falls, the heavier new particles decay into the lighter ones. Eventually, only the lightest supersymmetric particle (LSP) will be left. It can disappear only by pair annihilation. We must require that this pair annihilation is efficient enough to reduce the present day number density of the LSP to an acceptable level.

In this paper we give a detailed analysis of this question for the class of models which we favor: a minimal supersymmetric extension of the standard model, with supersymmetry broken via gravitational couplings to spontaneously broken supergravity [3]. This results in mass terms for the new particles and cubic scalar couplings which violate supersymmetry [4]. The mass scale which enters is the gravitino mass $m_{3/2}$, which we take to be O(10 to 100) GeV. Some relations among the masses of the new particles can be obtained by assuming that this theory is the low-energy sector of a supersymmetric grand unified theory [5]. We will make this assumption where necessary but many of our considerations are more general than the specific supersymmetric model which we emphasize.

In sect. 2 we discuss each of the new particles, and whether or not it could be the LSP. We conclude that by far the best candidate for the LSP is a neutral gauge/Higgs fermion (in general, these particles mix so that mass eigenstates contain both gauge fermion and Higgs fermion components).

In sect. 3 we present a detailed analysis of the annihilation rates of such particles so as to provide bounds on the parameters in the lagrangian. Some preliminary results of this analysis have already been given in ref. [6]. One favoured possibility is that the LSP is mostly a photino, and has a mass $\geqslant$ O($\frac{1}{2}$) GeV. Alternatively, the LSP may be mostly a shiggs which either has mass $\gtrsim m_b$ or m_t, or else weighs less than O(100) eV. The simplest models with such a light shiggs also contain a light axion which would need to be exorcised in a completely realistic model.

2. Which is the LSP?

Many LSP candidates can be ruled out immediately by simple arguments. Let us consider each new particle in turn, and whether or not it could be the LSP.

Charged, uncolored particles: winos, charged higgsinos, charged sleptons (selectron, smuon, stau). The possibility of a stable, charged particle was studied by Wolfram [7], who concluded that there should be a number density today of about

$$n \simeq 10^{-6}\left(\frac{M}{1\ \text{GeV}}\right) n_B, \tag{2.1}$$

where M is the mass of the charged particle, and n_B is the baryon-number density today. These particles would be well mixed with ordinary matter; their electromag-

netic interactions ensure that there is no way for them to separate from the matter which forms galaxies, stars, and planets. In particular, we should find such particles in terrestrial searches for anomalously heavy protons [7]. The results of ref. [8] tell us that for $M < 350$ GeV, there are no anomalous protons at a level $< 10^{-21} n_B$, well below that of eq. (2.1). Thus, a stable charged particle with $M < 350$ GeV can be excluded.

Gluino. A gluino should form new hadrons with strong interactions; at least one of these new hadrons must be stable if the gluino is the LSP. Calculations of the relic number density of heavy hadrons [7,9] yield

$$n \simeq 10^{-10} n_B , \tag{2.2}$$

independent of the new hadron's mass. If there is a stable, charge one hadron, its existence can be ruled out by the anomalous proton search [8] discussed above. If the only stable new hadrons are neutral, this argument does not apply. We expect that these hadrons are today bound to ordinary nuclei, but it is not clear which nuclei. Surely the precise nuclear interaction rates inside stars which determine the present day abundance of elements are different for nuclei with new hadrons stuck to them. Thus, although we know [10] that there are no anomalous isotopes of oxygen with mass less than 44 GeV at a level $n \simeq 10^{-17} n_B$, it is not entirely clear that this means that a 20 GeV gluino could not be the LSP. We do at least know from accelerator experiments [11] that the gluino mass exceeds 2 GeV.

In the absence of firm cosmological evidence against the gluino as the LSP, we turn to theoretical arguments. Neglecting mixing with Higgs fermions, the gluino-photino mass ratio is given by [5]

$$m_{\tilde{g}}/m_{\tilde{\gamma}} = \frac{3}{8} \frac{\alpha_3}{\alpha_2 \sin^2 \theta_w} , \tag{2.3}$$

to leading order in the renormalization group equations in any grand unified theory. Even after including arbitrary photino/higgsino mixing, the gluino is always heavier. Thus, in any susy GUT, the gluino is not the LSP.

Squarks. The same discussion is that for gluinos applies to searches for squarks in anomalous nuclei: they are inconclusive. Accelerator searches [12] tell us $M \gtrsim 17$ GeV.

Stable squarks might carry some of the baryon number of the universe, in which case they could be present at densities well above $10^{-10} n_B$. The reason is that, in GUTs, the cosmological baryon asymmetry (CBA) is produced at a temperature of at least 10^{10} GeV, when there was no distinction between squarks and quarks. At this time, half the CBA was stored in squarks. To decide if it is still there today, we must compute the reaction rate for the process squark + squark → quark + quark. The dominant contribution is from gluino exchange, which gives a cross section

times relative velocity of order

$$\langle \sigma v_{\rm rel} \rangle \simeq \alpha_3^2 / m_{\tilde{g}} m_s , \tag{2.4}$$

where m_s is the squark mass. Assuming that both $m_{\tilde{g}}$ and m_s are no larger than a few hundred GeV, standard estimates [7,9] tell us that the annihilation rate is large enough to reduce the excess squark density by several orders of magnitude. Thus the CBA flows from the squarks to the quarks, and the resulting relic squark density is acceptable.

Again we must turn to theoretical arguments. Models almost always predict [5] that squarks are heavier than sleptons, due to renormalization effects. Although this argument is less airtight than in the gluino case, we think it unlikely that a squark would be the LSP, and so we will not pursue this possibility.

Scalar neutrinos. There is no [13] cosmological argument against a sneutrino as the LSP.

Theoretically, we expect the sneutrino to be nearly degenerate with the scalar partner of the left-handed charged lepton in the same SU(2) weak doublet, but with the mass eigenvalues of the charged slepton $\tilde{\ell}$ split on either side of the sneutrino $\tilde{\nu}$ mass by an amount proportional to the mass of the charged lepton:

$$m_{\tilde{\ell}} = m_{\tilde{\nu}} \pm a m_{\ell} , \tag{2.5}$$

where a is a model-dependent constant. Thus, for example, we expect that there is a charged stau with a mass O(2) GeV below that of the tau sneutrino (and, indeed, all sneutrinos). This pattern of masses might be rearranged by electromagnetic radiative corrections in some models, so that it may be possible to arrange a sneutrino as the LSP. This possibility is discussed in ref. [13]; we will disregard it here.

Gravitino. The gravitino is a special case. Even if it is not the LSP, it decays so slowly that the LSPs in its decay products provide limits on the initial density of gravitinos [14].

As discussed by Weinberg [14], a 10 to 100 GeV gravitino cannot be the LSP – it annihilates much too slowly, with a cross section of order $1/M_{\rm p}^2$, where $M_{\rm p} = 1.2 \times 10^{19}$ GeV. A 10 to 100 GeV gravitino which is not the LSP would decay after nucleosynthesis; each gravitino would yield one LSP, but it would be much too cold for these LSPs to annihilate. The only apparent solution is inflation – this would remove the original number density of gravitinos (and all other particles); the weak coupling of gravitinos can prevent regenerating a large number density of them after inflation [15].

Weinberg has recently pointed out [16] that a process like photon + anything → gravitino + anything has a cross section of $\alpha N / M_{\rm p}^2$, where N is the number of degrees of freedom. In thermal equilibrium, this means a production rate Γ of order [17]

$$\Gamma \simeq \alpha N T_{\rm R}^3 / M_{\rm p}^2 , \tag{2.6}$$

at a temperature T_R, the temperature of the universe once it reheated after inflation. Γ is to be compared with the Hubble parameter at that time

$$H \simeq \sqrt{N}\, T_R^2/M_p\,, \tag{2.7}$$

in order to get a relative gravitino abundance

$$Y_{3/2} \simeq \Gamma/H \tag{2.8}$$

$$\simeq \alpha\sqrt{N}\, T_R/M_p\,. \tag{2.9}$$

Subsequent to the establishment of this primordial abundance, the photons in the microwave background reheated as the N degrees of freedom present at T_R annihilated. So, long after T_R but just before gravitino decay, the relative number density of gravitinos was depressed to

$$\frac{n_{3/2}}{n_\gamma} \simeq \frac{2Y_{3/2}}{N} \tag{2.10}$$

$$\simeq 2\alpha T_R/\sqrt{N}\, M_p\,. \tag{2.11}$$

Each of these gravitinos yield an LSP as it decays and these LSPs would still be around today. In order that they not dominate the mass of the universe, we must require (see also ref. [18])

$$Mn_{3/2} < \rho\,, \tag{2.12}$$

where ρ is the upper limit on the mass density of the universe. This cannot be used to bound M, since we do not know T_R. But using the bounds on M from sect. 3, we can get an upper limit on T_R:

$$T_R \lesssim \frac{\sqrt{N}\,\rho M_p}{2\alpha M n_\gamma}\,. \tag{2.13}$$

Using $N = 250$, $\alpha = \frac{1}{25}$, $n_\gamma = 400\ /\mathrm{cm}^3$, and the very conservative numbers (see sect. 3) $\rho = 2\times 10^{-29}\ \mathrm{g/cm}^3$, $M = \frac{1}{2}$ GeV, we get

$$T_R \lesssim 10^{14}\ \mathrm{GeV}\,, \tag{2.14}$$

while the more realistic numbers $\rho = 2\times 10^{-30}\ \mathrm{g/cm}^3$ and $M = 2$ GeV give

$$T_R \lesssim 3\times 10^{12}\ \mathrm{GeV}\,. \tag{2.15}$$

Thus, an inflationary solution to the gravitino problem cannot have a very high

reheating temperature*. Since the cosmological baryon asymmetry must be produced after inflation this constraint necessitates the existence of baryon-number violating particles with masses no larger than 10^{14} GeV, and more likely less than 10^{13} GeV.

Neutral gauge and Higgs fermions. These are the only remaining LSP candidates. There are no arguments – cosmological, experimental, or theoretical – against one of these particles being the LSP, and this is the case on which we will concentrate in sect. 3.

3. Neutral gauge and Higgs fermions

3.1. MASSES AND MIXING

Here we consider the likely possibility that the lightest susy particle is a colorless, neutral fermion. We start by reviewing the structure of the charged and neutral susy fermion mass matrices [6]. We consider a minimal susy model with two light doublets of Higgs chiral superfields H_1 and H_2 of weak hypercharge ± 1 respectively. The mass matrices for the charged and neutral susy fermions – gauginos and higgsinos – are determined by the lagrangian terms

$$L \ni \varepsilon\varepsilon_{ij}\tilde{H}_1^i\tilde{H}_2^j - M_2\tilde{W}_a\tilde{W}_a - M_1\tilde{B}\tilde{B}, \tag{3.1}$$

where W_a and B denote SU(2) and U(1) gauge superfields respectively, the tildes denote fermionic components and i, j (a) are doublet (triplet) SU(2) indices. We generally expect the parameters M_1, M_2 to be $O(M_W)$, while it is conceivable that ε might be very much smaller. We shall assume

$$M_1 = \frac{5}{3}\frac{\alpha_1}{\alpha_2}M_2, \tag{3.2}$$

where $\alpha_i = g_1^2/4\pi$, $i = 1,2,3$ are the gauge coupling constants, which holds to leading order in the renormalization group equations if $SU(2)\times U(1)$ is eventually embedded in a unifying non-abelian group. When combined with the conventional Higgs-gauge field couplings the full mass matrix for the left-handed charged fermion fields becomes

$$(\tilde{W}^+, \quad \tilde{H}_1^+)\begin{pmatrix} M_2 & g_2 v_2 \\ g_2 v_1 & -\varepsilon \end{pmatrix}\begin{pmatrix} \tilde{W}^- \\ \tilde{H}_2^- \end{pmatrix}, \tag{3.3}$$

* Eq. (2.13) does not constrain T_R if the LSP is a 100 eV higgsino (see subsect. 3.4). In this case, we can still derive a similar bound on T_R by requiring that the entropy produced by gravitino decay be acceptably small [17].

where $\langle 0|H^0_{1,2}|0\rangle = v_{1,2}$; $m_W^2 = \frac{1}{2}g_2^2(v_1^2 + v_2^2)$. This matrix is diagonalized by rotations through angles $\theta_\pm$ among the positively and negatively charged fields respectively, where

$$\tan\theta_\pm = \frac{b_\pm + \sqrt{b_\pm^2 + 4a_\pm}}{2a_\pm}, \tag{3.4a}$$

with

$$a_\pm = \begin{cases} M_2 g_2 v_1 - \varepsilon g_2 v_2 \\ M_2 g_2 v_2 - \varepsilon g_2 v_1 \end{cases}, \qquad b_\pm = M_2^2 - \varepsilon^2 \pm g_2^2(v_2^2 - v_1^2). \tag{3.4b}$$

The charged fermion masses are

$$\begin{aligned} m_1 &= M_2\cos\theta_+\cos\theta_- - g_2 v_2\cos\theta_+\sin\theta_- - g_2 v_1\sin\theta_+\cos\theta_- - \varepsilon\sin\theta_+\sin\theta_-, \\ m_2 &= M_2\sin\theta_+\sin\theta_- + g_2 v_2\sin\theta_+\cos\theta_- + g_2 v_1\cos\theta_+\sin\theta_- - \varepsilon\cos\theta_+\cos\theta_-. \end{aligned} \tag{3.5}$$

Fig. 1 shows mass contours for the lightest charged fermion $\chi^\pm$ for certain ranges of the parameters (M_2, ε) and representative values of v_1/v_2. Since H_1 gives masses to the charge-$\frac{2}{3}$ quarks and $m_c \gg m_s$, $m_t \gg m_b$, it seems plausible that $v_1 \geqslant v_2$. In figs. 1a, c, e, g (1b, d, f, h) v_1/v_2 takes the values 1, 2, 4, 8 while the sign of ε is taken positive (negative). We see that in much of this parameter space there is a light $\chi^\pm$ with $m_{\chi^\pm} < m_W$ and often even $m_{\chi^\pm} < 20$ GeV. The $m_{\chi^\pm} = 20$ GeV contour serves as a constraint on the parameters M_2 and ε since no new charged fermion has been seen in e^+e^- collisions with a mass less than about 20 GeV. We shall take this constraint seriously, although it has been argued [19] that a charged susy fermion with a mass somewhat lighter than 20 GeV might still have escaped detection.

In the neutral sector there are four susy fermions which mix, namely the $\tilde{W}^3$, $\tilde{B}^0$, $\tilde{H}_1^0$ and $\tilde{H}_2^0$. Their mixing matrix is [6]

$$\left(\tilde{W}^3, \tilde{B}^0, \tilde{H}_1^0, \tilde{H}_2^0\right) \begin{pmatrix} M_2 & 0 & -\sqrt{\tfrac{1}{2}}\, g_2 v_1 & \sqrt{\tfrac{1}{2}}\, g_2 v_2 \\ 0 & \frac{5}{3}\frac{\alpha_1}{\alpha_2} M_2 & \sqrt{\tfrac{1}{2}}\, g_1 v_1 & -\sqrt{\tfrac{1}{2}}\, g_1 v_2 \\ -\sqrt{\tfrac{1}{2}}\, g_2 v_1 & \sqrt{\tfrac{1}{2}}\, g_1 v_1 & 0 & \varepsilon \\ \sqrt{\tfrac{1}{2}}\, g_2 v_2 & -\sqrt{\tfrac{1}{2}}\, g_1 v_2 & \varepsilon & 0 \end{pmatrix} \begin{pmatrix} \tilde{W}^3 \\ \tilde{B}^0 \\ \tilde{H}_1^0 \\ \tilde{H}_2^0 \end{pmatrix}, \tag{3.6}$$

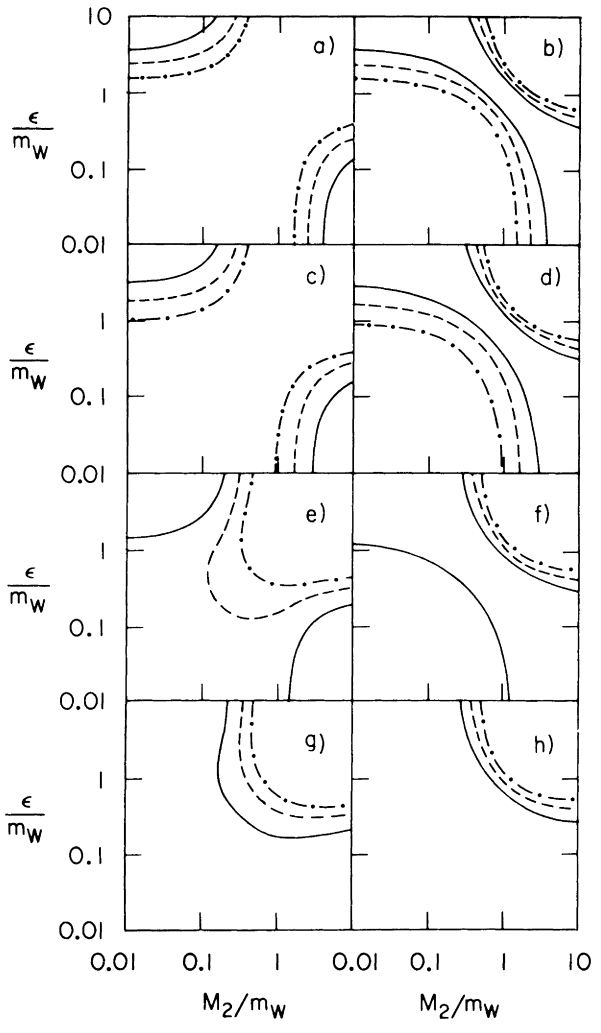


Fig. 1. Mass contours (solid = 20 GeV, dashed = 30 GeV, dot-dashed = 40 GeV) as functions of $|\varepsilon|$ and M_2 for the lightest eigenstate $\chi^{\pm}$ of the charged $\tilde{W}$, $\tilde{H}$ mass matrix (3.3), for different choices of v_1/v_2 and the sign of ε: (a) $v_1 = v_2$, $\varepsilon > 0$; (b) $v_1 = v_2$, $\varepsilon < 0$; (c) $v_1 = 2v_2$, $\varepsilon > 0$; (d) $v_1 = 2v_2$, $\varepsilon < 0$; (e) $v_1 = 4v_2$, $\varepsilon > 0$; (f) $v_1 = 4v_2$, $\varepsilon < 0$; (g) $v_1 = 8v_2$, $\varepsilon > 0$; (h) $v_1 = 8v_2$, $\varepsilon < 0$.

which in terms of the convenient linear combinations

$$\tilde{A}^0 \equiv \frac{v_1\tilde{H}_1^0 - v_2\tilde{H}_2^0}{v}, \qquad \tilde{S}^0 \equiv \frac{v_2\tilde{H}_1^0 + v_1\tilde{H}_2^0}{v}, \tag{3.7}$$

(where we have introduced $v \equiv \sqrt{v_1^2 + v_2^2}$) becomes

$$(\tilde{W}^3, \tilde{B}^0, \tilde{A}^0, \tilde{S}^0)\begin{pmatrix} M_2 & 0 & -\sqrt{\tfrac{1}{2}}\, g_2 v & 0 \\ 0 & \frac{5}{3}\frac{\alpha_1}{\alpha_2}M_2 & \sqrt{\tfrac{1}{2}}\, g_1 v & 0 \\ -\sqrt{\tfrac{1}{2}}\, g_2 v & \sqrt{\tfrac{1}{2}}\, g_1 v & \frac{-2v_1v_2}{v^2}\varepsilon & \frac{v_1^2 - v_2^2}{v^2}\varepsilon \\ 0 & 0 & \frac{v_1^2 - v_2^2}{v^2}\varepsilon & \frac{2v_1v_2}{v^2}\varepsilon \end{pmatrix}\begin{pmatrix} \tilde{W}^3 \\ \tilde{B}^0 \\ \tilde{A}^0 \\ \tilde{S}^0 \end{pmatrix}. \tag{3.8}$$

The general form of the orthogonal rotation that diagonalizes the matrix (3.6) is quite complicated and the Majorana mass eigenstate fields χ_i are in general complicated combinations of $\tilde{W}^3$, $\tilde{B}^0$, $\tilde{H}_1^0$ and $\tilde{H}_2^0$:

$$\chi_i \equiv \left(\alpha_i\tilde{W}^3 + \beta_i\tilde{B}^0 + \gamma_i\tilde{H}_1^0 + \delta_i\tilde{H}_2^0\right). \tag{3.9}$$

In the limit $\varepsilon \to 0$ there is a light higgsino state $\tilde{S}^0$ (as defined in eq. (3.7)) with mass equal to $2v_1v_2\varepsilon/v^2$. If this higgsino were extremely light with $m_{\tilde{S}_0} \leqslant$ O(100) eV*, its cosmology would be similar to that of a conventional neutrino, as discussed in subsect. 3.4. In the limit of $M_2 \to 0$ there is a light photino eigenstate:

$$\tilde{\gamma} \equiv \frac{g_1\tilde{W}^3 + g_2\tilde{B}^0}{\sqrt{g_1^2 + g_2^2}}, \qquad m_{\tilde{\gamma}} = \frac{8}{3}\frac{g_1^2}{g_1^2 + g_2^2}M_2. \tag{3.10}$$

From a purely cosmological standpoint, the photino like the higgsino could have a mass $\lesssim$ O(100) eV. However, in any conceivable model the photino and gluino masses are constrained to have roughly the same order of magnitude [20]:

$$\frac{m_{\tilde{g}}}{m_{\tilde{\gamma}}} = \frac{3}{8}\frac{\alpha_3}{\alpha_2 \sin^2\theta_W}, \tag{3.11}$$

if SU(3) × SU(2) × U(1) is embedded in a unifying non-abelian group, and the gluino is already known [11] to have a mass $\gtrsim 2$ GeV. Thus we discard the

* Simple models (EHNT, ref. [5]) with such a light higgsino also contain an unacceptable axion whose mass is around the geometric mean of ε and the gravitino mass. This may be exorcised in more complicated models whose Higgs boson masses are substantially increased relative to higgsino masses.

possibility of a photino in the eV range, and then we will see in subsect. 3.3 that cosmology imposes a lower bound on $m_{\tilde{\gamma}}$ of about $\frac{1}{2}$ GeV. In the limit where M_2 and ε are both small, the remaining two mass eigenstates are

$$\tilde{Z}^0_{\pm} \equiv \frac{g_1\tilde{B}^0 - g_2\tilde{W}^3 \pm \sqrt{g_1^2+g_2^2}\,\tilde{A}^0}{\sqrt{2(g_1^2+g_2^2)}}, \qquad m_{\tilde{Z}^0_{\pm}} \simeq m_Z = \sqrt{\tfrac{1}{2}(g_1^2+g_2^2)}\,v. \quad (3.12)$$

Mass contours for the lightest and second lightest neutral mass eigenstates χ and χ' for general M_2 and ε are plotted in figs. 2 and 3 respectively, with the same notation

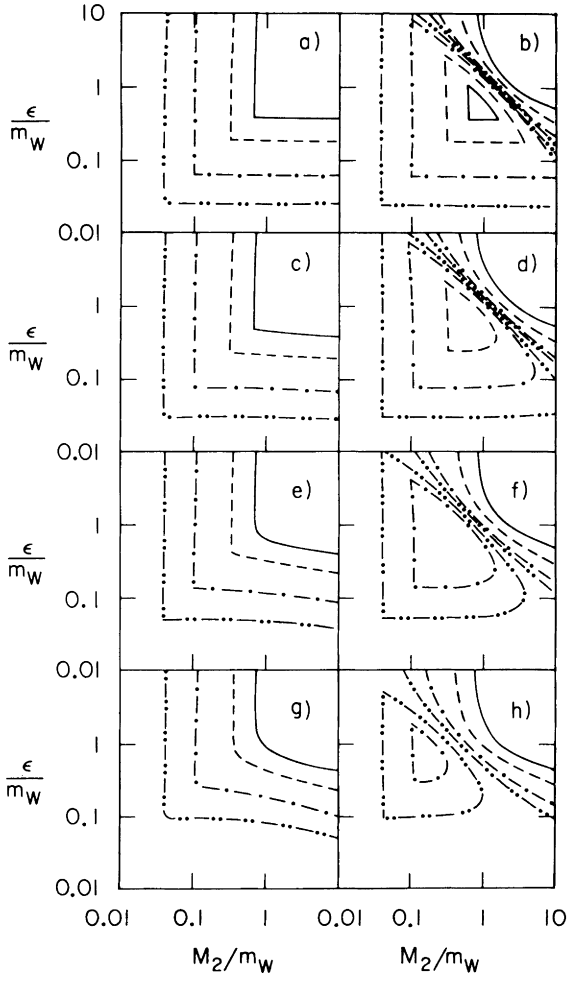


Fig. 2. Mass contours (solid = 30 GeV, dashed = 15 GeV, dot-dashed = 5 GeV, double-dot-dashed = 2 GeV) as functions of $|\varepsilon|$ and M_2 for the lightest eigenstate χ of the neutral $\tilde{W}$, $\tilde{H}$ mass matrix (3.6), for the same choices of v_1/v_2 and of the sign of ε as in fig. 1.

and conventions as in fig. 1. We observe that for most of parameter space there is one and often two eigenstates with mass < 30 GeV, except where M_2 and ϵ are both $\gtrsim m_W$.

3.2. COSMOLOGICAL MASS DENSITIES

Our primary task is to constrain the parameters M_2 and ϵ in the lagrangian (3.1) by requiring that the lightest neutral susy fermion χ have an acceptably low cosmological mass density ρ_χ. We know from the rate of expansion of the universe that $\rho_\chi \leqslant 2 \times 10^{-29}(\Omega h_0^2)$ g/cm^3, where Ω is the density in units of the closure

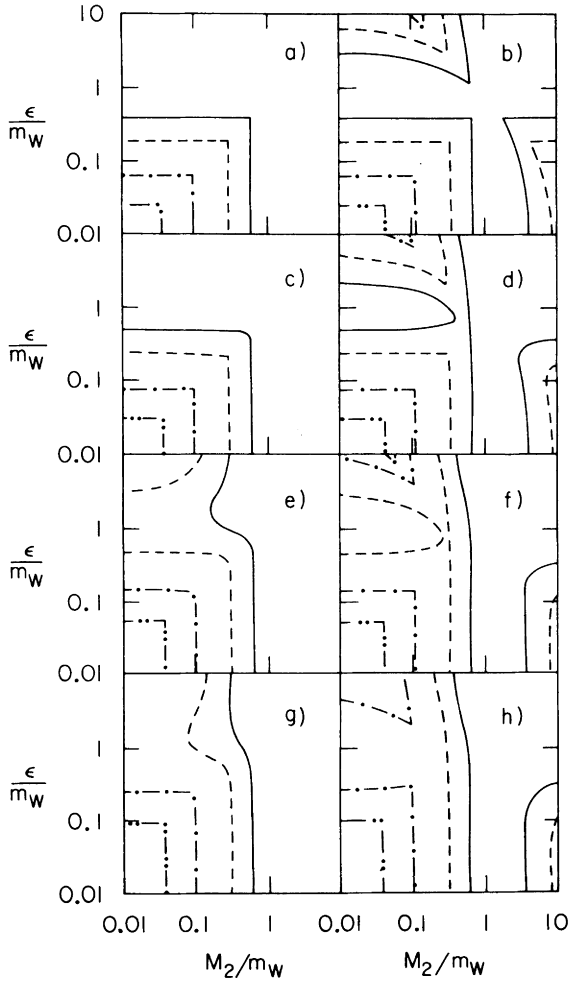


Fig. 3. Mass contours (same as fig. 2) as functions of $|\epsilon|$ and M_2 for the next-to-lightest eigenstate in the mass matrix (3.6), for the same choices of v_1/v_2 and of the sign of ϵ as in fig. 1.

density, and h_0 is the Hubble parameter in units of 100 $\mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$. It is reasonable to believe that $\Omega h_0^2 \lesssim 1$ implying that $\rho_\chi \lesssim 2 \times 10^{-29}$ g/cm^3. A more restrictive constraint follows from the plausible assumption that a non-relativistic susy fermion would participate in galaxy formation, in which case the limits on "dark matter" in galaxies [21] allow one to deduce that $\rho_\chi \leqslant 2 \times 10^{-30}(\Omega h_0^2)$ g/cm^3. We will present our cosmological bounds on the susy parameters M_2 and ε and the susy particle masses for both the conservative constraint $\rho_\chi \leqslant 2 \times 10^{-29}$ g/cm^3 from the observed expansion rate of the universe, and also the plausible bound $\rho_\chi \leqslant 2 \times 10^{-30}$ g/cm^3 based on the assumption of galactic clustering. Since all but the lightest mass eigenstate can decay, these constraints pertain to the lightest eigenstate only: either χ annihilates fairly efficiently as the universe cools, reducing the number density to an acceptably low level, or else χ is very light ($<$ O(100) eV) and is as cosmologically innocuous as a conventional neutrino.

To proceed quantitatively we review and extend the analysis of refs. [22,23] to include the general case of Majorana fermions. The rate equation for ordinary Dirac fermions [22] is

$$\frac{\mathrm{d}n}{\mathrm{d}t} = -3\frac{\dot{R}}{R}n - \langle \sigma v_{\mathrm{rel}} \rangle \left(n^2 - n_0^2\right), \tag{3.13}$$

where n is the actual number density at time t, n_0 is the number density of heavy fermions in thermal equilibrium, R is the cosmic scale factor, σ is the annihilation cross section, and v_{rel} is the relative velocity of the annihilating particles. For identical fermions, we might expect to see a factor of two multiplying the annihilation rate in (3.13), since two Majorana neutrals are destroyed by each annihilation event. However, there is also a factor of $\frac{1}{2}$ needed to prevent double counting in the thermal average over the initial velocity distribution of the annihilating particles, and thus (3.13) is valid for Majorana as well as for Dirac fermions.

Following standard methods [22] one can rewrite (3.13) in the convenient form

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \frac{m_\chi}{k^3}\left(\frac{8\pi^3 N_{\mathrm{F}} G}{45}\right)^{-1/2} \langle \sigma v_{\mathrm{rel}} \rangle \left(f^2 - f_0^2\right), \tag{3.13'}$$

where $x \equiv kT/m_\chi$, $f(x) = n/T^3$, $f^0(x) = n_0/T^3$, k is the Boltzmann constant, $G = 1/m_{\mathrm{P}}^2$ is Newton's constant and N_{F} counts the effective number of degrees of freedom at a given temperature [22].

The annihilation $\chi\chi \to \bar{\mathrm{f}}\mathrm{f}$ occurs through conventional Z-boson exchange and through the exchange of scalar partners of conventional fermions $\tilde{\mathrm{f}}_{\mathrm{L,R}}$. These interactions give rise to a low-energy effective lagrangian

$$\mathcal{L} = \sum_{\mathrm{f}} \bar{\chi}\gamma^\mu\gamma_5\chi\bar{f}\gamma_\mu\left(A_{\mathrm{f}}P_{\mathrm{L}} + B_{\mathrm{f}}P_{\mathrm{R}}\right)f, \tag{3.14}$$

where

$$A_f^Z = (\gamma^2 - \delta^2)\frac{g_1 \sin\theta_W + g_2 \cos\theta_W}{4M_Z^2}\left(\frac{1}{2} Y_{f_L} g_1 \sin\theta_W - T_{f_L}^3 g_2 \cos\theta_W\right), \quad (3.15a)$$

$$B_f^Z = (\gamma^2 - \delta^2)\frac{g_1 \sin\theta_W + g_2 \cos\theta_W}{8M_Z^2} Y_{f_R} g_1 \sin\theta_W, \quad (3.15b)$$

$$A_f^{\tilde{f}} = \frac{\left(T_{f_L}^3 \alpha g_2 + \frac{1}{2} Y_{f_L} \beta g_1\right)^2}{2m_{\tilde{f}_L}^2} - \begin{cases} \gamma^2 m_f^2 / 4v_1^2 m_{\tilde{f}_R}^2, & u, c, t \\ \delta^2 m_f^2 / 4v_2^2 m_{\tilde{f}_R}^2, & e, \mu, \tau, d, s, b, \end{cases} \quad (3.15c)$$

$$B_f^{\tilde{f}} = -\frac{\left(\frac{1}{2} Y_{f_R} \beta g_1\right)^2}{2m_{\tilde{f}_R}^2} + \begin{cases} \gamma^2 m_f^2 / 4v_1^2 m_{\tilde{f}_L}^2, & u, c, t \\ \delta^2 m_f^2 / 4v_2^2 m_{\tilde{f}_L}^2, & e, \mu, \tau, d, s, b, \end{cases} \quad (3.15d)$$

with $\frac{1}{2} Y = Q + T^3$ and $\alpha, \beta, \gamma, \delta$ defined as in (3.9). The sfermion exchange terms in (3.15c, d) proportional to fermion masses result from higgsino-fermion Yukawa couplings, which can be important if the χ is a nearly pure higgsino or when heavy quark modes are kinematically accessible*.

In the non-relativistic limit $v_{rel} \ll 1$, the effective interaction (3.14) leads to the $\chi\chi$ annihilation cross section

$$\langle \sigma v_{rel} \rangle = \frac{1}{2\pi} \sum_f \frac{p}{m_\chi} \left\{ \left(A_f^2 + B_f^2\right)\left[\frac{1}{6}\left(4m_\chi^2 - m_f^2\right) v_{rel}^2 + m_f^2\right] - 2A_f B_f m_f^2 \left(1 - v_{rel}^2/2\right) \right\}, \quad (3.16)$$

where $p = \sqrt{m_\chi^2 - m_f^2}$ is a final state 3-momentum. The v_{rel} dependence of (3.16) results from the P-wave barrier to the annihilation of Majorana fermions, as first emphasized by Goldberg [23]. Replacing v_{rel} in (3.16) by its thermal average ($v_{rel}^2 \simeq 6kT/m_\chi = 6x$), the cross section can be written

$$\langle \sigma v_{rel} \rangle = \tilde{a} + \tilde{b}x, \quad (3.17a)$$

where

$$\tilde{a} = \sum_f \theta(m_\chi - m_f) \frac{1}{2\pi} \frac{p}{m_\chi} m_f^2 (A_f - B_f)^2,$$

$$\tilde{b} = \sum_f \theta(m_\chi - m_f) \frac{1}{2\pi} \frac{p}{m_\chi} \left[\left(A_f^2 + B_f^2\right)\left(4m_\chi^2 - m_f^2\right) + 6A_f B_f m_f^2\right]. \quad (3.17b)$$

* There are in addition higgsino-gaugino interference terms which cannot be cast into an effective interaction of the form (3.14). We do not present these since there is no domain for which these interference contributions equal or exceed the gaugino-gaugino and/or higgsino-higgsino contributions (3.15).

The rate equation (3.13) takes the general form

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \left[\frac{m_\chi}{k^3}\left(\frac{8\pi^3 N_F G}{45}\right)^{-1/2}\right](\tilde{a}+\tilde{b}x)(f^2-f_0^2)$$

$$\equiv (a+bx)(f^2-f_0^2). \tag{3.18}$$

Following the approximate analytic approach of Lee and Weinberg [22], we expect the reduced number density $f(x)$ in (3.18) to remain approximately equal to its equilibrium value $f_0(x)$ until the temperature T drops to a freeze-out value T_f where the annihilation rate is equal to the rate of change in f_0:

$$\frac{\mathrm{d}f_0}{\mathrm{d}x} = (a+bx)f_0^2 \qquad \text{at} \qquad x = x_f, \tag{3.19}$$

and assume that thereafter it evolves approximately according to the equation

$$\frac{\mathrm{d}f}{\mathrm{d}x} = (a+bx)f^2, \tag{3.20}$$

subject to the initial condition $f(x_f)=f_0(x_f)$. Since x_f is generally $\ll 1$, we can use the non-relativistic approximation [22]

$$f_0(x) = 2k^3(2\pi x)^{-3/2}\mathrm{e}^{-1/x}, \tag{3.21}$$

to solve for the freeze-out temperature. Eqs. (3.19) and (3.21) together give

$$x_f = \frac{1}{\ln\left(a x_f^{1/2} + b x_f^{3/2}\right) + \ln\left[2k^3/(2\pi)^{3/2}\right]} \tag{3.22a}$$

$$= \frac{1}{\ln\left(\tilde{a} x_f^{1/2} + \tilde{b} x_f^{3/2}\right) - \frac{1}{2}\ln\left(16\pi^6 N_F G/45 m_\chi^2\right)}. \tag{3.22b}$$

We should emphasize that the non-relativistic approximation (3.16)–(3.22) is not valid when the annihilation cross section (3.16) is small enough for the scaled freeze-out temperature (3.22) to become $\gtrsim 1$, which is generally the case for a light higgsino state. We consider the case of a light higgsino separately in subsect. 3.4.

The present number density is obtained by integrating (3.20) from $x = x_f$ down to $x = 0$:

$$f(0) = \frac{1}{a x_f + \frac{1}{2} b x_f^2}. \tag{3.23}$$

The present mass density ρ_χ is given by

$$\rho_\chi = 0.8\left(\frac{T_\chi}{T_\gamma}\right)^3 T_\gamma^3 m_\chi \frac{1}{ax_f + \frac{1}{2}bx_f^2}, \tag{3.24a}$$

where $(T_\chi/T_\gamma)^3$ accounts for the subsequent reheating of the photon temperature with respect to the temperature of χ, due to the annihilation of particles with $m < x_f m_\chi$, and is tabulated [24] together with N_F in table 1. The fudge factor 0.8 is included to correct for the fact [22] that the analytic approximation (3.19) and (3.20) to the full rate equation (3.13) give a result which is approximately 25% too large. In terms of the coefficients $\tilde{a}$, $\tilde{b}$ appearing in the annihilation cross section, the mass density (3.24a) reads

$$\rho_\chi = 4.0 \times 10^{-40}\left(\frac{T_\chi}{T_\gamma}\right)^3\left(\frac{T_\gamma}{2.8°\mathrm{K}}\right)^3 N_F^{1/2}\left(\frac{\mathrm{GeV}^{-2}}{\tilde{a}x_f + \frac{1}{2}\tilde{b}x_f^2}\right)\frac{\mathrm{g}}{\mathrm{cm}^3}. \tag{3.24b}$$

Results obtained using this formula are discussed in subsect 3.3.

3.3. RESULTS

We have completed a numerical study of the two-dimensional space of parameters M_2, ε in order to distinguish the cosmologically allowed and disallowed domains. The following algorithm was used.

(i) M_2 and ε are selected and the physical mass eigenstates (3.9) and eigenmasses are obtained by diagonalizing the matrix (3.6).

TABLE 1

T_f	N_F	$(T_\chi/T_\gamma)^3$
$m_e - m_\mu$	$\frac{43}{8}$	2.75
$m_\mu - m_\pi$	$\frac{57}{8}$	3.65
$m_\pi - T_H$	$\frac{69}{8}$	4.41
$T_H - m_s$	$\frac{205}{8}$	13.1
$m_s - m_c$	$\frac{247}{8}$	15.8
$m_c - m_\tau$	$\frac{289}{8}$	18.5
$m_\tau - m_b$	$\frac{303}{8}$	19.4
$m_b - m_t$	$\frac{345}{8}$	22.1
$m_t - m_w$	$\frac{387}{8}$	24.8
$> m_w$	$\frac{423}{8}$	27.1

This table is adapted from ref. [24]. Most of the notation is described in the text, with the exception that T_H is the temperature above which it is supposed that hadrons should be described in terms of quark and gluon degrees of freedom.

(ii) The coefficients A_f and B_f in the effective lagrangian (3.14) for the lightest mass eigenstate χ are computed from eqs. (3.15).

(iii) The coefficients $\tilde{a}$ and $\tilde{b}$ in the annihilation cross section (3.17) are calculated.

(iv) The scaled freeze-out temperature x_f is determined iteratively using (3.22b), with N_F being updated using table 1 following each iteration.

(v) The cosmological mass density ρ_χ is calculated from (3.24b) and compared with the cosmological bounds $\rho_\chi \leqslant 2 \times 10^{-29}$ g cm^3 and $\rho_\chi \leqslant 2 \times 10^{-30}$ g/cm^3.

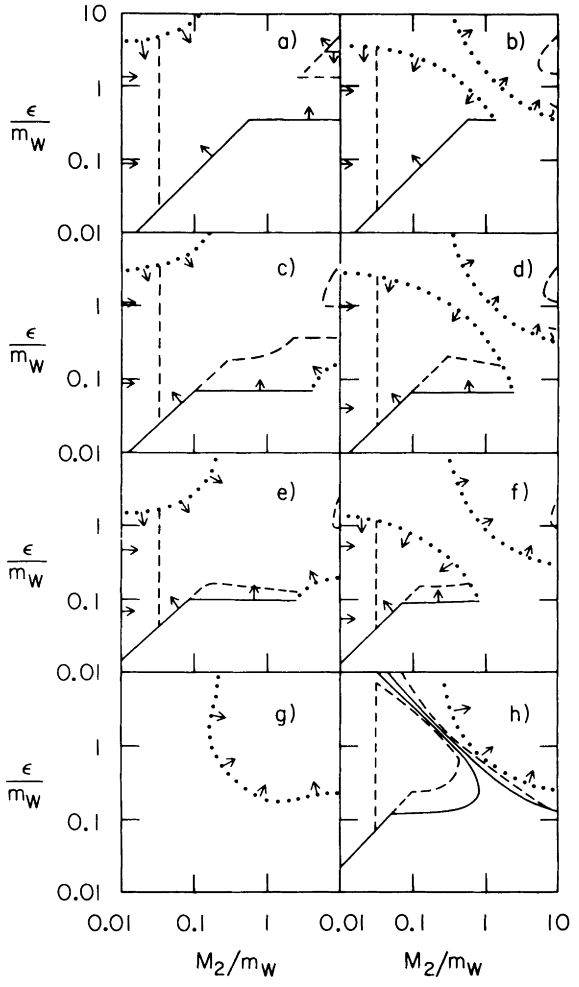


Fig. 4. Allowed domains of $|\varepsilon|$ and M_2 for the same choices of v_1/v_2 and of the sign of ε as in fig. 1. The solid curves are where $\rho_\chi = 2 \times 10^{-29}$ g/cm^3, while the dashed curves are for $\rho_\chi = 2 \times 10^{-30}$ g/cm^3. These results were obtained with $m_{\tilde{\ell}_L} = m_{\tilde{\ell}_R} = m_{\tilde{q}} = 20$ GeV or m_χ, whichever is larger.

Except in the case of a very light higgsino eigenstate ($\varepsilon < M_2 \ll m_W$) we find that $0.04 \lesssim x_f \lesssim 0.08$, which supports the non-relativistic approximation (3.16)–(3.23) based upon $x_f \ll 1$.

Our numerical results are exhibited in figs. 3–5 (preliminary early results were shown in ref. [6]). We divide the space of parameters into allowed versus disallowed regions, where the boundaries of acceptability corresponding to $\rho_\chi = 2 \times 10^{-29}$ g/cm^3 and to $\rho_\chi = 2 \times 10^{-30}$ g/cm^3 are drawn as solid and dashed lines respectively. Dotted lines represent the 20 GeV mass constraint on the lightest *charged* susy fermion mandated by PETRA and PEP. In figs. 4–6 the squark and slepton masses appearing in the $\chi\chi$ annihilation cross section (3.15b, c) are taken to be degenerate, and equal to 20, 40 and 100 GeV or m_χ, whichever is the larger, in figs. 4, 5 and 6

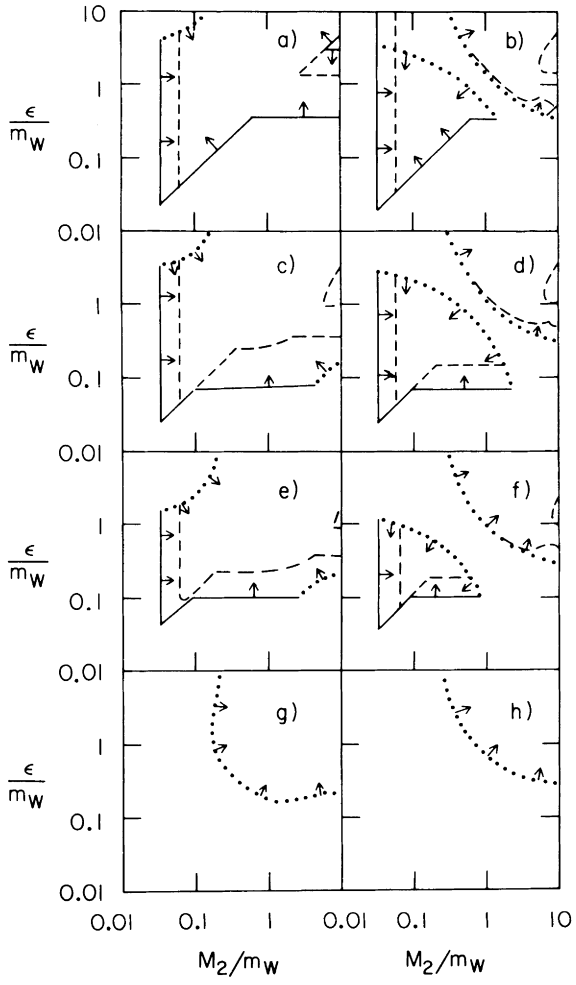


Fig. 5. As for fig. 4, except that $m_{\tilde{\ell}_L} = m_{\tilde{\ell}_R} = m_{\tilde{q}} = 40$ GeV has been used.

respectively. We begin our qualitative discussion of the results with fig. 4a, where $m_{\tilde{\ell}_L} = m_{\tilde{\ell}_R} = m_{\tilde{q}} = 20$ GeV, the vacuum expectation values v_1 and v_2 are equal, and ϵ is positive. There is a diagonal line traversing much of the figure from bottom left towards the top right, above which the lightest mass eigenstate χ is approximately a photino and below which χ is a pure $\tilde{S}^0$ higgsino state. Note that in the exact $v_1 = v_2$ limit, the Z-boson exchange contribution (3.15a, b) to the $\chi\chi$ annihilation cross section vanishes because $|\gamma| = |\delta|$. Therefore the $\tilde{S}^0$ can only annihilate through its Yukawa couplings to fermions, which are not large enough for the $\tilde{S}^0$ to annihilate to acceptably low levels unless $m_{\tilde{S}^0} > m_t$. Once the $\tilde{S}^0$ mass (see fig. 2a) exceeds that of the t-quark (taken to be 30 GeV on the basis of rumours [25] from UA1), its annihilation into $\bar{t}t$ pairs is sufficiently efficient that the $\tilde{S}^0$ becomes

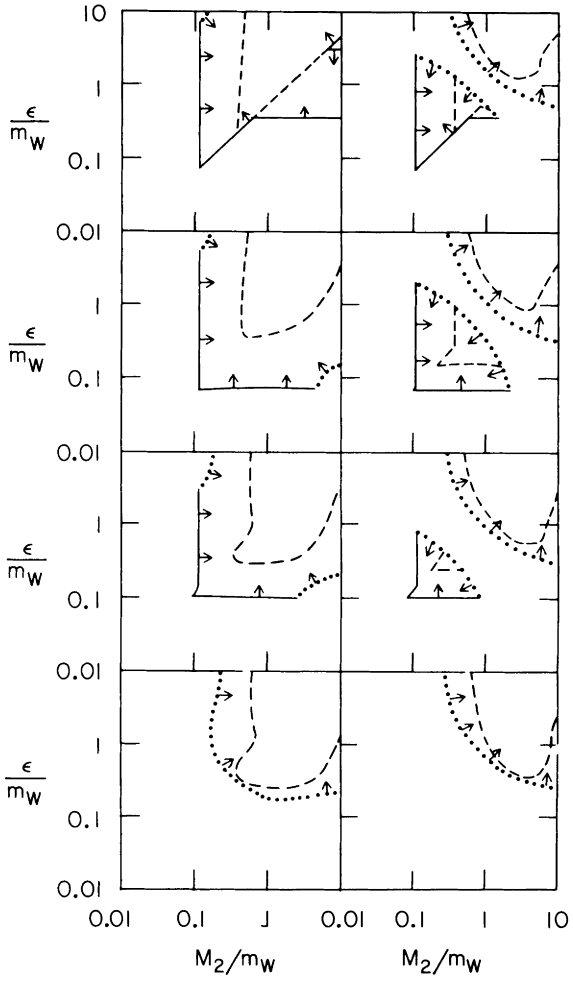


Fig. 6. As for fig. 4, except that $m_{\tilde{\ell}_R} = m_{\tilde{\ell}_L} = m_{\tilde{q}} = 100$ GeV has been used.

cosmologically allowed. This explains the solid horizontal line in fig. 4a which follows the 30 GeV contour in fig. 2a. Note that this boundary is sufficiently abrupt that the $\rho_\chi \leqslant 2 \times 10^{-29}$ g/cm^3 boundary (solid) and the $\rho_\chi \leqslant 2 \times 10^{-30}$ g/cm^3 boundary (dashed) nearly coincide.

Above the diagonal line in fig. 4a the lightest eigenstate χ is approximately a photino, which annihilates much more efficiently than the $\tilde{S}^0$ through its gauge couplings to fermions. The cosmological boundary $\rho \leqslant 2 \times 10^{-29}$ g/cm^3 occurs at a vertical line near the extreme left of the figure, corresponding to a photino mass of $m_{\tilde{\gamma}} \simeq \frac{1}{2}$ GeV (see fig. 2a). This result differs significantly from ref. [23], which excludes a photino lighter than about 2 GeV. The difference arises partly due to a factor of two discrepancy in the annihilation cross section (3.16), but primarily because we include masses for the light quarks, without which the $\chi\chi$ annihilation cross section would be purely P-wave below the charm-τ threshold. Our results were obtained using current quark masses $m_{u,d} = 10$ MeV and $m_s = 150$ MeV; the use of constituent quark masses gives a lower bound of $m_{\tilde{\gamma}}$ that is only slightly smaller. The lower bound of $m_{\tilde{\gamma}} \geqslant \frac{1}{2}$ GeV ($M_2 \geqslant 0.8$ GeV) is our most conservative bound, since it is derived using $\rho_\chi \leqslant 2 \times 10^{-29}$ g/cm^3 and sfermion masses of 20 GeV. The less conservative cosmological bound $\rho_\chi \leqslant 2 \times 10^{-30}$ g/cm^3 already forces the photino mass above the charm-τ threshold (dashed), and the use of heavier sfermion masses (figs. 5a, 6a) leads to considerably larger bounds.

If $v_1 \neq v_2$ as in figs. 4c, e, g, the constraint on the (now no longer pure $\tilde{S}^0$) higgsino mass is reduced from the t-quark mass to the b-quark mass (see figs. 2c, e, g). The reason is that there is now a significant contribution to the $\tilde{S}^0\tilde{S}^0$ annihilation cross section from Z-boson exchange. The b-quark channel is still needed to provide a significant S-wave contribution to the annihilation cross section (3.16). On the basis of figs. 4–6 we conclude generally that if the (almost pure) $\tilde{S}^0$ is the lightest susy fermion but heavier than about 100 eV, then it must be heavier than the b-quark.

The results for $\varepsilon > 0$ (figs. 4b, d, f, h) are very similar to those for $\varepsilon > 0$ except for the appearance of a narrow sheath of cosmologically forbidden territory (shown only in fig. 4h) due to an accidental zero in the mass matrix (3.6). As seen in figs. 1b, d, f, h, a similar line of zeros also appears in the charged fermion matrix (3.3), so that this tantalizing sheath is doubly excluded.

The most important change as v_1/v_2 increases (figs. 4c–h) is that the 20 GeV constraint on the lightest charged eigenstate becomes more restrictive. In figs. 4g, h we observe that if $v_1/v_2 \gtrsim 8$, much of the parameter space is excluded by the charged constraint, so that the cosmological bounds are almost irrelevant.

The constraints that result from setting the sfermion masses equal to 40, and 100 GeV or m_χ, whichever is the larger, are qualitatively similar (figs. 5, 6) except for an overall shrivelling of the cosmologically allowed regions. In particular, the lower bound on the photino mass from $\rho_\chi \leqslant 2 \times 10^{-29}$ g/cm^3 increases from $\frac{1}{2}$ GeV to 1.8 GeV for $m_{\tilde{\ell}_L} = m_{\tilde{\ell}_R} = m_{\tilde{q}} = 40$ GeV and to 5 GeV when $m_{\tilde{\ell}_L} = m_{\tilde{\ell}_R} = m_{\tilde{q}} = 100$ GeV. The analogous constraint from $\rho_\chi \lesssim 2 \times 10^{-30}$ g/cm^3 increases from 1.8 GeV to 3 GeV when $m_{\tilde{f}} = 40$ GeV, and to approximately 15 GeV if $m_{\tilde{f}} = 100$ GeV.

Figs. 7 and 8 are intended to illustrate the effects of varying the *relative* masses of right- and left-handed sleptons and of squarks. The effects of taking $m_{\tilde{\ell}_R} = m_{\tilde{\ell}_L} = m_{\tilde{q}} = 20$; $m_{\tilde{\ell}_R} = m_{\tilde{\ell}_L} = 20$, $m_{\tilde{q}} = 50$; $m_{\tilde{\ell}_R} = 20$, $m_{\tilde{\ell}_L} = 40$, $m_{\tilde{q}} = 100$ GeV (corresponding to three classes of models described in EHNT [5]) when $v_1 = 4v_2$ are shown respectively in a, b, c of figs. 7 and 8 for $\epsilon > 0$ and $\epsilon < 0$ respectively. We observe that the effect of changing the sfermion mass hierarchy is rather negligible, except for a slight increase in the photino mass limit with heavier squark masses (due to a suppression of the photino annihilation into light quarks). It is because these effects are small that we chose to focus our earlier presentation on the case of degenerate squarks and sleptons.

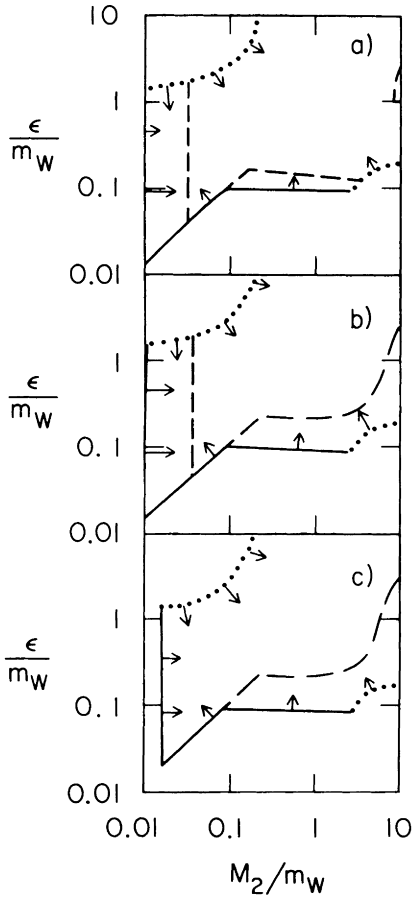


Fig. 7. The effects of varying the relative masses of the sfermions for $v_1 = 4v_2$ and $\epsilon > 0$: (a) $m_{\tilde{\ell}_R} = m_{\tilde{\ell}_L} = m_{\tilde{q}} = 20$ GeV; (b) $m_{\tilde{\ell}_R} = m_{\tilde{\ell}_L} = 20$ GeV, $m_{\tilde{q}} = 50$ GeV; (c) $m_{\tilde{\ell}_R}$20 GeV, $m_{\tilde{\ell}_L} = 40$ GeV, $m_{\tilde{q}} = 100$ GeV.

3.4. A VERY LIGHT HIGGS FERMION?

Finally we consider the case of a very light higgsino state, with mass $\lesssim$ O(100) eV. From a cosmological standpoint, the light higgsino looks very much like a neutrino. Its annihilation into light fermions is dominated by Z-boson exchange, with a relativistic annihilation cross section

$$\langle \sigma v_{\rm rel} \rangle = (\gamma^2 - \delta^2)^2 \frac{2G_F^2 s}{3\pi} \left(g_V^{e^2} + g_A^{e^2} \right), \tag{3.25}$$

where $g_V^e = -\frac{1}{2} + 2\sin^2\theta_W$ and $g_A^e = -\frac{1}{2}$. The higgsino falls out of thermal equilibrium when the interaction rate given by (3.25) falls below the expansion rate of

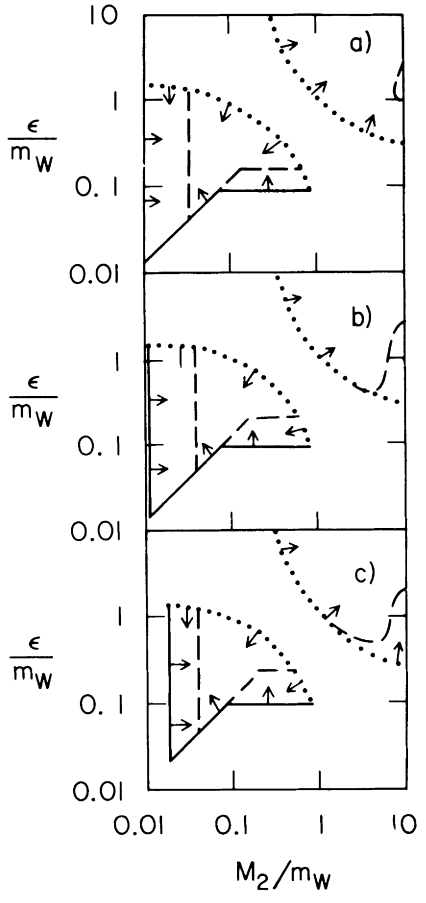


Fig. 8. As for fig. 7, but for the choices $v_1 = 4v_2$ and $\varepsilon < 0$.

the universe. This occurs at a temperature

$$T_{\mathrm{E}} = \left(\frac{1}{(\gamma^2 - \delta^2)^2} \frac{3\sqrt{10\pi}}{G_{\mathrm{F}}^2 N_{\mathrm{F}}^{1/4} m_{\mathrm{P}}} \right)^{1/3} . \tag{3.26}$$

Thus for higgsinos we find

$$T_{\mathrm{E}} = \frac{1.8\ \mathrm{MeV}}{(\gamma^2 - \delta^2)^{2/3}} . \tag{3.27}$$

If T_{E} (3.27) becomes larger than m_μ, $\mu^+\mu^-$ annihilation does not reheat the higgsinos, and their number density will be less than that of the neutrinos by a factor of $\frac{43}{57}$. In this case the permissible upper bound on m_{S^0} is a factor $\frac{57}{43}$ higher than would be allowed for neutrinos. We see from eq. (3.7) that

$$|\gamma^2 - \delta^2| = \left| \frac{v_1^2 - v_2^2}{v_1^2 + v_2^2} \right| , \tag{3.28}$$

which when combined with equation (3.27) tells us that the higgsinos avoid reheating by $\mu^+\mu^-$ annihilation if

$$\left| \frac{v_1^2 - v_2^2}{v_1^2 + v_2^2} \right| < 2 \times 10^{-3} . \tag{3.29}$$

Is it plausible that v_1^2 and v_2^2 could have such similar values? We can only answer this question by looking at models. In most models [5] that we know v_1 is considerably larger than v_2, and the condition (3.29) is not satisfied. There is one model (EHNT ref. [5]) in which v_1 and v_2 can be similar

$$\left| \frac{v_1^2 - v_2^2}{v_1^2 + v_2^2} \right| \approx \frac{g_2^2 \mu^2}{(g_2^2 + g_1^2) m_{\mathrm{W}}^2} , \tag{3.30}$$

but even in this case $\mu \gtrsim$ O(20) GeV (EHNT, ref. [5]), so eq. (3.29) tells us that $\mu^+\mu^-$ *does* reheat the higgsinos. We conclude that in all plausible models the relevant upper mass bound on the higgsinos is the same as that for neutrinos, namely $m_\chi \lesssim$ O(100) eV.

It is natural to suppose that as the Higgs fermion mixing parameter $\varepsilon \to 0$ in eq. (3.1), so also does the corresponding Higgs boson mixing parameter. The theory would then possess a Peccei-Quinn [26] U(1) symmetry. However, the physical spectrum then contains an axion. In the EHNT model of ref. [5] the physical Higgs

spectrum in this U(1) symmetry limit contains [27] an axion with a mass of around the geometric mean of ϵ and the gravitino mass, which is around an MeV. Such an axion would conflict with particle physics experiments and must therefore be exorcised. It is possible that this may be done in more complicated models by increasing the axion mass relative to the higgsino mass. For this reason we are reluctant to conclude that a light higgsino can be excluded by this indirect argument.

4. Summary

In this paper we have made a systematic and complete study of the cosmological constraints on supersymmetric theories extending the analysis of ref. [23] and updating ref. [6]. Our analysis is tailored preferentially, but not exclusively, to low-energy models [5] extracted from an underlying $N = 1$ supergravity theory. Where necessary, we have supplemented strictly cosmological constraints by clearly stated phenomenological assumptions. We have found that the lightest supersymmetric particle (LSP) is unlikely to be a charged wino, charged higgsino, slepton, sneutrino, gluino, squark, or gravitino. Most probably the LSP is a mixture of neutral gauge and Higgs fermions. We have explored the cosmologically allowed domains of parameter space for these particles, and our results are shown in figs. 4 to 8. The cosmologically allowed domains shrivel as the assumed masses of the sfermions are increased. We find one favoured class of solutions in which there is a mass eigenstate which is approximately a photino of mass above $\frac{1}{2}$ to 2 GeV, while the lightest (approximate) higgsino is generally heavier than the photino. Lighter higgsinos do not annihilate efficiently enough to suppress their present cosmological abundance below the acceptable limits unless they are heavier than the b- or t-quark. The cosmology of a very light higgsino is similar to that of a massive neutrino, and a higgsino weighing less than O(100) eV is equally compatible with observation.

References

[1] Proc. Supersymmetry versus experiment workshop, ed. D.V. Nanopoulos, A. Savoy-Navarro and C. Tao, CERN preprint TH-3311/EP-82/63 (1983);
I. Hinchliffe and L. Littenberg, Proc. D.P.F. Summer Study on elementary particle physics and future facilities, Snowmass 1982, eds. R. Donaldson, R. Gustafson and F. Paige (APS, 1982) p. 292

[2] G. Steigman, Proc. 1982 SLAC Summer Institute on particle physics, ed. A. Mosher (SLAC-259, 1983) p. 651

[3] E. Cremmer, B. Julia, J. Scherk, S. Ferrara, L. Girardello and P. van Nieuwenhuizen, Phys. Lett. 79B (1978) 231; Nucl. Phys. B147 (1979) 105;
E. Cremmer, S. Ferrara, L. Girardello and A. van Proeyen, Phys. Lett. 116B (1982) 231; Nucl. Phys. B212 (1983) 413

[4] J. Ellis and D.V. Nanopoulos, Phys. Lett. 116B (1982) 133;
R. Barbieri, S. Ferrara and C.A. Savoy, Phys. Lett. 119B (1982) 343;
H.P. Nilles, M. Srednicki and D. Wyler, Phys. Lett. 120B (1982) 346;
L.E. Ibáñez, Phys. Lett. 118B (1982) 73;
J. Ellis, D.V. Nanopoulos and K. Tamvakis, Phys. Lett. 121B (1983) 123

[5] K. Inoue, A. Kakuto, H. Komatsu and S. Takeshita, Prog. Theor. Phys. 67 (1982) 1889; 68 (1982) 927;
L.E. Ibáñez and G.G. Ross, Phys. Lett. 110B (1982) 215;
L. Alvarez-Gaumé, M. Claudson and M.B. Wise, Nucl. Phys. B207 (1982) 96;
J. Ellis, L.E. Ibáñez and G.G. Ross, Phys. Lett. 113B (1982) 283; Nucl. Phys. B221 (1983) 29;
L. Alvarez-Gaumé, J. Polchinski and M.B. Wise, Nucl. Phys. B221 (1983) 495;
L.E. Ibáñez and C. López, Phys. Lett. 126B (1983) 54;
J. Ellis, J.S. Hagelin, D.V. Nanopoulos and K. Tamvakis, Phys. Lett. 125B (1983) 275
[6] J. Ellis, J.S. Hagelin, D.V. Nanopoulos and M. Srednicki, Phys. Lett. 127B (1983) 233
[7] S. Wolfram, Phys. Lett. 82B (1979) 65
[8] P.F. Smith and J.R.J. Bennett, Nucl. Phys. B149 (1979) 525
[9] C.B. Dover, T.K. Gaisser and G. Steigman, Phys. Rev. Lett. 42 (1979) 1117;
D.A. Dicus and V.L. Teplitz, Phys. Rev. Lett. 44 (1980) 218
[10] R. Middleton, R.W. Zurmühle, J. Klein and R.V. Kollarits, Phys. Rev. Lett. 43 (1979) 429
[11] G.R. Farrar and P. Fayet, Phys. Lett. 76B (1978) 575; 79B (1978) 442;
B.A. Campbell, J. Ellis and S. Rudaz, Nucl. Phys. B198 (1982) 1;
G.L. Kane and J.P. Leveillé, Phys. Lett. 112B (1982) 227;
P.R. Harrison and C.H. Llewellyn Smith, Nucl. Phys. B213 (1983) 223;
CHARM Collaboration, F. Bergsma et al., Phys. Lett. 121B (1983) 429
[12] CELLO Collaboration, H. Behrend et al., Phys. Lett. 114B (1982) 287;
JADE Collaboration, W. Bartel et al., Phys. Lett. 114B (1982) 211;
MARK J Collaboration, D.P. Barber et al., Phys. Rev. Lett. 45 (1981) 1904;
TASSO Collaboration, R. Brandelik et al., Phys. Lett. 117B (1982) 365;
MARK II Collaboration, C.A. Blocker et al., Phys. Rev. Lett. 49 (1982) 517
[13] J.S. Hagelin, G.L. Kane and S. Raby, in preparation (1983)
[14] S. Weinberg, Phys. Rev. Lett. 48 (1982) 1303
[15] J. Ellis, A.D. Linde and D.V. Nanopoulos, Phys. Lett. 118B (1982) 59
[16] S. Weinberg, private communication and in preparation (1983)
[17] D.V. Nanopoulos, K.A. Olive, and M. Srednicki, Phys. Lett. 127B (1983) 30
[18] L. M. Krauss, Harvard University preprint, HUTP-83-A009 (Rev.) (1983)
[19] J.-M. Frère and G.L. Kane, Nucl. Phys. B223 (1983) 331
[20] J. Ellis and S. Rudaz, Phys. Lett. 128B (1983) 248
[21] S. Faber and J. Gallagher, Ann. Rev. Astron. Astrophys. 17 (1979) 135
[22] B.W. Lee and S. Weinberg, Phys. Rev. Lett. 39 (1977) 165
[23] H. Goldberg, Phys. Rev. Lett. 50 (1983) 1419
[24] K.A. Olive, D. Schramm and G. Steigman, Nucl. Phys. B180 (1981) 497
[25] UA1 Collaboration, D. Dibitonto, Talks at Recontre de Moriond and at 14th Int. Symp. on multiparticle dynamics, Granlibakken, Lake Tahoe (1983)
[26] R.D. Peccei and H.R. Quinn, Phys. Rev. Lett. 38 (1977) 1440; Phys. Rev. D16 (1977) 1791
[27] J. Ellis, J.S. Hagelin, D.V. Nanopoulos and K. Tamvakis, in preparation (1983)

Nuclear Physics B310 (1988) 693–713
North-Holland, Amsterdam

CALCULATIONS OF RELIC DENSITIES IN THE EARLY UNIVERSE

Mark SREDNICKI and Richard WATKINS

Department of Physics, University of California, Santa Barbara, CA 93106, USA

Keith A. OLIVE

School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455, USA

Received 24 May 1988

The cosmological relic densities of stable particles are calculated to high precision. The results depend on assumptions made about the quark–hadron phase transition. We discuss in detail the calculation of the thermally averaged annihilation cross section. We consider the specific cases of photinos, higgsinos, and Dirac and Majorana neutrinos.

1. Introduction

In recent years, many theories have been proposed that predict the existence of new stable particles that could account for the observed dark matter in the universe. These theories have free parameters which, because of the high energies and weak couplings involved, cannot currently be probed by terrestrial accellerators. This has led physicists to consider ways in which astrophysical data can be used to put constraints on the masses and couplings of dark-matter candidates.

One of the most important constraints is the requirement that the relic mass density of dark matter be compatible with present observations. In the early universe, these particles would have been present in large numbers in thermal equilibrium, and as the universe cooled they could reduce their density only through pair annihilation [1, 2, 3]. As their density decreases, however, it becomes more and more difficult for particles to find others with which to annihilate, and the comoving density can become constant, or "freeze out". This results in a relic density many orders of magnitude higher than that naively expected from thermodynamic arguments.

The purpose of this paper is to calculate the relic mass densities of stable particles as accurately as possible. Since we are dealing with particles which are not in thermal equilibrium, this involves tracing the evolution of the mass density with a precise rate equation. In sect. 2 we discuss this rate equation in detail. In sect. 3 we

0550-3213/88/$03.50 © Elsevier Science Publishers B.V.
(North-Holland Physics Publishing Division)

study the role played by the expansion of the universe and calculate the effective degrees of freedom in the universe as a function of temperature. In sect. 4 we discuss the thermally averaged annihilation cross section and introduce a new approximation which allows us to significantly improve on the accuracy of previous calculations. In sects. 5 and 6, we use our techniques to calculate the relic mass densities of photinos, higgsinos, and heavy neutrinos (both Dirac and Majorana). In an appendix we discuss the (often confusing) factors of two related to correctly averaging over initial spins and particle–antiparticle states.

2. The rate equation

The rate equation for a hypothetical stable particle χ is

$$\frac{\mathrm{d}n}{\mathrm{d}t} = -3\frac{\dot{R}}{R}n - \langle\sigma v_{\mathrm{rel}}\rangle\left(n^2 - n_0^2\right), \tag{1}$$

where n is the actual number density of χ particles, summed over spins and particle–antiparticle versions (see the appendix); n_0 is the number density the χ particles would have in thermal equilibrium at temperature T; R is the cosmic scale factor; σ is the annihilation cross section, properly averaged over initial spins and particle–antiparticle states (see the appendix); v_{rel} is the relative velocity of the annihilating particles; and $\langle\ \rangle$ implies a thermal average which we will explain in sect. 4.

It is convenient to replace time in the above equation by the photon temperature. For this purpose we define effective degrees of freedom $h(T)$ and $g(T)$ by

$$h(T) \equiv \frac{3}{2}\left(\frac{15}{\pi^2}\right)\frac{s(T)}{T^3}, \tag{2a}$$

$$g(T) \equiv 2\left(\frac{15}{\pi^2}\right)\frac{\rho(T)}{T^4}, \tag{2b}$$

where $s(T)$ and $\rho(T)$ are, respectively, the total entropy and energy densities of the universe, and T is the temperature of the photon gas. With these definitions photons contribute $+2$ to both $g(T)$ and $h(T)$; fully relativistic electrons and positrons together contribute $+\frac{7}{2}$; nonrelativistic particles (in chemical equilibrium) do not contribute. The total entropy S is proportional to hT^3R^3, and entropy conservation therefore implies

$$\frac{\dot{R}}{R} = -\frac{\dot{T}}{T} - \frac{1}{3}\frac{h'}{h}\dot{T}, \tag{3}$$

where $h' \equiv \mathrm{d}h/\mathrm{d}T$.

If we assume a standard Robertson–Walker cosmology, then we can express the time rate of change of the scale factor as

$$\frac{\dot{R}}{R} = \left(\tfrac{8}{3}\pi G\rho\right)^{1/2}. \tag{4}$$

This, along with eqs. (2b) and (3), allows us to put eq. (1) in the convenient form

$$\frac{\mathrm{d}q}{\mathrm{d}x} = M\left(\tfrac{4}{45}\pi^3 Gg\right)^{-1/2}\left(h + \tfrac{1}{3}Mxh'\right)\langle\sigma v_{\mathrm{rel}}\rangle\left(q^2 - q_0^2\right), \tag{5}$$

where M is the mass of a χ particle (assumed to be at least a few GeV), $x \equiv T/M$, $q \equiv n/T^3h$, and $q_0 \equiv n_0/T^3h$. It is important to remember that T is the temperature of the photon gas, which may differ from that of the χ particles (more on this below).

Eq. (5), which is our starting point for numerical analysis, differs somewhat from the one used in most previous calculations. The difference is that we have retained terms related to the change of h with temperature; in particular the second term on the right-hand side of eq. (3) has usually not been included (an exception is in ref. [4]). This has its greatest effect at $T \simeq 100$–400 MeV when the quark–hadron transition occurs, and at $T \simeq 0.5$ MeV when electrons and positrons annihilate away. The effect of the h' term at $T \simeq 0.5$ MeV can be estimated analytically, since by this late time the right-hand side of eq. (5) is negligible and q is essentially constant and equal to its value at $T = 0$. Thus $n(T) \simeq q(0)h(T)T^3$. When electrons and positrons annihilate, $h(T)$ decreases by a factor of approximately $\frac{4}{11}$. Therefore, so does $n(T)$. This reduction has been done "by hand" in previous calculations, whereas our numerical integration of eq. (5) does it automatically (and more accurately).

Other qualitative features of the solution $q(x)$ can be deduced from eq. (5). For $x \gg 1$, the χ particles are relativistic, and we expect them to be in chemical equilibrium with the photons so that $q = q_0$. In this regime $n_0 \sim T^3$ so $q_0 \sim$ constant. When $x \sim 1$, q_0 begins to fall like $\mathrm{e}^{-M/T}$, and the χ particles begin to annihilate. As their density decreases, however, it becomes harder for the χ's to find others to annihilate with, until at some temperature T_{f}, usually called the "freeze out" temperature, $q(x)$ ceases to track $q_0(x)$ and remains roughly constant as $T \to 0$. At this point the χ particles are no longer in *chemical* equilibrium with the photons. However, the χ particles can still be in *mechanical* equilibrium with the photon gas through scatterings with particles of other species, such as electrons. Because of this, in all cases of interest the photon gas and the χ gas stay at the same temperature down to temperatures of order a few MeV. By this time, $\langle\sigma v_{\mathrm{rel}}\rangle$ is essentially constant and $n_0 \ll n$. Therefore, the subsequent discrepancy between the photon

and χ temperatures does not affect the numerical solution of eq. (5), and we can ignore it.

To obtain the numerical solution of eq. (5), we integrate from $x \gg 1$ (where $q = q_0$) down to $x = 0$, thus finding $q(0)$. The present mass density of the χ particles is then simply given by

$$\rho_\chi = Mq(0)h(0)T_0^3, \tag{6}$$

where $T_0 = 2.7$ K is the present photon temperature, and $h(0)$ is computed in sect. 3. It is convenient to define, as usual, $\Omega_\chi = \rho_\chi/\rho_{\rm crit}$, where $\rho_{\rm crit} = 3H^2/8\pi G$ is the critical density, $H = (100 \text{ km s}^{-1} \text{ Mpc}^{-1})h$ is the present value of the Hubble parameter $\dot{R}/R$, and $0.5 \lesssim h \lesssim 1$ encodes the uncertainty in its value*. We then find

$$\Omega_\chi h^2 = (1.555 \times 10^8)(M/\text{GeV})h(0)q(0). \tag{7}$$

3. Effective degrees of freedom

In order to solve our rate equation, we must be able to compute the functions $g(T)$ and $h(T)$. Here we follow, with some modification, the treatment in ref. [5].

Suppose we have a particle species i present at temperature T whose interaction energy is small compared with its free particle energy. In general, the equilibrium density of particles of type i with momenta in a range d^3p centered on $\boldsymbol{p}$ is $f(E_i)\,\mathrm{d}^3p$, where

$$f(E_i) = \frac{\kappa_i}{(2\pi)^3}\left[\exp(E_i/T) \pm 1\right]^{-1}. \tag{8}$$

Here κ_i is the number of spin and particle–antiparticle states, $E_i = (m_i^2 + \boldsymbol{p}^2)^{1/2}$, and the plus (minus) sign is for fermions (bosons). In terms of integrals over $f(E_i)$, the contribution of species i to the energy and entropy densities is

$$\rho_i = \int \mathrm{d}^3p f(E_i)E_i, \qquad s_i = \int \mathrm{d}^3p f(E_i)\frac{3m_i^2 + 4\boldsymbol{p}^2}{3E_iT}. \tag{9),(10}$$

We then have

$$\rho = \sum_i \rho_i = \frac{\frac{1}{2}g(T)\pi^2T^4}{15}, \qquad s = \sum_i s_i = \frac{\frac{2}{3}h(T)\pi^2T^3}{15}. \tag{11),(12}$$

* The parameter h should not be confused with the entropy function $h(T)$.

This treatment assumes that all the particle species are in thermal equilibrium with the photons. This breaks down if the interaction rate of a particular species becomes of order the expansion rate of the universe. In this case that species is said to have decoupled, and its entropy and the entropy of the other particles (which continue to interact) are separately conserved. In particular, if a massive species annihilates and reheats the interacting particles, the decoupled species will not share in the released energy and so will differ in temperature from the interacting particles. The temperature of the decoupled particles is calculated from conservation of entropy, and is then used to find their contribution to $g(T)$ and $h(T)$.

Suppose a species i decouples at temperature $T = T_d$. (Of course, the decoupling temperature is actually a range of temperatures, but the following treatment proves more than accurate enough for our needs.) We know that $hT^3R^3 = \text{constant}$, and we can rescale R so that $hT^3R^3 \equiv 1$. Below T_d, species-i particles no longer interact with the other particles, so their entropy is conserved separately. Thus for $T < T_d$ we can write

$$h_i(T)T_i^3R(T)^3 = \gamma, \qquad h_o(T)T^3R(T)^3 = 1 - \gamma, \tag{13a,b}$$

where h_i is the contribution of species i to h, T_i is the temperature of species i, $h_o = \Sigma_{j \neq i} h_j$ (the subscript o means "other"), and the constant γ is specified by

$$\gamma = h_i(T_d)T_d^3R(T_d)^3, \qquad 1 - \gamma = h_o(T_d)T_d^3R(T_d)^3. \tag{14a,b}$$

Taking the ratio of eqs. (14a) and (14b) we find that

$$\gamma = \frac{1}{1 + h_o(T_d)/h_i(T_d)}. \tag{15}$$

Using $h(T) = 1/T^3R(T)^3$ along with eqs. (13b) and (15), we deduce that for $T < T_d$

$$\begin{aligned} h(T) &= \frac{1}{1-\gamma} h_o(T) \\ &= \left[1 + \frac{h_i(T_d)}{h_o(T_d)}\right] h_o(T). \end{aligned} \tag{16}$$

Furthermore from eqs. (13a), (13b), and (15) we find

$$\frac{T_i^3}{T^3} = \frac{h_i(T_d)h_o(T)}{h_o(T_d)h_i(T)}. \tag{17}$$

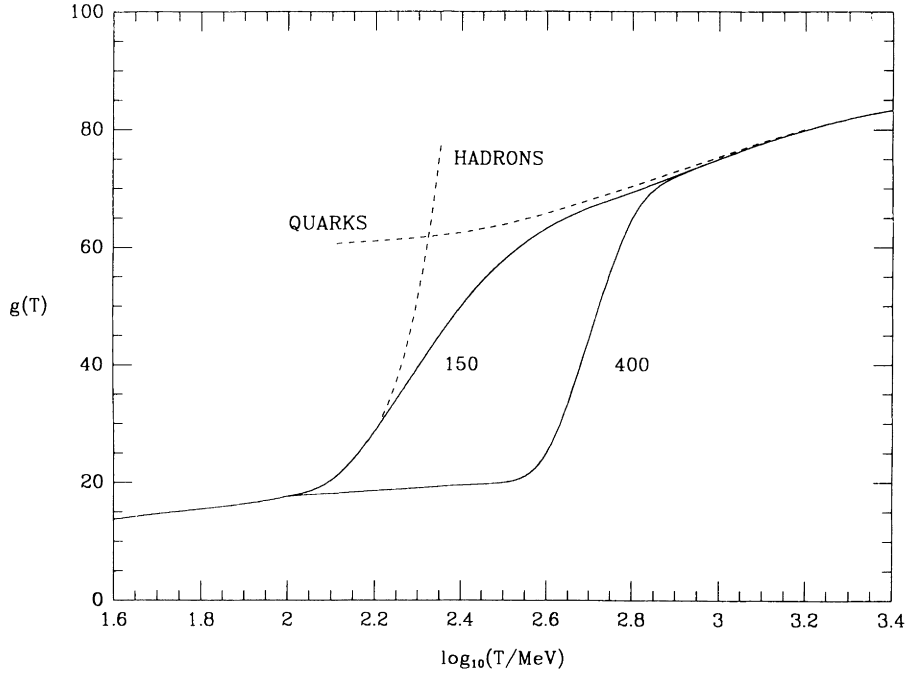


Fig. 1. The scaled energy density $g(T)$ for our two assumptions about the quark–hadron transition. The curve labeled "150" has this transition near 150 MeV, and the curve labeled "400" has it near 400 MeV. Curves for free quarks (with current algebra masses) and free hadrons are shown with dashed lines.

We need eq. (17) in order to compute $g(T)$

$$g(T) = g_o(T) + g_i(T_i)\frac{T_i^4}{T^4}\,. \tag{18}$$

We are concerned with the decoupling of light neutrinos which are fully relativistic at $T = T_d$. In this case both $g_i(T)$ and $h_i(T)$ are constants; each is equal to $\frac{7}{4}$ times the number of light-neutrino types.

If another species decouples at a lower temperature, the procedure above must be repeated, using eqs. (16) and (18) as the starting expressions for $h(T)$ and $g(T)$. To actually compute $g(T)$ and $h(T)$ at a given temperature T, we perform the integrals in eqs. (9) and (10) for each relevant species, and then do the sums in eqs. (11) and (12). We must decide whether we should be considering quarks or hadrons as particles to be included in the sum. We first consider temperatures well below the quark–hadron transition, $T \ll T_c \sim 100$–400 MeV. Here, we include the contributions from all the baryons and mesons*, as well as the leptons and the photons. At

* We used the tables in ref. [5]. Particles with masses above ~ 3 GeV contribute negligibly to $g(T)$ and $h(T)$ for $T \lesssim 400$ MeV.

temperatures of order a few MeV, neutrinos decouple and must be treated according to the procedure outlined above. We assumed [6] that electron neutrinos decouple at $T_d = 2.0$ MeV, and muon and tau neutrinos at $T_d = 3.5$ MeV. The temperatures differ since electron neutrinos interact with electrons via W exchange as well as Z exchange. We also took into account the fact that electrons are not fully relativistic at these temperatures. We ultimately find $h(0) = 3.915$ and $g(0) = 3.368$ for three generations of light neutrinos, and $h(0) = 4.553$ and $g(0) = 3.824$ for four generations of light neutrinos.

For temperatures approaching the quark–hadron transition, the particle density is so high that we no longer have an ideal gas of hadrons. As the interparticle distances become small, the hadrons begin to overlap and the interaction energy becomes important [7]. At even earlier times, the temperature is high enough so that the quarks become asymptotically free, and it is a good approximation to treat the matter as an ideal gas of quarks, gluons, leptons, W and Z bosons, and photons. At present we have no reliable model from which to obtain the values of g and h in the region between the free hadron phase and the free quark phase. It is believed that the quark–hadron phase transition takes place between 100 MeV and 400 MeV, with the lower end favored by recent results in lattice gauge theory [8]. We consider two illustrative cases, one with a quark–hadron transition near 150 MeV, and one

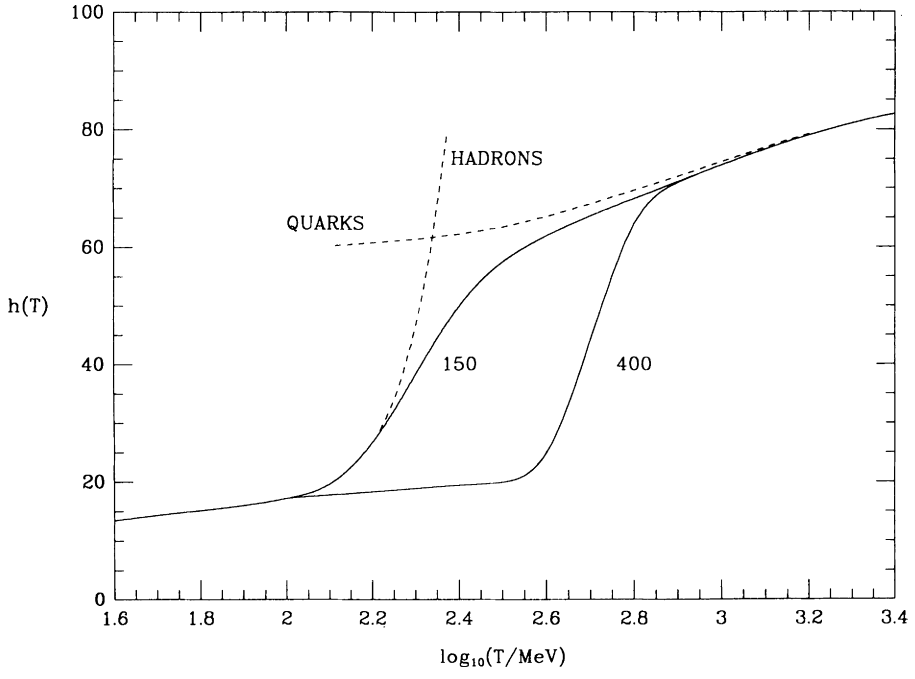


Fig. 2. Same as fig. (1), except for the scaled entropy density $h(T)$.

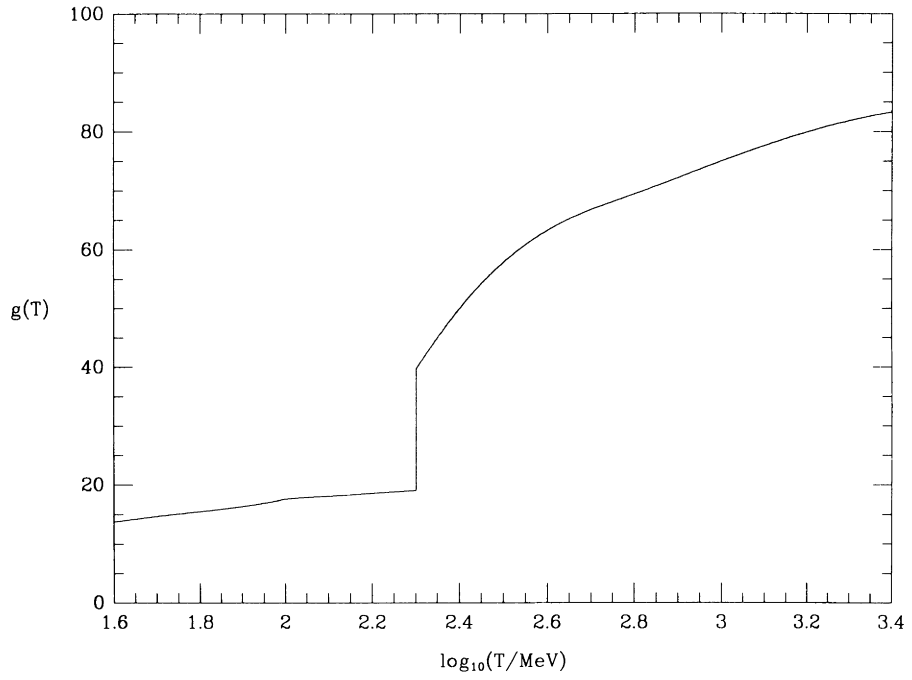


Fig. 3. The scaled energy density $g(T)$ assuming a first order quark–hadron transition at $T_c = 200$ MeV.

with this transition near 400 MeV. These two possibilities for $g(T)$ and $h(T)$ are shown in figs. 1 and 2, along with the free hadron and free quark curves. It is reasonably likely that the true curves lie between these two extremes, so that by calculating quantities using both choices we can infer the error introduced by our lack of knowledge of this regime*.

One interesting possibility (supported by the lattice results [8]) is that the quark–hadron phase transition was first order. This would imply that the temperature of the universe was kept constant during the transition by the energy released. If this were the case, then dh/dT would be infinite at the transition temperature, and one might worry that this could change the results radically from the cases that we describe above. This turns out not to happen. We performed a variety of calculations with the g curve in fig. 3 (and a corresponding h curve) and found that the results still lay between those obtained using the h and g curves of figs. 1 and 2.

* We should also note that the contribution of the χ particles themselves to $g(T)$ and $h(T)$ is not included in the figures since it varies with the assumed value of M. Furthermore, as discussed in sect. 2 we really only need to know $g(T)$ and $h(T)$ for $T \lesssim T_f$ where the χ contribution is completely negligible.

4. Thermal averaging

Let $|\mathscr{T}|^2$ be the absolute square of the reduced matrix element for the annihilation of two χ particles, summed over final spins and suitably averaged over initial spins and particle–antiparticle states (see the appendix for details). For fixed incoming momenta p_1 and p_2 we define as usual

$$\sigma v_{\rm rel} = \frac{1}{4E_1E_2}\int \mathrm{d\,LIPS}|\mathscr{T}|^2 , \tag{19}$$

where d LIPS (Lorentz Invariant Phase Space) is given by

$$\mathrm{d\,LIPS} = (2\pi)^4\delta^4\Big(p_1+p_2-\sum_j p_j\Big)\prod_i \frac{\mathrm{d}^3p_i}{(2\pi)^3 2p_i^0} . \tag{20}$$

The sum and product are each over the outgoing particles. Here we have assumed that Pauli blocking or enhancing factors for the phase space of the final state particles can be ignored. We will justify this in a moment.

Clearly $\int \mathrm{d\,LIPS}|\mathscr{T}|^2$ is Lorentz invariant, and therefore depends on the momenta only through the Mandelstam variable $s = -(p_1+p_2)^2$. It is convenient to define

$$w(s) \equiv \tfrac{1}{4}\int \mathrm{d\,LIPS}|\mathscr{T}|^2 = E_1E_2\sigma v_{\rm rel} . \tag{21}$$

We now suppose that the initial χ particles have an energy distribution $f(E)$, where $E = (\boldsymbol{p}^2 + M^2)^{1/2}$. We then define the thermal average as [9]

$$\langle \sigma v_{\rm rel}\rangle \equiv \frac{1}{n_0^2}\int \mathrm{d}^3p_1\,\mathrm{d}^3p_2\, f(E_1)f(E_2)\frac{1}{E_1E_2}w(s) , \tag{22}$$

where $n_0 = \int \mathrm{d}^3p f(E)$ has already made its appearance in eq. (5). Note, however, that $\langle \sigma v_{\rm rel}\rangle$ does not depend on the overall normalization of $f(E)$.

We know from our qualitative discussion that the χ-particles' density will begin to differ from its equilibrium value only after T falls below the freezing temperature $T_{\rm f}$. Typically $T_{\rm f} \sim \frac{1}{20}M$, so that for our purposes we need to know $\langle \sigma v_{\rm rel}\rangle$ only at temperatures where it is a very good approximation to take $f(E)$ to be a Boltzmann distribution regardless of the statistics of the particle. This is a great simplification, and justifies our ignoring any Pauli blocking/enhancing factors in eq. (19).

For a general cross section, the integrals in eq. (22) cannot be done analytically. Further, since this is a multiple integral, numerical solutions are expensive and difficult to obtain. The solution is to expand $\langle \sigma v_{\rm rel}\rangle$ in powers of $x \equiv T/M$. One can then do all the integrals analytically and obtain a relatively simple expression for

$\langle \sigma v_{\rm rel} \rangle$. Since, as discussed above, we are only interested in regimes where $x \lesssim \frac{1}{20}$, we need only take a few terms in the expansion to get very good accuracy. In practice we keep terms through order x^3.

To perform the expansion in x, we first write

$$s = -(p_1 + p_2)^2 = 2\bigl(M^2 + E_1 E_2 - p_1 p_2 \cos\theta\bigr), \tag{23}$$

where θ is the angle between $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$. This is the only dependence on angle in the integrand. Using the Boltzmann limit of eq. (8), $f(E) = \kappa(2\pi)^{-3} e^{-E/T}$ where κ is the number of spin states of the χ particle (see the appendix), we can write

$$\begin{aligned}
\langle \sigma v_{\rm rel} \rangle &= \frac{\kappa^2}{(2\pi)^6 n_0^2} \int d^3p_1\, d^3p_2\, e^{-E_1/T} e^{-E_2/T} \frac{1}{E_1 E_2} w(s), \\
&= \frac{\kappa^2}{8\pi^4 n_0^2} \int_0^\infty dp_1\, dp_2\, \frac{p_1^2 p_2^2}{E_1 E_2} e^{-E_1/T} e^{-E_2/T} \int_{-1}^{+1} d\cos\theta\, w(s), \\
&= \frac{\kappa^2}{8\pi^4 n_0^2} \int_M^\infty dE_1\, dE_2\, p_1 p_2\, e^{-E_1/T} e^{-E_2/T} \int_{-1}^{+1} d\cos\theta\, w(s),
\end{aligned} \tag{24}$$

where we have used $p\,dp = E\,dE$. We now make the following change of variables

$$E_a = M(1 + x y_a), \qquad p_a = M(2x)^{1/2}\bigl(y_a + \tfrac{1}{2} x y_a^2\bigr)^{1/2}, \tag{25}$$

where $a = 1, 2$ indexes the two incoming χ particles. We then find

$$\begin{aligned}
\langle \sigma v_{\rm rel} \rangle = \frac{\kappa^2 M^4 x^3 e^{-2/x}}{4\pi^4 n_0^2} \int_0^\infty dy_1\, dy_2 \bigl(y_1 + \tfrac{1}{2} x y_1^2\bigr)^{1/2} \bigl(y_2 + \tfrac{1}{2} x y_2^2\bigr)^{1/2} \\
\times e^{-y_1} e^{-y_2} \int_{-1}^{+1} d\cos\theta\, w(s).
\end{aligned} \tag{26}$$

We can now also rewrite eq. (23) as

$$\frac{s}{4M^2} = 1 + \tfrac{1}{2} x(y_1 + y_2) + \tfrac{1}{2} x^2 y_1 y_2 - x\bigl(y_1 + \tfrac{1}{2} x y_1^2\bigr)^{1/2} \bigl(y_2 + \tfrac{1}{2} x y_2^2\bigr)^{1/2} \cos\theta, \tag{27}$$

and expand the integrand of eq. (26) in powers of x. This involves making a Taylor expansion of $w(s)$ about $s/4M^2 = 1$. All the integrals can then be done in closed form. We also need to compute n_0, both because it appears in the definition of $\langle \sigma v_{\rm rel} \rangle$ and because it is present explicitly in our basic equation (5). Making the

change of variable (25), we get

$$n_0 = \int \mathrm{d}^3 p f(E)$$

$$= \left[\kappa(2\pi x)^{-3/2} \mathrm{e}^{-1/x} T^3\right] 2\pi^{-1/2} \int_0^\infty \mathrm{d}y (1+xy)\left(y + \tfrac{1}{2}xy^2\right)^{1/2} \mathrm{e}^{-y}$$

$$= \left[\kappa(2\pi x)^{-3/2} \mathrm{e}^{-1/x} T^3\right]\left[1 + \tfrac{15}{8}x + \tfrac{105}{128}x^2 - \tfrac{315}{1024}x^3 + \mathrm{O}(x^4)\right], \tag{28}$$

$$\frac{1}{n_0^2} = \frac{8\pi^3 \mathrm{e}^{2/x}}{\kappa^2 x^3 M^6}\left[1 - \tfrac{15}{4}x + \tfrac{285}{32}x^2 - \tfrac{2115}{128}x^3 + \mathrm{O}(x^4)\right]. \tag{29}$$

We ultimately arrive at the general formula

$$\langle \sigma v_{\mathrm{rel}} \rangle = \frac{1}{M^2}\left[w - \tfrac{3}{2}(2w - w')x + \tfrac{3}{8}(16w - 8w' + 5w'')x^2 \right.$$

$$\left. - \tfrac{5}{16}(30w - 15w' + 3w'' - 7w''')x^3 + \mathrm{O}(x^4)\right]_{s/4M^2=1}, \tag{30}$$

where primes denote derivatives with respect to $s/4M^2$ (rather than s itself), and w and its derivatives are all to be evaluated at $s/4M^2 = 1$.

Eq. (30) is one of our central results. In previous calculations, $\langle \sigma v_{\mathrm{rel}} \rangle$ has usually been truncated to w/M^2. The potential importance of the term linear in x was pointed out in ref. [10]. However, the coefficient of x has never before been correctly computed, and there have been no previous attempts to compute the quadratic and cubic terms.

In all cases of interest, the dominant annihilation is into particle–antiparticle pairs. Furthermore, the interaction is typically mediated by a heavy particle (such as a W, Z, scalar quark, etc.), and so can often be treated as a fundamental four-point interaction. In this case the phase-space integrals in eq. (21) can be easily done, and $w(s)$ takes the form

$$w(s) = \sum_f \left(1 - \frac{4m_f^2}{s}\right)^{1/2}\left[C_{0f} + C_{1f}\frac{s}{4M^2} + C_{2f}\left(\frac{s}{4M^2}\right)^2\right], \tag{31}$$

where m_f is the mass of the outgoing particle, C_{0f}, C_{1f}, and C_{2f} are dimensionless constants (which in general depend on the type of the outgoing particle–antiparticle pair), and the sum is over the allowed types of outgoing particle–antiparticle pairs.

Plugging this $w(s)$ into eq. (30) yields

$$\langle \sigma v_{\rm rel} \rangle = \sum_f \frac{1}{M^2}\left(1 - \frac{m_f^2}{M^2}\right)^{1/2} \left[a_f + b_f x + c_f x^2 + d_f x^3 + \mathrm{O}(x^4)\right], \tag{32}$$

where

$$a_f = C_{0f} + C_{1f} + C_{2f}, \tag{33a}$$

$$b_f = -\tfrac{3}{2}(2C_{0f} + C_{1f}) + \tfrac{3}{4}\beta_f(C_{0f} + C_{1f} + C_{2f}), \tag{33b}$$

$$c_f = \tfrac{3}{4}(8C_{0f} + 4C_{1f} + 5C_{2f}) - \tfrac{3}{8}\beta_f(9C_{0f} + 4C_{1f} - C_{2f}) - \tfrac{15}{32}\beta_f^2(C_{0f} + C_{1f} + C_{2f}), \tag{33c}$$

$$d_f = -\tfrac{15}{16}(10C_{0f} + 5C_{1f} + 2C_{2f}) + \tfrac{15}{32}\beta_f(21C_{0f} + 5C_{1f} + 3C_{2f}) + \tfrac{15}{64}\beta_f^2(15C_{0f} + 8C_{1f} + C_{2f}) + \tfrac{105}{128}\beta_f^3(C_{0f} + C_{1f} + C_{2f}), \tag{33d}$$

and $\beta_f \equiv m_f^2/(M^2 - m_f^2)$. The sum is again over the different types of particle–antiparticle pairs into which the χ's annihilate.

It should be noted that we have expanded $(1 - 4m_f^2/s)^{1/2}$ about $s(x = 0) = 4M^2$. Clearly this is not a valid expansion if $M \simeq m_f$. In cases of interest, m_f is just the mass of an ordinary quark or lepton into which χ annihilates. Since M is at least a few GeV, we need only worry about annihilations into the tau lepton and the charm, bottom, and top quarks. For M near the mass of one of these particles, the corresponding annihilations tend to contribute very little to the total cross section. To check this we performed a calculation for $M = m_{\rm b}$ using a valid expansion for the bottom quark contribution. The result was negligible compared with the terms from annihilation into lighter quarks and leptons. There is a narrow range of values of M just above $m_{\rm b}$ (and also $m_{\rm c}$ and m_τ) for which our expansion is not applicable; we estimate that this range is no more than ~ 0.2 GeV.

It is sometimes necessary to account for the finite mass of the exchanged particle to compute an accurate cross section. If the exchange is always in the s channel (as in the case of heavy neutrino annihilation through a virtual Z) eq. (31) is modified by the inclusion of a factor of $(1 - s/M_Z^2)^{-2}$. This, of course, then changes eqs. (32) and (33). The resulting corrected expressions are rather lengthy and uninformative, and we will not reproduce them here; we did, however, use them for our numerical work. If the exchange is in the t or u channels, eq. (31) itself must be replaced with a much more cumbersome expression. Also, if one considers the case of $M > M_{\rm W}$ (which we do not), then one must also include annihilation into $\mathrm{W}^+\mathrm{W}^-$ pairs [11].

5. Photinos, higgsinos, and Majorana neutrinos

In this paper we consider two generic cases for the χ particle: Majorana fermions and Dirac fermions. Majorana fermions include supersymmetric particles such as photinos and higgsinos. More precisely, the lightest (and therefore stable) particle in minimal supersymmetric extensions of the standard model is usually a linear combination of the fermionic partners of neutral gauge and Higgs bosons [12]. There are large regions of parameter space where it is possible to treat this lightest particle as a pure photino or higgsino, and ignore mixing effects. These cases of pure photino and higgsino are those we will analyze in detail in this paper. (More general cases could be handled by computing the appropriate values of C_{0f}, C_{1f}, and C_{2f}.) In this section we will also consider the case of a Majorana neutrino as the χ particle, although there is no particular reason for a Majorana neutrino to be stable. In sect. 6 we will consider a Dirac neutrino; the same caveat applies.

The effective lagrangian for a Majorana fermion χ to interact with an ordinary quark or lepton f can be written as

$$\mathscr{L}_{\text{eff}} = \sum_f \bar{\chi}\gamma^\mu\gamma_5\chi\bar{f}\gamma_\mu\left(V_f + A_f\gamma_5\right)f\,, \tag{34}$$

where χ is a Majorana spinor field and f is a Dirac spinor field. Using the usual techniques* we evaluate the reduced matrix element and extract C_{0f}, C_{1f}, and C_{2f}. We find

$$C_{0f} = \frac{2}{3\pi}M^4\left[7\xi_f A_f^2 - 2\xi_f V_f^2\right], \tag{35a}$$

$$C_{1f} = \frac{4}{3\pi}M^4\left[-\left(2+2\xi_f\right)A_f^2 - \left(2-\xi_f\right)V_f^2\right], \tag{35b}$$

$$C_{2f} = \frac{8}{3\pi}M^4\left[A_f^2 + V_f^2\right], \tag{35c}$$

where $\xi_f \equiv m_f^2/M^2$ and m_f is the mass of f. These expressions must be inserted in eq. (33), and the resulting $\langle\sigma v_{\text{rel}}\rangle$ summed over quarks and leptons f with mass less than the mass of the χ particle, M.

For pure photinos, we have

$$V_f = 0\,, \qquad A_f = 2\pi\alpha Q_f^2/M_{\text{sf}}^2\,, \tag{36}$$

where α is the fine structure constant, Q_f is the electric charge of f in units of the

* We can pretend that χ is a Dirac field provided that the resulting $|\mathscr{T}|^2$ is multiplied by a factor of four. See ref. [13] for details.

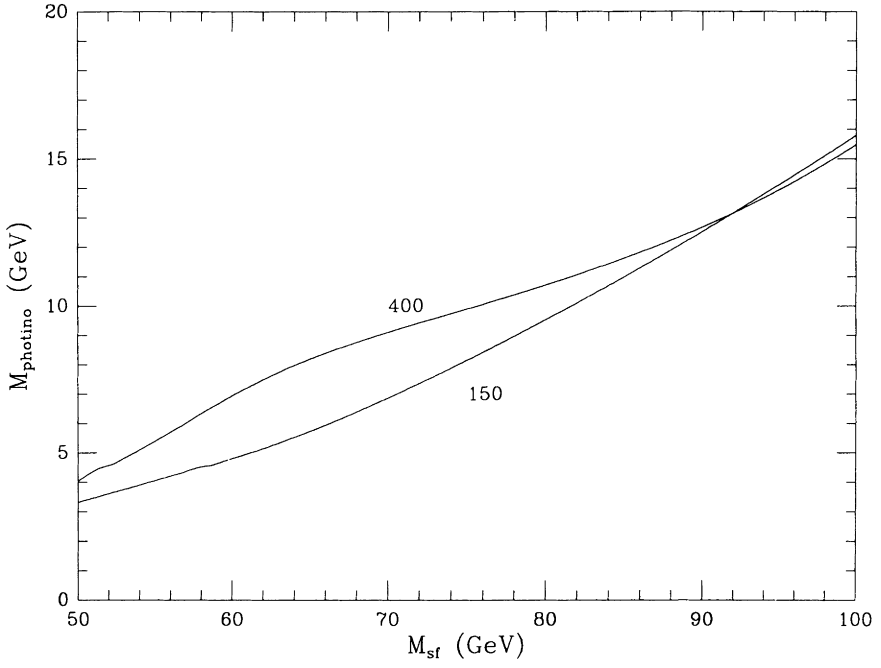


Fig. 4. The photino mass needed to have $\Omega_\chi h^2 = 0.25$ versus the (common) scalar quark and lepton mass for our two assumptions about the quark–hadron transition.

proton charge, and M_{sf} is the mass of the scalar partner of the quark or lepton f*. In our numerical work we did not correct eq. (31) for finite scalar masses. (The exchange is in the t channel, which leads to very messy expressions.) Furthermore, the corrections, which are of order M^2/M_{sf}^2, are no larger than those needed to correct for the mixing of the photino with other neutral gauge/Higgs fermions. Because we have not included these model-dependent effects, our results should be regarded as illustrative rather than definitive.

In fig. 4 and table 1 we have displayed the photino mass and the (common) scalar mass necessary to have $\Omega_\chi h^2 = 0.25$. We show the results for this value since $\Omega = 1$ is favored by inflation, while $h \simeq 0.5$ is favored by numerical simulations of galaxy formation when $\Omega = 1$. We show the results for the two cases of the quark–hadron transition depicted in figs. 1 and 2.

For both pure higgsinos and Majorana neutrinos, annihilation occurs through Z exchange in the s channel. In this case we have

$$V_f = \sqrt{\tfrac{1}{2}\lambda}\, G\left(T_f^{3L} - 2Q_f \sin^2\theta_W\right), \qquad A_f = \sqrt{\tfrac{1}{2}\lambda}\, G\left(-T_f^{3L}\right), \tag{37}$$

* Actually, there are separate scalar partners for the right- and left-handed parts of each quark and lepton. For simplicity we assume that all scalars have a common mass M_{sf}, and that there is negligible mixing between them.

where G is Fermi's constant, T_f^{3L} is the weak isospin of the left-handed part of the quark or lepton f, Q_f is the electric charge of f (in units of the proton charge), θ_W is the weak mixing angle, and λ is a numerical factor between zero and one. For higgsinos, $\lambda = \cos^2(2\alpha)$, where $\tan\alpha = V_2/V_1$ is the ratio of the two Higgs expectation values that appear in all supersymmetric extensions of the standard model. For higgsinos, there may also be contributions due to scalar exchange. We ignore these contributions (although they could alter our results if the scalar quark and lepton masses are of order M_Z or smaller) since the enlargement of the parameter space makes the presentation of the results cumbersome (see ref. [12]). For Majorana neutrinos, λ is some combination of Cabibbo-like mixing angles. We corrected eqs. (32) and (33) for the finite value of M_Z.

We show our results in two ways. First, in fig. 5 and table 2, we fix $\lambda = 1$, and display $\Omega_\chi h^2$ against the mass of the Majorana neutrino (or higgsino). Then, in fig. 6, we fix $\Omega_\chi h^2 = 0.25$, and display the needed value of λ against the mass.

6. Dirac neutrinos

In this section we consider a heavy, stable Dirac neutrino as the dark-matter particle. If such a particle is stable, it is likely that it carries a conserved quantum number and does not mix with ordinary neutrinos. The effective lagrangian for a Dirac fermion χ to interact with an ordinary quark or lepton f can be written as

$$\mathscr{L}_{\text{eff}} = \sum_f \bar{\chi}\gamma^\mu(1-\gamma_5)\chi \bar{f}\gamma_\mu(V_f + A_f\gamma_5)f, \tag{38}$$

where χ and f are both Dirac spinor fields. The constants V_f and A_f are again given by eq. (37), with $\lambda = 1$ if there is no mixing of the heavy neutrino χ with ordinary neutrinos. Using the usual techniques* we find

$$C_{0f} = \frac{1}{12\pi} M^4 \left[5\xi_f A_f^2 - \xi_f V_f^2\right], \tag{39a}$$

$$C_{1f} = \frac{1}{6\pi} M^4 \left[-(1+4\xi_f)A_f^2 - (1-2\xi_f)V_f^2\right], \tag{39b}$$

$$C_{2f} = \frac{2}{3\pi} M^4 \left[A_f^2 + V_f^2\right], \tag{39c}$$

where $\xi_f \equiv m_f^2/M^2$. These expressions must be inserted in eq. (33) and then corrected for the finite mass of the Z.

* Note that, when χ is a Dirac fermion, our σ is *one half* of the usual spin-averaged, particle–antiparticle annihilation cross section (see the appendix).

TABLE 1
The (common) scalar quark and lepton mass needed to have $\Omega_\chi h^2 = 0.25$ for a given photino mass, for our two assumptions about the quark–hadron transition.

$M_{\tilde{\gamma}}$ (GeV)	$T_c \simeq 150$ MeV M_{sf} (GeV)	$T_c \simeq 400$ MeV M_{sf} (GeV)
3.0	47.8	46.2
4.0	54.7	49.9
5.0	61.2	53.7
6.0	66.3	57.0
7.0	70.5	60.2
8.0	74.4	64.1
9.0	78.1	69.4
10.0	81.6	75.6
15.0	97.6	98.5
20.0	112	112

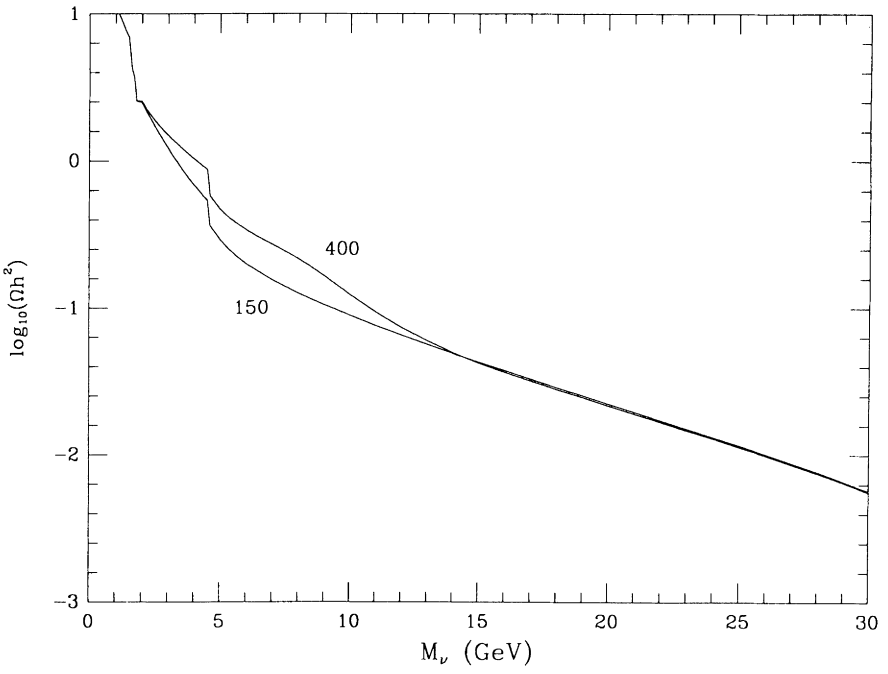


Fig. 5. The value of $\Omega_\chi h^2$ for a Majorana neutrino (or higgsino) with $\lambda = 1$ versus its mass for our two assumptions about the quark–hadron transition.

TABLE 2
The value of $\Omega_\chi h^2$ for a Majorana neutrino (or higgsino with $\lambda = 1$) for a given value of its mass, for our two assumptions about the quark–hadron transition.

M_{ν_M} (GeV)	$T_c \simeq 150$ MeV $\Omega_\chi h^2$	$T_c \simeq 400$ MeV $\Omega_\chi h^2$
3.0	1.22	1.50
4.0	0.685	1.03
5.0	0.291	0.472
6.0	0.200	0.341
7.0	0.155	0.271
8.0	0.126	0.216
9.0	0.106	0.165
10.0	0.0893	0.124
15.0	0.0431	0.0425
20.0	0.0224	0.0217
25.0	0.0117	0.0114
30.0	0.00569	0.00560

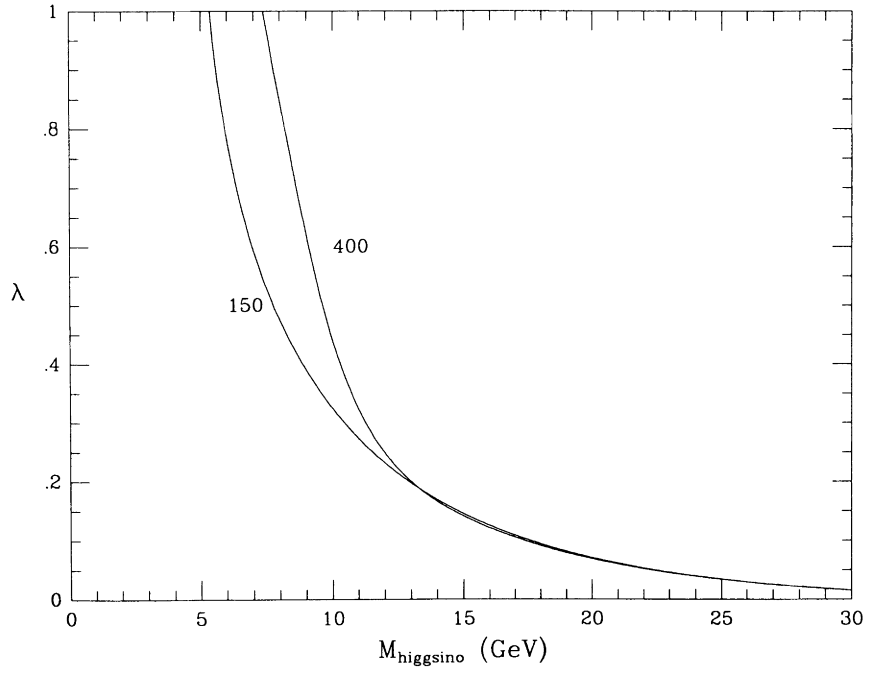


Fig. 6. The value of λ versus the higgsino mass for our two assumptions about the quark–hadron transition.

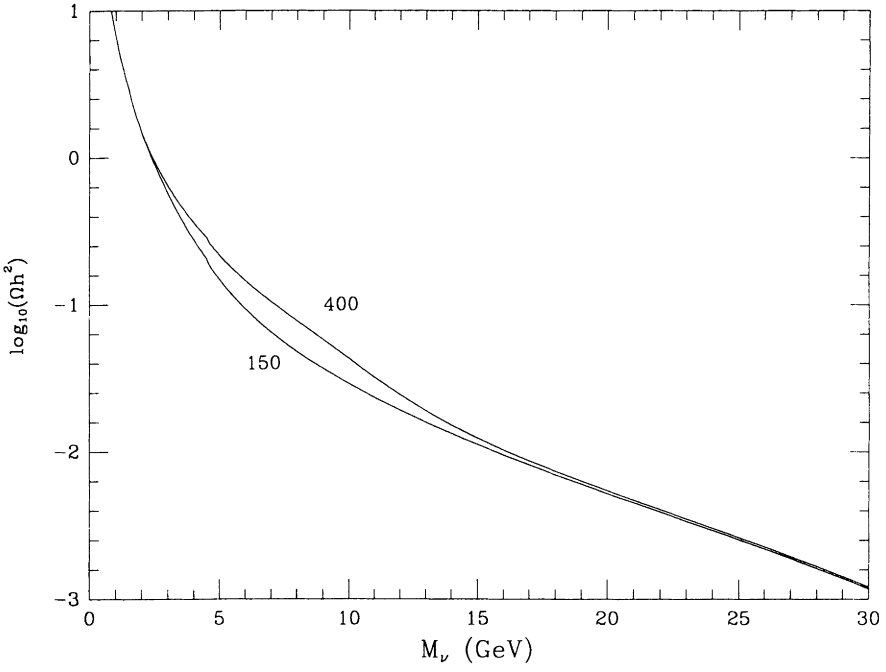


Fig. 7. The value of $\Omega_\chi h^2$ for a Dirac neutrino versus its mass for our two assumptions about the quark–hadron transition.

TABLE 3
The value of $\Omega_\chi h^2$ for a Dirac neutrino for a given value of its mass, for our two assumptions about the quark–hadron transition.

M_{ν_D} (GeV)	$T_c \simeq 150$ MeV $\Omega_\chi h^2$	$T_c \simeq 400$ MeV $\Omega_\chi h^2$
2.0	1.48	1.49
3.0	0.573	0.645
4.0	0.275	0.362
5.0	0.148	0.216
6.0	0.0933	0.145
7.0	0.0649	0.105
8.0	0.0479	0.0778
9.0	0.0368	0.0581
10.0	0.0290	0.0431
15.0	0.0112	0.0124
20.0	0.00521	0.00545
25.0	0.00254	0.00262

In fig. 7 and table 3, we have displayed the value of $\Omega_\chi h^2$ against the Dirac neutrino mass for the two cases of the quark–hadron transition we have discussed.

7. Conclusions

We have presented equations whose numerical solution gives accurate estimates of the mass density of hypothetical stable particles remaining after the Big Bang. We believe that all effects at the level of one percent or larger have been included. The greatest uncertainty is from our imprecise knowledge of the equation of state of quarks and hadrons at temperatures of a few hundred million electron volts.

We presented our numerical results for pure photinos, pure higgsinos, Majorana neutrinos, and Dirac neutrinos, using our two different assumptions about the quark–hadron transition. A direct consequence of these calculations is a lower limit on the mass of each of the particles we considered. For photinos, we find that $M > 4.80$ GeV (6.93 GeV) for a quark–hadron transition near 150 MeV (400 MeV) assuming that $M_{\rm sf} > 60$ GeV [14] and $\Omega_\chi h^2 < 0.25$. This is comparable to previous results [10, 15, 12]. Clearly the largest uncertainty comes from the details of the quark–hadron transition. For Majorana neutrinos (with $\lambda = 1$), $M > 5.34$ GeV (7.37 GeV), again comparable to previous results [16, 17]. For higgsinos, there is no model-independent lower limit on the mass, since it depends on the value of λ and on the scalar quark and lepton masses (which we took to be very large). Finally, for Dirac neutrinos, $M > 4.15$ GeV (4.70 GeV), which is somewhat higher than previous estimates [1–3, 17–19], due either to differences in the cross section itself or the method of thermal averaging. Our calculations also impact on any computations of astrophysical or laboratory signatures of dark-matter candidates. Our general technique should be useful when and if more concrete information on the nature of dark matter becomes available.

The work of M.S. and R.W. was supported in part by NSF Grant No. PHY-86-14185. M.S. also acknowledges the receipt of a Research Fellowship from the Alfred P. Sloan Foundation. The work of K.A.O. was supported in part by DOE Grant No. DE-AC02-83ER-40105, and by a Presidential Young Investigator Award.

Appendix

SPIN AND PARTICLE–ANTIPARTICLE AVERAGING

Suppose the χ particle comes in κ spin types, including antiparticles, so that Dirac fermions would have $\kappa = 4$ and Majorana fermions would have $\kappa = 2$. Then if n_i is the number density of the ith spin type, the contribution of annihilations to dn_i/dt can be written as

$$\frac{dn_i}{dt} = -\sum_{j=1}^{\kappa} \langle \sigma_{ij} v_{\rm rel} \rangle n_i n_j , \tag{40}$$

where σ_{ij} is the cross section for type i to annihilate with type j, $v_{\rm rel}$ is the relative velocity, and $\langle\ \rangle$ implies the thermal average of sect. 4.

It should be stressed that this formula is correct without any additional factors of two. For $i \neq j$, we can think of type j impinging on type i, and it should be clear that eq. (40) is correct as it stands. For $i = j$, there is an extra factor of two since two type-i particles are destroyed in each annihilation, but this is balanced by an extra factor of one half introduced to avoid double counting the number of interactions.

Summing eq. (40) over i we get

$$\frac{dn}{dt} = -\sum_{i,j=1}^{\kappa} \langle \sigma_{ij} v_{\rm rel} \rangle n_i n_j , \tag{41}$$

where $n \equiv \sum_{i=1}^{\kappa} n_i$ is the total number density of the χ particles. The equipartition theorem tells us that $n_i = n/\kappa$ for each spin type i, so we get

$$\begin{aligned} \frac{dn}{dt} &= -\frac{1}{\kappa^2} \sum_{i,j=1}^{\kappa} \langle \sigma_{ij} v_{\rm rel} \rangle n^2 \\ &= -\langle \sigma v_{\rm rel} \rangle n^2 , \end{aligned} \tag{42}$$

where we have defined the spin and particle–antiparticle averaged cross section

$$\sigma \equiv \frac{1}{\kappa^2} \sum_{i,j=1}^{\kappa} \sigma_{ij} . \tag{43}$$

It is this cross section which appears throughout the text.

We demonstrate for Dirac and Majorana fermions. A Dirac fermion has $\kappa = 4$: two spin states for each of the fermion f and the antifermion $\bar{f}$. If f carries a conserved quantum number, then $\sigma_{ff} = \sigma_{\bar{f}\bar{f}} = 0$. Furthermore, $\sigma_{f\bar{f}} = \sigma_{\bar{f}f}$, and $\sigma_{\bar{f}f}$ represents four different cross sections, corresponding to the two possible spin states for each of f and $\bar{f}$. Thus we get

$$\begin{aligned} \sigma &= \tfrac{1}{16} \sum_{\rm spins} (\sigma_{\bar{f}f} + \sigma_{f\bar{f}} + \sigma_{ff} + \sigma_{\bar{f}\bar{f}}) \\ &= \tfrac{1}{8} \sum_{\rm spins} \sigma_{\bar{f}f} \\ &= \tfrac{1}{2} \bar{\sigma}_{\bar{f}f} , \end{aligned} \tag{44}$$

where $\bar{\sigma}_{\bar{f}f}$ is the usual spin-averaged cross section.

Most papers about Dirac fermions as dark matter are written in terms of the density of just the fermions, $n_f = \frac{1}{2}n$. Thus

$$\frac{dn}{dt} = 2\frac{dn_f}{dt} = -\tfrac{1}{2}\langle \bar{\sigma}_{ff} v_{\rm rel}\rangle (2n_f)^2, \tag{45}$$

so that

$$\frac{dn_f}{dt} = -\langle \bar{\sigma}_{ff} v_{\rm rel}\rangle n_f^2, \tag{46}$$

which is the usual form of the rate equation for Dirac fermions in the literature. (After solving for n_f, the answer must be multiplied by two to get the relic density of the fermions *and* antifermions.)

For Majorana fermions there are no antiparticles, so $\kappa = 2$. In this case we have

$$\frac{1}{\kappa^2}\sum_{i,j=1}^{\kappa}\sigma_{ij} = \tfrac{1}{4}\sum_{\rm spins}\sigma_{ff} = \bar{\sigma}_{ff} \tag{47}$$

where $\bar{\sigma}_{ff}$ is the usual spin-averaged cross section. This again reproduces the equation used in earlier work. However, one should be aware of extra factors of two involved in calculating cross sections with Majorana spinors. These are thoroughly explained in ref. [13].

References

[1] P. Hut, Phys. Lett. B69 (1977) 85
[2] B. Lee and S. Weinberg, Phys. Rev. Lett. 39 (1977) 165
[3] M.I. Vysotskiĭ, A.D. Dolgov, and Ya.B. Zel'dovich, JETP Lett. 26 (1977) 188
[4] K. Griest and D. Seckel, Nucl. Phys. B283 (1987) 681; (E) B296 (1988) 1034
[5] K.A. Olive, D.N. Schramm, and G. Steigman, Nucl. Phys. B180 (1981) 497
[6] J. Yang, M.S. Turner, G. Steigman, D.N. Schramm, and K.A. Olive, Astrophys. J. 281 (1984) 493
[7] K.A. Olive, Nucl. Phys. B190 (1981) 483
[8] S. Gottlieb, W Liu, D. Toussaint, R.L. Renken, and R.L. Sugar, Phys. Rev. Lett. 59 (1987) 1513
[9] J. Bernstein, L.S. Brown, and G. Feinberg, Phys. Rev. D32 (1985) 3261
[10] H. Goldberg, Phys. Rev. Lett. 50 (1983) 1419
[11] K. Enqvist, K. Kainulainen, and J. Maalampi, Nucl. Phys. B, to be published
[12] J. Ellis, J.S. Hagelin, D.V. Nanopoulos, K.A. Olive, and M. Srednicki, Nucl. Phys. B238 (1984) 453
[13] H.E. Haber and G.L. Kane, Phys. Rep. 117 (1985) 75
[14] R.M. Barnett, H.E. Haber, and G.L. Kane, Nucl. Phys. B267 (1986) 625
[15] L.M. Krauss, Nucl. Phys. B227 (1983) 556
[16] L.M. Krauss, Phys. Lett. B128 (1983) 37
[17] E.W. Kolb and K.A. Olive, Phys. Rev. D33 (1986) 1202; (E) 34 (1986) 2531
[18] K. Sato and M. Kobayashi, Prog. Theor. Phys. 58 (1977) 1775
[19] D.A. Dicus, E.W. Kolb, and V.L. Teplitz, Phys. Rev. Lett. 39 (1977) 169

ERRATUM

Mark Srednicki, Richard Watkins and Keith A. Olive

Calculations of Relic Densities in the Early Universe, Nucl. Phys. B 310 (1988) 693.

In our analysis of relic densities of higgsinos and heavy neutrinos, we accounted for the finite mass of the Z^0 by multiplying the right hand side of eq. (31) by $(1 - s/M_z)^{-2}$. In addition, there are modifications to eq. (35) for higgsinos and Majorana neutrinos, and to eq. (39) for Dirac neutrinos. Eq. (35) should read

$$C_{0f} = \frac{2}{3\pi} M^4 \left[7\xi_f A_f^2 - 2\xi_f V_f^2 \right], \tag{35a}$$

$$C_{1f} = \frac{4}{3\pi} M^4 \left[-(2 + 2\xi_f + 12\xi_f \eta) A_f^2 - (2 - \xi_f) V_f^2 \right], \tag{35b}$$

$$C_{2f} = \frac{8}{3\pi} M^4 \left[(1 + 12\xi_f \eta^2) A_f^2 + V_f^2 \right], \tag{35c}$$

where $\eta \equiv M^2/M_z^2$. Eq. (39) should read

$$C_{0f} = \frac{1}{12\pi} M^4 \left[5\xi_f A_f^2 - \xi_f V_f^2 \right], \tag{39a}$$

$$C_{1f} = \frac{1}{6\pi} M^4 \left[-(1 + 4\xi_f + 12\xi_f \eta) A_f^2 - (1 - 2\xi_f) V_f^2 \right], \tag{39b}$$

$$C_{2f} = \frac{2}{3\pi} M^4 \left[(1 + 6\xi_f \eta^2) A_f^2 + V_f^2 \right], \tag{39c}$$

where again $\eta \equiv M^2/M_z^2$. These corrections are due to the $p^\mu p^\nu / M_z^2$ terms in the numerator of the Z^0 propagator [1,2].

For higgsinos and Majorana neutrinos, we then find that $\langle \sigma v_{\rm rel} \rangle$ at zero temperature is given by

$$\begin{aligned} \langle \sigma v_{\rm rel} \rangle_{x=0} &= \frac{1}{M^2} (1 - 4\eta)^{-2} \sum_f (1 - \xi_f)^{1/2} (C_{0f} + C_{1f} + C_{2f}) \\ &= \frac{2}{\pi} \sum_f (1 - \xi_f)^{1/2} m_f^2 A_f^2 \end{aligned}$$

where A_f is given by eq. (37). In particular, all dependence on M_z cancels. However, these formulas do not take into account the finite width $\Gamma_z \simeq 2.5$ GeV of the Z^0, and, therefore, do not apply in the very narrow range $|M - \frac{1}{2}M_z| \lesssim \Gamma_z$.

The corrections to eqs. (35), (39) make negligible changes in the values of $\Omega_\chi h^2$ for Dirac neutrinos with mass $M \leq 25$ GeV; we find $\Delta\Omega_\chi/\Omega_\chi < 10^{-3}$. For Majorana neutrinos (or higgsinos with $\lambda = 1$) with mass $M \leq 30$ GeV, the largest percentage correction is $\Delta\Omega_\chi/\Omega_\chi = 0.035$ for $M = 30$ GeV and $T_c \simeq 150$ MeV. These changes would be barely perceptible in our figures. For $M \lesssim 15$ GeV, the changes are smaller than the uncertainty induced by imprecise knowledge of the quark–hadron phase transition.

REFERENCES

[1] L. Bergstrom, Phys. Rev. D, in press.
[2] K. Griest, Phys. Rev. D 39 (1989) 3802.

Cosmic-Ray Antiprotons as a Probe of a Photino-Dominated Universe

Joseph Silk
Astronomy Department, University of California, Berkeley, California 94720, and Institute for Theoretical Physics, University of California, Santa Barbara, California 93106

and

Mark Srednicki
Physics Department, University of California, Santa Barbara, California 93106
(Received 8 June 1984)

Observational tests of the hypothesis that the universe is flat and dominated by dark matter in the form of massive photinos include the production of significant fluxes of cosmic rays and gamma rays in our galactic halo. Specification of the cosmological photino density and the masses of scalar quarks and leptons determines the present annihilation rate. The predicted number of low-energy cosmic-ray antiprotons is comparable to the observed flux.

PACS numbers: 94.40.Cn, 11.30.Pb, 14.80.Pb, 98.80.Bp

One of the most important predictions of inflationary cosmology is that the mean density Ω, in units of the critical Einstein–de Sitter value $\rho_{cr} = 3H_0^2/8\pi G$ (where H_0 is Hubble's constant), is unity.[1] Primordial nucleosynthesis constraints[2] limit the baryonic contribution Ω_b to be at most 0.2, and several nonbaryonic weakly interacting particle candidates have been proposed to make up the discrepancy. Viable scenarios for galaxy formation and clustering appear to require that the nonbaryonic matter be cold at epochs when the horizon first contained a mass comparable to that of a galaxy.[3] One of the more plausible particle candidates is the stable massive photino. We show here that there are noteworthy observational tests of the hypothesis that there is a critical cosmological density of gigaelectronvolt photinos.

Our derivation proceeds as follows. Specification of Ω suffices to fix the photino annihilation rate Γ, and specification of the various scalar quark and lepton masses then fixes the photino mass $m_{\tilde{\gamma}}$.[4,5] The low-energy limit of Γ then yields the annihilation rate in our galactic halo, which is assumed to consist of photinos. Accretion of photinos into galactic halos is unavoidable in a photino-dominated universe. We then find that observable annihilation products of ~ 3-GeV photinos in our halo include γ rays, cosmic-ray positrons, and, most significantly, low-energy cosmic-ray antiprotons. Following Gunn *et al.*[6] and Stecker,[7] Sciama[8] proposed that gamma rays would be produced in observable amounts by photino annihilations. Zel'dovich *et al.*[9] examined cosmic-ray production as a constraint on the mass of a stable, heavy, neutral lepton, but the possibility of antiparticle production at a significant level has not previously been considered.

To commence, we adopt a minimal supersymmetric model in which the lightest supersymmetric particle is a photino (with a negligible admixture of Higgs fermion). The thermally averaged annihilation cross section for $\tilde{\gamma}\tilde{\gamma} \to f\bar{f}$, where f is any quark or lepton, is[4,5]

$$\langle \sigma v \rangle = 8\pi\alpha^2 Q_f^4 r_f \left[\frac{m_f^2}{m_{sf1}^4} + \left(\frac{2T}{m_{\tilde{\gamma}}} \right) \left(\frac{m_{\tilde{\gamma}}^2}{m_{sf2}^4} - \frac{m_f^2}{m_{sf3}^4} \right) \right], \tag{1}$$

where Q_f and m_f are the charge and mass of f, $m_{sf1,2,3}$ are various functions[5] of the masses of the two scalar partners of f, and $r_f = (1 - m_f^2/m_{\tilde{\gamma}}^2)^{1/2}$. To obtain the total annihilation cross section, one should sum over all quarks and leptons with $m_f < m_{\tilde{\gamma}}$. By following standard methods,[10] this cross section can be used to compute a relic photino mass density. Requiring that this equal the critical density (for $H_0 = 50$ km/s·Mpc) gives a value for $\langle \sigma v \rangle$ at the freeze-out temperature T_*, when the annihilation rate per particle was equal to the expansion rate.[11] (For values of $\langle \sigma v \rangle$ of interest here, $T_* \approx \frac{1}{20} m_{\tilde{\gamma}}$.)[4,5] This needed value of $\langle \sigma v \rangle_*$ is plotted versus $m_{\tilde{\gamma}}$ as the solid line in Fig. 1, with the subsequent heating of the relic photon background by other species of particles[12] taken into account. We see that we need a total $\langle \sigma v \rangle$ of about 10^{-26} cm^3/s. Also plotted in Fig. 1 for $m_{sf1,2,3} = 50$ GeV is $\langle \sigma v \rangle_*$ as given by Eq. (1) (dashed line). The intersection of the two curves shows that scalar masses of 50 GeV and a photino mass of ~ 2.8 GeV would yield a critical density of relic photinos. Decreasing (increasing) the scalar masses would move the dashed curve up (down).

Given $m_{\tilde{\gamma}}$ and m_{sf1}, we can compute the annihila-

© 1984 The American Physical Society

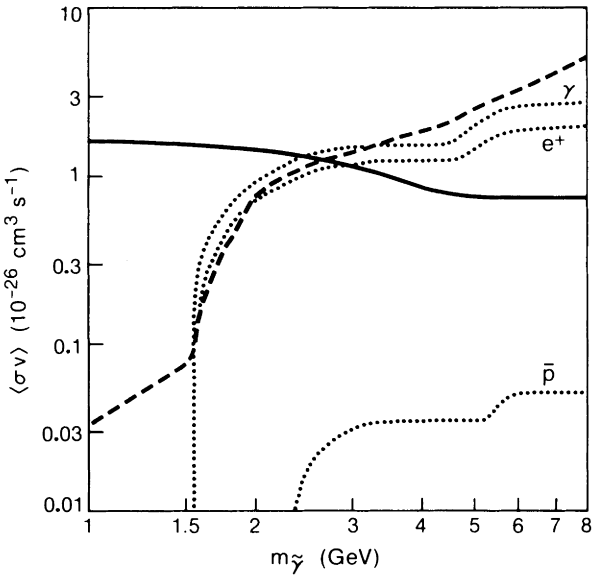


FIG. 1. The photino annihilation cross section necessary to have $\Omega = 1$ with $h = \frac{1}{2}$, as a function of photino mass (solid line); the annihilation cross section at the freeze-out temperature for 50-GeV quark and lepton scalars (dashed line); and the production cross sections for antiprotons, positrons, and gamma rays with 50-GeV scalars (dotted lines).

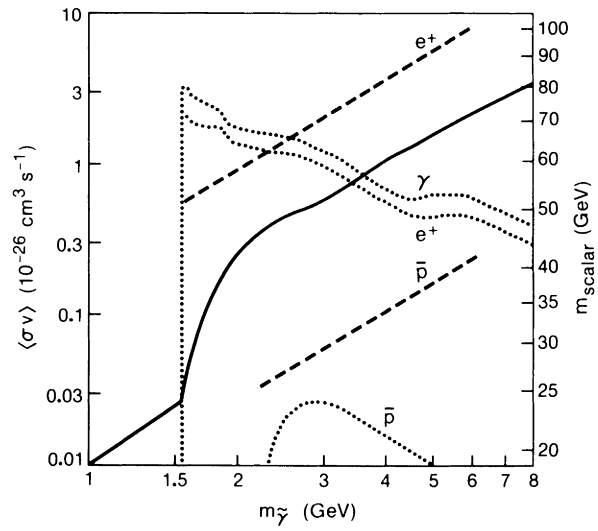


FIG. 2. The scalar quark and lepton masses (taken to be equal) as a function of photino mass needed to have $\Omega = 1$ with $h = \frac{1}{2}$ (solid line); the antiproton, positron, and gamma-ray production cross sections with these scalar masses (dotted lines); and the cross sections necessary to reproduce the observed antiproton and positron fluxes with our halo parameters (dashed lines).

tion rate today. For $m_{\tilde{\gamma}} \lesssim 1.5$ GeV, the annihilation rate is negligible. For 1.8 GeV $\lesssim m_{\tilde{\gamma}} \lesssim 5$ GeV, the primary annihilation products are τ leptons and charmed quarks. For $m_{\tilde{\gamma}} \gtrsim 5$ GeV, the bottom-quark mass, bottom quarks are also produced. These particles will then decay to ordinary particles. The annihilation products of most interest are e^+, $\bar{p}$, and γ. The total number of each type of particle produced per $\tilde{\gamma}\tilde{\gamma}$ annihilation can be estimated from experimental results[13] on $e^+e^- \to \tau^+\tau^-$, $\bar{c}c$, and $\bar{b}b$. We estimate the following yields of stable particles for each of the relevant channels:

$$\tilde{\gamma}\tilde{\gamma} \to \tau^+\tau^- \to 1.5(e^+e^-) + 0(\bar{p}p) + 1\gamma + 5.5\nu,$$
$$\tilde{\gamma}\tilde{\gamma} \to \bar{c}c \to 4(e^+e^-) + 0.2(\bar{p}p) + 7\gamma + 22\nu, \quad (2)$$
$$\tilde{\gamma}\tilde{\gamma} \to \bar{b}b \to 7.5(e^+e^-) + 0.3(\bar{p}p) + 13\gamma + 41\nu.$$

With these numbers, the total production cross sections for e^+, $\bar{p}$, and γ are plotted in Figs. 1 and 2. These results are sensitive to the *relative* masses of scalar quarks and leptons. Lighter scalar quarks and heavier scalar leptons would cause the e^+, $\bar{p}$, and γ lines in Fig. 2 to move up, but not by more than a factor of 3.

The energy distributions of the particles are harder to estimate. We guess that each γ has about twice the energy of each $e^\pm$ or ν ($\pi^0 \to \gamma\gamma$ vs $\pi^\pm \to e^\pm\nu\nu\nu$), and that the average kinetic energy of a p or $\bar{p}$ is about the same as a typical γ energy. Decay products of τ leptons (which account for 70% of all annihilation events below the b quark threshold of ~ 5 GeV and 58% above it, if all scalar masses are equal) have higher average energies than c or b quark decay products; we estimate $\epsilon_\gamma \simeq 0.2 m_{\tilde{\gamma}}$ for γ's from τ decay. Decay products of c and b quarks have a much softer spectrum; we estimate $\epsilon_\gamma \simeq 0.09 m_{\tilde{\gamma}}$ for γ's from c quark decay and $\epsilon_\gamma \simeq 0.05 m_{\tilde{\gamma}}$ for γ's from b quark decay.

Consider first the isotropic gamma-ray background. We integrate the volume emissivity $\epsilon_\gamma n_{\tilde{\gamma}}^2 \langle\sigma v\rangle_\gamma$ (where $n_{\tilde{\gamma}}$ is the average photino number density and $\langle\sigma v\rangle_\gamma$ is the production cross section for a single γ) over the red-shift range $1 \leq 1+z \leq \epsilon_\gamma/\epsilon_\gamma^0$, where $\epsilon_\gamma \simeq f_\gamma m_{\tilde{\gamma}}$, $f_\gamma \simeq 0.2$, and ϵ_γ^0 is the observed, red-shifted gamma-ray energy.

The resulting isotropic γ-ray background flux is[6,7]

$$F_\gamma = \int_0^{\epsilon_\gamma/\epsilon_\gamma^0} \frac{c}{4\pi H_0} dz\,(1+z)^{1/2} \rho_{cr}^2 \langle\sigma v\rangle_\gamma m_{\tilde{\gamma}}^{-2}$$
$$= 6.3\times 10^{-10} \langle\sigma v\rangle_{26} \Omega_{\tilde{\gamma}}^2 h^3 m_3^{-2} (\epsilon_\gamma/\epsilon_\gamma^0)^{3/2}$$
$$(\text{cm}^2\ \text{s}\ \text{sr})^{-1}, \quad (3)$$

where $\langle\sigma v\rangle_{26}=\langle\sigma v\rangle_\gamma/(10^{-26}\ \mathrm{cm^3\ s^{-1}})$ and $m_3=m_{\tilde\gamma}/(3\ \mathrm{GeV})$. This is some four orders of magnitude (with $m_{\tilde\gamma}\approx 3$ GeV) below the observed isotropic γ-ray flux above 100 MeV, which amounts to about[14] $0.9\times10^{-5}\ (\mathrm{cm^2\ s\ sr})^{-1}$.

However, cold dark matter cannot be prevented from accreting into our galactic halo. The most natural assumption is that our halo is made of photinos. The enhancement in flux from our halo relative to the isotropic background amounts to a factor $\sim(\rho_h/\rho)^2(a/ct_0)$, where ρ_h is the halo density within the uniform-density halo core, of radius a. For a more reliable estimate,[6] the halo density distribution is specified by an analytic fit to a self-consistent dynamical model, defined by measured rotation velocity $v_{200}\equiv v/(200\ \mathrm{km\ s^{-1}})$ and core radius $a_{10}=a/(10\ \mathrm{kpc})$, with ρ_h given by

$$\rho_h(r)=(v^2/4\pi Ga^2)(1+r^2/a^2)^{-1}; \tag{4}$$

hence the photino density in the halo core is

$$n_0=0.09v_{200}^2a_{10}^{-2}m_3^{-1}\ \mathrm{cm^{-3}}.$$

More sophisticated models[15] yield a similar density for $\rho_h(r)$ in the solar neighborhood. The halo γ-ray flux can now be expressed as

$$F_\gamma^h=\langle\sigma v\rangle_\gamma n_0^2a\int_0^\infty dx\,(x^2+a^2)^{-2}/4\pi$$
$$=1.7\times10^{-7}\langle\sigma v\rangle_{26}v_{200}^4a_{10}^{-3}m_3^{-2}$$
$$(\mathrm{cm^2\ s\ sr})^{-1}. \tag{5}$$

This cannot exceed the observed high-latitude γ-ray flux, whence we infer the weak bound

$$\langle\sigma v\rangle_{26}\lesssim 53m_3^2a_{10}^3v_{200}^{-4}. \tag{6}$$

A photino with mass ~ 3 GeV yields an integral γ-ray flux above 100 MeV that is at least an order of magnitude below that observed.

However, a substantial fraction of the γ-ray energy flux will be in rather energetic gamma rays (from the τ decay channel) with energy $\epsilon_\gamma\approx 0.2m_{\tilde\gamma}$. This may lead to a more significant prediction if $m_{\tilde\gamma}\gtrsim 3$ GeV, since the observed isotropic gamma-ray spectrum appears to be very steep above 100 MeV. In fact, if the number flux decreases as E^{-2} per unit of energy interval, as the data[14] suggest, over the range 100 MeV $\lesssim\epsilon\lesssim 1$ GeV, we infer that

$$\langle\sigma v\rangle_{26}\lesssim 2.1a_{10}^3v_{200}^{-4} \tag{7}$$

for $1.5\lesssim m_{\tilde\gamma}\lesssim 5$ GeV.

Cosmic-ray observations provide a more important constraint on halo photinos than do γ-ray observations. Cosmic rays are trapped in the halo by diffusion against magnetic field irregularities, and thus the observable flux is boosted by a considerable factor. The cosmic-ray diffusion time in the galactic disk is known to be of order 10^7 yr, but is likely to be considerably longer in the halo. The diffusion coefficient in most cosmic-ray confinement models[16] lies in the range $(1\text{–}10)\times10^{28}$ $\mathrm{cm^2\ s^{-1}}$. Writing $K=10^{29}K_{29}\ \mathrm{cm^2\ s^{-1}}$, we infer a confinement time scale $t_h\sim(a^2/3)K\sim10^8a_{10}^2K_{29}^{-1}$ yr. Hence the annihilation fluxes of cosmic rays are enhanced over the gamma-ray flux by a factor of order $(\langle\sigma v\rangle/\langle\sigma v\rangle_\gamma)ct_h/a$. We focus on the cosmic-ray antiprotons and positrons to obtain the most sensitive limits. Observations of p and e^+ in the range 0.1–1 GeV are subject to uncertainty because of solar modulation; however, the $\bar p/p$ and e^+/e^- ratios are reasonably well known.

There is some indication that primary cosmic-ray antiprotons may already have been observed below the kinematic threshold for secondary production of about 2 GeV. In the demodulated energy range of approximately 0.6–1.2 GeV, the observed $\bar p$ flux is[17]

$$F_{\bar p}\approx 3\times10^{-6}\ (\mathrm{cm^2\ s\ sr})^{-1}$$
$$(0.6\lesssim E\lesssim 1.2\ \mathrm{GeV}). \tag{8}$$

A secondary origin cannot be excluded for the antiprotons below 1 GeV, but the models become extremely contrived.[18] The antiproton flux predicted by halo photino annihilation is [with $\langle\sigma v\rangle_{27}=\langle\sigma v\rangle_{\bar p}/(10^{-27}\ \mathrm{cm^3\ s^{-1}})$]

$$F_{\bar p}^h=5\times10^{-6}\langle\sigma v\rangle_{27}v_{200}^4f_{0.1}a_{10}^{-2}K_{29}^{-1}m_3^{-2}$$
$$(\mathrm{cm^2\ s\ sr})^{-1}, \tag{9}$$

where the factor $f=0.1f_{0.1}$ corrects for the fact that only a fraction f of the annihilation $\bar p$ produced with a mean energy of about $0.09m_{\tilde\gamma}$ lies in the observed, demodulated energy range. Clearly, a reasonable choice of parameters, and $K_{29}\lesssim 1$, suffices to produce the observed low-energy $\bar p$ flux below 1 GeV.

The predicted positron flux is given by an expression similar to (9) but with a production cross section approximately 30 times higher than $\langle\sigma v\rangle_{\bar p}$ (Fig. 1). Now the observed flux is[18] $e^+/(e^++e^-)\approx 0.2$ over $0.1\lesssim E\lesssim 1$ GeV, while the absolute electron cosmic-ray density can be inferred from the synchrotron emissivity of the galaxy. The interstellar e^+ flux is approximately[19]

$$F_{e^+}=1\times10^{-3}\ (\mathrm{cm^2\ s\ sr})^{-1}$$
$$(0.2\lesssim E\lesssim 1\ \mathrm{GeV}). \tag{10}$$

The predicted flux of positrons amounts to [taking $\langle\sigma v\rangle_{26}=\langle\sigma v\rangle_{e^+}/(10^{-26}\,\mathrm{cm^3\,s^{-1}})$]

$$F^h_{e^+}=5\times10^{-4}\langle\sigma v\rangle_{26}v^4_{200}a^{-2}_{10}K^{-1}_{29}m^{-2}_3\ (\mathrm{cm^2\,s\,sr})^{-1}, \quad (11)$$

and is comparable to that observed.

In summary, we have found that reasonable values of the masses of scalar leptons and quarks can lead to a critical cosmological density of photinos and to predictions of observable fluxes of cosmic-ray antiprotons and positrons, and high-energy gamma rays. For scalar masses of ~ 50 GeV and a photino mass of ~ 3 GeV, plausible assumptions about the parameters of our dark halo yield an antiproton flux comparable to that observed below the threshold for secondary production in standard cosmic-ray propagation models.

We would like to thank R. M. Barnett and M. Witherell for discussions of charm particle production, and B. Achterberg for discussions of cosmic-ray diffusion. This work was supported in part by the Department of Energy under Contract No. AT0 3-82ER 40069, in part by the National Science Foundation through Grant No. PHY77-27084, supplemented by funds from NASA, and in part by the National Science Foundation through Grant No. PHY83-13324.

Note added.—That massive photinos make a good dark matter candidate was pointed out by S. Weinberg [Phys. Rev. Lett. **50**, 387 (1983)].

[1]A. Guth, Phys. Rev. D **23**, 347 (1981).

[2]J. Yang, M. Turner, G. Steigman, D. Schramm, and K. Olive, Astrophys. J. (to be published).

[3]M. Davis, C. Frenk, and S. White, to be published.

[4]H. Goldberg, Phys. Rev. Lett. **50**, 1419 (1983).

[5]J. Ellis, J. S. Hagelin, D. V. Nanopoulos, K. Olive, and M. Srednicki, Nucl. Phys. **B238**, 453 (1984).

[6]J. Gunn, B. Lee, I. Lerche, D. Schramm, and G. Steigman, Astrophys. J. **233**, 1015 (1978).

[7]F. W. Stecker, Astrophys. J. **233**, 1032 (1978).

[8]D. Sciama, to be published.

[9]Ya. B. Zel'dovich, A. A. Klypin, M. Yu. Khlopov, and V. M. Chechetkin, Yad. Fiz. **31**, 1286 (1980) [Sov. J. Nucl. Phys. **31**, 664 (1980)].

[10]G. Steigman, Annu. Rev. Nucl. Part. Sci. **29**, 313 (1979).

[11]B. Lee and S. Weinberg, Phys. Rev. Lett. **39**, 165 (1977).

[12]K. Olive, D. Schramm, and G. Steigman, Nucl. Phys. **B180**, 497 (1981).

[13]M. Roos *et al.* (Particle Data Group), Phys. Lett. **111B**, 1 (1982); G. S. Abrams *et al.*, Phys. Rev. Lett. **44**, 10 (1980); J. L. Siegrist *et al.*, Phys. Rev. D **26**, 969 (1982); M. W. Coles *et al.*, Phys. Rev. D **26**, 2190 (1982); M. S. Alam *et al.*, Phys. Rev. Lett. **51**, 1143 (1983), and references therein.

[14]B. C. Fichtel, G. Simpson, and D. Thompson, Astrophys. J. **222**, 833 (1978).

[15]J. Bahcall and R. Soneira, Astrophys. J. Suppl. **44**, 73 (1980).

[16]V. Ginzburg and V. Ptuskin, Rev. Mod. Phys. **48**, 161 (1976).

[17]A. Buffington, S. Schindler, and C. Pennypacker, Astrophys. J. **248**, 1179 (1981).

[18]For example, see R. Protheroe, Astrophys. J. **251**, 387 (1982); P. Kiraly, J. Szabelski, J. Wdowczyk, and A. W. Wolfendale, Nature **293**, 120 (1981); F. Stecker, R. Prothero, and D. Kazanas, Astrophys. Space Sci. **96**, 171 (1983); L. C. Tan and L. K. Ng, Astrophys. J. **269**, 751 (1983).

[19]For example, see R. Protheroe, Astrophys. J. **254**, 391 (1982).

New Limit on the Low-Energy Antiproton/Proton Ratio in the Galactic Cosmic Radiation

S. P. Ahlen,[1] S. Barwick,[2] J. J. Beatty,[1] C. R. Bower,[3] G. Gerbier,[2] R. M. Heinz,[3] D. Lowder,[2] S. McKee,[4] S. Mufson,[5] J. A. Musser,[4] P. B. Price,[2] M. H. Salamon,[2],[a] G. Tarlé,[4] A. Tomasch,[1] and B. Zhou[1]

[1]*Physics Department, Boston University, Boston, Massachusetts 02215*
[2]*Physics Department, University of California, Berkeley, California 94720*
[3]*Physics Department, Indiana University, Bloomington, Indiana 47405*
[4]*Physics Department, University of Michigan, Ann Arbor, Michigan 48109*
[5]*Astronomy Department, Indiana University, Bloomington, Indiana 47405*
(Received 3 December 1987)

We describe the results of a balloon-borne apparatus searching for low-energy antiprotons in the Galactic cosmic rays. For energies less than 640 MeV at the top of the atmosphere, no cosmic-ray antiprotons were observed. This yields an upper limit to the $\bar{p}/p$ ratio of 4.6×10^{-5} at the 85% confidence level.

PACS numbers: 98.60.Ce, 14.80.Ly, 98.60.Ln

Conventional models of cosmic-ray propagation predict the production of secondary antiprotons ($\bar{p}$'s) through collisions of high-energy cosmic rays with the interstellar medium.[1] At several gigaelectronvolts the predicted $\bar{p}/p$ ratio is of the order 10^{-4} and is expected to decrease with increasing energy as $E^{-0.6}$. For energies less than 1 GeV, secondary production is suppressed and results in a $\bar{p}/p$ ratio less than 10^{-5}. Thus the report[2] in 1981 of a cosmic-ray $\bar{p}/p$ ratio of $(2.2\pm0.6)\times10^{-4}$ for the energy interval 130 to 320 MeV (corrected to the top of the atmosphere) appeared to demand either a novel mechanism for $\bar{p}$ production or a radical revision of our understanding of cosmic-ray propagation. Many models of both categories have since been offered.[3] Of particular interest is the suggestion that low-energy $\bar{p}$'s are to be expected as annihilation products of photinos,[4-6] or Higgsinos,[6,7] if these particles constitute the missing dynamical mass of the Galactic halo. Although the details of these models depend on unknown mass parameters, distinct $\bar{p}$ spectral features are predicted which, if observed, could be interpreted as support for the role of these annihilating particles as the dark matter in the Galaxy and the Universe. This Letter describes the results of a balloon experiment (PBAR) which we have recently performed to measure the low-energy $\bar{p}/p$ ratio.

A schematic of PBAR appears in Fig. 1. S1 and S2 are scintillators used to identify the charge and velocity of cosmic-ray particles. S1 consists of three segments (each having dimensions $100\times25\times2.54$ cm^3) of fast plastic scintillator (Bicron BC420); S2 consists of two such segments. Each end of each segment is viewed through a bent, twisted light pipe by a fast (Phillips Model XP2020) photomultiplier tube (PMT). A relativistic muon yields 50 photoelectrons per PMT. DT is a drift-tube hodoscope which is placed in the bore of a split-coil superconducting magnet having a central field of 9 kG and a field nonuniformity of $\simeq$20%. The magnet design permits particles to pass through DT without coming in contact with the Dewar flash. DT determines the sign of the charge and the rigidity of cosmic rays by measuring the particle trajectory in the magnetic field. In conjunction with measurements of charge and velocity this permits the complete identification of the particles passing through PBAR. DT consists of 323 drift tubes in sixteen planes having tubes parallel to the magnetic field and eight planes having tubes perpendicular to the field. Each tube is 1.27 cm in diameter with a 20-μm wire strung down the center. A 50-50 mixture of Ar-C_2H_6 flows serially through the 323 tubes at slightly higher than ambient pressure. The total wall thickness traversed by a particle passing through DT is 0.28 g/cm^2. CK, a 15-cm-thick water Cherenkov counter, was not used for the analysis of data presented here.

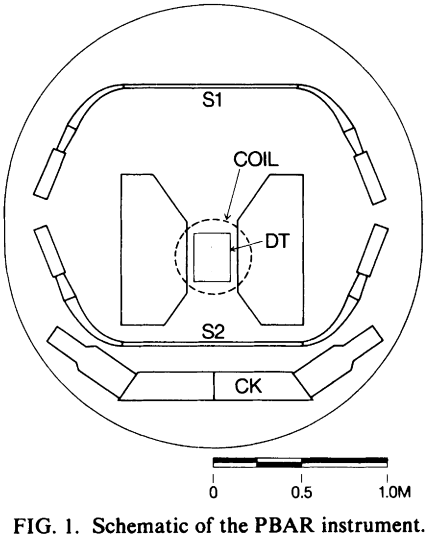


FIG. 1. Schematic of the PBAR instrument.

© 1988 The American Physical Society

PBAR was flown in an aluminum shell by balloon from Prince Albert, Saskatchewan, on 13–14 August 1987. The environment during the flight was ideal, the temperature deviating from 20 °C by no more than a few degrees and the gondola pressure maintaining good stability near 1 atm. PBAR achieved an altitude of 36 km for a duration of 10 h 35 min. The latitude varied between 53° 18′ N and 54° 22′ N, and the longitude between 104° 55′ W and 108° 04′ W. Science and housekeeping data were telemetered to a ground station for recording and on-line display, and commands were sent to the instrument as necessary. Particle candidates were required to pass a two-level trigger. A fast trigger was generated by a coincidence between any S1 PMT and any S2 PMT, which gated the electronics. A subsequent slow trigger was required from DT (any tube in the top two planes and any tube in the bottom two planes) to complete digitization of the event, which was then telemetered to the ground. The efficiency of this trigger was determined to be nearly 100% in preflight tests with cosmic-ray muons. Magnetic tapes of data recorded during the flight were distributed to the four institutional groups, who performed independent analyses. The results described below correspond only to events with energy less than 620 MeV/amu in DT.

Figure 2 shows the S2-S1 time-of-flight (TOF) resolution and the DT spatial resolution as determined from the flight data. Event positions along the length of any scintillator can be determined by two independent methods. The most accurate method uses the DT vector formation to extrapolate an event position on the scintillator. A second method uses the time difference between the left and right PMT on each scintillator. The TOF standard deviation can be determined from the relation $\sigma_{TOF} = (L/v)\sigma_{\Delta\tan\theta} = 160$ ps, where L is the separation of S1 and S2, $\Delta\tan\theta$ is the difference in track slopes as determined by DT vector information and PMT time differences in S1 and S2, and v is the effective propagation speed of light in the scintillators. Effective propagation speeds were determined from the measured slopes in plots of PMT time differences versus position extrapolated from DT. The width of the histogram in Fig. 2(a) was used to calculate that the PBAR timing resolution was 160 ps. Figure 2(b) shows the residuals of the drift tubes for fitted proton tracks in the magnetic field during the flight. The single-tube standard deviation is 109 μm averaged over all energies and angles.

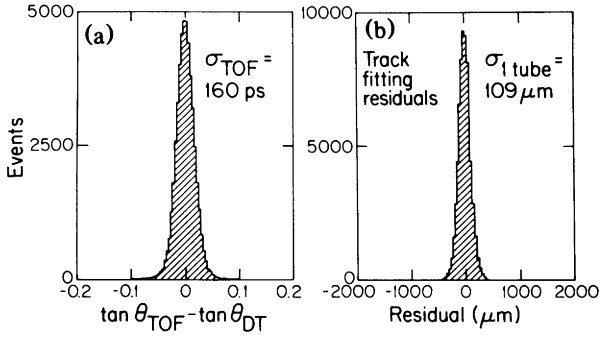


FIG. 2. (a) Histogram of track slope differences determined by drift tubes and scintillators. (b) Histogram of residuals of drift-tube position measurements.

In order to discriminate against particles with $|Z| > 1$, reduce backgrounds due to nuclear interactions and scattering, and maintain high efficiency for $\bar{p}$'s which annihilate in the detector, the following cuts were applied:

(1) Only one S1 segment has both PMT's triggered (eliminates showers).

(2) At least one S2 segment has two PMT's triggered (allows multiple hits due to annihilation pions produced in S2 and CK).

(3) The particle is downward moving (rejects albedo).

(4) At least one drift tube is hit in each of the top two and bottom two rows of DT, with at least six rows in both the top and bottom halves of DT parallel to **B** containing hits, and only one clean track is present in DT (clean trajectory data).

(5) The particle rigidity (from DT) is less than 1.5 GV and its speed is between 0.17c and 0.8c as determined by TOF (restricts energy range).

(6) The χ^2 for the fit of the track in the sixteen planes of tubes parallel to **B** is less than 41 (rejects scatters).

(7) The slopes of the track as measured by DT and the scintillators must be within 0.10 of each other (eliminates spurious TOF timing).

(8) The signals in the scintillator segments pointed to by the track in DT must exceed 100 photonelectrons (eliminates particles clipping scintillator corners), S1 $< 3S1_{calc}$ (eliminates $|Z| > 1$), and $S2 > S2_{calc}/6$ (eliminates corner clippers for large S2 pulse heights, but allows large signals for annihilation pions). The S_{calc}

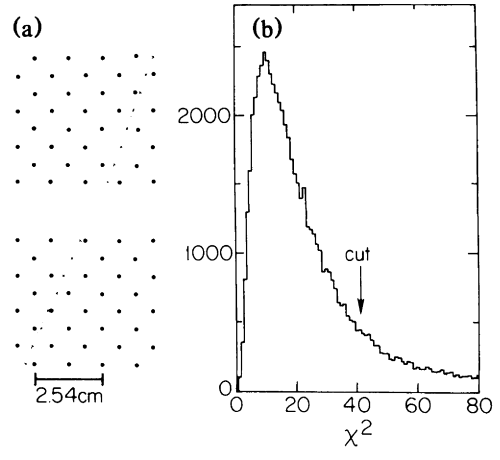


FIG. 3. (a) A track of a typical background event. (b) The χ^2 distribution for the fitted tracks.

values are those expected for a proton having the rigidity measured by DT.

An example of a hard-scatter background event is shown in Fig. 3(a). This is the track of a proton which scattered in the middle of DT, simulating the same sense of curvature as an antiproton. The χ^2 for this event is 76, and so it is rejected by our cut on goodness of fit at $\chi^2=41$, shown in Fig. 3(b).

For events which passed the above cuts, masses were calculated with the rigidity determined by DT and the velocity from TOF. Figure 4 shows the resulting mass histograms for positive and negative charge. Peaks for protons, positive and negative muons and/or pions, and possibly for negative and positive kaons are present. There are a total of 52000 protons in the mass range 600–1500 MeV/c^2. However, *there are no $\bar{p}$ candidates.* The mass spectrum for albedo (upward moving) protons is shown beneath the proton spectrum. Except for a slight shift in the mass peak, there is no evidence that these particles, which have the same sense of curvature as would $\bar{p}$'s, are treated any differently than downward moving protons. Furthermore, the ratio of negative to positive muons and/or pions is consistent with accelerator measurements.[8] Thus, it is apparent that there are no flaws in our track-fitting algorithms which would selectively remove $\bar{p}$'s from our data set. Nor can charge-sign asymmetry effects of detector response account for a bias against $\bar{p}$'s.[9]

In our calculating an upper limit to $\bar{p}/p$ from these data, the annihilation of $\bar{p}$'s in the atmosphere and instrument must be taken into account. A $\bar{p}$ which annihilates in the instrument (which would occur predominantly in CK) could be rejected in three ways: (i) multiple tracks in DT due to one or more secondary pions; (ii) multiple S1 slabs struck by one or more pions; and (iii) spurious S2 timing information for the S2 segment traversed by the $\bar{p}$, due to a pion's hitting the slab between its end and the $\bar{p}$ scintillation wave front. Mechanisms (i) and (ii) cause 6% of annihilations to be lost, while 5% are lost to (iii). Since 30% of $\bar{p}$'s annihilate in S2 or CK, the total correction to $\bar{p}/p$ is 1.03. A factor of 1.14 must be included to correct for $\bar{p}$'s which annihilate in the 8 g/cm^2 of material from the top of the atmosphere to the bottom of S1.[10] A factor of 1.13 must also be included to account for the enhancement of the proton flux due to secondary[11] and albedo protons at the top of the instrument, with an additional factor of 0.94 applied to account for protons absorbed in the atmosphere. The combined correction factor is 1.25, resulting in an 85%-confidence-level (C.L.) upper limit of $\bar{p}/p < 4.6 \times 10^{-5}$ at the top of the atmosphere in the energy range 205–640 MeV. The lower bound on the energy is due to the geomagnetic cutoff rigidity.

Solar-cycle modulation effects reduce the energy of cosmic rays entering the heliosphere[12,13] and are important at low energies.[14] The observed energy window for PBAR corresponds to 700–1135 MeV in interstellar

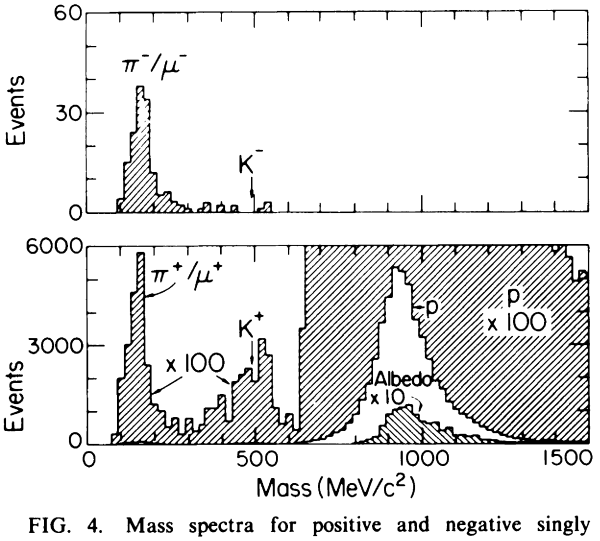


FIG. 4. Mass spectra for positive and negative singly charged particles. There are three histograms for the positive-curvature events. From top to bottom they are as follows: two for the downward-going particles at two different scales, and one for albedo particles at an expanded scale.

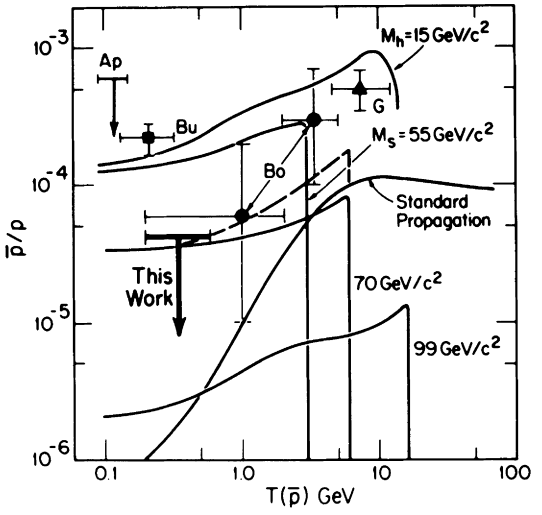


FIG. 5. Comparison of data with several models for the $\bar{p}/p$ ratio. The limits shown for this work are at the 85% confidence level so that they correspond to the upper limit of a 1σ error bar, Ap is from Ref. 17, Bo is from Ref. 19, Bu is from Ref. 2, and G is from Ref. 18. The standard propagation is a leaky-box calculation from Fig. 6 of Ref. 1b. Three curves for the photino model are shown, corresponding to photino masses of 3, 6, and 16 GeV/c^2, and scalar-fermion masses M_s of 55, 70, and 99 GeV/c^2, respectively. The Higgsino model (mass = 15 GeV/c^2) from Ref. 7 is also shown. These spectra have been modified to take solar modulation into account. The dashed line is the sum of the standard propagation and the 6-GeV photino curve.

space when estimated by the force-field approximation to solar modulation. For Ref. 2 the corresponding interval is 930–1120 MeV in interstellar space, allowing the two results to be directly compared. Satellite measurements from IMP-8 indicate no significant flux of solar protons above 100 MeV at the time of the PBAR flight.[15] Possible modulation effects which depend on charge sign,[16] although small, would tend to enhance the discrepancy between the present result and that of Ref. 2.

In Fig. 5 we compare data on $\bar{p}/p$ ratio with other data[2,17–19] and with several models. Each model has been corrected for the level of solar modulation present during our flight near solar minimum by integration of the spherically symmetric Fokker-Planck equations with a diffusion coefficient of the form given in Ref. 13 ($\phi = 500$ MV). We show Protheroe's standard cosmic-ray propagation model[1b] and several models involving the annihilation of exotic dark matter. The photino models require that photinos account for the missing mass of the Universe (with the assumption $\Omega = 1$), and for the dark matter in the galactic halo. The former constraint establishes a relation between the photino mass and the scalar fermion mass,[20] while the latter makes possible the calculation of $\bar{p}/p$ in the local interstellar medium.[5] This model cannot be correct unless the photino mass is greater than $\simeq 6$ GeV/c^2 and the scalar fermion mass is greater than $\simeq 70$ GeV/c^2. These values have not been excluded by accelerator searches[21] or by searches for high-energy neutrino emission from the Sun.[20]

Also shown in Fig. 5 is the ratio $\bar{p}/p$ for the Higgsino model proposed in Ref. 7, where Higgsinos of mass 15 GeV/c^2 account for the dark matter in the galactic halo. This model is inconsistent with our data, but would be consistent if the Higgsino mass were greater than 25 GeV, or if the mass density of the galactic halo were a factor of 2 smaller than that assumed in Ref. 7.

The results presented here represent the best limit to date on the ratio of antiprotons to protons in our galaxy. However, since the fraction of extragalactic cosmic rays in the neighborhood of the Earth at the energies considered here is likely to be no more than 2×10^{-5},[22] our results shed no light on the question of primary sources of antimatter outside of our galaxy.

This experiment was made possible by the efforts of personnel from the National Scientific Balloon Facility and by financial assistance from Boston University (B.U.), Indiana University (I.U.), and the University of Michigan (U.M.). We also acknowledge assistance from J. Bartlett, A. Buffington, R. Claxton, M. Gebhard, T. Karakashian, S. Lopez, H.-S. Park, J. Reynoldson, M. Solarz, and G. Turner. This work was supported by National Science Foundation Grants No. PHY-8519440 (B.U.), and No. PHY-8603225 and No. PHY-8702763 [University of California, Berkeley (U.C.B.)], U.S. Department of Energy Contract No. DE-AC02-76ER01112 (U.M.), NASA Contracts No. NGT-50014 (U.C.B.), No. NAS1-17820 (U.M.) and No. NAGW 1035 (I.U.), and Calspace Grants No. CS78-87 and No. CS08-86 (U.C.B.).

[a]Present address: Physics Department, University of Utah, Salt Lake City, UT 84112.

[1a]T. K. Gaisser and E. H. Levy, Phys. Rev. D **10**, 1731 (1974).

[1b]R. J. Protheroe, Astrophys. J. **251**, 387 (1981).

[2]A. Buffington, S. M. Schindler, and C. R. Pennypacker, Astrophys. J. **248**, 1179 (1981).

[3]S. A. Stephens and B. G. Mauger, Astrophys. Space Sci. **110**, 337 (1985); M. S. Turner, Nature (London) **297**, 379 (1982); F. W. Stecker, R. J. Protheroe, and D. Kazanas, Astrophys. Space Sci. **96**, 171 (1983); L. C. Tan and L. K. Ng, J. Phys. G **9**, 227 (1983); C. D. Dermer and R. Ramaty, Nature (London) **319**, 205 (1986).

[4]J. Silk and M. Srednicki, Phys. Rev. Lett. **53**, 624 (1984).

[5]F. W. Stecker, S. Rudaz, and T. F. Walsh, Phys. Rev. Lett. **55**, 2622 (1985).

[6]J. S. Hagelin and G. L. Kane, Nucl. Phys. B **B263**, 399 (1986).

[7]S. Rudaz and F. W. Stecker, University of Minnesota Report No. UMN-TH-606/87, 1987 (to be published).

[8]V. G. Grishin, Usp. Fiz. Nauk **127**, 51 (1979) [Sov. Phys. Usp. **22**, 1 (1979)].

[9]S. P. Ahlen, Rev. Mod. Phys. **52**, 121 (1980).

[10]T. Bowen and A. Moats, Phys. Rev. D **33**, 651 (1986).

[11]T. A. Tygg and J. A. Earl, J. Geophys. Res. **76**, 7445 (1971).

[12]L. J. Gleeson and W. I. Axford, Astrophys. J. **154**, 1011 (1968).

[13]M. Garcia-Munoz *et al.*, in *Proceedings of the Nineteenth International Cosmic Ray Conference, La Jolla, California, 1985,* edited by F. C. Jones, J. Adams, and G. Mason, NASA Conference Publication No. 2376 (U.S. GPO, Washington, D.C., 1985), Vol. 4, p. 409.

[14]J. S. Perko, Astron. Astrophys. **184**, 119 (1987).

[15]K. R. Pyle, private communication.

[16]P. Evenson, M. Garcia-Munoz, P. Meyer, K. R. Pyle, J. A. Simpson, Astrophys. J. **275**, L15 (1983).

[17]M. V. K. Apparao, Can. J. Phys. **46**, S654 (1968).

[18]R. L. Golden *et al.*, Phys. Rev. Lett. **43**, 1196 (1979).

[19]E. A. Bogomolov *et al.*, in *Proceedings of the Sixteenth International Cosmic Ray Conference, Kyoto, Japan, 1979,* edited by S. Miyake (Univ. of Tokyo Press, Tokyo, 1979), Vol. 1, p. 330; E. A. Bogomolov *et al.*, in *Proceedings of the Twentieth International Cosmic Ray Conference, Moscow, U.S.S.R., 1987,* edited by V. A. Kozyarivsky *et al.* (Nauka, Moscow, 1987).

[20]T. K. Gaisser, G. Steigman, and S. Tilav, Phys. Rev. D **34**, 2206 (1986).

[21]R. M. Barnett, H. F. Haber, and G. L. Kane, Nucl. Phys. **B267**, 625 (1986).

[22]S. P. Ahlen, P. B. Price, M. H. Salamon, and G. Tarlé, Astrophys. J. **260**, 20 (1982).

THE ASTROPHYSICAL JOURNAL, 296:679–684, 1985 September 15
© 1985. The American Astronomical Society. All rights reserved. Printed in U.S.A.

CAPTURE BY THE SUN OF A GALACTIC POPULATION OF WEAKLY INTERACTING, MASSIVE PARTICLES

WILLIAM H. PRESS AND DAVID N. SPERGEL
Harvard-Smithsonian Center for Astrophysics
Received 1985 February 4; accepted 1985 March 6

ABSTRACT

A previous calculation showed that hypothetical, massive particles, if present in the Sun with the right scattering cross section and in a concentration as low as 10^{-12} by number, could likely explain the solar neutrino discrepancy. We here suppose that the same hypothetical particles make up the missing dynamical mass in the Galaxy, providing a mass density between 0.01 and 0.1 $M_\odot$ pc^{-3} and with a velocity dispersion of $\sim$30 (disk) or 300 (halo) km s^{-1}. We compute the number of particles which would be captured by the Sun up to its present age by the capture mechanism of scattering off individual nucleons in the Sun and into bound orbits. For scattering cross sections not too far from the value of 4×10^{-36} cm^2, which is optimal for suppressing solar neutrinos, we find that just the right number can be captured. The solutions to the neutrino problem in the Sun and the missing mass problem in the Galaxy may thus be one and the same.

Subject headings: elementary particles — galaxies: Milky Way — neutrinos — Sun: interior

I. INTRODUCTION

In a previous paper (Spergel and Press 1985, hereafter Paper I) we calculated the effect of hypothetical, massive, weakly interacting particles on energy transport in the Sun. We found that for a broad mass range, $m_x = 5$–60 GeV, and for a certain optimal scattering cross section $\sigma_{\rm crit}(m_x)$, about 4×10^{-36} cm^2, a concentration of such particles in the Sun as low as 10^{-12} by number would substantially increase the Sun's effective thermal conductivity within the neutrino-producing central core. This could likely solve the observed solar neutrino discrepancy as reviewed by, e.g., Bahcall *et al.* (1982), Fowler (1982), or Bahcall (1985). Subsequent to writing Paper I, we also became aware of until now unpublished work by J. Faulkner and R. Gilliland (1985; summarized in Steigman *et al.* 1978), which reached similar conclusions.

As remarked in Paper I, the posited mass range of these "cosmions" is also one of current experimental and theoretical interest in particle physics. For example, the cosmion might be a supersymmetric photon (photino), in which case the posited cross section—which was derived in Paper I purely from astrophysical requirements—is entirely plausible for some particle-theory models (Krauss *et al.* 1985, hereafter Paper III).

Paper I did not attempt to explain the physical origin of the required concentration, say 10^{-12}. *If* cosmions are present in the Sun with a specified concentration, *then* they affect energy transport as calculated. The purpose of this paper, by contrast, is to *derive* the solar concentration of cosmions from the assumption that the cosmions form all, or a significant part of, the "missing" (that is, dark) mass in the Galaxy.

For some of the parameter space (cosmion mass and cross section) recommended by Paper I, we find, strikingly, that reasonable parameters for the missing mass (its density and velocity dispersion in the solar neighborhood) yield just the right concentration captured by the Sun so as to explain the solar neutrino discrepancy. The capture mechanism is scattering of cosmions in the Galactic population off individual nucleons in the Sun and into bound orbits. Subsequent scatterings rapidly put a captured cosmion into the approximately thermal distribution that was derived in Paper I. The Sun accumulates cosmions by capture over its lifetime, approximately 5×10^9 yr. The alternative mechanism of gravitational capture during the formation of the Sun, which was considered by Steigman *et al.* (1978), yields concentrations too small to be interesting.

If cosmions do not exist as actual particles, then this paper is a hypothetical exercise in the dynamics of accretion and capture. If cosmions *do* exist, then this paper attempts to make plausible the hypothesis that *the solutions to the neutrino problem in the Sun and the missing mass problem in the Galaxy are one and the same.* (L. M. Krauss and G. R. Blumenthal each have reached similar independent conclusions.)

The plan of this paper is as follows: In § II we calculate the rate of accretion onto the Sun, per phase-space volume, of an assumed Maxwellian distribution of Galactic cosmions, and delineate the interesting regions of phase space. In § III we estimate, parameterized by the cosmion mass and for each point in phase space, the probability that a cosmion scattering event will result in capture; this section is where the scattering kinematics enters. In § IV we integrate an appropriate distribution function of unbound orbits through an actual solar model, obtaining the scattering probability per accretion event. In § V, the results from other sections are combined and conclusions are drawn.

II. PHASE-SPACE RATE OF ACCRETION

The Sun has an escape velocity, from its surface, of about 618 km s^{-1}. We will be considering a range of mean-square velocities for the Galactic population of cosmions, but none higher than the order of 300 km s^{-1}. Therefore, the gravitational effects of the Sun are significant: it is incorrect to use only the geometrical cross section of the Sun, or only straight-line orbital trajectories. One way of getting the bookkeeping right is to reparameterize the velocity space of the cosmions, where their distribution is Maxwellian,

in terms of invariant integrals of orbits near (or through) the Sun, namely the total energy and total angular momentum per cosmion mass. We do this now.

We assume that the Galactic population of cosmions has a Maxwellian distribution in velocity space, so that the number per unit spatial volume with speed between v and $v + dv$ is

$$p(v)dv = n_x\left(\frac{3}{2\pi\bar{v}^2}\right)^{3/2} 4\pi v^2 \exp\left(\frac{-3v^2}{2\bar{v}^2}\right)dv\,, \tag{2.1}$$

where n_x is the number density of cosmions. In equation (2.1), $\bar{v}$ is the rms velocity of the cosmion distribution, related to the cosmion temperature T_x by

$$\bar{v}^2 = 3kT_x/m_x\,. \tag{2.2}$$

(We use $\bar{v}$ instead of T_x because Galactic populations are usually discussed in terms of velocity dispersions.)

Now consider a unit surface element located on a sphere of radius R centered on the Sun. We will at present let $R \to \infty$, so R should be thought of as very large. The spatial number density of cosmions with velocity between v and $v + dv$, and with angle from the normal between θ and $\theta + d\theta$ is equal to equation (2.1) multiplied by the factor $\frac{1}{2}d\cos\theta$. The flux dF (number per area per time) of such particles through the surface element is this, multiplied by an additional factor $v\cos\theta$. So in all,

$$dF = n_x\left(\frac{3}{2\pi\bar{v}^2}\right)^{3/2} \pi v^3 \exp\left(\frac{-3v^2}{2\bar{v}^2}\right)d(\cos^2\theta)dv\,, \tag{2.3}$$

with the limits of integration being

$$0 < v < \infty\,, \qquad 0 < \theta < \pi/2 \tag{2.4}$$

for inward-going particles.

The orbital invariants E and J are related to v and θ by

$$E \equiv \tfrac{1}{2}v^2\,, \qquad J \equiv vR\sin\theta\,. \tag{2.5}$$

One easily computes the Jacobian determinant relating $d(\cos^2\theta)dv$ to $dEdJ$ as

$$\left|\frac{\partial(\cos^2\theta,\, v)}{\partial(E,\, J)}\right| = \frac{J}{\sqrt{2}\,R^2E^{3/2}}\,. \tag{2.6}$$

Equations (2.6) and (2.3) combine to give, when multiplied by $4\pi R^2$, the total accretion rate (cosmions per unit time) from a phase-space volume $dEdJ$, which can be written as

$$d\mathscr{F} \equiv 4\pi R^2 dF = n_x\left(\frac{3}{2\pi\bar{v}^2}\right)^{3/2} \exp\left(\frac{-3E}{\bar{v}^2}\right)2\pi^2 dEdJ^2\,. \tag{2.7}$$

We must now determine (1) which regions of (E, J)-space correspond to orbits which intersect the Sun, and (2) for orbits which do intersect the Sun, which regions of (E, J)-space can kinematically yield scatters into bound orbits. We do the first task now; the second task is started in this section and completed in the next.

For Keplerian orbits outside the Sun, the perihelion (closest approach) distance is related to E and J by

$$r_{\text{peri}} = \left(\frac{J^2}{GM_\odot}\right)\Big/\left(1 + \sqrt{1 + 2\,\frac{J^2}{GM_\odot}\,\frac{E}{GM_\odot}}\right). \tag{2.8}$$

This equation has the two asymptotic limits

$$r_{\text{peri}} \approx \frac{J^2}{2GM_\odot} = \frac{v^2R^2\sin^2\theta}{2GM_\odot} \qquad \left[\frac{J^2E}{(GM_\odot)^2} \ll 1\right] \tag{2.9}$$

and

$$r_{\text{peri}} \approx \sqrt{\frac{J^2}{2E}} = R\sin\theta \qquad \left[\frac{J^2E}{(GM_\odot)^2} \gg 1\right], \tag{2.10}$$

corresponding to large- and small-angle gravitational deflections respectively. Note that equation (2.9) has no dependence on the energy E, since for negligible E all orbits from infinity are very nearly parabolic. The dividing curve between the regimes of large and small deflection is

$$\frac{J^2}{GM_\odot}\,\frac{E}{GM_\odot} = 1\,, \tag{2.11}$$

which is shown as the dashed hyperbola in Figure 1.

In Figure 1 we can also plot the boundary of orbits with perihelion less than the solar radius $R_\odot$. It follows from equation (2.8) that this is a straight line of positive slope, starting at abscissa 2 $R_\odot$; it is drawn in the figure as a short-dashed line. For capture, we

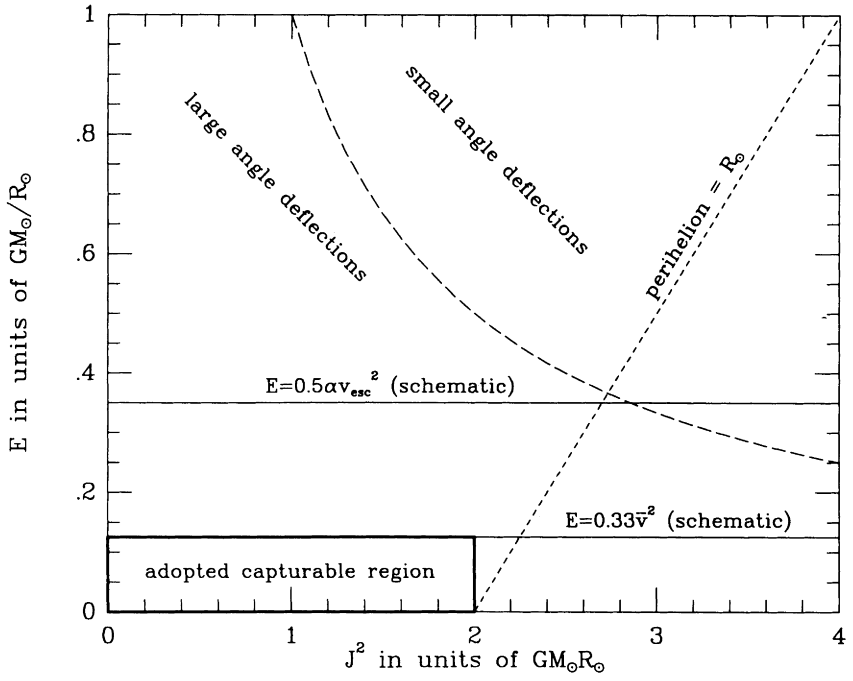


FIG. 1.—Phase space of unbound orbits near (and through) the Sun. A particle's orbit is determined by its energy at infinity per unit mass E and by the square of its angular momentum per unit mass J^2. Orbits to the left of the short-dash line have perihelions inside the Sun. Orbits well below the long-dash line undergo large gravitational deflections and all have effectively the same, zero-energy, shapes. The Maxwellian distribution of "cosmions" is concentrated below the line $E = 0.33\bar{v}^2$; to be kinematically capturable, a cosmion must also be below the line $E = 0.5\alpha v_{esc}^2$. The "adopted capturable region" is therefore bounded by the lower of the two solid lines (shown here schematically) and by the perihelion line. Cosmions in this region will be captured, on average, *if* they undergo a scattering event.

can obviously restrict our attention to cosmions with perihelion less than $R_\odot$. It is also the case that we can restrict our attention to cosmions with E (energy at infinity) less than some constant (of order unity) times v_{esc}^2: cosmions with larger energy would need to lose more of their kinetic energy in a scatter, to become bound, than is kinematically possible. In the figure, this upper limit is shown schematically as a horizontal line at $E = 0.5\alpha v_{esc}^2$, for some value of the constant α. The dimensionless factor α depends on the cosmion mass and on the temperature of the scattering baryons and will be estimated in detail in the next section.

Equation (2.7) shows that the total accretion rate, as a function of the E and J^2 variables plotted in Figure 1, is uniform in J^2 and is exponentially decreasing in E with an effective "scale height" of $\frac{1}{3}\bar{v}^2$. Since the cosmion flux falls off exponentially above this horizontal line, which is shown schematically in the figure, the line can serve as *another* upper bound on what we can call the "kinematically capturable phase space." We can approximate the accretion rate as being uniform below this line and zero above it.

In other words, the "arena" of phase space which must be considered is bounded on the "right" by the perihelion line, and "above" by the lower of the two limits (1) where capture is excluded by the kinematics of a scatter, and (2) where the Maxwellian cosmion density effectively falls off. For convenience, we approximate the perihelion line as being vertical, so that the "adopted capturable region" is as shown in the figure.

The principal result of this section can now be codified: the accretion rate of kinematically capturable cosmions onto the Sun (number per unit time) is approximately

$$\mathscr{F}_{\text{capturable}} = n_x\left(\frac{3}{2\pi\bar{v}^2}\right)^{3/2} 2\pi^2(2GM_\odot R_\odot)\,\min\left(\frac{1}{3}\bar{v}^2, \frac{\alpha}{2}v_{esc}^2\right). \tag{2.12}$$

It remains for us to estimate (1) the dimensionless factor α, which is the subject of § III; and (2) the fraction of "capturable" cosmions which undergo a scattering and which are therefore actually captured, which is the subject of § IV.

III. KINEMATICAL CAPTURE FACTOR

We want to calculate how energetic a cosmion can be and still be captured in a single scatter. That specific energy E, expressed as a fraction of the escape energy from the surface of the Sun, $\frac{1}{2}v_{esc}^2$ is defined to be the dimensionless factor α.

Of course, that energy is not strictly the same for all scattering events, since each scattering event selects from a statistical universe of two parameters: the scattering angle in the center of mass frame, and the depth within the Sun's gravitational potential at which the scatter takes place. In the same spirit by which we replaced the Maxwell exponential by its characteristic scale height, however, we can define a characteristic value for α which ought to give correct results to well within a factor of 2. We do this by computing the value α for an "average" collision, both in the sense that the energy loss is equal to the mean energy loss averaged over scattering

angles and in the sense that the scatter occurs at a typical average depth. Some more energetic cosmions will actually be captured; and some less energetic ones will actually fail to be captured. But since the probability distribution functions of scattering angle and depth are both mutually uncorrelated, and since the dependence of required or achieved energy loss on these parameters is not hugely nonlinear, the numbers which anomalously succeed or fail ought to compensate fairly closely.

Let $\overline{\Delta E}$ be the mean energy loss in a scatter. Let us consider first the case of scattering off baryons at zero temperature, i.e., at rest. Then simple kinematics gives

$$\overline{\Delta E} = \frac{2m_x m_p}{(m_x + m_p)^2}(1 - Q)E , \tag{3.1}$$

where m_x and m_p are the cosmion and baryon masses, and where Q is the differential cross section structure factor which was defined in equation (4.7) of Paper I. (For isotropic scattering, $Q = 0$.)

The equation for α therefore equates energy loss to energy at infinity,

$$\frac{2m_x m_p}{(m_x + m_p)^2}(1 - Q)\left(\alpha \frac{v_{esc}^2}{2} + \beta \frac{v_{esc}^2}{2}\right) = \alpha \frac{v_{esc}^2}{2} , \tag{3.2}$$

where β parameterizes the escape energy as a function of depth; $\beta = 1$ at the surface, and it increases inwards. At the half-mass radius of the Sun, which we will take as our "typical" scatter location, $\beta \approx 3.4$. The solution of equation (3.2) is

$$\alpha = \frac{2m_x m_p}{m_x^2 + m_p^2}\beta , \tag{3.3}$$

which is tabulated as a function of m_x/m_p in the first and second columns of Table 1.

The calculation of α including the effects of a finite temperature distribution of baryons is somewhat more involved. We can express the average energy loss as

$$\overline{\Delta E} = \frac{\text{energy loss per unit time}}{\text{collisions per unit time}} \equiv \frac{I_2}{I_1} , \tag{3.4}$$

where

$$I_2 = \int \langle \Delta E(\boldsymbol{v}_x, \boldsymbol{v}_p)\rangle f_p(v_p)\,|\boldsymbol{v}_p - \boldsymbol{v}_x|\,\sigma d^3 v_p , \tag{3.5}$$

and

$$I_1 = \int f_p(v_p)\,|\boldsymbol{v}_p - \boldsymbol{v}_x|\,\sigma d^3 v_p . \tag{3.6}$$

Here $f_p(v_p)$ is the baryons' velocity distribution function,

$$f_p(v_p) = \left(\frac{m_p}{2\pi k T_p}\right)^{3/2} \exp\left(\frac{-m_p v_p^2}{2kT_p}\right) , \tag{3.7}$$

and $\langle \Delta E(\boldsymbol{v}_x, \boldsymbol{v}_p)\rangle$ is the mean energy lost in a collision, averaged over scattering angle,

$$\langle \Delta E(\boldsymbol{v}_x, \boldsymbol{v}_p)\rangle = (1 - Q)\frac{m_x m_p}{(m_x + m_p)^2}\left[m_p v_p^2 - m_x v_x^2 + (m_x - m_p)v_x v_p \cos\chi\right] , \tag{3.8}$$

with $\boldsymbol{v}_x \cdot \boldsymbol{v}_p \equiv v_x v_p \cos\chi$.

The integral I_1 can be evaluated in terms of the dimensionless parameter $w \equiv m_p v_x^2/2kT_p$, giving

$$I_1 = \sigma v_x\left[\text{erf}\,(\sqrt{w})\left(1 + \frac{1}{2w}\right) + \frac{\exp(-w)}{\sqrt{\pi w}}\right] . \tag{3.9}$$

TABLE 1

m_x/m_p	$\alpha(T = 0)$	$\alpha(T = \frac{1}{2}T_{central})$
1	3.392	3.137
2	2.715	2.645
4	1.597	1.596
8	0.836	0.845
16	0.423	0.430
32	0.212	0.216
64	0.106	0.108

This has the asymptotic limits

$$I_1 = 2\sigma\sqrt{\frac{2kT_p}{\pi m_p}} \qquad (w \ll 1) \tag{3.10}$$

and

$$I_1 = \sigma v_x \qquad (w \gg 1)\,. \tag{3.11}$$

The integral I_2 can be evaluated from Appendix B of Paper I, with the result

$$I_2 = \frac{1}{4}(1-Q)\sigma m_x v_x^3 \frac{m_x m_p}{(m_x+m_p)^2}\left\{\frac{\mathrm{erf}\,(\sqrt{w})}{w^2}\left[\frac{4}{\mu}+1+w\left(\frac{8}{\mu}-4\right)-4w^2\right]+\frac{\exp -w}{\sqrt{\pi}\,w^{3/2}}\left[-4w+\frac{8}{\mu}-2\right]\right\}, \tag{3.12}$$

where $\mu \equiv m_x/m_p$.

Equations (3.4), (3.9), and (3.12) allow α to be computed, by setting v_x^2 equal to $(\alpha+\beta)v_{\rm esc}^2$, equating $\overline{\Delta E}$ to $\alpha v_{\rm esc}^2/2$, and solving the resulting implicit equation for α. Numerical results are given in Table 1, for a baryon temperature corresponding to the temperature at the half-mass radius of the Sun (also approximately equal to half the central temperature). The effects of finite temperature are visible but small.

In fact, for reasons discussed in Paper I, we are only interested in cosmions whose mass is $\gtrsim 5m_p$. In this regime, the two α columns of Table 1 can be approximately summarized by the fitting formula

$$\alpha \approx 2\beta\frac{m_p}{m_x} \approx \frac{7m_p}{m_x}\,. \tag{3.13}$$

IV. MEAN COLUMN DENSITY OF THE SUN

The previous two sections have calculated the accretion rate of "capturable" cosmions. The number that are actually captured depends on the fraction of capturable particles which actually undergo one or more scatters. We denote this fraction by f_σ. It is in the calculation of f_σ that the scattering cross section enters directly.

In terms of the local solar mass density ρ, we have

$$f_\sigma = \left\langle 1-\exp\left[-\int\frac{\sigma_x\rho}{m_p}\,dl\right]\right\rangle \approx \left\langle\int\frac{\sigma_x\rho}{m_p}\,dl\right\rangle, \tag{4.1}$$

where the last approximate equality is true when the cross section is small,

$$\sigma_x < \sigma_{\rm crit}\,, \qquad \sigma_{\rm crit} \equiv m_p R_\odot^2/M_\odot = 4.0\times10^{-36}\ \mathrm{cm}^2 \tag{4.2}$$

(compare eq. [2.7] of Paper I). In equation (4.1) the differential dl denotes arc length along an orbit, and the angle brackets denote averaging over capturable orbits.

Equations (4.1) and (4.2) can be combined as

$$f_\sigma \approx \frac{\sigma}{\sigma_{\rm crit}}\left\langle\int\frac{\rho}{M_\odot/R_\odot^3}\frac{dl}{R_\odot}\right\rangle. \tag{4.3}$$

Here one sees that the term in average brackets is a dimensionless structure factor of order unity which depends only on the run of density in the solar model. One evaluates that factor as follows: First note from § II that the regime of interest in (E, J)-space is strongly in the large deflection region. This means that, inside the Sun, a cosmion's orbit is well approximated by a zero-energy orbit (a parabola in the Keplerian case). Therefore, to good approximation, we need not sample orbits in both E and J, but only in J with $E = 0$.

Next define dimensionless variables

$$\hat{J}^2 \equiv J^2/GM_\odot R_\odot\,, \qquad \hat{r} \equiv r/R_\odot\,, \qquad \hat{u} \equiv 1/\hat{r}\,, \qquad \hat{M}(r) \equiv M(r)/M_\odot\,. \tag{4.4}$$

Then the standard equations for central force orbital motion inside the Sun can be written as

$$\frac{d^2\hat{u}}{d\theta^2}+\hat{u} = \frac{\hat{M}(r)}{\hat{J}^2} \tag{4.5}$$

with the starting values at the solar ingress ($\theta = 0$)

$$\hat{u} = 1\,, \qquad \frac{d\hat{u}}{d\theta} = \sqrt{\frac{2}{\hat{J}^2}-1}\,. \tag{4.6}$$

Here θ is polar angle in the orbital plane, i.e., differs from the true anomaly by only an additive constant.

One easily derives the relation

$$\frac{dl}{R_\odot} = \sqrt{\frac{\hat{r}^3}{\hat{J}^2}}\,d\theta\,, \tag{4.7}$$

from which the integral in equation (4.3) can conveniently be accumulated during numerical integration of equations (4.5)–(4.6) in θ.

We have done this for a standard solar model, averaging over orbits with $E = 0$ and $0 < J^2 < 2GM_\odot R_\odot$. The result is

$$f_\sigma \approx 0.89 \frac{\sigma}{\sigma_{\rm crit}} \qquad (\sigma < \sigma_{\rm crit}) . \tag{4.8}$$

Of course, for $\sigma > \sigma_{\rm crit}$, f_σ saturates at the value 1.

V. RESULTS

From equations (2.12), (3.13), and (4.8), and from the relations $v_{\rm esc}^2 = 2GM_\odot/R_\odot$ and $\rho_x \equiv m_x n_x$, we have the result

$$\left(\begin{array}{c}\text{solar abundance}\\ \text{by number}\end{array}\right) = \frac{\text{number captured}}{M_\odot/m_p} = \frac{m_p}{M_\odot}\tau_\odot \mathscr{F}_{\rm capturable} f_\sigma = 0.89\sqrt{\frac{3\pi}{2}}\frac{\rho_x (R_\odot^2 v_{\rm esc}\tau_\odot)}{M_\odot}\frac{v_{\rm esc}}{\bar v}\frac{\sigma}{\sigma_{\rm crit}}\frac{m_p}{m_x}\min\left(1, \frac{21}{2}\frac{m_p}{m_x}\frac{v_{\rm esc}^2}{\bar v^2}\right). \tag{5.1}$$

Here $\tau_\odot$ is the age of the Sun, 5×10^9 yr.

We want to consider the two possibilities, that cosmions provide the Galaxy's missing mass (1) in the Galactic disk, or (2) in an approximately spherical halo. In the former case, their mass density could be as large as $0.1\ M_\odot/{\rm pc}^{-3}$, while their velocity dispersion $\bar v$ could be as small as tens of kilometers per second (Bahcall 1984). In the latter case, the mass density is likely no larger than $0.01\ M_\odot/{\rm pc}^{-3}$, and the velocity dispersion should be on the order of 300 km s^{-1}.

For $\bar v < 300$ km s^{-1}, the "min" function in equation (5.1) is equal to 1 as long as $m_x < 42 m_p$, which pretty much covers the whole range of our interest (see Paper I). We therfore can drop that function henceforth. The physical interpretation is that virtually all of the Maxwell tail of cosmions is capturable. Evaluating the numerical factors in equation (5.1) gives

$$\left(\begin{array}{c}\text{solar abundance}\\ \text{by number}\end{array}\right) = 3 \times 10^{-9}\left(\frac{\rho_x}{1\ M_\odot/{\rm pc}^{-3}}\right)\frac{v_{\rm esc}}{\bar v}\frac{\sigma}{\sigma_{\rm crit}}\frac{m_p}{m_x} . \tag{5.2}$$

According to Paper I, an abundance of about 10^{-12} is required to reconcile the solar neutrino signature. We see that a factor of about 3000 is thus available to be divided up among the ratio factors in equation (5.2). The most stressful case is for the cosmion Galactic halo: If $m_x/m_p \approx 5$, and $\sigma \approx \sigma_{\rm crit}$, then we do succeed in capturing enough cosmions, but by only a factor of ~ 10. If σ is decreased, the number captured decreases, and the number required also increases (Paper I, eq. [2.11]), so that there is only a factor of ~ 3 leeway on the cross section; it must be very close to $\sigma_{\rm crit}$.

If the cross section is made *larger* than $\sigma_{\rm crit}$, then the cross section factor in equation (5.2) saturates at unity (since all kinematically capturable cosmions are in fact captured). However, the number of cosmions required, according to Paper I, increases; the net effect is to allow cross sections no larger than $\sim 10\sigma_{\rm crit}$. The overall conclusion for a spherical Galactic halo of cosmions is that a range of about a factor of 30 is allowed in the assumed (or "predicted") cross section, if the cosmions are both to solve the solar neutrino problem and be the Galactic missing mass.

If the cosmions are concentrated into the Galactic disk, then we can gain an extra factor of ~ 10 in their local density, and also an extra factor of ~ 10 in the factor $v_{\rm esc}/\bar v$. This allows the cross section to be decreased by a further factor of ~ 10, or increased by a further factor of ~ 100. The overall conclusion is that a range of about a factor of 3×10^4 of cross section is allowed. Alternatively, if the cross section is close to the optimal $\sigma_{\rm crit}$, one can tolerate some orders of magnitude of adverse phenomenology, for example the realistic possibility (Paper III) that cosmion annihilation cross sections are not negligible.

One must note that concentration of cosmions into the Galactic disk cannot have been achieved through dissipational processes like those that concentrate gas, since cosmions are collisionless. Dynamical processes at the time of Galaxy formation can yield a significantly flattened disk, but we do not know if a factor of ~ 10 is achievable.

We conclude that there is a small, but viable, region of parameter space in which a Galactic population of hypothetical, stable particles would naturally solve, simultaneously, both the problem of missing mass in the Galaxy and, by calculated capture into the Sun, the solar neutrino problem. The particle should have a mass between 5 and 60 GeV (Paper I), a scattering cross section on baryons within striking distance of 4×10^{-36} cm^2, a density in the solar neighborhood of 0.01–0.1 $M_\odot$ pc^{-3}, a velocity dispersion of $\sim$30–300 km s^{-1}, and (as will be shown in Paper III) a suppressed low-energy annihilation cross-section.

For discussions, we thank Katherine Freese, Lawrence Krauss, and John Bahcall. This work was supported in part by the National Science Foundation, grant no. PHY-83-06693. D. N. S. is supported by an NSF predoctoral fellowship.

REFERENCES

Bahcall, J. N. 1984, *Ap. J.*, **287**, 926.
———. 1985, in *Highlights in Astrophysics: Concepts and Controversies*, ed. S. Shapiro and S. Teukolsky (New York: Wiley), in press.
Bahcall, J. N., Huebner, W. F., Lubow, S. H., Parker, P. D., and Ulrich, R. K. 1982, *Rev. Mod. Phys.*, **54**, 767.
Faulkner, J., and Gilliland, R. L. 1985, *Ap. J.*, **299**, in press.
Fowler, W. A. 1982, in *Workshop on Science Underground* (Am. Inst. Phys. Conf. Proc., No. 96), p. 80.
Krauss, L. M., Freese, K., Spergel, D. N., and Press, W. H. 1985, *Ap. J.*, **299**, in press (Paper III).
Steigman, G., Sarazin, C. L., Quintana, H., and Faulkner, J. 1978, *A.J.*, **83**, 1050.
Spergel, D. N., and Press, W. H. 1985, *Ap. J.*, **294**, 663 (Paper I).

WILLIAM H. PRESS and DAVID N. SPERGEL: Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138

The Photino, the Sun, and High-Energy Neutrinos

Joseph Silk
Department of Astronomy, University of California, Berkeley, California 94720

and

Keith Olive
Theoretical Astrophysics Group, Fermi National Accelerator Laboratory, Batavia, Illinois 60510

and

Mark Srednicki
Physics Department, University of California, Santa Barbara, California 93106
(Received 29 April 1985)

If the Universe contains a nearly critical density of photinos which are also assumed to constitute the dark matter in our galactic halo, then gravitational trapping by the Sun and ensuing annihilation in the solar core yields a significant flux of ~ 250-MeV neutrinos. This results in about two neutrino-induced events per kiloton-year in an underground proton-decay detector.

PACS numbers: 96.60.Kx, 14.80.Pb, 96.40.Qr, 98.60.Ln

The photino, expected to be a stable, massive relic of supersymmetry, is a promising dark-matter candidate.[1] The photino mass range that may be relevant to interpretations[2] of CERN monojets (1–10 GeV) also can yield a critical closure density for the Universe.[3] A critical-density, spatially flat Friedmann cosmological model dominated by massive, weakly interacting particles reconciles several outstanding cosmological issues, including the isotropy of the microwave background,[4] the light-element abudances,[5] and predictions of inflationary cosmology.[6] However, a plausible weakly interacting dark-matter candidate is required, and the case for the photinos seems especially intriguing.

Two of us have recently noted[7] that if our galactic halo, known to consist largely of dark matter, is plausibly assumed to consist of the same dark matter that binds the Universe, then the identification of the photino as the dark-matter candidate leads to a unique, observable signature. For a photino mass $m_{\tilde{\gamma}} \sim 3$ GeV, photino annihilations in our halo were found to lead to a detectable flux of low-energy cosmic-ray antiprotons, below the kinematic threshold for secondary production by high-energy cosmic-ray interaction with interstellar matter.

In this Letter, we point out another signature of photino annihilations. The Sun gravitationally traps halo photinos,[8,9] which subsequently lose energy elastically and annihilate in the solar core. The yield of ~ 250-MeV neutrinos resulting from trapped-photino annihilations in the Sun is appreciable: it may be detectable in ongoing undergound experiments.

In the ensuing discussion, we will be interested in two cross sections involving photinos, the elastic-scattering cross section σ_E, and the annihilation cross section σ_A. The present mass density of photinos is basically determined by σ_A at the time annihilations freeze out. In order to achieve a cosmological mass density corresponding[10] to $\Omega = 1$, we need $\langle \sigma v \rangle_{A,F} \approx 10^{-26}$ cm^3 s^{-1} (for $h = \frac{1}{2}$) where $\langle \sigma v \rangle_{A,F}$ is the thermally averaged product of the annihilation cross section and relative velocity at the freezeout temperature. At lower velocities, the p-wave annihilation channel is suppressed and

$$\langle \sigma v_A \rangle = \langle \sigma v \rangle_{A,F} (1 + 0.04 m_{\tilde{\gamma}}^2)^{-1} \tag{1}$$

where $m_{\tilde{\gamma}}$ is in gigaelectronvolts and we have assumed that all scalar-quark and -lepton masses are equal. In terms of the low-energy annihilation cross section [Eq. (1)], we determine the elastic scattering cross section[11]:

$$\langle \sigma v \rangle_E = 0.17 \beta \langle \sigma v \rangle_A = 1.7 \times 10^{-27} \beta (1 + 0.04 m_{\tilde{\gamma}}^2)^{-1} \text{ cm}^3 \text{ s}^{-1}, \tag{2}$$

where $\beta = v/c$.

Photino annihilations in the Sun are, of course, subject to photinos first hitting the Sun and becoming trapped. Press and Spergel[9] have calculated the trapping rate and find that

$$\dot{N}_{\tilde{\gamma}}^{\text{trap}} = 9 \times 10^{28} n_{0.3} \sigma_{36} \bar{v}_{300}^{-1} m_{\tilde{\gamma}}^{-1} \text{ s}^{-1}, \tag{3}$$

where $n_{0.3} \simeq 0.3$ cm^{-3} and the average halo density of photinos $n_H = n_{0.3} m_{\tilde{\gamma}}^{-1}$, $\sigma_{36} = \sigma_E / (10^{-36} \text{ cm}^2)$ and $\bar{v}_{300} = \bar{v}/(300 \text{ km s}^{-1}) \approx 0.9$ is the rms velocity of a halo photino in the solar neighborhood [$= (3/2)^{1/2} \times$ galactic rotation velocity in the solar neighborhood for an isothermal dark halo model].

The number of photinos in the Sun can diminish in

© 1985 The American Physical Society

one of two ways: evaporation and/or annihilation. Steigman *et al.*[8] and Press and Spergel[12] have shown that unless $m_{\tilde{\gamma}} \lesssim 5$ GeV, evaporation is not important. Given the annihilation cross section [Eq. (1)], the annihilation rate is

$$\dot{N}_{\tilde{\gamma}}^{\text{ann}} = (4/3)\pi R_{\odot}^{3} n_{\tilde{\gamma}}^{2} \langle \sigma v \rangle_{A} f \approx 5\times 10^{54} (n_{\tilde{\gamma}}/n_{p})^{2} \langle \sigma v \rangle_{A,26} f \text{ s}^{-1}, \quad (4)$$

where $n_{\tilde{\gamma}}$ (n_p) is the mean photino (proton) density in the Sun, $\langle \sigma v \rangle_{A,26} = \langle \sigma v \rangle_{A}/(10^{-26} \text{ cm}^3 \text{ s}^{-1})$, and f is a density weighting factor ($f = \int n^2 \, dV / n^{-2} V$). Hence trapping can maintain a mean solar photino abundance

$$\frac{n_{\tilde{\gamma}}}{n_p} \simeq 1.3\times 10^{-13} n_{0.3}^{1/2} \left(\frac{\sigma_{36}}{\langle \sigma v \rangle_{A,26}} \right)^{1/2} \times m_{\tilde{\gamma}}^{-1/2} \bar{v}_{300}^{1/2} f^{-1/2}. \quad (5)$$

The annihilation rate is thus set by the trapping rate.

The only annihilation products that can escape from the solar interior are the high-energy neutrinos. These typically are expected[7] to have energies $\lesssim 0.05 m_{\tilde{\gamma}}$. It is now relatively straightforward to compute the expected flux of energetic neutrinos of type ν_i on earth:

$$\phi_{\nu_i} = \tfrac{1}{2} N_{\nu_i} \dot{N}_{\tilde{\gamma}}^{\text{trap}} / 4\pi (1 \text{ AU})^2 \text{ cm}^{-2} \text{ s}^{-1} \approx 16 N_{\nu_i} n_{0.3} \sigma_{36} \bar{v}_{300}^{-1} m_{\tilde{\gamma}}^{-1} \text{ cm}^{-2} \text{ s}^{-1}, \quad (6)$$

where N_{ν_i} is the number of neutrinos of type ν_i produced in each annihilation.

The most easily detectable process is the reaction $\bar{\nu}_e + p \rightarrow n + e^+$ which has a cross section $\sigma = 7.5 \times 10^{-38} [E_{\nu}/(1 \text{ GeV})]^2 \text{ cm}^2$. Given the flux from Eq. (6) we estimate the number of events in a water detector per kiloton-year,

$$N_E = \phi \sigma N_p, \quad (7)$$

where $N_p \simeq 6.7\times 10^{31}$ is the number of protons in one kiloton of H_2O. Thus we expect

$$N_E = 0.44 n_{0.3} \bar{v}_{300} m_{\tilde{\gamma}} \text{ events/kiloton year} \quad (8)$$

for $\bar{\nu}_e$'s with energy $E_{\bar{\nu}} \simeq 0.05 m_{\tilde{\gamma}}$ where[7] $N_{\bar{\nu}_e} = 2.5$ and $\sigma_{36} = 0.028$. Hence for $m_{\tilde{\gamma}} > 5$ GeV we expect $N_E > 2.2$ events/kiloton year. This is comparable to the expected event rate from atmospheric neutrinos[13-15] for $E \sim 250$ MeV. We note that there has been reported an excess of neutrino events around the 100–200-MeV range by the Kamiokande group.[13] Other sources of high-energy neutrinos have been discussed by Dar.[14]

It has been suggested[8,9,12] that a population of weakly interacting particles inside the Sun could solve the solar neutrino problem. However, in the case of photinos, the elastic scattering cross section appears to be too small for this effect to work.[11,16]

In summary, we predict that the existence of a dark galactic halo consisting of photinos leads to an appreciable flux of high-energy neutrinos on Earth due to photino annihilations in the Sun. We have estimated the event rate in a proton-decay-type water detector and found that the predicted event rate is comparable to atmospheric backgrounds. This effect depends only on the existence of dark matter composed of heavy ($m > 5$ GeV) weakly interacting particles, in the galactic halo. Our results carry uncertainties primarily in the estimate of the cross section and the number of neutrinos per decay channel, both of which depend on the masses of scalar quarks and leptons. We consider this in more detail elsewhere.[11] Here we specialized by considering photinos in particular but these results can be generalized to other forms of dark matter such as scalar neutrinos where this effect may be enhanced. In a forthcoming publication[11] we look at these other possibilities as well as a more detailed examination of the capability of detection.

One of us (M.S.) is supported in part by the National Science Foundation through Grant No. PHY83-13324. Another (J.S.) is supported in part by the U. S. Department of Energy through Grant No. 84-ER40161.

[1]S. Weinberg, Phys. Rev. Lett. **50**, 387 (1983).

[2]R. M. Barnett, H. E. Haber, and G. L. Kane, Lawrence Berkeley Laboratory Report No. 18990 (to be published).

[3]H. Goldberg, Phys. Rev. Lett. **50**, 1419 (1983); L. Krauss, Nucl. Phys. **B227**, 556 (1983); J. Ellis, J. S. Hagelin, D. V. Nanopoulos, K. A. Olive, and M. Srednicki, Nucl. Phys. **B238**, 453 (1984).

[4]N. Vittorio and J. Silk, Astrophys. J. **285** L39 (1984); J. R. Bond and G. Efstathiou, Astrophys. J. **285**, L44 (1984).

[5]J. Yang, M. Turner, G. Steigman, D. Schramm, and K. Olive, Astrophys. J. **281**, 493 (1984).

[6]A. Guth, Phys. Rev. D **23**, 347 (1981).

[7]J. Silk and M. Srednicki, Phys. Rev. Lett. **53**, 624 (1984).

[8]Trapping of weakly interacting dark matter by the Sun was first discussed by G. Steigman, C. L. Sarazin, H. Quintana, and J. Faulkner, Astron. J. **83**, 1050 (1978).

[9]W. H. Press and D. N. Spergel, Harvard Smithsonian Center for Astrophysics Report No. 2121 (to be published).

[10]Ω is the ratio of the cosmological density to the critical closure value in an Einstein–de Sitter cosmological model; h is Hubble's constant in units of 100 km s^{-1} Mpc^{-1}.

[11]M. Srednicki, K. A. Olive, and J. Silk, to be published.

[12]W. H. Press and D. N. Spergel, Harvard Smithsonian Center for Astrophysics Report No. 2107 (to be published).

[13]Y. Totusa, in *Proceedings of the Nineteenth Recontre de Moriond, La Plagne, France, 1984,* edited by J. Tran Thanh

Van (Editions Frontières, Paris, 1984) Vol. 1, p. 625; T. Suda, in *Proceedings of the Eleventh International Conference on Neutrino Physics and Astrophysics, Dortmund, West Germany, 1984,* edited by K. Kleinknecht and E. A. Paschos (World Scientific, Singapore, 1985).

[14]A. Dar, Technion Report No. PH-84-54 (to be published).

[15]T. Gaisser, T. Stanev, S. Bludman, and H. Lee, Phys. Rev. Lett. **51**, 223 (1983); A. Dar, Phys. Rev. Lett. **51**, 227 (1983).

[16]L. Krauss, K. Sreese, D. M. Spergel, and W. H. Press, to be published.

LIMITS ON THE FLUX OF ENERGETIC NEUTRINOS FROM THE SUN

J.M. LOSECCO [a], J.C. VAN DER VELDE [b], R.M. BIONTA [b], G. BLEWITT [c], C.B. BRATTON [d], D. CASPER [b,e], R. CLAUS [e], A. CIOCIO [e], B. CORTEZ [c], S. DYE [f], S. ERREDE [b], G. FOSTER [b], W. GAJEWSKI [g], K.S. GANEZER [g], M. GOLDHABER [h], T.J. HAINES [g], T.W. JONES [i], D. KIELCZEWSKA [j], W.R. KROPP [g], J.G. LEARNED [f], E. LEHMANN [c], J. MATTHEWS [b], H.S. PARK [b], F. REINES [g], J. SCHULTZ [g], S. SEIDEL [b,e], E. SHUMARD [b], D. SINCLAIR [b], H.W. SOBEL [g], J.L. STONE [e], L. SULAK [e], R. SVOBODA [f], G. THORNTON [b] and C. WUEST [g]

[a] *University of Notre Dame, Notre Dame, IN 46556, USA*
[b] *University of Michigan, Ann Arbor, MI 48109, USA*
[c] *California Institute of Technology, Pasadena, CA 91125, USA*
[d] *Cleveland State University, Cleveland, OH 44115, USA*
[e] *Boston University, Boston, MA 02215, USA*
[f] *University of Hawaii, Honolulu, HI 96822, USA*
[g] *University of California, Irvine, CA 92717, USA*
[h] *Brookhaven National Laboratory, Upton, NY 11973, USA*
[i] *University College, London WC1E 8BT, UK*
[j] *Warsaw University, PL-00-681 Warsaw, Poland*

Received 15 December 1986

A number of authors have proposed mechanisms by which the sun could be a strong source of energetic neutrinos. We search for an excess signal of penetrating neutral particles from the direction of the sun. We employ two data samples. One sample studies energies from 400 MeV to 2 GeV. The other studies ν_μ interactions above 2 GeV where the atmospheric background is lower. Our results are compared with the general background of atmospheric neutrinos from other directions. No significant excess has been found. These results can be used to set limits on possible dark matter candidates.

The nature of the dark matter implied by astronomical observations is not well understood. A number of authors [1] have pointed out that several (cold) dark matter candidates could give additional observable signals. They argue that dark matter would lose energy in passing through matter and so could become gravitationally trapped in astronomical objects such as the sun. Such dark matter would accumulate in the sun until the point where the rate for annihilation with other particles of the same type equalled the capture rate.

The products of annihilation would include neutrinos coming from the prompt decay of heavy flavors produced in the annihilation. Neutrinos from the decay of heavy flavors tend to have energies significantly above that of neutrinos coming from typical nuclear processes in the sun. In addition both muon and electron neutrinos would be produced with comparable rates.

A number of authors [1] have calculated the flux of neutrinos at earth due to such processes in the sun. The fluxes tend to be large; comparable or greater than the ambient fluxes of neutrinos produced by cosmic ray interactions in the earth's atmosphere.

The IMB detector [2] is an 8000 metric ton deep underground detector originally designed to search for proton decay. During 417 days of livetime 401 contained interactions were observed [3] inside the 3300 metric ton fiducial volume. These events have the properties expected from neutrino interactions. The energy of these interactions extends from 200 MeV to about 2 GeV. The energy, position, time, direction and track multiplicity of these events is measured. In addition 187 upward going muons [4]

0370-2693/87/$ 03.50 © Elsevier Science Publishers B.V. (North-Holland Physics Publishing Division)

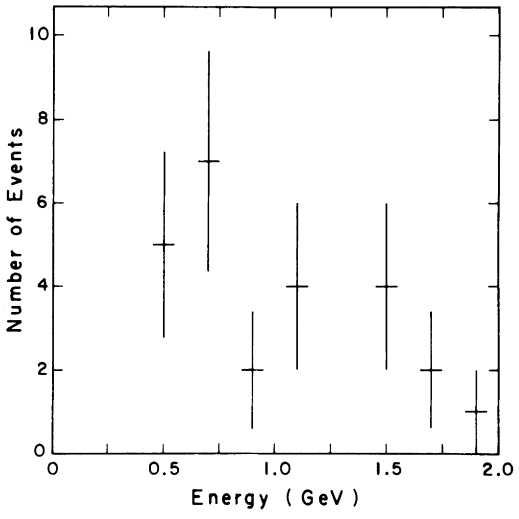


Fig. 1. Energy spectrum for all single prong contained neutrino interactions within 30° of the sun.

indicative of muon neutrino interactions in the shielding rock surrounding the detector were observed in a period of 396 days of livetime. This signal comes from neutrinos with energies in excess of 2 GeV. Only the direction, time and entry point are measured for these events. Since the interaction occurs in rock outside the detector the energy cannot be measured and the observation represents an integral of all flux above the threshold. Due to the significantly higher energy of the parent neutrinos the angle between the reconstructed muon and the neutrino is small. In general it is better than the angular resolution of the detector.

The global rate observed for both of these data samples are in good agreement with expectations for atmospheric neutrinos.

A search has been conducted for an excess rate from the direction of the sun. Fig. 1 shows the neutrino energy distribution for the 25 single prong contained events found within 30° of the sun. Fig. 2 shows the distribution for the total sample of 310 single prong events. The conversion from the measured visible energy to neutrino energy is an approximation [3] based on the simulated response of the detector to single prong neutrino interactions.

It is anticipated that any excess flux would be found at high energy. Table 1 lists the contents of the energy bins. The expectation is based on the solid angle subtended. A small excess is found above 1 GeV. The upper bound is computed [5] using the data sample at all other angles as a background estimate. This statistical technique [5] takes into account fluctuations in both the on source and off source measurements in deriving a 90% confidence level bound for an on source signal. The ratio of observed over expected for energies above 1 GeV is 1.66 ± 0.54. This is not significant. The total sample at all energies is slightly high. The observed over expected rate is 1.20 ± 0.25. This may be a manifestation of the detector acceptance. The detector is longer in the east–west direction than in any other. The sun follows an east–west

Table 1
Contained events within 30° of the sun.

Energy of events (MeV)	Total at all angles	Events near the sun	Expected number of events	Excess events 90% CL upper bound
400– 600	49	5	3.3 ± 0.5	6.3
600– 800	111	7	7.4 ± 0.7	10.3
800–1000	51	2	3.4 ± 0.5	3.3
1000–1200	35	4	2.3 ± 0.4	5.8
1200–1400	15	0	1.0 ± 0.3	2.3
1400–1600	18	4	1.2 ± 0.3	6.9
1600–1800	9	2	0.6 ± 0.2	4.7
1800–2000	12	1	0.8 ± 0.2	3.3
$\geqslant 2000$	10	0	0.7 ± 0.2	2.3
$\geqslant 1000$	99	11	6.6 ± 0.7	15.1
total	310	25	20.8 ± 1.2	31.2

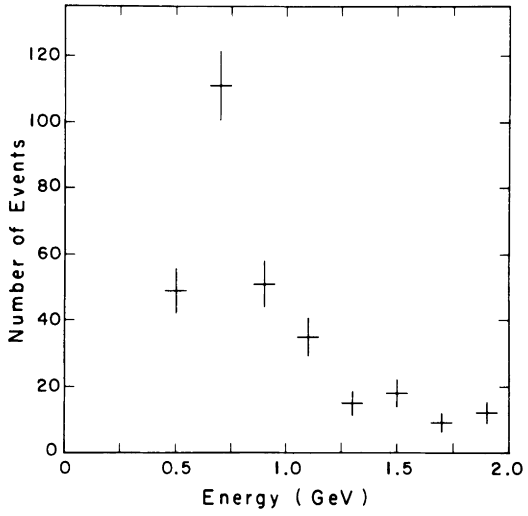


Fig. 2. Energy spectrum for all single prong contained neutrino interactions.

path and so is aligned with the detector's major axis. We can remove the detector acceptance by normalizing the high energy events to the overall flux in the suns direction. In that case we find a ratio of observed events to expected events for energies above 1 GeV of 1.38 ± 0.54. So there is no evidence for excess events at high energies.

Angular cuts consistent with the resolution and the reconstructed neutrino direction can be used to reduce the overall background. The angular resolution is energy dependent. In table 2 we list the angular resolution (hwhm) on a bin-by-bin basis. The angular resolution is determined by Monte Carlo studies of neutrino events taken from bubble chamber experiments. The resolution quoted includes both scattering and detector effects. Angular cuts of the quoted value have been used to get the limits in table 2. The background subtraction method of ref. [5] has been used to get the 90% CL bound. It is clear that the 30° angular cut chosen above is conservative for high energies ($E_\nu \geqslant 1$ GeV).

The bottom line of table 2 above indicates an efficiency corrected excess of less than 4.9 events/kiloton-year in the direction of the sun. This would be the sum of ν_e and ν_μ interactions. For explicit limits we use, as an example, the calculations of Gaisser, Steigman and Tilav [1]. In the context of their calculation this result can exclude

$$3\ \text{GeV} < m_{\nu_D} < 24\ \text{GeV}$$

and

$$m_{\nu_D} > 80\ \text{GeV}, \quad m_{\nu_M} > 80\ \text{GeV}, \quad m_{\tilde{\nu}} > 3\ \text{GeV},$$

and a region of

$$m_{\tilde{\gamma}} < 4\ \text{GeV}$$

and

$$m_S < 62\ \text{GeV}.$$

m_{ν_D} is the Dirac neutrino mass, m_{ν_M} is the Majorana neutrino mass, $m_{\tilde{\nu}}$ is the scalar neutrino mass, $m_{\tilde{\gamma}}$ is

Table 2
Contained events near the sun.

Energy of events (MeV)	Total at all angles	Angular resolution (hwhm)	Events near the sun	Excess events 90% CL upper bound
400– 600	49	51°	9	6.5
600– 800	111	47°	20	25.5
800–1000	51	41°	6	5.4
1000–1200	35	36°	6	7.5
1200–1400	15	29°	0	2.3
1400–1600	18	21°	1	3.4
1600–1800	9	16°	0	2.3
1800–2000	12	16°	1	3.7
⩾2000	10	16°	0	2.3
⩾1000	99	20°	5	6.5
total	310	18°	9	12.7

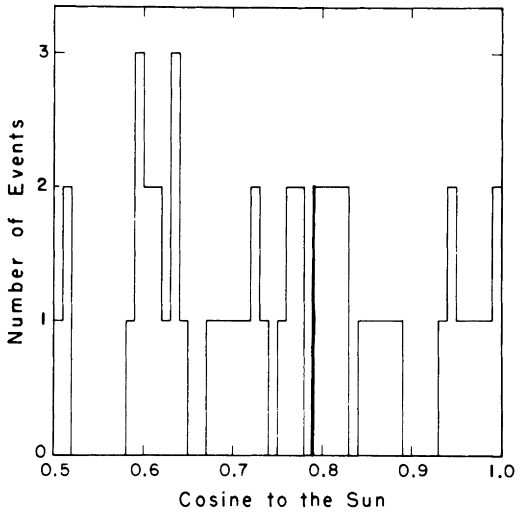


Fig. 3. Cosine of upward muons with respect to the sun. Only muons within 60° of the sun are plotted.

the photino mass and m_S is the scalar quark mass. Their are many assumptions embodied in their calculation. Our results can be more restrictive as tabulated above since the actual spectrum is also bounded.

The contained sample does not include events with energies in excess of 2 GeV. These tend to exit the detector and are removed by our filtering cuts. Instead we have a sample of upward muons which come from muon neutrino interactions in the rock below the detector. Electron neutrino interactions in the rock are generally not visible to us. Only muons travelling upward are used in this sample since penetrating muons coming from the surface make it impossible to isolate the energetic neutrino sample from above. This is not a limitation in studying the sun since the detector is operated 24 hours a day and the sun spends a good deal of the time below the horizon. This data sample has good angular resolution and correlates well with the parent neutrino direction. The exact correlation [4] depends on the parent neutrino energy but in general it is less than 7°.

Fig. 3 shows the distribution of events within 60° of the sun. Table 3 shows how the distribution varies with angle and normalizes the observation to the total sample observed at all angles.

The two events found nearest the sun are more than expected but no conclusions can be drawn from such small statistics. Caution is advised in interpreting these results since the data is an integral of all neutrinos above 2 GeV times an energy dependent detection efficiency. The efficiency tends to rise with energy due to the increase in neutrino cross section and the increase in effective detectors mass due to the high muon range. These two effects tend to offset the steep drop in the ambient neutrino flux at these energies.

The observed rate is compatible with the Volkova [6] prediction. A flux limit can be quoted from the table. Within 8° of the sun we have an *excess* of less than 4.64 times the atmospheric neutrino rate at 90% confidence level. Very crudely [6], we estimate that the flux is $\leqslant 3.7 \times E_\nu^{-3.69} \nu/(\mathrm{cm}^2\ \mathrm{s\ GeV})$.

To conclude we note that both of our results are higher than expected but are consistent with expectations from atmospheric neutrinos. No result is more than 1.4 standard deviations high. More statistics would be needed to determine if the excess is significant. These limits are still useful to place constraints on possible dark matter candidates.

Table 3
Upward muon events nearest the sun.

Minimum cosine	Maximum angle	Number of events	Fraction of total	Fraction limit 90% CL	Observed over expected
0.99	8°	2	0.011±0.008	0.029	2.14±1.52
0.98	11.5°	3	0.016±0.009	0.036	1.60±0.90
0.97	14°	4	0.021±0.011	0.043	1.40±0.73
0.96	16.25°	5	0.027±0.012	0.050	1.35±0.60
0.95	18.2°	6	0.032±0.013	0.056	1.28±0.52
0.90	25.8°	9	0.048±0.016	0.076	0.96±0.33
0.50	60°	51	0.273±0.033	0.319	1.09±0.13

This work was supported in part by the US Department of Energy. One of us (JML) was supported by a DOE Outstanding Junior Investigator grant. We would like to acknowledge stimulating conversations with J. Ellis, K. Olive, G. Steigman, G. Kane, T. Gaisser and J. Silk. JML would like to acknowledge the 1986 Lake Louise Winter Institute where this work was initiated.

References

[1] J. Silk, K. Olive and M. Srednicki, Phys. Rev. Lett. 55 (1985) 257;
B.A. Campbell, J. Ellis, K. Enqvist, D.V. Nanopoulos, J.S. Hagelin and K.A. Olive, Phys. Lett. B 173 (1986) 270;
M. Srednicki, K.A. Olive and J. Silk, High energy neutrinos from the sun and cold dark matter, UMN-TH-553/86, to be published;
J.S. Hagelin, K.W. Ng and K.A. Olive, Phys. Lett. B 180 (1986) 375;
T.K. Gaisser, G. Steigman and S. Tilav, Phys. Rev. D 34 (1986) 2206.

[2] R.M. Bionta et al., Phys. Rev. Lett. 51 (1983) 27;
S. Errede et al., Phys. Rev. Lett. 51 (1983) 245;
T.W. Jones et al., Phys. Rev. Lett. 52 (1984) 720;
B. Cortez et al., Phys. Rev. Lett. 52 (1984) 1092;
H.S. Park and G. Blewitt et al., Phys. Rev. Lett. 54 (1985) 22.

[3] J. LoSecco et al., Phys. Rev. Lett. 54 (1985) 2299;
G. Blewitt et al., Phys. Rev. Lett. 55 (1985) 2114;
T. Haines et al., Phys. Rev. Lett. 57 (1986) 1986.

[4] R. Svoboda et al., An upper limit on the flux of extraterrestrial neutrinos, UC Irvine preprint, Astrophys. J. (April 15, 1987), to be published;
R.C. Svoboda, A search for astrophysical point sources of neutrinos using a large underground water Cerenkov detector, Ph.D. Thesis, University of Hawaii at Manoa (1985), unpublished.

[5] R.F. Protheroe, Astron. Express 1 (1984) 33.

[6] L.V. Volkova, Sov. J. Nucl. Phys. 31 (1980) 784.

SEARCH FOR DARK MATTER AND NEUTRINO SOURCES WITH THE FREJUS DETECTOR

FREJUS COLLABORATION

presented by B. KUZNIK

Bergische Universität GH Wuppertal
Fachbereich 8
Gaußstr. 20
5600 Wuppertal , GERMANY

L.A.L.-Université Paris Sud
Bât.200
91405 ORSAY Cedex , FRANCE



C. Arpesella[2] , P. Bareyre[4] , R. Barloutaud[4] , J.Becker[5] , K.H. Becker[5] , Y. Benadjal[2] , Ch. Berger[1] , D. Blum[2] , A. Bory[4] , C. Bourdarios[2] , G. Chardin[4] , H.J. Daum[5] , B. Degrange[3] , G. Deuzet[2] , B. Dudelzak[2] , J.Ernwein[4] , P.Eschstruth[2] , G. Gerbier[4] , J.F.Glicenstein[4] , A. Hofmann[1] , S. Jullian[2] , W. Kolton[4] , B. Kuznik[2,5] , D. Lalanne[2] , F. Laplanche[2] , C. Longuemare[2] , Y. Minet[3] , R. Mayer[5] , H. Meyer[5] , L. Mosca[4] , L. Moscoso[4] , R. Möller[5] , U. Nguyen-Khac[3] , C. Paulot[2] , F. Raupach[1] , P. Schleper[1] , M. Schubnell[6] , P. Serri[3] , G. Szklarz[2] , S. Tisserant[3] , B. Tripp[3] , Y. Wei[5] , P. Wintgen[5]

Aachen[1] , Orsay[2] , Palaiseau[3] , Saclay[4] , Wuppertal[5]

ABSTRACT

A high energy neutrino flux from the sun is expected if the sun contains significant amounts of dark matter in the form of weak interacting massive particles, which may be gravitationally bound and annihilate in the solar core.

High energy gamma ray sources could be also sources for high energy neutrinos due to the production in hadronic cascades. A search has been done for high energy neutrinos coming from the sun, Cyg X3, the galactic center and different other galactic objects.

1) INTRODUCTION :

The dark matter problem in cosmology was discussed over a long time. There are different available dark matter candidates of the particle zoo which are distinguished in three different categories by galaxy formation models (1).

1) Cold dark matter , $v \ll c$

2) Warm dark matter , $v \simeq 0.1\ c$

3) Hot dark matter , $v \simeq c$

The picture of galaxies and cluster formation as seen in the universe is in good agreement with the hypothesis of cold dark matter in the biased galaxy formation model (2).

Cold dark matter candidates ($m \geq$ GeV, massive neutrinos, photinos, scalar neutrinos) could be a source of high energy neutrinos. It has been suggested that those candidates produced in the Big Bang and contained in our own galaxy can be captured by the sun and accumulate in the solar core. This matter may annihilate and produce a flux of high energy neutrinos (3,4,5). This flux is predicted to be large enough to be observed in deep underground detectors searching for proton decay.

The search for sources in the universe is the search for photon and neutrino sources, these are the only available particles for astronomy over long distances (assuming the neutrinos are stable), because the galactic magnetic field will distore the path of charged particles and other neutral particles will decay before they reach the earth.

All high energy gamma ray sources will be also sources of high energy neutrinos, if they are produced in the decay of neutral pions. There are also high energy charged pions due to the production mechanism. Their decay will be the source for high energy neutrinos. It has been suggested that the center of our galaxy could be a source of high energy neutrinos (6), and there are calculations that show that objects like Cxy X3 could probably be seen in a large detector (7). It is possible that, if the density of matter in the direction of the source is much bigger than the radiation length, those objects could only be seen in the neutrino signal. Their detection would give us information about the processes occuring in the center of stars or galaxies.

We have performed an analysis of high energy neutrino events obtained with the Frejus nucleon decay detector and set limits on the high energy neutrino event rate from the sun (per Kt x year) and the high energy neutrino fluxes from Cyx X3, the galactic center and different other galactic objects .

2) FREJUS EXPERIMENT

The Frejus laboratory is located in the roadtunnel connecting France (Modane) and Italy (Bardonecchia) in the Alps. The geographical coordinates are 45.2^0 north and 6.7^0 east. It is at a distance of 6.5 km from the entrance of the tunnel with an overburden of $\geq$ 1.450 meter rock (density of 2.75 g/cm^3). Muons which reach the detector traverse an average of 4800 kg/cm^2. The measured flux of crossing muons is 4.2/day m^2 .

The detector is a high resolution iron calorimeter with Geigertubes for the triggering and Flashchambers for the tracking of the particles. The size of the detector is 6m x 6m x 12.5m and he has a total mass of 912 tons. The detector was build for nucleon decay and we have not the possibility to make a fast time of flight measurement, so it is not possible to distinguish upgoing muons (produced in neutrino interactions in the rock) from downgoing muons, except if the muon stops inside the detector.

For the general nucleon decay analysis we have defined a fiducial volume (554 tons) by excluding the outer 50 cm region near the detector faces this corresponds to one hadronic interaction length or seven radiation length (a more complete description is given in ref.8 and ref.9)

3) ANALYSIS

The data have been taken continously from the beginning in 1984 (240 tons) up to 18-11-1986. From june 1985 with the full detector size of 912 tons. The whole livetime corresponds to an exposure of 0.88 Kt years fiducial mass. In that time we have recorded 124 neutrino events with the vertex inside the fiducial volume and we estimated 127 (15 % systematic error) from our Monte Carlo (10) taking into account the atmospheric neutrino / antineutrino flux from T.K.Gaisser (11).

For the present analysis all CC neutrino events with visible energy $E_{vis} \geq 2.$ GeV and visible momentum $p_{vis} \geq 1.25$ GeV/c are included. Some events (mainly muon neutrino interactions , Table 1) are not completely contained in the detector volume and only a lower limit for the visible energy and visible momentum can be given. Also 7 other events are taken into account, which pass the selection criteria. 4 of them have the vertex outside the fiducial volume (50 cm region) and the other 3 events are upward going stopping muons produced in neutrino interactions in the rock.

The resulting data sample corresponds to an effective exposure of 1.14 Kt x year (1.3 x 0.88 Kt x year , 1.3 correction factor for events with vertex outside the fiducial volume) and contains 31 events with an estimated angular resolution of 15^0 (Monte Carlo for quasielastic interactions and resonant single pion production up to W = 2 GeV, will be better when taking into account the deep inelastic part).

The visible energy distribution is shown in fig 1.

A comparison with the Monte Carlo estimation for the events with vertex contained and visible energy $E_{vis} \geq 2.$ GeV is shown in table 1.

Table 1) Comparison of the event rate (vertex in the fiducial volume) with Monte Carlo estimation $E_\nu \geq 2.$GeV without cuts and with experimental cuts $E_{vis} \geq 2.$ GeV and $p_{vis} \geq 1.25$ GeV

Neutrino reaction	Experiment	Monte Carlo	Monte Carlo with cuts
$N_{CC}(\nu_\mu + \bar{\nu}_\mu)$	14	28	15
$N_{CC}(\nu_e + \bar{\nu}_e)$	10	10.6	10

For the analysis we have calculated the direction of origin of the neutrinos in terms of declination (^{0}) and right ascension (RA(h)), the result is shown in fig 2.

The projections to the declination and RA is shown in fig 3a , 3b and compared with the estimation of a Monte Carlo which takes into account the angular dependence of the atmospheric neutrino / antineutrino flux (12). They are in close agreement.

4)NEUTRINOS FROM THE SUN

To search for neutrinos of solar origin we compare the position of the sun in declination and right ascension with the position of origin in the sky of all neutrino events by taking the respective difference of the coordinates, this fixes the sun in the center of fig 4. The corresponding 15^0 angular resolution bin is also shown.

We have no event in the 15 ^{0} errorbox of the sun which corresponds to the energy range $E_{vis} \geq 2.$ GeV.

The calculated event rate upper limits for neutrinos comming from the sun is

2 events / kt year (90 % c.l.) .

This shows that our search for high energy neutrinos $E_{vis} \geq 2.$ GeV coming from

the sun is not limited by the background of atmospheric neutrinos, we estimated in within the angular resolution 0.2 electron neutrino events and 0.3 muon neutrino events per Kt x year. The limit is only determined by the observation time.

As pointed out by the different calculations (3,4,5)the flux ratio of electron neutrinos to muon neutrinos will be sometimes equal or will be in a defined proportion. This is a possibility to distinguish the neutrino signal produced by different cold dark matter candidates, if a signal is found in the future.

5) POINT SOURCES

The correlation of the direction of origin in the sky (Fig.2) with the center of our galaxy and with Cyg X3 is shown in cos(theta) distributions in fig 5a and fig 5b. The position of the objects is at cos(theta) = 1, angular resolution $\cos(15^0) = 0.966$.

The neutrino luminosity of a piont like source is related to the number of events

$$L_\nu = \frac{N\nu * 4 * \pi * d^2}{< \frac{\sigma}{E} > * M * t * N}$$

M = effective mass of the detector

d = distance to the source

$< \frac{\sigma}{E} > = .50 \times 10^{-38}\ cm^2/\text{GeV}$ (assuming $\nu \equiv \bar{\nu}$)

N = Avogadro number

t = running time

By assuming a source distance the luminosity limits (90 % c.l.) can be calculated for the different astrophysical objects. The results are listed in table 2.

Table 2) Neutrino luminosity upper limits (90 % c.l.) for different galactic objects

Source	N candidates	luminosity $\leq$	distance	declination
CygX-3	1	$9.6\ 10^{41}$ ergs/sec	12 kpc	41^0
Gal center	1	$6.7\ 10^{41}$	10 kpc	-25^0
HerX-1	1	$1.7\ 10^{41}$	5 kpc	25^0
Crab	0	$1.7\ 10^{40}$	2 kpc	24^0
SS433	1	$6.0\ 10^{40}$	3 kpc	7^0

6) CONCLUSION :

The search for neutrinos of solar origin shows no event in the 15 ^{0} errorbox of the sun with visible energy $E_{vis} \geq 2.$ GeV. This could be compared to the predictions based on SUSY particles as dark matter candidates as given in the references. Specialy for the calculation of Gaisser, Steigman and Tilav (3) we can rule out the following mass region for the dark matter candidates at 90 % confidence level.

1) Dirac neutrinos from 4 GeV to 32 GeV and above 65 GeV

2) Majorana neutrinos from 6 GeV to 32 GeV and above 75 GeV

3) Scalar neutrinos above 4 GeV complete

4) Photinos from 4 GeV to 12 GeV

With further improvements in the statistic it will be possible to set better limits, $\geq$ 25 GeV for photinos. Our limits may also be very usefull to constrain other models for dark matter candidates in our galaxy.

No evidence for neutrino sources in the sky was found. Up to now all measurements are in close agreement with the estimated event rates from atmospheric neutrino flux .

1) Absolute event rate
2) Event rate with visible energy $E_{vis} \geq 2$ GeV
3) Distribution in the sky plot

REFERENCES

1) J.R.Primack , SLAC - PUB - 3387 July 1984 (T / E / AS)
2) J.R.Primack , SCIPP 86 / 65 June 1986 T
3) T.K.Gaisser , G.Steigman , S.Tilav , Phys. Rev. D 34,8,2206 (1986)
4) J.S.Hagelin , K.W.NG , K.A.Olive , Phys. Letters B 180,4,375 (1986)
5) M.Srednicki , K.A.Olive , J.Silk , Nucl. Phys. B 279,804 (1987)
6) V.S.Berezinsky , Proceedings of the 16th ICRC 10,332 (1979)
7) K.Brecher , G.Chanmugan , Proceedings of the 19th ICRC 1,103 (1985)
8) F.Laplanche , 21st Rencontre de Morion , Edition Frontieres
Progress in Elektroweak Interactions , 505 (1986) edited by Tran Than Van
9) J.Ernwein , 6th Workshop of Grand Unification , 18-20 Apr. 1985 Minnesota
10) C.Longuemare LAL Orsay , Internal Report Taup 3/1986
11) T.K.Gaisser , 11th Conf. on Neutrino and Astrophysics
Dortmund (1984) edited by K.Kleinknecht
12) T.K.Gaisser , Private communication

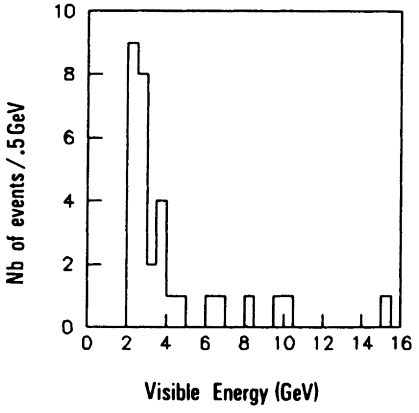


Fig. 1) Visible energy distribution, $E_{vis} \geq 2$ GeV $p_{vis} \geq 1.25$ GeV. 14 events contained,14 vertex contained,3 are stopping muons

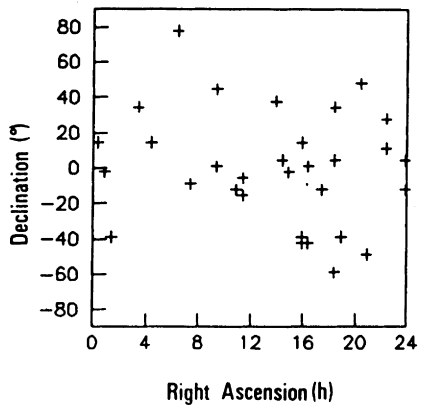


Fig. 2) Origin of the neutrino events in the sky , the estimated angular resolution is 15^{0}.

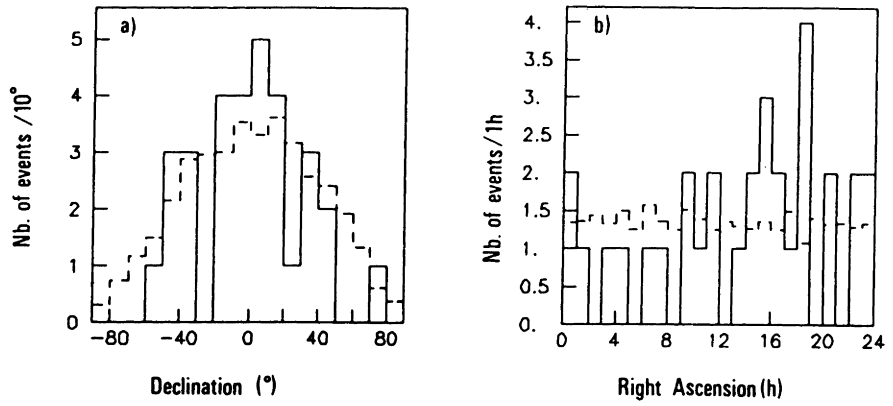


Fig. 3 a,b) Projections of the skyplot (Fig. 2) to the declination and right ascension and the comparison with the Monte Carlo estimation (dashed line).

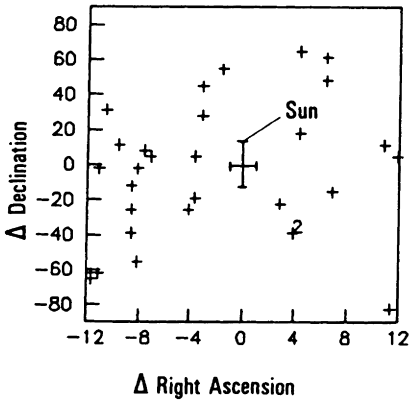


Fig. 4) Delta correlation plot of the neutrino events with the position of the sun in the sky (t = event time)

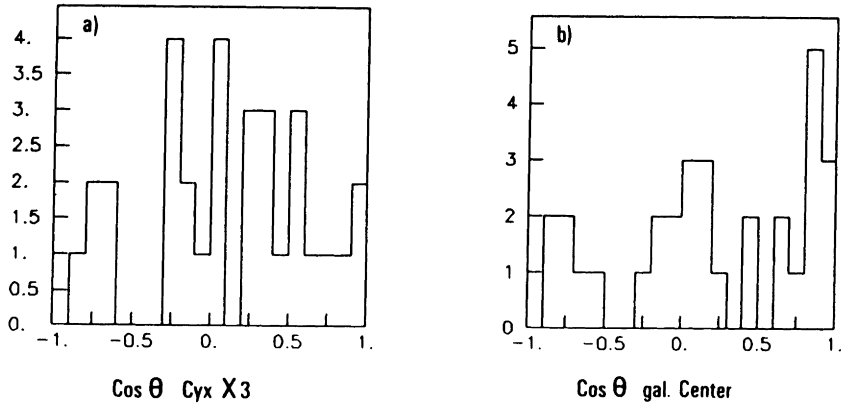


Fig. 5 a,b) Correlation of the neutino events with Cyx X3 and the galactic center

SEARCH FOR HIGH ENERGY NEUTRINOS FROM DARK MATTER ANNIHILATION IN THE SUN

KAMIOKANDE II COLLABORATION

K.S.Hirata, T.Kajita, T.Kifune, M.Nakahata, K.Nakamura,
S.Ohara, Y.Oyama, N.Sato, M.Takita and Y.Totsuka
Institute for Cosmic Ray Research, University of Tokyo, JAPAN

A.Suzuki, K.Takahashi and T.Tanimori
National Laboratory for High Energy Physics (KEK), JAPAN

M.Koshiba
Department of Physics, University of Tokai, JAPAN and CERN, Geneva, Switzerland

T.Suda
Department of Physics, University of Kobe, JAPAN

K.Miyano and M.Yamada
Department of Physics, University of Niigata, JAPAN

E.W.Beier, L.R.Feldscher, E.D.Frank, W.Frati, S.B.Kim,
A.K.Mann, F.M.Newcomer, R.Van Berg and W.Zhang
Department of Physics, University of Pennsylvania, U.S.A.

Presented by M.Takita

Abstract

Recent theoretical studies have indicated that weakly interacting massive particles(WIMP'S) might be dark matter candidates explaining "the missing mass probrem". It is suggested that these WIMP's should be captured and accumulated in the Sun core and then annihilate with their anti-particles producing high energy neutrinos($\geq$1GeV). A search is made for high energy neutrinos from the Sun, based on the 2 different data samples, 342 fully-contained neutrino events(2.82 kton·yr exposure) and 120 neutrino-induced upward-going muons(165.3 m^2·yr exposure) recorded by the KAMIOKANDE detector. No significant excess signal of high energy neutrinos has been observed from the direction of the Sun, which sets mass limits on some of the WIMP's as a dark matter candidate.

Introduction

The problem derives from the astronomical observations suggesting that the total mass of the Universe be "dark". The rotation curves in the various galaxies indicate that the "luminous mass" only accounts for ~20 % of the total mass of the Universe. On the other hand, the standard calculation of the primordial nucleosynthesis, based on the abundance of the light elements (d,^{3}He) in the Universe, constrains the baryon density less than 20 % of the critical density.

To solve this apparent discrepancy, many sorts of non-baryonic particles have been proposed as a candidate explaining "the missing mass". They are categorized into 2 groups;

"hot" particles → light massive ν O(10 eV),

"cold" particles → axion O(10^{-5}eV),

heavy massive ν(Dirac or Majorana) $\geq$1GeV,

lightest SUSY particle(photino or scalar ν) $\geq$1 GeV.

The "hot" particles and axions are omitted from further discussion, because they are too light to produce high energy ordinary neutrinos. These massive "cold" particles floating near the Sun, if exist, will undergo elastic scattering with the nucleons losing their energy and then be gravitationally captured in the Sun core. They start annihilating with their antiparticle producing high energy neutrinos. The equilibrium state finally comes when the annihilation rate becomes half as frequent as the capture rate. The Sun might serve, therefore, as a constant source of high energy neutrinos. A recent calculation[1] of the high energy neutrino flux from the Sun suggests that these high energy neutrinos be detectable by massive(~1kton) underground detectors dedicated to the detection of nucleon decays. The only conceivable background source is atmospheric neutrinos[2),3] produced in the atmosphere of the Earth. As atmospheric neutrinos have an roughly isotropic angular distribution to the Sun direction, what we expect is to observe high energy neutrinos from the Sun above isotropic atmospheric neutrino background.

Two different data samples are adopted for the analysis:

1) Fully-contained neutrino events.

2) Upward-going muon events produced by high energy neutrinos.
The definition of "fully-contained" events is that both event vertex and energy of them are contained in the detector, ensuring correct energy measurement.

The KAMIOKANDE detector

The KAMIOKANDE experiment, located 1000 m underground (2700 m.w.e.) in the KAMIOKA mine(300 km west of Tokyo) in Japan, is divided into 2 phases; KAM I and KAM II. The KAM I detector is a massive water Čerenkov calorimeter which is composed of 3000 metric tons of pure water cylinder, viewed by one thousand of 20"ϕ photomultipliers(PMT's), in a steel tank(15.6 m ϕ × 16 m in height). The photosensitive area covers 20% of the entire inner surface. The KAM II detector is an upgraded version of the KAM I detector, improved with a 4π solid angle anticounter and a new dead-time-free electronics system. The detector details will be found elsewhere[4),5)]. For this analysis of fully-contained neutrino events, only KAM II data is used. The fiducial mass in this analysis is defined by software to be 2140 tons of cylindrical water column (14.44 mϕ×13.08 m), all the photosensitive volume. On the other hand, both KAM I and KAM II data are adopted for the analysis of the neutrino-induced upward-going muons. The nominal detection area is 165 m^2 (150 m^2) for KAM I (KAM II). The trigger threshold for an electron is 30(8.5) MeV/c for KAM I (KAM II) at 50 % efficiency. Event energy is measured in unit of photoelectron(p.e.) and a 1MeV/c electron corresponds to energy deposit of 3.4 p.e. The cosmic-ray muon rate at the detector site is ~0.4Hz.

Analysis of fully-contained neutrino events[6)]

KAM II started data-taking in January 1986 and has accumulated data during a total exposure time of 2.82 kton·yr. Raw data are reduced by the following criteria:
i) Total photoelectron number in the anti-counter $\leq$ 20.
ii) Total number of lit PMT's(threshold=0.3 p.e.) in the anti-counter $<$ 5.
iii) Total photoelectron number in the main detector $>$ 90.

The criteria i) or ii) reject the entering particles such as cosmic-ray muons, while the criterion iii) eliminates low energy γ from the radio active contamination in the water

and from the surrounding rock. After scanning and spatial reconstruction of events, a total of 342 fully-contained events are selected. Since our Monte Carlo study shows that a low energy fully-contained event does not sufficiently remember the direction of its parent neutrino, one more criterion is added:

iv) Total photoelectron number in the main detector > 3000.

The final event sample in this analysis thus contains a total of 72 fully-contained events, of which 45 are single prong events and 27 are multi prong ones. Most of them are induced by ν_e or $\overline{\nu}_e$, since energetic muons exit the detector.

The Monte Carlo estimation shows that a 1 GeV ν_e ($\overline{\nu}_e$) yields 2700 p.e., with the energy resolution obeying to $35\%/\sqrt{\mathrm{E}_{\nu_e}(GeV)}$. Therefore the total photoelectron number can be converted to the corresponding ν_e energy. As the sample contains multi-ring events, The determination of the direction vector **X** of an event, in the present analysis, is specially done as follows:

$$X = \sum_i (p.e.)_i \cdot e^{|\mathbf{r_i}|/L} \cdot \frac{1}{f(\theta_i)} \cdot \frac{\mathbf{r_i}}{|\mathbf{r_i}|} \quad ,$$

where $(p.e.)_i$ is the photoelectron number recorded by the i-th PMT with a displacement vector $\mathbf{r}_i$ from the vertex position, L is the attenuation length of Čerenkov lights in water($\sim$ 50m) and $f(\theta_i)$ is the acceptance correction to the i-th PMT.

The Monte Carlo study exhibits that the angular resolution (detector response + scattering angle) is 35° at E_{ν_e}=2 GeV, decreases down to 10° at E_{ν_e}=10 GeV. Fig.-1 shows a two-dimensional plot for the directional cosine to the Sun ($\cos\theta$) versus E_{ν_e}. Considering the angular resolution, the angular window to the Sun is conservatively set to ($\cos\theta > 0.8$), and then the S/N ratio becomes 11/6.8. One can conclude that there is no significant excess event from the Sun at present.

Analysis of neutrino-induced upward-going muons[6)]

Upward-going muons are produced by very high energy neutrinos interacting with rock layer around the detector. A total of 120 "through-going" muons, with their energies more than 1.7 GeV corresponding to a path length in the detector of 7 m, have been recorded by the KAM I[7)] and KAM II[8)] detectors, during a total effective

exposure time of 165.3 $m^2 \cdot yr$, from July 1983 to October 1987. Fig.-2(a) shows their angular distribution to the Sun, together with an expanded view Fig.-2(b) near the Sun direction. Setting an angular window of 10°-in-radius circle to the Sun, no event has been observed from the Sun. This gives a 90 % confidence-level flux upper limit:

$$4.4 \times 10^{-14} cm^{-2} sec^{-1} / \epsilon \ ,$$

where ϵ is the detection efficiency within the angular window cut above, depending on neutrino energies. The angular resolution due to the event reconstruction error is estimated to be $\Delta_{\theta 1}$=2.1° using Monte Carlo upward-going penetrating muons. On the other hand, the scattering angle $\Delta_{\theta 2}$ between a neutrino and its daughter muon and the angular deviation $\Delta_{\theta 3}$ of the muon while traversing the terrestrial rock are dependent on the energy spectrum of parent neutrinos. Thus the detection efficiency ϵ can be calculated by the information above, $\Delta_{\theta 1}$, $\Delta_{\theta 2}$ and $\Delta_{\theta 3}$.

Mass limits for various dark matter candidates[6)]

The neutrino energy spectra originating from the annihilation of various dark matter candidates , calculated by the authors of Ref.-1), are adopted with a modification, namely, the present analysis adopts the LUND model for the hadronization processes of heavy quarks and the decay of τ leptons, while the treatment of Ref.-1) was based on the simple parton model resulting in favor of higher energy neutrinos. In order to extract the best limits, the angular window to the Sun and the energy cut are determined according to each dark matter candidate and its mass for the analysis of the 72 fully-contained events. In case of the upward-going muons, only the angular window to the Sun is set, depending also on each dark matter candidate and its mass. Preliminary mass limits on various dark matter candidates are summarized in Table 1 It should be noted here that the mass limits in Table 1 are model dependent.

We gratefully acknowledge the cooperation of the KAMIOKA Mining and Smelting Co. This work was supported by the Japanese Ministry of Education, Science and Culture, the United States Department of Energy, and The University of Pennsylvania Research Fund.

References

1) T.K.Gaisser, G.Steigman and S.Tilav, Phys.Rev. D34(1986)2206.
2) T.K.Gaisser, T.Stanev and G.Barr, Bartol Report No. BA-88-1.
3) T.K.Gaisser and T.Stanev, Phys.Rev. D30(1984)985.
4) K.Arisaka et al., J.Phys.Soc.Jpn. 54(1985)3213 and T.Kajita et al., J.Phys.Soc.Jpn. 55(1986)711 for KAM I.
5) K.Hirata et al., Phys.Rev.Lett. 58(1987)1490 for KAM II.
6) N.Sato, Doctoral Thesis, "Search for High Energy Neutrinos from Dark Matter Annihilation in the Sun", Univ. of Tokyo, 1987.
7) Y.Oyama et al., Phys.Rev.Lett. 56(1986)991.
8) Y.Oyama et al., Phys.Rev.Lett. 59(1987)2604.

Table 1. 90 % confidence-level excluded mass ranges for various dark matter candidates.

Dark Matter Candidate	Excluded Mass Range(GeV/c^2)	
	Fully-Contained Events	Upward-going Muons
Majorana ν	NO[1)]	> 76
Dirac ν	10 - 20	> 60
scalar ν_e	3 - 25	NO
scalar ν_μ	NO	> 3
scalar ν_τ	4 - 25	?[2)]
photino	NO	NO

1) The symbol "NO" indicates that no limit is set at 90% C.L.
2) The symbol "?" indicates that the analysis has not been done yet.

Fig.-1. $\cos\theta$ (Sun direction) versus energy of the 72 fully-contained events which survive all the selection criteria. $\cos\theta=1$ indicates the Sun direction. The energy scale is expressed both in unit of total p.e. number and ν_e equivalent energy . The energies of the 6 events on the upper edge of the plot exceeds 15000 p.e.

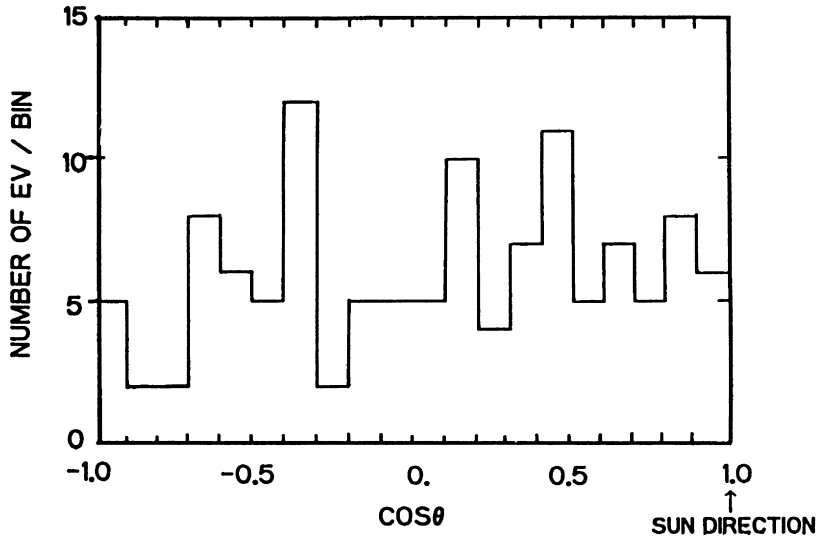


Fig.-2(a). cosθ distribution to the Sun of 120 upward-going muons. The cosθ bin size in the histogram is set to 0.1. cosθ=1 indicates the Sun direction.

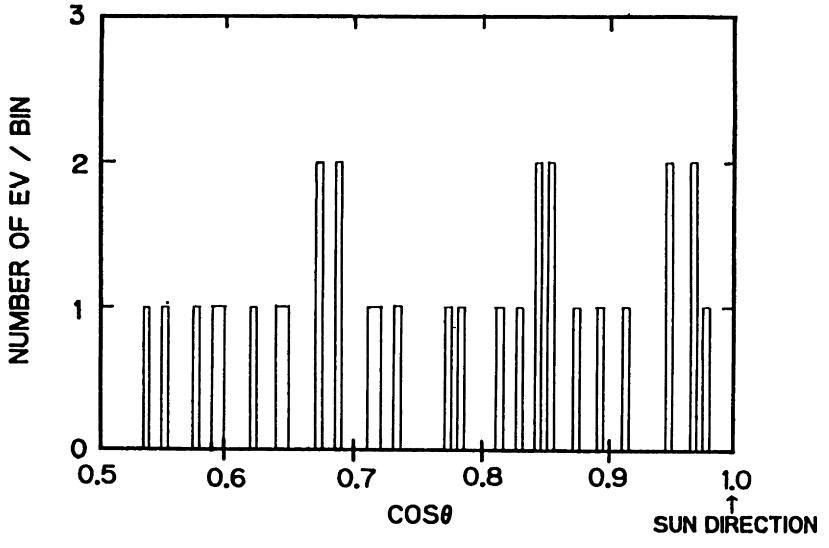


Fig.-2(b). An expanded view of Fig.-2(a) above, especially near the Sun direction, with a finer cosθ bin size of 0.005.

IMPLICATIONS FOR DARK MATTER PARTICLES OF SEARCHES FOR ENERGETIC SOLAR NEUTRINOS

John ELLIS, R.A. FLORES
CERN, CH-1211 Geneva 23, Switzerland

and

S. RITZ
Physics Department, University of Wisconsin, Madison, WI 53706, USA

Received 28 August 1987

We investigate whether recent searches for energetic solar neutrinos give new constraints on candidate dark-matter particles such as photinos $\tilde{\gamma}$ and higgsinos $\tilde{h}$, in view of the astrophysical and particle physics uncertainties. Previous quark model estimates of the elastic cross sections responsible for $\tilde{\gamma}$ or $\tilde{h}$ capture in the sun can be improved using hyperon decay data, but are thrown into doubt by recent European Muon Collaboration data on the spin-dependent proton structure function, which reduce the estimate of $\sigma(\tilde{\gamma}p)$ by a factor up to four. We emphasize the need to include charm and bottom quark fragmentation effects when calculating the flux of neutrinos from dark matter annihilation events, which may reduce the observable flux by a factor ~ 2. Even the most stringent upper limits from the Fréjus experiment still allow a local $\tilde{\gamma}$ or $\tilde{h}$ density considerably above the range of 0.20–0.43 GeV/cm^3 favoured by astrophysics, and no range of $m_{\tilde{\gamma}}$ or $m_{\tilde{h}}$ can yet be excluded.

There is an emerging consensus that non-dissipative dark matter is prevalent in galactic halos [1]. One of the prime possibilities is that the halo dark matter consists of some species of massive, non-relativistic, neutral, weakly-interacting particle. A possible candidate for this dark-matter particle is the lightest supersymmetric particle (LSP), which could be a photino $\tilde{\gamma}$, higgsino $\tilde{h}$ or sneutrino $\tilde{\nu}$ [2]. Methods of detecting dark-matter particles are now being actively pursued [3]. The most convincing would be direct detection of the elastic scattering of dark-matter particles off nuclei in the laboratory [4], unsuccessful searches for which have already been used to exclude sneutrinos in the range $16\ \text{GeV} < m_{\tilde{\nu}} < 1\ \text{TeV}$ [5]. An elegant but less direct method [6] is based on the observation that certain candidate dark-matter particles have some probability of being trapped in the sun after losing energy through elastic scattering [7,8]. Evaporation is negligible for masses above 2.5 GeV [9,10], in which case the solar population is stabilized by annihilations. Such annihilations, e.g., $\tilde{\gamma}\tilde{\gamma} \to f\bar{f}$, can produce energetic neutrinos via the prompt decays of heavy flavours: $f = \tau, c, b, t$. This idea has been pursued extensively [11–15] and varying estimates given of the resulting ν flux.

Data are now becoming available from underground detectors [16–18], most recently the Fréjus experiment [18], both on contained ν interactions and on upward-going muons induced by ν_μ interactions in rock. These provide upper limits on the energetic solar ν flux which can be confronted with different dark-matter predictions. Some candidates seem to be clearly excluded by such analyses, e.g., $\tilde{\nu}_\mu$ and $\tilde{\nu}_e$ with masses $\gtrsim 4$ GeV [13,14]. However, the confrontation is much more delicate for some candidates, such as the $\tilde{\gamma}$ and $\tilde{h}$, for which the experimental upper limits can be comparable with the ν event estimate. For $\tilde{\gamma}\tilde{\gamma}$ annihilation, this is approximately

0370-2693/87/$ 03.50 © Elsevier Science Publishers B.V.
(North-Holland Physics Publishing Division)

$$R_{\nu} \simeq 2.7\times 10^{-2} \times\left(\frac{\rho_{\mathrm{L}}}{0.3\ \mathrm{GeV\ cm^{-3}}}\right)\left(\frac{300\ \mathrm{km\ s^{-1}}}{\bar{v}}\right) \times \mathrm{O}(m_{\tilde{\gamma}}-m_{\mathrm{E}}) f(m_{\tilde{\gamma}}/m_{\mathrm{c}})\left(\frac{\sigma(\tilde{\gamma}\mathrm{p}\rightarrow\tilde{\gamma}\mathrm{p})}{10^{-40}\mathrm{cm}^2}\right) \times \sum_{\mathrm{i}} a_{\mathrm{i}} \sum_{\mathrm{f}} B_{\mathrm{f}} \langle Nz\rangle_{\mathrm{fi}} \quad \text{events/kt yr} \qquad (1)$$

for contained ν events, where ρ_{L} is the local halo density, $\bar{v}$ is the three-dimensional velocity dispersion of the dark-matter particles, m_{c} is a critical mass [8], m_{E} is an evaporation mass ~ 2.5 GeV [10], the a_{i} are simple numerical factors in the cross section for ν_{i} interactions which depend on the composition of the detector, B_{f} is the branching ratio for $\tilde{\gamma}\tilde{\gamma}\rightarrow \mathrm{f}\bar{\mathrm{f}}$ annihilation, N is the number of neutrinos produced in each annihilation and z the scaled energy fraction $E_{\nu_{\mathrm{i}}}/m_{\tilde{\gamma}}$. In this paper we refine previous numerical estimates [6,11–15] based on (1), survey the inherent uncertainties, and study whether the present data [16–18] can be used to give significant constraints on $m_{\tilde{\gamma}}$ or $m_{\tilde{\mathrm{h}}}$.

Among the uncertainties we discuss are the following. (a) The principal uncertainty in the numerical prefactor in (1) is due to the shape of the dark-matter velocity distribution, but is not very large. (b) Model studies [19] indicate that the local dark-matter density ρ_{L} is probably between 0.2 and 0.43 GeV/cm^3, with a preferred value $\rho_{\mathrm{L}}=0.3$ GeV/cm^3. (c) The best estimate of $\bar{v}$ is 300 km s^{-1} with an error of about 10%. (d) The factor written as $f(m_{\mathrm{c}}/m_{\tilde{\gamma}})$ is often approximated [7] by $\min(1, m_{\mathrm{c}}/m_{\tilde{\gamma}})$, but is better approximated by a calculation due to Gould [8]. (e) Estimates of $\sigma(\tilde{\gamma}\mathrm{p}\rightarrow\tilde{\gamma}\mathrm{p})$ depend principally on estimates of the squark masses and on the proton matrix elements of $\bar{\mathrm{q}}\gamma_{\mu}\gamma_5\mathrm{q}$ operators. The overall scale of squark and slepton masses $m_{\tilde{\mathrm{q}}}$ and $m_{\tilde{\mathrm{l}}}$ can be fixed by requiring the relic photino mass density $\rho_{\tilde{\gamma}}$ to equal the critical density ρ_{c} required for closure: $\Omega_{\tilde{\gamma}} \equiv \rho_{\tilde{\gamma}}/\rho_{\mathrm{c}} = 1$ [20]. However, the ratios $m_{\tilde{\mathrm{q}}}/m_{\tilde{\mathrm{l}}}$ are in general model-dependent and $\neq 1$. Previous estimates [4,6,11–14,21] of $\langle \mathrm{p}|\bar{\mathrm{q}}\gamma_{\mu}\gamma_5\mathrm{q}|\mathrm{p}\rangle$ have mainly been based on the naive quark model, but these can be refined using experimental data. The matrix elements of flavour-octet currents can be estimated on the basis of hyperon β-decay measurements [22] [#1], which differ somewhat from the naive quark model estimates. However, until recently the matrix element of the singlet current had not been measured, and it was generally fixed by assuming that $\langle \mathrm{p}|\bar{\mathrm{s}}\gamma_{\mu}\gamma_5\mathrm{s}|\mathrm{p}\rangle \simeq 0$ [24,25]. This assumption has recently been called into question by the European Muon Collaboration (EMC) data [26] on the spin-dependent proton structure function $g_1^{\mathrm{p}}(x,Q^2)$, whose integral is in theory proportional to the same combination of $\langle \mathrm{p}|\bar{\mathrm{q}}\gamma_{\mu}\gamma_5\mathrm{q}|\mathrm{p}\rangle$ as the $\tilde{\gamma}\mathrm{p}$ scattering matrix element. The EMC value of the integral at $Q^2 \gtrsim 3$ GeV^2 is about $\frac{1}{2}$ the naive quark model prediction, and indicates that $\langle \mathrm{p}|\bar{\mathrm{s}}\gamma_{\mu}\gamma_5\mathrm{s}|\mathrm{p}\rangle \neq 1$. The $\tilde{\gamma}\mathrm{p}$ scattering process explores these matrix elements at scales $m_{\tilde{\mathrm{q}}}^2 \sim 10^4$ GeV^2, but we do not expect significant variation between $Q^2 \sim 3$ GeV^2 and these scales, though there might be significant Q^2-dependence at lower Q^2 [27]. The EMC data suggest the possibility that previous estimates [4,6,11–14,21] of $\sigma(\tilde{\gamma}\mathrm{p}\rightarrow\tilde{\gamma}\mathrm{p})$ may need to be reduced by a factor ~ 4. Finally, we emphasize [15] that quark fragmentation effects should be taken into account in estimating $\langle Nz\rangle_{\mathrm{fi}}$ in (1), and that these reduce naive estimates [6,11–14] of c and b quark contributions by factors of 2–3. For completeness, we also present corresponding results for higgsinos $\tilde{\mathrm{h}}$.

We use the recent Fréjus data [18] to give upper limits on the local photino density $\rho_{\tilde{\gamma}}$ under different theoretical assumptions on $m_{\tilde{\mathrm{l}}}/m_{\tilde{\mathrm{q}}}$, on Ω and on the $\langle \mathrm{p}|\bar{\mathrm{q}}\gamma_{\mu}\gamma_5\mathrm{q}|\mathrm{p}\rangle$ matrix elements. These are generally above the astrophysical estimate [19] of ρ_{L} except under extreme assumptions, so that there is no safe bound on $m_{\tilde{\gamma}}$ or $m_{\tilde{\mathrm{h}}}$ yet. However, future data on upward-going muons may soon sharpen considerably the confrontation between theory and experiment.

The rate for trapping of cold dark-matter particles $\tilde{\mathrm{X}}$ by the sun is [7,8]

$$\Gamma_{\mathrm{T}} = a(\tfrac{3}{2}\pi)^{1/2} N_{\odot}(v_{\odot}^2/\bar{v}) n_{\tilde{\mathrm{X}}} f_{\mathrm{G}} \times \sum_{\mathrm{i}} x_{\mathrm{i}} \sigma(\tilde{\mathrm{X}}\mathrm{i}\rightarrow\tilde{\mathrm{X}}\mathrm{i}) f(m_{\tilde{\mathrm{X}}}/m_{\mathrm{c_i}}) \qquad (2)$$

where $N_{\odot}$ is the number of nuclei in the sun, $v_{\odot} = 617$ km s^{-1} is the escape velocity from the surface of the sun, $\bar{v}$ is the local velocity dispersion in the halo, $n_{\tilde{\mathrm{X}}}$ is the number density of $\tilde{\mathrm{X}}$ particles in the solar neighbourhood, and x_{i} is the fractional solar abun-

[#1] See ref. [23] for a review.

dance by number of nuclear species i. The factor $a=0.89$ for a nuclear species whose number density traces the mass density of the sun [7]. Since $\tilde{\gamma}$ and $\tilde{h}$ do not have coherent cross sections for scattering off heavy nuclei, the sum over species i in (2) can be truncated to protons p, for which this tracing assumption is good and $x_p=0.86$. The factor $f_G\simeq 1.5$ corrects [8] the estimate of ref. [7] for the motion of the sun through the galactic halo and for a numerical slip. The remaining kinematic factors $f(m_{\tilde{X}}/m_{c_i})$ express the mismatching of the $\tilde{X}$ and nuclear masses. These are often approximated [7] by

$$f(m_{\tilde{X}}/m_{c_i})=\min(1,m_{\tilde{X}}/m_{c_i})\,, \tag{3a}$$

where

$$m_{c_i}=\frac{1}{3\beta}\left(\frac{m_i^2+m_{\tilde{X}}^2}{m_i}\right)\frac{\bar{v}^2}{v_\odot^2}\,,$$

$$\beta\simeq 3.4\,, \tag{3b}$$

but this is not [8] very accurate for $m_{\tilde{X}}/m_{c_i}\sim 1$. In the case of protons, a better approximation [15] is

$$f(m_{\tilde{X}}/m_c)=\left(\frac{m_c^a}{m_c^a+m_{\tilde{X}}^a}\right)^{1/a}\,,$$

$$a=1.5\,,\quad m_c=20.4\ \text{GeV}\,, \tag{4}$$

which is accurate to within $\sim 5\%$.

The population of relic particles $\tilde{X}$ in the sun is controlled by evaporation [9] and annihilation. The evaporation rate is calculated [10] to be

$$C_E=0.8\times 10^{-5}x^2\exp(-32.9x)\times\left(\frac{\sigma(\tilde{X}p\to\tilde{X}p)}{10^{-40}\text{cm}^2}\right)\ \text{s}^{-1}\,, \tag{5a}$$

where

$$x=\left(\frac{m_{\tilde{X}}}{3.75\ \text{GeV}}\right)\left(\frac{1.4\times 10^7\text{K}}{T_{\tilde{X}}}\right), \tag{5b}$$

with $T_{\tilde{X}}$ the effective temperature of the isothermal distribution of $\tilde{X}$ particles within the sun. Because of the rapidly varying exponential factor in (5a), evaporation dominates over annihilation for $m_{\tilde{X}}$ below some critical evaporation mass m_E, but can be neglected by comparison to annihilation when $m_{\tilde{X}}>m_E$ [#2]. This explains the $\theta(m_{\tilde{X}}-m_E)$ factor in (1), where the evaporation mass $m_E\sim 2.5$ GeV for $\tilde{X}=\tilde{\gamma}$ and $\tilde{h}$ using our estimates of $\sigma(\tilde{X}p\to\tilde{X}p)$ and $T_{\tilde{X}}$. When $m_{\tilde{X}}>m_E$, the annihilation rate can be deduced from $\Gamma_T(2)$ by dividing by two for self-conjugate particles such as the $\tilde{\gamma}$ and $\tilde{h}$, and the resulting flux at the earth is obtained by dividing $4\pi R^2$, where R is the earth–sun distance. In this way, we obtain the solar contributions to the prefactors in the observable neutrino interaction rate (1).

The main astrophysical uncertainty in (1) is in ρ_L. If one assumes [19] that galaxies form in pre-existing potential wells of dark matter, then an adiabatic model [28] can be used to bound ρ_L. The main constraints are that the circular velocity in the galaxy not fall below 200 km s^{-1} at large radii ~ 60 kpc, and that the ratio F of luminous to total mass inside the region where the gas cools to form visible galaxies obey $F\gtrsim 0.05$. One then obtains [19]

$$0.20\lesssim\rho_L/\text{GeV cm}^{-3}\lesssim 0.43\,, \tag{6a}$$

if the shape of the potential well before galaxy formation is that expected for cold dark matter [#3]. For more likely values of the parameters

$$0.30\lesssim\rho_L/\text{GeV cm}^{-3}\lesssim 0.43\,, \tag{6b}$$

and the typical estimate $F\simeq 0.1$ implies $\rho_L\simeq 0.3$ GeV cm^{-3}. Needless to say, these are dynamical bounds and the lower bound therefore applies to particle dark matter only if it makes up all the mass of the halo. If some of the halo dark matter are e.g. little Jupiters or black holes [29], then the particle dark-matter density could be below the lower bounds (6). Also, meaningful confidence levels cannot be assigned to the bounds on ρ_L because the systematic uncertainties are large. For these reasons, we will prefer to use the experimental constraints to provide upper bounds on the particle dark matter contribution to ρ_L.

Next comes the uncertainty in $\bar{v}$. If one assumes that the velocity dispersion of the halo particles is homogeneous and isotropic, i.e., the halo is an isothermal sphere, then "hydrostatic" equilibrium implies $\bar{v}^2=\frac{3}{2}v_\infty^2$, where v_∞ is the asymptotic circular velocity. If one assumes a constant circular velocity outside the solar radius, then $v_\infty=v_c$, the total circular velocity for which the IAU recommended value

[#2] The width of the transition region is ~ 200 MeV, and is ignored in our subsequent discussion.

[#3] The lower bound is relaxed by only 15% if one relaxes this assumption.

[30] is $v_c = 220$ km s^{-1} and a straight mean of the observations yields [30] $v_c = 222 \pm 20$ km s^{-1}, implying $\bar{v} = 270 \pm 25$ km s^{-1}. If one does not assume a constant circular velocity, but assumes that $v_\infty = v$ (60 kpc), observations suggest that probably $200 \lesssim v_\infty/\text{km s}^{-1} \lesssim 260$ [#4], implying $245 \lesssim \bar{v}/\text{km s}^{-1} \lesssim 320$. The halo particles cannot have an exactly isothermal distribution, because they are compressed during galaxy formation [28] and distorted by the potential of the disc. These effects on $\bar{v}$ have yet to be quantified, but appear to be small [33]. Motivated by the fact that higher $\bar{v}$ implies lower R_v (1), we will take $\bar{v} = 300$ km s^{-1} for our analysis. It is justified to take the highest value of $\bar{v}$ that is allowed by the uncertainty in v_c in deriving a conservative bound on masses.

Next we turn to uncertainties in estimating $\sigma(\tilde{\gamma}p \to \tilde{\gamma}p)$, which takes the following form [6,21] if the supersymmetric partners of the left- and right-handed squarks have equal masses, $m_{\tilde{q}_L} = m_{\tilde{q}_R} = m_{\tilde{q}}$:

$$\sigma(\tilde{\gamma}p \to \tilde{\gamma}p) = \frac{3}{\pi} \frac{m_{\tilde{\gamma}}^2 m_p^2 (4\pi\alpha)^2}{(m_{\tilde{\gamma}} + m_p)^2} M^2 ,$$

$$M = \sum_q \frac{Q_q^2}{m_{\tilde{q}}^2} \Delta q \qquad (7a)$$

where $Q_q = \frac{2}{3}$ for u, c, $-\frac{1}{3}$ for d, s and Δq is defined by

$$\langle p | \tilde{q} \gamma_\mu \gamma_5 q | p \rangle = \Delta q \cdot 2 s_\mu , \qquad (7b)$$

where s_μ is the proton spin. Differences between $m_{\tilde{q}_L}$ and $m_{\tilde{q}_R}$ can arise from different renormalizations of these parameters by interactions at scales $\gtrsim m_W$, and from differences in Higgs vacuum expectation values, but we disregard such possibilities here. The overall scale of the sparticle masses as functions of $m_{\tilde{\gamma}}$ can be fixed by requiring closure density for photinos: $\Omega_{\tilde{\gamma}} = 1$ [#5]. If $m_{\tilde{\ell}} = m_{\tilde{q}}$ (EM), one needs masses [6,13,14]

$$m_{\tilde{q}} \simeq 56 \text{ GeV} \cdot (1 + 0.04 m_{\tilde{\gamma}}^2)^{1/4} \qquad (8)$$

[#4] The Magellanic stream [31] indicates a high v (60 kpc) $\simeq 240$ km s^{-1}, whilst globular cluster data [32] indicate $\simeq 200$ km s^{-1}.

[#5] We assume a present-day Hubble expansion rate $H_0 = 50$ km/s/Mpc. If H_0 is larger (smaller) one needs larger (smaller) values of $m_{\tilde{\gamma}}$.

to get $\Omega_{\tilde{\gamma}} = 1$. However, in most models $m_{\tilde{\ell}}$ and $m_{\tilde{q}}$ are renormalized differently by interactions at scales $\gtrsim m_W$, and a universal value m_0 at some high GUT scale yields different masses. For example, in a minimal supergravity model (MSM), one finds [20]

$$m_{\tilde{q}}^2 \simeq m_0^2 (1 + 6.4\xi^2) ,$$
$$m_{\tilde{e}_R}^2 \simeq m_0^2 (1 + 0.15\xi^2) ,$$
$$m_{\tilde{e}_L}^2 \simeq m_0^2 (1 + 0.52\xi^2) , \qquad (9)$$

where $\xi \equiv m_{1/2}/m_0$ and

$$m_{\tilde{\gamma}} \simeq 0.47 \xi m_0 , \quad m_{\tilde{g}} \simeq 2.8 \xi m_0 . \qquad (10)$$

Assuming these mass relations, we use eqs. (8) and (11) of ref. [12] to fix the values of $m_{\tilde{q}}$ needed to obtain any specific value of $\Omega_{\tilde{\gamma}}$. One needs masses given approximately [20] by

$$\text{MSM}_1 : \quad m_{\tilde{q}} \simeq 44 \text{ GeV} + 5.5 m_{\tilde{\gamma}} \qquad (11)$$

to get $\Omega_{\tilde{\gamma}} = 1$, and [20]

$$\text{MSM}_{1/4} : \quad m_{\tilde{q}} \simeq 24 \text{ GeV} + 5.3 m_{\tilde{\gamma}} \qquad (12)$$

to get $\Omega_{\tilde{\gamma}} = \frac{1}{4}$. We will use EM(8), MSM_1(11) and $\text{MSM}_{1/4}$(12) as examples for the subsequent extraction of model-dependent upper bounds on $\rho_{\tilde{\gamma}}$ and bounds on $m_{\tilde{\gamma}}$.

We now consider uncertainties in the matrix element factors Δq (7). Neutron decay tells us that

$$g_A = \Delta u - \Delta d = 1.25 . \qquad (13)$$

The naive quark model is often used [4,6,11–14,21] to set $\Delta s \simeq 0$ and fix Δu and Δd separately. An improved estimate can be made by using flavour SU(3) to determine Δu and Δd by relating them to the F and D parameters measured in hyperon β-decays [22,23]:

$$\Delta u - \Delta d = F + D ,$$
$$\Delta u + \Delta d - 2\Delta s = 3F - D , \qquad (14)$$

where $F/D = 0.63$ [#6], and it is again conventionally assumed [24] that $\Delta s \simeq 0$ yielding

$$\text{H}\beta : \quad \Delta u = 0.96(6) , \quad \Delta d = -0.28(4) \quad [\Delta s = 0] . \qquad (15)$$

[#6] A previous estimate [21] used an out-of-date value for F/D.

Recent EMC data [26] on polarized μp scattering cast doubt on the assumption that $\Delta s \simeq 0$. They measure the spin-dependent structure function $g_1^p(x,Q^2)$, for which there is a QCD sum rule

$$\int_0^1 dx\, g_1^p(x,Q^2) = \frac{1}{2}\left(\sum_q Q_q^2 \Delta q\right)[1+O(\alpha_s/\pi)]$$

$$\simeq \tfrac{1}{2}(\tfrac{4}{9}\Delta u + \tfrac{1}{9}\Delta d + \tfrac{1}{9}\Delta s)\,. \qquad (16)$$

The $O(\alpha_s/\pi)$ correction in (16) is $(-1\cdot\alpha_s/\pi)$ for non-singlet combinations [34] of the Δq, and $(-\tfrac{1}{3}\cdot\alpha_s/\pi)$ for the singlet combination [35]: these do not give a large Q^2-dependence for $Q^2 > 3$ GeV^2. Note that EMC measures the same combination of the Δq as appears in $\sigma(\tilde{\gamma}p\to\tilde{\gamma}p)$ (7a) if all the $m_{\tilde{q}}$ are approximately equal as we expect. The prediction using Hβ (15) is

$$\int_0^1 dx\, g_1^p(x,Q^2) = 0.20$$

$$\simeq \tfrac{1}{2}(\tfrac{4}{9}\Delta u + \tfrac{1}{9}\Delta d + \tfrac{1}{9}\Delta s)\,, \qquad (17)$$

whereas the EMC measure [26]

$$\int_0^1 dx\, g_1^p(x,Q^2)\Big|_{Q^2 \geq 3\mathrm{GeV}^2}$$

$$= 0.113 \pm 0.012 \pm 0.025\,. \qquad (18)$$

If one neglects any possible dependence on the renormalization scale $\mu \simeq Q$, this can be combined with (14) to yield

$$\text{EMC:}\quad \Delta u = 0.70(8)\,,\quad \Delta d = -0.54(2)\,,$$

$$\Delta s = -0.25(7)\,, \qquad (19)$$

each with errors $\pm 0.08(4)$ if the statistical and systematic errors in (18) are added in quadrature [#7].

Different estimates of the factor $(\sum Q_q^2 \Delta q)^2$ appearing in eq. (7a) are shown in table 1. The reduction in the observed value (18) compared to that previously expected (17) suggests a reduction in the prediction of $\sigma(\tilde{\gamma}p\to\tilde{\gamma}p)$ by a factor ~ 4. As already mentioned, we do not expect significant renormalization scale (μ) dependence that decouples the EMC measurements at $\mu^2 \simeq Q^2 \geq 3$ GeV^2 from the values of the Δq at $\mu^2 \simeq m_{\tilde{q}}^2 \sim 10^4$ GeV^2 required in $\tilde{\gamma}p$ scattering. Moreover, the EMC Collaboration sees no evidence for any Q^2-dependence [#8]. Until the EMC result is confirmed, one should nevertheless perhaps regard the range (15), (19) as a legitimate uncertainty in dark-matter detection estimates.

We now consider the possibility that the dark-matter particle is a higgsino $\tilde{h}$, for which [21]

$$\sigma(\tilde{h}p\to\tilde{h}p) = \frac{6 m_{\tilde{h}} m_p^2}{\pi(m_{\tilde{h}}+m_p)^2} G_F^2 \cos^2(2\beta)\left(\sum_q I_{3q}\Delta q\right)^2\,, \qquad (20)$$

where $I_{3u} = +\tfrac{1}{2}$, $I_{3d,s} = -\tfrac{1}{2}$, Δq is given by (7b), (15) and (19) as before, and β is given by $\tan\beta \equiv v_2/v_1$ where v_2 (v_1) is the VEV of the Higgs giving masses to the u (d,s) quarks. Notice that the squark masses do not affect $\sigma(\tilde{h}\tilde{h}\to f\bar{f})$ or $\sigma(\tilde{h}p\to\tilde{h}p)$ (20), but it is possible [2] to find an acceptable choice of β which gives $\Omega_{\tilde{h}} = 1$ whenever $m_{\tilde{h}} > m_1 \approx 6$ GeV.

The factors a_i in (1) are simple numerical constants in νN scattering cross sections. For n and p targets they are [15]

$$a_{\nu p} = 4.2\,,\quad a_{\bar{\nu}p} = 4.1\,,$$

$$a_{\nu n} = 9.5\,,\quad a_{\bar{\nu}n} = 2.2\,, \qquad (21)$$

corresponding to

$$\text{isoscalar target:}\quad a_\nu = 6.85\,,\quad a_{\bar{\nu}} = 3.15\,,$$

$$\text{Fe} = 26p + 30n:\quad a_\nu = 7.04\,,\quad a_{\bar{\nu}} = 3.09\,,$$

$$H_2O = 10p + 8n:\quad a_\nu = 6.56\,,\quad a_{\bar{\nu}} = 3.26\,. \qquad (22)$$

The factors B_f in (1) are the branching ratios for $\tilde{X}\tilde{X}\to f\bar{f}$ which are determined by the couplings in the particular model studied. We use the Monte Carlo procedure detailed in ref. [15] to model the annihilation events in the sun and calculate the factors

[#7] The estimates (19) correspond to a fraction $-0.09(1) \pm 0.24(2)$ of the proton spin being carried by quarks. They also correspond to a singlet/octet matrix element ratio of $-0.13(3) \pm 0.37(1)$ times the quark model value previously estimated assuming $\Delta s = 0$.

[#8] Jaffe [27] has pointed out that there could in principle be significant μ-dependence between $\mu^2 = 3$ GeV^2 and $\mu^2 \ll 1$ GeV^2. This is irrelevant to $\sigma(\tilde{\gamma}p\to\tilde{\gamma}p)$, but could invalidate the derivation of the individual Δq in (19). This introduces an uncertainty in the subsequent estimate of $\sigma(\tilde{h}p\to\tilde{h}p)$, which is fortunately not large.

Table 1
Eestimates of $\bar{M}^2 \equiv (\sum_q Q_q^2 \Delta q)^2$.

Assumptions	p	n	References
$\tilde{u}$ only, naive quark model	0.24	0.005	[4,11]
$\tilde{u}+\tilde{d}$, $\Delta s=0$, old F/D	0.11	0.007	[21]
$\tilde{u}+\tilde{d}$, $\Delta s=0$, $F/D=0.63$	0.16	0.0004	this paper using ref. [24]
$\tilde{u}+\tilde{d}+\tilde{s}$, EMC	0.05	0.04	this paper using ref. [26]

$\langle Nz \rangle_{fi}$ in (1). Effects due to the solar medium are important. Relativistic light hadrons and muons are stopped before they decay, so that only the semi-leptonic decays of heavy quarks and τ decays contribute to the high-energy neutrino flux [11,15]. In general, one might also worry about neutrino interactions with the sun and dE/dx effects on heavy hadrons, but these are negligible for the range of particle masses considered here [15]. Each annihilation event gives two fundamental fermions in the final state with energies equal to $m_{\tilde{X}}$. If the annihilation is into a pair of quarks, fragmentation and other final-state QCD effects are important. Heavy-flavour hadrons will not have the "beam" energy $m_{\tilde{X}}$, and the subsequent neutrino spectra are softer than one would naively expect. This fragmentation effect has been modelled

Table 2
Model-dependent estimates of R_ν, R_μ. These numbers and the Fréjus curves in the figures are calculated for $E_\nu > 2$ GeV [18]. The IMB curves in the figures are for $0.2 < E_\nu < 2$ GeV [16].

	$m_{\tilde{X}}$ (GeV)	Model					
		A	B	C	D	E	F
		$\Omega=1$ MSM_1 $H\beta$	$\Omega=1$ MSM_1 EMC	$\Omega=1$ EM_1 $H\beta$	$\Omega=1/4$ $MSM_{1/4}$ $_H\beta$	$\Omega=1$ $H\beta$	$\Omega=1$ EMC
R_ν (events/kt yr)	3	0.46	0.15	0.51	1.91	–	–
	4	1.56	0.50	1.80	6.10	–	–
	6	1.83	0.59	2.28	6.24	1.16	1.69
	8	0.82	0.26	1.03	2.74	0.61	0.89
	10	0.58	0.19	0.81	1.76	0.59	0.86
	15	0.22	0.072	0.39	0.57	0.33	0.49
	20	0.11	0.034	0.23	0.23	0.22	0.32
	30	0.028	0.009	0.09	0.05	0.10	0.15
R_μ (events/100 m^2 yr)	3	0.12				–	
	4	0.44				–	
	6	0.68				0.35	
	8	0.37				0.21	
	10	0.31				0.23	
	15	0.17				0.17	
	20	0.11				0.14	
	30	0.04				0.09	

carefully [15] using a Monte Carlo [36] which reproduces correctly the spectra of charged leptons and other secondary particles in e^+e^- annihilation. It reduces $\langle Nz\rangle_{fi}$ by a factor ~ 3 for the case f=c, and ~ 2 for the case f=b. We will present and use results for $i=\nu_e,\nu_\mu,\bar{\nu}_e,\bar{\nu}$ only.

Our results are shown in table 2 and the figures. Table 2 makes predictions for contained high-energy

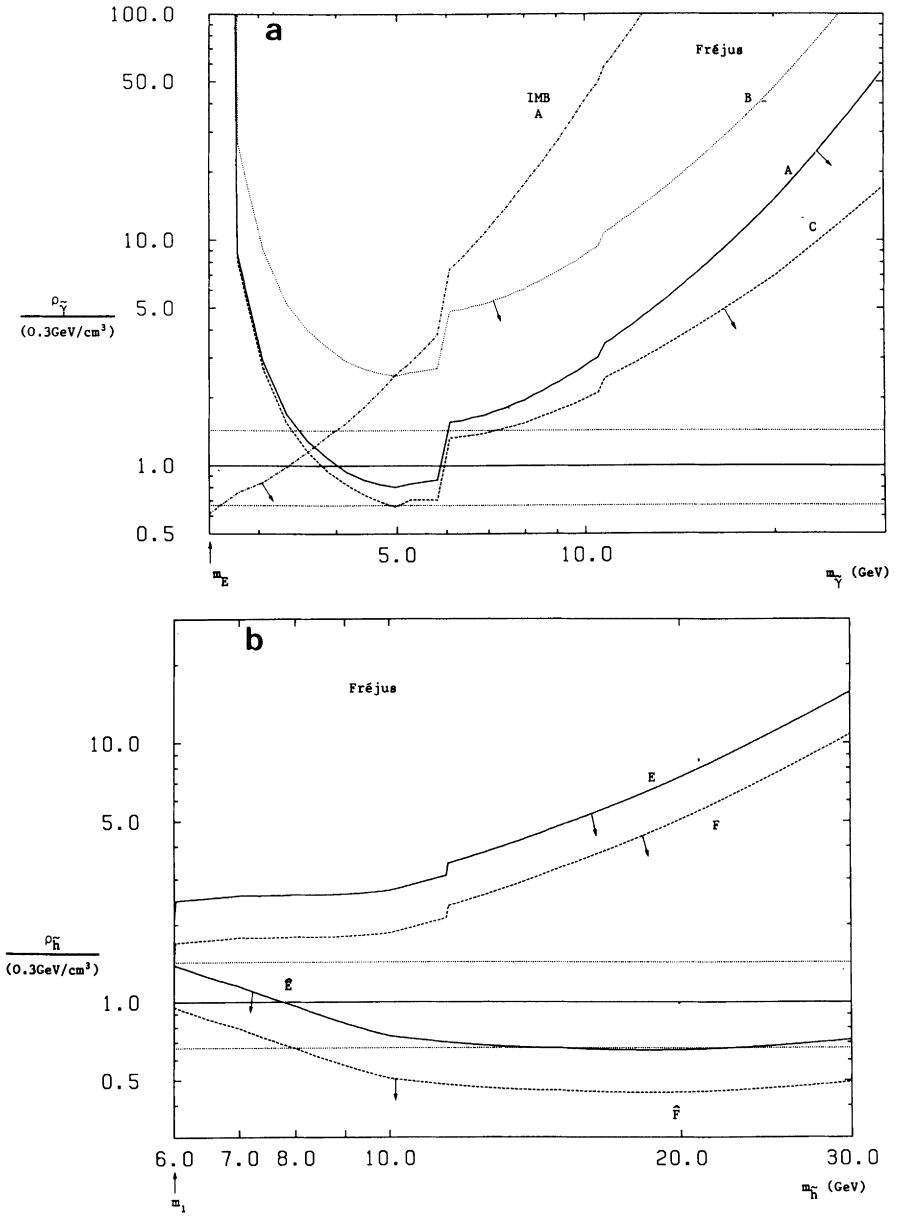


Fig. 1. Upper limits on $\rho_{\tilde{X}}$: (a) $\tilde{X}=\tilde{\gamma}$, using Fréjus data [18] under model assumptions (A), (B) and (C), and IMB data [16] under assumption (A); (b) $\tilde{X}=\tilde{h}$ using Fréjus data [18] under model assumptions (E), (F). The $\hat{E}$ and $\hat{F}$ curves are obtained from eq. (20) by removing the model-dependent factor of $\cos^2(2\beta)$. The horizontal lines indicate the astrophysically preferred value and probable limits [19] on $\rho_{\tilde{X}}$.

solar neutrino event rates R_ν for the following different models:

(A) $\tilde{X}=\tilde{\gamma}$: $\Omega_{\tilde{\gamma}}=1$, MSM_1 masses and Hβ estimates of Δq,

(B) $\tilde{X}=\tilde{\gamma}$: $\Omega_{\tilde{\gamma}}=1$, MSM_1 masses and EMC estimates of Δq,

(C) $\tilde{X}=\tilde{\gamma}$: $\Omega_{\tilde{\gamma}}=1$, EM_1 masses $m_{\tilde{q}}=m_{\tilde{q}}$ and Hβ estimates of Δq,

(D) $\tilde{X}=\tilde{\gamma}$: $\Omega_{\tilde{\gamma}}=\frac{1}{4}$, $MSM_{1/4}$ masses and Hβ estimates of Δq,

(E) $\tilde{X}=\tilde{h}$: $\Omega_{\tilde{h}}=1$, Hβ estimates of Δq,

(F) $\tilde{X}=\tilde{h}$: $\Omega_{\tilde{h}}=1$, EMC estimates of Δq.

We see that only in case (D) does the rate predicted for $\rho_{\tilde{X}}=0.2$ GeV cm^{-3} and $\bar{v}=300$ km s^{-1} come above the upper limit of $R_\nu<1.6$ events kt^{-1} yr^{-1} provided by the Fréjus experiment [18]. An upper limit on contained events is also available from the IMB experiment [16] [#9]. The Kamioka experiment is also preparing an upper limit on R_ν [17], but this is not yet available in a form which can be compared with the astrophysically expected [19] halo density (6a). Fig. 1a shows upper limits from the Fréjus experiment [18] on $\rho_{\tilde{\gamma}}$ in cases (A), (B) and (C) above. Also shown is the upper limit from the IMB experiment [16] on $\rho_{\tilde{\gamma}}$ in case (A). The horizontal lines indicate the expected range of the halo density [19], which is not yet excluded by the underground experiments for any value of $m_{\tilde{\gamma}}$ [#10]. Fig. 1b shows upper limits from the Fréjus experiment [18] on $\rho_{\tilde{h}}$ in cases (E) and (F) above, where again we see that no region of $m_{\tilde{h}}$ can be excluded in a model-independent way. Since the expected sparticle masses decrease when $\Omega_{\tilde{\gamma}}$ is decreased [see (11) and (12)], resulting in an increase in $\sigma(\tilde{\gamma}p\to\tilde{\gamma}p)$ (7a) and hence an increase in R_ν (1), we might expect to get a non-trivial bound on $m_{\tilde{\gamma}}$ if $\Omega_{\tilde{\gamma}}$ is sufficiently small. This effect was already discernible in case (D) of table 2, where $\Omega_{\tilde{\gamma}}=\frac{1}{4}$ and the Hβ estimates of Δq are used. It is also exhibited in fig. 2, as a domain of

[#9] The IMB experiment [16] also quotes an upper limit on the upward-going μ flux, but this is not yet competitive with upper limits on the ν flux from contained events at Fréjus.

[#10] The vertical glitches in these curves correspond to $\tilde{\gamma}\tilde{\gamma}$ annihilations in the early Universe at temperatures $T_A\sim\frac{1}{20}m_{\tilde{\gamma}}$ higher/lower than the hadron phase transition temperature, taken here to be $T_c=300$ MeV. The values of $m_{\tilde{q}}$ required to get $\Omega_{\tilde{\gamma}}=1$ are different for $T_A\gtrless T_c$, causing a change in $\sigma(\tilde{\gamma}p\to\tilde{\gamma}p)$. Because T_c is uncertain, so also are the exact values of $m_{\tilde{\gamma}}$ where the glitches occur.

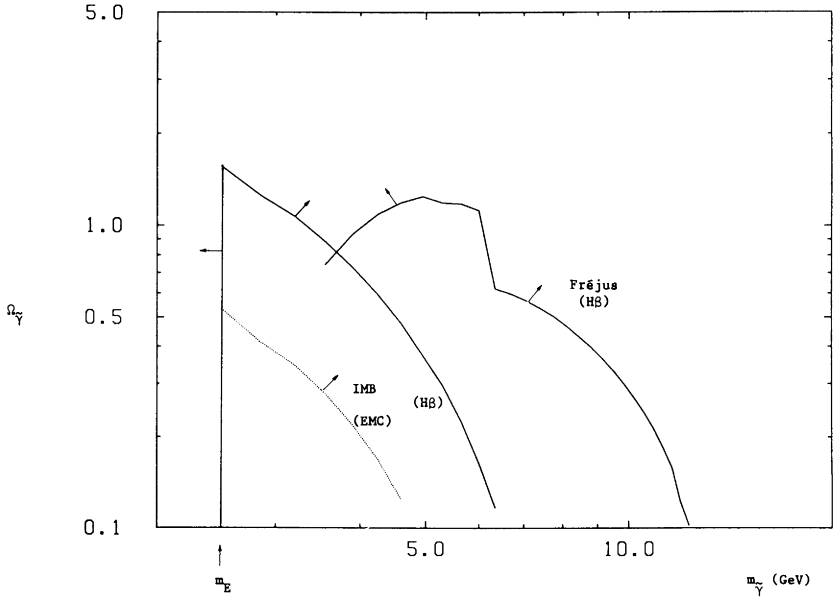


Fig. 2. Domain of the $(m_{\tilde{\gamma}},\Omega_{\tilde{\gamma}})$ plane excluded by the IMB [16] and Fréjus data [18] assuming MSM masses (9) and (10) and the Hβ and EMC estimates of Δq (15).

the $(m_{\tilde{\gamma}}, \Omega_{\tilde{\gamma}})$ plane which can be excluded assuming MSM masses (9), (10) and Hβ estimates of Δq (15). Note, however, that even this exclusion zone is greatly reduced if the EMC estimates of Δq are used instead.

So far, we have concentrated on the contained event signal for high-energy solar neutrinos. Alternatively, one can look for upward-going muons from ν_μ collisions in rock [6,14]. With the same assumptions and uncertainties as in the derivation of eq. (1), we estimate

$$R_\mu \simeq 1.1 \times 10^{-2} \times \left(\frac{\text{area}}{100\ \text{m}^2}\right)\left(\frac{\rho_{\text{L}}}{0.3\ \text{GeV cm}^{-3}}\right)\left(\frac{300\ \text{km s}^{-1}}{\bar{v}}\right) \times O(m_{\tilde{\gamma}} - m_{\text{E}}) f(m_{\tilde{\gamma}}/m_{\text{c}}) \left(\frac{\sigma(\tilde{\gamma}\text{p} \to \tilde{\gamma}\text{p})}{10^{-40}\text{cm}^2}\right) m_{\tilde{\gamma}} \times \sum_i a_i b_i \sum_f B_f \langle N z^2 \rangle_{fi} ,$$

$$b_{\nu_\mu} = 0.5 , \quad b_{\bar{\nu}_\mu} = 0.67 , \qquad (23)$$

for the rate of such through-going muon events. The existing Fréjus [18], IBM [16] and Kamioka [17] experiments are unlikely to their current limits on R_ν (1) by large factors, and the IMB limit [16] on R_μ (23) is not yet competitive. However, future experiments with large underground μ detectors such as MACRO [37] could make significant improvements in the upper limit on R_μ. Therefore, we have included in table 2 predictions for R_μ in the models (A) and (E) introduced above. This is likely to be in the future the best way to establish lower bounds on $m_{\tilde{\gamma}}$ and $m_{\tilde{h}}$ or to find indirect evidence for supersymmetric dark-matter particles.

We thank K.A. Olive, R. Petronzio and D. Seckel for discussions of particle physics and astrophysics related to this paper, and V. Papavassiliou and T. Sloan for discussions about the EMC data. The work of S.R. is supported by the US Department of Energy Contract Number DE-AC02-76ER 00881 and the US National Science Foundation grant number INT-8313994 for travel. We acknowledge the use of the Wisconsin VAX8800 computer for the numerical calculations.

References

[1] See, e.g., ed. S.M. Faber, Nearly normal galaxies (Springer, Berlin, 1987).
[2] H. Goldberg, Phys. Rev. Lett. 50 (1983) 1419; J. Ellis, J.S. Hagelin, D.V. Nanopoulos, K.A. Olive and M. Srednicki, Nucl. Phys. B 238 (1984) 453.
[3] See, e.g., P. Smith, in: Cosmology, astronomy and fundamental physics, eds. G. Setti and L. Van Hove (ESO–CERN, Geneva, 1986) p. 237.
[4] M. Goodman and E. Witten, Phys. Rev. D 31 (1985) 3059.
[5] G. Gelmini, reported in: Neutrino masses and neutrinos astrophysics, eds. V. Barger et al. (World Scientific, Singapore, 1987) p. 337.
[6] M. Srednicki, K.A. Olive and J. Sil, Nucl. Phys. B 279 (1987) 804.
[7] W.H. Press and D.N. Spergel, Astrophys. J. 296 (1985) 679.
[8] A. Gould, SLAC report, SLAC-PUB-4226 (1987).
[9] L.M. Krauss, M. Srednicki and F. Wilczek, Phys. Rev. D 33 (1986) 2079.
[10] K. Griest and D. Seckel, Nucl. Phys. B 283 (1987) 681; A. Gould, SLAC report, SLAC-PUB-4162 (1987).
[11] T.K. Gaisser, G. Steigman and S.Z. Tilav, Phys. Rev. D 34 (1986) 2206.
[12] B.A. Campbell, J. Ellis, K. Enqvist, D.V. Nanopoulos, J.S. Hagelin and K.A. Olive, Phys. Lett. B 173 (1986) 270.
[13] J.S. Hagelin, K.-W. Ng and K.A. Olive, Phys. Lett. B 180 (1987) 375.
[14] K.-W. Ng, K.A. Olive and M. Srednicki, Phys. Lett. B 188 (1987) 138.
[15] S. Ritz and D. Seckel, CERN preprint TH.4627 (1987).
[16] IMB Collab., J.M. Losecco et al., Phys. Lett. B 188 (1987) 388.
[17] Y. Totsuka, University of Tokyo preprint UT-ICEPP-87-02 (1987).
[18] Fréjus Collab., presented by B. Kuznik, Orsay preprint LAL-87-21 (1987).
[19] R.A. Flores, CERN preprint CERN-TH.4736 (1987).
[20] J. Ellis, J.S. Hagelin and D.V. Nanopoulos, Phys. Lett. B 158 (1985) 26.
[21] G.L. Kane and I. Kani, Nucl. Phys. B 277 (1986) 525.
[22] M. Bourquin et al., Z. Phys. C 21 (1983) 27.
[23] J.-M. Gaillard and G. Sauvage, Ann. Rev. Nucl. Part. Sci. 34 (1984) 351.
[24] J. Ellis and R.L. Jaffe, Phys. Rev. D 9 (1974) 1444.
[25] V.M. Belyaev, B.L. Ioffe and Y.I. Kogan, Phys. Lett. B 151 (1985) 290.
[26] EMC Collab., J. Ashman et al., CERN-EP preprint, in preparation, presented by V. Papavassiliou and T. Sloan at 1987 EPS High energy physics Conf. (Uppsala).
[27] R.L. Jaffe, Phys. Lett. B 193 (1987) 101.
[28] G. Blumenthal, S.M. Faber, R.A. Flores and J. Primack, Astrophys. J. 301 (1986) 27.
[29] D. Hegyi and K.A. Olive, Astrophys. J. 303 (1986) 56.
[30] F.J. Kerr and D. Lynden-Bell, Mon. Not. R. Astron. Soc. 221 (1986) 1023.

[31] T. Murai and M. Fujimoto, Pub. Astr. Soc. Japan 32 (1980) 581;
D.N.C. Lin and D. Lynden-Bell, Mon. Not. R. Astron. Soc. 198 (1982) 707.
[32] F.D.A. Hartwick and W.L.W. Sargent, Astrophys. J. 221 (1978) 512.
[33] J. Primack, private communication.
[34] J. Kodaira, S. Matsuda, T. Muta, T. Uematsu and K. Sasaki, Phys. Rev. D 20 (1979) 67;
J. Kodaira, S. Matsuda, K. Sasaki and T. Uematsu, Nucl. Phys. B 159 (1979) 99.
[35] J. Kodaira, Nucl. Phys. B 165 (1979) 129.
[36] T. Sjöstrand, University of Lund preprint LUTP 85-10 (1985).
[37] M.J. Longo, in: Neutrino masses and neutrino astrophysics, eds. V. Barger et al. (World Scientific, Singapore, 1987) p. 463.

SOLAR NEUTRINO SEARCHES AND COLD DARK MATTER

Keith A. OLIVE
School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455, USA

and

Mark SREDNICKI
Department of Physics, University of California, Santa Barbara, CA 93106, USA

Received 30 November 1987

We consider new data on high energy neutrinos ($E_\nu > 2$ GeV) obtained from the Frejus and Kamioka detectors as well as previous data from the IMB detector and compare the experimental limits on neutrinos from the sun to our expectations based on a cold dark matter hypothesis. We also extend previous work to include a wide variety of cold dark matter candidates.

Postulating the existence of dark matter in the Universe solves a host of astrophysical and cosmological problems. (For a review, see ref. [1].) However, until an experimental determination of the existence and identity of a dark matter candidate, dark matter scenarios remain hypothetical. One approach to searching for dark matter is to look for indirect signatures of its presence. A promising idea is to look for an enhanced high energy neutrino flux from the sun due to annihilations within of dark matter particles [2]. In this letter, we utilize new data from underground "proton decay" detectors to place constraints on properties of dark matter candidates. We also extend previous work to include a wider variety of candidates.

Many scenarios for solving dark matter problems involve a cold dark matter particle X which is massive, $m_X \gtrsim 1$ GeV. If present in the galactic halo, these particles will become trapped [3–5] in the sun due to their elastic scatterings with protons and other nuclei. Inside the sun, they would annihilate and produce a flux of high energy neutrinos [2,6–12]. If the neutrino flux is sufficiently large compared with the atmospheric background flux, a signal could be detected in present underground experiments. Indeed, several experimental groups are searching for such a signal [13–15]. To date, there is no evidence for an enhanced flux in the direction of the sun. The best we can do is place limits on the neutrino flux and hence on the properties of X.

There are many possibilities for dark matter particles: photinos [16,17], higgsinos [17], Majorana neutrinos [18,19], Dirac neutrinos [20,19], and sneutrinos [21]. The capture rate [4,5] of X's in the sun and the flux [6–8] of neutrinos resulting from the annihilations of X's have been calculated. The integrated flux of solar neutrinos, $\bar{S}$, is given by

$$\bar{S} = \tau n_N \int dE\, V_F(E) \times \left(\sigma_\nu(E) \frac{d\Phi_{S,\nu}}{dE} + \sigma_{\bar{\nu}}(E) \frac{d\Phi_{S,\bar{\nu}}}{dE} \right), \quad (1)$$

where τ is the observation time, n_N is the number density of nucleons in the detector, $V_F(E)$ is the fiducial volume of the detector, $\sigma_{\nu(\bar{\nu})}(E)$ is the interaction cross section per nucleon for a neutrino (antineutrino) and $d\Phi_{S,\nu(\bar{\nu})}/dE$ is the predicted differential flux of solar neutrinos (antineutrinos) integrated over solid angle. A similar formula gives $\bar{A}$, the integrated flux of atmospheric neutrinos; one simply replaces $d\Phi_S/dE$ with $d\Phi_A/dE$. We would like to compare the solar flux to the atmospheric flux. Towards this end we define the ratio

0370-2693/88/$ 03.50 © Elsevier Science Publishers B.V.
(North-Holland Physics Publishing Division)

$$r \equiv \bar{S}/\bar{A}. \tag{2}$$

There are at least two advantages to setting limits on r rather than the number of solar neutrino events per kiloton-year. First, any systematic errors in the overall normalization of the experimental rates will cancel when computing r. Second, the experimental limit on r may be computed just by knowing the numbers of events of a given type; it is not necessary to know the observation time, fiducial volume, etc. It is, however, necessary to have a theoretical determination of the atmospheric flux in order to predict r for a given dark matter candidate. We use the calculation of Volkova [22]; Volkova's results are in agreement with the data to date. (For a review, see ref. [23].)

The calculation of $\bar{S}$ and $\bar{A}$, as given by eq. (1), depends on the type of neutrino event we wish to consider: contained events, or through-going muons produced in the surrounding rock. Previously [10], we calculated limits on r based on the IMB data [13] for contained events with 1 GeV $< E_\nu <$ 2 GeV and through-going muon events with $E_{\nu_\mu} > 2$ GeV. We compared these limits to the predictions resulting from photinos, higgsinos, and sneutrinos as dark matter. Here we will use data from Kamioka [14] and Frejus [15] as well as IMB to derive a limit on r for contained events with $E_\nu > 2$ GeV. We will also compute the expected value of r for this case for a large variety of dark matter candidates.

Because the data we are using involves relatively few events, any limit we quote only makes sense up to some confidence level which depends on our statistical analysis. We use an analysis suggested by Protheroe [24] which is based on bayesian statistics [25]. Let T be the total number of events of a specified type (i.e., contained, neutrino energy within a certain range, etc.). Let C be the number of events of this type due to neutrinos coming from within a cone of solid angle $4\pi p$ centered on the sun. Given the values of $\bar{A}$ and $\bar{S}$, the probability of observing C events in the solar cone with a total of T events is the product of two Poisson distributions, one for events inside the cone, and one for events outside:

$$P(C, T; \bar{A}, \bar{S}) = \left(\frac{(\bar{S}+p\bar{A})^C \exp[-(\bar{S}+p\bar{A})]}{C!} \right) \times \left(\frac{[(1-p)\bar{A}]^{T-C} \exp[-(1-p)\bar{A}]}{(T-C)!} \right). \tag{3}$$

Of course, what we really want to do is use the observed values of C and T to get information on $\bar{A}$ and $\bar{S}$. All we know about $\bar{A}$ and $\bar{S}$, a priori, is that neither can be negative. We therefore assume that, a priori, $\bar{A}$ and $\bar{S}$ are equally likely to have any non-negative values. Then, according to Bayes' theorem [25], after doing the experiment (and finding C events in the solar cone with a total of T events) we can interpret $P(C, T; \bar{A}, \bar{S})$ as an unnormalized joint probability distribution for $\bar{A} \geq 0$ and $\bar{S} \geq 0$. The 90% confidence level (CL) upper limit on r, r_{90}, is then found by demanding that the probability that $\bar{S}$ is less than $r_{90}\bar{A}$ be exactly 0.9:

$$\frac{\int_0^\infty d\bar{A} \int_0^{r_{90}\bar{A}} d\bar{S}\, P(C, T; \bar{A}, \bar{S})}{\int_0^\infty d\bar{A} \int_0^\infty d\bar{S}\, P(C, T; \bar{A}, \bar{S})} = 0.9. \tag{4}$$

By changing the integration variables from $\bar{A}$ and $\bar{S}$ to $\bar{A}+\bar{S}$ and r, we can do the integral over $\bar{A}+\bar{S}$ and rewrite eq. (4) as

$$\frac{\int_0^{r_{90}} dr\, (1+r)^{-2}(r+p)^C(1+r)^{-T}}{\int_0^\infty dr\, (1+r)^{-2}(r+p)^C(1+r)^{-T}} = 0.9. \tag{5}$$

The integrals in eq. (5) can be done in closed form; a convenient statement of the result is

$$F(\lambda_{90}) = (0.1)\, F(p), \tag{6}$$

where

$$F(x) \equiv (1-x)^{T-C+1} \sum_{S=0}^{C} \frac{(T-C+S)!}{(T-C)!S!} x^S, \tag{7}$$

and

$$r_{90} = \frac{\lambda_{90} - p}{1-\lambda_{90}}. \tag{8}$$

Here $\lambda = (\bar{S}+p\bar{A})/(\bar{S}+\bar{A})$ is the probability that any one event is in the solar cone. Eq. (5) has a particularly simple solution when there are no events in the solar cone ($C=0$):

$$r_{90} = 10^{1/(T+1)} - 1. \tag{9}$$

Note that in this case, r_{90} is independent of p. This may seem counterintuitive. The point is that if $\bar{S}=0$, the expected value of C is pT. So we only expect to find $C=0$ if $pT \lesssim 1$. Further reducing p, so that $pT \ll 1$, does not give us any more information; in fact, one could argue that it gives us *less*, since $C=0$ is now even closer to the expectation value pT. (A

conventional [26], as opposed to bayesian, statistical analysis reflects this: it yields a value of λ_{90} which is always independent of p, and so r_{90} [see eq. (8)] *increases* as p decreases for fixed C and T.)

Previously, we had analyzed the IMB data for the following types of events:

(1) Contained events with 1 GeV $< E_\nu <$ 2 GeV, of which there were 11 events within 30° of the sun ($p=0.067$) out of a total of 89 events, yielding [10] $R \equiv r/p < 2.1$ at the 90% CL or

$$r < 0.14$$

(IMB, contained, 1 GeV $< E_\nu <$ 2 GeV). (10)

(2) Through-going muon events with $E_\nu < 2$ GeV, of which there were two events within 8° of the sun ($p=0.005$) out of a total of 187 events, yielding [10] $R < 4.8$ at the 90% CL or

$$r < 0.024$$

(IMB, through-going muons, $E_\nu > 2$ GeV). (11)

(3) Contained events with $E_\nu > 2$ GeV, of which there were none within 30° of the sun out of a total of 10 events, yielding a 90% CL limit of

$$r < 0.23$$

(IMB, contained, $E_\nu > 2$ GeV). (12)

(This number was not given in ref. [10]; we include it here for completeness.)

Recently, both the Kamioka and Frejus Collaborations reported data in searches for a high energy solar neutrino excess. For contained events with $E_\nu > 2$ GeV, Kamioka reports a total of 23 events, none in the solar direction, giving a 90% CL limit of

$$r < 0.10$$

(Kamioka, contained, $E_\nu > 2$ GeV). (13)

Frejus reports a total of 31 events, again with none in the solar direction, giving a 90% CL limit of

$$r < 0.075$$

(Frejus, contained, $E_\nu > 2$ GeV). (14)

Note that we do not need to know the angular resolutions of the detectors, since r_{90} is independent of p when $C=0$.

To combine data for contained events with $E_\nu > 2$ GeV from the three experiments, we need to modify eq. (3). The right-hand side is replaced by a multiple product of Poisson distributions (two for each experiment). Since each experiment will have a different observing time and fiducial volume, the values of $\bar{A}$ and $\bar{S}$ will differ among the experiments; however, the ratio $r=\bar{S}/\bar{A}$ will be universal (provided we are considering events of the same type). If p_i, C_i, and T_i are the corresponding quantities for each experiment, then the integrand in eq. (5), in both the numerator and denominator, is replaced by

$$(1+r)^{-2} \prod_i (r+p_i)^{C_i} (1+r)^{-T_i}. \quad (15)$$

The product is over the different experiments. If $C_i=0$ for all experiments, as it does for the data in the three experiments for contained events with $E_\nu > 2$ GeV, one can simply use eq. (9) with $T=\sum_i T_i$. In this case, there are a total of 64 events, yielding the 90% CL limit

$$r < 0.036$$

(combined, contained, $E_\nu > 2$ GeV). (16)

We note that these limits on r can be converted to limits on solar events per kiloton-year (for a given detector) by multiplying by the predicted number of atmospheric events per kiloton-year (for that detector).

Let us now compare these limits with our expectations for the various dark matter candidates. In tables 1–6, we show our predictions for r for the three event types we have discussed: (1) contained events

Table 1
Photinos.

$m_{\tilde{\gamma}}$ (GeV)	r		
	case 1	case 2	case 3
4	0.11	0.014	0.097
6	0.032	0.015	0.077
10	0.0067	0.016	0.054
20	–	0.0066	0.013
40	–	0.0023	0.0023
limit	0.14	0.024	0.036

Table 2
Higgsinos.

$m_{\tilde{H}}$ (GeV)	r		
	case 1	case 2	case 3
6	0.045	0.0091	0.052
10	0.0084	0.010	0.040
20	–	0.0057	0.012
40	–	0.0023	0.0025
limit	0.14	0.024	0.036

Table 3
Majorana neutrinos.

m_{ν_M} (GeV)	r		
	case 1	case 2	case 3
6	0.055	0.011	0.064
10	0.023	0.028	0.11
20	0.0058	0.051	0.11
40	–	0.078	0.082
limit	0.14	0.024	0.036

Table 4
Dirac neutrinos.

m_{ν_D} (GeV)	r		
	case 1	case 2	case 3
4	–	0.022	0.087
6	–	0.044	0.12
10	–	0.10	0.17
20	–	0.30	0.24
40	–	0.70	0.28
limit	0.14	0.024	0.036

Table 5
Scalar electron neutrinos.

$m_{\tilde{\nu}_e}$ (GeV)	r		
	case 1	case 2	case 3
4	–	–	3.7
6	–	–	6.3
10	–	–	11
20	–	–	20
40	–	–	26
limit	0.14	0.024	0.036

Table 6
Scalar muon neutrinos.

$m_{\tilde{\nu}_\mu}$ (GeV)	r		
	case 1	case 2	case 3
4	–	0.50	1.4
6	–	1.3	2.3
10	–	3.8	4.1
20	–	13	9.5
40	–	35	18
limit	0.14	0.024	0.036

with 1 GeV $< E_\nu < 2$ GeV; (2) through-going muon events with $E_{\nu_\mu} > 2$ GeV; and (3) contained events with $E_\nu > 2$ GeV. To calculate $\bar{A}$, we use the differential atmospheric fluxes derived by Volkova [22]. To calculate $\bar{S}$ for photinos, higgsinos, and Majorana neutrinos, we take the differential fluxes $d\Phi_S/dE$ from ref. [8]. For solar neutrinos, $d\Phi_{S,\nu_e}/dE = d\Phi_{S,\bar{\nu}_e}/dE = d\Phi_{S,\nu_\mu}/dE = d\Phi_{S,\bar{\nu}_\mu}/dE$. These fluxes are integrated over the appropriate energy range. For Dirac neutrinos and sneutrinos, we consider only the monochromatic flux of neutrinos resulting from annihilation into a single $\nu\bar{\nu}$ pair, and use the total flux calculated in ref. [6]. The differential flux in these cases is just $d\Phi_S/dE = \Phi_S \delta(E - m_X)$. For photinos, we assume that all scalar quark and lepton masses are equal, and adjust the value of this common mass so that $\Omega = 1$ with $h = 0.5$ [#1]. For higgsinos, we adjust the ratio of Higgs field expectation values V_1/V_2 so that $\Omega = 1$ with $h = 0.5$. These adjustments, in turn, affect the scattering cross sections which are used to compute the rate at which the sun traps dark matter particles. For heavy neutrinos (Majorana or Dirac) and sneutrinos, we do not adjust any parameters other than the mass. In the case of heavy neutrinos, this means that Ω will vary with the mass; it will equal one for $m_{\nu_M} \simeq 6$ GeV in the Majorana case and for $m_{\nu_D} \simeq 4.5$ GeV in the Dirac case [19,27], but drop as these masses increase. We also note that a higgsino with V_1/V_2 equal to zero or infinity (called "generic" in ref. [8]) is equivalent to a Majorana neutrino. For sneutrinos with $m_{\tilde{\nu}} \lesssim 7$ GeV, the scattering cross sec-

[#1] Ω is the mass density of relic particles in units of the critical density $\rho_c = 1.054 \times 10^{-5}$ GeV cm^{-3}, and h is the Hubble parameter in units of 100 km s^{-1} Mpc^{-1}.

tion is independent of the annihilation cross section, and we can have $\Omega=1$; for larger values of $m_{\tilde{\nu}}$, the annihilation cross section rises due to Z^0 exchange, and Ω decreases [6]. We make use of the detailed results of a forthcoming numerical study of the rate equation governing relic stable particle densities [27]. In some cases these results differ substantially from the approximate formulas we used previously. The greatest uncertainty in the results comes from our lack of knowledge about the quark/hadron phase transition, and is most significant for particles with masses of 5 GeV to 10 GeV. To be conservative, we have assumed a transition temperature near 100 MeV, which minimizes the expected solar fluxes relative to those we would find using a larger transition temperature. We have also employed Gould's corrections [5] to the Press–Spergel [4] trapping rate; in addition to an overall correction factor of 1.5, we used the solid line in Gould's fig. 3 for photinos, higgsinos, and Majorana neutrinos, and the dashed line for Dirac neutrinos and sneutrinos.

Comparing the values in the tables, we see that the expected value of r exceeds the experimental 90% CL upper limit in case 2 or case 3 for a variety of dark matter candidates: photinos with 4 GeV $\lesssim m_{\tilde{\gamma}} \lesssim 10$ GeV; higgsinos with 6 GeV $\lesssim m_{\tilde{H}} \lesssim 10$ GeV; Majorana neutrinos with $m_{\nu_M} \gtrsim 6$ GeV; Dirac neutrinos with $m_{\nu_D} \gtrsim 4$ GeV; and sneutrinos (either electron type or muon type) with $m_{\tilde{\nu}} \gtrsim 4$ GeV. However, the uncertainties involved make it difficult to make a strong case against any of the candidates except sneutrinos. An example of the type of uncertainty involved is the local density of dark matter. The trapping rate – and hence the annihilation rate and the neutrino flux – depend on the combination of $n_X/\bar{v}$, where n_X is the local number density for X's and $\bar{v}$ is their RMS velocity. We have taken $n_X = (0.3\ \text{GeV}/m_X)\ \text{cm}^{-3}$ and $\bar{v}=300$ km/s. Previously [10] we noted that uncertainties in both n_X and $\bar{v}$ would allow the ratio to vary from 0.14 to 2.3 times our nominal value. Recently, the question of local density has been reexamined [28] and it was suggested that perhaps the uncertainty range is much smaller: 0.7–1.4 times our nominal value. Other studies [29] involving a flattened halo allow for larger densities of up to a factor of three times our nominal value. Another uncertainty [12] is the effect of fragmentation on the annihilation products. Such an effect might lower the neutrino fluxes by a factor of two. We think the overall uncertainty is at least a factor of four.

Looking back at the tables, we see that for photinos and higgsinos, the predicted values of r never exceed the experimental limits by a factor of four or more. For heavy neutrinos and sneutrinos, we still exceed the limits by a factor of four for larger values of the mass. However, it is worth mentioning again that these cases are somewhat artificial in that we do not expect $\Omega=1$ for these candidates except for specific values of the mass. At those values, we have the excess factor of four that we would like only for sneutrinos (both electron and muon type). Furthermore, if Ω drops below about 0.05, there would not be enough dark matter to form galactic halos; this occurs for Majorana neutrinos if $m_{\nu_M} \gtrsim 35$ GeV and for Dirac neutrinos if $m_{\nu_D} \gtrsim 16$ GeV [27].

It is also important to note that we are using a 90% confidence level. If we consider the 95% CL instead, the limit on r in case 3, eq. (16), becomes $r<0.047$; at the 99% CL, the limit becomes $r<0.073$.

In conclusion, the non-observation of a high energy solar neutrino flux appears to rule out scalar muon neutrinos and scalar electron neutrinos as dark matter. The other dark matter candidates are not presently ruled out when one takes into account the various uncertainties in the input parameters (such as n_X), but certain ranges of this parameter space can be excluded. The viability of photinos and higgsinos as dark matter will be further challenged by future data.

We would like to thank David Caldwell and Rick Watkins for helpful discussions. K.A.O. is supported in part by DOE Grant No. DE-AC02-83ER-40105, and by a Presidential Young Investigator Award. M.S. is supported in part by NSF Grant No. PHY86-14185, and by an Alfred P. Sloan Research Fellowship.

References

[1] D.N. Schramm, Nucl. Phys. B 252 (1985) 53.
[2] J. Silk, K.A. Olive and M. Srednicki, Phys. Rev. Lett. 55 (1985) 257.
[3] G. Steigman, C. Sarazin, H. Quintana and J. Faulkner, Astron. J. 83 (1978) 1050.
[4] W.H. Press and D.N. Spergel, Astrophys. J. 296 (1985) 67.
[5] A. Gould, Astrophys. J. 231 (1987) 571.

[6] M. Srednicki, K.A. Olive and J. Silk, Nucl. Phys. B 279 (1987) 804.
[7] T.K. Gaisser, G. Steigman and S. Tilav, Phys. Rev. D 34 (1986) 2206.
[8] J.S. Hagelin, K.-W. Ng and K.A. Olive, Phys. Lett. B 180 (1986) 375.
[9] K. Greist and D. Seckel, Nucl. Phys. B 283 (1987) 681.
[10] K.-W. Ng, K.A. Olive and M. Srednicki, Phys. Lett. B 188 (1987) 138.
[11] S. Ritz and D. Seckel, CERN preprint TH-4627 (1987).
[12] J. Ellis, R.A. Flores and S. Ritz, Phys. Lett. B 198 (1987) 393.
[13] IBM Collab., J. LoSecco et al., Phys. Lett. B 188 (1987) 388.
[14] Kamioka Collab., Y. Totsuka, University of Tokyo preprint UT-ICEPP-87-02 (1987).
[15] Frejus Collab., B. Kuznik, Orsay preprint LAL 87-21 (1987).
[16] H. Goldberg, Phys. Rev. Lett. 50 (1983) 1419; L.M. Krauss, Nucl. Phys. B 227 (1983) 556.
[17] J. Ellis, J. Hagelin, D.V. Nanopoulos, K.A. Olive and M. Srednicki, Nucl. Phys. B 238 (1984) 453.
[18] L.M. Krauss, Phys. Lett. B 128 (1983) 37.
[19] E.W. Kolb and K.A. Olive, Phys. Rev. D 33 (1986) 1202; D 34 (1986) 2531 (E).
[20] P. Hut, Phys. Lett. B 69 (1977) 85; B. Lee and S. Weinberg, Phys. Rev. Lett. 39 (1977) 165.
[21] L.E. Ibáñez, Phys. Lett. B 137 (1984) 160; J. Hagelin, G.L. Kane and S. Raby, Nucl. Phys. B 241 (1984) 638.
[22] L.V. Volkova, Sov. J. Nucl. Phys. 31 (1980) 784.
[23] D.H. Perkins, Annu. Rev. Nucl. Part. Sci. 34 (1984) 1.
[24] R.J. Protheroe, Astron. Express 1 (1984) 33.
[25] S.A. Schmitt, Measuring uncertainty: an elementary introduction to bayesian statistics (Addison-Wesley, Reading, 1969).
[26] Particle Data Group, M. Aguilar-Benitez et al., Review of particle properties, Phys. Lett. B 170 (1986) 52.
[27] M. Srednicki, R. Watkins and K.A. Olive, in preparation.
[28] R. Flores, CERN preprint TH-4736/87 (1987).
[29] J. Binney, A. May and J. Ostriker, Mon. Not. R. Astron. Soc. 226 (1987) 149.

Experimental Tests of the "Invisible" Axion

P. Sikivie
Physics Department, University of Florida, Gainesville, Florida 32611
(Received 13 July 1983)

Experiments are proposed which address the question of the existence of the "invisible" axion for the whole allowed range of the axion decay constant. These experiments exploit the coupling of the axion to the electromagnetic field, axion emission by the sun, and/or the cosmological abundance and presumed clustering of axions in the halo of our galaxy.

PACS numbers: 14.80.Gt, 11.30.Er, 95.30.Cq

Some time ago, it was shown that the strong CP problem[1] can be solved[2] by the introduction of a light pseudoscalar particle,[3] called the axion a. The properties of the axion depend mainly on the magnitude v of the vacuum expectation value that spontaneously breaks the $U_{PQ}(1)$ quasisymmetry which was postulated by Peccei and Quinn[2] and of which the axion is the pseudo-Goldstone boson. The axion mass and its couplings to ordinary particles are all inversely proportional to v. As far as the solution to the strong CP problem is concerned, the value of v is arbitrary.[4] Past experiments,[5] attempting to produce and detect axions in the laboratory, essentially rule out values of v near 250 GeV. Moreover the range 250 GeV $\lesssim v \lesssim 10^8$ GeV is ruled out by considering the effect axions have on stellar evolution.[6] Stars emit too many axions for those values of v. The axion with $v \gtrsim 10^8$ GeV is so weakly coupled[4] that it has been called "invisible." It was thought, incorrectly I believe, that such an axion solves the strong CP problem in a manner which is free of presently observable consequences.

Actually, the consideration of the cosmological implications[7,8] of axion models has already gone some way towards making a misnomer of the expression "invisible" axion. First, axion models are afflicted with cosmologically unacceptable domain walls[7] unless special precautions are taken to avoid this.[9] Second, it was shown that the present cosmological axion energy density is too large[8] unless $v \lesssim 10^{12}$ GeV. On the other hand, axions may have a useful cosmological role to play with regard to the problem of galaxy formation.[7,10] First, the primordial density perturbations from which galaxies evolved may have been produced by the presence of axionic domain walls for a limited time period in the early universe.[7] Second, axions may be the stuff the dark halos[11] of galaxies are made of.[10] Because of their very large primordial phase-space density, axions cluster easily and if $v \gtrsim 10^{10}$ GeV, axions are abundant enough to provide all the halo matter. In that case, the axion density near the sun's location is about

$$\rho_{a,\mathrm{halo}} \simeq \frac{10^{-24}\ \mathrm{g}}{\mathrm{cm}^3} \simeq \frac{0.5\times 10^{12}\ \mathrm{axions}}{\mathrm{cm}^3}\left(\frac{v}{10^{10}\ \mathrm{GeV}}\right)\frac{1}{r}, \quad (1)$$

where we have used the following expression for the axion mass ($\hbar = c = 1$ everywhere):

$$m_a = 1.24\times 10^{-3}\ \mathrm{eV}[(10^{10}\ \mathrm{GeV})/v]r = \frac{2\pi}{10^{-1}\ \mathrm{cm}}\left(\frac{10^{10}\ \mathrm{GeV}}{v}\right)r = \frac{2\pi}{\frac{1}{3}\times 10^{-11}\ \mathrm{sec}}\left(\frac{10^{10}\ \mathrm{GeV}}{v}\right)r. \quad (2)$$

r is a model dependent number of order $N/6$, where N is the number of vacua of the axion model.[7,9] These halo axions have velocities $\beta \lesssim 10^{-3}$, and thus would form a highly degenerate Bose gas with quantum state occupation numbers averaging $(10^{17}/r^4)[v/(10^{10}\ \mathrm{GeV})]^4$. If the cosmological axions did not cluster into galactic halos [possibly because galactic clusters condensed before galaxies did], the axion density on earth should be the average cosmological one which, according to the result of Ref. 8, is about

$$\frac{0.2\times 10^6\ \mathrm{axions}}{\mathrm{cm}^3}\left(\frac{v}{10^{10}\ \mathrm{GeV}}\right)^{13/6}\frac{1}{r}.$$

If their typical velocity is the virtual velocity in galactic clusters, $\beta \simeq 5\times 10^{-3}$, their average quantum state occupation number is about $0.4(10^9/r^4)[v/(10^{10}\ \mathrm{GeV})]^{31/6}$. The first two experiments proposed below would attempt to detect the axions of cosmological origin. The third experiment would attempt to detect the axions emitted by the sun. From the work of Fukugita, Watamura, and Yoshimura,[6] one obtains the following result for the solar axion flux on earth:

$$f_{a,\odot} \simeq \frac{0.8\times 10^{13}\ \mathrm{axions}}{\mathrm{sec\ cm}^2}\left(\frac{10^8\ \mathrm{GeV}}{v}\right)^2. \quad (3)$$

© 1983 The American Physical Society

The solar axions have energies of order one kiloelectronvolt.

To make the axion visible, we will exploit the coupling of the axion to the electromagnetic field and the fact that we have available in the laboratory large static magnetic or electric fields, or large oscillating ones with frequencies of order the axion mass (2). We also have very sensitive devices to monitor the electromagnetic field. The effective action density of axions and photons is

$$\mathcal{L} = -\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{e^2N}{12\pi^2}\frac{a}{v}F_{\mu\nu}\tilde{F}^{\mu\nu} + \tfrac{1}{2}\partial_\mu a\partial^\mu a - \tfrac{1}{2}m_a{}^2a^2[1+O(a^2/v^2)], \tag{4}$$

where $\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta}$, $F_{\alpha\beta} \doteq \partial_\alpha A_\beta - \partial_\beta A_\alpha$, and where we have assumed grand unification of the strong and electroweak interactions with the unrenormalized $\sin^2\theta_w{}^0 = \frac{3}{8}$. The action density (4) has recently been used to study[12] the long-range interactions [$\sim$(distance)$^{-2}$] of electric and magnetic charges with axionic domain walls, which behave towards them like polarizable surfaces. The classical equations of motion derived from (4) are

$$\nabla\cdot\vec{E} = \frac{e^2N}{3\pi^2v}\vec{B}\cdot\nabla a, \quad \nabla\times\vec{B} - \frac{\partial\vec{E}}{\partial t} = \frac{e^2N}{3\pi^2v}\left[\vec{E}\times\nabla a - \vec{B}\frac{\partial a}{\partial t}\right], \quad \Box a = \frac{e^2N}{3\pi^2v}\vec{E}\cdot\vec{B} - m_a{}^2a. \tag{5}$$

The axion haloscope.—According to (4) or (5), axions will convert to photons in a strong inhomogeneous magnetic field $B_0(\vec{x})$. The inhomogeneity is necessary because three-momentum must be provided for the transition to occur. We find the following cross section for a detector of volume V:

$$\sigma = \frac{1}{16\pi^2|\vec{\beta}_a|}\left(\frac{e^2N}{3\pi^2v}\right)^2\sum_\lambda\int d^3k_\gamma\,\delta(E_\gamma - E_a)\left|\int_V d^3x\, e^{i\vec{q}\cdot\vec{x}}\vec{B}_0(\vec{x})\cdot\vec{\epsilon}(\vec{k}_\gamma,\lambda)\right|^2, \tag{6}$$

where $\vec{q} = \vec{k}_\gamma - \vec{k}_a$ and the sum is over the photon polarization. Multiplying (6) by the axion flux of the Milky Way halo, one obtains the rate

$$\frac{\text{No. of photons}}{\text{time}} \simeq \frac{1.6}{10^6\ \text{sec}}\frac{V}{1\ \text{cm}^3}\left(\frac{B_0}{1\ \text{T}}\right)^2 R(m_a)\left(\frac{N}{6r}\right)^2, \tag{7}$$

where $R(m_a)$ is a measure of the detector's response,

$$R = \frac{E_a}{V}\int\frac{d^3k_\gamma}{(2\pi)^3}\,\delta(k_\gamma - E_a)\sum_\lambda\left|\int_V d^3x\, e^{i\vec{q}\cdot\vec{x}}\vec{b}(\vec{x})\cdot\vec{\epsilon}(\vec{k}_\gamma,\lambda)\right|^2, \tag{8}$$

and $\vec{b}(\vec{x}) = B_0{}^{-1}\vec{B}_0(\vec{x})$. One can choose to have large values of $R(m)$ over a small frequency band near m_0 by making the spatial dependence of $\vec{B}_0(\vec{x})$ periodic over many wavelengths $2\pi/m_0$. If $R = 10$, $B_0 = 10$ T, and $V = (30\ \text{cm})^3$, there are about 45 $a\to\gamma$ events in the detector per second. Antennas exist which can detect single microwave quanta. For the signal not to be swamped by thermal noise it will be necessary to cool the volume V to below approximately $\frac{1}{40}m_a \simeq (0.36\ \text{K})[(10^{10}\ \text{GeV})/v]r$. One way to obtain the desired inhomogeneity of the magnetic field is to embed grains or wires of a superconducting metal in a material transparent to microwave radiation. When the detector is cooled below the critical temperature, the magnetic flux lines will be expelled from the superconducting loci and hence will be made inhomogeneous. If the cosmological axions did not cluster into galactic halos, the rate (7) is reduced by $0.4\times10^{-6}\,[v/(10^{10}\ \text{GeV})]^{7/6}$.

To probe large values of v (small values of the axion mass) it may be advantageous to use a variable-frequency cavity in the axion haloscope experiment. One would attempt to tune the cavity's frequency ω exactly to the energy of the Milky Way axions,

$$m_a < \omega < m_a + \tfrac{1}{2}m_a\beta^2 = m_a[1+O(10^{-6})]. \tag{9}$$

If Eq. (9) is satisfied, the power due to axion $\to$ photon conversion into the lowest TM mode of a rectangular cavity of length d and square cross section, in which there is a longitudinal static magnetic field B_0, is

$$P = (1.5\times10^{-15}\ \text{W})\frac{1}{N^2}\left(\frac{N}{6r}\right)^4\left(\frac{v}{10^{12}\ \text{GeV}}\right)^2\left(\frac{B_0}{10\ \text{T}}\right)^2 kd, \tag{10}$$

where $k = (\omega^2 - m_a{}^2)^{1/2}$.

Finally, let us take note of the interest of trying to devise an experiment which exploits the very high quantum degeneracy of the axions in the halo of our galaxy. In particular, emission of axions with energies between m_a and $m_a[1+O(10^{-6})]$ is enormously stimulated {the rate is multiplied by a factor $10^{25}[v/(10^{12}\ \mathrm{GeV})]^4$}, and hence may become measurable. It is necessary, however, that the emitter does not equally well absorb axions; otherwise, stimulated emission and stimulated absorption will cancel each other, as is the case with a cavity or with any other *harmonic* oscillator coupled to the axion field.

The axion helioscope.—The idea here is the same as the previous one but now applied to the solar axion flux. In a strong magnetic field, solar axions convert to x rays. The change in three-momentum is

$$q_z = \frac{1}{2}\frac{m_a^2}{E_a} \simeq \frac{2\pi}{16\ \mathrm{cm}}\left(\frac{10^8\ \mathrm{GeV}}{v}\right)^2\left(\frac{1\ \mathrm{keV}}{E_a}\right)r^2. \tag{11}$$

Consider a detector of length L in the direction $\vec{n}$ of the sun, inside of which there is a transverse magnetic field $\vec{B}_0 = \hat{t}B_0\cos[(2\pi/d)\vec{n}\cdot\vec{x}]$. The response (10) is

$$R = \frac{E_a L}{8\pi}\left[\frac{\sin(2\pi/d - q_z)L/2}{(2\pi/d - q_z)L/2} + \frac{\sin(2\pi/d + q_z)L/2}{(2\pi/d + q_z)L/2}\right]^2. \tag{12}$$

R can be huge and compensate for the smallness of the solar axion flux. (It is of course unessential for this that the spatial dependence of $\vec{B}_0$ is exactly a cosine.) Multiplying (8) by (3) one finds the rate

$$\frac{\text{No. of x rays}}{\text{time}} \simeq \frac{6\times10^{-3}}{\text{sec}}\frac{SL^2}{1\ \mathrm{m}^4}\left(\frac{10^8\ \mathrm{GeV}}{v}\right)^4\left(\frac{B_0}{10\ \mathrm{T}}\right)^2 N^2\frac{8\pi R}{E_a L}, \tag{13}$$

where S is the area of the detector perpendicular to $\vec{n}$.

In conclusion, the "invisible" axion hypothesis can be tested experimentally, contrary to what was at first believed. Relatively simple experiments can provide new information about physics at very high energies, near the grand unification mass scale. If the axion exists, we will have new powerful tools to study the sun and the galaxy.

This research was supported in part by U. S. Department of Energy Contract No. DE-AS05-81-ER40008. I am very much indebted to N. Sullivan for his interest and for his patience in explaining experimental realities to me. I am also grateful to H. Georgi and C. Hooper for stimulating conversations. Finally, I thank the Harvard Physics Department for their hospitality while part of this work was being done.

[1]A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Y. S. Tyuplin, Phys. Lett. 59B, 85 (1975); G. 't Hooft, Phys. Rev. Lett. 37, 8 (1976), and Phys. Rev. D 14, 3432 (1976); R. Jackiw and C. Rebbi, Phys. Rev. Lett. 37, 172 (1976); C. G. Callan, R. F. Dashen, and D. J. Gross, Phys. Lett. 63B, 334 (1976).

[2]R. D. Peccei and H. Quinn, Phys. Rev. Lett. 38, 1440 (1977), and Phys. Rev. D 16, 1791 (1977).

[3]S. Weinberg, Phys. Rev. Lett. 40, 223 (1978); F. Wilczek, Phys. Rev. Lett. 40, 279 (1978).

[4]J. Kim, Phys. Rev. Lett. 43, 103 (1979); M. Shifman, A. Vainshtein, and V. Zakharov, Nucl. Phys. B166, 493 (1980); M. Dine, W. Fischler, and M. Srednicki, Phys. Lett. 104B, 199 (1981); H. P. Nilles and S. Raby, Stanford Linear Accelerator Center Report No. SLAC-PhB-2743, 1981 (unpublished).

[5]For a recent review, see A. Zehnder, SIN Report No. PR-83-03, 1983 (to be published).

[6]D. Dicus, E. Kolb, V. Teplitz, and R. Wagoner, Phys. Rev. D 18, 1829 (1978), and Phys. Rev. D 22, 839 (1980); K. Sato and H. Sato, Prog. Theor. Phys. 54, 1564 (1975); M. Fukugita, S. Watamura, and M. Yoshimura, Phys. Rev. Lett. 48, 1522 (1982).

[7]P. Sikivie, Phys. Rev. Lett. 48, 1156 (1982).

[8]L. Abbott and P. Sikivie, Phys. Lett. 120B, 133 (1983); J. Preskill, M. Wise, and F. Wilczek, Phys. Lett. 120B, 127 (1983); M. Dine and W. Fischler, Phys. Lett. 120B, 137 (1983).

[9]For a recent review and list of references, see P. Sikivie, University of Florida Report No. UFTP-83-6, 1983 (to be published).

[10]J. Ipser and P. Sikivie, Phys. Rev. Lett. 50, 925 (1983); F. W. Stecker and Q. Shafi, Phys. Rev. Lett. 50, 928 (1983); M. S. Turner, F. Wilczek, and A. Zee, to be published.

[11]J. P. Ostriker, P. J. E. Peebles, and A. Yahil, Astrophys. J. Lett. 193, L1 (1974); J. Einasto, A. Kaasik, and E. Saar, Nature (London) 250, 309 (1974).

[12]P. Sikivie, to be published.

ERRATA

EXPERIMENTAL TESTS OF THE "INVISIBLE" AXION. P. Sikivie [Phys. Rev. Lett. 51, 1415 (1983)].

Please replace Eq. (10) and the sentence preceding it by the following: "When m_a equals the frequency of one of the TM_{nm0} modes of a rectangular cavity (the longitudinal direction is that of the static magnetic field $\vec{B}_0$) with n and m odd, the power into that mode from axion – photon conversion is

$$P_{nm} = (3\times10^{-18}\ \mathrm{W})\left(\frac{V}{10^5\ \mathrm{cm}^3}\right)\left(\frac{B_0}{10\ \mathrm{T}}\right)^2 \frac{1}{n^2m^2}\left(\frac{m_a}{1.24\times10^{-5}\ \mathrm{eV}}\right)\left(\frac{N}{6r}\right)^2 \min(1, 10^{-6}Q_{nm}), \tag{10}$$

where Q_{nm} is the corresponding quality factor and V is the volume of the cavity."

Detectability of certain dark-matter candidates

Mark W. Goodman and Edward Witten
Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08544
(Received 7 January 1985)

We consider the possibility that the neutral-current neutrino detector recently proposed by Drukier and Stodolsky could be used to detect some possible candidates for the dark matter in galactic halos. This may be feasible if the galactic halos are made of particles with coherent weak interactions and masses $1-10^6$ GeV; particles with spin-dependent interactions of typical weak strength and masses $1-10^2$ GeV; or strongly interacting particles of masses $1-10^{13}$ GeV.

Dark galactic halos[1] may be clouds of elementary particles so weakly interacting or so few and massive that they are not conspicuous. Many dark-matter candidates have been proposed. Magnetic monopoles are one dark-matter candidate accessible to experimental search,[2] and the same seems to be true for axions.[3] On the other hand, massive neutrinos are a popular dark-matter candidate which seems very difficult to detect except under very favorable conditions.[4] For many other dark-matter candidates considered in the literature, no practical experiments have been proposed.

Recently, Drukier and Stodolsky proposed[5] a new way of detecting solar and reactor neutrinos. The idea is to exploit elastic neutral-current scattering of nuclei by neutrinos (a mechanism that is also believed to play an important role in supernovas).[6] The detector will consist of superconducting grains of radius a few microns embedded in a nonsuperconducting material in a magnetic field. The grains are maintained just below their superconducting transition temperature. A scattered neutrino will impart a small recoil kinetic energy to the nucleus it scatters from (of order 1—100 eV in the experiments considered in Ref. 5). Such a small energy deposit can make a tiny superconducting grain go normal, permitting the magnetic flux to collapse into the grain and producing an electromagnetic signal in a read-out circuit. The principle of such a detector has already been demonstrated.[7]

In this paper, we will calculate the sensitivity of the detector considered in Ref. 5 to various dark-matter candidates. Although this detector is not very sensitive to halo neutrinos (with their tiny masses and interaction rates), it has, as we will see, a useful sensitivity to some other dark-matter candidates. We also mention some other detection schemes.

We will consider three classes of dark-matter candidates: particles with coherent weak couplings; particles with spin-dependent couplings of roughly weak strength; and particles with strong interactions. If a detector sensitive to 1 event/kg day can be built, useful limits can be placed on these particles in the mass ranges $1-10^6$ GeV, $1-10^2$ GeV, and $1-10^{13}$ GeV, respectively (see Table I). The main difficulty in detecting these particles comes from backgrounds of radioactivity and cosmic rays, which we do not attempt to estimate here; such estimates were made in Ref. 5.

Let us first discuss the lower limit on detectable masses. If a halo particle of mass m and velocity v scatters from a target nucleus of mass M, the recoil momentum is at most $2mv$ and the recoil kinetic energy is at most $\epsilon=(2mv)^2/2M$. A reasonable value of v is $v=200$ km/sec. The lightest nucleus considered in Ref. 5 is aluminum, with $A=27$ and $M\simeq 27$ GeV. There seems to be a reasonable chance of building a detector sensitive to $\epsilon\sim 50-100$ eV (considerably more optimistic possibilities are discussed in Ref. 5). For $\epsilon\gtrsim 50-100$ eV, we need $m\gtrsim 1-2$ GeV, and this is the lower limit on the mass of detectable halo particles. It is important to note, though, that much larger values of m, say $m\gtrsim 100$ GeV, are also of interest in the dark-matter searches we envision. Thus values of ϵ up to 10—100 keV are of interest.

Consider elastic scattering of halo particles of mass m by target nuclei of mass M. The elastic scattering cross section is $\sigma=[m^2M^2/\pi(m+M)^2]\,|\mathcal{M}|^2$, assuming the invariant amplitude $\mathcal{M}$ is a constant (independent of angles) at low energy. If ρ is the mass density of halo parti-

TABLE I. Some experiments using the detector in Ref. 5. The spallation, reactor, and solar neutrino experiments were considered in Ref. 5. The event rate given for the spallation source refers to "reactor on." The supernova experiment of Ref. 5, which involves detection of a pulse, is not comparable to the others and is not included.

Experimental source	Event rate in $kg^{-1}\,day^{-1}$	Recoil energy range
Spallation source	10^2-10^3	10—100 keV
Reactor	10	50—500 eV
Solar neutrinos		
pp cycle	$10^{-3}-10^{-2}$	1—10 eV
^{7}Be	$10^{-2}-5\times10^{-2}$	5—50 eV
^{8}B	$10^{-3}-10^{-2}$	100 eV—3 keV
Galactic halo		
coherent $m\sim 2$ GeV	50—1000	10—100 eV
$m\gtrsim 100$ GeV	up to 10^4	10—100 keV
Spin dependent		
$m\sim 2$ GeV	0.1—1	10—100 eV
$m\gtrsim 100$ GeV	up to 1	10—100 keV

©1985 The American Physical Society

cles, their number density is ρ/m, and the flux is $F=\rho v/M$, where v is the mean velocity. The interaction rate in a detector with K target nuclei is $R=KF\sigma$. For a one-kilogram detector and target nuclei of Z protons, N neutrons, $N+Z=A$, the number of target nuclei is $K=6.0\times10^{26}/A$. The counting rate per kilogram of detector per day is

$$R=\frac{5.8\text{ events}}{\text{kg day}}\left[\frac{\bar{\sigma}}{10^{-38}\text{ cm}^2}\right]\left[\frac{\rho}{10^{-24}\text{ g/cm}^3}\right]\times\left[\frac{V}{200\text{ km/sec}}\right], \quad (1)$$

where for later convenience we define

$$\bar{\sigma}=(\sigma/A)(1\text{ GeV/m}) .$$

Now let us consider in turn particles of the three classes envisioned above.

(i) Particles with coherent weak interactions

We first assume that the unknown halo particle, X, has vector couplings to Z bosons and scatters from nuclei by Z exchange (the process considered in Ref. 5 for neutrinos). One dark-matter candidate with this property is the scalar partner of the neutrino in supersymmetric theories.[8] For such particles, the axial couplings, which only produce small spin-dependent effects, can be neglected. In the nonrelativistic limit, the weak scattering amplitude is

$$\mathcal{M}=\frac{g^2+g'^2}{M_Z{}^2}J_X^0J_T^0=4\sqrt{2}G_FJ_X^0J_T^0 ,$$

where J_X^0 and J_T^0 are the zero components of the weak neutral current of the X particles and the target nucleus.

For the neutral X particle, J_X^0 depends only on the hypercharge. If X is a fermion, its left- and right-handed components may have separate hypercharge Y_L and Y_R. Then $J_X^0=\bar{Y}/2$, where $\bar{Y}=\frac{1}{2}(Y_L+Y_R)$. (In the special case $Y_L=-Y_R$, there is not coherent coupling to Z bosons. Then the X particle has only spin-dependent interactions with nuclei, and its phenomenology will resemble that of photinos, discussed later.) If X is a boson, let $\bar{Y}$ be its weak hypercharge; then again $J_0^X=\bar{Y}/2$.

For a nucleus with N neutrons and Z protons, let $\bar{N}=N-(1-4\sin^2\theta)Z$; then $J_T^0=\bar{N}/4$. Hence the scattering amplitude is $\mathcal{M}=G_F\bar{Y}\bar{N}/\sqrt{2}$, and the cross section is $\sigma=[m^2M^2/2\pi(m+M)^2]G_F{}^2\bar{Y}^2\bar{N}^2$ so the quantity $\bar{\sigma}$ defined earlier is

$$\bar{\sigma}=(2.0\times10^{-35}\text{ cm}^2)\left[\frac{4mM}{(m+M)^2}\right]\bar{Y}^2\left[\frac{\bar{N}}{100}\right]^2 . \quad (2)$$

Comparing to our previous formula for the event rate, we see rates of order 10^4/kg day are obtained if $m\sim M$, $\bar{Y}\sim1$, $\bar{N}\sim100$. For a target such as lead, the event rate is at least 1/kg day for $m\lesssim10^7$ GeV. Equation (2) ignores the finite size of the nucleus, which is unimportant as long as $kR\ll1$. For lead this does not hold, and the effect is to lower the maximum detectable mass to about 10^6 GeV. See Table II for typical reaction rates.

TABLE II. Event rates in $\text{kg}^{-1}\text{day}^{-1}$ for coherent weak interactions with various detector materials and various values of the mass M_X of the unknown particles. A correction to account for loss of coherence at high momentum has been included.

M_X (GeV)	Al	Ge	Sn	Pb
1	27	90	170	370
10	150	700	1400	3.4×10^3
10^2	120	1400	4700	6.0×10^3
10^3	18	340	1700	6.0×10^2
10^4	2	40	200	58
10^5		4	21	5.8
10^6			2	0.6

(ii) Particles with spin-dependent interactions

We now consider the possibility that the dark matter consists of particles that interact with nuclei only via spin-dependent forces. As an interesting and representative example, we will consider the possibility[9] that the dark matter consists of photinos. This possibility has often been considered on the hypothesis that the photino mass is very small ($\ll1$ GeV), but the mass range of interest to us has also been considered.[10]

Photinos interact with quarks via the exchange of scalar quarks (Fig. 1). If there is important mixing between left- and right-handed scalar quarks, the photino gets coherent couplings to quarks from Fig. 1(b). However, this is unfavored in most models. We first assume mixing is unimportant, so that the photino has only spin-dependent interactions, from Fig. 1(a).

Let Q be the light quark (u or d) whose scalar partners are lightest. In general, the scalar partners of the left- and right-handed components of Q may have different masses, but we consider first the case where they have a common mass $M_{\tilde{Q}}$. Let q be the electric charge of Q. As explained in Ref. 11, Fig. 1(a) corresponds to an amplitude $\mathcal{M}=(q^2/M_{\tilde{Q}}{}^2)\bar{\tilde{\gamma}}\gamma_\mu\gamma_5\tilde{\gamma}\bar{Q}\gamma^\mu\gamma^5Q$, where $\tilde{\gamma}$ is the pho-

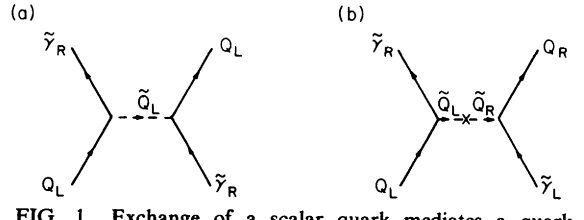


FIG. 1. Exchange of a scalar quark mediates a quark-photino interaction. Part (a), which is always present, gives a spin dependent force; (b), which is present only if there is strong mixing of left- and right-handed scalar quarks, gives a coherent interaction.

tino field. In the nonrelativistic limit, the time component of the axial-vector current is negligible, and $\bar{\tilde{\gamma}}\gamma\gamma_5\tilde{\gamma}$ reduces to $2\mathbf{S}_{\tilde{\gamma}}$, where $\mathbf{S}_{\tilde{\gamma}}$ is the photino spin. As for the quark current, by the Wigner-Eckardt theorem $\bar{Q}\gamma\gamma_5 Q$ may be replaced by $2\lambda\mathbf{J}$ where $\mathbf{J}$ is the nuclear spin and λ is an unknown constant that we will estimate shortly. The scattering amplitude is hence $\mathcal{M}=(4q^2/M_{\tilde{Q}}{}^2)\lambda\mathbf{S}_{\tilde{\gamma}}\cdot\mathbf{J}$. Averaging over initial spins and summing over final spins gives $\langle|\mathcal{M}|^2\rangle=(4q^4/M_{\tilde{Q}}{}^4)\lambda^2 J(J+1)$, where J is the magnitude of the nuclear spin.

It is straightforward to combine this expression with previous formulas to find that the event rate is

$$R=\frac{1.1\text{ events}}{\text{kg day}}\left[\frac{100\text{ GeV}}{M_{\tilde{Q}}}\right]^4\frac{4M_{\tilde{\gamma}}M_{\text{Nuc}}}{(M_{\tilde{\gamma}}+M_{\text{Nuc}})^2}\left[\frac{q}{\frac{2}{3}e}\right]^4[\lambda^2J(J+1)]\left[\frac{\rho}{10^{-24}\text{ g/cm}^3}\right]\left[\frac{\langle v\rangle}{200\text{ km/sec}}\right]. \tag{3}$$

Hence an event rate of order 1/kg day arises for $q=\frac{2}{3}e$, $M_{\tilde{Q}}=100$ GeV, or for $q=\frac{1}{3}e$, $M_{\tilde{Q}}=50$ GeV, assuming $\lambda^2J(J+1)\sim 1$.

To estimate the value of λ, we must calculate matrix elements of $\bar{u}\gamma\gamma_5 u$ and $\bar{d}\gamma\gamma_5 d$ for nuclear states. We first consider these matrix elements for individual nucleons, and then use the nuclear shell model to relate these to nuclear matrix elements.

The isovector matrix element is given by the relation

$$\langle p|\bar{u}\gamma\gamma_5 u-\bar{d}\gamma\gamma_5 d|p\rangle=2g_A\langle p|\mathbf{S}|p\rangle\,,$$

where $\mathbf{S}$ is the spin operator and experimentally $g_A\simeq 1.2$. For the isosinglet combination we rely on the quark model, since there seems to be no experimental information. In the quark model, the isosinglet combination is simply twice the spin operator,

$$\langle p|\bar{u}\gamma\gamma_5 u+\bar{d}\gamma\gamma_5 d|p\rangle=2\langle p|\mathbf{S}|p\rangle\,.$$

With analogous results for neutrons, we have

$$\begin{aligned}
&\langle p|\bar{u}\gamma\gamma_5 u|p\rangle=(1+g_A)\langle p|\mathbf{S}|p\rangle\,,\\
&\langle p|\bar{d}\gamma\gamma_5 d|p\rangle=(1-g_A)\langle p|\mathbf{S}|p\rangle\,,\\
&\langle n|\bar{u}\gamma\gamma_5 u|n\rangle=(1-g_A)\langle n|\mathbf{S}|n\rangle\,,\\
&\langle n|\bar{d}\gamma\gamma_5 d|n\rangle=(1+g_A)\langle n|\mathbf{S}|n\rangle\,.
\end{aligned} \tag{4}$$

Spin-dependent interactions can only occur for nuclei with nonzero spin; this usually arises (in the ground state) only for odd-even nuclei. To find the ground-state nuclear matrix elements we use the nuclear shell model and describe the ground state in terms of a single extra nucleon or nucleon hole with definite L, S, and J quantum numbers in a spherically symmetric background nuclear potential. Empirically, this description is good only near filled shells or "magic numbers."

From Eq. (4) we expect the rate corresponding to scalar up-quark exchange with odd-Z nuclei or down-quark exchange with odd-N nuclei to be roughly 100 times greater than that for scalar down-quark exchange with odd-Z nuclei or up-quark exchange with odd-N nuclei. $\lambda^2J(J+1)$ is of order 1 in the most favorable cases [see Table III for the values of $\lambda^2J(J+1)$ in some cases of interest]. A complicated variation in the interaction rate as a function of Z and N, and vanishing interaction rate for spinless nuclei, is the signal for spin-dependent interactions.

Now let us discuss how Eq. (3) is modified under different assumptions about the scalar quark masses. If the left- and right-handed scalar quarks have different masses $M_{\tilde{Q}_L}$ and $M_{\tilde{Q}_R}$ (but mixing between them is unimportant), the factor of $(1/M_{\tilde{Q}})^4$ in (3) is simply replaced by

$$\frac{1}{4}\left[\frac{1}{M_{\tilde{Q}_L}{}^2}+\frac{1}{M_{\tilde{Q}_R}{}^2}\right]^2.$$

If mixing between $\tilde{Q}_L$ and $\tilde{Q}_R$ is important then the coherent interactions given by Fig. 1(b) will dominate. If the scalar-quark-mass eigenstates are $\tilde{Q}_1=(\cos\beta\tilde{Q}_L+\sin\beta\tilde{Q}_R)$, $\tilde{Q}_2=(-\sin\beta\tilde{Q}_L+\cos\beta\tilde{Q}_R)$ with masses M_1, M_2, then Fig. 1(b) gives the amplitude

$$\mathcal{M}=-\frac{q^2}{4}\sin 2\beta\left[\frac{M_1{}^2-M_2{}^2}{M_1{}^2M_2{}^2}\right]\bar{\tilde{\gamma}}\tilde{\gamma}\bar{Q}Q\,.$$

Let $\mathcal{N}=\langle\text{Nuc}|\bar{Q}Q|\text{Nuc}\rangle$ be the expectation value of $\bar{Q}Q$ in the target nucleus; according to the simple quark model, $\mathcal{N}=2N+Z$ (or $\mathcal{N}=2Z+N$) if Q is a d (or u) quark. The event rate turns out to be

TABLE III. Estimates of $\lambda^2J(J+1)$ for some isotopes. For each isotope, two values are given, depending on whether the scalar quark considered is $\tilde{u}$ or $\tilde{d}$. The last column gives the shell-model description assumed for the nucleus in question, and an estimate of the reliability of the shell model applied to this particular nucleus.

Isotope	$\tilde{u}$	$\tilde{d}$	Shell-model input (quality)
^{27}A	0.42	0.0035	$D_{5/2}$ proton hole (good)
^{29}Si	0.0075	0.91	$S_{1/2}$ neutron (good)
^{69}Ga, ^{71}Ga	0.50	0.0042	$P_{3/2}$ proton (fair)
^{73}Ge	0.0031	0.37	$G_{9/2}$ neutron (good)
^{111}Cd, ^{113}Cd, ^{115}Sn, ^{117}Sn, ^{119}Sn	0.0075	0.91	$S_{1/2}$ neutron (poor)
^{139}Ln	0.24	0.0019	$G_{7/2}$ proton (poor)
^{207}Pb	0.0008	0.10	$P_{1/2}$ neutron hole (excellent)

$$R=\frac{700\text{ events}}{\text{kg day}}\left[\frac{\mathcal{N}}{100}\right]^2\sin^2 2\beta\left[\frac{(100\text{ GeV})^4(M_1{}^2-M_2{}^2)^2}{M_1{}^4M_2{}^4}\right]\left[\frac{q}{\frac{2}{3}e}\right]^4\frac{4M_{\tilde{\gamma}}M_{\text{Nuc}}}{(M_{\tilde{\gamma}}+M_{\text{Nuc}})^2}\left[\frac{\rho}{10^{-24}\text{ g/cm}^3}\right]\left[\frac{\langle v\rangle}{200\text{ km/sec}}\right]. \tag{5}$$

Because of the coherence (the factor of $\mathcal{N}^2$), this effect can dominate the spin-dependent interaction by several orders of magnitude if it is present.

Aside from the detector proposed in Ref. 5, an interesting possibility is to detect dark-matter particles via inelastic rather than elastic scattering from nuclei. For instance, ^{169}Tm has a $\frac{1}{2}^+$ ground state and a $\frac{3}{2}^+$ excitation at 8.4 keV. A dark-matter particle with $m\gtrsim 40$ GeV has enough kinetic energy to excite this transition. The excitation could readily be excited by particles like photinos with spin-dependent interactions. The signal would be the 8.4-keV x-ray photon from decay of the excited state, emitted with a lifetime $t_{1/2}=4.0\times10^{-9}$ sec. Attenuation of photons at this energy might make this difficult to observe, however.

To estimate the cross section for inelastic scattering is a difficult problem in nuclear physics. Apart from phase space, it involves the ratio

$$R=\langle\tilde{N}\,|\,\bar{Q}\gamma_i\gamma_5 Q\,|\,N\rangle/\langle N\,|\,\bar{Q}\gamma_i\gamma_5 Q\,|\,N\rangle\ ,$$

where $|\,N\rangle$ and $|\,\tilde{N}\rangle$ are the nuclear ground state and excited state. To get a rough idea of R, we may look at the experimental value of the reduced matrix element M for the decay $\tilde{N}\to N+\gamma$, defined by

$$\Gamma_{fi}=\frac{4\alpha k^3}{3m_p{}^2}\,|\,M\,|^2\ .$$

Insofar as the quark spin, which appears in $\bar{Q}\gamma_i\gamma_5 Q$, enters the magnitude dipole transition $\tilde{N}\to N+\gamma$, M and R may be similar. For ^{169}Tm, $|\,M\,|^2=1.65$, suggesting that the elastic and inelastic cross sections may be comparable. In Table IV, we list some examples of nuclear excitations with relatively low excitation energy and large values of M.

Inelastic scattering may eventually be an interesting approach to studying particles with spin-dependent interactions. It is less likely to be interesting for particles with spin-independent interactions since the inelastic cross sections are in that case several orders of magnitude smaller due to lack of coherence.

TABLE IV. Some inelastic transitions and their magnetic dipole matrix elements. Relatively favorable cases (low excitation energy and high matrix element) are ^{169}Tm and ^{187}Os.

Isotope	J^P Ground state	J^P Excited state	Excitation energy (keV)	$\lvert M\rvert^2$
^{155}Gd	$\frac{3}{2}^-$	$\frac{5}{2}^-$	60.02	0.10
^{157}Gd	$\frac{3}{2}^-$	$\frac{5}{2}^-$	54.54	0.02
^{159}Tb	$\frac{3}{2}^+$	$\frac{5}{2}^+$	57.99	0.12
^{165}Ho	$\frac{7}{2}^-$	$\frac{9}{2}^-$	94.7	0.13
^{169}Tm	$\frac{1}{2}^+$	$\frac{3}{2}^+$	8.4	1.65
^{173}Yn	$\frac{5}{2}^-$	$\frac{7}{2}^-$	7.9	2.00
^{183}W	$\frac{1}{2}^-$	$\frac{3}{2}^-$	46.484	0.31
^{187}Os	$\frac{1}{2}^-$	$\frac{3}{2}^-$	9.8	2.49

(iii) Strongly interacting particles

There are many possible candidates for a dark-matter particle X that would have strong interactions. Examples would be the H particle[12] if it is stable, or a bound state of ordinary quarks and gluons with a heavy stable quark, scalar quark, gluino, or colored technibaryon. There might be more exotic possibilities as well.

One limit on X comes from experiments[13] looking for exotic, heavy nuclear isotopes corresponding to bound states of X with ordinary nuclei. Failure to find such isotopes suggests that, if X is cosmologically abundant, its low-energy interaction with ordinary nuclei is repulsive. We assume X interacts with nuclei at low energies via a square-well potential of radius R (probably a repulsive square well for the reason just noted). The low-energy limit of the cross section is $\sigma=4\pi R^2$. The nuclear radius R is approximately $R=1.25A^{1/3}\times10^{-13}$ cm, so $\sigma\sim2.0A^{2/3}\times10^{-25}$ cm^2. With this cross section, it can be seen that X particles with $v\sim200$ km/sec are stopped by the earth if $M_X\lesssim10^7$ GeV. They are largely stopped by the atmosphere if $M_X\lesssim100$ GeV. (In making these estimates, one must note that the X particle is significantly slowed only after colliding with a mass of nuclei comparable to its own mass.) The counting rate for the detector in Ref. 5 is

$$R=\frac{1.15\times10^{14}\text{ events}}{\text{kg day}}A^{-1/3}\epsilon\left[\frac{1\text{ GeV}}{M_X}\right]\times\left[\frac{\rho}{10^{-24}\text{ g/cm}^3}\right]\left[\frac{v}{200\text{ km/sec}}\right]. \tag{6}$$

Here ϵ is a parameter that measures the stopping power of the earth or the atmosphere. For $M_X>10^7$ GeV, $\epsilon=1$; for $10^7\text{ GeV}>M_X>10^2$ GeV, $\epsilon\simeq\frac{1}{2}$ due to depletion of upward moving particles; for $M_X<10^2$ GeV, $\epsilon\ll1$. For detectors such as ^{27}Al or ^{207}Pb, $\frac{1}{3}\geq A^{-1/3}\geq\frac{1}{6}$, so if a counting rate of 1/kg day is detectable, then strongly interacting dark-matter candidates are detectable at masses up to 2–4×10^{13} GeV. At masses much less than 10^{13} GeV, counting rates would be so large that even a prototype of the detector envisaged in Ref. 5 would probably place useful bounds.

Strongly interacting particles, with their observable mean-free path (~20 cm) and low velocity ($\beta\lesssim10^{-3}$) offer good opportunities for background rejection which might make more conventional detection schemes feasible.

A distinctive signal in a NaI crystal would be a pair of events with energy deposit ~ 10 keV (~ 10 photons detected) separated by ~ 20 cm and by ~ 1 μsec. Together with anticoincidence shielding for cosmic rays, one could probably achieve sensitivities comparable to the detector in Ref. 5, and perhaps significantly better. In fact, a large range of masses for strongly interacting dark-matter particles is probably already ruled out by the simple observation that NaI does not "glow in the dark." The crystal ball experiment, for example, probably rules out masses 10^2 GeV $< M_X < 10^6$ GeV.[14]

For values of $M_X \lesssim 100$ GeV, the stopping power of the atmosphere becomes important, and the X particle probably could not be detected at sea level. Their kinetic energy would be too low to excite the detector in Ref. 5 or a NaI crystal. One way to circumvent this problem would be to put a detector high in the atmosphere with a balloon, or on a satellite above the atmosphere. One would then have to be careful to avoid being swamped by primary cosmic rays.

Another interesting possibility is to detect strongly interacting particles by calorimetry. It can be estimated that in liquid helium, strongly interacting halo particles deposit energy at a rate

$$J = \frac{2.5\times 10^{-7}\ \mathrm{W}}{\mathrm{kg}}\epsilon \left[\frac{1\ \mathrm{GeV}}{M_X}\right]\left[\frac{\rho}{10^{-24}\ \mathrm{g/cm^3}}\right] \times \left[\frac{v}{200\ \mathrm{km/sec}}\right]. \quad (7)$$

Heat leaks as low as 10^{-10} W/kg have been demonstrated.[15] At sea level, it may be possible to bound strongly interacting particles in a mass range near 10^2–10^3 GeV, but this range is probably already ruled out. Satellite-based calorimetry might also be feasible for $M_X \lesssim 10^2$ GeV, but the heat flux from cosmic rays would be comparable to that from dark-matter particles, and would be difficult to sort out.

Apart from the detector of Ref. 5, another idea for an experiment that might be sensitive to the particles discussed in this paper has recently been suggested.[16]

We would like to thank R. S. Raghavan for pointing out the relevance of Ref. 5, and to thank K. DeConde, D. Osheroff, P. J. E. Peebles, and D. Marlow for discussions. This work was supported in part by National Science Foundation Grant No. PHY-80-19754.

[1]S. M. Faber and J. S. Gallagher, in *Annual Reviews of Astromony and Astrophysics,* edited by G. Burbridge, D. Layzer, and J. G. Phillips (Annual Reviews, Palo Alto, California, 1979).

[2]See, for example, B. Cabrera, Phys. Rev. Lett. **48**, 1378 (1982); S. Errede *et al., ibid.* **51**, 1933 (1983); F. Kajino *et al., ibid.* **52**, 1373 (1984).

[3]P. Sikivie, Phys. Rev. Lett. **51**, 1415 (1983).

[4]A. de Rújula and S. Glashow, Phys. Rev. Lett. **45**, 942 (1980); E. Witten, in *Neutrino Oscillation Workshop,* edited by N. Baggett (Brookhaven National Laboratory, Upton, N.Y., 1981).

[5]A. Drukier and L. Stodolsky, Phys. Rev. D **30**, 2295 (1985).

[6]D. Z. Freedman, D. N. Schramm, and D. L. Tubbs, Annu. Rev. Nucl. Sci. **27**, 167 (1977).

[7]A. K. Drukier, Nucl. Instrum. Methods **201**, 77 (1981), and references therein.

[8]L. E. Ibañez, Phys. Lett. **137B**, 160 (1984); J. S. Hagelin, G. L. Kane, and S. Raby, Nucl. Phys. **B241**, 638 (1984).

[9]N. Cabibbo, G. Farrar, and L. Maiani, Phys. Lett. **105B**, 155 (1981).

[10]J. Silk and M. Srednicki, Phys. Rev. Lett. **53**, 624 (1984).

[11]H. E. Haber and G. L. Kane, Phys. Rep. **117**, 75 (1985).

[12]R. L. Jaffe, Phys. Lett. **38**, 195 (1977).

[13]P. F. Smith and J. R. J. Bennett, Nucl. Phys. **B149**, 525 (1979).

[14]D. Marlow (private communication).

[15]F. Pobell, Physica **109&110B**, 1485 (1982).

[16]B. Cabrera, L. Krauss, and F. Wilczek, University of California at Santa Barbara Report, 1984 (unpublished).

PHYSICAL REVIEW D

PARTICLES AND FIELDS

THIRD SERIES, VOLUME 33, NUMBER 8 15 APRIL 1986

Possibility of detecting heavy neutral fermions in the Galaxy

Ira Wasserman

Center for Radiophysics and Space Research, Cornell University, Ithaca, New York 14853

(Received 22 May 1985; revised manuscript received 31 October 1985)

It is shown that heavy neutral fermions in the galactic halo could produce numerous detectable low-energy events in a "thermal" neutrino detector if the heavy-fermion–nucleon vector coupling is comparable in strength to the vector weak interaction. The conditions under which a detectable event rate could arise for fermions with purely axial-vector couplings are also discussed. In a silicon detector heavy-fermion events would be concentrated at low energies, and could be distinguished from solar or supernova neutrino events, which are expected to have a flatter energy spectrum. A measurement of the energy spectrum of heavy-fermion events could lead to a determination of the fermion mass, subject to astrophysical uncertainties concerning the velocity distribution of halo particles.

Recently, a new type of neutrino detector, which relies on the idea that even small neutrino energy losses ($\Delta E_\nu \gtrsim 1$ keV) in cold material ($T \sim 1-10$ mK) with a small specific heat could produce measurable temperature changes, has been proposed.[1] In the proposed detection scheme, the recoil energy from neutrino-electron or neutrino-nucleus scattering, which is rapidly thermalized, is responsible for heating the detector. Cabrera, Krauss, and Wilczek[1] have estimated that solar *pp* or Be^7 neutrinos could produce measurable events at a rate ~ 1/ton day in a Si detector, and that a burst of supernova neutrinos would produce ~ 10 simultaneous events in a 10-ton detector.

The purpose of this paper is to examine the possibility that such a detector can be used to observe heavy neutral fermions ($m \gtrsim 1$ GeV) in the Galaxy. Such particles, it has been suggested, could be a substantial component of the cosmological missing mass,[2,3] and would be expected to condense gravitationally, in particular, into galactic halos.[4] We will assume here that the extended massive halo believed to surround the Galaxy[5] is dominated by heavy fermions. The detection rates estimated here will therefore be upper bounds, with strict equalities pertaining only to the astrophysically interesting but possibly optimistic scenario in which heavy fermions are a significant constituent of the galactic halo. Similar calculations have been carried out independently by Goodman and Witten,[6] and Drukier, Freese, and Spergel.[6]

Heavy halo fermions, with typical momenta

$$p_{\rm rms} = m v_h = 1\ \text{MeV}\, m(\text{GeV}) \left[\frac{v_h}{10^{-3}c}\right] = (2\times 10^{-11}\ \text{cm})^{-1} m(\text{GeV}) \left[\frac{v_h}{10^{-3}c}\right] \quad (1)$$

for a halo velocity dispersion v_h, may interact coherently with detector nuclei of mass number A provided

$$m \lesssim 130\, A^{-1/3} \left[\frac{v_h}{10^{-3}c}\right]^{-1} \text{GeV} . \quad (2)$$

Assuming Eq. (2) to be satisfied, as must be the case if heavy fermions are to contribute significantly to the cosmological mass density,[2,3] coherent scattering with detector nuclei will be the dominant energy-loss mechanism.

We shall adopt a largely phenomenological description of the heavy-fermion–nucleus interaction. We write the heavy-fermion–nucleon interaction Lagrangian (in the point-interaction limit) as ($N = p$ for protons, n for neutrons)

$$\mathcal{L}_{f-N} = \sqrt{2}\, G_f J_f^\mu J_{N\mu} , \quad (3)$$

where the heavy-fermion current is

$$J_f^\mu = \bar{\psi}_{f,L} \gamma^\mu \psi_{f,L} , \quad (4)$$

©1986 The American Physical Society

for a left-handed Dirac or Majorana heavy fermion field $\psi_{f,L}$. Ignoring second-class currents we get,[7-9] for small momentum transfers,

$$\langle N' \mid J_N^\mu \mid N \rangle = \langle N' \mid J_N^\mu \mid N \rangle_V + \langle N' \mid J_N^\mu \mid N \rangle_A \quad (5)$$

for the matrix element of the nucleon current between states $\mid N \rangle$ and $\mid N' \rangle$, where in the nonrelativistic limit the vector (V) and axial-vector (A) components reduce to

$$\langle N' \mid J_N^0 \mid N \rangle_V = \frac{-i}{(2\pi)^3} u_{N'}^\dagger [g_V^{(N)}(0) + f_V^{(N)}(0)] u_N \quad (6)$$

and

$$\langle N' \mid \mathbf{J}_N \mid N \rangle_A = \frac{1}{(2\pi)^3} u_{N'}^\dagger [g_A^{(N)}(0) \sigma] u_N \quad (7)$$

to leading order in $v/c \ll 1$. Equations (5)–(7) are of course identical to what is found in weak-interaction theory,[8,9] with the exception that here we regard the zero-momentum-transfer form factors $g_V^{(N)}(0)$, $f_V^{(N)}(0)$, and $g_A^{(N)}(0)$ as phenomenological parameters with unknown values (but presumably ~ 1 in general).

For simplicity, let us consider the interaction of a heavy fermion with an atomic nucleus, such as $_{14}Si^{28}$, with total angular momentum $J=0$. For this case the axial-vector interaction, which transforms as a vector under rotations, makes no contribution to the nuclear matrix element. For coherent scattering it is then straightforward to show that[10]

$$d\sigma_V^{(D)} = \frac{G_f^2}{4\pi} [\tilde{g}_V^{(p)} Z + \tilde{g}_V^{(n)} (A-Z)]^2 m_R^2 \frac{dT}{T_{\max}} \quad (8a)$$

for Dirac fermions, and

$$d\sigma_V^{(M)} = \frac{G_f^2}{\pi} [\tilde{g}_V^{(p)} Z + \tilde{g}_V^{(n)} (A-Z)]^2 \times m_R^2 v^2 \left[1 - \frac{T}{T_{\max}}\right] \frac{dT}{T_{\max}} \quad (8b)$$

for Majorana fermions, where T is the nuclear recoil energy, whose maximum value is

$$T_{\max} = \frac{4mM_A}{(m+M_A)^2} E_f = \frac{2m^2 M_A}{(m+M_A)^2} v^2 \quad (9)$$

for an incident fermion kinetic energy $E_f = \frac{1}{2} m v^2$ and nuclear mass M_A. In Eq. (8), Z is the nuclear charge and A the total nucleon number,

$$\tilde{g}_V^{(N)} \equiv g_V^{(N)}(0) + f_V^{(N)}(0) \quad (N = n,p) \quad (10)$$

and

$$m_R = \frac{mM_A}{m+M_A} \quad (11)$$

is the reduced mass. Integrating over T gives the total scattering cross sections

$$\sigma_V^{(D)} = \frac{G_f^2}{4\pi} [\tilde{g}_V^{(p)} Z + \tilde{g}_V^{(n)} (A-Z)]^2 m_R^2 = 4.4 \times 10^{-39} [\tilde{g}_V^{(p)} Z + \tilde{g}_V^{(n)} (A-Z)]^2 \left[\frac{G_f}{G_F}\right]^2 \left[\frac{m_R}{1\ \text{GeV}}\right]^2 \text{cm}^2 \quad (12a)$$

and

$$\sigma_V^{(M)} = \frac{G_f^2}{2\pi} [\tilde{g}_V^{(p)} Z + \tilde{g}_V^{(n)} (A-Z)]^2 m_R^2 v^2 = 8.7 \times 10^{-45} [\tilde{g}_V^{(p)} Z + \tilde{g}_V^{(n)} (A-Z)]^2 \left[\frac{G_f}{G_F}\right]^2 \left[\frac{m_R}{1\ \text{GeV}}\right]^2 \left[\frac{v}{10^{-3}c}\right]^2 \text{cm}^2, \quad (12b)$$

where $G_F \approx (290\ \text{GeV})^{-2}$ is the Fermi weak coupling constant. For nominal values of the various parameters in Eq. (12), and $\tilde{g}_V^{(p)} \sim \tilde{g}_V^{(n)} \sim 1$, $\sigma_V^{(D)}$ is a good deal larger than, while $\sigma_V^{(M)}$ is comparable to, either neutrino-electron or neutrino-nucleus scattering cross sections for low-energy neutrinos ($E_\nu \sim 1$ MeV).

If $\tilde{g}_V^{(N)} = 0$ then the heavy-fermion–nucleon interaction is purely axial vector, and the cross section for (coherent) scattering of heavy fermions by a $J=0$ nucleus vanishes identically. This will be the case, for example, for photino-nucleus scattering, if the left- and right-handed scalar partners of the u and d quarks are equally massive.[11] For $\tilde{g}_V^{(N)} = 0$ but $g_A^{(N)} \neq 0$ we find, for scattering by a nucleus with $J \neq 0$,

$$d\sigma_A^{(D)} = \frac{G_f^2}{4\pi} \frac{|\mathcal{J}_A|^2}{2J+1} m_R^2 \frac{dT}{T_{\max}} \quad (13a)$$

for Dirac fermions, and

$$\frac{d\sigma_A^{(M)}}{dT} = 4 \frac{d\sigma_A^{(D)}}{dT} \quad (13b)$$

for Majorana fermions[12] where, using the Wigner-Eckart theorem[13]

$$\frac{\mathcal{J}_A}{\sqrt{2J+1}} = \left[\frac{J+1}{J}\right]^{1/2} \left(\left\langle J,J \left| g_A^{(p)}(0) \sum_p \sigma_z^{(p)} \right| J,J \right\rangle + \left\langle J,J \left| g_A^{(N)}(0) \sum_n \sigma_z^{(n)} \right| J,J \right\rangle \right). \quad (14)$$

For nuclei with a single additional nucleon outside other-

wise filled levels, $J=l\pm\frac{1}{2}$ and

$$\frac{\mathcal{J}_A}{\sqrt{2J+1}}=g_A^{(N)}(0)\left[\frac{J+1}{J}\right]^{1/2}\quad(J=l+\tfrac{1}{2})$$
$$=-g_A^{(N)}(0)\left[\frac{J}{J+1}\right]^{1/2}\quad(J=l-\tfrac{1}{2})\qquad(15)$$

for $N=n$ or p, while for nuclei with a single missing nucleon inside otherwise filled levels we get the same results multiplied by an overall factor -1 (Ref. 14). Integrating Eq. (13) over T gives the total cross sections

$$\sigma_A^{(D)}=\frac{G_f^2}{4\pi}\frac{|\mathcal{J}_A|^2}{2J+1}m_R^2$$
$$=4.4\times10^{-39}\frac{|\mathcal{J}_A|^2}{2J+1}\left[\frac{G_f}{G_F}\right]^2\left[\frac{m_R}{1\text{ GeV}}\right]^2\text{cm}^2\qquad(16a)$$

and

$$\sigma_A^{(M)}=4\sigma_A^{(D)}\qquad(16b)$$

which are much larger than low-energy neutrino-electron or neutrino-nucleus scattering cross sections unless G_f/G_F is appreciably smaller than unity.

In order to obtain estimated event rates for galactic fermions, we shall adopt a spherically symmetric isothermal model for the galactic halo.[15] The halo will be assumed to be nonrotating, so that the Solar System (and hence fermion detector) moves through the halo with a peculiar speed v_r, the rotation speed of the Sun about the galactic center. The event rate per detector nucleus for recoil energies between T and $T+dT$ is then $(dR/dT)dT$, where

$$\frac{dR}{dT}=\frac{(\rho_0/m)v_h}{\sqrt{2\pi}}\left[\frac{v_h}{v_r}\right]$$
$$\times\int_{v_{\min}(T)/v_h}^{\infty}du\,u^2\frac{d\sigma}{dT}(e^{-u_-^2/2}-e^{-u_+^2/2})\qquad(17)$$

with $d\sigma/dT$ a function of both recoil energy T and $u\equiv v/v_h$ in general [cf. Eqs. (8) and (13)], and $u_\pm\equiv u\pm v_r/v_h$. In Eq. (17) ρ_0 is the local mass density of the halo and

$$v_{\min}^2(T)=\frac{T(m+M_A)^2}{2m^2M_A}\,.\qquad(18)$$

The halo velocity dispersion is

$$v_h^2=\frac{GM_0/R_0}{(-d\ln\rho/d\ln r)_{r=R_0}}\,,\qquad(19)$$

where M_0 is the total halo mass within $r=R_0$, the solar (and hence detector) halo radius, and $-d\ln\rho/d\ln r=2$ for the singular isothermal sphere.[15] Recent determinations[16] are consistent with $\rho_0\approx0.01M_\odot$ pc^{-3} and $M_0\approx6\times10^{10}M_\odot$ at $R_0\approx8$ kpc, with

$$(-d\ln\rho/d\ln r)_{r=R_0}\approx2\,,$$

implying[17] $v_h\approx130$ km sec$^{-1}\approx0.43\times10^{-3}c$. The rotation speed will be taken to be[16] $v_r\approx250$ km sec^{-1}. Event rates will only be estimated here for these fiducial values of v_r and v_h. Other contributions to Earth's peculiar motion through the halo (for example, the Solar System's oscillatory movement perpendicular to the galactic disk and the orbital motion of Earth around the Sun) will not be taken into account in this paper.

Substituting Eq. (8) in Eq. (17) gives

$$\frac{dR_V^{(D)}}{dT}=\frac{G_f^2\rho_0(M_A/m)[\tilde{g}_V^{(p)}Z+\tilde{g}_V^{(n)}(A-Z)]^2}{2(2\pi)^{3/2}v_h}$$
$$\times F_V^{(D)}\left[\frac{v_{\min}(T)}{\sqrt{2}\,v_h},\frac{v_r}{\sqrt{2}\,v_h}\right],\qquad(20a)$$

for Dirac fermions where

$$F_V^{(D)}(x,y)=\frac{\sqrt{\pi}}{4y}[\text{erf}(x+y)-\text{erf}(x-y)]\qquad(20b)$$

and

$$\frac{dR_V^{(M)}}{dT}=\frac{4G_f^2\rho_0(M_A/m)v_h[\tilde{g}_V^{(p)}Z+\tilde{g}_V^{(n)}(A-Z)]^2}{(2\pi)^{3/2}}$$
$$\times F_V^{(M)}\left[\frac{v_{\min}(T)}{\sqrt{2}\,v_h},\frac{v_r}{\sqrt{2}\,v_h}\right],\qquad(20c)$$

for Majorana fermions where

$$F_V^{(M)}(x,y)$$
$$=\frac{1}{2y}\left[\frac{\sqrt{\pi}}{2}(\tfrac{1}{2}+y^2-x^2)[\text{erf}(x+y)-\text{erf}(x-y)]\right.$$
$$\left.+\left[\frac{x+y}{2}\right]e^{-(x-y)^2}-\left[\frac{x-y}{2}\right]e^{-(x+y)^2}\right],\qquad(20d)$$

with erf(z) the usual error function. For Dirac fermions integrating Eq. (20a) over T gives a total event rate Γ_V per unit detector mass M_D ($\Gamma_V/M_D=R_V/M_A$):

$$\frac{\Gamma_V^{(D)}}{M_D}=\frac{2G_f^2\rho_0v_h[\tilde{g}_V^{(p)}Z+\tilde{g}_V^{(n)}(A-Z)]^2}{(2\pi)^{3/2}}\frac{m/M_A}{(1+m/M_A)^2}G_V^{(D)}\left[\frac{v_r}{\sqrt{2}v_h}\right]$$
$$=2.5\times10^3\text{ ton}^{-1}\text{ day}^{-1}\left[\frac{G_f}{G_F}\right]^2[\tilde{g}_V^{(p)}Z+\tilde{g}_V^{(n)}(A-Z)]^2\frac{m/M_A}{(1+m/M_A)^2}\,,\qquad(21a)$$

where

$$G_V^{(D)}(y)=\frac{1}{2y}[\sqrt{\pi}(\tfrac{1}{2}+y^2)\mathrm{erf}(y)+y\,e^{-y^2}]\approx 1.53 \tag{21b}$$

for $y=v_r/\sqrt{2}\,v_h\approx 1.36$. The corresponding event rate for Majorana fermions is

$$\Gamma_V^{(M)}=\frac{16G_f^2\rho_0 v_h^3[\tilde{g}_V^{(p)}Z+\tilde{g}_V^{(n)}(A-Z)]^2}{(2\pi)^{3/2}}G_V^{(M)}\left[\frac{v_r}{\sqrt{2}\,v_h}\right]$$
$$=8.0\times 10^{-3}\ \mathrm{ton}^{-1}\ \mathrm{day}^{-1}\left[\frac{G_f}{G_F}\right]^2[\tilde{g}_V^{(p)}Z+\tilde{g}_V^{(n)}(A-Z)]^2\frac{m/M_A}{(1+m/M_A)^2}\ , \tag{21c}$$

where

$$G_V^{(M)}(y)=\frac{1}{4y}[\sqrt{\pi}(y^4+3y^2+\tfrac{3}{4})\mathrm{erf}(y)+y(y^2+\tfrac{5}{2})e^{-y^2}]\approx 3.16\ . \tag{21d}$$

The average energy deposited per Dirac fermion event is

$$\langle T\rangle_V^{(D)}=\frac{4m^2M_A v_h^2}{(m+M_A)^2}H_V^{(D)}\left[\frac{v_r}{\sqrt{2}\,v_h}\right]=1.6\ \mathrm{keV}\left[\frac{m}{1\ \mathrm{GeV}}\right]\frac{m/M_A}{(1+m/M_A)^2}\ , \tag{22a}$$

where

$$H_V^{(D)}(y)=\frac{(y^4+3y^2+\frac{3}{4})\mathrm{erf}(y)+(y/\sqrt{\pi})(y^2+\frac{5}{2})e^{-y^2}}{2[(y^2+\frac{1}{2})\mathrm{erf}(y)+(y/\sqrt{\pi})e^{-y^2}]}\approx 2.07\ , \tag{22b}$$

while the average energy deposited per Majorana fermion event is

$$\langle T\rangle_V^{(M)}=\frac{4m^2M_A v_h^2}{(m+M_A)^2}H_V^{(M)}\left[\frac{v_r}{\sqrt{2}\,v_h}\right]=1.4\ \mathrm{keV}\left[\frac{m}{1\ \mathrm{GeV}}\right]\frac{m/M_A}{(1+m/M_A)^2}\ , \tag{22c}$$

where

$$H_V^{(M)}(y)=\frac{\sqrt{\pi}(y^6+15y^4/2+45y^2/4+\frac{15}{8})\mathrm{erf}(y)+(y^5+7y^3+33y/4)e^{-y^2}}{3[\sqrt{\pi}(y^4+3y^2+\frac{3}{4})\mathrm{erf}(y)+(y^3+5y/2)e^{-y^2}]}\approx 1.88\ . \tag{22d}$$

Integrated event rates at recoil energies $T\geq T_0$ are given in the Appendix. For Dirac fermions Eq. (A2) implies about 5×10^2 events with $T\geq 1$ keV per day in 1 ton of silicon if $G_f=G_F$, $\tilde{g}_V^{(p)}=\tilde{g}_V^{(n)}=1$, and $m=2$ GeV, and about 1×10^5 events per day if instead $m=5$ GeV, while Eq. (21a) implies total event rates, integrated over all recoil energies, equal to roughly 10^5 and 3×10^5 $\mathrm{ton}^{-1}\mathrm{day}^{-1}$, respectively. For Majorana fermions, Eq. (A3) implies about 4×10^{-3} events with $T\geq 1$ keV per day in 1 ton of silicon if $G_f=G_F$, $\tilde{g}_V^{(p)}=\tilde{g}_V^{(n)}=1$, and $m=2$ GeV, and about 4×10^{-1} events per day if instead $m=5$ GeV, while Eq. (21c) implies total event rates of about 4×10^{-1} $\mathrm{ton}^{-1}\mathrm{day}^{-1}$ and 8×10^{-1} ton day^{-1}, respectively. The event rate at $T\geq 1$ keV in a Si^{28} detector reaches its largest value, at[18] $m\approx 30$ GeV. If $G_f=G_F$ and $\tilde{g}_V^{(p)}=\tilde{g}_V^{(n)}=1$, the peak event rate is roughly 5×10^5 $\mathrm{ton}^{-1}\mathrm{day}^{-1}$ for Dirac fermions and 1.5 $\mathrm{ton}^{-1}\mathrm{day}^{-1}$ for Majorana fermions. (Since the average energy deposited is roughly 10 keV for Majorana or Dirac fermions with $m\approx 30$ GeV, the total event rate is essentially the same as the event rate at $T\geq 1$ keV.) Smaller event rates would of course be recorded if $G_f<G_F$. For example, taking $\tilde{g}_V^{(p)}=\tilde{g}_V^{(n)}=1$, Eq. (A2) implies $\lesssim 1$ event per day at $T\geq 1$ keV for Dirac fermions with $m\gtrsim 5$ GeV if $G_f\lesssim 10^{-3}G_F$, corresponding to $M_B\gtrsim 3000$ GeV for $G_f\equiv(e/M_B)^2$.

For heavy fermions whose coupling to nucleons are purely axial vector, the event rate in a pure Si^{28} detector would vanish identically. However, naturally occurring silicon consists of 4.7% Si^{29}, which has one neutron outside otherwise filled shells, implying a nonzero scattering cross section.[19] Substituting Eq. (13) into Eq. (17) gives, for Dirac fermions,

$$\frac{dR_A^{(D)}}{dT}=\frac{G_f^2\rho_0(M_A/m)}{2(2\pi)^{3/2}v_h}\frac{|\mathcal{J}_A|^2}{2J+1}F_A\left[\frac{v_{\min}(T)}{\sqrt{2}\,v_h},\frac{v_r}{\sqrt{2}\,v_h}\right], \tag{23a}$$

and

$$\frac{dR_A^{(M)}}{dT}=4\frac{dR_A^{(D)}}{dT}\ , \tag{23b}$$

where

$$F_A(x,y)=F_V^{(D)}(x,y)\ . \tag{23c}$$

Integrating Eq. (23) over T gives a total event rate per unit detector mass

$$\frac{\Gamma_A^{(D)}}{M_D}=\frac{2G_f^2\rho_0 v_h}{(2\pi)^{3/2}}\frac{f_J|\mathcal{J}_A|^2}{2J+1}\frac{m/M_A}{(1+m/M_A)^2}G_A\left[\frac{v_r}{\sqrt{2}v_h}\right]=2.5\times10^2\ \text{ton}^{-1}\text{day}^{-1}\left[\frac{G_f}{G_F}\right]^2\left[\frac{f_J}{0.1}\right]\frac{|\mathcal{J}_A|^2}{2J+1}\frac{m/M_A}{(1+m/M_A)^2}, \tag{24a}$$

for Dirac fermions, and $\Gamma_A^{(M)}/M_D=4\Gamma_A^{(D)}/M_D$ for Majorana fermions, where

$$G_A(y)=G_V^{(D)}(y)\approx1.53\ , \tag{24b}$$

and f_J is the fraction of the detector mass in $J\neq0$ nuclei. The average energy deposited per event is

$$\langle T\rangle_A=\frac{19m^2M_Av_h^2}{3(m+M_A)^2}H_A\left[\frac{v_r}{\sqrt{2}v_h}\right]$$
$$=1.6\ \text{keV}\left[\frac{m}{1\ \text{GeV}}\right]\frac{m/M_A}{(1+m/M_A)^2}\ , \tag{25a}$$

for both Majorana and Dirac fermions where

$$H_A(y)=H_V^{(D)}(y)\approx2.07\ . \tag{25b}$$

Integrated event rates at recoil energies $T\geq T_0$ may be found using the Appendix. For Dirac fermions Eq. (A2) implies[20] about 2×10^{-2} events with $T\geq1$ keV per day in 1 ton of silicon if $g_A^{(n)}(0)=1$, $G_f=G_F$, and $m=2$ GeV, and about 2×10^1 events per day if instead $m=5$ GeV while Eq. (24) implies total event rates, integrated over all recoil energies, equal to roughly 2×10^1 and 5×10^1 $\text{ton}^{-1}\text{day}^{-1}$, respectively. The event rate at $T\geq1$ keV reaches its largest value at $m\approx30$ GeV. For Dirac fermions, the peak event rate is roughly 9×10^1 $\text{ton}^{-1}\text{day}^{-1}$ if $g_A^{(n)}(0)=1$, $G_f=G_F$, and $f_J=0.047$. (As before, the total event rates and event rates at $T\geq1$ keV are essentially the same for $m\approx30$ GeV.) Event rates for Majorana fermions are, from Eq. (23b), a factor of 4 larger. All of the event rates estimated above could be considerably larger for $G_f\gg G_F$, and vice versa.

While the main source of uncertainty in the rates estimated above is the largely unconstrained microphysics of the heavy fermion interactions with nucleons, it is important to bear in mind some astrophysical ambiguities as well. The main difficulty arises from the essentially unknown velocity distribution of the halo particles, which we have taken to be a continuous Maxwell-Boltzmann distribution with a single velocity dispersion v_h. Even if our implicit assumption of a locally isotropic velocity distribution is correct, one would expect a sharp cutoff for speeds greater than the local escape speed $v_e(r_0)$. Using a truncated rather than continuous Maxwell-Boltzmann distribution would result in generally lower event rates. However, as we shall argue below, we do not expect that including a velocity cutoff will substantially alter our results, at least for moderately large fermion masses.

A precise value of $v_e(r_0)$ is not known because it depends critically on the extent of the flat portion of the galactic rotation curve.[21] [Indeed, for a truly unbounded isothermal sphere, $v_e(r)$ is infinite at all r, and even for a tidally limited isothermal profile, $v_e(r)/v_h$ may be very large.] If the galactic rotation curve is flat out to $r\gtrsim2r_0$, which is not implausible, then[21] $v_e(r_0)/v_r\gtrsim[2(\ln2+1)]^{1/2}$ or[22] $v_e(r_0)\gtrsim460$ km sec^{-1}. At low values of m, where $v_{\min}(T_0)/v_h\sim$ a few, heavy fermion events at $T\geq T_0$ are mainly due to particles with halo speeds near $v_{\min}(T_0)-v_r$, the lowest possible value. As long as $[v_{\min}(T_0)-v_r]^2$ is significantly smaller than $v_e^2(r_0)$ our estimated event rates should be reasonably accurate. For $m=2$ GeV, $[v_{\min}(1\ \text{keV})-v_r]^2\approx6.6v_h^2$, as opposed to $v_e^2(r_0)\gtrsim12.5v_h^2$, so this condition is satisfied.[23] At larger values of m, the event rate above 1 keV is due to a larger range of halo speeds. Roughly speaking we expect our estimated event rates to be adequate provided $v_{\min}^2(T_0)\lesssim v_e^2(r_0)+v_r^2$, which is the same as requiring that the typical halo kinetic energy needed to produce an event not exceed the minimum kinetic energy for escape. Based on this criterion, our calculated event rates should already be reliable at $m=5$ GeV [$v_{\min}^2(1\ \text{keV})\approx3.9v_h^2$ as opposed to $v_e^2(r_0)+v_r^2\gtrsim16.2v_h^2$] and should be very accurate at larger values of m [e.g., $v_{\min}^2\ (1\ \text{keV})\approx v_h^2$ for $m\approx10$ GeV]. All in all, the average event rates calculated above are probably reasonably accurate over the entire range of fermion masses considered ($m\gtrsim2$ GeV), and are certainly dependable estimates at $m\gtrsim5$ GeV.

Conservatively, then, we conclude that galactic heavy fermions would produce numerous detectable events in a one-ton thermal detector of the type described in Ref. 1 if the heavy neutral fermion couplings to nucleons are at least comparable in strength to the weak interaction and if the fermion mass $m\gtrsim2$–3 GeV. This would certainly be the case for sufficiently heavy neutral leptons, which may be either Majorana or Dirac fermions. For photinos, which are generally thought to be Majorana fermions,[3,11] vector couplings roughly as strong as the weak interaction could arise for left- and right-handed scalar-quark masses satisfying[24] $|1/m_L^2-1/m_R^2|\sim(100\ \text{GeV})^{-2}$, while far stronger axial-vector couplings, corresponding to $m_L\approx m_R\ll100$ GeV, cannot be ruled out.[3]

Heavy fermion events would be concentrated at low energies, and could be distinguished from low-energy neutrino events by their energy spectrum and their possibly large frequency of occurrence. The optimum observing strategy for simply detecting halo fermions would be to use a detector enriched in $J\neq0$ nuclei. Such a detector would be especially sensitive to Dirac heavy leptons, through the coherent vector interaction, but would also register substantial event rates for Majorana heavy leptons or photinos, for which the axial-vector coupling would dominate. Because the expected energy spectrum is essentially the same for Majorana and Dirac fermions in such a detector (ignoring the relatively small number of events due to the suppressed coherent vector coupling of Majorana fermions), one could not distinguish between these two

possibilities for the fermion mass eigenstates from observations with a single $J\neq 0$ nuclear species. For this purpose one could compare the results of observations made with $J\neq 0$ nuclei of different nuclear charges Z and nucleon members A, since only Dirac fermions scatter coherently to a significant degree. Alternatively, one could use a detector enriched in $J\neq 0$ nuclei to distinguish between Majorana and Dirac fermions, as the expected energy spectra [for example, Eqs. (20a), (20b) and (20c), (20d) for an isothermal halo model] are reasonably different for the two possibilities.

Assuming that the isothermal distribution accurately describes the galactic halo locally,[25] one could in principle determine the fermion mass m from a fit of Eqs. (20) and/or (23) to the observed energy spectrum of heavy fermion events. Moreover, such an analysis would yield independent determinations of v_h and v_r (more precisely, Earth's peculiar velocity relative to the halo), which would be of great interest astrophysically. [For a truncated isothermal distribution, one could also obtain information on $v_e(r_0)$, although the dependence of the predicted event rates on this parameter is weak at $m \gtrsim 2$–3 GeV.] The significance of the resulting derived parameters would be determined by the energy resolution of the data and by the appropriateness of the isothermal halo model.

This research was supported in part by National Science Foundation Grant No. AST 84-15162 and by the Alfred P. Sloan Foundation. It is a pleasure to acknowledge useful discussions with D. F. Chernoff, M. Harwit, J. Houck, E. E. Salpeter, and M. D. Weinberg.

APPENDIX: EVENT RATES AT $T \geq T_0$

The total event rate at recoil energies above a threshold value T_0 is

$$\frac{\Gamma(\geq T_0)}{M_D}=\frac{1}{M_A}\int_{T_0}^{\infty} dT \frac{dR}{dT}\ , \tag{A1}$$

where dR/dT is given by either Eqs. (20) or Eq. (23). Plugging Eqs. (20a) and (20b) into Eq. (A1) gives

$$\frac{\Gamma_V^{(D)}(\geq T_0)}{M_D}=\frac{\Gamma_{V,0}^{(D)}}{M_D}K_V^{(D)}\left[\frac{v_{\min}(T_0)}{\sqrt{2}\,v_h},\frac{v_r}{\sqrt{2}\,v_h}\right], \tag{A2a}$$

where

$$\frac{\Gamma_{V,0}^{(D)}}{M_D}=\frac{2G_f{}^2\rho_0 v_h[\tilde{g}_V^{(p)}Z+\tilde{g}_V^{(n)}(A-Z)]^2}{(2\pi)^{3/2}}\frac{m/M_A}{(1+m/M_A)^2} \tag{A2b}$$

and

$$K_V(x,y)=\frac{1}{2y}\left[\frac{\sqrt{\pi}}{2}(y^2-x^2+\tfrac{1}{2})[\mathrm{erf}(x+y)-\mathrm{erf}(x-y)]+\left[\frac{x+y}{2}\right]e^{-(x-y)^2}-\left[\frac{x-y}{2}\right]e^{-(x+y)^2}\right]. \tag{A2c}$$

Taking $x=0$ in Eq. (A2) gives Eqs. (21a) and (21b). Plugging Eqs. (20c) and (20d) into Eq. (A1) gives

$$\frac{\Gamma_V^{(M)}(\geq T_0)}{M_D}=\frac{\Gamma_{V,0}^{(M)}}{M_D}K_V^{(M)}\left[\frac{v_{\min}(T_0)}{\sqrt{2}\,v_h},\frac{v_r}{\sqrt{2}\,v_h}\right], \tag{A3a}$$

where

$$\frac{\Gamma_{V,0}^{(M)}}{M_D}=\frac{16G_f{}^2\rho_0 v_h{}^3}{(2\pi)^{3/2}}[\tilde{g}_V^{(p)}Z+\tilde{g}_V^{(n)}(A-Z)]^2\frac{m/M_A}{(1+m/M_A)^2} \tag{A3b}$$

and

$$K_V^{(M)}(x,y)=\frac{1}{8y}\{\sqrt{\pi}(y^4+3y^2+x^4-2x^2y^2-x^2+\tfrac{3}{4})[\mathrm{erf}(x+y)-\mathrm{erf}(x-y)]+[(x+y)(y^2-x^2+\tfrac{3}{2})+y]e^{-(x-y)^2}-[(x-y)(y^2-x^2+\tfrac{3}{2})-y]e^{-(x+y)^2}\}\ . \tag{A3c}$$

Taking $x=0$ in Eq. (A3) gives Eqs. (21c) and (21d). Plugging Eq. (23) into Eq. (A1) gives

$$\frac{\Gamma_A^{(D)}(\geq T_0)}{M_D}=\frac{\Gamma_{A,0}^{(D)}}{M_D}K_V^{(D)}\left[\frac{V_{\min}(T_0)}{\sqrt{2}\,v_h},\frac{v_r}{\sqrt{2}\,v_h}\right], \tag{A4a}$$

where

$$\frac{\Gamma_{A,0}^{(D)}}{M_D}=\frac{2G_f{}^2\rho_0 v_h}{(2\pi)^{3/2}}\frac{f_J|\mathcal{J}_A|^2}{2J+1} \tag{A4b}$$

and $K_V^{(D)}(x,y)$ is given by Eq. (A2b). Taking $x=0$ in Eq. (A4) gives Eq. (24).

[1]B. Cabrera, L. M. Krauss, and F. Wilczek, Phys. Rev. Lett. **55**, 25 (1985); A. Drukier and L. Stodolsky, Phys. Rev. D **30**, 2295 (1984).

[2]B. W. Lee and S. Weinberg, Phys. Rev. Lett. **39**, 165 (1977); D. A. Dicus, E. N. Kolb, and V. L. Teplitz, *ibid.* **39**, 168 (1977); J. E. Gunn, B. W. Lee, I. Lerche, D. Schramm, and G. Steigman, Astrophys. J. **233**, 1015 (1978).

[3]H. Goldberg, Phys. Rev. Lett. **50**, 1419 (1983); P. Binétry, G. Girandi, and P. Salati, Nucl. Phys. **B237**, 285 (1984); J. Ellis, J. S. Hagelin, D. V. Nanopoulos, K. Olive, and M. Srednicki, *ibid.* **B238**, 453 (1984); J. Silk and M. Srednicki, Phys. Rev. Lett. **53**, 624 (1984).

[4]J. R. Bond, A. S. Szalay, and M. S. Turner, Phys. Rev. Lett. **48**, 1636 (1982); P. J. E. Peebles, Astrophys. J. **258**, 415 (1982); J. R. Bond and A. S. Szalay, *ibid.* **274**, 443 (1983).

[5]J. P. Ostriker and P. J. E. Peebles, Astrophys. J. **186**, 467 (1973).

[6]M. W. Goodman and E. Witten, Phys. Rev. D **31**, 3059 (1985); A. Drukier, K. Freese, and D. Spergel, report, 1985 (unpublished).

[7]T. D. Lee, *Particle Physics and Introduction to Field Theory* (Harwood Academic, New York, 1981), pp. 607–612.

[8]S. Weinberg, Phys. Rev. D **5**, 1412 (1972).

[9]E. D. Commins and P. H. Bucksbaum, *Weak Interactions of Leptons and Quarks* (Cambridge University Press, Cambridge, 1983), Chap. 4.

[10]The derivation is analogous to the derivation of the coherent scattering cross section for (relativistic) neutrinos interacting with nuclei. See D. Z. Freedman, Phys. Rev. D **9**, 1389 (1974).

[11]P. Fayet, Phys. Lett. **86B**, 272 (1979); see also Ellis *et al.* (Ref. 3), and references therein.

[12]B. Kayser and R. E. Shrock, Phys. Lett. **112B**, 137 (1982).

[13]We are tacitly assuming that the nuclear shell model is an adequate approximation. The Wigner-Eckart theorem is discussed in general in K. Gottfried, *Quantum Mechanics* (Benjamin, New York, 1966), Sec. 36.3. Particular applications to the nuclear shell model are discussed in A. deShalit and H. Feshbach, *Theoretical Nuclear Physics,* Volume 1: Nuclear Structure (Wiley, New York, 1974), Appendix A.

[14]See deShalit and Feshbach (Ref. 13).

[15]See, for example, S. Chandrasekhar, *Principles of Stellar Dynamics* (Dover, New York, 1960), Sec. 5.8.

[16]J. N. Bahcall and R. M. Soneira, Astrophys. J. Suppl. **44**, 73 (1980).

[17]It is interesting to note that the velocity dispersions of various halo objects, although generally anisotropic, are in good agreement with the result of Eq. (19). See K. C. Freeman, in *The Milky Way Galaxy,* edited by H. van Woerden, R. J. Allen, and W. B. Burton (I.A.U. Symposium No. 106) (Reidel, Dordrecht, 1985). A recent study of the kinematics of 29 distant RR Lyrae stars in the halo is consistent with an isotropic velocity dispersion of 109 ± 16 km sec^{-1}. See A. Saha, Astrophys. J. **289**, 310 (1985). The best fit of a model with an isotropic velocity dispersion to the kinematics of 66 globular clusters gives $v_h \approx 118$ km sec^{-1} at $r=9$ kpc, but is not isothermal ($v_h \propto r^{0.2}$). See C. Frenk and S. D. M. White, Mon. Not. R. Astron. Soc. **193**, 295 (1980). These results indicate the range of uncertainty in galactic halo models, but support the value $v_h \approx 130$ km sec^{-1} adopted here, at least crudely.

[18]Such large values of m may, however, result in a very low present-day cosmological abundance of heavy fermions. See Refs. 2 and 3. Solar System constraints may also rule out Dirac neutrinos with $m \gtrsim 10$ GeV. See L. M. Krauss, M. Srednicki, and F. Wilczek, Report No. HUTP-85/A007, 1985 (unpublished); and K. Freese, CFA Report No. 2108, 1985 (unpublished).

[19]See, for example, H. A. Enge, *Introduction to Nuclear Physics* (Addison Wesley, Reading, 1966), Appendix 6.

[20]Although Si^{29} is a deformed nucleus, and may only be accurately understood in terms of an axisymmetric nuclear potential (Nilsson Hamiltonian) rather than a spherically symmetric shell model potential, we use the matrix element for its $2S_{1/2}$ parent state, $|\mathcal{J}_A|^2/(2J+1)=(J+1)/J=3$, in estimating event rates. That this may be reasonably accurate is indicated by the fact that even relatively large distortions of the $2S_{1/2}$ parent shell-model state lead to little change in the orbital energy. See Enge (Ref. 18) Sec. 6.8 and deShalit and Feshbach (Ref. 13) Sec. VI.10, and references therein.

[21]The escape velocity at radius r is given by $v_e^2(r) = -2[\phi(r)-\phi(r_T)]$, where r_T is the tidal radius of the halo, beyond which halo particles can no longer be considered bound to the Galaxy. In terms of the rotation velocity $v_r(r)$,

$$v_e^2(r)=2\int_r^{r_T} dr' v_r^2(r')/r' .$$

Assuming a flat rotation curve out to $r=\lambda r_0$, and $r_T \gg \lambda r_0$ gives $v_e^2(r_0)/v_r^2 > 2(\ln\lambda+1)$, since $v_r^2(r) > v_r^2(\lambda r_0/r)$ at $\lambda r_0 < r < r_T$. See Ref. 15.

[22]Observations of high-velocity stars in the solar neighborhood imply $v_e(r_0) \gtrsim 400$–450 km sec^{-1}. See D. Lynden-Bell and D. N. C. Lin, Mon. Not. R. Astron. Soc. **181**, 37 (1977), and references therein.

[23]Perhaps more relevant is a comparison of the Boltzmann factors $f_{\min}=\exp\{-[v_{\min}(T_0)-v_r]^2/2v_h^2\}$ and $f_e=\exp[-v_e^2(r_0)/2v_h^2]$. For $m=2$ GeV and $T_0=1$ keV, $f_{\min}/f_e \gtrsim 20$, indicating that our calculated event rate at 2 GeV is accurate to much better than a factor of 2 for our adopted galactic parameters. However, $f_{\min}/f_e \lesssim 1$ for $m \lesssim 1.7$ GeV, indicating that the results at low m are sensitive to the values of v_h, v_r, and $v_e(r_0)$ used. For larger masses, $m \gtrsim 5$ GeV, the dependences on these parameters are expected to be relatively weak, for reasons discussed in the text.

[24]Parity violation in nuclear interactions may constrain $|1/m_L^2-1/m_R^2| \lesssim (100\ \text{GeV})^{-2}$ unless u and d scalar quarks are themselves degenerate in mass. See M. Suzuki, Phys. Lett. **115B**, 40 (1982), and also J. Ellis and D. V. Nanopoulos, *ibid.* **110B**, 44 (1982) for a discussion of mass degeneracy among various flavors of scalar quarks. For photino masses $m \gtrsim 5$ GeV, the event rate due to vector coupling is still appreciable for $G_f/G_F \gtrsim 10^{-2}$, corresponding very roughly to $|1/m_L^2-1/m_R^2| \gtrsim (1000\ \text{GeV})^{-2}$ for u and d scalar quarks.

[25]It is perhaps worth noting that the local velocity distribution on laboratory space and time scales need not be even approximately Maxwellian even if an isothermal model adequately describes the large-scale properties of the galactic halo as a whole. This is because violent relaxation of collisionless halo particles leaves their fine-grained phase-space density unchanged even though it may cause the coarse-grained distribution to approach a (truncated) Maxwellian. The separation between the length scales on which the coarse and fine-grained distributions apply is not well understood theoretically. If this critical length scale is large compared to $v_h t_{\text{expt}} \approx 10^{15}$ cm t_{expt}(yr) for an experiment lasting a time t_{expt}, then the incident heavy fermions could even be monoenerget-

ic, in which case the observed energy spectrum would be directly proportional to the differential cross sections, Eqs. (8a), (8b), (13a), and (13b). It is, of course, true that the event rates and energy spectra in an isothermal halo would, over sufficiently long time scales, approach the expressions derived in this paper.

Cross sections, relic abundance, and detection rates for neutralino dark matter

Kim Griest
Astronomy and Astrophysics Center, Enrico Fermi Institute, The University of Chicago, Chicago, Illinois 60637
and NASA/Fermilab Astrophysics Center, Fermi National Accelerator Laboratory, Batavia, Illinois 60510
(Received 6 June 1988)

Neutralino annihilation and elastic scattering cross sections are derived which differ in important ways from previous work. These are applied to relic abundance calculations and to direct detection of neutralino dark matter from the galactic halo. Assuming the neutralino is the lightest supersymmetric particle and that it is less massive than the Z^0, we find relic densities of neutralinos greater than 4% of critical density for almost all values of the supersymmetric parameters. We constrain the parameter space by using results from the DESY e^+e^- storage ring PETRA (chargino mass less than 23 GeV) and the ASP detector at the SLAC storage ring PEP, and then assuming a critical density of neutralinos, display event rates in a cryogenic detector for a variety of models. A new term implies "spin-independent" elastic scattering even for these Majorana particles and inclusion of propagator momenta increases detection rates by 10–300% even for pure photinos. Z^0-squark interference leads to very low detection rates for some values of the parameters. The new term in the elastic cross section dominates for heavy, mostly spinless materials and mitigates the negative interference cancellations in light materials, except for the pure photino or pure Higgsino cases where it does not contribute. In general, the rates can be substantially different from the pure photino and pure Higgsino special cases usually considered.

I. INTRODUCTION

One of the most interesting unsolved problems in physics today is the identity of the dark matter (DM) known to exist in galactic halos. Several lines of reasoning lead one to hypothesize that the DM may not be ordinary "baryonic" material but rather may consist of some, as yet undiscovered, elementary particle. If so, it is likely that substantial quantities of these particles exist in our Galaxy's halo at the position of the Earth and it may be possible to detect these particles as they pass through detectors in laboratories on Earth.[1–4] Indirect methods, such as detecting the products of DM particle-antiparticle annihilation have also been suggested, the most promising of which use the enhancement in density which results from capture of DM particles into the body of the Sun or Earth.[5,6]

Crucial to all detection schemes are the DM particle cross sections. A zero elastic scattering cross section off ordinary matter would mean no possibility of direct direction and no capture of particles into the Earth or Sun. Particle-antiparticle annihilation cross sections determine, in part, the relic abundance of DM particles and also the flux of detectable annihilation products. Since there is little astrophysical data which bear on these cross sections we must turn to particle-physics models to make predictions. A wide spectrum of candidate particles have been proposed, but the most interesting are those which were originally proposed for nonastrophysical reasons, and only subsequently turned out to be suitable as DM candidates. The most popular particles in this class are the neutrino, the axion, and the lightest supersymmetric particle (LSP). In this paper we consider only the LSP.

Supersymmetry has been immensely popular among theorists for the past decade. It seems to be an essential ingredient of theories which unify gravity with the other interactions and in its low-energy manifestation provides an elegant solution to the hierarchy problem.[7] Supersymmetry has also been popular among experimentalists, but although much effort has been expended in the search, no supersymmetric particles have been discovered. A major goal of the next generation of particle accelerators will be to discover or set limits on supersymmetric particles. In the minimal supersymmetric extension of the standard model which we consider here, there is a new fermionic (bosonic) superpartner for every bosonic (fermionic) ordinary particle, so there is a doubling of the number of standard-model particles. The masses of these particles are undetermined in the most general scheme, but if supersymmetry is to solve the hierarchy problem, some superpartners must exist with masses below roughly a TeV. In most versions of these models there is also a conserved multiplicative quantum number $R = -1^{2S+3B+L}$. This R parity implies that the lightest superpartner is stable and therefore a DM candidate. The identity and cosmology of the LSP was considered in 1984 by Ellis *et al.*,[8] and since then, the lightest neutralino ($\tilde{\gamma}$), a linear combination of the photino, Z-ino, and two neutral Higgsinos, has been considered the likely candidate for LSP (Ref. 9). In considering detectability, most authors have, however, considered only the pure photino and pure Higgsino, two special cases of the general neutralino. If the neutralino is very light, then one might expect a reasonably pure photino or Higgsino, but there are no strong theoretical or experimental reasons to expect such a light LSP, and as the mass of the neutralino increases, a pure photino or Higgsino becomes more and more unlikely.

©1988 The American Physical Society

One of the most remarkable characteristics of neutralinos is that over most of the supersymmetric parameter space a relic density $\rho_{\tilde{\chi}}$ near critical density is obtained. We emphasize that a value of $\Omega_{\tilde{\chi}}$ (defined as $\rho_{\tilde{\chi}}/\rho_{\text{crit}}$) near 1, is therefore very natural, and this fact adds motivation to the search for neutralino dark matter (see also Ref. 8). In fact, even if neutralinos do not constitute the DM responsible for galactic rotation curves, we still expect $\Omega_{\tilde{\chi}} \geq 0.04$ in our galactic halo as long as they exist and are the LSP. So if low-energy supersymmetry does exist we expect to find a substantial fraction of the mass of the Universe in neutralinos and it is therefore worth making the effort to detect them.

In this paper we recalculate the cross sections for neutralinos and find several important differences with respect to earlier work.[8] Using these cross sections we show that direct detection event rates in cryogenic detectors can be very different from the pure photino and pure Higgsinos special cases. We also find that as the neutralino becomes more massive pure photinos and especially pure Higgsinos become very unlikely. Including a new term in the elastic scattering cross section and using a method due to Shifman, Vainshtein, and Zakharov[10] which includes the effects of heavy quark loops, we find a piece of the elastic cross section which is proportional to the mass of the scattering nucleus. This means neutralinos might be detectable with mostly spinless material such as mercury, germanium, or helium, even though Majorana particles are usually thought to have only "spin-dependent" interactions.

Inclusion of several new terms in the annihilation cross section, as well as a different sign interference term, changes relic abundance results for neutralinos, and the inclusion of previously ignored effects such as propagator momenta changes the results even for pure photinos. Finally, since the same couplings are involved in the elastic, annihilation, and production cross sections[11] we can extract information from various accelerator experiments. In particular, large areas of neutralino parameter space are ruled out by experiments such as UA1 (Ref. 12), the DESY storage ring PETRA (Ref. 13), and the ASP detector at the SLAC storage ring PEP (Ref. 14) and we find that neutralinos with mass less than about 5 GeV are unlikely.

The plan of this paper is as follows. In Sec. II we define the models under consideration and the parameter space to be explored. We then calculate the $\bar{\tilde{\chi}}\tilde{\chi}\bar{q}q$ matrix element (q is any quark or lepton) and the annihilation cross section ($\bar{\tilde{\chi}}\tilde{\chi} \to \bar{q}q$). For use in relic abundance calculations we also present the nonrelativistic expansion including the possibly important effect of propagator momenta. Exploiting crossing symmetry, we use the annihilation matrix element to find the elastic scattering ($\tilde{\chi}q \to \tilde{\chi}q$) and production ($e^+e^- \to \bar{\tilde{\chi}}\tilde{\chi}$) cross sections. For use in comparison with ASP results, we finally present the cross section $\sigma(e^+e^- \to \bar{\tilde{\chi}}\tilde{\chi}\gamma)$ in the soft-photon limit.

In Sec. III we rederive and extend some of the results from Sec. II using an effective-Lagrangian approach. In particular, we find the elastic scattering cross section of neutralinos off nuclei rather than neutralinos off quarks and show that due to heavy quark loops, a "spin-independent" interaction exists for the general neutralino. This interaction is proportional to the Z-ino component and so does not contribute to pure photino or pure Higgsino scattering. We also point out a difference (with respect to earlier work) of the sign of an interference term.

In Sec. IV we use the annihilation cross section to calculate the relic abundance of neutralinos and show that $\Omega_{\tilde{\chi}} \sim 1$ is very natural. We also show that apart from near the Z^0 pole a very low relic density of neutralinos is unlikely. We also describe and implement several consistency requirements such as the neutralino being lighter than the squarks and the charginos.

In Sec. V we use the elastic cross section to find the rate at which DM neutralinos interact with nuclei in a cryogenic detector. We show that the new (scalar) term can give enhancements of several orders of magnitude for very heavy nuclei. We also point out that with the standard "spin-dependent" (axial-vector) cross section very low event rates generically occur due to negative interference between Z^0 and squark exchange. These cancellations are substantially mitigated by the new scalar term even for light elements. We reduce parameter space throughout by requiring that $\Omega_{\tilde{\chi}} = 1$. We also mention that the inclusion of propagator momenta can be important, especially when the neutralino mass nears the squark mass. This is illustrated for the pure photino case where, for a neutralino mass of 30 GeV, the event rate with propagator momenta included is a factor of 2 higher than without. We emphasize that the event rates for the general neutralino can be quite different from the pure photino and pure Higgsino special cases, and that since detection should be aimed at the general neutralino, all terms in the cross section should be taken into account when selecting detector material.

In Sec. VI we show how accelerator experiments can compliment dark-matter searches. In particular, the PETRA result[13] that $m_{\tilde{\chi}^\pm} < 23$ GeV eliminates large regions of parameter space and as these limits improve (or the chargino is detected) the chargino constraint will become more and more important in deciding the detectability of neutralino DM. In addition, the ASP limit[14] $\sigma(e^+e^- \to \gamma + \text{missing}) < 0.03$ pb is reexamined for the case where "missing" is a pair of neutralinos, and again parts of parameter space are eliminated. While the PETRA constraint eliminates mostly states with a substantial Z-ino component, the ASP constraint rules out light, most photino states, and together the constraints rule out almost all states with mass less than about 5 GeV.

Finally, in Sec. VII we consider the effect of relaxing several simplifying assumptions made during the first six sections. We consider the effect the new European Muon Collaboration (EMC) spin-dependent structure functions[15] have on direct detection, and find a sensitivity to the composition of the neutralino and to shell-model parametrization of the nuclei whenever the axial-vector term dominates the elastic cross section. For shell-model "neutrons" there is a general lowering of rates, except for

photinolike states which are greatly enhanced. The change in rates for shell model "protons" is not as large except again for photinolike states which are somewhat suppressed. However, while details can vary substantially, the general picture remains unchanged. We also consider the effect of nondegenerate selectron and squark masses ($M_{\tilde{l}}$ and $M_{\tilde{q}}$, respectively). We show results for $M_{\tilde{q}}=3M_{\tilde{l}}$ which, along with the previously considered $M_{\tilde{q}}=M_{\tilde{l}}$, bracket most of the models we surveyed from the literature. Event rates are in general 1 or 2 orders of magnitude lower, since annihilation through the slepton channels ($\propto M_{\tilde{l}}^{-4}$) can be strong, resulting in weaker couplings to achieve $\Omega_{\tilde{\chi}}=1$, and these weaker couplings remain for the elastic scattering which proceeds via the heavier squark ($\propto M_{\tilde{q}}^{-4}$) exchange. Section VIII sums up the paper.

II. CROSS SECTIONS

In this section we calculate the matrix element and cross section for the annihilation ($\bar{\tilde{\chi}}\tilde{\chi}\to\bar{q}q$) of two neutralinos into a fermion and antifermion, and then use crossing symmetry to find the elastic scattering ($\tilde{\chi}q\to\tilde{\chi}q$) and the anomalous single-photon (ASP) ($e^+e^-\to\bar{\tilde{\chi}}\tilde{\chi}\gamma$) cross sections. Here q stands for any quark or lepton and e^- is the electron.

Throughout we will use the minimal supersymmetric extension of the standard model as described in Ref. 7 (see especially the appendixes) and Ref. 16. (We also use the conventions and notation of Ref. 7.) This is a group of models which is minimal, in the sense that it contains the minimum number of new particles, and general, in the sense that it has all possible *CP*-conserving soft supersymmetry-breaking terms included in the Lagrangian. In these models there exist four neutralinos which are linear combinations of the supersymmetric partners of the neutral W, the B, and the two neutral Higgs bosons. These can also be characterized as the photino, Z-ino, and two neutral Higgsinos. Only the lightest will be stable (we assume a conserved R parity and also that the lightest neutralino is the LSP) and we denote it as

$$\tilde{\chi}=Z_{11}\tilde{B}+Z_{12}\tilde{W}^3+Z_{13}\tilde{H}_1+Z_{14}\tilde{H}_2\ , \tag{1}$$

where the Z_{ij} are the elements of the real orthogonal matrix which diagonalizes the neutralino mass matrix; that is, if $Z_{11}=Z_{12}=0$, $\tilde{\chi}$ is a pure Higgsino, if $Z_{11}=\cos\theta_w$, $Z_{12}=\sin\theta_w$, $\tilde{\chi}$ is a pure photino, and if $Z_{11}=-\sin\theta_w$, $Z_{12}=\cos\theta_w$, $\tilde{\chi}$ is a pure Z-ino. The assumption of *CP* conservation ensures that the Z_{ij} are real.

The neutralino masses and the Z_{ij}'s are fully determined by four parameters: $\tan\beta$, μ, M, and M', where $\tan\beta=v_2/v_1$ is the ratio of Higgs vacuum expectation values,[17] M and M' are soft supersymmetry-breaking parameters, and μ is a supersymmetric Higgs-boson mass. Throughout, we make the standard[7] simplification $M'=\frac{5}{3}M\tan^2\theta_w$ which is true if the theory is eventually embedded in any simple grand unified group. Overall then we have three undetermined parameters: $\tan\beta$, M, and μ, and it is this parameter space we explore. We consider several representative values of $\tan\beta$ and M and μ in the wide range $0\le M,\mu\le 1$ TeV.

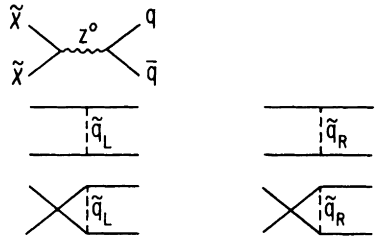


FIG. 1. Feynman diagrams contributing to neutralino ($\tilde{\chi}$) interaction with quarks or leptons (q). The first diagram shows Z^0 exchange while the last four show left and right chiral squark (or slepton) exchange.

For a neutralino of mass m_χ less than the Z^0 mass m_Z, the annihilation, elastic scattering, and production ($e^+e^-\to\bar{\tilde{\chi}}\tilde{\chi}$) processes are all given by the five Feynman diagrams shown in Fig. 1. There are four squark (or slepton) exchange diagrams and one Z^0 exchange. The Feynman rule for the left chiral $\tilde{\chi}q\tilde{q}_L$ vertex is $-ig\sqrt{2}(a\epsilon P_R+bP_L)$ and the rule for the right chiral $\tilde{\chi}q\tilde{q}_R$ vertex is $-ig\sqrt{2}(aP_L-\epsilon cP_R)$ where g is the electroweak coupling constant, $P_R=\frac{1}{2}(1+\gamma_5)$, etc., and

$$\begin{aligned} a&=m_q d_q/(2m_W)\ ,\\ b&=T_{3L}Z_{j2}-\tan\theta_w(T_{3L}-e_q)Z_{j1}\ ,\\ c&=\tan\theta_w e_q Z_{j1}\ . \end{aligned} \tag{2}$$

Here $d_q=Z_{j3}/\cos\beta$ for down-type quarks or leptons and $d_q=Z_{j4}/\sin\beta$ for up types, m_q is the mass of the quark or lepton, T_{3L} is the weak isospin, e_q is the charge, $\sin^2\theta_w=0.23$, and ϵ is the sign of the neutralino mass eigenvalue. The a factor corresponds to a Higgsino-type coupling, while the b and c factors are the photino–Z-ino couplings. The $Z^0\tilde{\chi}\bar{\tilde{\chi}}$ Feynman rule is $(ig/2\cos\theta_w)(Z_{13}^2-Z_{14}^2)\gamma^\mu\gamma_5$ and the $Z^0q\bar{q}$ rule is as usual $(-ig/\cos\theta_w)\gamma^\mu(c_LP_L+c_RP_R)$. Here $c_L=T_{3L}-e_q\sin^2\theta_w$ and $c_R=-e_q\sin^2\theta_w$. More details on the Feynman rules and techniques for handling Majorana fermions can be found in the Appendixes of Ref. 7 and in Ref. 16.

Since our results differ in the sign of interference terms and include new terms we present our matrix element before summing and squaring in the Appendix. The complete summed and squared matrix element including nondegenerate left and right chiral squark masses is also included in the Appendix. The effect of left-right chiral splitting has been considered previously[1,18] and is expected to be small. In order to simplify the formulas we will not consider this effect and set $M_{\tilde{q}L}=M_{\tilde{q}R}=M_{\tilde{q}}$. In this case we have, for the annihilation channel,

$$
\begin{aligned}
|\mathcal{M}|^2 = 16g^4 \Bigg\{ & (u'^2+v'^2)\left[\frac{(k_1p_1)^2}{(M_{\tilde{q}}^2-t)^2}+\frac{(k_1p_2)^2}{(M_{\tilde{q}}^2-u)^2}-\frac{m_{\tilde{X}}^2(k_1k_2)}{(M_{\tilde{q}}^2-t)(M_{\tilde{q}}^2-u)}\right] \\
& +2w'^2\left[\frac{(k_1p_1)^2+m_q^2m_{\tilde{X}}^2}{(M_{\tilde{q}}-t)^2}-\frac{(p_1k_1)^2+(p_1k_2)^2-(k_1k_2)(p_1p_2)+m_q^2m_{\tilde{X}}^2}{(M_{\tilde{q}}^2-t)(M_{\tilde{q}}^2-u)}+\frac{(k_1p_2)^2+m_q^2m_{\tilde{X}}^2}{(M_{\tilde{q}}^2-u)^2}\right] \\
& +2u'v'\left[\frac{m_q^2m_{\tilde{X}}^2}{(M_{\tilde{q}}^2-t)^2}+\frac{m_q^2m_{\tilde{X}}^2}{(M_{\tilde{q}}^2-u)^2}-\frac{m_q^2(p_1p_2)}{(M_{\tilde{q}}^2-t)(M_{\tilde{q}}^2-u)}\right] \\
& +4w'(u'+v')m_qm_{\tilde{X}}\left[\frac{(k_1p_1)}{(M_{\tilde{q}}^2-t)^2}+\frac{(k_1p_2)}{(M_{\tilde{q}}^2-u)^2}-\frac{(k_1p_1)+(k_1p_2)}{2(M_{\tilde{q}}^2-t)(M_{\tilde{q}}^2-u)}\right] \\
& +\frac{(Z_{13}^2-Z_{14}^2)^2}{\cos^4\theta_w(m_Z^2-s)^2}\{\tfrac{1}{4}(c_L^2+c_R^2)[(k_1p_1)^2+(k_1p_2)^2-m_{\tilde{X}}^2(k_1k_2)]+\tfrac{1}{2}c_Lc_Rm_q^2[(p_1p_2)-2m_{\tilde{X}}^2]\} \\
& +\frac{Z_{13}^2-Z_{14}^2}{2\cos^2\theta_w(m_Z^2-s)}\Bigg[\left[\frac{(k_1p_1)^2}{M_{\tilde{q}}^2-t}+\frac{(k_1p_2)^2}{M_{\tilde{q}}^2-u}\right](2v'c_R-2u'c_L) \\
& \qquad +\left[\frac{1}{M_{\tilde{q}}^2-t}+\frac{1}{M_{\tilde{q}}^2-u}\right]\{(u'c_L-v'c_R)m_{\tilde{X}}^2(k_1k_2) \\
& \qquad\qquad +m_q^2(u'c_R-v'c_L)[2m_{\tilde{X}}^2-(p_1p_2)]\} \\
& \qquad +2m_qm_{\tilde{X}}w'(c_R-c_L)\left[\frac{2(k_1p_2)-(k_1p_1)}{M_{\tilde{q}}^2-u}+\frac{2(k_1p_1)-(k_1p_2)}{M_{\tilde{q}}^2-t}\right]\Bigg]\Bigg\}, \qquad (3)
\end{aligned}
$$

where $m_{\tilde{X}}$ is the neutralino mass, p_1 and p_2 are the incoming $\tilde{\chi}$ four-momenta, k_1 and k_2 are the outgoing fermion momenta, and $(p_1p_2)=p_1^0p_2^0-\mathbf{p}_1\cdot\mathbf{p}_2$ denotes a four-momentum dot product. The symbols s, t, and u are the usual Mandelstam invariants, $u'=a^2+b^2$, $v'=a^2+c^2$, and $w'=\epsilon a(b-c)$.

The matrix element for the elastic scattering process $\tilde{\chi}q\to\tilde{\chi}q$ with momenta labeled $p_1+k_1\to p_2+k_2$, respectively, can be found from Eq. (3) by crossing $k_2\to -k_1$, $k_1\to k_2$, and $p_2\to -p_2$, so $t\to u$, $u\to s$, and $s\to t$. Likewise, the matrix element for the production process $e^+e^-\to\tilde{\chi}\bar{\tilde{\chi}}$ (momenta labeled $p_1+p_2\to k_1+k_2$) is found by $k_1\leftrightarrow p_2$ and $k_2\leftrightarrow p_1$. With these substitutions the labels m_q and $m_{\tilde{X}}$ remain unchanged.

To find the annihilation cross section we must integrate Eq. (3) over the center-of-mass scattering angle. This is tedious due to the angular dependence in the propagator momenta, but we do not actually need this result because we are only interested in using the annihilation cross section to calculate relic abundances. For this purpose one needs a thermally averaged cross section, which is most easily obtained by expanding the relativistic cross section in powers of the relative velocity v. We do this before integrating and keep terms up to order v^2. This method is also satisfactory for calculating neutralino annihilation in the Sun or galactic halo since again the collisions are nonrelativistic. We find the annihilation cross section to be

$$
\begin{aligned}
\sigma_{\text{ann}}v = \sum_q \frac{4}{\pi}G_F^2c_qm_{\tilde{X}}^2\beta'(y'^4\{ & (u'^2+v'^2)[z^2+(a_1+r_1)v^2]+4w'^2[1+(a_2+r_2)v^2] \\
& +4w'(u'+v')z[1+(\tfrac{1}{4}x^2+r_3)v^2]+2u'v'z^2[1+(a_4+r_4)v^2]\} \\
& +(Z_{13}^2-Z_{14}^2)^2x'^4[\tfrac{1}{4}(c_L^2+c_R^2)(z^2+a_1v^2)+\tfrac{1}{2}c_Lc_Rz^2(-1-a_4v^2)] \\
& +(Z_{13}^2-Z_{14}^2)x'^2y'^2\{(v'c_R-u'c_L)[z^2+(a_1+z^2r_5)v^2]+(u'c_R-v'c_L)z^2[1+(a_4+r_5)v^2] \\
& -2w'(c_L-c_R)z[1+(\tfrac{1}{4}x^2+r_5)v^2] \\
& +v^2\beta'^2r[-\tfrac{2}{3}(u'c_L+v'c_R)+w'(c_L-c_R)z]\}), \qquad (4)
\end{aligned}
$$

where

$$
a_1=\tfrac{2}{3}-\tfrac{5}{12}z^2+\tfrac{1}{4}z^2x^2,\quad a_2=\tfrac{1}{4}(2-z^2+x^2),\quad a_4=\tfrac{1}{4}(-3+x^2),
$$

$$
r_1=\frac{r}{3}(-4+z^2+4r-3rz^2-rz^4),\quad r_2=\frac{r}{3}(-5+2z^2+3r\beta'^2+2rz^2\beta'^2),
$$

$$
r_3=\frac{r}{3}(-3+5r\beta'^2-2\beta'^2),\quad r_4=\frac{r}{3}(-3+5r\beta'^2),\quad r_5=\frac{r}{3}(-\tfrac{3}{2}+r\beta'^2),
$$

where G_F is the Fermi constant, c_q is a color factor, 3 for quarks, 1 for leptons, $x'^2=m_Z^2/[(m_Z^2-s)^2+\Gamma_Z^2 m_Z^2]^{1/2}$ is the Z^0 pole factor, Γ_Z is the Z^0 width, $z=m_q/m_\chi$, $\beta'=(1-z^2)^{1/2}$, and $y'^2=m_W^2/(M_{\tilde{q}}^2+\beta'^2 m_\chi^2)$ is the squark mass suppression including the propagator momentum. The propagator momenta factor $r=m_\chi^2/(M_{\tilde{q}}^2+\beta'^2)$ is usually small, as is $x^2=\frac{1}{2}z^2/(1-z^2)$. When $m_\chi \to m_q$ however, $z\to 1$, $x\to\infty$, and the expansion breaks down. This occurs only very near mass thresholds, so we avoid evaluating Eq. (4) right at threshold. The effect of ignoring propagator momenta can be found by taking the limits $y'\to m_W/M_{\tilde{q}}$ and $r\to 0$ in Eq. (4).

As a check, we can consider the pure photino (or pure Higgsino) limit of Eq. (4). A pure photino has $Z_{13}=Z_{14}=a=0$, $Z_{11}=\cos\theta_w$, $Z_{12}=\sin\theta_w$ and $b=c=e_q\sin\theta_w$. The annihilation cross section becomes

$$\sigma_{\rm ann} v(\text{photino})=\sum_q \frac{8\pi\alpha^2 c_q e_q^4 m_\chi^2\beta'}{(M_{\tilde{q}}^2+m_\chi^2\beta'^2)^2}[z^2+\tfrac{1}{3}v^2(1-\tfrac{7}{4}z^2+\tfrac{3}{4}z^2x^2-2r+2r^2-rz^2+r^2z^2-3r^2z^4)] \ . \tag{5}$$

In the no-propagator momenta limit we recover the well-known[5,11,19] formula for the photino. We can also recover a Higgsino by taking $b=c=Z_{11}=Z_{12}=0$, and $Z_{13}=\sin\beta$, $Z_{14}=\cos\beta$. We perform the thermal average in the standard manner by replacing v^2 with $6T/m_\chi$ where T is the temperature, although this is not strictly correct to this order in v^2 (Ref. 20).

The above neutralino cross section, Eq. (4), differs from previous work[8] in several ways. First, it includes the effect of propagator momenta which can be important if the neutralino is near in mass to the squark or slepton. Even for the pure photino, this effect can be important as is discussed in Sec. V and shown in Fig. 10 below. Second, several new terms appear which can be important or not depending on parameters. Third, we find opposite signs for the b^2 and c^2 terms, an important effect when Z^0-squark interference occurs.

To find the elastic scattering cross section we cross the matrix element as described above and integrate over the center-of-mass scattering angle. Since we are interested in the scattering of neutralinos from the halo of our Galaxy, and the velocity of halo particles is known to be $v_{\rm halo}\sim 10^{-3}$, we can take an extreme nonrelativistic limit. This greatly simplifies the formulas. In particular, $s\to(m_\chi+m_q)^2$, $t\to 2\,|\,\mathbf{p}_1\,|^2(1-\cos\theta^*)$, $u\to(m_\chi-m_q)^2$, and so if we can ignore the small mass of the squark, we have $s\approx u$ and $t\approx 0$. Taking the limit $M_{\tilde{q}L}=M_{\tilde{q}R}=M_{\tilde{q}}$ we find

$$\sigma_{\rm el}=\frac{8G_F^2 m_\chi^2 m_q^2}{\pi(m_\chi+m_q)^2}\{3[(2a^2+b^2+c^2)\bar{x}_q^2+\tfrac{1}{2}(c_R-c_L)(Z_{13}^2-Z_{14}^2)]^2+4\bar{x}_q^4 a^2(b-c)^2\} \ , \tag{6}$$

where $\bar{x}_q^2=m_W^2/(M_{\tilde{q}}^2-m_\chi^2)$. We are actually more interested in scattering off nuclei than off quarks, and this will be discussed in the next section. Taking the pure photino limit we find

$$\sigma_{\rm el}(\text{photino})=\frac{48\pi\alpha^2 e_q^4 m_\chi^2 m_q^2}{(m_q+m_\chi)^2(M_{\tilde{q}}^2-m_\chi^2)^2} \ , \tag{7}$$

which, apart from the newly included propagator correction to the scalar fermion mass, reproduces the well-known formula.[5,11] A pure Higgsino result can also be easily found.

Finally, we wish to be able to include constraints on the parameter space from the ASP experiment.[14] This experiment measured the cross section for e^+e^- going to photon plus missing energy at the PEP storage ring at SLAC ($\sqrt{s}=29$ GeV). With their luminosity, the standard-model background from $e^+e^-\to\bar{\nu}\nu\gamma$ is predicted to be 0.03 pb, and from their limit of $\sigma(e^+e^-\to\gamma+\text{missing})\le 0.06$ pb, we get a limit on the production of neutralinos of $\sigma(e^+e^-\to\tilde{\bar{\chi}}\tilde{\chi}\gamma)\le 0.03$ pb (Ref. 21). In the soft-photon limit valid here, the desired cross section can be calculated from $\sigma(e^+e^-\to\tilde{\bar{\chi}}\tilde{\chi})$ using the formula[22]

$$\frac{d\sigma(e^+e^-\to\gamma X\bar{X};s)}{dx\,dy}\simeq\frac{2\alpha}{\pi}\frac{[(1-\frac{1}{2}x)^2+\frac{1}{4}x^2y^2]}{x(1-y^2)}\sigma(e^+e^-\to X\bar{X};\hat{s}) \ , \tag{8}$$

where s is the Mandelstam variable, $x=2E_\gamma/\sqrt{s}$ is the dimensionless photon energy, $y=\cos\theta$ is the angle between the beam and the photon, and $\hat{s}=s(1-x)$. Performing the crossing described above, setting the electron mass to zero and taking the limit $M_{\tilde{q}L}=M_{\tilde{q}R}=M_{\tilde{q}}$ we find

$$\sigma(e^+e^-\to\tilde{\bar{\chi}}\tilde{\chi})=\frac{2}{3\pi}G_F^2 s\left[1-\frac{4m_\chi^2}{s}\right]^{3/2}[x_q^4(b^4+c^4)+x_q^2x'^2(Z_{13}^2-Z_{14}^2)(c_R c^2-c_L b^2)+\tfrac{1}{4}x'^4(Z_{13}^2-Z_{14}^2)^2(c_L^2+c_R^2)] \ , \tag{9}$$

where $x_q = m_W/M_{\tilde{q}}$ and we have not included the propagator momenta here. The pure photino limit of Eq. (9) agrees with previous work.[23] (See also Ref. 24.)

III. EFFECTIVE LAGRANGIAN AND ELASTIC SCATTERING OFF NUCLEI

The cross sections presented in Sec. II are for neutralino interactions with quarks or leptons, while for direct detection, or capture into the body of the Earth or Sun, the interaction between neutralinos and nuclei is needed. For this purpose, it is useful to rederive some of the above using an effective Lagrangian approach. Since previous calculations[8] have used this approach it is also useful for comparison purposes.

We start from the supersymmetric Lagrangian interaction terms given in Refs. 7 and 16, and consider only the limit of heavy scalar fermions. We have

$$L_{\text{eff}} = \frac{-4g^2}{M_{\tilde{q}L}^2}\bar{\tilde{\chi}}(aP_R + bP_L)q\bar{q}(aP_L + bP_R)\tilde{\chi} - \frac{4g^2}{M_{\tilde{q}R}^2}\bar{\tilde{\chi}}(cP_R - aP_L)q\bar{q}(cP_L - aP_R)\tilde{\chi} - \frac{g^2}{2m_W^2}(Z_{13}^2 - Z_{14}^2)\bar{q}\gamma^\mu(c_L P_L + c_R P_R)q\bar{\tilde{\chi}}\gamma_\mu\gamma_5\tilde{\chi} , \tag{10}$$

where q is the quark field, and the other symbols were defined in Sec. II. To get Eq. (10) in a more useful form we perform Fierz transformations on the first two terms. Using the fact[7] that for Majorana fermions $\bar{\tilde{\chi}}\gamma_\mu\tilde{\chi}=0$ we find

$$L_{\text{eff}} = \frac{g^2}{2m_W^2}[\bar{\tilde{\chi}}\gamma_\mu\gamma_5\tilde{\chi}\bar{\gamma}^\mu(V' + A'\gamma_5)q + 2a\epsilon(bx_{qL}^2 - cx_{qR}^2)(\bar{\tilde{\chi}}\tilde{\chi}\bar{q}q + \bar{\tilde{\chi}}\gamma_5\tilde{\chi}\bar{q}\gamma_5 q)] , \tag{11}$$

where

$$\begin{aligned} V' &= -\tfrac{1}{2}(c_R + c_L)(Z_{13}^2 - Z_{14}^2) + x_{qL}^2(b^2 - a^2) + x_{qR}^2(a^2 - c^2) , \\ A' &= \tfrac{1}{2}(c_L - c_R)(Z_{13}^2 - Z_{14}^2) - x_{qL}^2(a^2 + b^2) - x_{qR}^2(a^2 + c^2) , \end{aligned} \tag{12}$$

and where $x_{qL} = m_W/M_{\tilde{q}L}$, etc. Later we will include the effect of propagator momenta by substituting $\bar{x}_{qL}$ for x_{qL}, etc. For the limiting case of pure photino or pure Higgsino, Eq. (11) agrees with the corresponding limiting cases of Ellis *et al.* For the general neutralino, however, there are several differences. First, there is a new term which is not of the form of an axial-vector coupling. In the limit of equal left and right chiral squark masses, it is proportional to $(b-c)$, the Z-ino coupling, and to a, the Higgsino coupling. This term may have been ignored previously because Higgsinos, like Higgs bosons, couple proportional to mass, and for quarks or leptons this is usually quite small. However, for low-energy elastic scattering off nuclei, Higgs bosons,[10] as well as Higgsinos couple proportional to the nuclei mass, not the quark mass, and so this new scalar term can be important. Second, the sign of the b^2 and c^2 terms [or equivalently the $(Z_{13}^2 - Z_{14}^2)$ and a^2 terms, since the overall sign is arbitrary] differ from those of Ellis *et al.* Since in annihilation, for example, the a^2 term is usually small, the primary effect of this sign difference is negative Z^0-squark interference, where before there was positive interference (and vice versa). We shall show that this interference results in low direct detection rates for parts of parameter space.

The effective Lagrangian, Eq. (11) can be used to calculate the annihilation, elastic, and production cross sections in the standard manner. For the annihilation cross section there are several terms which appear in the complete cross section, Eq. (4), which do not appear in the cross section calculated from the effective Lagrangian. For the elastic cross section in the extreme nonrelativistic limit we find that the pseudoscalar term does not contribute, and from the complete calculation, Eq. (6), we also see that there is no interference between the axial-vector and scalar terms. In fact, Eq. (11) reproduces Eq. (6) except for propagator momenta.

We now use Eq. (11) to find scattering off of nuclei. This is done by first finding $\sum_q |\langle N|L_{\text{eff}}|N\rangle|^2$, where $|N\rangle$ is the nucleus wave function, which is assumed to be a sum of nucleon wave functions $|n\rangle$, which are in turn assumed to be sums of quark wave functions. The axial-vector piece of the elastic cross section can be evaluated as in Goodman and Witten[1] and we find the elastic scattering cross section off a nucleus of mass m_N to be

$$\sigma_{\text{el}} = \frac{24m_{\tilde{\chi}}^2 m_N^2 G_F^2}{\pi(m_{\tilde{\chi}} + m_N)^2}\left[\tfrac{4}{3}\lambda^2 J(J+1)\left[\sum_{u,d,s} A'\Delta q\right]^2 + \left[\frac{2m_N}{27m_W}\right]^2\left[\sum_{c,b,t}(b-c)x_q^2 d_q\right]^2\right] , \tag{13}$$

where J is the total spin of the nucleus and the sums are over the indicated quarks. The first term agrees with Ref. 1 in the photino limit (see also Ref. 19) and the second term is new and requires some explanation. In the above, we followed Goodman and Witten,[1] and Refs. 25 and 18 in defining

$$\lambda=\tfrac{1}{2}\{1+[s_p(s_p+1)-l(l+1)]/[J(J+1)]\} \tag{14}$$

from the one-particle nuclear shell model[26] and the Landé formula, where l is the shell-model angular momentum and s_p is the proton (or neutron) spin. We also follow Refs. 15, 25, and 18 in defining $\langle p \mid \bar{q}\gamma_\mu\gamma_5 q \mid p \rangle = 2\Delta q \mathbf{s}_q$, where $\mathbf{s}_q$ is the spin of quark q and Δq measures the fraction of the proton spin carried by quark q. Under some assumptions, the EMC group[15] gives $\Delta u=0.746$, $\Delta d=-0.508$, and $\Delta s=-0.226$, while the flavor-SU(3) quark model predicts $\Delta u=0.97$, $\Delta d=-0.28$, and $\Delta s=0$. The EMC results are still controversial, so for most of our work we use the flavor-SU(3) values. In Sec. VII we show the effect of using the EMC values. Note that for simplicity we left out vector pieces in Eq. (11) which can be important if there is significant left and right chiral squark mixing. These terms have been discussed in Refs. 1 and 18 and are expected to be small.

In deriving the second term of Eq. (13) we modified slightly a technique described in Shifman, Vainshtein, and Zakharov[10] and used recently by Raby and West.[27] For coherent scattering of a neutralino off a nucleus we need to find

$$\left\langle N \left| \sum_q 2a(b-c)x_q^2\bar{q}q \right| N \right\rangle \propto \left\langle N \left| \sum_q T_{3L}x_q^2 d_q m_q \bar{q}q \right| N \right\rangle , \tag{15}$$

where $\langle N \mid$ is the nucleus state and the sum is over all the quarks, both valence and sea. Using the "heavy-quark expansion" for the charm, bottom, and top quarks Shifman, Vainshtein, and Zakharov, write

$$m_q\bar{q}q \simeq -\frac{2}{3}\frac{\alpha_s}{8\pi}G^a_{\mu\nu}G^a_{\mu\nu}+O\left(\frac{\alpha_s^2}{m_q^2}\right)$$

and by including the anomaly in the trace of the quark energy-momentum tensor $\theta_{\mu\mu}$ they find

$$m_N\bar{\Psi}_N\Psi_N=\langle N \mid \theta_{\mu\mu} \mid N \rangle \simeq -\frac{9\alpha_s}{8\pi}\langle N \mid G^a_{\mu\nu}G^a_{\mu\nu} \mid N \rangle . \tag{16}$$

Physically, this last equation says that the mass of the nucleon (and therefore the nucleus) comes from the light-quark anomaly. Since the light quarks in Eq. (15) are quite light, we follow Shifman, Vainshtein, and Zakharov in ignoring them and find

$$\left\langle N \left| \sum_q 4aT_{3L}x_q^2\bar{q}q \right| N \right\rangle \simeq \left(\frac{4m_N}{27m_W}\right)\sum_{c,b,t} T_{3L}x_q^2 d_q \simeq \frac{2m_N x_q^2}{27m_W}\left(\frac{2Z_{14}}{\sin\beta}-\frac{Z_{13}}{\cos\beta}\right) , \tag{17}$$

where in the last step we made the simplifying assumption that all quarks have the same mass. Using Eq. (17), one finds Eq. (13) in a straightforward manner.

The essence of the above derivation is that neutralinos can couple to the gluons in the nucleon via a loop involving heavy quarks and a squark. In the limit of very heavy squarks this becomes the same loop as Shifman, Vainshtein, and Zakharov discuss for the Higgs-boson-exchange case. Since the Higgsino, like the Higgs boson, couples proportional to the mass of the quark in the loop, the quark mass cancels out and one is left with a coupling to the gluons which does not involve (to lowest order) the heavy-quark mass. As mentioned above, the light-quark glue gives rise to the mass of the nucleon, and so the final coupling is proportional not to the quark mass, but to the nucleon mass. For dark-matter neutralinos from the galactic halo we can sum coherently over all the nucleons in the nuclei and arrive at a coupling proportional to the mass of the nucleus. Actually, when a heavy neutralino scatters off of a heavy nucleus some loss of coherence is possible and we take this effect into account in Sec. V. We do not claim that the above cross section, Eq. (13) is exact, but it shows that "spin-independent" cross sections exist for Majorana particles. Uncertainties include the extent to which the charm, bottom, and top quarks contribute equally, the extent to which the strange quark contributes, the possibility of additional generations of quarks, and higher-order contributions, both in the heavy-quark and the heavy-squark expansions.

Finally note that we have included the effect of propagator momenta for the elastic cross section, Eq. (13), by using $\bar{x}_q$, instead of x_q in A' as derived in the complete elastic cross section. This can have a substantial effect if m_χ is near $M_{\tilde{q}}$, as illustrated in Sec. V.

IV. RELIC ABUNDANCE AND LSP CONSTRAINTS

Using Eq. (4) the present-day mass density of neutralinos can be calculated. This "Lee-Weinberg" calculation[28] has been the subject of many papers and we will use the method described in Refs. 29 and 30, which is an extension of the methods described in Ref. 31. This is an approximate analytic solution to the Boltzman equation $\dot{n}=-3Hn-\langle\sigma v\rangle_{\rm ann}(n^2-n_0^2)$, which governs the number density n of particles as the Universe cools. Here H is the Hubble parameter, n_0 is the equilibrium number density of neutralinos, and $\langle\sigma v\rangle_{\rm ann}$ is the thermally averaged annihilation cross section derived in Sec. II. The basic idea is that when the temperature T of the radiation in the Universe falls below the mass of the neutralino, the neutralino number density is suppressed by a Boltzman factor $e^{-E/T}$ and falls rapidly. But since the interaction of neutralinos with ordinary matter is weak, there comes a time (denoted as freeze-out) after which annihilation of neutralinos becomes improbable and the neutralinos present then are, for the most part, still present today. Our approximation uses an accurate method of finding the freeze-out temperature, includes properly the effect of changing degrees of freedom, and is accurate to better than 5% (as compared to a numerical integration). The reader is referred to Refs. 29 and 30 for more details.

Figures 2(a)–2(c) show "scatter plots" of $\Omega_{\tilde{\chi}}h^2$ versus the mass m_{χ} of the neutralino, where $h=H/(100$ km sec^{-1}Mpc^{-1}). These figures were made by selecting a value of $M_{\tilde{q}}$ and tanβ and considering a grid of points in the (M,μ) plane. For each set of the four parameters M, μ, tanβ, and $M_{\tilde{q}}$ we find m_{χ} and $\langle \sigma v \rangle_{\text{ann}}$, and then the relic abundance $\Omega_{\tilde{\chi}}h^2$. Each $\times$ in the figure represents a different supersymmetric model, and the collection of $\times$'s indicate the range of relic abundance possible. There are several interesting features in these figures.

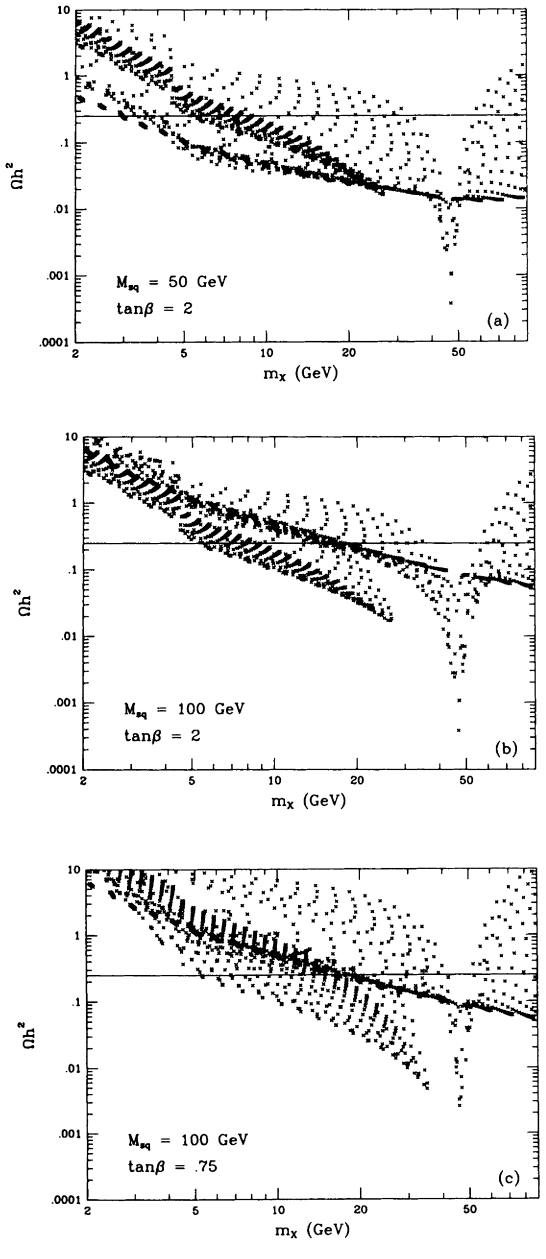


FIG. 2. Relic abundances of neutralinos vs the neutralino mass. Each $\times$ shows a different supersymmetric model. The horizontal line denotes critical density ($\Omega=1$) for $h=\frac{1}{2}$. Supersymmetric models with parameters in the wide range $0\leq M\leq 1$ TeV and $0\leq\mu\leq 1$ TeV are shown for several values of squark mass $M_{\tilde{q}}$ and tanβ.

First, notice that over a very wide range of supersymmetric parameters almost no models give $\Omega_{\tilde{\chi}}h^2<0.01$ (the lines across the figures indicate $\Omega_{\tilde{\chi}}=1$ for our preferred value $h=\frac{1}{2}$), the primary exception being at $m_{\chi}\sim m_Z/2$, where annihilation is greatly enhanced due to the Z^0 pole. The values of $M_{\tilde{q}}$ and tanβ in Figs. 2(a) and 2(c) were selected because they give low values for $\Omega_{\tilde{\chi}}h^2$; values of $M_{\tilde{q}}$ of less than 50 GeV being inconsistent with results from the UA1 experiment at CERN (Ref. 12). The "wedge"'s on the left of the figures consist of models where the neutralino contains substantial Higgsino content and move upward in $\Omega_{\tilde{\chi}}h^2$ as tanβ moves away from 1. The "lines"'s through the Z^0 pole consist of models with a photinolike neutralino and also move upward in $\Omega_{\tilde{\chi}}h^2$, this time as the (assumed degenerate) scalar-fermion masses increase. Recalling that the total luminous matter in the Universe contributes $\Omega\simeq 0.1$, the observed dynamical (dark) mass contributes $\Omega\simeq 0.1$–0.3, and nucleosynthesis limits baryonic matter to $0.015\leq\Omega h^2\leq 0.16$, we see that if low-energy supersymmetry exists and the neutralino is the LSP, it almost certainly is a significant component of the Universe. This is true whether or not it is the main component of dark matter in galactic halos. Since m_{χ} is in the GeV range, neutralinos constitute cold dark matter and should cluster in galactic halos, so we consider it likely that galaxies contain as much or more mass in the form of neutralinos as in the form of baryons. This likelihood alone is enough to justify experimental attempts to detect neutralino DM. It is also seen from Fig. 2 that a wide range of parameters predict a critical density of neutralinos ($\Omega_{\tilde{\chi}}h^2=\frac{1}{4}$) and that much of parameter space is ruled out by requiring $\Omega_{\tilde{\chi}}\leq 1$. Since $\Omega_{\tilde{\chi}}=1$ arises naturally and since $\Omega=1$ is desirable for several theoretical reasons, in this paper we will use the constraint $\Omega_{\tilde{\chi}}h^2=\frac{1}{4}$ to reduce the parameter space which must be explored. As mentioned in the Introduction, galactic neutralinos may be detectable and many of the items discussed in this paper may be relevant even if $\Omega_{\tilde{\chi}}<1$, especially since Fig. 2 suggests that $\Omega_{\tilde{\chi}}$ is likely to be greater than 4%. In fact, lowering $\Omega_{\tilde{\chi}}$ requires an increase in coupling strengths causing neutralinos to interact more strongly in a detector.

To begin to develop intuition about the neutralino parameter space we show in Figs. 3(a)–3(d) contours of $\Omega_{\tilde{\chi}}h^2=\frac{1}{4}$ in the (M,μ) plane for several values of tanβ and $M_{\tilde{q}}$. For typical values of tanβ and $M_{\tilde{q}}$ there is a "closed curve" in the middle of the figure and a "hyperbola" in the upper right-hand corner. The contours are sometimes broken; small gaps resulting from the finite grid size in our contour finding program, and large gaps showing areas where no solution exists. No solution may exist when a neutralino mass eigenvalue crossing occurs and the resulting discontinuity in neutralino composition causes a discontinuity in relic abundance. For brevity, we will call a piece of broken contour a "strand." These

strands will be the building blocks of most of the figures in the remainder of this paper. We map the strands onto various other parameter spaces and use consistency and accelerator experiment results, which constrain the parameter space, to chop pieces off the strands. This is not, perhaps, the ideal way to present our results, but four- (at least) dimensional parameter space is difficult to display. Note that Figs. 3(a) and 3(b) show the $\mu > 0$ case while Figs. 3(c) and 3(d) show the $\mu < 0$ case.

As an important first example, consider the consistency constraint that the neutralino be the LSP. In a given model $M_{\tilde{q}}$ is specified and so any values of M, μ, and $\tan\beta$ which result in $m_{\chi} > M_{\tilde{q}}$ are ruled out. For example, all the models in Fig. 2(a) (with $M_{\tilde{q}} = 50$ GeV) to the right of $m_{\chi} = 50$ GeV are inconsistent with the neutralino being the LSP and should not be considered. Throughout we discard regions of parameter space which do not satisfy $m_{\chi} < M_{\tilde{q}}$.

Another important consistency constraint comes from the charginos, the supersymmetric partners of the charged Higgs bosons. The neutralino parameters M, μ, and $\tan\beta$ determine uniquely the masses $m_{\tilde{\chi}^{\pm}}$ and couplings of the two charginos. Accelerator experiments constrain these masses and this will be discussed in Sec. VI, but here we only consider the consistency constraint $m_{\chi} < m_{\tilde{\chi}^{\pm}}$. Figures 4(a) and 4(b) show contours of $m_{\chi} = m_{\tilde{\chi}^{\pm}}$ for several values of $\tan\beta$ and $\mu > 0$. The areas between the two contours correspond to $m_{\tilde{\chi}^{\pm}} < m_{\chi}$ and are therefore inconsistent. Figure 4(a) shows the $m_{\chi} = m_{\tilde{\chi}^{\pm}}$ contours on top of the $\Omega_{\tilde{\chi}} h^2 = \frac{1}{4}$ contour for one value of $M_{\tilde{q}}$ and $\tan\beta$. Note that much of one strand lies between the $m_{\chi} = m_{\tilde{\chi}^{\pm}}$ contours. This is a generic feature when $\mu > 0$ and cuts out considerable parameter space. (For $\mu < 0$, we find $m_{\chi} < m_{\tilde{\chi}^{\pm}}$ almost everywhere and this constraint has little effect.) This loss of parameter space will show up later as breaks or gaps in the projected strands. Figure 4(b) shows the $m_{\chi} = m_{\tilde{\chi}^{\pm}}$ contours for several values of $\tan\beta$.

V. DIRECT DETECTION

In this section we apply the elastic scattering cross section, Eq. (13), to scattering from various elements and

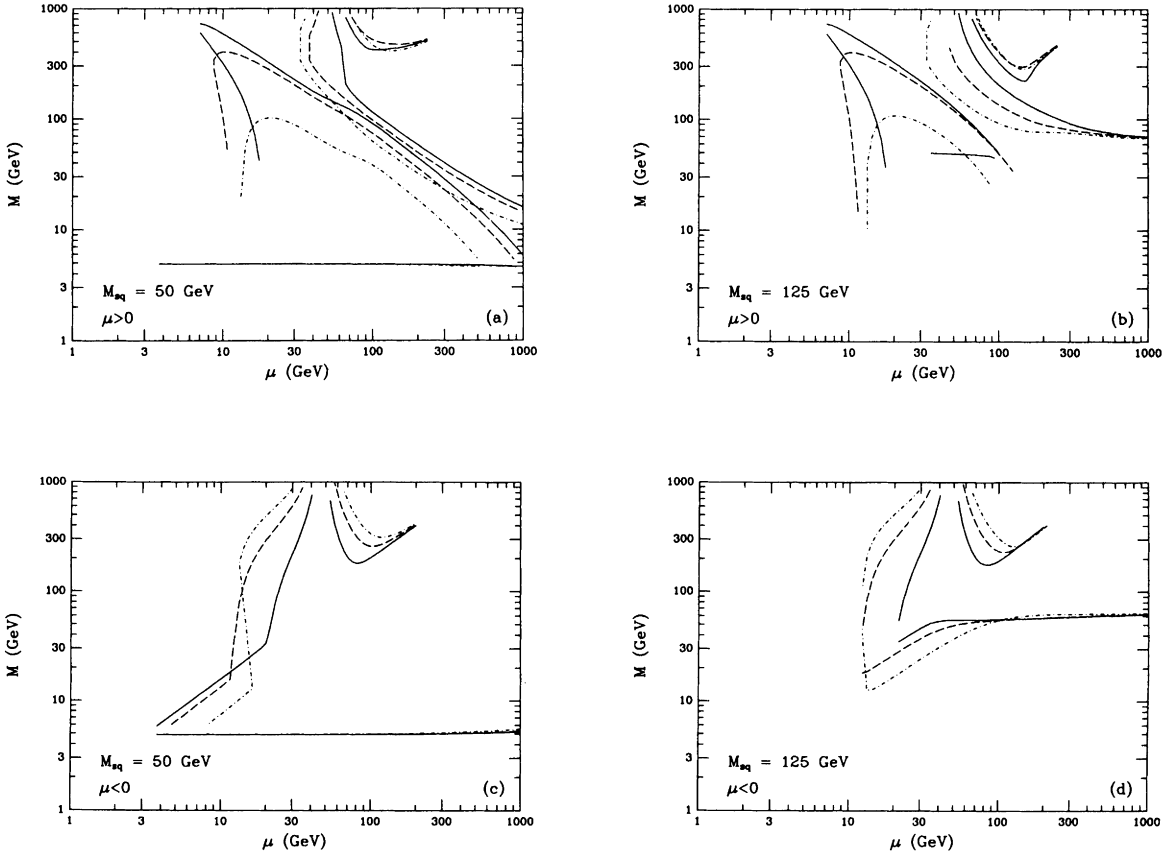


FIG. 3. Contours of $\Omega_{\tilde{\chi}} h^2 = \frac{1}{4}$ in the (M,μ) plane for various values of $M_{\tilde{q}}$ and $\tan\beta$. (a) and (b) show the $\mu > 0$ case while (c) and (d) show $\mu < 0$. Solid lines indicate $\tan\beta = \frac{3}{4}$, dashed lines indicate $\tan\beta = 2$, and dotted-dashed lines indicate $\tan\beta = \frac{1}{4}$.

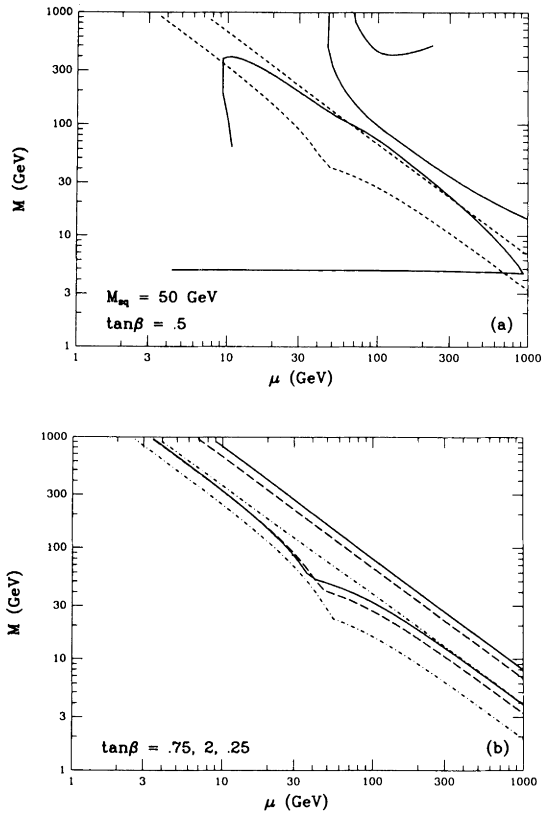


FIG. 4. Consistency constraint: contours of $m_X = m_{\tilde{\chi}^\pm}$ in the (M,μ) plane for various values of $M_{\tilde{q}}$ and $\tan\beta$, where $m_{\tilde{\chi}^\pm}$ is the mass of the lightest chargino. The areas between the contours are inconsistent with the neutralino as LSP. (a) shows the $\Omega_{\tilde{\chi}} h^2 = \frac{1}{4}$ contour (solid lines) being cut by the $m_X = m_{\tilde{\chi}^\pm}$ contour (dashed lines) for $M_{\tilde{q}} = 50$ GeV and $\tan\beta = \frac{1}{2}$. (b) shows contours for several values of $\tan\beta$. See Fig. 3 caption for the code for lines. Only $\mu > 0$ cases are shown.

give the rates for neutralino interaction in a cryogenic detector. This new class of detectors plans to measure the small (order keV) energy deposited when a particle from the galactic halo hits a nucleus in the detector. Ionization detectors now operating with energy thresholds of order 5 keV have already ruled out Dirac neutrinos with masses greater than 20 GeV as the major component of the galactic halo.[32] Many groups are developing new nonionization cryogenic detectors which will operate at lower temperatures to reduce the background and lower the energy threshold.

The rate of detection in a cryogenic detector is given in Ref. 3 (and verified by us) as

$$R = \left[\frac{8}{3\pi}\right]^{1/2} \eta_t \eta_v \eta_c \frac{\langle v \rangle_{\text{halo}} \rho_{\text{halo}} \sigma_{\text{el}}}{m_N m_X} , \tag{18}$$

where σ_{el} is the elastic cross section, Eq. (13), $\langle v \rangle \simeq 270$ km/sec is the average dispersion velocity in the halo, η_v is a correction due to the motion of the Sun and the Earth, and η_t corrects for the energy threshold. Including only the motion of the Sun for simplicity, we find $\eta_v \simeq 1.3$. The factor η_c estimates the loss of coherence which occurs at high momentum transfer when neutralinos start to interact with individual nucleons rather than the nucleus as a whole. Using the form factor from Ref. 3,

$$\eta_c = \alpha^{-2}\left[\frac{1}{2} + p'^2 + \frac{p' e^{-p'^2}}{\sqrt{\pi}\,\text{erf}p'}\right]^{-1} \times \left[1 - \left[\frac{\exp\left[\frac{-p'^2\alpha^2}{1+\alpha^2}\right]\text{erf}\left[\frac{p'}{\sqrt{1+\alpha^2}}\right]}{\sqrt{1+\alpha^2}\,\text{erf}p'}\right]\right], \tag{19}$$

where

$$\alpha^2 \simeq \frac{0.83(m_X/40\text{ GeV})^2}{(m_X/m_N+1)^2}[(m_N/200\text{ GeV})^{1/3}+0.06]^2 \times (\langle v \rangle/270\text{ km sec}^{-1}) , \tag{20}$$

and $p' = \sqrt{3/2} v_{\text{Sun}}/\langle v \rangle \simeq 1$. For convenience we give a plot of η_c as a function of m_N and m_X in Fig. 5. It is a substantial correction for heavy DM particles ($m_X > 50$ GeV) and for heavy nuclei ($m_N > 100$ GeV). Note that for simplicity, we have taken a zero-energy threshold throughout ($\eta_t = 1$). The actual value of η_t depends upon detector design and is most important for light nuclei and light DM particles.

In Fig. 6 we show the event rates in a mercury detector (natural abundance) plotted against the neutralino mass for several supersymmetric models. Figure 6(a) is the total rate, while Fig. 6(b) shows the axial-vector rate (without the new scalar term) and Fig. 6(c) shows the scalar interaction rate alone. Figures 6(a)–6(c) show the $\mu > 0$ case, while Figs. 6(d)–6(f) show the $\mu < 0$ case. We chose mercury not because it is an especially promising element, but because it emphasizes the possible importance of the new scalar term. It is heavy, and since $\lambda^2 J(J+1) = \frac{1}{12}$ for the 17% of Hg which has a spin, it

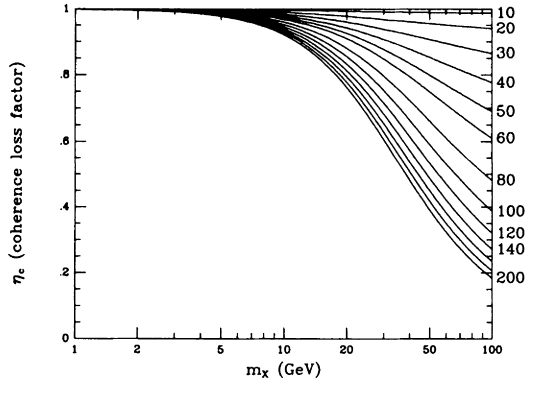


FIG. 5. Coherence loss factor η_c as a function of neutralino and nucleus mass. The various lines are labeled on the right by the nucleus mass in GeV.

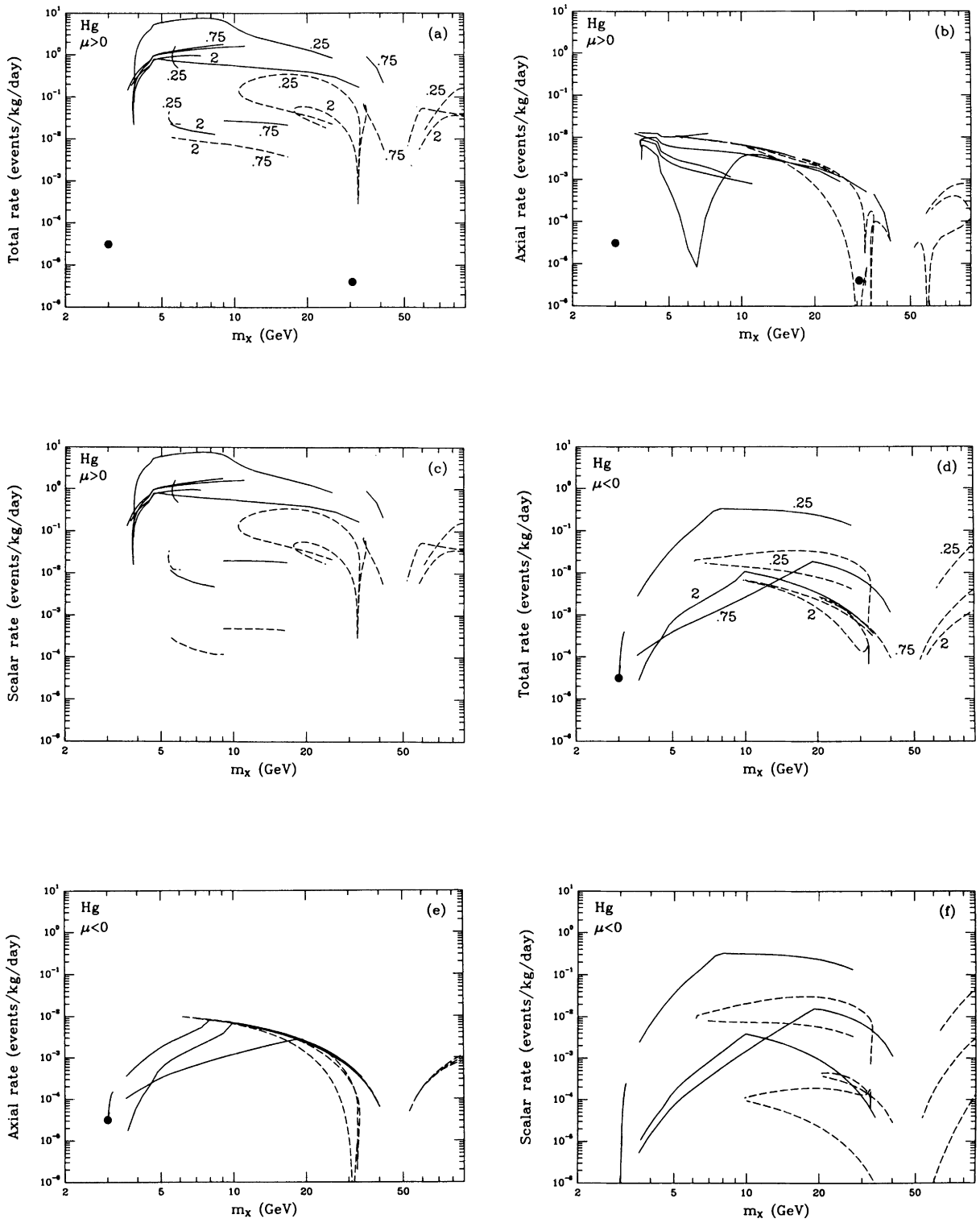


FIG. 6. Event rates in a cryogenic detector made of mercury as a function of neutralino mass. All points correspond to $\Omega_{\tilde{\chi}} h^2 = \frac{1}{4}$ for $\tan\beta = 0.25$, 0.75, or 2 and $M_{\tilde{q}} = 50$ or 125 GeV. Lines in (a) and (d) are labeled by $\tan\beta$. The solid lines indicate $M_{\tilde{q}} = 50$ GeV while the dashed lines indicate $M_{\tilde{q}} = 125$ GeV. (a)–(c) show the $\mu > 0$ case while (d)–(f) show the $\mu < 0$ case. (a) and (d) show the total rate, (b) and (e) show the contribution from the axial-vector term alone, (c) and (f) show the new scalar term contribution alone.

has very low rates for the pure axial-vector coupling. Figure 7 is the same as Fig. 6 but for fluorine [7(a) and 7(b)], a light element which is 100% a spin-$\frac{1}{2}$ isotope with favorable shell parameters [$\lambda^2 J(J+1)=\frac{3}{4}$] and for thallium [7(c)], a heavy element with 100% $\lambda^2 J(J+1)=\frac{3}{4}$. For fluorine the axial-vector term dominates almost everywhere while for thallium the scalar and axial-vector contributions are comparable. The lines in the figure cor-

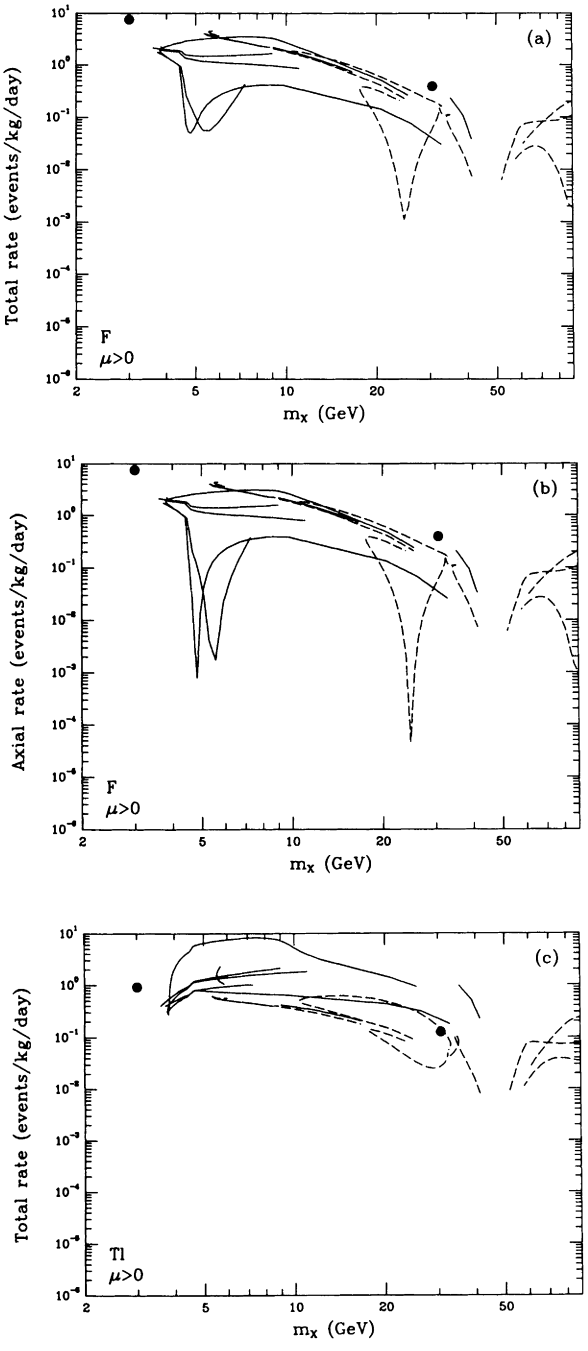


FIG. 7. Same as Fig. 6 for detectors made of fluorine and thallium. Only the $\mu > 0$ cases are shown.

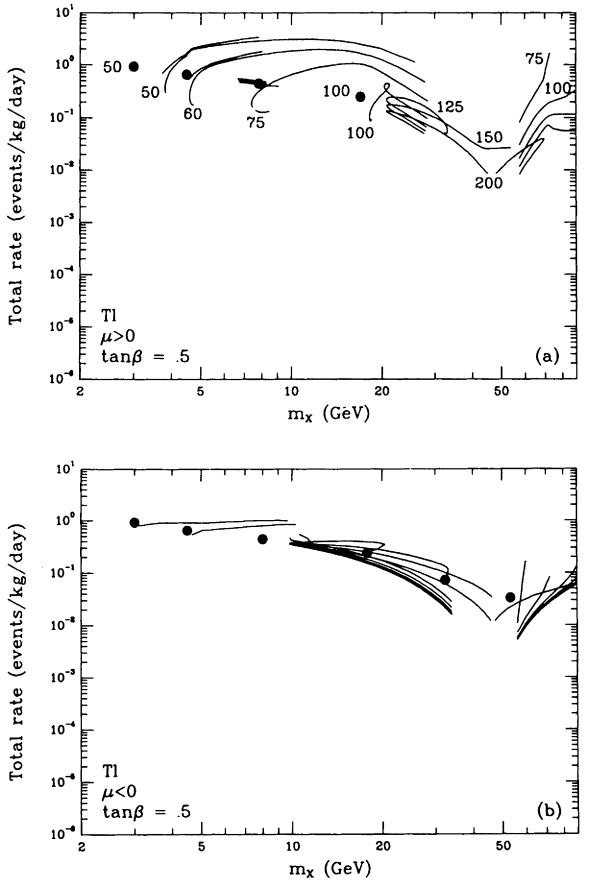


FIG. 8. Effect of varying squark mass: Event rates in a thallium detector as a function of neutralino mass for $\tan\beta=\frac{1}{2}$ and $M_{\tilde{q}}=50$, 60, 75, 100, 125, 150, and 200 GeV. The lines and "pure photino blobs" are labeled with the squark mass in GeV. For $\mu>0$ (a) no blobs were found for $M_{\tilde{q}} \geq 125$ GeV, and for $\mu<0$ (b) no blob was found for $M_{\tilde{q}}=200$ GeV.

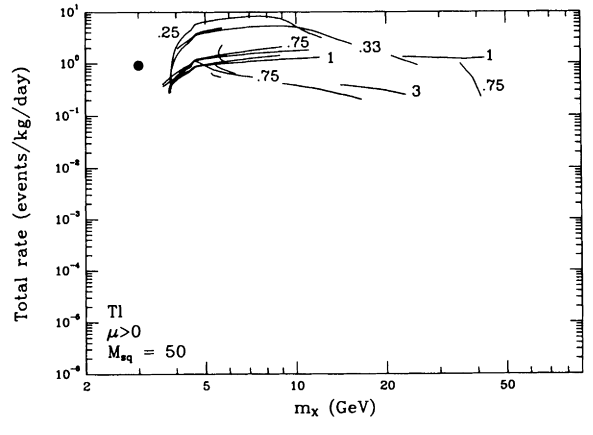


FIG. 9. Effect of varying $\tan\beta$: Event rates in a thallium detector as a function of neutralino mass for $M_{\tilde{q}}=50$ and $\tan\beta=0.25$, 0.33, 0.75, 1, and 3. The lines are labeled by $\tan\beta$.

respond to "strands" in the (M,μ) plane for which $\Omega_{\tilde{\chi}}h^2=\frac{1}{4}$ as discussed in Sec. IV. Various values of $M_{\tilde{q}}$ and $\tan\beta$ are shown. All squark and slepton masses have been taken degenerate; the nondegenerate case will be discussed in Sec. VII.

The rates shown here are easily scaled to other elements with the same one-particle-shell-model structure (Hg is a shell-model "neutron," while F and Tl are "proton" 's). The axial-vector rate is scaled by

$$\frac{[\lambda J(J+1)]_N}{[\lambda J(J+1)]_{\rm Hg}}\frac{m_N}{m_{\rm Hg}}\frac{(m_{\rm Hg}+m_X)^2}{(m_N+m_X)^2}\frac{f_N}{f_{\rm Hg}}\ , \tag{21}$$

where f_N is the relative abundance of the relevant isotope, while the scalar term is scaled by

$$\frac{m_N^3}{m_{\rm Hg}^3}\frac{(m_{\rm Hg}+m_X)^2}{(m_N+m_X)^2}\frac{(\eta_c)_N}{(\eta_c)_{\rm Hg}}\ . \tag{22}$$

For germanium, which is modeled as a "neutron," for example, these factors are roughly 4 and 0.4, respectively, for light neutralinos. For elements which are not describable by single-particle protons or neutrons the sum over $A'\Delta q$ must be redone to get the axial-vector rate. The scalar rate is still given by the formula above.

Figures 6 and 7 show several important features. First note the wide range in rates possible at a given m_X which comes from considering the general neutralino rather than just the pure photino. (See Ref. 18 for another recent discussion of this point.) A large percentage of the bottoms of the strands in the loops in Fig. 3 represent almost pure photino states and these strands cluster in the large "pure photino" blobs seen at around 3 and 30 GeV. The rate for a pure photino depends only on $M_{\tilde{q}}$ so there is one blob for $M_{\tilde{q}}=50$ GeV and one for $M_{\tilde{q}}=125$ GeV. The variation in rate is perhaps overemphasized with Hg where, as a shell-model neutron, the axial-vector rate is extremely small, due in part to the quark model flavor-SU(3) spin structure functions. (See Ref. 33 for the rates for Hg with EMC structure functions.) Thallium is also heavy, but is a shell-model proton and has favorable shell-model parameters and shows much less spread in the total rate. However, the axial-vector rate alone drops below 10^{-4} events kg^{-1} day^{-1} in several places due to the cancellations described below.

Next note the large dip in rate at $m_X\simeq m_Z/2\simeq 45$ GeV. The Z^0 pole makes annihilation of 45-GeV neutralinos very efficient, so small values of the couplings are needed for $\Omega_{\tilde{\chi}}=1$. These small couplings remain in the elastic cross section, where the Z^0 pole is not present, resulting in low rates. A neutralino with $m_X\simeq m_Z/2$ would be difficult to detect directly, though its suppression here implies an enhancement in Z^0 decay or accelerator production. In Fig. 6(b) (axial-vector term) there are also large dips in event rate at $m_X\simeq 6$ and 30 GeV. These are another generic feature and come about from negative Z^0-squark interference. (There is, in addition, cancellation among the terms in the sum over $A'\Delta q$.) The values $m_X\simeq 6$ and 30 are not special, and as $M_{\tilde{q}}$ and $\tan\beta$ are varied these cancellations occur for all values of m_X. These low rates are mitigated by the contribution from the scalar term as seen in Figs. 6(a) and 7(a). Even for fluorine, where the scalar term is small elsewhere, it dominates here and gives a minimum value for the elastic cross section. Another thing to note is that the pure photino blobs become lines as $M_{\tilde{q}}$ is varied, and that they do not move from Fig. 6(a) to Fig. 6(b) showing that the scalar term does not contribute to pure photino elastic scattering.

In order to get a feel for the four-dimensional parameter space we show in Fig. 8 the effect of varying $M_{\tilde{q}}$ with $\tan\beta$ fixed, and in Fig. 9 the effect of varying $\tan\beta$ with $M_{\tilde{q}}$ fixed. Keep in mind that reduction of the parameter space will result in further "chopping" of the strands when we include constraints from accelerator experiments in the next section.

Finally, we note that the inclusion of propagator momenta has had a significant effect when m_X neared $M_{\tilde{q}}$. The largest effect was in the annihilation cross section which decreased, thereby requiring an *increased* coupling strength for a given $\Omega_{\tilde{\chi}}$. The increased coupling strength increased the event rate, which was further increased by the inclusion of propagator momenta in the elastic cross section. For the general neutralino, the curves in Figs. 6 and 7 differ when propagator momenta are ignored, but the trend is hard to see. For the pure photino, however, the requirement $\Omega_{\tilde{\chi}}h^2=\frac{1}{4}$ fixes the squark mass as a function of m_X, which in turn fixes the event rate. In Fig. 10 we show the event rate in a fluorine cryogenic detector both with and without propagator momenta included. At $m_X\simeq 10$ GeV, the rate is 30% higher with propagator momenta, at $m_X\simeq 30$ GeV it is double, and by 90 GeV it is more than five times higher. If no new channels or particle thresholds appear, we note that the rate for pure photino detection reaches a minimum at about $m_X=130$ GeV and then starts to rise again as m_X approaches $M_{\tilde{q}}$.

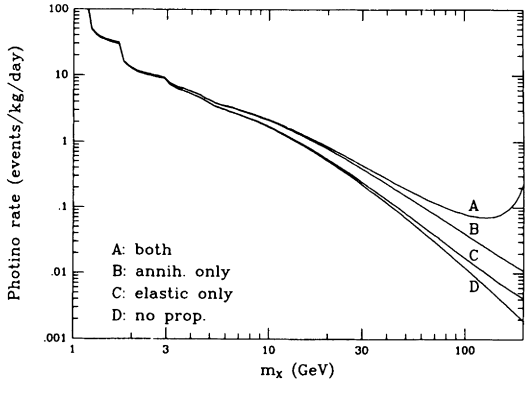


FIG. 10. Effect of propagator momenta for the pure photino. For each value of photino mass m_X, $M_{\tilde{q}}$ is found so that $\Omega_{\tilde{\chi}}h^2=\frac{1}{4}$ and the corresponding event rate in a fluorine cryogenic detector displayed. Curves show the effect of propagator momenta both in the annihilation cross section and in the elastic scattering cross section. Flavor-SU(3) spin structure functions were used.

VI. ACCELERATOR CONSTRAINTS

We have so far used cosmological predilection ($\Omega_{\tilde{\chi}}=1$) and consistency requirements ($M_{\tilde{q}}, m_{\tilde{\chi}_\pm} > m_{\tilde{\chi}}$) to reduce the rather large parameter space in which neutralino dark matter lives. In our selection of values of μ, M, $\tan\beta$, and $M_{\tilde{q}}$ we have also implicitly used theoretical prejudice ($\tan\beta$ not too different from 1, soft supersymmetry-breaking parameters less than 1 TeV) and accelerator results[12] ($M_{\tilde{q}} > 50$ GeV). In this section we will further restrict the parameter space by including the result from PETRA (Ref. 13) that $m_{\tilde{\chi}_\pm} > 23$ GeV and the result from ASP (Ref. 14) that $\sigma(e^+e^- \to \bar{\tilde{\chi}}\tilde{\chi}\gamma) \lesssim 0.03$ pb. The first experiment restricts parameter space because the same three parameters μ, M, and $\tan\beta$ which determine the neutralino mass and couplings also determine the chargino mass and couplings. (The charginos are the supersymmetric partners of the two charged Higgs bosons.) The ASP experiment is relevant because the $e^+e^- \to \bar{\tilde{\chi}}\tilde{\chi}$ cross section is related to the $\bar{\tilde{\chi}}\tilde{\chi} \to e^+e^-$ cross section. In fact, in general, accelerator experiments are important for particle DM detection since the DM particles may first be found there, and because negative results bear on the feasibility of direct detection. Since limits from experiments will continue to improve it is important to know what parts of parameter space are affected.

Figure 11 shows contours of $m_{\tilde{\chi}_\pm}=23$ GeV in the (μ, M) plane for several values of $\tan\beta$. Figure 11(a) shows $\mu > 0$ and Fig. 11(b) shows $\mu < 0$. The areas between the contours in Fig. 11(a) have $m_{\tilde{\chi}_\pm} < 23$ GeV and are ruled out, while the areas outside the contours are ruled out in Fig. 11(b). The effect this has on our event-rate plots in Figs. 6(a) and 6(d) is shown in Fig. 12, where the ruled out areas are marked with x's. For $\mu > 0$ many of the models with $m_{\tilde{\chi}} < 10$ GeV have been eliminated. Pure photinos, however, are relatively unaffected by this constraint (note the "blob" at 3 GeV is not ruled out) since they occur mostly near the $M=0$ axis, and the affected areas between the contours in Fig. 11(a) contain mostly models where the neutralino has substantial Higgsino and Z-ino components. We note that if the lim-

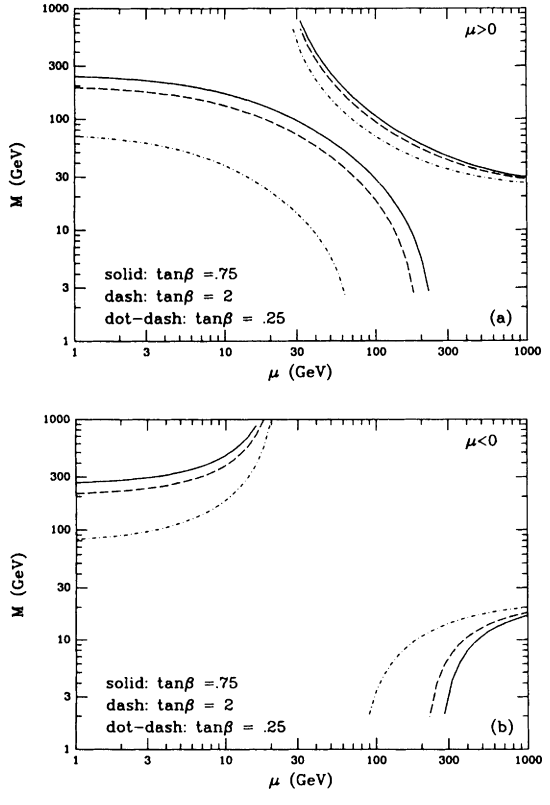


FIG. 11. PETRA constraint. Contours of $m_{\tilde{\chi}_\pm}=23$ GeV in the (M,μ) plane are shown for several values of $\tan\beta$. (a) shows the $\mu > 0$ case where the areas *between* the contours are ruled out due to $m_{\tilde{\chi}_\pm} < 23$ GeV. (b) shows the $\mu < 0$ case where the areas *outside* the contours are ruled out.

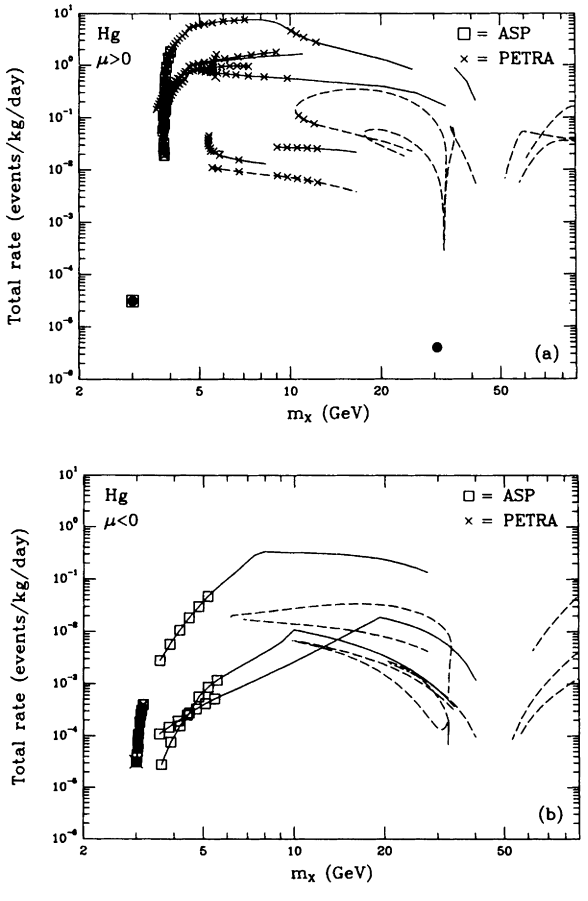


FIG. 12 Effect of PETRA and ASP constraints. Event rates in a mercury detector as a function of the neutralino mass [same as Figs. 6(a) and 6(d)] with areas ruled out by the PETRA and ASP constraints marked. (a) shows the $\mu > 0$ case while (b) shows the $\mu < 0$ case. Areas ruled out by PETRA are marked with x's and areas ruled out by ASP are marked with large boxes.

it on the chargino mass was pushed to $m_{\tilde{\chi}^{\pm}} > 40$ GeV, most of the parameter space with $m_{\chi} < 20$ GeV (except for pure photinos) would be ruled out. Conversely, if a chargino was discovered with $m_{\tilde{\chi}^{\pm}} < 40$ GeV we would have a good idea of the neutralino mass and couplings. Figure 12(a) shows the $\mu > 0$ case, while Fig. 12(b) shows the $\mu < 0$ case. For $\mu < 0$ the chargino constraint is not nearly as important because much less area is ruled out.

The ASP experiment,[14] performed at the PEP storage ring at SLAC, and similar experiments,[22] in principle actually produce neutralinos. These are neutrino counting experiments whose primary aim is to limit the cross section $\sigma(e^+e^- \to \gamma\bar{\nu}\nu)$, where the photon is detected and the presence of the neutrinos is inferred from their missing energy. For neutralinos with mass below the ASP threshold ($\sim\sqrt{s}/2 \sim 14.5$ GeV), the corresponding process $e^+e^- \to \bar{\tilde{\chi}}\tilde{\chi}\gamma$ can take place and the ASP limit of $\sigma(e^+e^- \to \gamma + \text{missing}) \leq 0.06$ pb (90% confidence level) can be used to rule out those areas of neutralino parameter space which would exceed this. Actually, the neutrino production process must exist as a background [$\sigma(e^+e^- \to \bar{\nu}\nu\gamma) \sim 0.03$ pb] and the rate for production of exotic particles is then $\sigma_{\text{exotic}} < 0.03$ pb. The ASP group[14] has used this limit to produce restrictions on pure photino parameter space, that is, the selectron mass, and find $M_{\tilde{l}} > 62$ GeV for $m_{\chi} = 0$. However, a Bayesian analysis of the same data[34,14] results in the weaker bound $\sigma_{\text{exotic}} < 0.046$ pb, giving $M_{\tilde{l}} > 58$ GeV for $m_{\chi} = 0$. Both statistical methods have problems, and since a recent report[35] combines Bayesian analysis and the results from other experiments to arrive at a number near 68 GeV for the selectron mass, we will use the stronger (and simpler) limit $\sigma_{\text{exotic}} < 0.03$ pb here.

We have integrated Eq. (8) over the ASP acceptance $20 < \theta < 160$ degrees, $E_{\max} > 12$ GeV, and $p_T > 0.8$ GeV/c, and show in Fig. 13 contours of $\sigma(e^+e^- \to \bar{\tilde{\chi}}\tilde{\chi}\gamma) = 0.03$ pb for $\mu > 0$. The areas below the long contours and within the wedges at the right are ruled out at the 90% confidence level. The contours for the $\mu < 0$ case are very similar, the only differences being that the long contours extend straight through to the right edge, and that there are no "wedges"'s. In contrast with the PETRA chargino constraint, we see that for $\mu > 0$ this experiment affects mostly photino neutralinos and has little effect on the rest of parameter space. (See also the recent report of Dress, Kim, and Tata.[24]) The effect of the ASP limits on the event rate versus m_{χ} plots of Fig. 6 is shown in Fig. 12 where the ruled out areas are marked with big boxes. For $\mu > 0$ the pure photino blob and a few other points are affected, while for $\mu < 0$, virtually every neutralino with $m_{\chi} < 5$ GeV is ruled out. We see that the PETRA and ASP constraints work in a complimentary fashion to rule out most very light ($m_{\chi} \leq 5$ GeV) neutralinos. As new experiments are performed we expect these con-

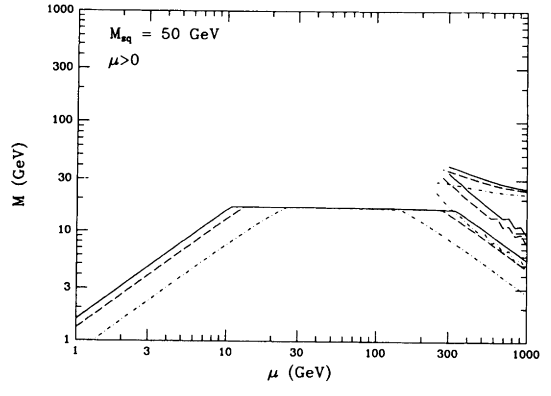


FIG. 13. ASP constraint. Contours of $\sigma_{\text{ASP}} = 0.03$ pb (see text) in the (M,μ) plane are shown for $M_{\tilde{q}} = 50$ and several values of tanβ. The code for the lines is the same as in Fig. 3. The areas below the long contours and inside the "wedges" at the far right are ruled out. Only the $\mu > 0$ case is shown; the $\mu < 0$ case is very similar except the horizontal contours continue straight across to $\mu = 1$ TeV and there are no "wedges." There are no solutions for $M_{\tilde{q}} = 125$ GeV and so no constraints exist for that case.

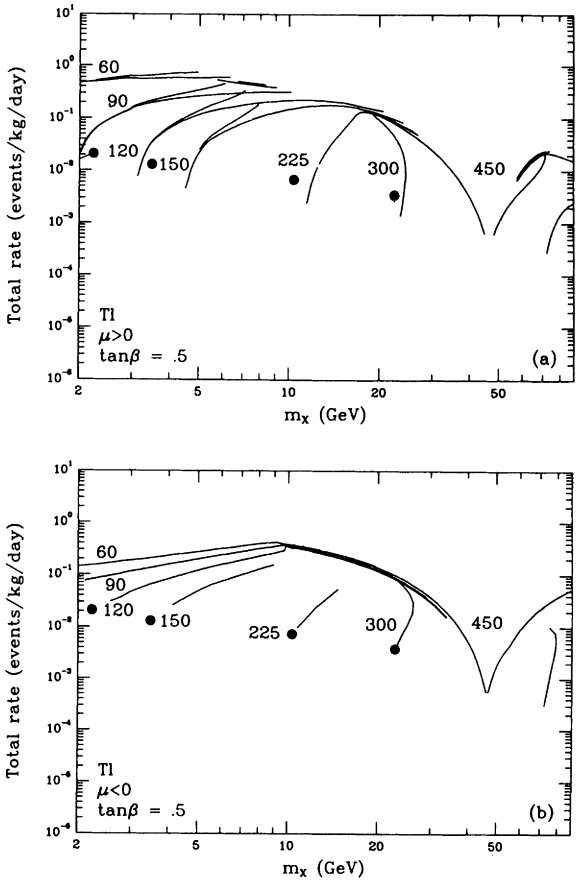


FIG. 14. Effect of nondegenerate squarks and sleptons. Event rates in a thallium detector as a function of the neutralino mass with $M_{\tilde{q}} = 3M_{\tilde{l}}$ instead of $M_{\tilde{q}} = M_{\tilde{l}}$ (see Fig. 8). Lines and blobs are labeled by the value of the squark mass in GeV. In comparing with Fig. 8 the relevant mass is actually the selectron mass, one-third of the labeled mass. (a) shows the $\mu > 0$ case while (b) shows the $\mu < 0$ case.

straints to strengthen and eventually rule out (or discover) the light neutralino. Unfortunately, we see that current experiments give little information on heavier neutralinos and this situation is unlikely to change in the near future.

VII. MODEL DEPENDENCE AND OTHER UNCERTAINTIES

In this section we consider the effect of relaxing a few of the many simplifying assumptions we made throughout the bulk of this paper. Specifically, we consider the effect of nondegenerate scalar-fermion masses and the effect of using the EMC rather than the flavor-SU(3) proton spin structure functions.

The assumption of equal-mass squarks and sleptons played an important role in producing reasonable cryogenic detector event rates. There are six left chiral squarks, six right chiral squarks, and an equal number of sleptons, all of which, in principle, can have different masses. In addition, there can be off-diagonal terms in the mass matrices which result in mixing. These effects have been discussed previously,[1,18] so to limit the possibilities, we will here consider only splitting between the squarks and sleptons, assuming degeneracy among the squarks and among the sleptons themselves. The off-diagonal terms are expected to be small, and the effect of varying squark masses, while possibly important, is very analogous to squark-slepton splitting. To get an idea of the range of models possible, we surveyed approximately 20 supersymmetric models[36] which used renormalization-group techniques to predict the scalar-fermion mass splittings. These included "superstring-inspired," "heavy top supergravity," and "light top supergravity" models among others. The range of splitting predicted varied substantially from model to model; however, all the models we surveyed were contained within two extreme cases: $M_{\tilde{q}}=M_{\tilde{l}}$ and $M_{\tilde{q}}=3M_{\tilde{l}}$. We used $M_{\tilde{q}}=M_{\tilde{l}}$ throughout this paper, so we show here the case $M_{\tilde{q}}=3M_{\tilde{l}}$ to hopefully bracket the effect of scalar-fermion splitting.

In Fig. 14 we show the total event rates in a mercury detector for $M_{\tilde{q}}=3M_{\tilde{l}}$ and the same value of $\tan\beta$ displayed in Fig. 8. The values of $M_{\tilde{q}}$ in Fig. 8 correspond roughly to values of $M_{\tilde{l}}$ in Fig. 14, with $M_{\tilde{q}}$ in Fig. 14 being three times larger. The curves have moved and overall there is a reduction in rate by 1–2 orders of magnitude. As explained in the Introduction this is to be expected. We see that substantial scalar-fermion splitting, if it exists, would make neutralino dark-matter detection more difficult.

Recent measurements[15] of the spin structure of the proton can be interpreted as giving the surprising result that very little of the spin of the proton is carried by the quarks. While the measurement and interpretation are still controversial, the results, if true, have important implications for neutralino detection. In particular, the matrix element $\langle p \mid \bar{q}\gamma^{\mu}\gamma_5 q \mid p \rangle = 2\Delta q s_q$ is found to differ substantially from the flavor-SU(3) quark-model expectations. The effect of this difference is illustrated in Fig. 15 where we show the analogues of Figs. 6(b) and 7(a) using

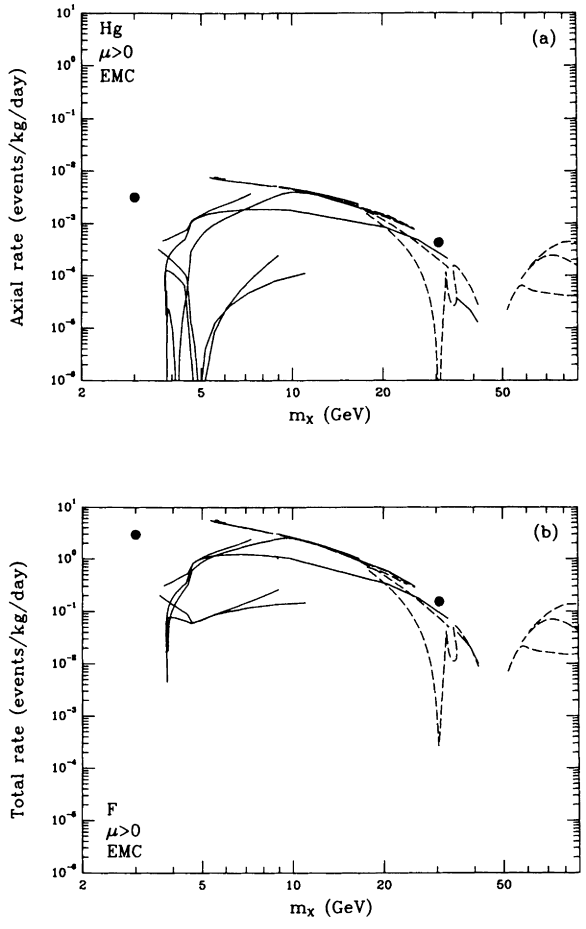


FIG. 15. Effect of EMC structure functions. Same as Fig. 6(b) (a) and Fig. 7(a) (b) with EMC rather than flavor-SU(3) spin-dependent structure functions.

EMC rather than flavor-SU(3) values for the Δq's. The rates are substantially changed whenever the axial-vector term dominates. The scalar term is unaffected. Since the scalar term dominates the rate for mercury, in Fig. 15(a) we show the axial-vector rate only [see Fig. 6(b)]. For fluorine the axial-vector term dominates so here we show the total rate [see Fig. 7(a)]. We see that the magnitude and sign of the change caused by using EMC depends sensitively on both the neutralino composition and the nuclear shell-model parametrization. Photinolike states (especially the pure photino blobs) seem to be strongly affected, the rate being either strongly enhanced for shell-model neutrons like Hg, or suppressed for shell-model protons like F. Note that the positions of the Z^0-squark interference cancellations have moved and that for shell-model neutrons there is a general lowering of the rate. The effect of the EMC structure functions on direct detection of neutralinos has also been discussed recently in Ref. 18. This uncertainty in spin structure functions will have to be resolved before accurate determinations of neutralino detection rates can be made.

Finally we mention some other uncertainties and limi-

tations bearing on the results of this paper. First, we note we have ignored the top quark and the Higgs bosons throughout. These particles almost certainly exist in some form, but their masses and properties are unknown. The cross sections and some of our conclusions would change for some values of the top mass. Ellis *et al.*[8] in fact, included a 30-GeV top quark in their original calculations, and this explains some of our differences. The Higgs bosons (there are five in the minimal supersymmetric models) are even more problematic since both masses and mixing parameters are unknown and again, their inclusion could make important changes in our results. Next, we have throughout only considered neutralinos less massive than the Z^0 boson. Heavier neutralinos could exist, but new annihilation channels open up and the problem of ignoring the top quark and Higgs bosons is exacerbated. Certainly, the trend is that detection becomes more difficult as the neutralino mass increases, but a very heavy LSP pushes the masses of the other supersymmetric particles even higher and makes a supersymmetric solution to the hierarchy problem more unlikely. Another limitation is that we considered only the minimal supersymmetric extension of the standard model. Although this is the most well-studied class of models, many other models have been proposed, and the addition of new particles and couplings can change our predictions drastically. In fact, there may even be particle dark matter whose only interaction with ordinary matter is gravitational. Finally, and most importantly, we have throughout made the assumptions that low-energy supersymmetry exists, and that the neutralino is the LSP.

VIII. CONCLUSIONS

In this paper we covered a wide range of topics having to do with the detection of neutralino dark matter. We started by calculating the neutralino annihilation, elastic scattering, and production cross sections and then used these to find the relic abundance of neutralinos and the rate of interaction of galactic neutralinos with a cryogenic detector. New features included additional terms in the cross sections, the inclusion of propagator momenta, and the sign of interference terms. We emphasized that over most of the supersymmetric parameter space a near-critical density of neutralinos exists, and that therefore they are likely to constitute a significant fraction ($\Omega_{\tilde{\chi}} \geq 0.04$) of the galactic halo. This is true whether or not they are, in fact, the dark matter. With regards to direct detection, a new term in the elastic cross section coming from heavy-quark loops gives neutralinos a "spin-independent" interaction which can be important for heavy materials, and reduces the Z^0-squark negative interference in almost all cases. Pure photinos and pure Higgsinos are not affected by the new scalar term.

We then considered the complementary nature of direct detection DM searches and accelerator experiments, showing how results from the PETRA, ASP, and UA1 experiments rule out large areas of parameter space. These constraints in turn affect the detectability of neutralinos. Finally, we considered the effect of nondegenerate squarks and sleptons on our results, showing a reduction in direct detection rates for the most extreme cases of 1 or 2 orders of magnitude, and also the effect of the new EMC proton spin structure functions on the rates, showing again a substantial change.

Overall, we found that the neutralino makes a superb particle dark-matter candidate, and that the pure photino and pure Higgsino special cases usually considered do not represent well the breadth of possibilities. These particles have a chance of being detectable, either directly or in accelerator experiments and we encourage our experimental colleagues to make strong efforts in this direction.

ACKNOWLEDGMENTS

We would like to acknowledge very useful conversations with P. Arnold, W. Bardeen, R. Flores, H. Haber, B. Sadoulet, D. Seckel, and M. Turner. This work was supported in part by the DOE (at Chicago). The computing was supported in part by the DOE and NASA (at Fermilab).

APPENDIX

In this appendix we list the complete matrix element, including propagator momenta and nondegenerate squark (or slepton) masses. We also show the initial matrix elements corresponding to the Feynman diagrams in Fig. 1.

For the annihilation process ($\bar{\tilde{\chi}}\tilde{\chi} \to \bar{q}q$) labeled by the four-momenta ($p_1 + p_2 \to k_1 + k_2$) the five graphs of Fig. 1 have matrix elements

$$\mathcal{M}_a = \frac{-2g^2}{M_{\tilde{q}L}^2 - t} \bar{u}(k_1)(a\epsilon_i P_L + bP_R)u(p_1)\bar{v}(p_2)(a\epsilon_i P_R + bP_L)v(k_2) ,$$

$$\mathcal{M}_b = \frac{-2g^2}{M_{\tilde{q}R}^2 - t} \bar{u}(k_1)(-aP_R + \epsilon_i cP_L)u(p_1)\bar{v}(p_2)(-aP_L + \epsilon_i cP_R)v(k_2) ,$$

$$\mathcal{M}_c = \frac{-2g^2}{M_{\tilde{q}L}^2 - u} \bar{v}(p_2)(aP_L + \epsilon_i bP_R)v(k_1)\bar{u}(k_2)(aP_R + \epsilon_i bP_L)u(p_1) , \tag{A1}$$

$$\mathcal{M}_d = \frac{-2g^2}{M_{\tilde{q}R}^2 - u} \bar{v}(p_2)(-\epsilon_i aP_R + cP_L)v(k_1)\bar{u}(k_2)(-\epsilon_i aP_L + cP_R)u(p_1) ,$$

$$\mathcal{M}_z = \frac{-g^2(Z_{13}^2 - Z_{14}^2)}{2\cos^2\theta_w (m_Z^2 - s)} \bar{v}(p_2)\gamma^\mu\gamma_5 u(p_1)\bar{u}(k_1)\gamma_\mu(c_L P_L + c_R P_R)v(k_2) ,$$

where ϵ_i is the sign of the neutralino mass eigenvalue and the other symbols are defined in Sec. II.

The total matrix element is $\mathcal{M}=\mathcal{M}_z+\mathcal{M}_c+\mathcal{M}_d-\mathcal{M}_a-\mathcal{M}_b$. The total matrix element squared is the sum of the following terms:

$$|\mathcal{M}_a|^2=\frac{16g^4}{(M_{\tilde{q}L}^2-t)^2}\{(a^2+b^2)^2(k_1p_1)(k_2p_2)+4a^2b^2m_q^2m_X^2+2m_qm_Xab(a^2+b^2)\epsilon[(k_1p_1)+(k_2p_2)]\}\ ,$$

$$|\mathcal{M}_b|^2=\frac{16g^4}{(M_{\tilde{q}R}^2-t)^2}\{(a^2+c^2)^2(k_1p_1)(k_2p_2)+4a^2c^2m_q^2m_X^2-2m_qm_Xac(a^2+c^2)\epsilon[(k_1p_1)+(k_2p_2)]\}\ ,$$

$$|\mathcal{M}_c|^2=\frac{16g^4}{(M_{\tilde{q}R}^2-u)^2}\{(a^2+b^2)^2(k_1p_2)(k_2p_1)+4a^2b^2m_X^2m_q^2+2m_qm_Xab(a^2+b^2)\epsilon[(k_1p_2)+(k_2p_1)]\}\ ,$$

$$|\mathcal{M}_d|^2=\frac{16g^4}{(M_{\tilde{q}R}^2-u)^2}\{(a^2+c^2)^2(k_1p_2)(k_2p_1)+4a^2c^2m_q^2m_X^2-2m_qm_Xac(a^2+c^2)\epsilon[(k_1p_2)+(k_2p_1)]\}\ ,$$

$$|\mathcal{M}_z|^2=\frac{16g^4(Z_{13}^2-Z_{14}^2)^2}{\cos^4\theta_w(m_Z^2-s)^2}\{\tfrac{1}{4}(c_L^2+c_R^2)[(k_1p_2)(k_2p_1)+(k_1p_1)(k_2p_2)-m_X^2(k_1k_2)]+\tfrac{1}{2}c_Lc_Rm_q^2[(p_1p_2)-2m_X^2]\}\ ,$$

$$2\,\mathrm{Re}\mathcal{M}_a\mathcal{M}_b^*=\frac{32g^4}{(M_{\tilde{q}L}^2-t)(M_{\tilde{q}R}^2-t)}[a(c-b)(k_1p_1)+m_qm_X\epsilon(bc-a^2)][a(c-b)(k_2p_2)+m_qm_X\epsilon(bc-a^2)]\ ,$$

$$\begin{aligned}2\,\mathrm{Re}\mathcal{M}_a\mathcal{M}_c^*=\frac{16g^4}{(M_{\tilde{q}L}^2-t)(M_{\tilde{q}L}^2-u)}&\{2a^2b^2[(p_1k_2)(p_2k_1)+(k_1p_1)(k_2p_2)-(k_1k_2)(p_1p_2)+m_q^2m_X^2+m_q^2(p_1p_2)]\\&+(a^4+b^4)m_X^2(k_1k_2)+m_qm_Xab\epsilon(a^2+b^2)[(p_2k_1)+(p_1k_2)+(p_1k_1)+(p_2k_2)]\}\ ,\end{aligned}$$

$$\begin{aligned}2\,\mathrm{Re}\mathcal{M}_a\mathcal{M}_d^*=\frac{16g^4}{(M_{\tilde{q}L}^2-t)(M_{\tilde{q}R}^2-u)}&(-2a^2bc[(p_1k_2)(p_2k_1)+(k_1p_1)(k_2p_2)-(k_1k_2)(p_1p_2)+m_q^2m_X^2]\\&+a^2(b^2+c^2)m_X^2(k_1k_2)+m_q^2(a^4+b^2c^2)(p_1p_2)\\&+m_qm_Xa\epsilon\{b(a^2+c^2)[(p_2k_1)+(p_1k_2)]-c(a^2+b^2)[(p_1k_1)+(p_2k_2)]\})\ ,\end{aligned}\tag{A2}$$

$$\begin{aligned}2\,\mathrm{Re}\mathcal{M}_b\mathcal{M}_c^*=\frac{16g^4}{(M_{\tilde{q}R}^2-t)(M_{\tilde{q}L}^2-u)}&(-2a^2bc[(p_1k_2)(p_2k_1)+(k_1p_1)(k_2p_2)-(k_1k_2)(p_1p_2)+m_q^2m_X^2]\\&+a^2(b^2+c^2)m_X^2(k_1k_2)+m_q^2(a^4+b^2c^2)(p_1p_2)\\&+m_qm_Xa\epsilon\{-c(a^2+b^2)[(p_2k_1)+(p_1k_2)]+b(a^2+c^2)[(p_1k_1)+(p_2k_2)]\})\ ,\end{aligned}$$

$$\begin{aligned}2\,\mathrm{Re}\mathcal{M}_b\mathcal{M}_d^*=\frac{16g^4}{(M_{\tilde{q}R}^2-t)(M_{\tilde{q}R}^2-u)}&\{2a^2c^2[(p_1k_2)(p_2k_1)+(k_1p_1)(k_2p_2)-(k_1k_2)(p_1p_2)+m_q^2m_X^2+m_q^2(p_1p_2)]\\&+(a^4+c^4)m_X^2(k_1k_2)-m_qm_Xac\epsilon(a^2+c^2)[(p_2k_1)+(p_1k_2)+(p_1k_1)+(p_2k_2)]\}\ ,\end{aligned}$$

$$2\,\mathrm{Re}\mathcal{M}_c\mathcal{M}_d^*=\frac{32g^4}{(M_{\tilde{q}L}^2-u)(M_{\tilde{q}R}^2-u)}[a(c-b)(k_1p_2)+m_qm_X\epsilon(bc-a^2)][a(c-b)(k_2p_1)+m_qm_X\epsilon(bc-a^2)]\ ,$$

$$\begin{aligned}2\,\mathrm{Re}\mathcal{M}_a\mathcal{M}_z^*=\frac{8g^4(Z_{13}^2-Z_{14}^2)}{\cos^2\theta_w(m_Z^2-s)(M_{\tilde{q}L}^2-t)}&\{2(b^2c_L-a^2c_R)(k_1p_1)(k_2p_2)+2(a^2c_L-b^2c_R)m_q^2m_X^2\\&-(a^2c_L-b^2c_R)m_q^2(p_1p_2)-(b^2c_L-a^2c_R)m_X^2(k_1k_2)\\&+ab\epsilon m_qm_X(c_R-c_L)[(p_1k_2)+(p_2k_1)-2(k_1p_1)-2(k_2p_2)]\}\ ,\end{aligned}$$

$$\begin{aligned}2\,\mathrm{Re}\mathcal{M}_b\mathcal{M}_z^*=\frac{8g^4(Z_{13}^2-Z_{14}^2)}{\cos^2\theta_w(m_Z^2-s)(M_{\tilde{q}R}^2-t)}&\{2(a^2c_L-c^2c_R)(k_1p_1)(k_2p_2)+2(c^2c_L-a^2c_R)m_q^2m_X^2\\&-(c^2c_L-a^2c_R)m_q^2(p_1p_2)-(a^2c_L-c^2c_R)m_X^2(k_1k_2)\\&+ac\epsilon m_qm_X(c_R-c_L)[-(p_1k_2)-(p_2k_1)+2(k_1p_1)+2(k_2p_2)]\}\ ,\end{aligned}$$

$$2\,\mathrm{Re}\mathcal{M}_c\mathcal{M}_z^* = \frac{8g^4(Z_{13}^2 - Z_{14}^2)}{\cos^2\theta_w (m_Z^2 - s)(M_{\tilde{q}L}^2 - u)} \{2(a^2c_R - b^2c_L)(k_1p_2)(k_2p_1) + 2(b^2c_R - a^2c_L)m_q^2 m_{\tilde{X}}^2$$
$$-(b^2c_R - a^2c_L)m_q^2(p_1p_2) - (a^2c_R - b^2c_L)m_{\tilde{X}}^2(k_1k_2)$$
$$+ab\epsilon m_q m_X(c_R - c_L)[2(p_1k_2) + 2(p_2k_1) - (k_1p_1) - (k_2p_2)]\}\ ,$$

$$2\,\mathrm{Re}\mathcal{M}_d\mathcal{M}_z^* = \frac{8g^2(Z_{13}^2 - Z_{14}^2)}{\cos^2\theta_w (m_Z^2 - s)(M_{\tilde{q}R}^2 - u)} \{2(c^2c_R - a^2c_L)(k_1p_2)(k_2p_1) + 2(a^2c_R - c^2c_L)m_q^2 m_{\tilde{X}}^2$$
$$-(a^2c_R - c^2c_L)m_q^2(p_1p_2) - (c^2c_R - a^2c_L)m_{\tilde{X}}^2(k_1k_2)$$
$$+ac\epsilon m_q m_X(c_R - c_L)[-2(p_1k_2) - 2(p_2k_1) + (k_1p_1) + (k_2p_2)]\}\ ,$$

where all symbols were defined in Sec. II.

---

[1]M. W. Goodman and E. Witten, Phys. Rev. D **31**, 3059 (1985); I. Wasserman, *ibid.* **33**, 2071 (1986).

[2]A. K. Drukier, K. Freese, and D. N. Spergel, Phys. Rev. D **33**, 3495 (1986).

[3]K. Freese, J. Frieman, and A. Gould, Phys. Rev. D **37**, 3388 (1988).

[4]K. Griest, Phys. Rev. D **37**, 2703 (1988).

[5]J. Silk, K. A. Olive, and M. Srednicki, Phys. Rev. Lett. **53**, 624 (1985); T. K. Gaisser, G. Steigman, and S. Tilav, Phys. Rev. D **34**, 2206 (1986); J. S. Hagelin, K. W. Ng, and K. A. Olive, Phys. Lett. B **180**, 375 (1987); M. Srednicki, K. A. Olive, and J. Silk, Nucl. Phys. **B279**, 804 (1987).

[6]L. M. Krauss, M. Srednicki, and F. Wilczek, Phys. Rev. D **33**, 2079 (1986); K. Freese, Phys. Lett. **167B**, 295 (1986).

[7]H. E. Haber and G. L. Kane, Phys. Rep. **117**, 75 (1985).

[8]J. Ellis, J. S. Hagelin, D. V. Nanopoulos, K. A. Olive, and M. Srednicki, Nucl. Phys. **B238**, 453 (1984).

[9]Other possible LSP candidates include the sneutrino and gravitino. They are more difficult to consistently arrange as dark matter and we will not consider them here.

[10]M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, Phys. Lett. **78B**, 443 (1978); A. I. Vainshtein, V. I. Zakharov, and M. A. Shifman, Usp. Fiz. Nauk. **130**, 537 (1980) [Sov. Phys. Usp. **23**, 429 (1980)].

[11]K. Griest and B. Sadoulet (unpublished).

[12]C. Albajar *et al.*, Phys. Lett. B **198**, 261 (1987).

[13]W. Bartel *et al.*, Z. Phys. C **29**, 505 (1985); H.-J. Behrend *et al.*, *ibid.* **35**, 181 (1987).

[14]C. Hearty *et al.*, Phys. Rev. Lett. **58**, 1711 (1987); G. Bartha *et al.*, *ibid.* **56**, 685 (1986).

[15]J. Ashman *et al.*, Phys. Lett. B **206**, 364 (1988).

[16]J. F. Gunion and H. E. Haber, Nucl. Phys. **B272**, 1 (1986).

[17]We follow the convention in which H_1 gives mass to the down-type quarks and H_2 gives mass to the up type. This is the opposite of the Ellis *et al.* (Ref. 8) convention, implying $Z_{13} \leftrightarrow Z_{14}$ and $\tan\beta \rightarrow v_1/v_2$. Also, our μ corresponds to their $-\epsilon$.

[18]J. Ellis and R. A. Flores, CERN Report No. CERN-TH. 4911/87, 1988 (unpublished).

[19]G. L. Kane and I. Kani, Nucl. Phys. **B277**, 525 (1986); B. A. Campbell, *et al.*, Phys. Lett. B **173**, 270 (1986).

[20]K. A. Olive, M. Srednicki, and R. Watkins, University of Minnesota Report No. UMN-TH-646/88 (unpublished).

[21]A Bayesian analysis gives instead $\sigma(e^+e^- \rightarrow \tilde{\bar{X}}\tilde{X}\gamma) \lesssim 0.046$ pb, but both statistical methods have problems so we use the simpler one.

[22]H.-J. Behrend *et al.*, Phys. Lett. B **176**, 247 (1986).

[23]P. Fayet, Phys. Lett. **117B**, 461 (1982); K. Grassie and P. N. Pandita, Phys. Rev. D **30**, 22 (1984).

[24]M. Dress, M. C. Kim, and X. Tata, Phys. Rev. D **37**, 784 (1988).

[25]J. Ellis, R. A. Flores, and S. Ritz, Phys. Lett. B **198**, 393 (1987).

[26]A. de-Shalit and I. Talmi, *Nuclear Shell Theory* (Academic, New York, 1963).

[27]S. Raby and G. B. West, Phys. Lett. B **202**, 47 (1988).

[28]B. W. Lee and S. Weinberg, Phys. Rev. Lett. **39**, 165 (1977); Ya. B. Zeldovich, L. B. Okun, and S. Pikel'ner, Usp. Fiz. Nauk. **87**, 113 (1965) [Sov. Phys. Ups. **8**, 702 (1966)].

[29]K. Griest and D. Seckel, Nucl. Phys. **B283**, 681 (1987); **B296**, 1034(E) (1988).

[30]K. Griest, Ph.D. thesis, University of California, Report No. SCIPP 87/92, 1987.

[31]E. W. Kolb and K. A. Olive, Phys: Rev. D **33**, 1202 (1986); R. J. Scherrer and M. S. Turner, *ibid.* **33**, 1585 (1986).

[32]D. O. Caldwell *et al.*, University of California, Santa Barbara Report No. UCSB-HEP-87-13 (unpublished); S. P. Ahlen *et al.*, Phys. Lett. B **195**, 603 (1987).

[33]K. Griest, Report No. FERMILAB-Pub-88/52-A, 1988 (unpublished).

[34]C. Hearty (private communication).

[35]J. Grivaz, in *The Standard Model: Supernova 1987A*, proceedings of the XXIInd Rencontre de Moriond, Starbursts, and Galaxy Evaluation, Les Arcs, France, 1987, edited by J. Tran Thanh Van (Editions Frontíeres, Gif-sur-Yvette, 1987), p. 115.

[36]L. E. Ibanez and J. Mas, Nucl. Phys. **B286**, 107 (1987); L. E. Ibanez, C. Lopez, and C. Munoz, *ibid.* **B256**, 218 (1985); L. E. Ibanez and C. Lopez, *ibid.* **B233**, 511 (1984).

Fermilab-Pub-89/139-A

ERRATUM

Kim Griest

Cross Sections, Relic Abundance, and Detection Rates for Neutralino Dark Matter, Phys. Rev. D 38 (1988) 2357.

At least three typographical errors appear in eq. (4), the annihilation cross section of neutralinos into fermions. In the last line $-\frac{2}{3}$, should read $\frac{2}{3}$, and $v'c_R$ should read $-v'c_R$. The definition of r should read $r = m_X^2/(M_{\tilde{q}}^2 + m_X^2\beta')$, not as given.

More importantly, it was pointed out by Lars Bergstrom that in our paper we (inadvertently) left out the $-p^\mu p^\nu/m_Z^2$ part of the Z^0 propagator. Including this gives additional terms in the matrix element and the annihilation cross section. For much of the supersymmetric parameter space the resulting corrections are very small, but for pure Higgsinos a maximum correction of around 9% can occur. There are no corrections to the elastic-scattering cross sections, the pure photino limit of the annihilation cross section or to the production ($e^+e^- \to \tilde{\chi}\tilde{\chi}$) cross sections. The changes to the annihilation cross section can result in up to a 9% change in the dark-matter detection rates. However, as Bergstrom and Snellman [1] point out, when using this annihilation cross section for present day annihilation of dark matter in the halo (or in the body of the earth or sun), the additional terms can be quite important. In particular, at the Z^0 pole ($m_X = \frac{1}{2}m_Z$), the term proportional to $(m_q/m_X)^2$ cancels. Since the remaining terms are proportional to $v^2 \approx 10^{-6}$, there is a very large dip in the cross section here rather than a large enhancement which the published cross section would predict. This dip does not occur for annihilation in the early universe where $v^2 \approx \frac{1}{3}$. The following corrections should be made.

The squared matrix element [eq. (3)] should include the additional terms

$$
\begin{aligned}
|\mathcal{M}'|^2 = 16g^4 \Bigg\{ & \frac{(Z_{13}^2 - Z_{14}^2)^2 (c_L - c_R)^2}{\cos^4\theta_w \,(m_Z^2 - s)^2} \frac{m_X^2 m_q^2}{m_Z^4} \left(\frac{s^2}{4} - \frac{s m_Z^2}{2} \right) + \frac{(Z_{13}^2 - Z_{14}^2)(c_L - c_R)}{4\cos^2\theta_w \,(m_Z^2 - s)} \frac{m_X m_q}{m_Z^2} \\
& \times \Bigg[\left(\frac{1}{M_{\tilde{q}}^2 - t} + \frac{1}{M_{\tilde{q}}^2 - u} \right) \left[w' s^2 + 2(u' + v') s m_X m_q \right] \\
& + \left(\frac{1}{M_{\tilde{q}}^2 - t} - \frac{1}{M_{\tilde{q}}^2 - u} \right) (4w') \left[(p_1 k_1)^2 - (p_1 k_2)^2 \right] \Bigg] \Bigg\},
\end{aligned}
\tag{1}
$$

where all symbols were defined in the paper.

The non-relativistic expansion of the neutralino annihilation cross section (eq. 4) should include the additional terms

$$\sigma'_{\rm ann} v = \sum_q \frac{4}{\pi} G_F^2 c_q m_X^2 \beta' \Bigg\{ \left(Z_{13}^2 - Z_{14}^2\right)^2 x'^4 \frac{(c_L - c_R)^2 z^2}{4} \frac{m_X^2}{m_Z^2}$$
$$\times \left[\frac{16 m_X^2}{m_Z^2} - 8 + 2v^2 \left(\frac{2m_X^2}{m_Z^2}(1 + x^2) - x^2 \right) \right]$$
$$+ \left(Z_{13}^2 - Z_{14}^2\right)(c_L - c_R) x'^2 y'^2 z \frac{m_X^2}{m_Z^2}$$
$$\times \left[\left(2 + \tfrac{1}{2} v^2 \left(x^2 - 2r + \tfrac{4}{3}\beta'^2 r^2\right)\right) \right.$$
$$\left. \times (4w' + 2z(u' + v')) + 2w' v^2 - \tfrac{4}{3} w' v^2 \beta'^2 r \right] \Bigg\}. \qquad (2)$$

In the limit $v^2 = 0$, relevant for present day annihilation, the total cross section is simply

$$\sigma_{\rm ann}^{\rm tot}(v^2 = 0) = \sum_q \frac{4}{\pi} G_F^2 c_q m_X^2 \beta'$$
$$\times \left[y'^2 (2w' + z(u' + v')) + \tfrac{1}{2} x'^2 \left(Z_{13}^2 - Z_{14}^2\right)(c_R - c_L) z \left(1 - 4\frac{m_X^2}{m_Z^2}\right) \right]^2. \qquad (3)$$

Finally, for completeness, eq. (A1) should include

$$\mathcal{M}'_z = \frac{-2 m_X m_q g^2 (c_R - c_L)\left(Z_{13}^2 - Z_{14}^2\right)}{2\cos^2\theta_w \left(m_Z^2 - s\right) m_Z^2} \bar{v}(p_2)\gamma_5 u(p_1)\bar{u}(k_1)\gamma_5 v(k_2), \qquad (A1')$$

and eq. (A2) should contain the following terms

$$|\mathcal{M}'_z|^2 = \frac{16 g^4 \left(Z_{13}^2 - Z_{14}^2\right)^2}{\cos^4\theta_w \left(m_Z^2 - s\right)^2} (c_R - c_L)^2 \frac{m_X^2 m_q^2}{m_Z^4} \left(\frac{s^2}{4}\right),$$

$$2\,\mathrm{Re}\,\mathcal{M}_z \mathcal{M}'^*_z = -\frac{8 g^4 \left(Z_{13}^2 - Z_{14}^2\right)^2}{\cos^4\theta_w \left(m_Z^2 - s\right)^2} (c_R - c_L)^2 \frac{m_X^2 m_q^2 s}{m_Z^2},$$

$$2\,\mathrm{Re}\,\mathcal{M}_a \mathcal{M}'^*_z = \frac{-8 g^4 \left(Z_{13}^2 - Z_{14}^2\right)(c_R - c_L) m_X m_q}{\cos^2\theta_w \, m_Z^2 \left(m_Z^2 - s\right)\left(M_{\tilde{q}L}^2 - t\right)}$$
$$\times \left(2\epsilon ab \left[\frac{s^2}{4} + (p_1 k_1)^2 - (p_1 k_2)^2 \right] + s m_q m_X (a^2 + b^2) \right),$$

$$2\,\mathrm{Re}\,\mathcal{M}_b\mathcal{M}_z'^{*} = \frac{-8g^4\left(Z_{13}^2 - Z_{14}^2\right)(c_R - c_L)m_X m_q}{\cos^2\theta_w\, m_Z^2\left(m_Z^2 - s\right)\left(M_{\tilde{q}R}^2 - t\right)}$$

$$\times\left(-2\epsilon ac\left[\frac{s^2}{4} + (p_1k_1)^2 - (p_1k_2)^2\right] + sm_q m_X(a^2 + c^2)\right), \qquad \text{(A2')}$$

$$2\,\mathrm{Re}\,\mathcal{M}_c\mathcal{M}_z'^{*} = \frac{-8g^4\left(Z_{13}^2 - Z_{14}^2\right)(c_R - c_L)m_X m_q}{\cos^2\theta_w\, m_Z^2\left(m_Z^2 - s\right)\left(M_{\tilde{q}L}^2 - u\right)}$$

$$\times\left(2\epsilon ab\left[\frac{s^2}{4} - (p_1k_1)^2 + (p_1k_2)^2\right] + sm_q m_X(a^2 + b^2)\right),$$

$$2\,\mathrm{Re}\,\mathcal{M}_d\mathcal{M}_z'^{*} = \frac{-8g^4\left(Z_{13}^2 - Z_{14}^2\right)(c_R - c_L)m_X m_q}{\cos^2\theta_w\, m_Z^2\left(m_Z^2 - s\right)\left(M_{\tilde{q}R}^2 - u\right)}$$

$$\times\left(-2\epsilon ac\left[\frac{s^2}{4} - (p_1k_1)^2 + (p_1k_2)^2\right] + sm_q m_X(a^2 + c^2)\right).$$

Note that these should be added with signs appropriate for $\mathcal{M} = \mathcal{M}_z + \mathcal{M}_z' - \mathcal{M}_a - \mathcal{M}_b + \mathcal{M}_c + \mathcal{M}_d$.

We would like to thank Marc Kamionkowski for help with this erratum. This work was supported in part by the DoE (at Chicago) and by the DoE and NASA (grant NAGW − 1340) (at Fermilab).

References

[1] L. Bergstrom, preprint USITP-88-12 (1988);
L. Bergstrom and H. Snellman, Phys. Rev. D 37 (1988) 3737.

LIMITS ON COLD DARK MATTER CANDIDATES FROM AN ULTRALOW BACKGROUND GERMANIUM SPECTROMETER

S.P. AHLEN [a], F.T. AVIGNONE III [b], R.L. BRODZINSKI [c], A.K. DRUKIER [d,e], G. GELMINI [f,g,1] and D.N. SPERGEL [d,h]

[a] *Department of Physics, Boston University, Boston, MA 02215, USA*
[b] *Department of Physics, University of South Carolina, Columbia, SC 29208, USA*
[c] *Pacific Northwest Laboratory, Richland, WA 99352, USA*
[d] *Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA*
[e] *Applied Research Corp., 8201 Corporate Dr., Landover MD 20785, USA*
[f] *Department of Physics, Harvard University, Cambridge, MA 02138, USA*
[g] *The Enrico Fermi Institute, University of Chicago, Chicago, IL 60637, USA*
[h] *Institute for Advanced Study, Princeton, NJ 08540, USA*

Received 5 May 1987

An ultralow background spectrometer is used as a detector of cold dark matter candidates from the halo of our galaxy. Using a realistic model for the galactic halo, large regions of the mass–cross section space are excluded for important halo component particles. In particular, a halo dominated by heavy standard Dirac neutrinos (taken as an example of particles with spin-independent Z^0 exchange interactions) with masses between 20 GeV and 1 TeV is excluded. The local density of heavy standard Dirac neutrinos is <0.4 GeV/cm^3 for masses between 17.5 GeV and 2.5 TeV, at the 68% confidence level.

Galactic rotation curves suggest that most of the matter in the universe is non-luminous [1]. A variety of arguments suggest that this matter may be non-baryonic [2]. This letter discusses the use of an ultralow background germanium spectrometer as a detector of cold dark matter particles interacting with Ge nuclei. Since only ^{73}Ge, with a natural abundance of 7.8%, has a non-zero spin, our best bounds apply to spin-independent (s.i.) interactions. Bounds on dark matter candidates coupling to baryons through Z^0 exchange, like stable massive Dirac neutrinos [3] and scalar neutrinos [4], are presented. Our results exclude a halo dominated by particles with scattering cross section $\sigma^{\text{s.i.}} = \sigma_{\text{weak}}$ with masses 20 GeV $\leqslant m \leqslant 1$ TeV (their local density is <0.4 GeV/cm^3 for 17.5 GeV $\leqslant m \leqslant 2.5$ TeV at the 68% confidence level) and apply to s.i. reactions in the range of $\sigma^{\text{s.i.}} \simeq 10^{-1}\sigma_{\text{weak}}$ to $\sigma^{\text{s.i.}} \approx 10^{-28}$ cm^2 (for which the dark matter particles would be stopped in the earth's crust before arriving at the detector) where σ_{weak} is the weak scattering cross section of a standard heavy Dirac neutrino from a Ge nucleus. This range includes neutral technibaryons, recently proposed as dark matter candidates [5], having cross sections $\simeq 10\sigma_{\text{weak}}$, which are, therefore, excluded for masses larger than 16 GeV. The ^{73}Ge in the detector with $s = 9/2$, allows us to obtain a bound on particles with spin-dependent (s.d.) interactions, which case applies to particles in the range $\sigma \simeq 10^4 \sigma^{\text{s.d.}}_{\text{weak}}$ to $\sigma \approx 10^{-28}$ cm^2 (where $\sigma^{\text{s.d.}}_{\text{weak}}$ corresponds to a standard heavy Majorana neutrino).

The measurement of the nuclear recoil, due to the scattering of heavy weakly interacting massive particles (WIMPs), requires a detector with a low energy threshold and excellent background rejection [6–8]. In this paper, the use of a germanium diode detector to search for dark matter is discussed. The low band gap (0.69 eV at 77 K) and high efficiency for converting electronic energy loss to electron–hole (e–h) pairs (2.96 eV per e–h pair at 77 K) make germanium detectors probably the best existing detectors

[1] On leave of absence from Department of Physics, University of Rome II, Via Orazio Raimondo, I-00173 Rome, Italy.

0370-2693/87/$ 03.50 © Elsevier Science Publishers B.V.
(North-Holland Physics Publishing Division)

of low velocity recoiling nuclei. Their low intrinsic noise (≈ 500 eV for equivalent electron energies of order 10 keV) make them ideal for the search for the rare phenomena considered in this paper as well as other exotica such as neutrinoless double beta decay.

An important parameter for the interpretation of count rates of recoiling nuclei in Ge detectors is the relative efficiency factor (R.E.F). This is the ratio of the number of e–h pairs produced by an incident electron with energy, T, to the number of e–h pairs produced by an incident Ge nucleus with the same energy. The R.E.F. is equal to 1 only for $T \gg 1$ MeV. For example, a uranium nucleus with energy 6 MeV, incident on a silicon diode, produces a signal equivalent to a 2 MeV electron. This "pulse height defect" is attributed to (i) the energy loss of the ion in the electrode or detector dead layer; (ii) electron–hole recombination in regions of intense energy loss and/or poor collection of electrons; and (iii) competition between electronic and nuclear stopping powers. The first effect is not important for Ge recoils produced internally. Since the electronic energy loss is small for low velocities, the second effect is also not significant. Effect (iii) will therefore dominate in our case.

The R.E.F. for Ge detectors was calculated by evaluating the fraction of primary Ge recoil energy lost in electronic collisions. The electronic energy loss of secondary and higher order recoil nuclei was accounted for, and current electronic and nuclear stopping cross sections [9] have been used. Results are shown in fig. 1 (as function of the Ge nuclei energy recoil). The solid curve is based on the assumption that electronic energy loss is proportional to velocity at arbitrarily small velocities (i.e. the band gap is ignored), while the dashed curve is the result for a kinetic threshold energy of 0.27 keV (a Ge nucleus with this energy transfers a maximum

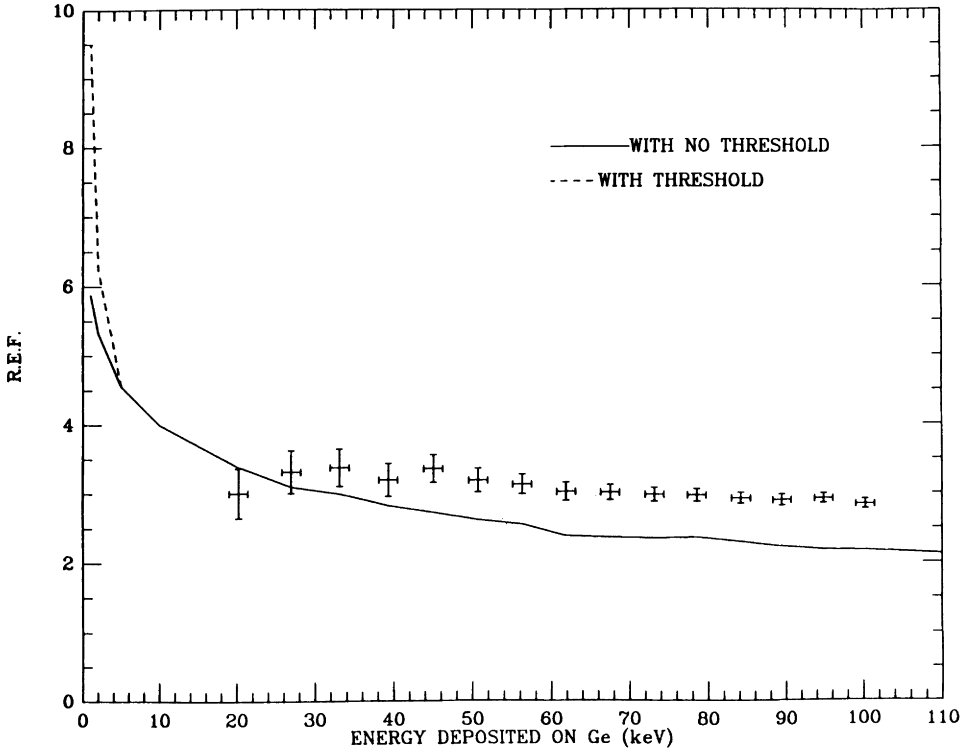


Fig. 1. The R.E.F. (the ratio of the number of electron hole-pairs produced by an incident electron to the number produced by a recoiling Ge nucleus with the same energy) is shown as function of Ge recoil energy. The solid curve is based on the assumption that there is no kinematic threshold, while the dashed curve shows the results for a kinematic recoil energy threshold of 0.27 keV. The theoretical curve is compared to data from Chasman et al. [10].

energy of 0.7 eV to an electron in a direct collision, i.e. the band gap is regarded as a sharp threshold). In this way the sensitivity of the calculations to threshold phenomena is tested. The theoretical values were compared to data from Chasman et al. [10] who measured e–h yields of neutron induced recoil nuclei within a Ge detector (fig. 2). Good agreement between the calculation and the data gives us confidence that the theoretical predictions should be valid down to several keV. A similar experiment, on silicon instead of Ge, by Sattler [11], and other calculations of the expected response in Ge and Si, by Linhard et al. [12], support our results. Linhard's calculations, which do not include higher than second-order nuclear collisions and use older data on cross sections, give lower values for the R.E.F. of Ge than the present calculation, by at most 12% for nuclear recoil energies between 20 and 100 keV.

The PNL/USC group has developed a 135 cm^3 intrinsic Ge detector [13,14] having a background reduced by about three orders of magnitude over conventional low background gamma-ray spectrometers. The detector is located in the Homestake mine at a depth equivalent to 4000 m of water to eliminate the cosmic ray induced background. The detector cryostat is constructed from high-purity copper and is surrounded by 11 tons of lead, sheet cadmium and neutron moderator, to eliminate the radioactive background and neutrons from the rock. The inner shield was made from high purity copper, when the 14 d of data used in this work were taken. These data were selected because they correspond to a period of decreased level of mining operations in the vicinity of the detector. This resulted in fewer microphonic noise pulses. For this analysis, the absence of low energy noise is more important than the quantity of data. However, 1000 h of data are shown in fig. 2 to better display the X-ray and γ-ray peaks used for energy calibration.

The energy threshold was reduced to an incident electron energy of 4 keV. According to fig. 1, this permitted the detection of Ge nuclei–WIMP scatterings with nuclear recoil energies greater than 15 keV. The count rates of 1000 h of low energy data are shown in fig. 2, as function of incident electron energy (to be multiplied by the appropriate R.E.F to get the corresponding nuclear recoil energy). The following photon peaks, clearly in evidence, were

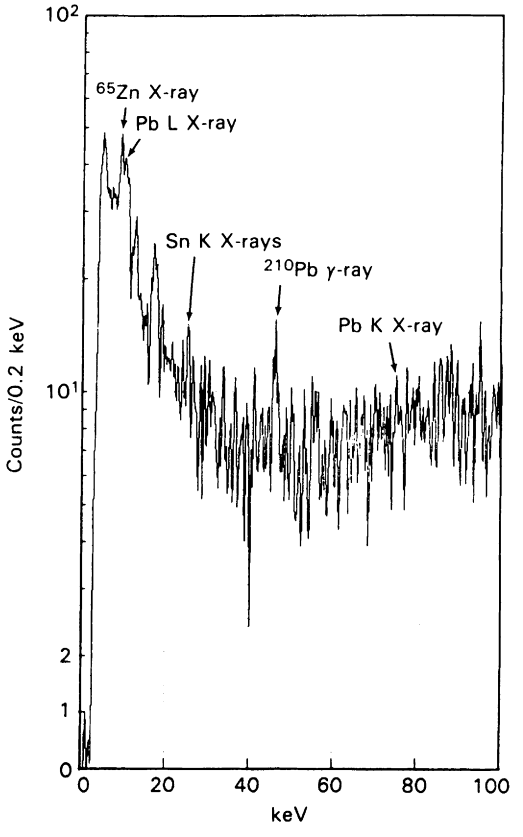


Fig. 2. 1000 h of data from the Ge spectrometer are shown. The width of each channel is 0.2 keV. The identified peaks result from the decay products of radioactivity in the exposed solder.

used to calibrate the detector: the 8.9805 keV Cu X-ray from the electron capture of ^{65}Zn, the 10.551 keV Ga X-ray from the electron capture of ^{68}Ge plus Pb-L, the 25.2713 keV Sn K X-ray, the 46.503 keV γ-ray in ^{210}Bi from the decay of ^{210}Pb and the 74.969 keV Pb X-ray. The Pb and Sn X-rays came from a solder connection in the proximity of the detector. The Cu and Ga X-rays are sums of X-rays from all shells, since the sources were cosmogenically created within the crystal. Each line was fitted with a modified gaussian to obtain its centroid. Then the ratio of the channel to energy of the peak was fitted to a quadratic function. There is a 0.1% non-linearity. The spectrum was then adjusted until the best fit was achieved. The non-linearity accounts for about one channel in one thousand. The maximum error in the calibration at 4 keV is one half of a channel, which is 0.1 keV.

The observed count rate can be compared to the rate predicted if the halo was comprised of WIMPs. The model of the halo used assumes a local halo density of 0.01 solar mass/pc^3 = 0.38 GeV/cm^3 [15], and an isotropic (in the halo rest frame) gaussian distribution function of velocities of the halo particles, $f(v)$, with an RMS of 250 km/s and a maximum of 550 km/s [8]. It is assumed that the halo, slowly rotates with a local velocity of 70 km/s [16], like the galactic spheroid. This is a conservative assumption which reduces relative velocities since the Sun moves around the galaxy in the same sense, at 250 km/s; the halo may not rotate at all. The maximum halo velocity may be higher [15] in which case the limits extend to lower masses. The local halo density is known with a factor of 2 uncertainty. The predicted rate of recoils having energy T and producing a signal in the detector, $R_p(T)$, was calculated according to

$$R_p(T) = n_X \Delta T \int_{v_{min}(T)}^{v_{max}} \frac{d\sigma}{dT}(v, T) f(v) v \, d^3 v , \qquad (1)$$

where ΔT is the range of recoil energies detected in a given channel centered in T and n_X is the local density of dark matter. Standard Dirac neutrinos were chosen as an example of coherently interacting particles. The cross section as a function of recoil energy, $d\sigma/dT$, for an incident standard Dirac neutrino (that interacts through Z^0 exchange) depends upon its mass, m_x, and velocity, v, according to

$$\frac{d\sigma}{dT} \simeq \frac{G_F^2 m_N c^2}{8\pi v^2} [Z(1 - 4\sin^2\theta_W) - N]^2$$

$$\times \left[1 + \left(1 - \frac{T}{E_x} \right)^2 - \frac{m_N T + m_x^2}{E_x^2} \right]$$

$$\times \exp(-m_N 2TR^2/3\hbar^2) , \qquad (2)$$

where m_N is the mass of the nucleus, Z and N are the atomic and neutron numbers respectively, T is the recoil energy, G_F is the weak coupling constant, θ_W is the weak mixing angle and $E_x = m_x(c^2 + v^2/2)$. When the de Broglie wavelength $1/q$ of the momentum transfered in the recoil $g = (2m_N T)^{1/2}$, is smaller than the nucleus, the assumption of a coherent interaction with a point-like nucleus is no longer valid, and the finite size of the nucleus must be included [17]. The exponential in eq. (2) is a nuclear form factor corresponding to a gaussian density distribution of nucleons, in a nucleus of radius $R = 1.2A^{1/3}$ fm, and atomic number A [17]. A better model of the nucleonic density, such as a Woods–Saxon density [18], would produce a form factor decreasing less steeply with T. Note that for small energy transfers, $T \ll 11$ keV, such that $qR \ll 1$, the WIMP interacts with a point-like nucleus, and there is no loss of coherence. The halo model used is conservative. If the maximum halo velocity was closer to the local escape velocity of ~ 750 km/s [15], the lower limit of the excluded mass range would be 10 GeV.

The observed count rate in the germanium detector can be used to obtain limits on the density of interacting dark matter particles in the halo. The integral in eq. (1) was evaluated over all velocity phase space and bounds on n_X, at the 68% and 95% confidence levels, were obtained for every T. The most restrictive T dependent bounds were taken as the final bounds on n_X. The best bounds come from values of T near threshold, because of the rapid decrease of the predicted rate with increasing T (notice that in the non-relativistic limit $(d\sigma/dT)^{NR} = \sigma^{NR}/T_{max}$, where σ^{NR} does not depend on T, and for a given T one must sum over all $T_{max} > T$). Because of their importance in this analysis, the number of counts in the first ten channels (of 0.2 keV each), of the 14 d of data used, starting at 4.0 keV are given. They were respectively 4.53, 9.65, 19.38, 20.66, 15.78, 16.33, 16.84, 13.46, 12.01, 10.64. Fig. 3 shows the limits for standard heavy Dirac neutrinos, which have s.i. Z^0 exchange interactions. Fig. 3 can be used for other s.i. vectorial interactions by multiplying the vertical axis by the ratio $(\sigma_{weak}/\sigma^{s.i.})$. For example for neutral technibaryons this ratio would be approximately 0.1 and for sneutrinos this ratio is 0.5. Similar curves for s.d. interactions will be given in a subsequent paper.

The spectrum has a smooth continuum contribution, due mainly to Compton-scattered background γ-rays. The low energy peaks in the present data are primarily due to the presence of the ^{210}Pb decay chain, in a solder connection in direct line-of-sight to the surface of the detector. The solder and an indium contact have recently been removed, and the inner shield has been upgraded by the use of 449 y old lead in the place of the super-pure copper, which had some

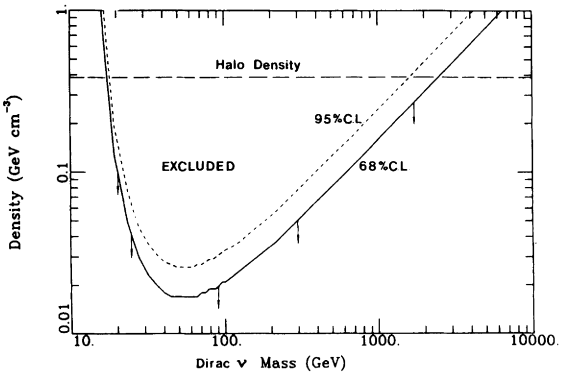


Fig. 3. The maximum halo density of heavy standard Dirac neutrinos (as an example of particles with weak spin-independent interactions) is shown, consistent with the observed count rate, as a function of their mass. The solid line shows the 68% confidence level and the dashed line shows the 95% confidence level.

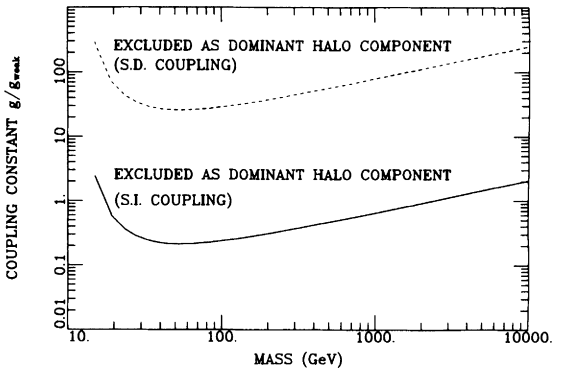


Fig. 4. The regions in mass–cross section space excluded at the 68% confidence level are shown. The halo cannot be composed of particles that interact with nuclei through spin-independent interactions whose coupling constant (normalized to the coupling of massive Dirac neutrinos to baryons) lies above the solid line. Nor can the halo be composed of particles that interact with nuclei through spin-dependent interactions whose coupling constant (normalized as above) lies above the dashed line.

cosmogenic radioactive contamination. In this way the background has been reduced by about a factor of ten at around 50 keV. The reduction of background appears to improve with decreasing energy; however, in an attempt to reduce background from ^{40}K, data are being acquired in the present phase with the field-effect transistor removed from the cryostat. This configuration increases the low energy noise (while not reducing the ^{40}K significantly). New data on dark matter will not be available untill the next phase of the experiment.

When the data for this work were taken, the energy threshold was set at 4 keV because of microphonic noise at lower energies. The shape of the low energy X-ray lines suggests that $\Delta E(\mathrm{FWHM}) \approx 500$ eV in this region. The strong increase of noise below $E_{\mathrm{th}} = 4$ keV is largely due to microphonics engendered by mining operations. Hardware and software have recently been developed to reduce this noise and permit lowering the energy threshold to 1 keV. In the near future, rejection/detection of the existence of coherently interacting particles of mass >8 GeV should be feasible. Detection could be confirmed by the expected modulation in the signal due to the earth's motion relative to the halo [8].

It will be difficult to reduce the energy threshold below 1 keV, thus the detection of particles of lower mass will require cryogenic detectors. The germanium detector is also not sensitive to particles like the photino that couple through s.d. weak interactions.

Our main results are shown in fig. 4. The range of mass and cross section of particles excluded as main components of the halo are shown. The ratio g/g_{W} is defined as $(\sigma/\sigma_{\mathrm{weak}})^{1/2}$ where σ_{weak} is the cross section for standard heavy Dirac neutrinos.

A limit on the solar axion flux has also been derived from the USC/PNL germanium spectrometer [19].

F.T.A. is supported by NSF grant PHY-8405654, R.L.B. by DOE Contract DE-AC06-76RLO-1830, G.G. by NSF grants PHY-82-15249 and PHY-85-21588 and DOE grant DE-AC02-82ER-40073, and DNS by NSF grants PHY-83-06693 and PHY-82-17352. We would like to thank R. Glauber and T. Jaroszewicz for useful discussions.

References

[1] S.M. Faber and J.S. Gallagher, Annu. Rev. Astron. Astrophys. 17 (1979) 135;
D. Burnstein and V.C. Rubin, Astrophys. J. 297 (1985) 423.
[2] D.J. Hegyi and K.A. Olive, Phys. Lett. B 126 (1983) 28; Astrophys. J. 303 (1986) 56.
[3] J. Bagger, S. Dimopoulos, E. Masso and J. Reno, Phys. Rev. Lett. 54 (1985) 2199;
E. Kolb and K. Olive, Phys. Rev. D 33 (1986) 1202; D 34 (1986) 2531.
[4] J.S. Hagelin, G.L. Kane and S. Raby, Nucl. Phys. B 241 (1984) 648; L.E. Ibáñez, Phys. Lett. B 137 (1984) 160.
[5] S. Nussinov, Phys. Lett. B 165 (1985) 55.

[6] M.W. Goodman and E. Witten, Phys. Rev. D 31 (1985) 3059; I. Wasserman, Phys. Rev. D 33 (1986) 2071.
[7] A.K. Drukier and L. Stodolsky, Phys. Rev. D 30 (1984) 2295.
[8] A.K. Drukier, K. Freese and D.N. Spergel, Phys. Rev. D 33 (1986) 3495.
[9] W.D. Wilson, L.G. Haggmark and J.P. Biersack, Phys. Rev. B 15 (1977) 2485.
[10] C. Chasman, K.W. Jones and R.A. Ristinen, Phys. Rev. Lett. 15 (1965) 245.
[11] R.A. Sattler, Phys. Rev. 138 (1965) A1815.
[12] J. Linhard, V. Nielsen, M. Scharff and P.V. Thompsen, Kgl. Danske Videnskab Selskab Mat. Fys. Medd. 33, No. 10 (1963).
[13] R.L. Brodzinski, D.P. Brown, J.C. Evans Jr., W.K. Hensley, J.H. Reeves, N.A. Wogman, F.T. Avignone III and H.S. Miley, Nucl. Instrum. Methods A 239 (1985) 207.
[14] F.T. Avignone III, R.L. Brodzinski, J.C. Evans Jr., W.K. Hensley, H.S. Miley and J.H. Reeves, Phys. Rev. C 34 (1986) 666.
[15] J. Caldwell and J.P. Ostriker, Astrophys. J. 251 (1981) 61.
[16] J. Bahcall and S. Casertano, Astrophys. J. 308 (1986) 347.
[17] D.Z. Freedman, Phys. Rev. D 9 (1974) 1389; D. Tubbs and D.N. Schramm, Astrophys. J. 201 (1975) 467.
[18] E.g. R.C. Barrett and D.F. Jackson, Nuclear sizes and structure (Clarendon, Oxford, 1979).
[19] F.T. Avignone III, R.L. Brodzinski, S. Dimopoulos, A.K. Drukier, G. Gelmini, B.W. Lynn, D.N. Spergel and G.D. Starkman, Limits on solar axions from the ultralow background germanium spectrometer, SLAC-Pub. 3872 and PNL-SA-14132(1986), Phys. Rev. D, to be published.

Laboratory Limits on Galactic Cold Dark Matter

D. O. Caldwell, R. M. Eisberg, D. M. Grumm, and M. S. Witherell
Department of Physics, University of California, Santa Barbara, California 93106

B. Sadoulet
Physics Department, University of California, Berkeley, California 94720

and

F. S. Goulding and A. R. Smith
Lawrence Berkeley Laboratory, Berkeley, California 94720
(Received 13 November 1987; revised manuscript received 16 May 1988)

Interesting limits are set on candidates for cold-dark-matter particles in the halo of our Galaxy from their interaction with a very-low-background Ge detector used to search for double-β decay. Dirac neutrinos constituting all of dark matter are excluded for masses between 12 GeV/c^2 and 1.4 TeV/c^2. There are slightly better limits on magninos and cosmions, proposed massive particles which also explain the solar-neutrino problem but which interact more strongly with Ge. In addition, millicharged shadow matter is ruled out as the main form of dark matter.

PACS numbers: 98.60.Ln, 14.60.Gh, 14.80.Pb, 96.60.Kx

Clearly one of the most important questions facing basic physics is the nature of the majority of the matter in the Universe. Since perhaps 90% of matter is detected only through its gravitational interactions, knowledge of what constitutes this dark matter would have widespread importance for astrophysics and cosmology. Since it is unlikely[1] that dark matter consists of particles as they are known in the standard model, the discovery of their nature would also point the way to important new particle physics. Useful constraints on candidates for dark matter can already be set by apparatus used to search for neutrinoless double-β decay. In order to approach lifetime limits[2] of 10^{24} yr on this process in ^{76}Ge, it has been necessary to take extreme measures to reduce backgrounds. When the energy threshold for the Ge detectors is made as low as possible, the resulting background rates can be used to set limits on the scattering of the dark-matter candidates from the Ge nuclei.[3] We report here limits on some specific models for dark matter, as well as giving some more general constraints from measurements with the UCSB/LBL double-β-decay detector.[4]

The apparatus has been described elsewhere,[4] but some relevant features will be mentioned here. Although up to eight Ge detectors have been used for double-β decay, the data of concern here comes from just one detector of about 160-cm^3 fiducial volume (0.9 kg). The Ge detectors were surrounded by ten blocks of NaI(Tl) scintillator, which formed a complete anticoincidence shield of 15 cm thickness. Compton scattering, which provides the main source of radioactivity background at low energies, is suppressed not only by the vetoing of the scattered γ in the NaI or another Ge detector, but also because many of the γ's come from cascade decays, and the detection of one of these other γ's would also eliminate the event. The threshold at 30-keV equivalent electron energy for the NaI counters is high enough so that the small energy deposits by the dark-matter candidates do not cause a veto signal, except for the case of large cross section. The NaI detectors are surrounded by borated polyethylene to degrade and capture neutrons and then by a 20-cm shield of 99.999% pure virgin lead.

Much effort has been devoted to the finding of structural materials of low radioactivity and to the development of fabrication procedures to reduce the amount of such materials. In order also to avoid losing γ's before they could be vetoed, it was particularly necessary to keep structural material to a minimum inside the NaI cavity and to have that used of as low an atomic number, Z, as possible. Thus the Ge detectors do not have the usual thermal shield, but instead an overlying metallization 1000 Å thick. They are enclosed in cylindrical cans of electroformed Cu with walls 0.25 mm thick, with each set of two cans attached to a similarly made flat box which passes through a slot in the NaI and serves as the vacuum cryostat, evacuated by ion pumps and containing the cold finger to keep the Ge at liquid-nitrogen temperature. Each 4-mm-thick cold finger is made from single-crystal Si, which has about twice the conductivity of Cu at this temperature, is of very high purity, and has low Z, reducing the absorption of γ rays. Before fabrication, a kilogram or more of material to be used was measured with a low-background Ge spectrometer with picocurie sensitivity, and after fabrication the piece was tested again.

When the apparatus was constructed, the emphasis was on achievement of low background and good Ge resolution near 2.041 MeV, the region to be searched for

© 1988 The American Physical Society

a peak from neutrinoless double-β decay in ^{76}Ge. However, when it was decided to look for dark matter, we found that there was a rapidly rising background below about 400 keV. This was due to the presence of about half a gram of In, which undergoes a 486-keV β^- decay with a half-life of 4×10^{14} yr! When the In was removed from one detector, the background for that detector became flat down to about 14 keV at a level of $\frac{1}{2}$ count/(keV kg day), except for some x-ray peaks. A part of that spectrum is shown in Fig. 1, and it will be noticed that at lower energies there is a rise in the counting rate due mainly to Ga (10.4 keV), and a little to Zn (9.7 keV) x-ray peaks which result from the ^{68}Ge-^{68}Ga decay chains.[2] The counting-rate drop at the low end of the x-ray peaks begins to be offset by the rapidly rising noise rate which currently limits measurements to about 3 keV. The noise has three sources: inadequate infrared shielding (since we avoided using a thermal shield on the Ge), electronic noise, and microphonics. Although further improvements in background will occur, certainly because of decay of the ^{68}Ge-^{68}Ga activity and probably also because of detector improvements, we already have lower backgrounds in the low-energy region than have heretofore been available. Hence, we have utilized those data to set useful limits on several dark-matter candidates.

An interesting speculation[5] which has motivation from string theory is that another world may occupy the space in which we live and that this shadow universe interacts with us only gravitationally. However, Holdom[6] has pointed out that if any connector particle exists, no matter how heavy, then the shadow world has a massless photon which mixes with our photon, providing a way in which shadow photons can interact with us through a very weak electric charge. Thus shadow matter would emit mainly shadow photons, which we could not detect, but also our photons with strength $\propto\epsilon^2$, where perhaps $\epsilon\sim10^{-3}$. Goldberg and Hall[7] have investigated whether such millicharged shadow matter could be a candidate for the dark matter. Although their model is constrained by big-bang cosmology, galactic astrophysics, and searches for fluxes of dark-matter particles, including limits set by another Ge detector,[8] there appeared to be a range of allowable masses of the shadow proton versus the shadow electron. The Ge limits[8] came from observations both at the surface and at a depth of 4000 m of water equivalent (mwe) where the count rates were $\simeq100$ kg^{-1} day^{-1} for Ge nucleus recoil energies which give ionization energies of 20–150 keV. Over this same energy range our count rates were about $\frac{1}{3}$ of that, but what is really significant in this case is that our apparatus is at a depth of only about 600 mwe. The 4000-mwe result[8] was being satisfied[7] by the requirement that the interaction cross section, and hence the mass, of the shadow proton be sufficiently large so that too few such particles survived to that depth to be counted. Scaling the Goldberg-Hall calculations to 600 mwe gives a result that requires a further increase in cross section by about an order of magnitude. This, then, effectively wipes out the allowable region in shadow proton versus shadow electron mass. The only faint possibility remaining is if the shadow electron's mass is less than the value of $\frac{1}{3}$ GeV/c^2 at which the calculations break down, there being no predictions in that region with which to make comparisons.

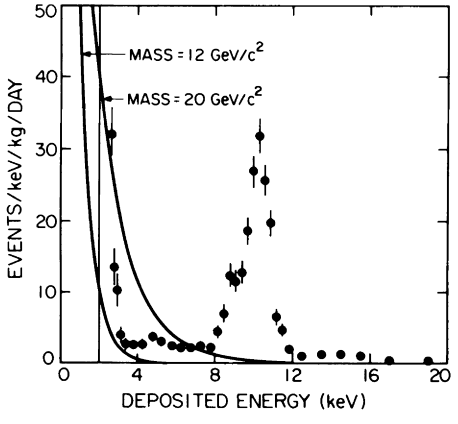


FIG. 1. Ionization energy detected at the low end of the operating range of one Ge counter in a four-week run. Below the Ga and Zn x-ray peaks the noise is almost flat until 3 keV. The threshold was set at 2.0 keV. The two curves show the contributions expected for particles of mass 12 and 20 GeV/c^2, which interact via Z^0 exchange. If we take into account the almost flat background in the 3–8-keV region, the 12-GeV/c^2 curve is ruled out at the 95% confidence level.

Another candidate for dark matter is the class of weakly interacting massive particles (WIMP's) which interact via Z^0 exchange. This vectorial, spin-independent interaction is coherent in a nucleus and so depends on the square of the number of neutrons, N^2. If the coupling of these particles is similar to that of Dirac neutrinos and their mass is greater than a few GeV/c^2, the resulting large annihilation cross section necessitates, at masses above a few GeV/c^2 particle-antiparticle asymmetry for them to be responsible for all of dark matter.[9] An asymmetry similar to that of baryons would require a mass of the order of 10 GeV/c^2.

For such a particle of mass m scattering from a Ge nucleus of mass M, the average energy deposition at low energy is $\langle E_d\rangle = m^2M\langle v^2\rangle/(m+M)^2$, where $\langle v^2\rangle$ is the mean square speed of m. For these low values of E_d, the recoiling nucleus ionizes very inefficiently and gives an ionization signal much smaller ($\sim\frac{1}{4}$) than that of an electron of the same kinetic energy. Fortunately, accurate measurements[10] of this ionization efficiency have been made by neutron scattering in Ge down to an equivalent electron energy of 2 keV ($E_d\approx10$ keV), and the results agree very well with the theory of Lindhard *et*

al.[11] It is the equivalent electron energy which is plotted in Fig. 1.

To compare the measured rates in Fig. 1 with the spectrum expected for a given mass of Dirac neutrino from the halo of our Galaxy, we took a halo density of 0.3 GeV/cm^3, which is 5.3×10^{-25} g/cm^3, and a Maxwellian velocity distribution with a root mean square average of 270 km/s without any truncation (which could exist if the Galaxy escape velocity were relatively low). For the relevant period (March) we assumed the velocity of the Earth with respect to the halo to be 230 km/s. With use of the known[3] interaction cross section and the parametrization of Lindhard's model[11] by Robinson,[12] the solid curves of Fig. 1 were generated for several values of m. If we assume a flat background in the region 3–8 keV, masses above 12 GeV/c^2 are excluded. The mass value is actually double valued for a fixed halo density, so that there is an upper bound of 1.4 TeV for the region of excluded masses. If we decrease the halo density by 30%, the lower mass limit increases from 12.0 to 12.5 GeV/c^2. Changing the rms velocity to 240 km/s increases the limit to 12.2 GeV/c^2.

There is an interesting variant on the standard Dirac neutrino which Raby and West have proposed[13] in order to have the same particle not only be the main component of dark matter but also explain the solar-neutrino problem. This apparent deficit of ^{8}B neutrinos coming from the sun[14] could be explained[15] by WIMP's being trapped in the Sun and transporting energy from the core to larger solar radii, thereby slightly cooling the central region from which most of the neutrinos originate. This mechanism requires cross sections with protons of about 10^{-36} cm^2, a low enough annihilation rate, and masses roughly in the range of 4 to 15 GeV/c^2. Below 4 GeV/c^2 the particles evaporate too rapidly from the surface and above about 15 GeV/c^2 they do not travel far enough from the core to cool it efficiently. Such particles could also explain[15] the discrepancy between observations of solar oscillations and the standard solar model, but the required properties are not possessed by conventional or even previously proposed particles.[16] The Raby-West[13] solution is to postulate a fourth-generation Dirac neutrino, called a magnino, since it would have a large magnetic moment arising from the existence nearby in mass of its charged partner and associated Higgs particle. The resulting electromagnetic interaction with protons has an appropriate cross section $\simeq10^{-36}$ cm^2, much higher than the weak-interaction cross section. For a heavy nucleus like Ge, however, the electromagnetic interaction becomes only comparable to the weak interaction. In addition, the interference term becomes model dependent, since it involves the charge radius r and magnetic moment μ of the magnino. The complete cross section has been calculated for us by Raby and West. With parameters ($\mu=1/8\pi^2$, $r^2=\mu^2/8m^2$, where m is the magnino mass) suggested by the authors, the resulting mass limit is 11 GeV/c^2, which is not very different from that for a particle with only weak interactions. Although the statement is often made that WIMP's with cross sections large enough to account for the solar-neutrino problem should be easy to detect in laboratory experiments, this model at least shows that that is not necessarily so. However, as the mass of either the incident particle or the target nucleus is diminished, the relative effect of the electromagnetic interaction becomes more important.

In contrast, there are models for WIMP's to solve the solar-neutrino problem in which the large cross section on protons is reflected in a large cross section on Ge as well. These include the cosmion of Gelmini, Hall, and Lin,[17] which interacts with light quarks via a heavy colored scalar, another fourth-generation neutrino of Raby and West,[18] which has a stronger interaction because of a light neutral Higgs boson, and the neutrino from broken E_6 of Ross and Segrè,[19] which gets its needed interaction from an additional Z'. A single mass limit of <9 GeV/c^2 is determined from the data for all of these models because the cross sections are large enough so that the limit is set by the kinematic threshold.

Since such models may be ephemeral, it is important to display the results of this experiment in a way which may have wider eventual application. Such information is shown in Fig. 2, which is an exclusion plot in dark-particle mass versus interaction cross section, primarily of that particle with Ge, although the upper limit is for the cross section on material above the detector. In the region of large masses, the upper limit on cross section corresponds to the case in which the energy deposition in

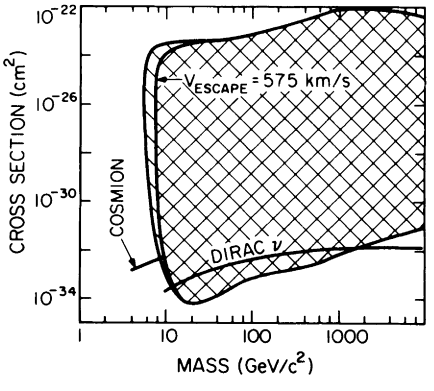


FIG. 2. Mass exclusion plot for particular cross sections for the interaction with Ge (or, for the upper boundary, with the Earth) using data from the present experiment. The larger shaded region is excluded if the escape velocity is infinite; the smaller region, if the escape velocity is 575 km/s. All limits are 2σ. The curve near the bottom of the plot shows the cross section as a function of mass for a Dirac neutrino. The short line with the label cosmion shows the expected cross section for cosmions with mass between 4 and 9 GeV/c^2.

the NaI is sufficient to veto the event. At lower masses, particles with large enough cross section with the Earth are slowed before they reach the apparatus so that they produce insufficient recoil energy to be observed. The left edge of the exclusion region is determined by the data near threshold. Note the small difference in this region which occurs if a 575-km/s escape velocity is assumed to truncate the Maxwellian velocity distribution. The rise with mass of the lower edge comes from a combination of the decreasing number density of the incident particles (since the mass density of the halo is fixed) and of an increasing contribution of the background due to the widening of the expected energy distribution. We have taken into account the loss of coherence at high masses, assuming a radius of 4×10^{-13} cm for the Ge nucleus. Above about 50-GeV mass the background at energies above the Ga x ray is setting the limit. At high masses, the lower limit is proportional to mass. All mass limits given here are at the 2-standard-deviation level. In all cases the usual assumption has been made that the considered particles constitute all of dark matter.

Figure 2 shows that existing Ge detectors set useful limits on dark-matter candidates. We have shown here that one class of proposed dark matter, millicharged shadow matter, is effectively eliminated and that the masses of particles with an initial asymmetry in the universe, such as Dirac neutrinos, magninos, and cosmions, are being constrained. While the present apparatus will give improved results after some time, more promising is the development of Si detectors which can cover the full range (down to 4 GeV) of interest for WIMP's which could also solve the solar-neutrino problem.[20]

This work was supported in part by the U.S. Department of Energy, and the results could not have been obtained without the invaluable help of D.L. Hale, D. A. Landis, N. W. Madden, and R. H. Pehl.

[1]D. Hegyi and K. A. Olive, Phys. Lett. **126B**, 28 (1983); J. Yang *et al.*, Astrophys. J. **281**, 493 (1984); G. R. Blumenthal *et al.*, Nature (London) **311**, 517 (1984).

[2]D. O. Caldwell *et al.*, Phys. Rev. Lett. **59**, 419 (1987).

[3]M. W. Goodman and E. Witten, Phys. Rev. D **31**, 3059 (1985).

[4]F. S. Goulding *et al.*, IEEE Trans. Nucl. Sci. **31**, 285 (1984), and **32**, 463 (1985); D. O. Caldwell *et al.*, Phys. Rev. Lett. **54**, 281 (1985), and Phys. Rev. D **33**, 2737 (1986).

[5]D. J. Gross *et al.*, Phys. Rev. Lett. **54**, 502 (1985).

[6]R. Holdom, Phys. Lett. **166B**, 196 (1985).

[7]H. Goldberg and L. J. Hall, Phys. Lett. B **174**, 151 (1986).

[8]S. P. Ahlen *et al.*, Phys. Lett. B **195**, 603 (1987).

[9]E. W. Kolb and K. A. Olive, Phys. Rev. D **33**, 1202 (1986); K. Greist and D. Seckel, Nucl. Phys. **B283**, 681 (1987).

[10]C. Chasman, in *Penetration of Charged Particles in Matter—A Symposium* (National Academy of Sciences, Washington DC, 1970), p. 16, and references therein.

[11]J. Lindhard *et al.*, Kgl. Dan. Vidensk. Selsk., Mat.-Fys. Medd. **33**, 10, 14 (1963).

[12]M. T. Robinson, in *Nuclear Fusion Reactors,* edited by J. L. Hall and J. H. C. Maple (British Nuclear Energy Society, London, 1970), p. 364.

[13]S. Raby and G. B. West, Nucl. Phys. **B292**, 793 (1987), and Phys. Lett. **B 194**, 557 (1987), and **200**, 547 (1988).

[14]J. K. Rowley, B. T. Cleveland, and R. Davis, Jr., in *Solar Neutrinos and Neutrino Astronomy—1984,* edited by M. L. Cherry *et al.*, AIP Conference Proceedings No. 126 (American Institute of Physics, New York, 1985), p. 1.

[15]J. Faulkner and R. L. Gilliland, Astrophys. J. **299**, 994 (1985); L. M. Krause, Harvard University Report No. HUTP-85/A008a, 1985 (unpublished); W. H. Press and D. N. Spergel, Astrophys. J. **296**, 679 (1985); R. L. Gilliland, J. Faulkner, W. H. Press, and D. N. Spergel, Astrophys. J. **306**, 703 (1986).

[16]L. M. Krause *et al.*. Astrohpys. J. **299**, 1001 (1985); J. Silk, K. Olive, and M. Srednicki, Phys. Rev. Lett. **55**, 257 (1985).

[17]G. B. Gelmini, L. J. Hall, and M. J. Lin, Nucl. Phys. **B281**, 726 (1987).

[18]S. Raby and G. B. West, Phys. Lett. B **202**, 47 (1988).

[19]G. G. Ross and G. C. Segrè, Phys. Lett. B **197**, 45 (1987).

[20]B. Sadoulet, J. Rich, M. Spiro, and D. O. Caldwell, Astrophys. Lett. **324**, 75 (1988).