PONTIFICIAE ACADEMIAE SCIENTIARVM SCRIPTA VARIA

48

ASTROPHYSICAL COSMOLOGY

PROCEEDINGS OF THE STUDY WEEK ON COSMOLOGY AND FUNDAMENTAL PHYSICS

September 28 - October 2, 1981

EDITED BY

H. A. BRÜCK, G. V. COYNE AND M. S. LONGAIR

PONTIFICIA ACADEMIA SCIENTIARVM

EX AEDIBVS ACADEMICIS IN CIVITATE VATICANA

MCMLXXXII

PONTIFICIAE ACADEMIAE SCIENTIARVM SCRIPTA VARIA
48

ASTROPHYSICAL COSMOLOGY

PROCEEDINGS OF THE STUDY WEEK ON COSMOLOGY AND FUNDAMENTAL PHYSICS

September 28 - October 2, 1981

EDITED BY
H. A. BRÜCK, G. V. COYNE AND M. S. LONGAIR

PONTIFICIA
ACADEMIA
SCIENTIARVM

EX AEDIBUS ACADEMICIS IN CIVITATE VATICANA

MCMLXXXII

Distributed by:

Specola Vaticana
I - 00120 Città del Vaticano
Europe

CONTENTS

I. INTRODUCTION

II. LARGE-SCALE STRUCTURE OF THE UNIVERSE

III. EVOLUTION OF GALAXIES

IV. EVOLUTION OF QUASARS, RADIO GALAXIES AND THE X-RAY BACKGROUND RADIATION

V. PRIMORDIAL NUCLEOSYNTHESIS AND THE ORIGIN OF GALAXIES

VI. THE VERY EARLY UNIVERSE AND PARTICLE PHYSICS

VII. COSMOLOGY AND FUNDAMENTAL PHYSICS: CONCLUDING REMARKS

Editorial note: By previous agreement the manuscripts of the papers that follow were submitted to the editors some time after the Study Week. Thus they take into account many of the comments made during the discussions, some of which are printed here following the papers. The sequence of papers in this volume is somewhat different from that in which they were actually given. The editors agreed not to require strict uniformity among the authors in such matters as style, spelling or abbreviation of references.

PREFACE

It is a great pleasure to present this volume which contains the papers and discussions of the Study Week on "Cosmology and Fundamental Physics", held at the Casina Pius IV, seat of the Pontifical Academy of Sciences.

The Academy traces its origin back to 1603, the date of the foundation hy Federico Cesi of the "Accademia dei Lincei", an institution which harbored among its members Galileo Galilei. In 1847 the Academy was reformed by Pius IX and became an official part of the Roman State. In 1936 Pope Pius XI transformed the "Accademia dei Nuovi Lincei", established by Pope Pius IX, into the Pontifical Academy of Sciences and its membership became a roster of international scientists elected without any discrimination of religion or race.

One of the activities of the Pontifical Academy of Sciences consists in the organization of Study Weeks, 18 of which have already been held. The Annals of the present Study Week will certainly go far beyond the limits of our walls, as the words pronounced by His Holiness Pope John Paul II, during the Audience he gave to the group of Academicians and scientists gathered in Castelgandolfo on the 3rd of October 1981 will resound throughout the world scientific community.

The present Study Week was organized by the Pontifical Academicians Hermann A. Brück and Father G. V. Coyne, S. J., together with Prof. Martin Rees of Cambridge and Prof. Malcolm Longair of Edinburgh. They were able to bring together some of the most important contributors to this extraordinary, interesting, exciting and active field of the astronomical sciences. I wish to express to them my gratitude and hope that

this volume will create as much interest and work as have the Annals of the two previous Study Weeks on topics of astronomy.

It is my pleasure also to acknowledge and to thank the members of the Specola Vaticana for the help given during the organization and the realization of the Study Week. I wish to extend my thanks also to Mrs. Brück of the University of Edinburgh.

I am very much indebted to the Specola Vaticana for the important collaboration and responsibility they have taken in the diffusion of the present Proceedings.

My thanks are also due to Mrs. Massa for the excellent transcription of the tape records.

It is my desire also to emphasize the work done and the assistance given to the organization of the meeting by Father Enrico di Rovasenda, Director of the Chancellery, by Mrs. Michelle Porcelli-Studer, Secretary of the Academy and Mr. Silvio Devoto, which permitted the perfect development of this arduous task.

CARLOS CHAGAS
President of the Pont. Academy of Sciences

LIST OF PARTICIPANTS

Prof. J. AUDOUZE, Institut d'Astrophysique du CNRS, 98 bis, Bd. Arago, F-75014, Paris, France.

Prof. H.A. Brück, Department of Astronomy, University of Edinburgh, Royal Observatory, Edinburgh, EH9 3HJ, U.K.

Dr. G.V. COYNE, S.J., Specola Vaticana, I-00120, Vatican City State.

Dr. M. DAVIS, Department of Astronomy, University of California, Berkeley, CA 94720, U.S.A.

Dr. S. FABER, Lick Observatory, Board of Studies in Astronomy and Astrophysics, University of California, Santa Cruz, CA 95064, U.S.A.

Prof. L.Z. FANG, Astrophysics Research Division, University of Science and Technology of China, Hefei, Anhui, Peoples Republic of China.

Prof. L. GRATTON, Laboratorio di Astrofisica, Casella Postale 67, 00044 Frascati (Roma), Italy.

Prof. J.E. GUNN, Princeton University Observatory, Princeton, NJ 08544, U.S.A.

Prof. S.W. HAWKING, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge, CB3 9EW, U.K.

Prof. L. LEPRINCE-RINGUET, Ecole Polytechnique, Laboratoire de Physique, 17 Rue Descartes, Paris V, France.

Prof. M.S. LONGAIR, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, U.K.

Prof. D. LYNDEN-BELL, Institute of Astronomy, The Observatories, Madingley Road, Cambridge CB3 0HA, U.K.

Prof. J.H. OORT, Sterrewacht Leiden, Wassenaarseweg 78, Leiden, Holland.

Prof. J.P. OSTRIKER, Princeton University Observatory, Princeton, NJ 08544, U.S.A.

Prof. P.J.E. PEEBLES, Joseph Henry Laboratories, Physics Department, Princeton University, Princeton, NJ, 08544, U.S.A.

Prof. M.J. REES, Institute of Astronomy, The Observatories, Madingley Road, Cambridge, CB3 0HA, U.K.

Prof. M. SCHMIDT, Palomar Observatory, California Institute of Technology, Pasadena, CA 91125, U.S.A.

Prof. D.W. SCIAMA, Department of Astrophysics, University of Oxford, South Parks Road, Oxford, OX1 3RQ, U.K.

Prof. G. SETTI, Istituto di Radioastronomia, Università di Bologna, 40126, Bologna, Italy.

Dr. J. SILK, Department of Astronomy, University of California, Berkeley, CA 94720, U.S.A.

Prof. G. SWARUP, Radio Astronomy Centre, Tata Institute of Fundamental Research, Ootacamund, 643001, India.

Prof. G.A. TAMMANN, Astronomisches Institut der Universität Basel, CH-4102, Binningen, Switzerland.

Prof. H. VAN DER LAAN, Sterrewacht Leiden, Wassenaarseweg 78, Leiden, Holland.

Prof. S. WEINBERG, Department of Theoretical Physics, Harvard University, Cambridge, MA, 02138, U.S.A.

Prof. V. WEISSKOPF, Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, 02139, U.S.A.

Prof. L. WOLTJER, European Southern Observatory, 8046 Garching bei München, Federal Republic of Germany.

The meetings were attended also by Dr. M.T. Brück of the University of Edinburgh and Drs. R.P. Boyle, J. Casanovas and M.F. McCarthy of the Vatican Observatory who took notes of the various discussions and edited the transcripts of tape records.

PAPAL AUDIENCE

On October 3, 1981 His Holiness John Paul II graciously received the participants in the Study Week on Cosmology and Fundamental Physics during an audience granted to the Pontifical Academy of Sciences in the Hall of the Swiss in the Apostolic Palace at Castel Gandolfo. On that occasion the following addresses were delivered.

ADDRESS OF PROFESSOR CARLOS CHAGAS
PRESIDENT OF THE ACADEMY

Sainteté,

L'audience que vous accordez aujourd'hui à l'Académie Pontificale des Sciences se réalise dans un climat de joie profonde, qui s'unit aux sentiments d'admiration et de respect filial qui marquent toujours les occasions solennelles au cours desquelles vous recevez les Académiciens Pontificaux et les scientifiques qui viennent participer à nos travaux. Nous tous nous avons suivi avec anxiété et espérance Votre longue maladie, et maintenant nous nous sentons heureux de pouvoir jouir de Votre présence et Vous exprimer notre pleine réjouissance.

Ces beaux moments, sans ombrage, que nous passons dans le calme de Castelgandolfo, nous font oublier ceux de profonde angoisse éprouvés après l'acte de barbarie dont Vous avez été victime. Nous sommes heureux de pouvoir — dans un monde perturbé par la violence, l'incompréhension, les égoïsmes et les inégalités — venir écouter à nouveau auprès de Vous la bonne parole qui nous donnera l'encouragement nécessaire pour renforcer et guider l'action qui anime nos vies.

Nous croyons tous que la science et la technique ont un rôle significatif à jouer pour le progrès de la société. Science et technique mises au service de l'humanité peuvent alléger la condition humaine de ses pesanteurs. Elles le feront en abordant les problèmes de l'Univers inorganique ou vivant — du microcosmos et du macrocosmos — lesquels suscitent la curiosité de chaque être vivant depuis l'aube de notre civilisation et dont la connaissance pourra permettre une amélioration de son sort. Elles créeront aussi des procédés qui donneront aux travailleurs des outils plus propres à leur dignité personnelle et à leur position au sein de la collectivité. Vous nous avez enseigné, Sainteté, dans votre très récente Encyclique « Laborem exercens » que « les actions accomplies par chaque travailleur doivent toutes servir à la réalisation de sa personne, à l'achèvement de la vocation comme être humain qui lui est

propre en raison de sa personnalité ». C'est aussi le devoir de la science et de la technique d'assurer en particulier aux plus dépourvus les conditions de nutrition, d'habitation, de communication et de transport que le simple fait d'exister exige. Science et technique pourvoient les moyens pour prévenir et guérir les maladies et assurer la multiplication indispensable de la production agricole. Elles créeront en outre un espace plus grand pour le loisir, espace dans lequel trouveront plus de place la prière, la méditation spirituelle ou philosophique, les arts, la poésie, l'invocation créatrice, la fantaisie et le sport.

Science et technique, éprises du progrès humain, peuvent contribuer au rêve d'établir un réseau de vraie solidarité entre les peuples et un langage commun de paix et de compréhension.

C'est à cette fin que Votre prédécesseur Benoît XV a conçu l'idée de l'internationalisation de l'ancienne « Accademia dei Nuovi Lincei », que Pie XI a bien voulu rendre effective. Votre Académie l'a poursuivie, et depuis la très généreuse décision de Paul VI de me nommer Président, j'essaie de la réaliser, en suivant les pas de ceux qui m'ont précédé grâce à l'appui de Paul VI et de Vous-même, Sainteté, et à l'aide de mes confrères et de mes collaborateurs.

Tâche en même temps hardie et pleine de récompenses, parce qu'elle nous met au service de l'homme, donc une tâche culturelle et chrétienne.

En réunissant autour de nos tables de travail des scientifiques provenant de tous les horizons disciplinaires et géographiques, nous pouvons aborder les problèmes les plus variés dans le but d'accroître les connaissances et perfectionner la condition existentielle de l'humanité.

C'est en reconnaissant l'unité du cosmos et l'unité de l'humanité que nos travaux se déroulent.

La Semaine d'Etude qui vient de se terminer a été vouée à une réflexion sur « La cosmologie et la physique fondamentale ». Permettez-moi, Saint-Père, de rendre hommage à ce propos à la mémoire de l'Abbé Lemaître, Président de l'Académie Pontificale des Sciences de 1960 à 1966, un des responsables de la création du concept de l'explosion de l'atome primaire, dont les travaux marquent l'histoire de la science moderne.

Simultanément, un Groupe de travail s'est occupé des « Perspectives d'immunisation des maladies parasitaires » Sujet plein d'attirance, non seulement parce qu'il intéresse fortement les pays du tiers monde, mais aussi parce qu'il démontre l'importance de la science fondamentale pour la solution des problèmes pratiques et nous fait reconnaître la vérité des

propositions de Pasteur, lorsqu'il affirmait que « la science et ses applications ne font qu'un tout ».

Aujourd'hui même, nous commençons à traiter en Séance plénière le thème « L'impact de la biologie moléculaire sur la société ». Thème d'actualité, plein de promesses mais aussi de difficultés. Dans quelques jours, un autre Groupe de travail traitera les « Effets conséquents à un bombardement atomique ». Les paroles que Vous avez prononcées à diverses occasions, Très Saint-Père, concernant les dangers d'une guerre atomique, suffisent à expliquer l'intérêt que nous tous portons au sujet qui sera traité.

L'ampleur de nos desseins se justifie par les arguments les plus divers. Je citerai seulement l'intérêt que Vous-même, Sainteté, portez à l'intégration de la science à la culture qui reflète l'image de chaque peuple, ainsi que Vous l'avez dit à l'UNESCO il y a un peu plus d'une année.

Je vous annonce avec joie, Très Saint-Père, que les nouveaux membres de l'Académie que Vous avez nommés sont présents à cette Audience. Je me permets de citer leurs noms: Anatole Abragam de la France, Christian Anfinsen des Etats-Unis, Werner Arber de la Suisse, Ennio De Giorgi de l'Italie, Manfred Eigen de l'Allemagne, André Lichnerowicz de la France, Mambillikalathil Menon des Indes, Thomas Odhiambo du Kenya, Max Perutz de l'Angleterre, Bernard Pullman de la France, Stanley Runcorn de l'Angleterre, Abdus Salam du Pakistan, Janos Szentágothai de l'Hongrie et l'Académicien Honoraire Silvio Ranzi de l'Italie. En se joignant à notre Académie, je suis sûr qu'ils rendront le plus grand service à l'oeuvre que nous devons accomplir.

Je désire encore Vous demander Sainteté, avec Votre bienveillance, de remettre au Prof. Jean-Marie Lehn de Strasbourg la médaille d'or Pie XI. En le choisissant parmi d'autres candidats, jeunes comme lui et comme lui très valables, le Conseil de l'Académie a voulu couronner l'oeuvre d'un scientifique qui, après des travaux remarquables dans le domaine de la Chimie organique, s'est dédié à l'étude de la photochimie qui se présente comme une passerelle par laquelle la science pourra passer pour utiliser l'énergie solaire.

Sainteté, en Vous remerciant de nous avoir reçus et en Vous demandant la Bénédiction Apostolique, je ne peux m'empêcher de Vous exprimer à nouveau notre joie et de Vous dire que Votre sérénité au cours de ces longues semaines qui se sont écoulées, nous a enseigné à vivre plus près des êtres humains et d'une façon plus chrétienne.

ALLOCUTION OF HIS HOLINESS JOHN PAUL II

Monsieur le Président,
Messieurs les Académiciens,
Mesdames, Messieurs,

1. *Le programme des travaux que votre Président a présenté, et dont j'avais déjà connaissance avant cette rencontre, montre la grande vitalité de votre Académie, son intérêt pour les problèmes les plus aigus de la science contemporaine et pour le service de l'humanité. J'ai déjà eu l'occasion de vous dire, lors d'une autre séance solennelle, combien l'Eglise estime la science pure: elle est — disais-je — « un bien, digne d'être très aimé, car elle est connaissance et donc perfection de l'homme dans son intelligence... Elle doit être honorée pour elle-même, comme une partie intégrante de la culture » (Discours à l'Académie Pontificale des Sciences, 10 novembre 1979).*

Avant d'aborder les problèmes dont vous avez déjà discuté ces jours-ci et ceux que vous vous proposez maintenant d'étudier, permettez-moi de remercier chaleureusement votre illustre Président, le Professeur Carlos Chagas, des félicitations qu'il a bien voulu m'exprimer au nom de toute votre Assemblée, pour avoir retrouvé mes forces physiques, grâce à la miséricordieuse Providence de Dieu et à la compétence des médecins qui m'ont soigné. Et je suis heureux de profiter de cette occasion pour dire ma particulière gratitude à Messieurs les Académiciens qui, de toutes les parties du monde, m'ont adressé leurs voeux et m'ont assuré de leurs prières.

2. Pendant cette Semaine d'études, vous vous penchez sur le problème de la « Cosmologie et physique fondamentale », avec la participation de savants du monde entier, depuis les deux Amériques jusqu'à l'Europe et à la Chine. Ce sujet se rattache à des thèmes déjà traités par l'Académie Pontificale des Sciences au cours de son histoire prestigieuse. Je veux parler ici des sessions sur les micro-séismes, sur les populations stellaires, sur les radiations cosmiques, sur les noyaux des galaxies, sessions qui se sont déroulées sous la présidence du Père Gemelli, de Monseigneur Lemaître, et aussi du Père O'Connell auquel j'adresse mes voeux les plus fervents en demandant au Seigneur de l'assister dans son épreuve de santé.

La cosmogonie et la cosmologie ont toujours suscité un grand intérêt chez les peuples et dans les religions. La Bible elle-même nous parle de l'origine de l'univers et de sa constitution, non pas pour nous fournir un traité scientifique mais pour préciser les justes rapports de l'homme avec Dieu et avec l'univers. L'Ecriture Sainte veut simplement déclarer que le monde a été créé par Dieu, et pour enseigner cette vérité elle s'exprime avec les termes de la cosmologie en usage au temps de celui qui écrit. Le livre sacré veut en outre faire savoir aux hommes que le monde n'a pas été crée comme siège des dieux, comme l'enseignaient d'autres cosmogonies et cosmologies, mais qu'il a été créé au service de l'homme et à la gloire de Dieu. Tout autre enseignement sur l'origine et la constitution de l'univers est étranger aux intentions de la Bible: celle-ci ne veut pas enseigner comment a été fait le ciel, mais comment on va au ciel.

Toute hypothèse scientifique sur l'origine du monde, comme celle d'un atome primitif d'où dériverait l'ensemble de l'univers physique, laisse ouvert le problème concernant le commencement de l'univers. La science ne peut par elle-même résoudre une telle question: il y faut ce savoir de

l'homme qui s'élève au-dessus de la physique et de l'astrophysique et que l'on appelle la métaphysique; il y faut surtout le savoir qui vient de la révélation de Dieu. Il y a trente ans, le 22 novembre 1951, mon prédécesseur le Pape Pie XII, parlant du problème de l'origine de l'univers lors de la Semaine d'études sur le problème des micro-séismes organisée par l'Académie Pontificale des Sciences s'exprimait ainsi: « En vain attendrait-on une réponse des sciences de la nature, qui déclarent au contraire loyalement se trouver devant une énigme insoluble. Il est également certain que l'esprit humain versé dans la méditation philosophique pénètre plus profondément dans le problème. On ne peut nier qu'un esprit éclairé et enrichi par les connaissances scientifiques modernes, et qui envisage avec sérénité ce problème, est conduit à briser le cercle d'une matière totalement indépendante et autonome — parce que ou incréée ou s'étant créée elle-même — et à remonter jusqu'à un Esprit créateur. Avec le même regard limpide et critique dont il examine et juge les faits, il y entrevoit et reconnaît l'oeuvre de la Toute-Puissance créatrice, dont la vertu, suscitée par le puissant "fiat" prononcé il y a des milliards d'années par l'Esprit créateur, s'est déployée dans l'univers, appelant à l'existence, dans un geste de généreux amour, la matière débordante d'énergie ».

3. *Je me réjouis vivement, Messieurs les Académiciens, du thème que vous avez choisi pour votre Session plénière qui commence aujourd'hui même: « L'impact de la biologie moléculaire sur la société ». J'apprécie les avantages qui résultent — et qui peuvent résulter encore — de l'étude et des applications de la biologie moléculaire, complétée par d'autres disciplines comme la génétique et son application technologique dans l'agriculture et dans l'industrie, et aussi, comme on l'envisage, pour le traitement de diverses maladies, dont certaines de caractère héréditaire.*

J'ai une ferme confiance dans la communauté scientifique mondiale, et d'une manière toute particulière dans l'Académie Pontificale des Sciences, certain que grâce à elles les progrès et les recherches biologiques, comme du reste toute autre recherche scientifique et son application technologique, s'accompliront dans le plein respect des normes morales, en sauvegardant la dignité des hommes, leur liberté et leur égalité. Il est nécessaire que la science soit accompagnée et contrôlée par la sagesse qui appartient au patrimoine spirituel permanent de l'humanité et qui s'inspire du dessein de Dieu inscrit dans la création avant d'être ensuite annoncée par sa Parole.

Une réflexion qui s'inspire de la science et de la sagesse de la communauté scientifique mondiale doit éclairer l'humanité sur les conséquences — bonnes et mauvaises — de la recherche scientifique, et spécialement de celle qui concerne l'homme, afin que, d'une part, on ne se fixe pas sur des positions anticulturelles qui retardent le progrès de l'humanité, et que d'autre part on n'offense pas ce que l'homme a de plus précieux: la dignité de sa personne, destinée à un vrai progrès dans l'unité de son être physique, intellectuel et spirituel.

4. *Un autre sujet a retenu ces jours-ci l'attention de certains d'entre vous, savants éminents de diverses parties de la terre convoqués par l'Académie Pontificale des Sciences: c'est celui des maladies parasitaires qui frappent les pays les plus pauvres du monde et sont un grave obstacle à la promotion de l'homme dans le cadre harmonieux de son bien-être physique, économique et spirituel. Les efforts en vue d'éliminer le plus possible les fléaux provoqués par les maladies parasitaires dans une bonne partie de l'humanité sont inséparables de ceux qu'il faut faire en faveur du développement socio-économique des mêmes populations. Les*

hommes ont normalement besoin d'une santé suffisante et d'un minimum de biens matériels pour pouvoir vivre dignement selon leur vocation humaine et divine. C'est pour cela que le Christ Jésus s'est tourné avec un amour infini vers les malades et les infirmes, et qu'il a guéri miraculeusement quelques-unes des maladies dont vous vous êtes occupés ces jours derniers. Que le Seigneur inspire et assiste l'activité des savants et des médecins qui consacrent leur recherche et leur profession à l'étude et au besoin des infirmités humaines, spécialement des plus graves et des plus humiliantes!

5. *A côté du thème des maladies parasitaires, l'Académie a abordé le problème d'un fléau d'une ampleur et d'une gravité catastrophiques qui pourrait atteindre la santé de l'humanité si un conflit nucléaire venait à éclater. Outre la mort d'une bonne partie de la population mondiale, un conflit nucléaire pourrait provoquer des effets incalculables sur la santé des générations présentes et futures.*

L'étude pluri-disciplinaire que vous vous apprêtez à accomplir ne pourra pas ne pas constituer pour les Chefs d'Etat un rappel de leurs immenses responsabilités et susciter dans l'humanité entière une soif toujours plus ardente de concorde et de paix: cette aspiration vient du plus profond du coeur humain, et aussi du message du Christ qui est venu apporter la paix aux hommes de bonne volonté.

En vertu de ma mission universelle, je veux me faire encore une fois l'interprète du droit de l'homme à la justice et à la paix, et de la volonté de Dieu qui désire voir tous les hommes sauvés. Et je renouvelle l'appel que je lançais à Hiroshima le 25 février dernier: « Engageons-nous solennellement, ici et maintenant, à ne plus jamais permettre (et encore moins rechercher) que la guerre soit un moyen de résoudre les conflits. Promettons à nos frères en humanité de travailler sans nous lasser au désarmement et a la condamna-

tion de toutes les armes atomiques. Remplaçons la domination et la haine par la confiance mutuelle et la solidarité ».

6. *Parmi les efforts à accomplir pour la paix de l'humanité, il y a celui qui vise à garantir à tous les peuples l'énergie nécessaire à leur développement pacifique. L'Académie s'est occupée de ce problème durant la Semaine d'études de l'année dernière. Je suis heureux de pouvoir remettre aujourd'hui la Médaille d'or de Pie XI à un savant qui a contribué d'une manière notable, par sa recherche dans le domaine de la photochimie, à l'utilisation de l'énergie solaire. Il s'agit du Professeur Jean-Marie Lehn, du Collège de France et de l'Université de Strasbourg, auquel j'exprime mes vives félicitations.*

A vous tous, Messieurs, j'adresse mes sincères compliments pour le travail que vous accomplissez dans la recherche scientifique. Je prie le Dieu Tout-Puissant de vous bénir, vous, vos familles, ceux qui vous sont chers, vos collaborateurs, et toute l'humanité pour laquelle, par des routes diverses mais convergentes, vous et moi accomplissons la mission qui nous a été confiée par Dieu.

INTRODUCTORY REMARKS

H. A. BRÜCK

This conference has been opened by Professor Chagas, the President of the Pontifical Academy, who has spoken to you about the Academy's work and in particular about the institution of its Study Weeks.

I should like to join him first of all in welcoming this distinguished body of scientists. We are very happy that so many of you have been able to come. We are extremely sorry that none of our Soviet colleagues could accept the Academy's invitations, but we are pleased to have received a paper from Professor Zeldovich which will be read to us by Professor Martin Rees. We much regret that Professor Freeman Dyson who intended to take part was taken ill on his journey here and cannot be with us.

Perhaps I may be allowed to say a few words about the way in which this particular Study Week has come about. You have heard from Professor Chagas that Study Weeks devoted to a variety of topical scientific subjects were started by the Pontifical Academy in 1948 as a major part of its activities. Two of them were directly concerned with astronomy. In 1957 we had a Study Week on the problem of Stellar Populations, and in 1970 we had a second one in which Nuclei of Galaxies were discussed. Both of these arose from the initiative of Father O'Connell, the former Director of the Vatican Observatory, who, you will be sad to hear, is unfortunately too ill to be here with us this morning. Several of you have attended one or both of those earlier Study Weeks which, I believe it would be right to say, have been eminently successful.

It seemed to me that the time had come for a third astronomical Study Week, and the field of Cosmology appeared to present a particularly appropriate subject for this Academy. Apart from its obvious topical interest, a Study Week on present-day cosmology would be a tribute to one of the earliest members of this Academy and its President from 1960 until his death six years later. I am speaking, of course, of George Lemaître who has been called by one of our participants in this Study Week "the Father of Big-Bang Cosmology".

When the Council of the Academy had formally agreed to hold this particular Study Week, I thought it best, not being an expert in the field, to consult with Sir Martin Ryle, an old friend and fellow member of the Academy. Martin Ryle suggested that I should discuss the project with Malcolm Longair, his close colleague who at the time was still in Cambridge. I also approacbed Martin Rees at the Institute of Astronomy in Cambridge, and it was he who made the important suggestion that it would be interesting and useful if the proposed subject of cosmology would be linked with that of fundamental physics. Our small group was joined by Father George Coyne, the present Director of the Vatican Observatory, who undertook to look after many of the practical problems including the recording of the scientific discussions in the course of this week. Father Coyne has also agreed to be responsible for the important task of the eventual printing of the proceedings.

Since February of this year when Coyne, Longair, Rees and myself met in Edinburgh, the major part of the preparation and planning of the actual scientific programme has heen in the hands of Malcolm Longair, and a look at the programme in front of you shows you how splendidly he has performed his task.

Here in Rome the innumerahle detailed preparations necessary for the smooth running of our Study Week have been most ably carried out by Father Enrico di Rovasenda, O.P., the Director of the Academy's Chancellery and Mme. Michelle Porcelli-Studer and their staff. I am certain that we shall all be very well looked after during the week.

You see from the scientific programme that you will be fairly hard worked in nine sessions during four and a half days. There will be a break

on Wednesday afternoon when you will have a chance to visit the Vatican Museums. On Saturday morning there will be an audience of Pope John Paul II for both tbe participants in the Study Week and the members of the Pontifical Academy who will later start their biannual Plenary Session. The audience will be at the Pope's summer residence at Castel Gandolfo in the Alban Hills where you will also have a chance to visit the Vatican Observatory.

We are now probably ready to start the scientific programme and I ask Malcolm Longair to take the Chair for the first Session.

SCIENTIFIC PAPERS

I.

INTRODUCTION

INTRODUCTORY SURVEY

MARTIN J. REES
Institute of Astronomy
Madingley Road, Cambridge

The theme of this Study Week, Cosmology and Fundamental Physics, is a timely one. Hopes are high that we are starting to understand how the universe evolved; and the key physical processes, particularly during the earliest and densest stages, involve high energy particle physics (as well as, of course, gravitational theory). But the title of this conference can be interpreted also in a second sense: the initial conditions of the universe, the boundary conditions of our physical world, are *in themselves* a crucial part of fundamental physics.

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

Just how good is the evidence for large-scale homogeneity and isotropy? How did galaxies form, and what determines their present morphology? Quantitative attempts to answer these questions, by studying galaxies, the distribution of galaxies and the background radiation, will be the main aim of our discussions. But let me offer now a sketchy overview, starting with recent epochs and working back as far as the threshold of quantum gravity at 10^{-43} secs (or our threshold of credulity, if that comes sooner).

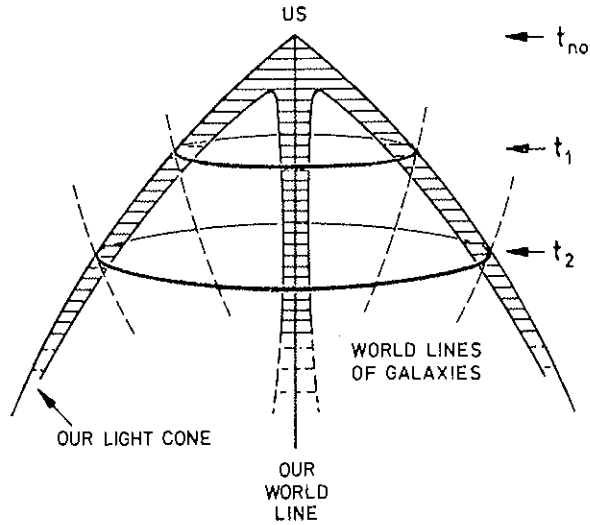


FIG. 1. Schematic space-time diagram showing world line of our Galaxy and our past light cone. The only regions of space-time concerning which we have direct evidence are those shaded in the diagram, which lie either close to our own world line (inferences on the chemical and dynamical history of our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

1 - GALAXIES

To the cosmologist, or at least to the cosmographer, entire galaxies are just "markers" or test particles scattered through space which indicate how the material content of the universe is distributed and how it is moving. But it is these "small-scale" deviations from homogeneity, groups of galaxies and their constituent members, which constitute the main subject matter of astronomy. And it is a prime goal of our subject to understand how the universe has evolved from an initial dense fireball 10^{10} yrs ago to its present state, where galaxies dominate the large-scale cosmic scene.

At the IAU Symposium in Tallin in 1977 Zeldovich said "It will only be a few years before the origin and evolution of galaxies is understood." What has happened since 1977 may have tempered his optimism. But

there have certainly been tremendous observational improvements in our data on galaxies, along with some progress in understanding their present state. A few years ago we were worse than ignorant about some crucial points. We "knew" some things that were not so. Let me mention two such things in particular. It used to be tacitly accepted that elliptical galaxies owed their shape to rotational flattening. But the observations now show that the rotation rates are often too slow. The internal dynamics are more complex. Some elliptical galaxies may be triaxial — bars spinning end over end — and the random component of the stellar velocities may generally be anisotropic. A second tacit assumption was that the luminous content of galaxies dominated their dynamics, i.e. that the total gravitating mass is distributed in the same way as the stars we see. But we are now told that the luminous content of galaxies may be swamped by ten times as much dark matter. This material forms a halo around massive isolated galaxies and provides the virial mass in rich clusters.

Progress in understanding tbe structure, dynamics, and morphology of individual galaxies depends on quantifying the processes summarized in the simple flow chart in Figure 2. There is obviously a gradual conversion of gas into stars over the lifetime of a galaxy. Our poor under-

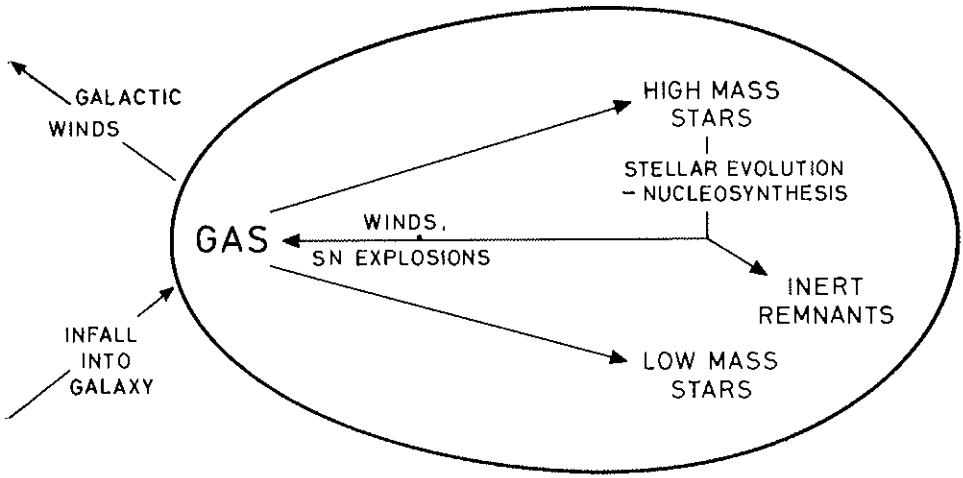


FIG. 2. The key processes involved in galactic evolution (illustrated by this flow diagram) are the "efficiency" of gas → star conversion and the initial mass function. These are both uncertain, but are crucial determinants of morphological type and of evolution in the colours and luminosity of galaxies.

standing of star formation is the main stumbling-block preventing us from quantifying this. We do not know what determines the initial mass function (IMF); still less do we know how the IMF depends on physical conditions. Nor do we know much about the efficiency of star formation: this determines how much gas can be turned into stars on the free fall timescale, and bow efficiently enriched gas can be recycled into new stars.

There has been less progress in understanding how galaxies first formed. Galaxy formation straddles the interface between cosmology and astrophysics. It occurred at a cosmic epoch very different from the present, and thus falls in the cosmologist's province. On the other hand, once galaxies have formed, the phenomena within them that interest astrophysicists proceed more or less regardless of the broader cosmic context. (At least this is true unless the universe eventually collapses on top of them). We do not know why the most conspicuous luminous entities in the universe are aggregates of 10^{11} stars, with dimensions $\sim 10^4$ parsecs. Even worse, we do not know whether the explanation we are seeking lies within the province of the astrophysicist or the cosmologist. Are these characteristic dimensions a direct consequence of initial conditions? Or is there a physical reason for this preferred scale, just as we now know there is a physical reason why all stars have a mass within an order of magnitude of the Chandrasekar mass? There are some straightforward physical ideas, based on the cooling, collapse and fragmentation of massive gas clouds, that predict a characteristic mass and radius which seem relevant to galaxy formation. If these ideas indeed have something in them, there is no more need to relate the masses of galaxies to a preferred scale of initial cosmic irregularities than to invoke a preferred fluctuation scale in the interstellar medium to explain the masses of stars. (But the characteristic scales of the *biggest* galaxies must to some extent be a function of cosmic epoch. If we came back in 10^{11} years, we should find that our Galaxy and Andromeda would have merged into a single elliptical system, and that the entire content of a cluster like Coma would be an amorphous cD galaxy).

It is thus not clear whether galaxies are permanent structures manifesting some "magic mass" for which we should seek a physical explanation (maybe this is so for spirals and disc systems, even if not for ellipticals). Even if so, it is unclear whether this mass scale, and the corresponding length scale, stem from local physics, or are consequences of selective growth or damping mechanisms in the early universe.

2 - Initial spectrum of fluctuations

Scenarios for galaxy formation all postulate that the early universe was not completely smooth. Some initial irregularities must have given rise to bound systems, which then either themselves turned into galaxies, or triggered the formation of galaxies by an indirect route.

Two classes of perturbation can be envisaged: isothermal (or entropy) perturbations, in which tbe radiation pressure is unaltered; or adiabatic (isentropic) perturbations in which the photon/baryon ratio is unperturbed. Isothermal perturbations are essentially "frozen in" before recombination; adiabatic perturbations oscillate before recombination, and all scales below $\sim 10^{15}\,M_\odot$ are attenuated. A general perturbation whose oscillatory component is damped by viscosity can leave behind an isothermal component. The oscillatory behaviour and damping of these various modes prior to recombination has been extensively discussed in the literature. After recombination, when the gravitational instability of the matter is opposed only by gradients in the gas pressure (less than radiation pressure by a factor $\sim 10^8\,\Omega^{-1}_{\rm baryon}$), all scales of $\gtrsim 10^6\,M_\odot$ grow, at least until bound systems form and generate enough energy to heat and reionise the gas. To obtain a bound system by the present time, the necessary amplitude at recombination must still be at least 10^{-3}. The growth of perturbations on the mass scales of galaxies is inhibited before recombination by the effects of radiation pressure and viscosity; this means that amplitudes of $\gtrsim 10^{-3}$ may be necessary even at the (earlier) epoch when such scales are first encompassed within the particle horizon. (In models involving massive neutrinos, when growth can occur before recombination, one can get by with perturbations whose amplitude on entering the horizon is somewhat smaller).

The two basic "scenarios" for galaxy (and cluster) formation, which are certain to come up repeatedly during the Study Week, are summarized in Figure 3. These pictures are very different: they contrast most starkly if we compare what the universe would have been like at, say, $z = 20$. In the hierarchical picture (requiring entropy perturbations) bound systems would already exist. Indeed, nearly all the initial mass could have been incorporated in pregalactic stars and Population III objects, which could have already augmented or distorted the microwave background spectrum. These pregalactic objects may amplify perturbations on large scales, giving rise to galaxies by secondary effects. On the other hand, in the adiabatic picture essentially nothing of interest has happened by

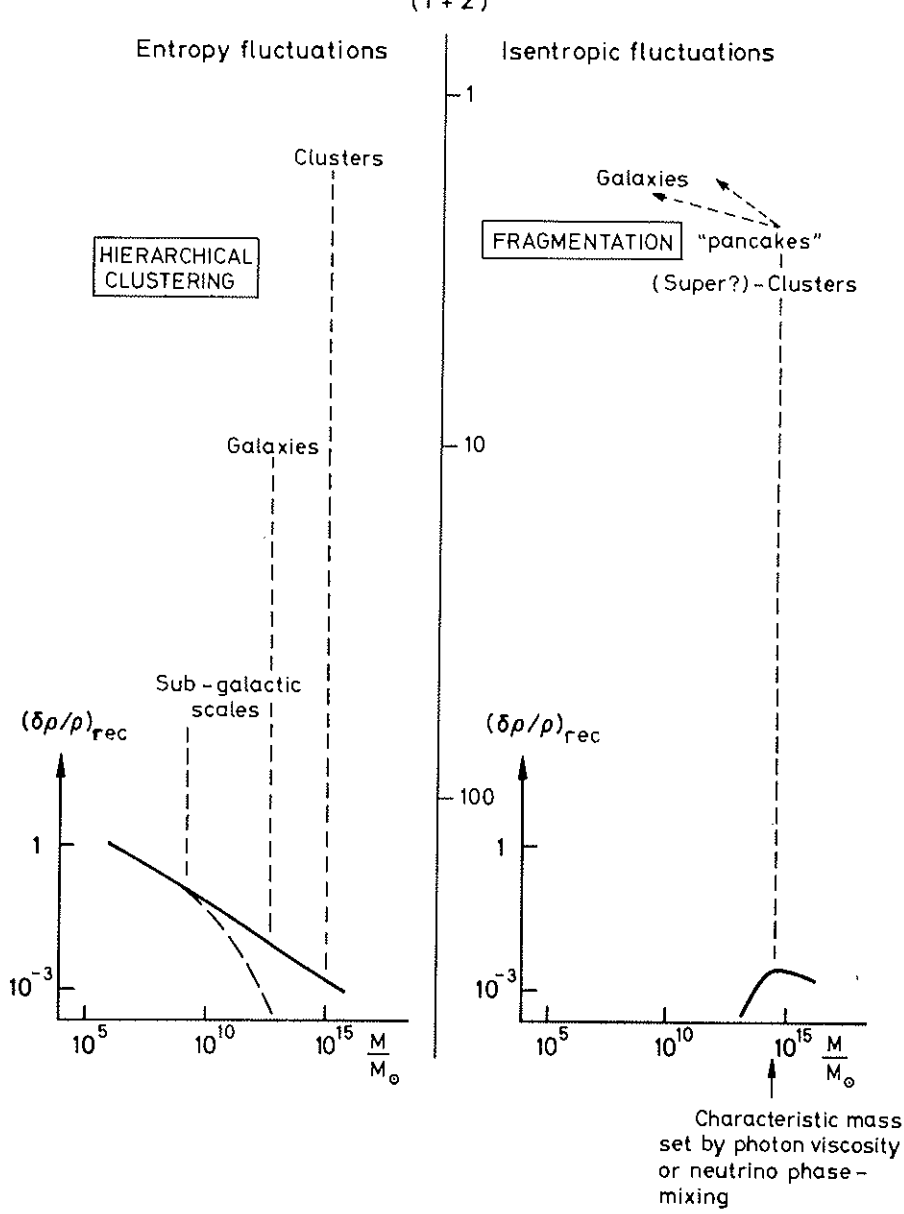


FIG. 3. At the bottom of the diagram are shown two possible forms of the density perturbation spectrum at t_{rec} ($z \simeq 1000$). Entropy fluctuations at the left may have a spectrum extending down to $\sim 10^6\,M_\odot$, objects in the mass range 10^6-$10^8\,M_\odot$ may condense out soon after t_{rec}, and galaxies and clusters then form later via hierarchical build-up of subunits. If the subunits generate energy, and give rise to "secondary" perturbations, then galaxies and clusters could form even if there were a negligibly small ($\ll 10^{-3}$) amplitude on the relevant scales at t_{rec} (as in the dashed spectrum). At the right is the contrasting case of adiabatic perturbations. The first scales to condense out would be on a scale of galaxy clusters or larger, because $(\delta\rho/\rho)_{rec}$ would have been attenuated on smaller scales by pre-recombination damping processes. If massive neutrinos play a key role in galaxy formation, the scenario resembles that shown on the right, though the characteristic mass below which attenuation occurs is then set by phase mixing of collisionless neutrinos rather than by photon viscosity.

$z = 20$. Density contrasts are still less than order unity, and the universe would still be in the form of expanding neutral gas.

The details of the large-scale distribution of galaxies, and the issue of whether there is a preferred scale, may offer some clues to which picture is the right one. The universe undoubtedly looks smoother on larger scales: no-one is gainsaying this general trend. Recent redshift surveys suggest, however, that the fluctuations on scales of 10 to 100 Mpc are larger than previously thought. The key question is whether the data are consistent with gravitational clustering of individual (galactic) masses or whether there is evidence for gas-dynamical dissipative processes, leading to giant sheetlike or filamentary structures, and to a preferred mass on supercluster scales. Everyone would agree that on scales of galaxies and below dissipation has been crucial, but on sufficiently large scales any observed inhomogeneities must have been induced by gravitational instability alone. But what is the transition scale between these regimes? Is it individual galaxies, clusters, or still larger? The papers by Davis, Oort and Peebles will help to clarify this.

3 - Hidden mass

The evidence is now strong that 90% of the material in the universe is in some form, not stars and not gas, which is less dissipative and less centrally condensed than the stellar distribution. One would like to know how universal is this ratio of 10:1 between dark and luminous matter. Does this ratio prevail over all scales exceeding, say, 1 megaparsec? Are the voids empty of all matter or just of luminous galaxies?

One must of course be cautious about inferring the distribution of *all* mass from observations of luminous objects which may comprise only 10% of the total amount of gravitating material — the more so because within individual galaxies the dark material is less centrally concentrated than the visible stars and gas. However, while it is easy to envisage how such segregation could have occurred in the course of an individual galaxy's formation and evolution, it is less easy to envisage how the two kinds of matter could be segregated on scales exceeding a few megaparsecs. Thus on large scales it is probably justifiable to regard galaxies as valid tracers for the mass distribution. If large "voids" really exist, which cannot be explained by gravitational clustering, one's mind should be open to more radical possibilities: Could galaxy formation from primordial gas have been inhibited in some large regions? Could the early universe have split into "domains" with different baryon/photon ratios?

It needs no great ingenuity to invent possible forms of hidden mass-"massonium". Among them are these:

a) *Neutrinos of rest mass ~ 10 ev.* If neutrino rest masses are non-zero, then it is a straightforward consequence of the hot Big Bang model that they contribute a density parameter $\Omega_\nu \simeq 0.04\ (H_0/50\ \mathrm{km/sec/Mpc})^{-2}\ (m_\nu)_{\mathrm{ev}}$. This is because their number density is comparable to that of the photons in the microwave background; i.e. they outnumber the baryons by a factor $\sim 10^8\ \Omega^{-1}_{\mathrm{baryon}}$. These primordial neutrinos would have cooled adiabatically with the expanding universe, and would become sufficiently slow-moving that they could participate in gravitational clustering. We shall hear more about this from Sciama later in the week.

b) *Low-mass stars ("Jupiters") of mass $\lesssim 0.1\ M_\odot$.* Such objects in our own halo might be discernible as faint red or infrared stars with high proper motion.

c) *Remnants of massive stars.* Black holes could in principle have formed as the endpoint of the evolution of now-defunct massive stars which formed either in the early history of galaxies or at a possible pregalactic era of activity ($10 \lesssim z \lesssim 10^3$). There are then some other constraints. Limits on the extragalactic night sky brightness imply that any stars with masses (say) 5 to 100 $M_\odot$ must have formed and completed their bright evolutionary phases at $z \gtrsim 10$ (unless their optical and ultraviolet emission is absorbed by dust and re-emitted in the infra-red). Such considerations do not, however, constrain a "Population III" of massive stars that formed at $z \simeq 100$.

d) *Remnants of supermassive stars.* Here the background light constraint is more ambiguous, because supermassive stars may collapse before having gone through a prolonged phase of hydrogen burning. However there are other ways of detecting them. For instance, there would be conspicuous accretion effects in our galaxy if individual masses exceeded $\sim 10^6\ M_\odot$. Also, dynamical friction would be important for $\gtrsim 10^7\ M_\odot$. (This would cause the dark objects to concentrate towards the centres of galaxies, contrary to the aim of the game which is that they should be predominant outside the luminous material). A new possibility is that gravitational lensing may have detectable effects if the hidden mass is in massive black holes — the characteristic angular scale of the images, for a path length of the order of the Hubble radius, is $\sim 10^{-3}\ (M/10^6\ M_\odot)^{1/2}$ arc sec.

Among these options, which obviously are not exhaustive, my favourites are numbers a) or d). We shall hear a good deal from other

speakers about the remarkable possibility that the universe is neutrino-dominated, and that all the baryons studied by astronomers, in stars, galaxies and gas, amount to no more than a trace of "sediment" in a neutrino-dominated cosmos. I hope to say a bit more later in the week about supermassive objects and pregalactic stars. In the meantime, it should quench hopes of rapid progress to note that more than 90% of the universe may be in unknown entities whose individual masses are uncertain by 70 orders of magnitude.

4 - Discrete objects at high z

Crucial in clarifying our ideas on galactic evolution are observations of galaxies out to $z = 1$, a look-back time $> 50\%$ of the time since the Big Bang. These observations are crucial also for classical cosmology and the measurement of the deceleration parameter q. The evolutionary correction will only be understood when we have better data on high-z galaxies, and only then will there be any progress in "geometrical" cosmology. The emphasis among cosmologists seems to have shifted, owing to the realization that evolutionary corrections are large and uncertain, towards attempts to determine Ω rather than q. However in the long run one would like to determine *both* of these quantities, because it is only in the standard Friedmann models that one can assume $\Omega = 2\,q$. Ideally, one could test general relativity on a cosmic scale by seeing if this relation is indeed fulfilled.

To probe back to earlier epochs, one must utilize active galactic nuclei, detectable out to redshifts of order 5. If quasars are indeed active galactic nuclei, one can infer that galaxies must have been assembled, at least to the extent of already having well-defined centres, by the epoch corresponding to $z = 3$. Quasars and radio galaxies display a strong evolutionary trend, indicating that young nuclei had a greater propensity to give rise to active outbursts. This could mean that there is more uncondensed gas in a young galaxy available for fuelling a central compact object.

Even if quasars are not properly understood, their optical spectta provide valuable probes for the intervening gas. Most of the absorption-line systems displaying large velocities relative to the quasar are due to intervening clouds apparently unrelated to the quasar. The amount of material needed to give absorption features is rather small. They could be due to cool filaments in the halos of young galaxies, or to clouds

embedded in a hot galactic medium. The distribution of absorption lines in quasar spectra thus provides, in principle, evidence on the process whereby gas gradually condenses into galaxies or is ejected from galaxies.

The most remote quasars so far observed are at distances such that we are looking back 80% of the time to the Big Bang when we study them. Even so, we remain completely ignorant about the whole range of redshifts between 5 and 1000, (or, equivalently, the range of times between 10^6 years and 10^9 years) as illustrated in Figure 4. During this era the contents of the universe must have gradually transformed from the almost homogeneous gas decoupling from the fireball at recombination into bound galaxies or dark discrete objects (cf. Figure 3). This era has received disproportionally little attention from theorists compared to that lavished on the first million years, or even the first minutes or microseconds.

I have mentioned that one might distinguish the scenarios for galaxy formation by studying the details of clustering at the present epoch. But a more direct method would be to probe redshifts greater than 5. One would like to know whether discrete objects, such as quasars, or maybe young galaxies in the infrared, may be detectable at these redshifts.

5 - The early universe

Where the initial fluctuations come from is still a mystery. There is still no fully plausible idea which explains why our universe contains inhomogeneities of large enough amplitude to yield the presently-observed structure, yet is not so "chaotic" that the overall "Robertson-Walker" geometry is invalid. We can make direct inferences about the amplitude of fluctuations on the scale of galaxies and clusters; but it might offer some clues to the underlying mechanism if we had evidence about the spectrum beyond this mass range. The angular structure of the microwave background can in principle reveal Doppler shifts and gravitational perturbations at the epoch of last scattering, which may be at a redshift as large as 1000. In any theory these perturbations would be linear on scales exceeding 10^8 solar masses; they may extend up to much larger scales even to those of superclusters. These fluctuations could have been imprinted as initial conditions or via exotic processes at very early times. Any quadrupole dependence (or indeed any fluctuations on angles exceeding a few degrees) corresponds to scales exceeding the horizon size at last scattering. These are therefore "acasual" fluctuations, in the context of a classical Friedmann

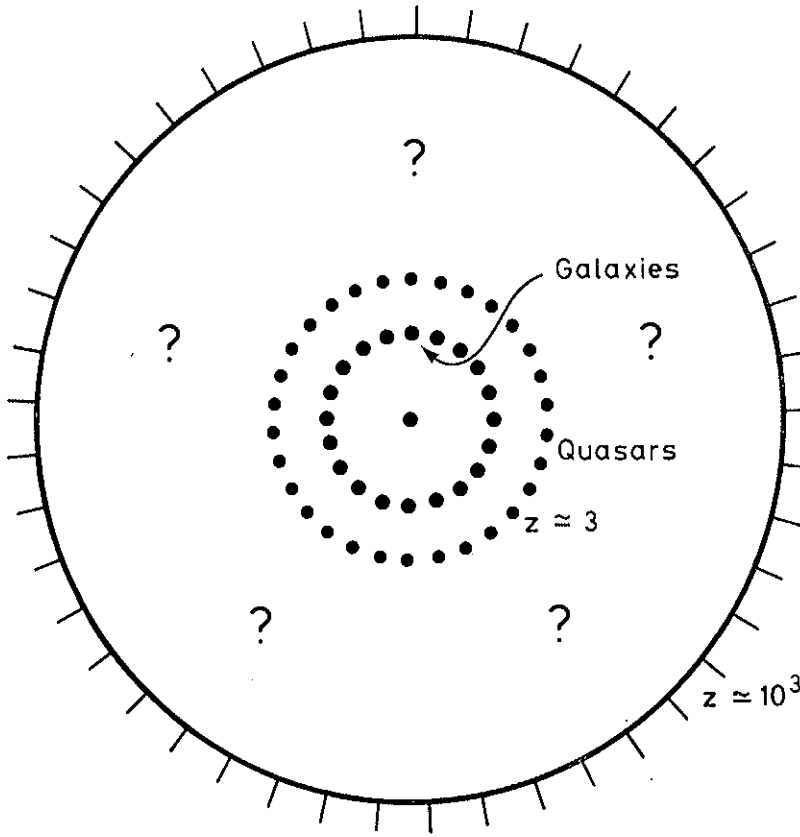


FIG. 4. This diagram illustrates various "redshift shells" in terms of the Robertson-Walker radial coordinate r. The unknown region $5 \lesssim z \lesssim 1000$ corresponds to the era $10^6 \lesssim t \lesssim 10^9$ years. This can be probed by studying fluctuations in the microwave background; if we are fortunate, young galaxies or protogalaxies may be detected as discrete sources in the infrared or X-ray band. Further evidence on this region would help to distinguish between the two scenarios for galaxy formation shown in Figure 3.

model. Microwave background measurements are now tantalizingly close to the level of precision where a wealth of positive anisotropy measurements may become available.

The microwave background is, of course, our main observational link with the physics of the early universe. The microwave photons do not provide such direct evidence of redshifts $\gg 1000$. In the standard picture the close equilibrium between photons and electrons, and the large scattering optical depth, would have erased all traces of events at earlier

epochs. But there are other processes where equilibrium breaks down at early stages. The most familiar example is the way in which the neutron-proton ratio "freezes out" at a temperature of about 1 MeV. As is well known, the most compelling reason for being serious about extrapolating the hot Big Bang model back to $t \lesssim 1$ second is that the simplest assumptions (i.e. homogeneity, isotropy, no new physics, Einstein's relativity, etc.) yield a calculated helium abundance in gratifying accordance with observations. Most cosmologists are so convinced of the tenets of primordial nucleosynthesis that they feel some confidence in constraining Ω_{baryon} on the basis of abundance data on ^{3}He and D, and even in drawing conclusions about particle physics, such as limits on the number of neutrino species.

6 - The very early universe indeed

The success of primordial nucleosynthesis has emboldened some physicists to extrapolate the hot Big Bang back to still earlier times, where the physics is more uncertain. The only mandatory stopping-place for such extrapolations is the Planck time, where quantum gravity effects are crucial. If the expansion had indeed followed a Friedmann model ever since the threshold of classical cosmology, kT would exceed 1 GeV for the first microsecond. During the initial stages the particle energies would sweep down through the entire range of interest to theoretical high energy physicists, including of course the ultra-high energies unattainable by any feasible terrestrial accelerator. In effect the universe provides us with a giant but cheap accelerator (or at least one which is not being charged to us). However, it shut down ten billion years ago. The only surviving fossils of the high energy era will be related to processes which fell out of equilibrium at that stage. For the first 10^{-36} seconds the thermal energies were high enough that the massive X-boson postulated in GUT theories would have existed. Following a prescient suggestion of Sacharov, many people have recently explored the exciting possibility that the baryon content of the universe — the ratio of the number of photons to the number of baryons — can be explained in terms of GUT theories.

To generate a non-zero baryon number requires three things:

a) *Temperatures such that kT exceeds $m_x c^2$.* This happens for $t \lesssim 10^{-36}$ seconds in a hot Friedmann cosmology;

b) *C and CP must be violated.* This is permitted in GUTs, and is necessary in order that baryons be favoured over anti-baryons;

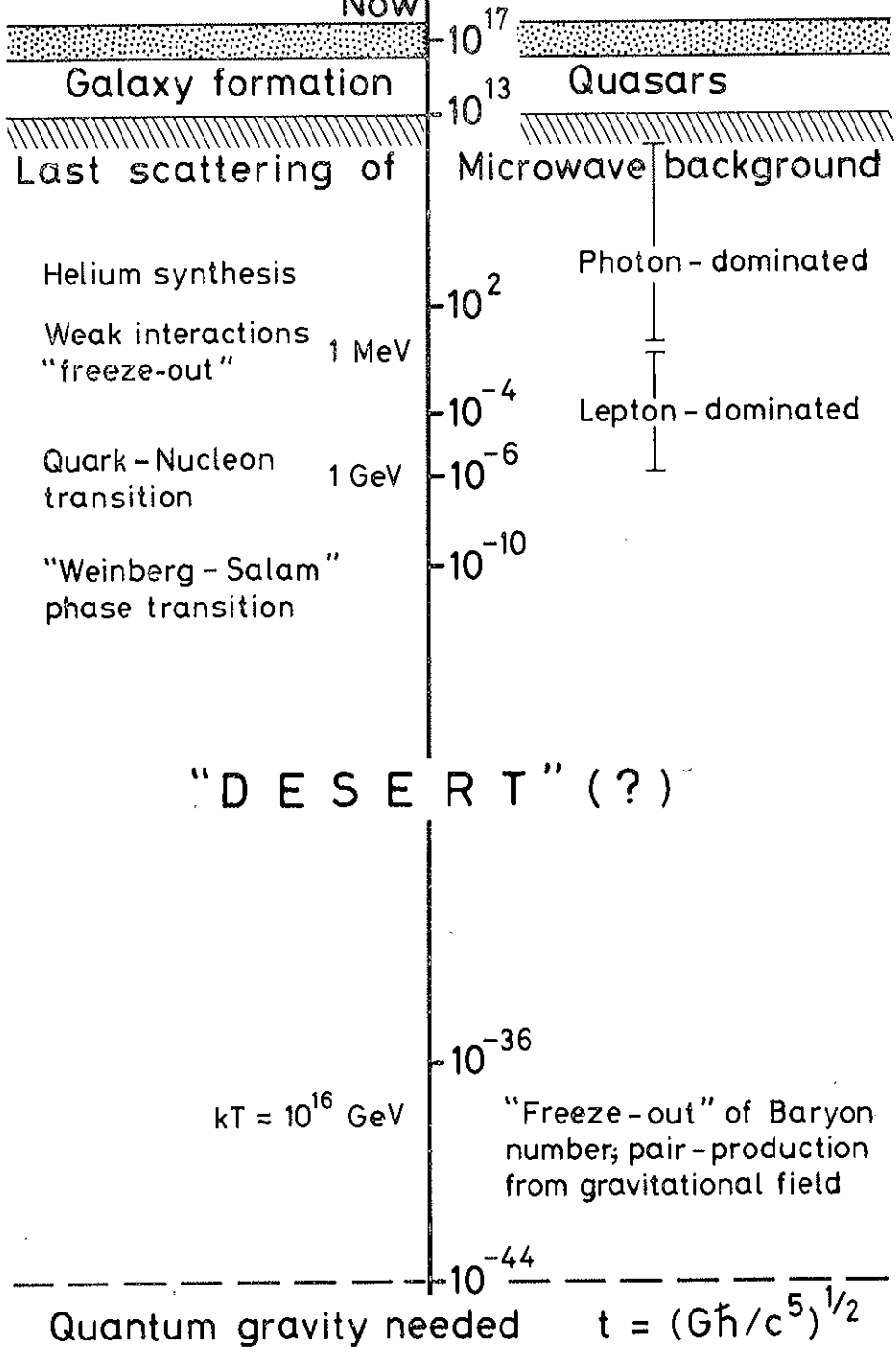


FIG. 5. This diagram illustrates, in terms of logarithmic time, various key physical stages in the expansion of a standard Big Bang model. 60 "decades" separate us from the Planck time. Observations of individual sources, even if they extend out to $z \simeq 5$, permit us to probe only the last decade (stippled region of diagram); the last scattering of the microwave background may have occurred when the universe had only $\sim 10^{-4}$ of its present age: primordial nucleosynthesis yields evidence on physical conditions when $t \simeq 1$ sec. The crucial consequences

c) *Thermal equilibrium must be violated.* This happens because the rates are slower than the expansion rate of the universe, and is an essential requirement (just as the slowness of the weak interactions at the nucleosynthesis era causes the matter to emerge as hydrogen and helium rather than being transmuted all the way into iron).

The value of the net baryon excess can be computed, given a specific GUT, and given also the CP-violation parameter. This work is not yet on the same footing as the nucleosynthesis of helium and deuterium. It is perhaps at the same level as nucleosynthesis was in the pioneering days of Gamow and Lemaitre. But if it could he firmed up it would represent an extraordinary triumph. The mixture of radiation and matter characterizing our universe would not be "ad hoc" but would be a consequence of the simplest initial conditions. Also, as well as vindicating a GUT, it would vindicate an extrapolation in one bound, based on a Friedmann model, right back to the threshold of classical cosmology — almost back to the Planck time. On a logarithmic scale, this is a bigger extrapolation from the nucleosynthesis era than is involved in going to that era from the present time (see Figure 5). It would also place constraints on dissipative processes arising from viscosity, phase transitions, black hole evaporation etc., which might occur as the universe cools through the "desert" between 10^{15} and 100 GeV. Although these ideas are still speculative, one could perhaps be more assertive and claim the "prediction" of the photon/baryon ratio as one of the few likely observational tests of GUTs.

The Robertson-Walker hypersurfaces of our universe are smooth only in the sense that the oceans are smooth: superposed on a well-defined mean curvature there are small "ripples", whose origin is entirely unknown. In some people's mind it is however not the small-scale roughness, but the large-scale smoothness, of the universe which poses the main mystery. Why should the universe have been set up at the Planck time with hypersurfaces whose radius of curvature was at least 30 orders of magnitude larger than the horizon then was? We can rather feebly say that this was essential for astrophysical evolution, in that the universe would otherwise have recollapsed quickly, without ever departing from thermal equilibrium, or else expanded so fast that gravitational clustering never occurred. But that seems an inadequate explanation. Moreover, we do not know why the universe is quite accurately "Robertson-Walker". Why did it not avail itself of the other degrees of freedom open to it?

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

In GUT schemes there is a first-order phase transition at a temperature of order 10^{15} GeV. This transition could occur via the formation of "bubbles" whose walls eventually collide with others; alternatively the phase change (though delayed) may occur homogeneously. The differing energy densities of the two vacua would be equivalent to different values of the cosmical constant, or Λ-term. Λ is now very small, corresponding to a mass energy density of less than 10^{-30} gms/cm^3 (or, in other units, to 10^{-44} GeV^4). One would expect the *change* in the effective Λ associated with symmetry breaking to be of the order of the energy density at the grand unified temperature, i.e. 10^{60} GeV^4. Why things should be fine-tuned to a precision better than one part in 10^{100}, so that the *post-transition* Λ-term is so small, is still a mystery. But if the dynamics of the very early universe is effectively dominated hy the Λ-term, then the scale-factor inflates exponentially as in De Sitter cosmology. The idea of an exponential growth phase has appealing consequences. In particular, it suggests that the universe might be large because the curvature radius of the hypersurfaces of homogeneity, growing in proportion to the scale factor, could exponentiate many times during this phase. The large scale of the universe, or its "flatness", could then be accounted for.

There are difficulties with this scheme. In particular, it is not clear whether the universe can remain stuck in the exponential phase for long enough to grow by a gigantic factor and still, afterwards, achieve a "graceful exit" and turn back to a Friedmann phase. Note also that for this scheme to work the heat released during the phase transition must raise kT above $m_X c^2$ so that baryon synthesis can occur after the exponential phase. Optimists might hope that this model could not only account for the universe's overall homogeneity, but also enable microscopic effects to initiate the fluctuations needed to give galaxies and clusters. It would indeed be remarkable if the entire observable universe could have sprung from a microscopic fluctuation. Since no non-zero conserved quantities

are involved (if baryons are not conserved), this almost *ex nihilo* origin for the universe is at least physically consistent.

It is interesting that the Λ-term, regarded by most relativists as an ugly appendage to the theory of relativity, has now been resuscitated in a more modern guise, and plays an important role in recent ideas involving symmetry breaking. Another important consequence of symmetry breaking would be the production of magnetic monopoles. Indeed, naive estimates suggest that monopoles will be produced in greater profusion than is consistent with the persistence of large-scale cosmic magnetic fields. Some mechanism for suppressing or diluting the monopole density is needed. Symmetry-breaking may lead also to the formation of "domain walls" or "strings". The former can be ruled out on the grounds that the effective mass energy of the walls would be too large to be compatible with present-day cosmological constraints. One-dimensional singularities — "strings" — cannot however be excluded. It is unclear whether the topology of the strings would allow them to contract and disappear, or whether on the other hand they are stretched as the universe expands. In the latter case they could contribute $\Omega = 10^{-4}$ even at the present time, and might be the seed fluctuations that trigger galaxy formation.

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

7 - Concluding comments

The underlying assumption that there was a "hot Big Bang" is likely to figure prominently in our discussions. But we should bear in mind that this is still just a hypothesis, certainly not a dogma. It seems consistent

with the data we have, but these data are sparse and ambiguous by the standards of some other sciences. The hot Big Bang model seems more plausible, and less contrived, than any equally specific alternative model for the early universe. This claim would probably be granted even by those who would rate the theory's chance of long-term survival as less than 50%. To focus attention primarily on this model need not imply blinkered dogmatism or blind adherence to conventional physics. The hot Big Bang concept offers a good framework for our cosmological discussions; and it is by exploring its consequences more fully and confronting it with all new data that we can most quickly decide wbether it must indeed be abandoned, or whether it can be developed into a fuller picture of our universe's 10 to 20 billion year history.

It would be easy to conclude this talk with a long list of unanswered questions. I will however restrict myself to questions on which we can reasonably *expect*, and not merely hope for some enlightenment during this Study Week. I have thought of seven such questions.

1. Do we yet know the Hubble constant with a precision better than 25%?
2. Does the large-scale distribution of galaxies yield evidence for preferred scales of superclustering, or for dissipative non-gravitational effects on these scales?
3. On what dimensions are galaxies, and their light distribution, a good tracer for the hidden mass, and what can we infer about the nature of the hidden mass?
4. What is the relation between quasar and galactic evolution?
5. Can we probe the redshift range $z=5$ to 1000 (the epoch $10^6 \lesssim t \lesssim 10^9$ yrs) by finding young galaxies, background fluctuations, etc.?
6. Could improved observations of helium, deuterium, etc. constrain the hot Big Bang model further?
7. Can comprehensible physical processes at $10^{-44} \lesssim t \lesssim 10^{-36}$ seconds account for the overall homogeneity but small-scale roughness of the universe?

Perhaps this last is the most basic question of all. A universe with large dimensions and long timescales, 10^{60} Planck times, is a prerequisite for the cosmic evolution that has led to our existence; overall isotropy and smoothness, and the symmetry and simplicity that ensue, seem a prerequisite for our making at least some headway in our role as cosmologists.

DISCUSSION

OSTRIKER

I wish to ask a question on your characterisation of the differences between adiabatic and isothermal models for the development of structure. You remarked that structure develops at very different epochs in the two pictures. But is it not possible to have the formation of the presently observed galaxies occur in the era $z = 2$ to 10 in *both* pictures? This determining, or even observing, of the period of galaxy formation would not, unfortunately, help us to determine which if either of the two standard models is correct.

REES

I completely agree. It is on other mass-scales that the scenarios differ most. In the "entropy fluctuation" picture many generations of pregalactic systems smaller than galaxies may have evolved, and clustered hierarchically, before galaxies form; clusters of galaxies, in this picture, form *after* individual galaxies. In the "adiabatic" picture, *nothing* of interest happens before galaxies form, but the new-born galaxies *would already be clustered.* However, an era of galaxy formation in the range $z = 2$ to 10 is compatible with both schemes.

II.

LARGE SCALE STRUCTURE OF THE UNIVERSE

H_0, q_0 AND THE LOCAL VELOCITY FIELD

ALLAN SANDAGE
Mount Wilson and Las Campanas Observatories
of the
Carnegie Institution of Washington

and

G.A. TAMMANN
Astronomisches Institut der Universität Basel
European Southern Observatory, Garching

1 - Introduction and historical summary

A) *Form of the Expansion Law*

Beginning with Hubble's (1929) announcement of a linear relation between redsbift and distance, and continuing into the mid-1970's, a major reason to observe galaxies at all distances and in all directions was to establish the *form* of the expansion velocity field in all distance regimes. It was of particular interest to test if the increase of velocity with distance is linear, because this type of field is the only one with two fundamental properties that must be met if the set of perfectly homogeneous Friedmann models is even a first approximation; namely: (1) for any observer in the manifold the observed velocity-distance relation has the identical appearance as that for every other observer at every other place, both in form and in absolute rate; (2) a linear field is the only one that permits a singularity in time (i.e. reversal of all velocities brings all points to a common origin at a common time).

Tests of the linearity were made using cluster galaxies (either the first ten, or the first ranked) as standard candles by Hubble and Humason (1931) and Humason (1936) at Mount Wilson in tbe 1930's, and at Palomar after 1950 (Hubble 1953; Humason, Mayall and Sandage 1956;

Sandage 1972). Until recently the effort was to make observations to ever larger distances, because the goal then was not only to test the form of the field locally, but also to measure the deceleration directly via the look-back-time so as to carry out the hope of Gauss to determine the intrinsic geometry of space *by experiment*. The method (Hubble 1938; Mattig 1958; Sandage 1961, 1963) has been to use the connection between geometry and energy-density provided by General Relativity to estimate the Riemannian curvature itself.

It is now generally understood that this very direct approach of obtaining the kinetic energy-density relative to that of the gravitational field for the measurement of q_0 may be unconvincing for yet some time to come because of luminosity evolution of the standard candles by both the evolution of the stellar content of E galaxies, and perhaps by mergers of cluster galaxies that should make the near and far cluster samples differ.

The end result of the far-field work to date has been to establish the linearity of the expansion with good precision for velocities larger than $v_0 = 3000$ km s^{-1}. Yet with the above *caveat*, there is no clear deviation from linearity of the World Picture (in Milne's terminology) caused by the deceleration (making the World Picture differ from the World Map by *more* than just the finite light-travel time).

It is this failure of the look-back approach which suggested that a local test could be made of the strength of gravity by putting limits on the deviation of the velocity-distance relation of nearby galaxies due to the effect of the Virgo North Galactic Density Anomaly (Sandage, Tammann and Hardy 1972; Silk 1974; Peebles 1976). Hence, since about 1975 the problem has been two-fold: (1) to search for a systematic deviation from linearity for distances that are under the control of the Virgo complex; (2) to determine the value of the mean random motion about the systematic flow, whatever the local pattern may be. The first problem is described by many authors as determining the local "infall velocity" toward the Virgo cluster core.

B) *The Velocity-Distance Relation in Various Distance Regimes*

i. *Large Distances.* Figure 1 shows the observational data for 82 clusters whose redshifts are plotted versus the apparent magnitude of the first ranked cluster members, corrected for the effects of absorption and redshift. The requirement for a linear velocity-distance relation is shown by the line of slope 5, provided that such galaxies have the same intrinsic

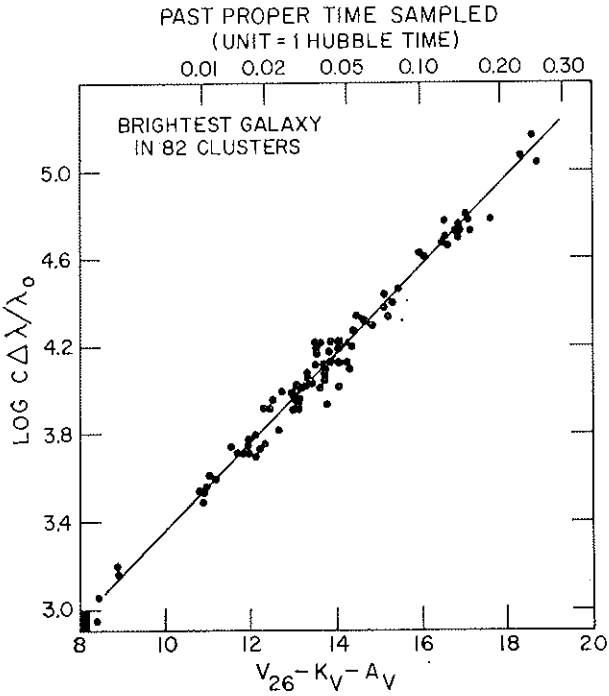


FIG. 1. Correlation of apparent magnitude of first ranked cluster galaxies with redshift of their parent clusters (Sandage 1972). The magnitudes are corrected to isophotal aperture of 26 mag arcsec^{-2} and for the effects of redshift and absorption. The line of slope 5 satisfies the data, showing that the form of the expansion velocity field is linear. The box in the lower left is the very local domain discussed in Sec. 4.

luminosity. The data encompass the redshift range 1000 km s$^{-1} < v_0 \lesssim 140\,000$ km s^{-1} and a look-back time of $\sim 0.3\,H_0^{-1}$ (if $q_0 = 0$) and show, over this range, that the expansion is linear.

The nature of the scatter in Figure 1 shows no increase in the $\Delta \log z$ residuals (read vertically) with increasing faintness. Hence, $\Delta z/z \cong$ constant but, as this would require that velocity dispersion should increase with distance, most if not all scatter in Fig. 1 must rather be due to spread in absolute magnitude, which (read horizontally) is σ (M) $\cong$ 0.3 mag. Hence, an upper limit to the random component of the mean random velocities of cluster centers is $\sigma\,(\Delta z/z) \cong 0.16$. This is an interesting upper limit at redshifts of log $cz = 3.5$ (or $v_0 \simeq 3000$ km s^{-1}), where there are several points in Figure 1, because it corresponds to $c\Delta z \cong 500$ km s^{-1}. The true random

motions are certainly less than this and, hence, less than any systematic motion of $\sim$ 600 km s^{-1} relative to the universe itself given by our anisotropy due to the 3 K radiation (Smoot and Lubin 1979; Boughn, Cheng and Wilkinson 1981).

ii. *Intermediate Distances.* The linearity of tbe expansion field over the smaller distance interval embraced by $3000 < v < 10\,000$ km s^{-1} (or 60 Mpc $< r <$ 2000 Mpc) is sampled with more detail in Figure 2 where data for first ranked E galaxies in groups as well as in clusters are plotted. Again a line of slope 5 fits the data well, including the solid dot at $V_c = 8.2$ that is the Virgo Cluster. As before a study of the residuals shows that $\sigma(\Delta z/z) < 0.2$ is conservative, which puts an upper limit of $\sigma(c\Delta z) = 300$ km s^{-1} for the random motion of group centers as close as $v_0 = 1500$ km s^{-1}. The true value of the random velocity is clearly lower than this [except in the unlikely event that the dispersion in absolute magnitude is actually $\sigma(M) = 0.0$].

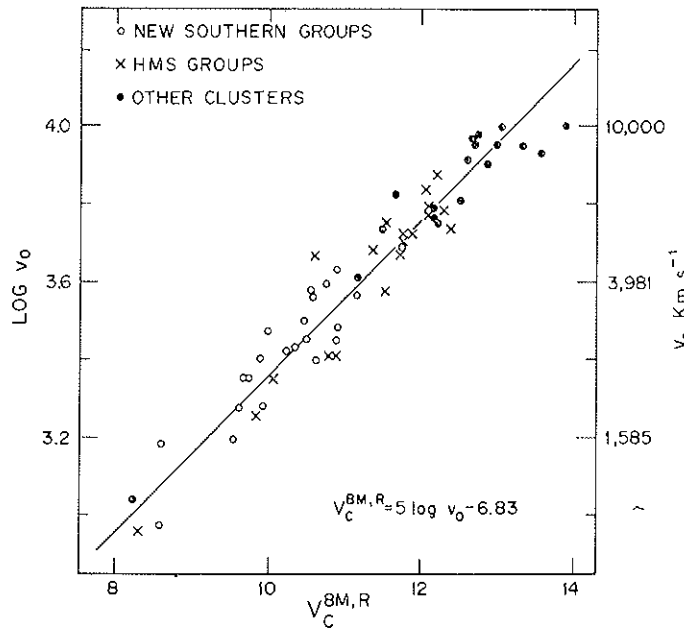


FIG. 2. The Hubble diagram using data from nearby groups added to those from the great clusters (Sandage 1975). As in Figure 1 magnitudes are corrected to an isophotal aperture, for redshift and absorption effect, and further for cluster richness effect.

To guard against objections that the absolute magnitude of the first-ranked cluster galaxy may be a special case and, therefore, a poor standard candle (despite the smallness of the scatter in Figures 1 and 2), Weedman (1976) obtains the same result for the mean of the five brightest cluster galaxies shown in Figure 3, using nuclear magnitudes (suitably defined in his discussion). A line of slope 5 is drawn, again showing the linearity of the expansion field between 1000 and 10 000 km s^{-1}.

As before the Virgo Cluster point lies close to the mean line, brought inward from larger distances, showing that any systematic velocity of the Virgo cluster core relative to the universe as a whole (in a Machian sense of a fixed frame, defined by the global distribution of matter) is small, with limits of only a few hundreds of km s^{-1}. It is the smallness of this value, and also the near zero random motions of other local individual galaxies (Sec. 4), that is perhaps the most astonishing fact of the Hubble flow. Although at various times we have claimed random velocities as low as $\sigma(\Delta V) < 50$, or even 25 km s^{-1} (cf. Tammann, Sandage and Yahil 1979) it would even be astonishing if it were as low as 100 km s^{-1}. On this point Zeldovich once remarked: "One of the most unexpected ob-

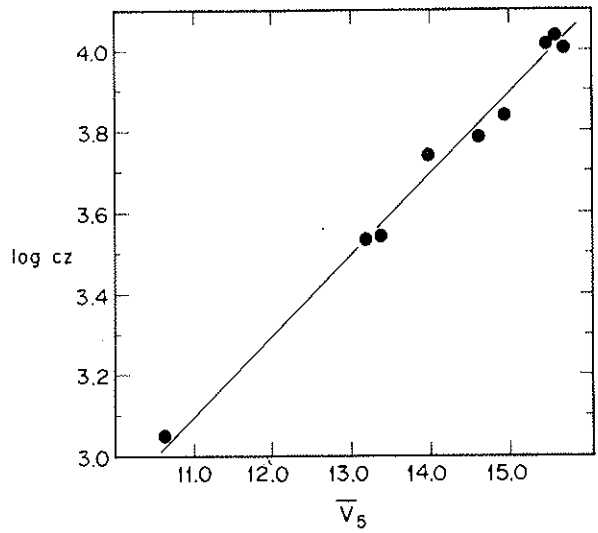


FIG. 3. Weedman's (1976) Hubble diagram for the mean of the five brightest cluster galaxies using nuclear magnitudes. The Virgo cluster point in the lower left shows, as do Figures 1 and 2, that no large velocity perturbation occurs for the Virgo cluster itself relative to the much more distant reference frame.

servational results for a theoretician is that the expansion velocity field is so noiseless in the presence of such great inhomogeneities in the mass distribution". It is, of course, this very nearly noiseless condition that is just the evidence why either $q_0 \cong 0$ (gravity is unimportant) or, conversely, why q_0 could be very large, if a uniformly distributed homogeneous substratum mass were to make the apparent density inhomogeneities less significant. However, in the latter case the Hubble diagram would be expected to deviate so significantly from linearity in its World Picture at large redshifts that this deviation in the observed Hubble diagram (Mattig 1958) would be readily detected. As this does not occur, an upper limit of $q_0 < 5$ seems secure.

iii. *Nearby Distances.* Still closer, within $v_0 \cong 1500$ km s^{-1}, we enter the realm where large systematic velocity perturbations due to the Virgo complex might be expected if q_0 is not close to zero.[1]

We show in Figure 4 (Tammann and Kraan 1978) the velocity-distance relation within 25 Mpc. Plotted are earlier results on galaxies with individual distance indicators (Tammann 1977) and the Virgo cluster with the distance taken from Sandage and Tammann (1976). In addition three other galaxy groups are plotted whose distances relative to Virgo are known (Visvanathan and Sandage 1977). A linear velocity-distance relation is shown with a Hubble constant of $H_0 = 55$ km s^{-1} Mpc. Although the adopted distances need some revision, the scatter read in velocity is small, of the order of 100 km s^{-1}, even this locally.

In Sec. 4 we shall examine the region of the lower left for distances less than $\sim$ 10 Mpc so as to study the perturbation due to Virgo in more detail. The important point to note in Figure 4 is that the perturbation is still small and the Zeldovich wonder becomes even deeper for these regions within the Virgo complex. A new result from the present work in Sec. 4 is that the Virgo density contrast has, in fact, little effect even on the very local velocity field.

C) *Absolute Calibration of the Expansion Rate* (H_0)

Already in 1926 Hubble had a calibration of the mean absolute magnitude of the brightest resolved stars in a few nearby galaxies (M 31,

[1] Recall that 2 q_0 is the ratio of gravitational potential energy to kinetic energy in the expansion, hence, if $q_0 < 1/2$, the kinetic energy dominates and, in the $q_0 = 0$ limit, no perturbation of the velocity field due to Virgo is expected.

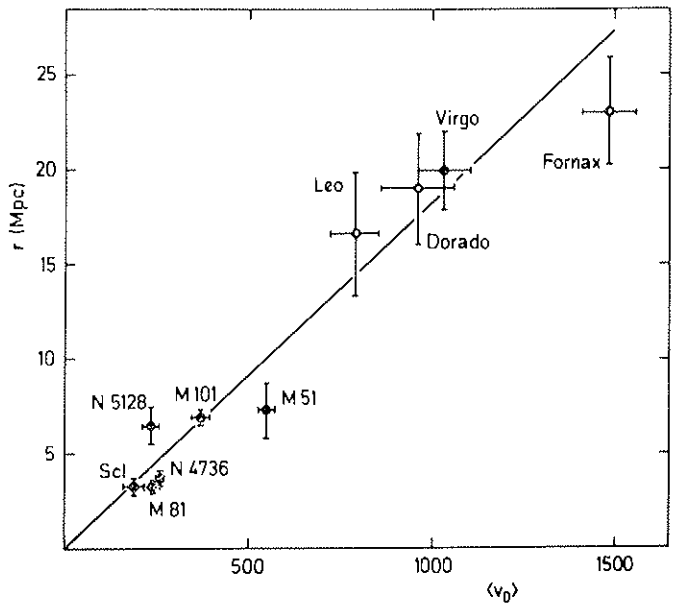


FIG. 4. The very local velocity-distance relation using distances to nearby groups, and relative distances of other more remote groups to Virgo. The even nearer regime for V < 500 km s^{-1} is discussed in Sec. 4.

M 33, LMC, SMC, NGC 6822 and, strangely, M 101; cf. Table XV of Hubble 1926) whose distances he already had from Cepheids. There he derived M_B (brightest star) = − 6.1, a value which was used in his discovery paper of the velocity-distance relation (Hubble 1929) and in his final calibration of 1936.

His best value of the expansion rate in the original 1929 paper was $H_0 = 513 \pm 60$ km s^{-1} Mpc^{-1}. This was revised slightly (Hubble 1936b, Table VI) to 526 km s^{-1} Mpc^{-1} in 1936 and this became the *ex cathedra* value until observational problems belonging to different roots began to appear in the early 1950's.

Hubble believed that he had resolved the brightest stars in spirals in the Virgo cluster beginning at $m_{pg} = 19.0$ in NGC 4321, 19.2 in NGC 4254, etc. (Table III in Hubble 1936a). His calibration M_B (star) = − 6.3 (corrected for Malmquist bias) led him to a Virgo cluster modulus of $(m-M) = 26.8$, fully compatible with $H_0 \cong 525$ km s^{-1} Mpc^{-1}, using a velocity to the Virgo cluster of 1230 km s^{-1}.

In the early 1950's Hubble began a program to strengthen the measurement of distances to the fundamental calibrating galaxies by beginning the Cepheid campaign on the M 81/NGC 2403 group. As Baade (1952) was failing to find RR Lyrae variables in M 31 with the 200-inch reflector, indicating an error even in the first step of the distance scale, Hubble recalled that "the spectroscopists at Mount Wilson believed that my early distances to galaxies were too large [rather than too small] because neither Adams, Joy, nor Merrill had ever seen a star brighter than $M_B = -3$ mag.". At that time the claim of $M_B = -6.3$ in other galaxies was unsupported by any other foundation and hence the value of H_0 even then was in dispute. The principal argument then was that Hubble's assumed stellar magnitudes were too bright, but a second telling point in the opposite sense (that his distances were much too small) was that the time of $H_0^{-1} = 1.9 \times 10^9$ years was far too short compared with the radioactive age of the Earth's crust of 3×10^9 years determined by Holmes in the 1930's. Hubble's 1926 calibration of the distance scale was used as late as 1950. Distances, given by Hubble to Holmberg (1950) for his discussion of nearby groups, are listed in Table 1 to compare with modern values. The largest 1950 distance shown there is D = 2.3 Mpc to the Virgo cluster.

The developments since 1950 that have caused a progressive increase in the distance scale occurred via a series of steps that include: (1) Baade's (1952) proof that the zero point of the Cepheid P-L relation was to he revised brightward by ~ 1.5 mag; (2) Stebbins, Whitford and Johnson's (1950) demonstration that the Mount Wilson magnitude scales beyond $B \cong 16$ need a continuous adjustment faintward by as much as 2 magnitudes at $m_{pg} \cong 19$; (3) new observations of the Cepheids in M 31 hy Baade and Swope (1963); (4) a revised calibration of the P-L and then the P-L-C relation, using newly discovered Cepheids in galactic clusters; (5) discovery and photometry of Cepheids in NGC 2403 in the M 81/NGC 2403 group; (6) use of brightest red stars as distance indicators; (7) discovery that most of Hubble's "brightest stars" in galaxies from his 1936a list fainter than $m \cong 19$ are actually H II regions and/or associations and hence that he had no fundamental data that were based on primary indicators for redshifts larger than ~ 400 km s^{-1}; (8) proof that galaxies even of a given van den Bergh luminosity class have a large range in intrinsic luminosity whose mean $< M >$ is strongly affected by the Malmquist bias in any magnitude-limited sample.

Study of these problems has suggested new values for the moduli.

TABLE 1 - *Comparison of Hubble's 1950 distances with distances adopted here.**

GALAXY (1)	$(m\text{-}M)_{AB}$ Hubble (1950) (2)	D (Hubble) (Mpc) (3)	$(m\text{-}M)_{AB}$ Now (1982) (4)	D (Now) Mpc (5)	D (Now)/D (1950) (6)
LMC	17.1	0.22	18.91	0.52	2.4
SMC	17.3	0.25	19.35	0.71	2.8
M33	22.3	0.24	24.68	0.82	3.4
M31	22.4	0.23	24.78	0.67	2.9
NGC 6822	21.6	0.16	25.03	0.62	3.9
IC 1613	22.0	0.23	24.55	0.77	3.3
WLM	22.3	0.25	25.54	1.28	5.1
M81/NGC 2403 group	24.0	0.52	27.60	3.25	6.2
M101 group	24.0	0.52	29.2	6.92	13.3
Virgo	26.8	2.29	31.7	21.9	9.6

* Sources for Hubble's 1950 distances are Holmberg (1950).
Sources for present distances are Sandage and Tammann (1974a, 1982).
Note that the apparent blue moduli are given in columns (2) and (4) but the "true" distances in columns (3) and (5), using the adopted absorption corrections listed in the cited references.

These are given in column (4) of Table 1. The ratio of the distances in columns (3) and (5) is given in column (6), where the most striking feature is the stretching of the scale, starting as a factor of ~ 2 within the Local Group, and reaching ~ 10 at the Virgo cluster.

2 - H_0 FROM THE BRIGHTEST RED SUPERGIANTS AND SUPERNOVAE

It is evident that in spite of the smallness, as discussed in Sec. 1, of the streaming motions superimposed on the expansion field and of the random peculiar motions of galaxies outside of the large clusters, non-Hubble motions of a few hundreds km s^{-1} could have an appreciable effect on the local value of H_0 and that, therefore, H_0 must be determined at distances corresponding to $v_0 > 5000$ km s^{-1} to obtain the true large-scale value.

In a series of papers (Sandage and Tammann 1974a, b, c; 1975, 1976, 1982; hereafter referred to as ST I-VI) a method was presented whereby H_0 (cosmic) was obtained from several distance indicators which were used consecutively to reach distances where $v_0 \approx 5000$ km s^{-1}. This method relied quite heavily on the small luminosity scatter of spiral galaxies of a fixed van den Bergh luminosity class. This small dispersion has since been questioned (Tammann, Yahil and Sandage 1979; Sandage and Tammann 1981) which weakened the route to H_0. It is therefore helpful that a new and independent way to H_0 (cosmic) can be presented here.

Since the first derivation of $H_0 = 50$ to 55 km s^{-1} Mpc^{-1} occasionally higher values of up to $H_0 = 100$ have been suggested. Reasons why these high values are believed to be unreliable are given in the Appendix.

The present route to H_0 relies on only three distance indicators: Cepheids, brightest red supergiants, and supernovae of type I (hereafter abbreviated as SNe I). These distance indicators fulfill two fundamental criteria: (1) it can be shown independently that they are indeed stable distance indicators; (2) they are particularly insensitive to selection bias. This is true in the case of Cepheids because they follow tightly the period-luminosity-color relation with $\sigma(M) < 0.1$ mag (Sandage and Tammann 1969, 1971). SNe I at maximum B light have a luminosity scatter of at most $\sigma(M) < 0.4$ mag. This translates, under somewhat simplified assumptions, into a systematic magnitude difference (Malmquist effect) between a magnitude-limited and a distance-limited sample of only $\Delta M = 1.38\,\sigma^2 < 0.22$ mag, such that a local calibration of the absolute B magnitude at maximum leads to a systematic distance *under*estimate of distant SNe I of not more than 10%. Finally, the brightness of the brightest red supergiants is apparently governed by a physical guillotine effect, possibly due to mass loss of very massive, evolved stars. Such a fiducial upper luminosity limit is, of course, independent of any selection effect, and hence of distance.

A) *Absolute Magnitudes of Brightest Stars*

Distances to twelve late-type galaxies are known from Cepheids (cf. ST I); five of these lie within the Local Group and the remaining ones (including here the dwarf galaxy Ho IX) are members of the M 81-NGC 2403 group, the distance of which follows from the Cepheids in NGC 2403. [The argument of group membership does not require that the group is bound; it implies only that the galaxies lie roughly at the same distance from the observer, which can be demonstrated independently to within

± 20% (Tammann, Sandage and Yahil 1980)]. To these galaxies the dwarf galaxy Sextans A can now be added (Sandage and Carlson 1982). In addition the very faint magnitude at which Cepheids are found in M 101 sets a firm lower limit to the distance modulus of the M 101 group of $(m-M)^0 > 29.0$, and probably > 29.2, a value which is also required by the brightest blue stars in M 101 (cf. also Humphreys 1980b) and its companions, NGC 5474 and NGC 5585 (ST III). The distance of the M 101 group is included here for completeness and has no consequences for the following results.

A first indication that the Cepheid distances are nearly correct comes from the fact that other authors, using quite different reduction procedures, have arrived at only slightly smaller distances. The distances derived by de Vaucouleurs (1978a, b) agree on the average for Local Group galaxies to within 16% and, for galaxies in the M 81 group, to within 12%. Van den Bergh's (1977) distances for 6 Local Group members are smaller in the mean by not more than 12%.

It is even more significant that new ways to obtain the distances of the Magellanic Clouds confirm the adopted Cepheid distance moduli to within ± 0.2 mag (i.e. 10% in distance). This new evidence comes from Graham's (1973, 1975, 1977) RR Lyrae stars, from infrared magnitudes of Cepheids (Martin, Warren and Feast 1979; Feast 1977), from Mira stars in the LMC (Glass and Evans 1981) and from OB stars (Crampton 1979; Crampton and Greasly 1981).

The zero point of the P-L-C relation of Cepheids still rests, via the calibration in open clusters, on the old Hyades modulus of $(m-M)^0 = 3.03$ (van Bueren 1952; Wayman, Symms and Blackwell 1965; Eggen 1979). If a larger distance modulus of $(m-M)^0 = 3.30$ (e.g. van Altena 1974; Hanson 1975; cf. McAlister 1977) were adopted, the distance increase would enter with almost full weight into the Cepheid distances. However, the resultant distance increase of 13% (0.27 mag) would be difficult to reconcile with Graham's RR Lyrae stars in the Magellanic Clouds (cf. Tammann, Sandage and Yahil 1980), and also with the LMC Cepheids of Martin, Warren and Feast (1979), whose zero point rests at least partially on Balona's method, similar to the Baade-Wesselink method, to derive parallaxes from pulsating atmospheres. We, therefore, conservatively base the following discussion on the old Hyades modulus.

The data for the three brightest blue and red stars in galaxies with distances known from Cepheids are compiled in Table 2. Column (2) lists the apparent blue distance modulus, column (3) the adopted galactic absorption in the blue, column (4) gives the absolute blue magnitude of the galaxy

TABLE 2 - *The Brightest Blue and Red Stars from Cepheid Distances.*

Galaxy (1)	$(m\text{-}M)_{AB}$ (2)	A_B (3)	$M_B^{0,i}$ (gal) (4)	m_B (3) (5)	M_B^0 (3) (6)	m_V (3) (7)	M_V^0 (3) (8)
Solar Neighb.	. . .	. . .	(− 18.5)[1]	. . .	− 8.84[2]	. . .	− 7.97[3]
LMC	18.91[4]	0.32[4]	− 18.44	9.46[5]	− 9.45	11.07[6]	− 7.76
SMC	19.35[4]	0.08[4]	− 16.81	10.63[5]	− 8.72	11.77[7]	− 7.56
M33	24.68[4]	0.12[4]	− 18.87	15.72[8]	− 8.96	16.70[8]	− 7.95
NGC 6822	25.03[4]	1.08[4]	− 15.84	16.89[9]	− 8.14	16.94[9]	− 7.82
IC 1613	24.55[4]	0.12[4]	− 14.71	16.68[10]	− 7.87	16.95[10]	− 7.57
Sextans A	25.67[11]	0.07[11]	− 13.97	17.88[11]	− 7.79	18.09[11]	− 7.53
NGC 2403	27.80[4]	0.24[4]	− 19.37	18.27[12]	− 9.53	20.07[13]	− 7.67
NGC 2366	27.76[4]	0.19[4]	− 16.53	18.97[13]	− 8.78	. . .	. . .
NGC 4236	27.58[4]	0.02[4]	− 18.21	19.22[13]	− 8.36	. . .	. . .
IC 2574	27.60[4]	0.04[4]	− 16.87	19.77[13]	− 7.83	(20.0)[13]	(− 7.6)
Ho II	27.67[4]	0.11[4]	− 16.58	19.64[13]	− 8.03	(20.4)[13]	(− 7.2)
Ho I	27.63[4]	0.07[4]	− 14.40	19.73[13]	− 7.90	. . .	. . .
Ho IX	27.63[14]	0.07[14]	− 13.45	19.56[14]	− 8.07	(19.5)[14]	(− 8.1)
M 101	29.2 [15]	0.00[15]	− 21.31	18.99[16]	− 10.21	$\geqslant$ 21.2 [17]	$\geqslant$ − 8.0
NGC 5474	29.2 [15]	0.00[15]	− 18.19	20.6 [17]	− 8.6	. . .	. . .
NGC 5585	29.2 [15]	0.00[15]	− 18.26	20.9 [17]	− 8.3	. . .	. . .
							− 7.72 ± 0.06
							σ = 0.17

Sources - [1] Assuming for our Galaxy $M_B \approx -21.^m0$ and that the solar neighborhood is representative for 1/10 of the Galaxy. [2] Mean for ρ Cas, ζ Sco, and HD 134959 (Humphreys 1978). [3] Mean for μ Cep, KY Cyg, and KW Sgr (Humphreys 1978). [4] Sandage and Tammann 1974a (ST I). [5] Feast, Thackeray and Wesselink 1960. [6] Humphreys 1979a. [7] Humphreys 1979b. [8] Humphreys and Sandage 1980. [9] Kayser 1967; Humphreys 1980a. [10] Sandage and Katem 1976; Humphreys 1980a. [11] Sandage and Carlson 1982. [12] Sandage and Tammann 1974b (ST II); Humphreys 1980b. [13] Sandage and Tammann 1974b (ST II). [14] Sandage 1982. [15] Sandage and Tammann 1974c (ST III), 1981. The distance limit is required by the absence of Cepheids in M101. [16] Sandage and Tammann 1974c (ST III); Humphreys 1980b. [17] Sandage and Tammann 1974c (ST III).

corrected for galactic and intrinsic absorption (as listed by Kraan-Korteweg and Tammann 1979), columns (5) and (6) list the mean blue apparent and absolute magnitudes, respectively, of the three brightest blue stars, and columns (7) and (8) give the available data on the mean visual apparent and absolute magnitudes, respectively, of the three brightest red stars.

The magnitudes of the brightest stars in Table 2 are corrected for galactic absorption, but not for internal absorption in the parent galaxy. The latter procedure seems to be justified for two reasons: (1) the brightest stars in a galaxy as seen from an outside observer are expected to be biased in favor of low internal absorption; (2) the internal absorption may increase the magnitude scatter, but does not introduce a systematic effect, as long as the reduction is done consistently. The only stars in Table 2 which are corrected for the full amount of absorption are those in the solar neighborhood; here the three brightest red supergiants appear indeed to be somewhat bright.

The values M_V^0 (3) from Table 2 are plotted in Figure 5 against the absolute magnitude $M_B^{0,i}$ (gal) of the parent galaxies. It is striking that M_V^0 (3) does not depend on the galaxy luminosity, and that the mean value of $< M_V^0 (3) > = -7.72 \pm 0.06$ ($\sigma = 0.17$) applies with little scatter to all calibrating galaxies, from the extreme dwarf galaxy Ho IX ($M_B = -13.45$) to at least NGC 2403 ($M_B = -19.37$), and probably also to the supergiant spiral M 101 ($M_B = -21.31$). This shows by experiment that the upper luminosity limit of red supergiants is a powerful distance indicator, independent of the physical nature of this limit. (It could, for instance, be caused by mass loss whose importance increases with the mass of an evolving star, but the astronomical result of nearly constant M_V (3) is independent of its astrophysical explanation).

The calibrated value of $< M_V^0 (3) >$ can hardly be considered controversial. De Vaucouleurs (1978a) found from his distance scale for the one brightest red supergiant $< M_V^0 (1) > = -7.6 \pm 0.2$, which translates into $< M_V^0 (3) > = -7.4 \pm 0.2$ (cf. Table 3). The latter value is merely 0.3 mag fainter than the present calibration. For this reason it is believed that the present route to H_0 leaves discrepancies between different extragalactic distance scales at the level of only $\sim 15\%$.

Figure 5 shows also the values M_B^0 (3) for the three brightest blue stars plotted against the galaxy absolute magnitude. They show a pronounced correlation, as in the original calibration (ST II), which is naturally explain-

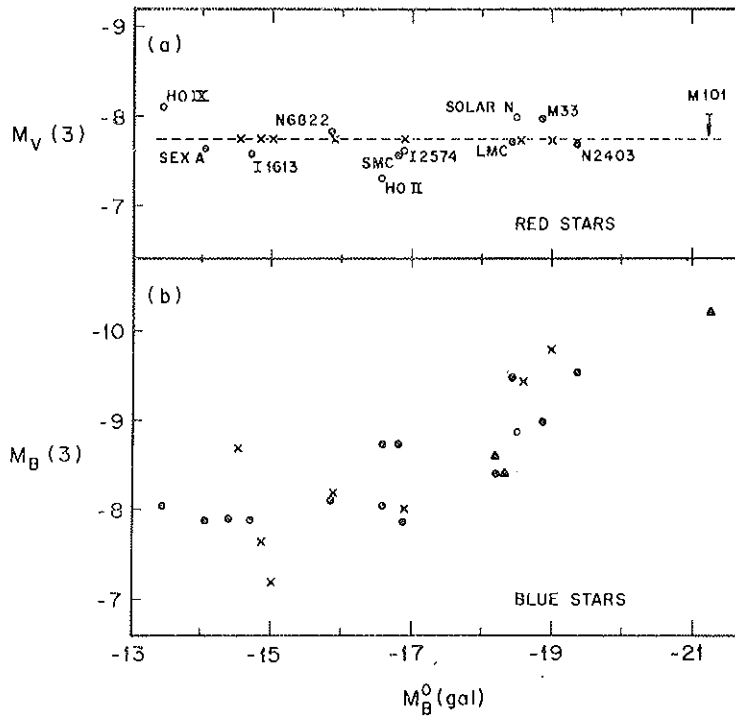


FIG. 5. Upper panel: The mean absolute V magnitude of the three brightest red supergiants plotted against the absolute B magnitude of the parent galaxy. M_V (3) does not correlate with the galaxy luminosity.

Lower panel: The same plot for the mean absolute B magnitude of the three brightest blue stars. Note that here the luminosity of the brightest stars increases with the luminosity of the parent galaxy for M_B^0 (gal) < -17. Dots are calibrating galaxies, open circles are calibrators of lower weight, triangles are members of the M 101 group; the crosses are galaxies whose distances are determined here by requiring that their brightest red supergiants fit the mean relation in the upper panel.

ed as a statistical effect of a normal, tailing-off luminosity function. The correlation begins in galaxies brighter than M_B (galaxy) $\simeq -17$, hence blue stars in such galaxies are nearly useless as distance indicators because the correlation line is almost 45°, which is the well-known degenerate condition for which faint, nearby indicators imitate bright distant ones exactly.

The calibrated mean luminosity of the three brightest red supergiants is used in the next Section to derive distances for seven other nearby galaxies.

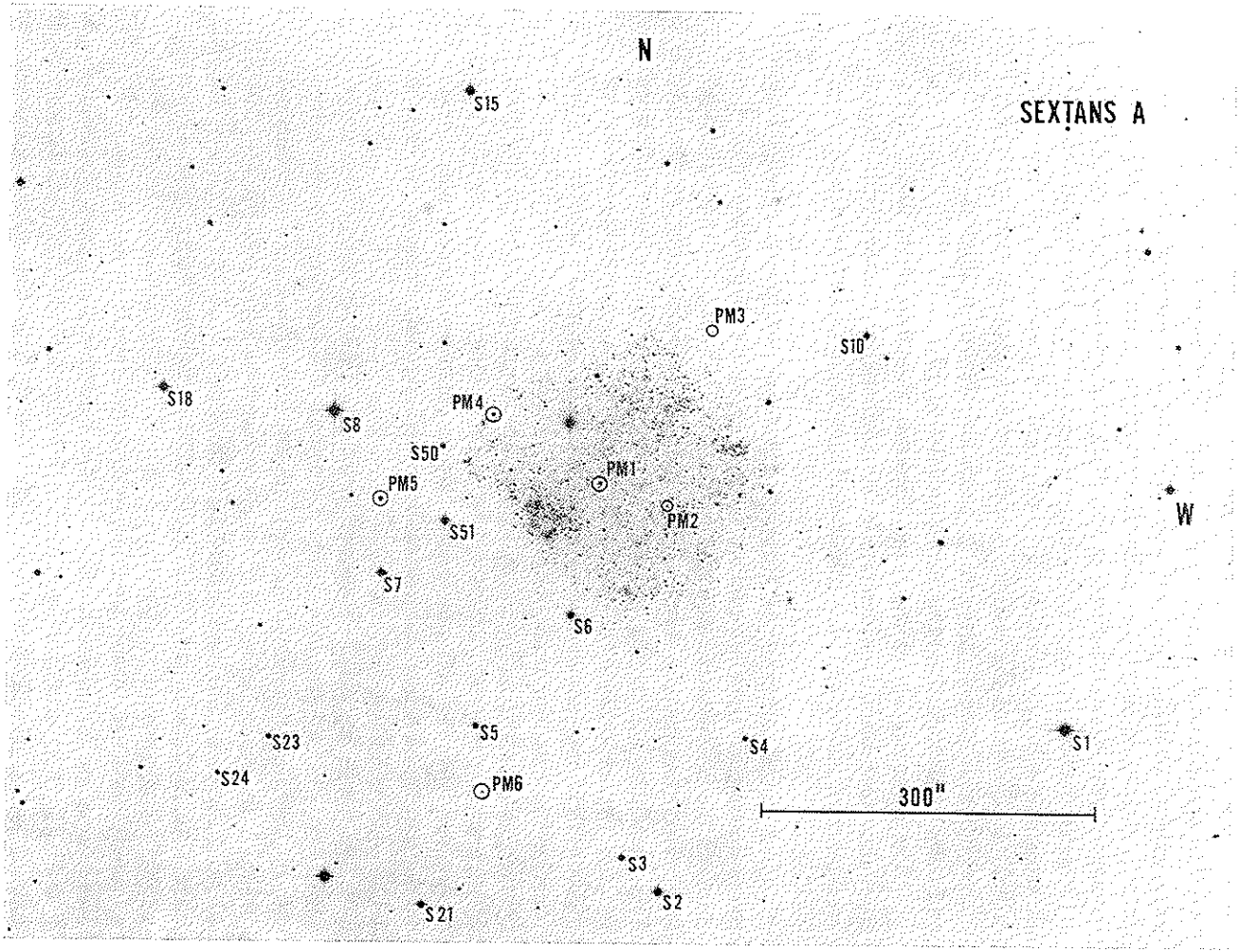
B) *New Distances to Seven Nearby Galaxies*

For seven nearby, highly resolved, late-type galaxies plates in B and V were secured with the 5 m Hale and the 2.5 m Du Pont telescopes. Photoelectric magnitude sequences were set up around each galaxy, and these sequences were extended down to the plate limits using plates taken with a Pickering-Racine wedge. Color-magnitude diagrams were established for all measurable bright stars to $V < 22$, $B < 23$ by means of step scale estimates. Details of the program are published elsewhere (Sandage 1982). Photographs of several program galaxies are shown in Figures 6 to 13, and color-magnitude diagrams in Figures 14 to 19.

The identification of the three brightest blue stars with $(B-V) < 0.4$ from the color-magnitude diagram is straightforward, because the chance occurrence of an equally blue foreground star is highly improbable at these faint levels.

The identification of the three brightest red supergiants is helped by the following of their properties, known from the calibrators: (i) the colors of the brightest red supergiants have $(B-V) > 1.6$; at these red colors foreground stars become scarce; (2)most or all red supergiants are variable. The variability affects the accuracy of these stars as distance indicators but, as the experiment of the calibrators with a magnitude scatter of $\sigma\,(M) = 0.17$ mag shows, the deterioration is by no means severe; (3) for most of the program stars plates exist spanning a time interval of ~ 25 years; this is sufficient to efficiently exclude foreground stars on the basis of their proper motions, which can be unambiguously detected by blinking techniques; (4) the red supergiants have a clear tendency to concentrate in the regions of active star formation, which strongly reduces the area in which they are likely to occur.

The mean apparent V magnitude of the three brightest red supergiants, m_V (3), are listed in Table 3, column 2. The galactic absorption A_V, as determined from the precepts of the RSA (Sandage and Tammann 1981), is shown in column 3. An absorption-free polar cap for $b > 50°$, adopted there, is now well established from new high-latitude X-ray, far ultraviolet and radio data (cf., Heiles 1980; Bohlin, Savage and Drake 1978). The absolute calibration of the red stars, obtained in the last section, then leads to the true distance moduli given in column 4. The mean apparent B magnitude of the three brightest blue stars in column 5 leads to their absolute magnitudes in column 6 via the true distance moduli and a galactic absorption of $A_B = 4/3\ A_V$. The galaxy absolute magnitudes in column 7 are calculated analogously to those in Table 2.


N
SEXTANS A
S15
PM3
S10
S18
PM4
S8
S50
PM1
PM5
PM2
W
S51
S7
S6
S5
S23
S4
S1
S24
PM6
300"
S3
S2
S21

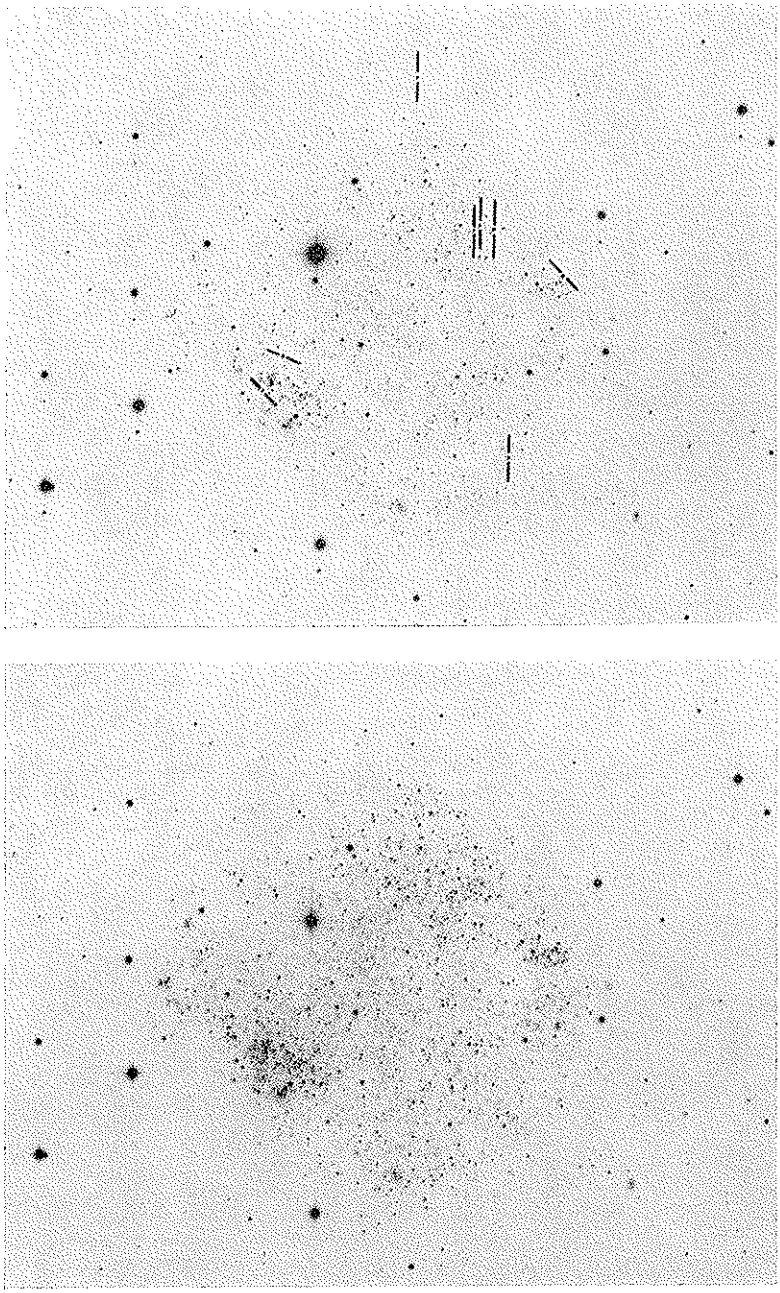


FIG. 7. Blue (right) and yellow (left) photographs of Sextans A. Some bright red stars are marked. The V plate is taken with a Pickering-Racine wedge, which produces secondary images 4.m98 fainter than the primary images.

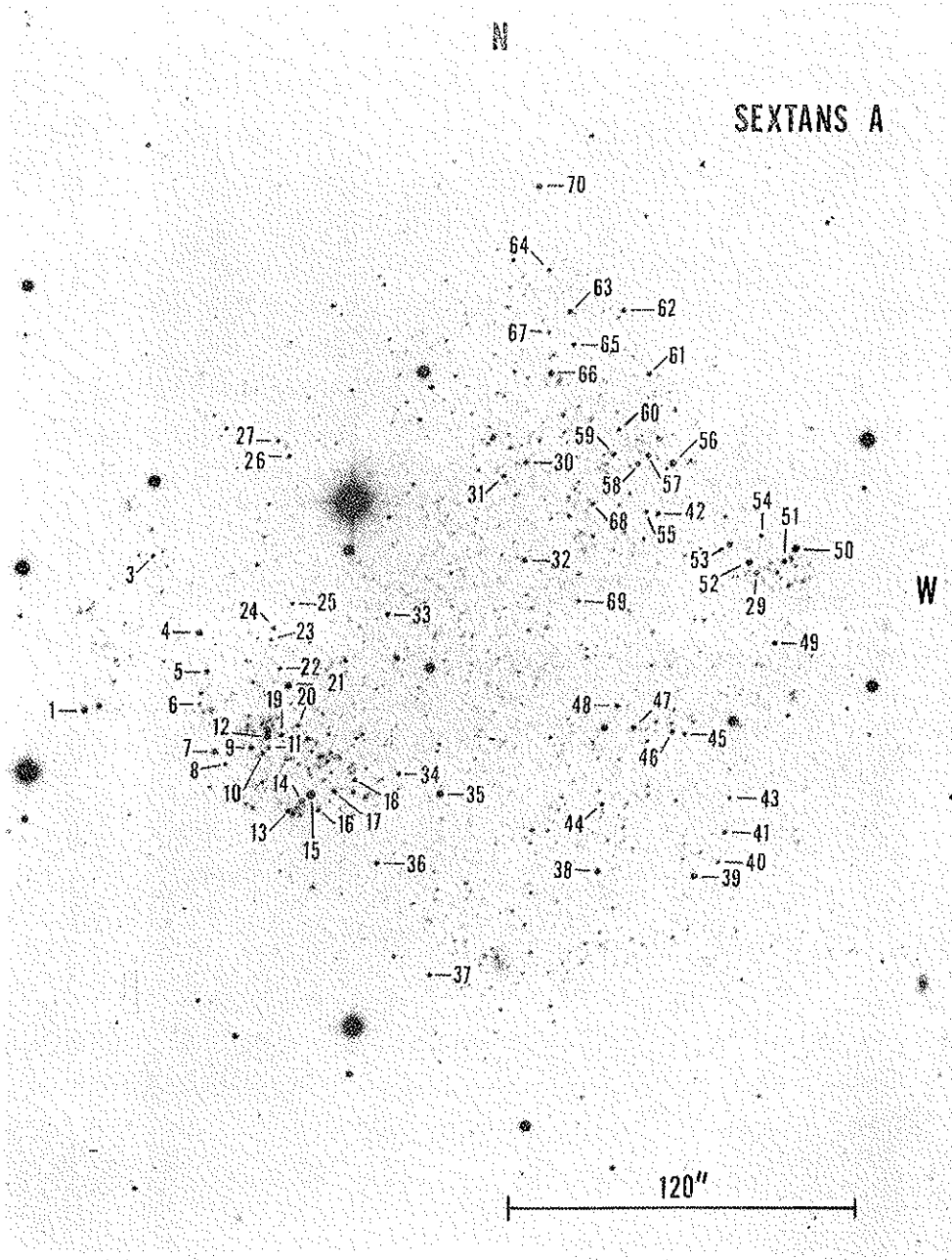


FIG. 8. Yellow (103aD + GC 495) photograph of Sextans A, taken with the Palomar 5 m-Hale reflector and with a Pickering-Racine wedge. The brightest stars in V are marked.

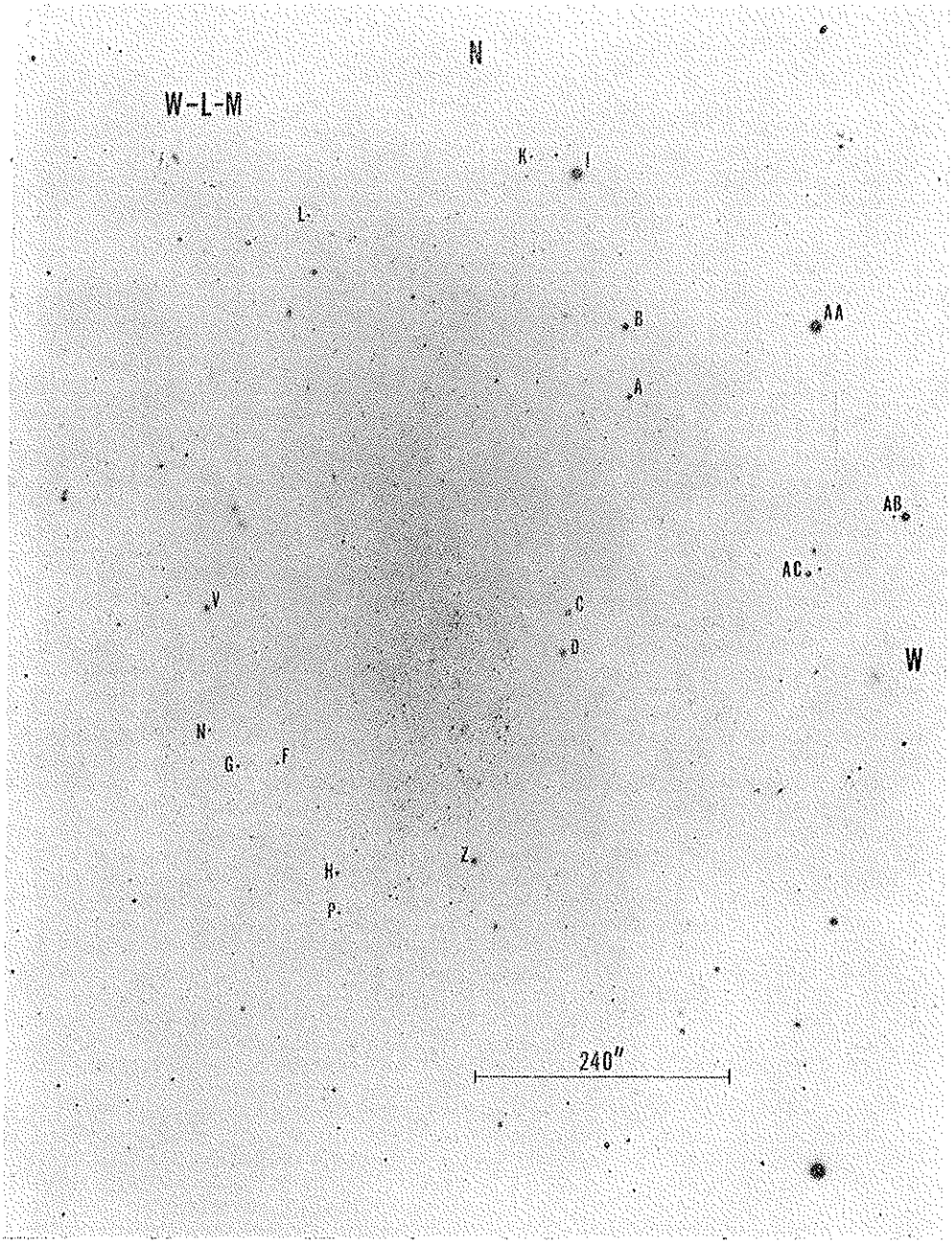
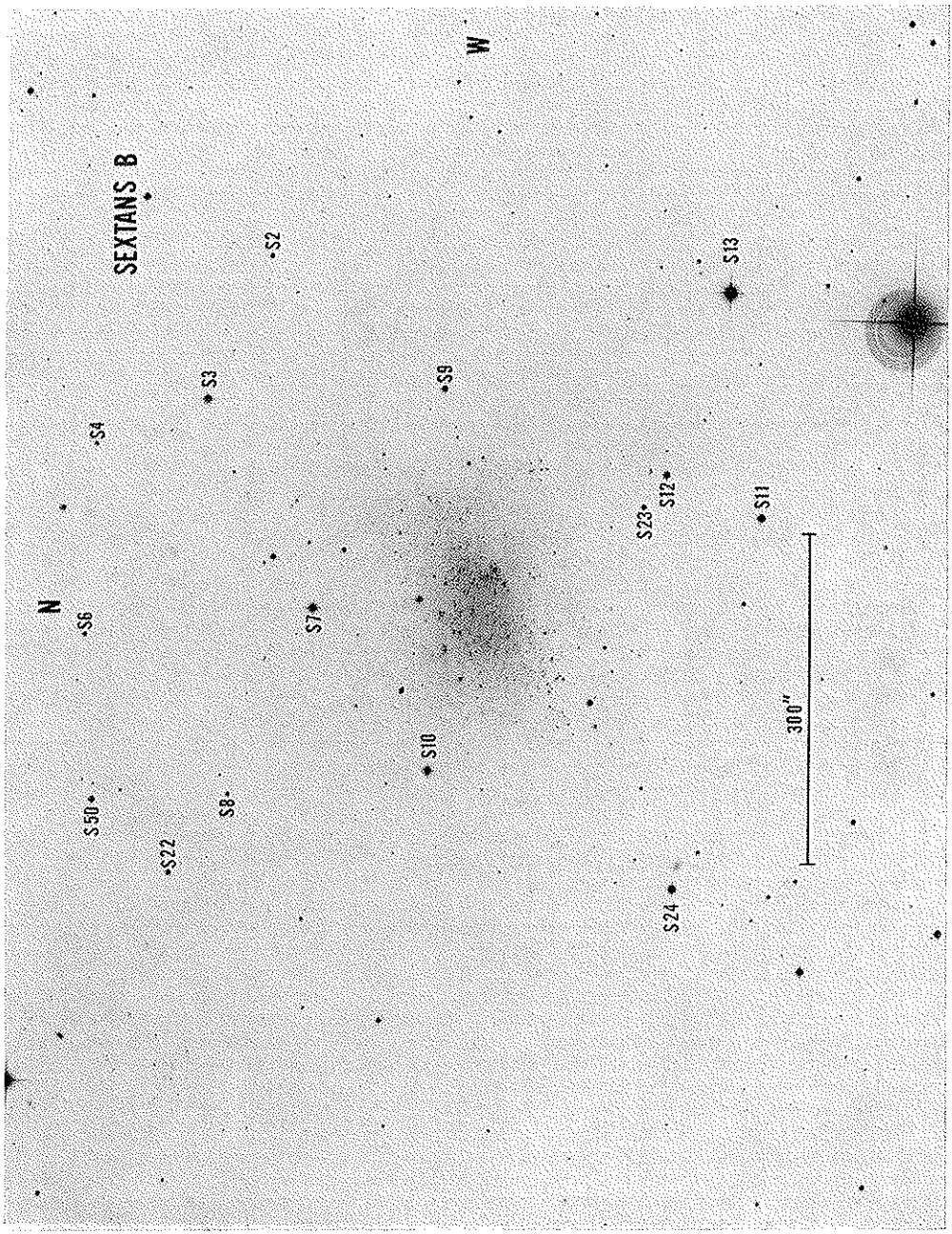


FIG. 9. Blue photograph of Wolf-Lundmark-Melotte (W-L-M), taken with the Palomar 5 m-Hale reflector. The photoelectric sequence is marked.


SEXTANS B
N
W
S50
S6
S4
S22
S8
S3
S2
S7
S10
S9
S13
S12
S23
S11
S24
300"

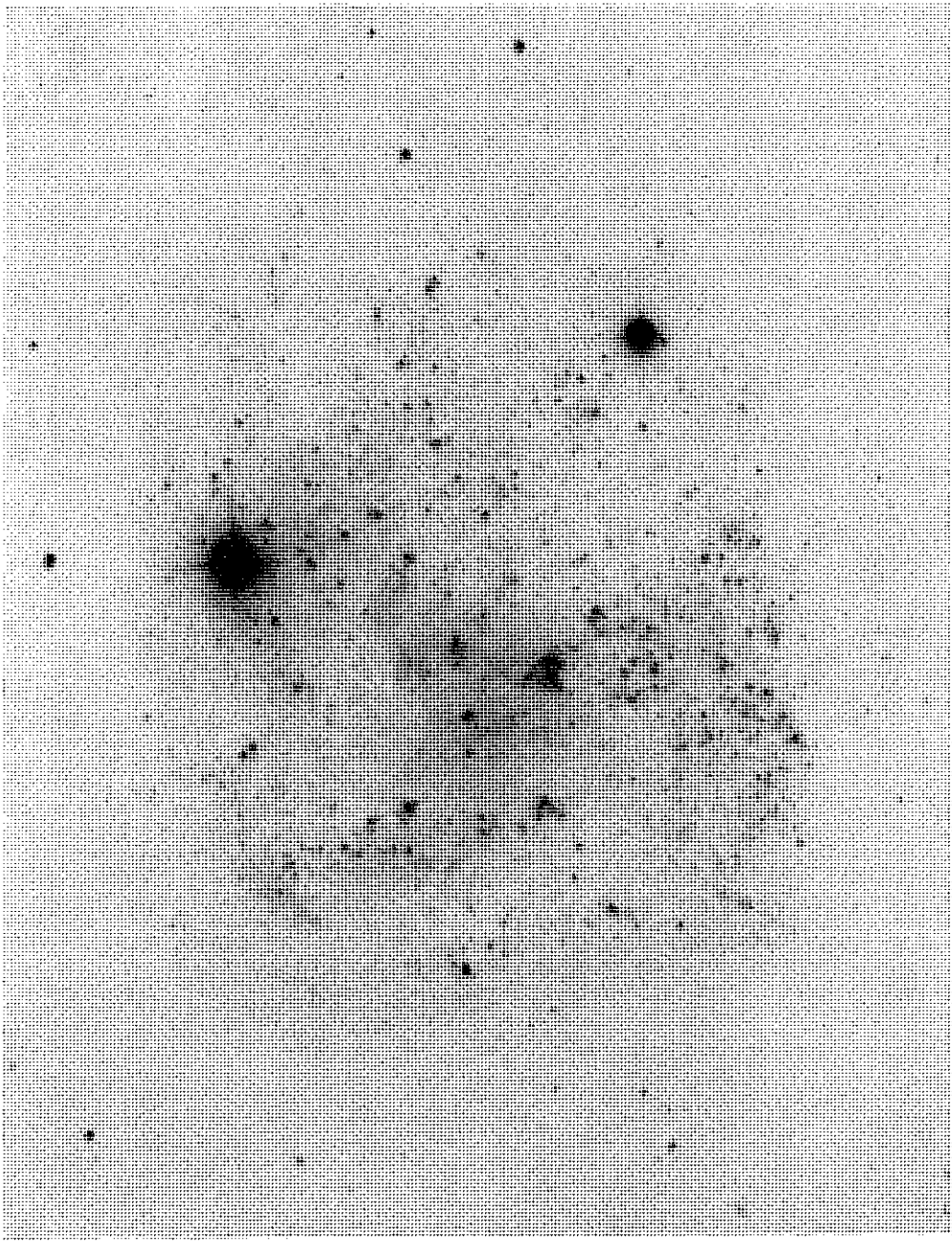
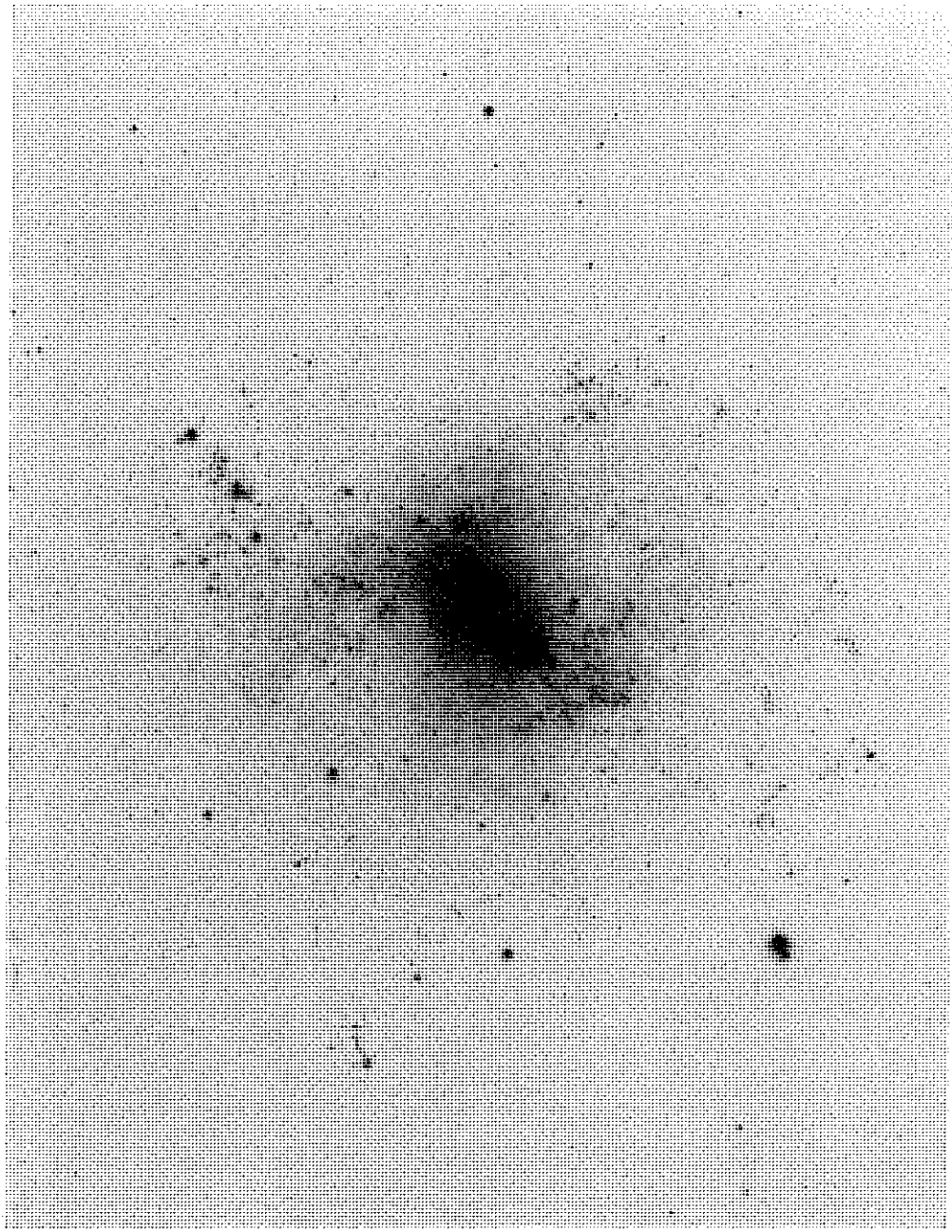


FIG. 11. Blue photograph of IC 4182, taken with the Palomar 5 m-Hale reflector. The galaxy was the site of SN 1937 *c*.



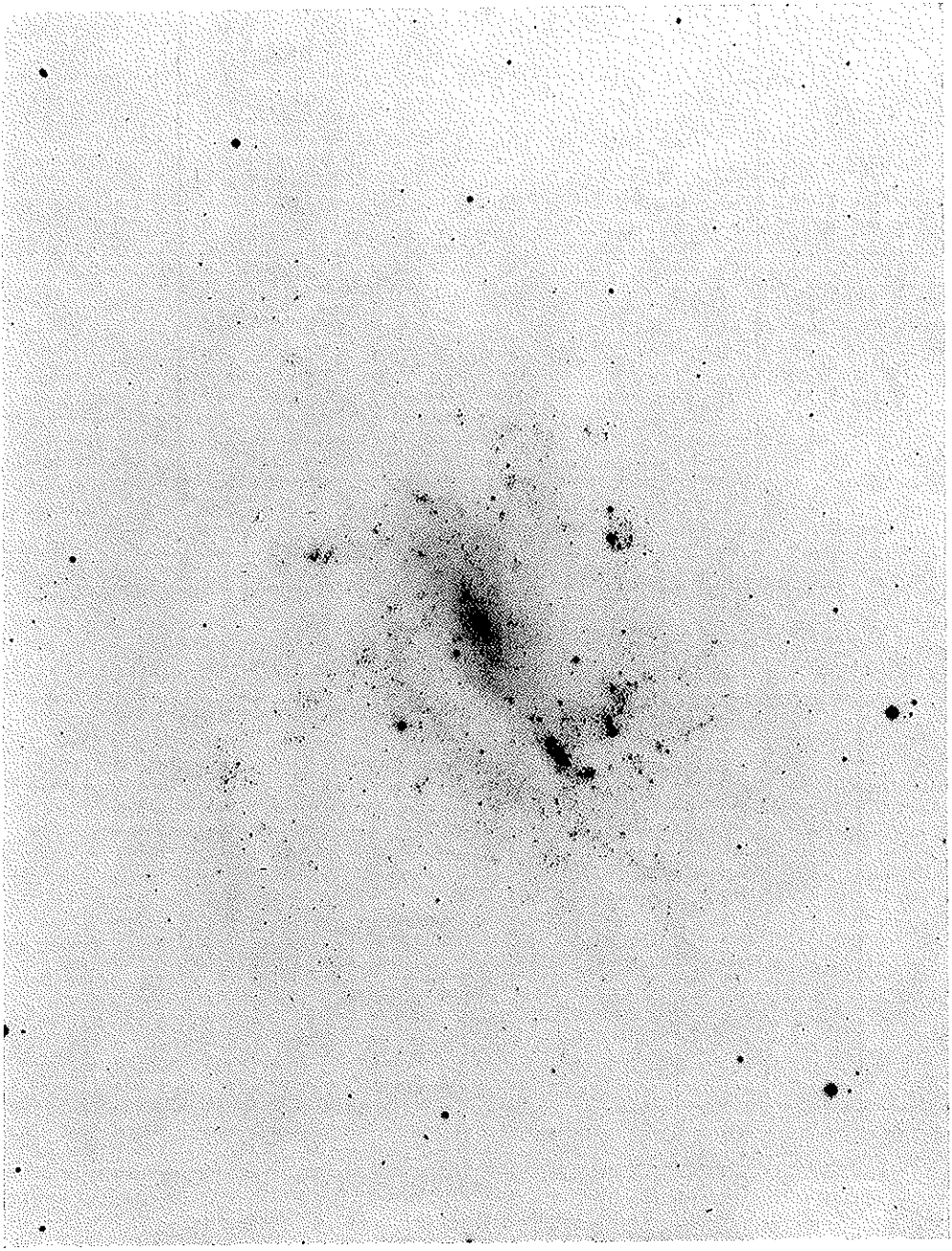


FIG. 13. Blue photograph of NGC 4395, an easily resolved companion to NGC 4214, taken with the Palomar 5 m-Hale reflector.

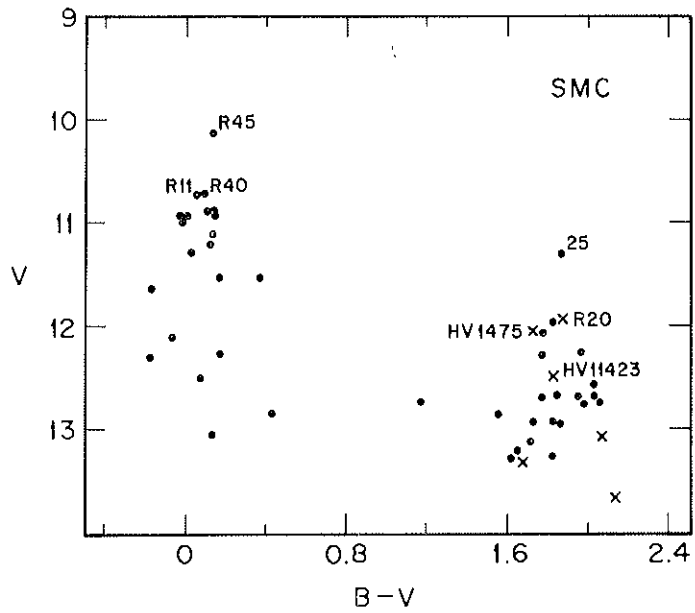


FIG. 14. Color-magnitude diagram of the brightest stars in the calibrating SMC. Variables are shown as x's

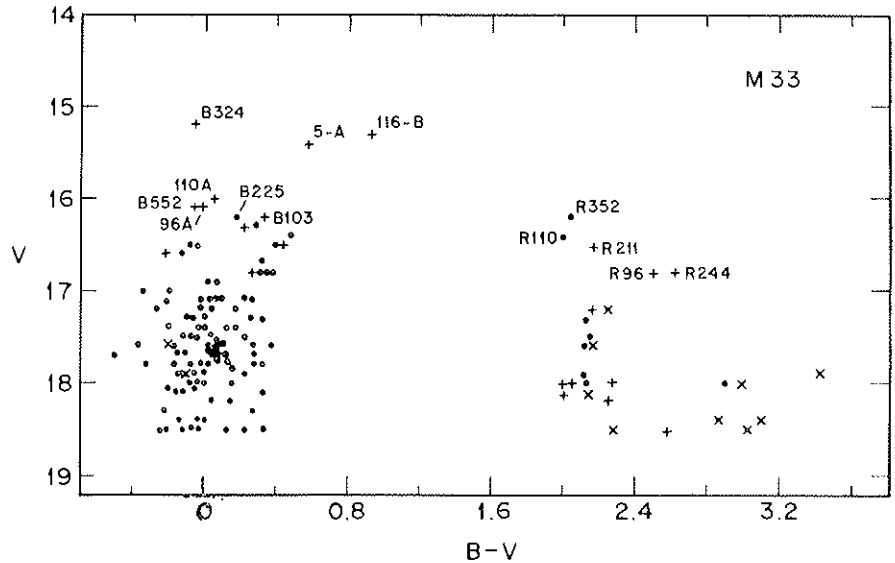


FIG. 15. Color-magnitude diagram of the calibrating galaxy M 33. Variables are shown as x's, stars with known spectrum as crosses.

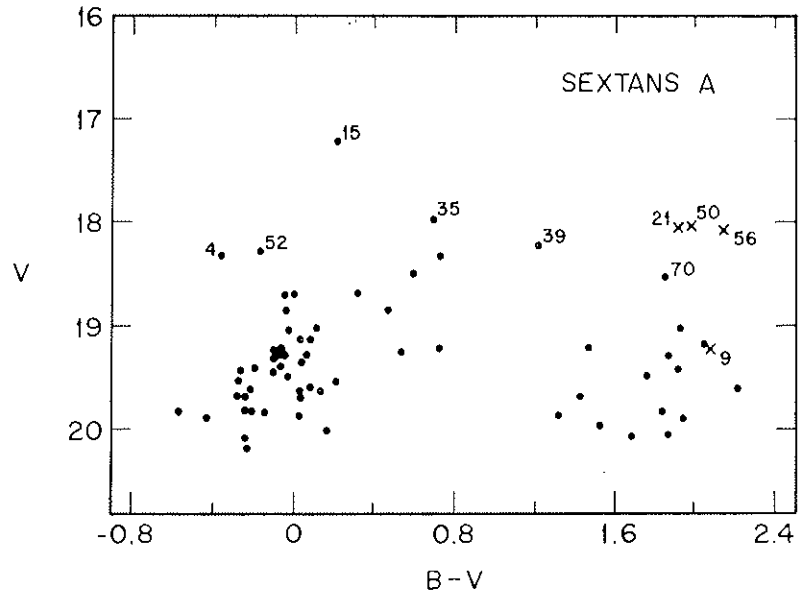


FIG. 16. Color-magnitude diagram of the calibrating galaxy Sextans A. Variables are shown as x's.

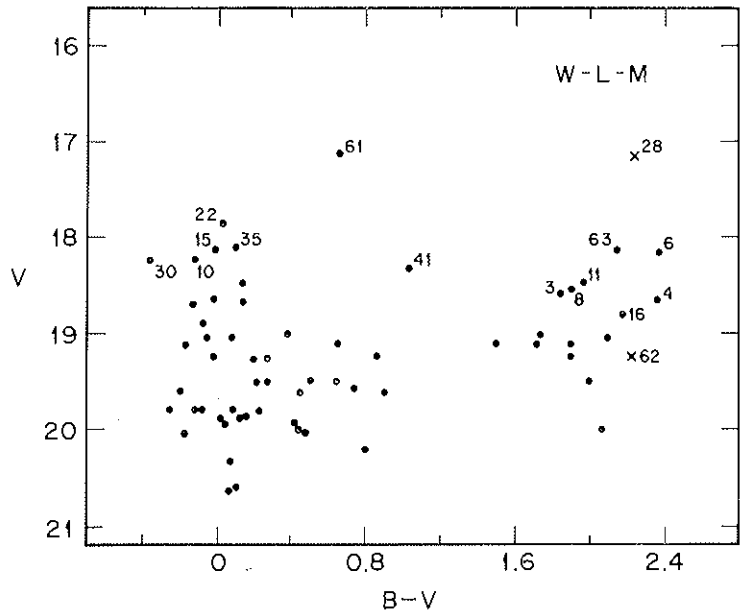


FIG. 17. Color-magnitude diagram of W-L-M. Variables are shown as x's.

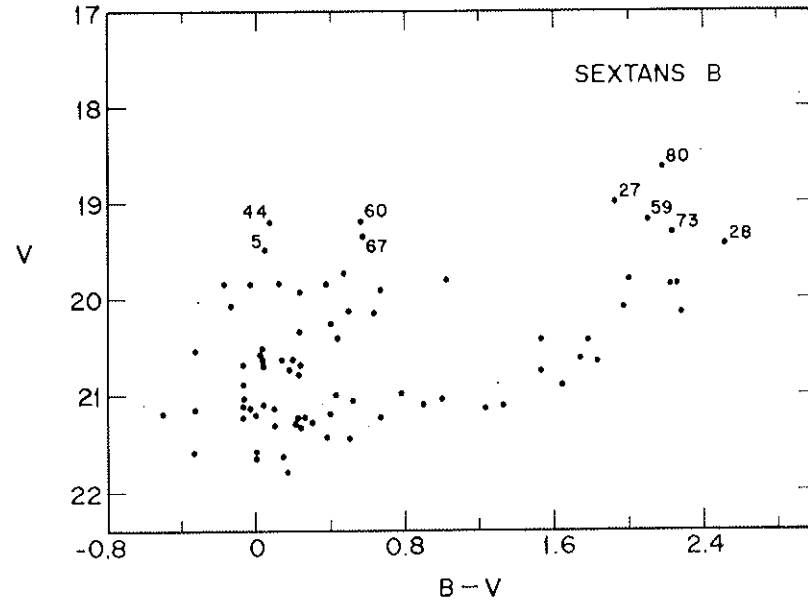


FIG. 18. Color-magnitude diagram of Sextans B.

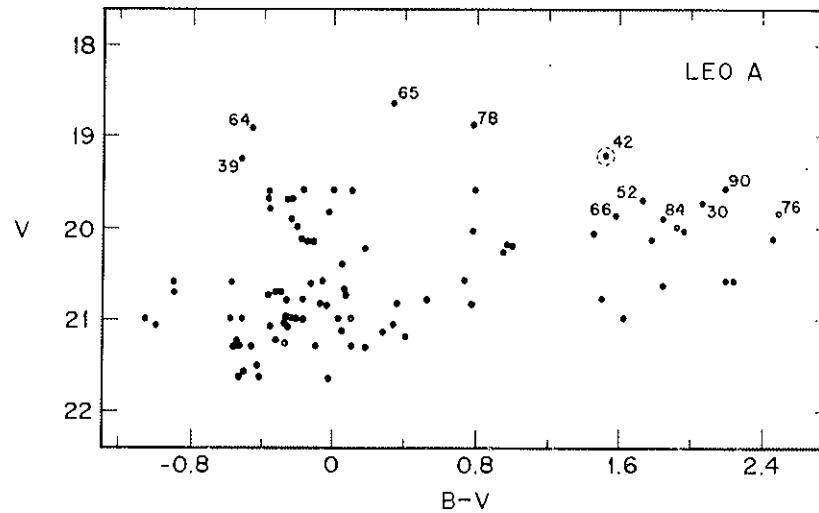


FIG. 19. Color-magnitude diagram of Leo A. Star No. 42 is a foreground star on the basis of its proper motion.

TABLE 3 - *New Distances from Red Supergiants.*

Galaxy (1)	m_V (3) (2)	A_V (3)	$(m\text{-}M)^0$ (4)	m_B (3) (5)	M_B^0 (3) (6)	$M_B^{0,i}$ (gal) (7)
W-L-M	17.82	0.00	25.54	17.94	— 7.60	— 14.81
Sextans B	18.94	0.05	26.64	19.50	— 7.17	— 15.00
Leo A	19.68	0.00	27.40	18.75	— 8.65	— 14.49
Pegasus	20.52	0.05	28.19	20.10	— 8.15	— 15.84
IC 4182	20.49	0.00	28.21	20.24	— 7.97	— 16.87
NGC 4395	21.1	0.00	28.82	19.29	— 9.53	— 18.47
NGC 4214	21.3[a]	0.00	29.02	19.23	— 9.79	— 18.99

[a] The brightest red supergiant has m_V (1) = 21.^{m}1. The mean difference m_V (3) — m_V (1) = 0.^{m}19 (σ = 0.^{m}14) has been applied.

The absolute magnitudes of the brightest blue and red stars in Table 3 are plotted in Figure 5 as crosses. While for the red stars the crosses lie on the mean line by construction, the resulting position of the blue stars within the scatter-range of the calibrators is satisfactory, and gives some additional support to the adopted distance moduli.

The listed moduli and absolute magnitudes are believed to be accurate to 0.2 to 0.3 mag. The error is somewhat larger for NGC 4214, since it has a remarkably high surface brightness, making the photometry of individual stars relatively difficult. Fortunately NGC 4214 has both in projection and velocity space a close, highly resolved and low-surface brightness neighbor, NGC 4395, which almost certainly lies at essentially the same distance as NGC 4214. For this galaxy pair, which belongs to a whole group of galaxies (Kraan-Korteweg and Tammann 1979; their group B 4), a mean distance of $(m-M)^0 = 29.02 \pm 0.3$ is therefore adopted.

C) *Relative Calibration of Type I Supernovae as Distance Indicators*

1. SNe I as Standard Candles

Problems of determining the magnitudes of supernovae at maximum and the internal absorption in galaxies of all types have been reviewed recently (Tammann 1981). Available data had suggested earlier (Kowal

1968) that the dispersion in $< M_B^{max} >$ may be small in well-observed SNe I and, hence, that they might be good relative distance indicators. Once calibrated, they would then become good absolute indicators.

It is now clear that SNe I *appear* to have different absolute magnitudes at maximum, although they represent with all likelihood the same physical phenomenon in E and S0 galaxies on the one hand and spiral galaxies on the other. The SNe I in spirals are consistently fainter and redder and they have larger luminosity scatter than their counterparts in E/S0 galaxies. The reason is simply internal absorption in spiral galaxies, which is absent in E/S0 galaxies. Analysis of 17 SNe I in E/S0 galaxies, either of known redshift or internal to the Virgo and Coma cluster, shows that $\sigma (M_B^{max}) = 0.58$ mag for the complete sample, and $\sigma (M) = 0.43$ mag for a subsample of the 9 best observed SNe I (Tammann 1981). The magnitudes at maximum, corrected for galactic absorption, of the 17 SNe I in Table 5 were taken from a recent compilation (Cadonau and Tammann 1981) of previous data (Barbon, Capaccioli and Ciatti 1975; Pskovskii 1977; Kowal 1978; Branch and Bettis 1978). A conversion of the m_{pg} magnitudes to the B system of $m_B - m_{pg} = 0.24$ mag at maximum has been adopted. The m_B^0 (max) values are plotted in the Hubble diagram of Figure 20, where the six SNe I in Virgo cluster E galaxies and the five in the Coma cluster are combined, respectively.

The SNe I follow a line of slope 5 with high accuracy. Since it is independently known from the discussion in Sec. 1.B that the expansion is linear, Fig. 20 shows that $< M_B^{max} >$ for SNe I is stable to within the scatter quoted above. A least-square fit, using m_B^0 (max) as an independent variable, and forcing a slope of 5, requires:

$$< M_B^{max} > = (- 19.73 \pm 0.24) + 5 \log (H_0/50) \ . \qquad (1)$$

For the 6 SNe I in the Virgo cluster a true recession velocity of $v_0 = 1200$ km s^{-1} has been assumed. This value comes from the observed mean velocity of the cluster of $< v_0 > = 967 \pm 50$ km s^{-1} (Kraan-Korteweg 1982) and an allowance of ~ 233 km s^{-1} for our infall velocity toward the cluster (cf. Sec. 4). An error of ± 200 km s^{-1} of the adopted infall velocity affects the mean absolute magnitude of the 17 SNe I by only ± 0.2 mag. This error has been already allowed for in equation (1). Another check on the stability of $< M_B^{max} >$, which is independent of all velocity data, comes from the fact that the six SNe I in Virgo cluster E galaxies have $< M_B^{max} > = 12.02 \pm 0.18$, again with a small scatter of $\sigma = 0.43$.

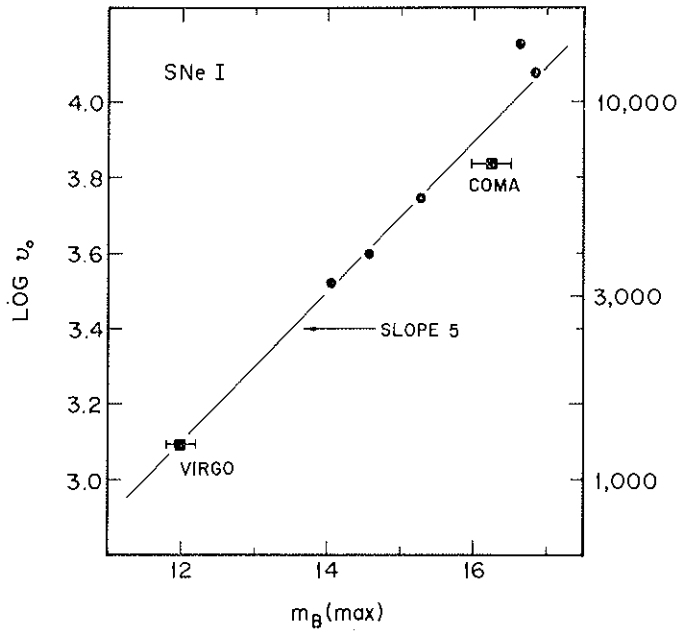


FIG. 20. The Hubble diagram for supernovae of type I at maximum B light. The six SNe in the Virgo cluster and the five SNe in the Coma cluster are combined, respectively. The low-velocity SN 1939a is not plotted. The best fit to the points agrees very well with a line of slope 5, which is required for standard candles.

In view of the poor-quality photometry available for most SNe, it is likely that much of the luminosity scatter is due to observational errors, and that the true scatter is considerably less than $\sigma = 0.4$ mag. This expectation has recently been confirmed in a most convincing manner by Elias, Frogel, Hackwell and Persson (1981). They have obtained extended infrared ligbt curves for three SNe I, two of which lie in the same (!) cD galaxy NGC 1316 and one in the spiral galaxy NGC 4536. The complex light curves are identical to within < 0.1 mag, which is of the same order as the observational accuracy. This not only indicates that SNe I are good standard candles, but that their photometric properties are not dependent on the type of their parent galaxies. SNe I in spiral galaxies are, therefore, equally useful standard candles, once the effect of their internal absorption can be controlled.

D) *The Absolute Calibration of Type I Supernovae and the Value of* H_0

Two of the galaxies in Table 3 with known distances have produced SNe I with well observed light curves, i.e. IC 4182 (SN 1937c) and NGC 4214 (SN 1954a). Although SN 1954a has shown some spectral pecularities, it is no doubt of type I (Oke and Searle 1974).

The relevant data for the two SNe are compiled in Table 4. Baade's (1938, 1941) maximum of $m_{pg}^{max} = 8.2$ for SN 1937c translates to $m_B^{max} = 8.44$ (column 3). The available color information does not allow a correction for internal absorption and, because the SN lies outside the main body of IC 4182, we have conservatively adopted zero absorption. (Note that any absorption correction would lead to a lower value of H_0). Wild's (1960) photometry of SN 1954a, together with the standard light curve of Barbon, Ciatti and Rosino (1973a), gives $m_B^{max} = 10.18$. To estimate the absorption we have taken the shape of the (B-V) light curve from the well observed SN 1972e in NGC 5253 (Ardeberg and de Groot 1973; Lee, Wamsteker and Wisniewski 1972) and the zero point of the color curve from the presumably absorption-free SN 1970j which occurred in NGC 7619, an E galaxy at $b = -48°$ (Barbon, Ciatti and Rosino 1973b). This detour is necessary because the published color curves are not independent of SN 1954a. The resulting reddening for SN 1954a is $E_{B-V} = 0.22 \pm 0.08$, which then leads (assuming $A_B/E_{B-V} \approx 4$ for SNe) to a blue absorption of $A_B \approx 0.88 \pm 0.32$ and to the corrected value of m_B^0 (max) in Table 4, column 3. In column 4 the distance moduli of the parent galaxies are repeated from Table. 3. The resulting mean absolute magnitude of SNe I at maximum in column 5, M_B^0 (max) $= -19.74 \pm 0.19$ can now be inserted into equation (1) to give

$$5 \log (H_0/50) = -0.01 \pm 0.31 \ . \qquad (2)$$

TABLE 4 - *Luminosity Calibration of SNe I using NGC 4214 and IC 4182.*

Name (1)	Galaxy (2)	m_B^0 (max) (3)	$(m\text{-}M)^0$ (4)	M_B^0 (max) (5)
SN 1937c	IC 4182	8.44 ± 0.1	28.21 ± 0.2	-19.77 ± 0.22
SN 1954a	NGC 4214	9.30 ± 0.3	28.92 ± 0.3	-19.62 ± 0.42
				-19.74 ± 0.19

Solving for H_0 leads to

$$H_0 = 50 \pm 7 \text{ km s}^{-1} \text{ Mpc}^{-1} . \quad (3)$$

Because equation (1) is mainly defined by SNe I with $3000 < v_0 < 12\,000$ km s^{-1}, this is clearly the global value of H_0, well beyond the effect of any local velocity anisotropies.[2]

There are two lines of argument which suggest that the present calibration of M_B^0 (max) must be nearly correct:

(1) With this value, the two galactic SNe (presumably of type I) of Tycho (SN 1572) and Kepler (SN 1604) are at distances of 4.0 and 3.2 kpc from us, respectively. These distances are well within range of the published values. SN 1885a in M 31 (S And) must either have been dimmed by 0.6 mag of internal absorption, or it was not of type I (cf. Tammann 1981 for discussion on both these points).

(2) The calibration is also in satisfactory agreement with presently available expansion parallaxes of SNe I and SNe II. The quoted values for SNe I are within the limits $-20.5 < M_B^{max} < -19.12$ (Branch 1977, 1981; Arnett 1981). For the type II SN 1979c in the Virgo cluster Branch *et al.* (1981) found a distance modulus of $(m-M)^0 = 31.8$, which then requires, with $< m_B^{max} > = 12.02 \pm 0.18$ for the six SNe I in Virgo E/S0 galaxies (see Table 5), $M_B^{max} = -19.8$.

E) *Some Additional Distances*

In Table 5 we list data for 17 SNe I in E/S0 galaxies. The six SNe I in Virgo cluster E galaxies have, as just quoted, m_B^0 (max) $= 12.02 \pm 0.18$ (cf. Tammann 1981). This value combined with the SN calibration in Table 4 gives a Virgo cluster modulus of $(m-M)^0 = 31.67 \pm 0.26$.

From the infrared photometry of SNe I (Elias *et al.* 1981) one knows further that the two SNe I in NGC 1316 were 0.4 ± 0.1 mag fainter than the one in the Virgo cluster galaxy NGC 4536. This leads to a distance modulus of the Fornax cluster, of which NGC 1316 is the brightest member, of $(m-M)^0 = 32.07 \pm 0.28$.

Furthermore the infrared photometry of SN 1972e (Kirshner *et al.* 1973) in NGC 5253 shows it to be brighter by 3.9 ± 0.2 mag than the

[2] Note that with no absorption correction for SN 1954c one would have obtained M_B^0 (max) $= -19.55 \pm 0.19$ and $H_0 = 54$, which statistically is the same result.

TABLE 5 - m_B^0 (max) *of 17 SNe I in E/S0 Galaxies.*

SN	Galaxy	v_0	m_B^0 (max)	Remarks
1939a	NGC 4636	853	12.57*	
1919a	NGC 4486	1200	12.37*	Virgo cluster
1939b	NGC 4621	1200	12.15	Virgo cluster
1957b	NGC 4374	1200	12.53	Virgo cluster
1960r	NGC 4382	1200	12.08	Virgo cluster
1961h	NGC 4564	1200	11.35	Virgo cluster
1965i	NGC 4753	1200	11.72*	Virgo cluster
1972j	NGC 7634	3350	14.05	
1970j	NGC 7619	3983	14.58	
1975b	Anon.	5536	15.27:*	Perseus cluster
1961d	Anon.	6890	16.72	Coma cluster
1962a	Anon.	6890	15.79	Coma cluster
1962i	Anon.	6890	17.07*	Coma cluster
1963c	Anon.	6890	15.85	Coma cluster
1963m	Anon.	6890	15.75:*	Coma cluster
1968h	Anon.	12020	16.85*	
1962e	Anon.	14455	16.63*	

* Observations of somewhat lower quality.

two SNe I in NGC 1316. The resulting distance modulus is hence $(m-M)^0 = 28.17 \pm 0.34$. A different route makes use of the *optical* data of the two SNe I (SN 1895b and SN 1972e) in NGC 5253. They had $< m_B^{max} > = 8.37$ (Tammann 1981), with probably little internal absorption. The galactic absorption is $A_B = 0.13$ (cf. RSA). With this value and the calibration in Table 4, one obtains $(m-M)^0 = 27.98 \pm \sim 0.30$. The mean distance modulus is then $(m-M)^0 = 28.06 \pm 0.22$. The distance of NGC 5253 is interesting because it is probably a member of the NGC 5128 group, which contains besides NGC 5128 (Cen A) itself also NGC 5236 (M 83) and others (cf. Kraan-Korteweg and Tammann 1979).

The above distances are compiled in Table 6 and will be used in Sec. 4 together with other data to map the very local velocity field.

TABLE 6 - *Additional Distances from SNe I.*

Galaxy		v_0[1] (km s^{-1})	Aggregate	$< v_0 >$ (km s^{-1})	$(m\text{-}M)^0$	r (Mpc)
NGC	5253	147	NGC 5128 group	232±21[2]	28.06±0.22	4.1
NGC	4536	1646	Virgo cluster	967±50[3]	31.67±0.26	21.7
NGC	1316	1713	Fornax cluster	1486±76[1]	32.07±0.28	25.9

[1] RSA (Sandage and Tammann 1981).
[2] Tammann and Kraan (1978).
[3] Kraan-Korteweg (1982). The value is in perfect agreement with $< v_0 > = 978 \pm 51$ km s^{-1} which is the result of Mould, Aaronson and Huchra (1980), after their published value of 1019 km s^{-1} is reduced to the presently adopted centroid of the Local Group (Yahil, Tammann and Sandage 1977).

3 - THE VALUE OF q_0

A) *The Classical Deceleration*

i. *Principle*: Because we look back in time as we look out in space, we could, in principle, measure the deceleration of the expansion directly by comparing the World Map "now" with the World Map "then". But the problem is complicated because, due to the finite velocity of light, we do not see the World Map, but only the World Picture.

In all homogeneous and isotropic Friedmann universes, the velocity-distance relation of the World Map is strictly linear at all times and has the same rate everywhere at any given cosmic time. But, since the World Picture cuts the World Map at different cosmic times, the observed velocity-distance relation will appear to be non-linear at large distances, even in the absence of deceleration. This pseudo non-linearity must, of course, be subtracted before finding any true deceleration. Suppose now we add mass to the World, causing a true deceleration. The velocity of any given galaxy will now progressively decrease with time. This adds an additional non-linearity to the World Picture which we seek to measure by some means, for example via the Hubble "velocity-distance" diagram. But the problem is quite complicated because we never measure a "distance" but merely redshifts, apparent luminosities, and angular diameters of

galaxies. Hence, the practical problem reduces to assembling the equations of the homogeneous Friedmann models into relations between observables alone to obtain even a first approximation.

These complications are, of course, accounted for in the treatment of tbe equations that relate the metric of space-time to the energy-density and, hence, to relations between m, q_0 and z only. Before 1958 this problem of the World Map versus the World Picture in terms of redshift and deceleration was developed by among others Tolman, Milne, McCrea, Heckman (1942), Robertson (1955), McVittie (1956), Davidson (1959). The solutions there were given by series expansions in the look-back time (i.e. redshift) for all classes of homogeneous cosmological models. Besides the general philosophical difficulty of not being related directly to the physics (but only to the kinematics) of the models, these Taylor expansions were generally useful only for low redshifts.

A great advance was made in 1958 by Mattig's (1958, 1959) codification of the problem by an elegant reduction that permitted all the fundamental connections between theory and observation to be expressed in closed form. It is remarkable that the former series-expansion treatments in terms of R, $\dot{R}$ and $\ddot{R}$ could be replaced in exact closed form, using only the two quantities: $H_0 \equiv \dot{R}_0/R_0$, measured at the present epoch; and $q_0 \equiv -\dot{R}_0/R_0H_0^2$, also valid for the present epoch. The number q_0 is, amongst other things: (a) a measure of the three-dimensional space curvature R^{-1} of the 4 space-time manifold, given by $kc^2/R^2 = H_0^2\,(2\,q_0 - 1)$; (b) the mean mass-density in units of $4\,\pi\,G/3\,H_0^2$; and (c) twice the ratio of gravitational potential energy to the kinetic energy of the expansion. It is, therefore, remarkable that all the equations which connect the observables of magnitude, number counts, angular diameters, and time scale with redshift contain only H_0, q_0 and z, suppressing any concept of "distance".

ii. *Practice*: Unfortunately, this direct approach to measure the deceleration via Mattig's (m, z, q_0) equation, i.e., via the generalized Hubble diagram, has been largely disappointing to date, despite new data on magnitudes of first-ranked cluster galaxies to redshifts of $z \rightarrow 1$ (Gunn and Oke 1975; Kristian, Sandage and Westphal 1978). The difficult problem is still to know the evolutionary corrections, due either to luminosity changes of the standard candles or to mergers. The summary at the last astronomical Vatican Study Week [*Nuclei of Galaxies*, by Spinrad (p. 45) and by others] is essentially still the *status quo*, with, to be sure, heroic recent ef-

forts by Gunn, Faber, Spinrad, Tinsley and their colleagues to determine $d\,(\mathrm{mag})/dt$ due to luminosity evolution in the look-back time. The answer from this approach, using evolutionary corrections, is that q_0 is still determined only within the older well-known wide range of $0 < q_0 < 1$.

The prospects of a definitive answer via this route are still unknown, yet the appeal of this approach to the geometry of space directly, in the spirit of Gauss, Riemann, Schwarzschild and Hubble, remains so great that the route will surely continue to be used. Possibly, SNe I standard candles, observed at redshifts of $z \approx 0.5$ with the Space Telescope, will bring the necessary progress (cf. Tammann 1979). The route to q_0 via other classical tests, as the count-magnitude relation and the diameter-redshift relation (Sandage 1961), poses equally formidable problems due to evolutionary effects.

B) *Other Routes to q_0*

In view of the great difficulties to measure the deceleration directly, it is important that q_0 be obtained by measuring the mean mass density ρ_0, because in a pressure-free Friedmann universe the two quantities are connected by:

$$q_0 = \frac{4\,\pi\,G\,\rho_0}{3\,H^2_0}\,. \qquad (4)$$

There are several routes to measure ρ_0 (see also Fall 1979; Peebles 1979; Saslaw and Aarseth 1981), of which three will be mentioned here.

i. *q_0 from the Mean Luminosity Density.* The mean large-scale luminosity density residing in E and S0 galaxies is $\mathcal{L}^{E}_{B} = 2.2\times10^{7}\,L_\odot\,\mathrm{Mpc}^{-3}$, and that in spiral galaxies is (after correction for internal absorption $\mathcal{L}^{S}_{B} = 8.8\times10^{7}\,L_\odot\,\mathrm{Mpc}^{-3}$ (Yahil, Sandage and Tammann 1980a). The dynamically determined mass-to-light ratio (equally corrected for internal absorption) of pairs and groups of galaxies is $\mathcal{M}/L_B \approx 25$ (Faber and Gallagher 1979). Because pairs and groups are dominated by spiral galaxies, this value may be taken as typical for the mass associated with spirals. Large clusters of galaxies contain mainly E and S0 galaxies. The virial theorem gives for such clusters $\mathcal{M}/L_B \approx 325$ (Faber and Gallagher 1979). Combining in a first approximation the above luminosity densities with the appropriate mass-to-light ratios gives $\rho_0 = 9.4\times10^{9}\,\mathcal{M}_\odot\,\mathrm{Mpc}^{-3} = 6.4\times10^{-31}\,\mathrm{g\,cm}^{-3}$, which with equation (4) leads to $q_0 \approx 0.07$. Although this determination

may easily be off by a factor of 2, it is clear that the matter which binds pairs, groups and clusters of galaxies cannot close the universe. (Note our *extreme* assumption of $\mathcal{M}/L_B = 325$ for E and S0 "field" galaxies; hence their value of q_0 is an *upper* limit).

ii. ρ_0 *from our Virgocentric Motion.* Density fluctuations in the universe must induce peculiar motions. One can therefore determine the excess mass contained in these fluctuations, if one can measure the deviations from a pure Hubble flow; i.e., if the density contrast $\Delta\rho/\rho$ of the fluctuations is known, the mean mass density ρ_0 follows (Sandage, Tammann and Hardy 1972; Silk 1974; Sandage 1975).

From the discussion in Sec. 4 we adopt here our peculiar infall velocity towards the Virgo cluster to be $v_{vc} = 220 \pm 50$ km s^{-1}. This velocity must reflect the gravitational effect of the excess mass within a sphere centered on the Virgo cluster and the Local Group at its periphery. Compared to the large-scale field, the density contrast within that sphere is found to be $\Delta\rho/\rho_0 = 3.0 \pm 0.2$ (Yahil, Sandage and Tammann 1980a). From counts of galaxies with $12 < m < 13$ mag in the North and South Polar Caps it follows that $\Delta\rho/\rho_0 = 2.4 \pm 0.2$ (Sandage, Tammann and Hardy 1972), which is independent of any redshift information. From a deep redshift survey Davis *et al.* (1980) determined, after adjustement to an infall velocity of 220 km s^{-1}, $\Delta\rho/\rho_0 = 2.8 \pm 0.3$. We adopt the latter value as the mean of the three determinations. With $v_{vc} = 220 \pm 50$ km s^{-1} and $\Delta\rho/\rho_0 = 2.8 \pm 0.3$ one finds (cf. Yahil, Sandage and Tammann 1980b; Yahil 1980) $q_0 = 0.05 \pm 0.04$. This method measures all baryonic and non-baryonic matter which is clustered like visible galaxies.

iii. *Additional Constraints on* q_0. The yield of the light isotopes ^{2}D, ^{3}He, ^{4}He and ^{7}Li during the primordial fireball nucleosynthesis at a temperature of $T \approx 10^9$ K depends on the matter density ρ at that epoch (e.g. Wagoner 1973). The density which gives the proper yield to agree with the observed abundances can be scaled down to the present cosmic background radiation temperature ($T \approx 3$ K) to give a determination of the present density ρ_0 contained in baryons. In particular the yield of ^{2}D is quite sensitive to the baryon density. The observed abundance of ^{2}D, with a moderate allowance for destruction in stars, gives a relative abundance of $X_D > 3 \times 10^{-5}$ which requires (Yang *et al.* 1979) $0.02 < q_0 < 0.04$.

This determination of q_0, which is further supported by ^{3}He, can be invalidated by postulating a non-zero lepton number (Yahil and Beaudet 1976); however, in this case quite specific requirements must be met (David

and Reeves 1980; Audouze 1982). Moreover the universally observed ^{4}He abundance of $Y \geqslant 0.24$, which is relatively insensitive to the density, becomes then a result of chance, whereas in canonical Big Bang models it is a necessity (Olson and Silk 1978).

It is noteworthy that the baryonic density from nucleosynthesis arguments agrees very well with the total matter density clumped like galaxies. Therefore, if there is to be mass sufficient to close the universe, it must be in the form of invisible, non baryonic matter which does not follow the clustering of galaxies.

A possibility would be massive neutrinos. The number of one type of relic neutrinos plus antineutrinos is about 100 cm^{-3} (cf. Sciama 1982). Hence, to account for $\sim 90\%$ of the closure density of 4.7×10^{-30} g cm^{-3}, a neutrino must bave a mass of $\sim 4 \times 10^{-32}$ g ≈ 20 eV (assuming that all the neutrino mass is in one type of neutrino). With this mass, however, neutrinos would tend to cluster in galaxies and even more so in clusters of galaxies (Gunn 1982). It is then not clear why the above dynamical determinations of q_0 give only a fraction of the closure density and how, therefore, most neutrinos can avoid clustering like galaxies. One explanation would be if the closing mass were due to several types of neutrino, each of which had sufficiently low mass to escape the concentrations of baryonic mass.

C) *Time Scale Determination*

The time scale test, made by comparing H_0^{-1} (determined directly via local distances and expansion velocities) with globular cluster ages, increased by the galaxy formation time, has become a relatively powerful way to determine q_0 for Friedmann universes where $\Lambda = 0$.

Let the age of the universe be t_0. The equations for all Friedmann models give an explicit closed form between t_0, H_0 and q_0 alone as

$$H_0 t_0 = f(q_0) \ , \tag{5}$$

where $1 < f(q_0) < 0.57$ for q_0 in the interval $0 < q_0 < 1$ (Sandage 1961, Table 8 and equations 61, 62 and 65).

Let the age of the globular clusters be T (0) and the time from the Friedmann "beginning" to the beginning of the Galaxy (and hence to the formation of the clusters) be ΔT. Clearly

$$t_0 = \mathrm{T}(0) + \Delta \mathrm{T} = H_0^{-1} f(q_0) \ . \tag{6}$$

We identify the formation time of galaxies with the upper redshift limit of quasar redshifts for the obvious reason that quasars, being events in the nuclei of galaxies, signal a lower limit to the time of their birthday by tbe largest redshift they are observed to have. Let tbis maximum redshift be z_m. Hence,

$$R_{now}/R_{QSS} = 1 + z_m \tag{7}$$

gives the ratio of the scale factor now to tbe scale at the time when quasars began. Once known, z_m gives this ratio. Then, for any q_0, the time ΔT via the relevant $R = g(t, q_0)$ Friedmann curve is found. The geometry is shown in Figure 21.

Consider two characteristic cases of $q_0 = 0$ and $q_0 = 1/2$, using $T(0) = (17 \pm 2) \times 10^9$ years (Sandage 1982) and $H_0 = 50$ km s^{-1} Mpc^{-1}, justified earlier from the velocity-distance calibration.

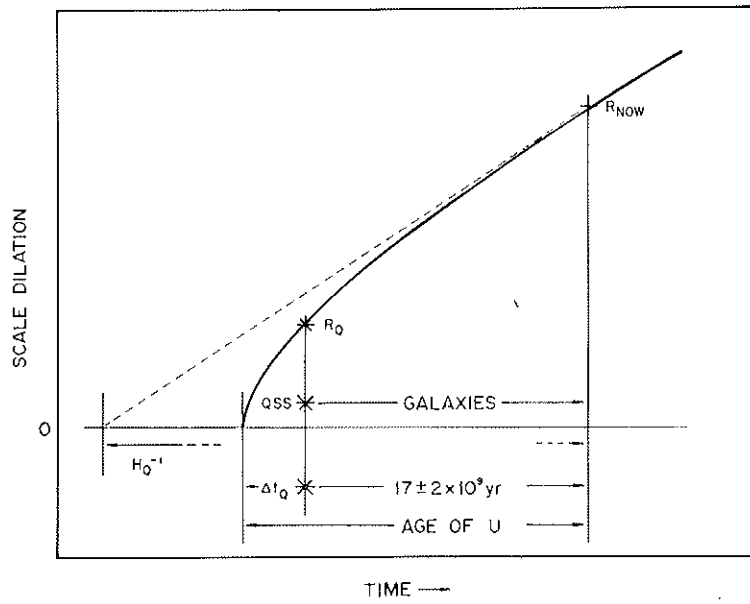


FIG. 21. The time scale test for $t(H_0, q_0)$ for Friedmann models. The Friedmann $R(t, q_0)$ curve from the exact solution is the heavy solid line. The tangent to it at the present epoch (labelled R_{NOW}) gives the time H_0^{-1} at its intersection at the $R = 0$ origin. To the age of the galaxies at $(17 \pm 2) \times 10^9$ years must be added the formation age ΔT_Q (determined from quasars), from which the equations in the text follow.

For $q_0 = 0$ and $z_m = 4$ (cf. Osmer 1982), for the maximum quasar redshift, eq. 3 gives $R_N/R_Q = 5$, hence $\Delta T = 0.2\ H_0^{-1}$. For the $q_0 = 1/2$ case, $R(t) \sim t^{2/3}$, hence, at $R_N/R_Q = 5$, $\Delta T = 0.09\ t_0$, or $\Delta T = 0.06\ H_0^{-1}$.

From equation (6)

$$t_0 H_0^{-1} = f(q_0) = H_0 [\ T(0) + \Delta T\ ]\ , \tag{8}$$

which, because we know that q_0 lies between 0 and 1/2 by the argument of the last section, $\Delta T \cong 0.12\ H_0^{-1}$ as an average, useful in an iteration procedure.

Therefore,

$$f(q_0) = H_0 [\ (17 \pm 2) \times 10^9\ ] + 0.12\ . \tag{9}$$

Changing $H_0 = 50 \pm 5$ km s^{-1} Mpc^{-1}, known directly from Sec. 2, into time units of $H_0^{-1} = (19.5 \pm 2) \times 10^9$ years gives

$$f(q_0) = 0.99 \pm 0.15\ , \tag{10}$$

where the 1 σ lower limit would therefore be $f(q_0) = 0.84$. Hence, using Table 8 of the quoted reference gives

$$0 < q_0 < 0.1 \tag{11}$$

at the 1 σ error limit [we ignore a negative q_0 implied by the 1 σ upper limit of $f(q_0) = 1.14$ from equation (10), since we do not consider any model where the cosmological constant is anything but zero].

The value in equation (11) agrees well with $q_0 = 0.05 \pm 0.02$, which was found consistently from different methods in Sec. 3 B.

4 - The very local velocity field

A) *The Present Data*

Data on the distances and velocities of very nearby galaxies give information concerning: (1) the random motion of "field" galaxies; (2) any, presumably Virgo-cluster-induced, streaming or shear velocities of appropriate scale length; and (3) the very local expansion rate and hence the gravitational effect of the Local Group.

The possibilities of such an investigation are explored in the present pilot program. The available set of nearby galaxies with reliably and homogeneously determined distances is very restricted and, hence, no high-quality results can yet be expected. The following application, however, establishes already some restrictions on the kinematic parameters of the immediate extragalactic neighborhood.

In Table 7 we have compiled the galaxies with distances known from red supergiants (Table 3) and from SNe (Table 6). The galaxies NGC 300 and NGC 55 are added on the grounds of recent work on their brightest stars (Graham 1981). The brightest stars in NGC 7793 are fainter by 1 mag than those in the two last-mentioned galaxies (RSA; Sandage 1982), from which results its distance in Table 7.

The recession velocities v_0 in column 4 of Table 7 are taken from

TABLE 7 - *Local Galaxies with Known Distances.*

Galaxy	r (Mpc)	θ (Virgo)	v_0 (km s^{-1})	v_0 (corr) $v_{vc}=220$ (km s^{-1})	(excl. LG) $v_{vc}=440$ (km s^{-1})	v_0 (corr) $v_{vc}=220$ (km s^{-1})	(incl. LG) $v_{vc}=440$ (km s^{-1})
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
W-L-M	1.3	169°	−5	−17	−29	106	52
Sextans A	1.3	48	117	114	112	237	193
NGC 300	1.4	156	128	119	110	225	180
NGC 55	1.4	154	115	107	98	213	168
Sextans B	2.1	45	132	128	123	175	154
NGC 7793	2.2	158	239	224	209	267	237
Leo A	3.0	31	−10	−14	−18	9	−3
NGC 2403	3.3	63	299	329	361	348	374
NGC 5253	4.1	47	146	156	167	168	175
Pegasus	4.3	145	61	41	21	52	28
IC 4182	4.4	24	344	315	287	316	294
NGC 4395	5.8	20	304	253	203	259	207
NGC 4214	6.4	23	290	243	195	248	198
M 101	6.9	41	372	399	426	403	429
Virgo Cl.	21.7	—	960	1187	1407	1187	1407
Fornax Cl.	25.9	131	1486	1435	1384	1435	1384

recent compilations (RSA; Kraan-Korteweg and Tammann 1979). The velocities are corrected in columns 5 and 6 under the assumption of a consistent Virgocentric flow model (Yahil, Sandage and Tammann 1980b, 1980c), which is characterized by a Virgocentric infall velocity of the Local Group of $v_{vc} = 220$ km s^{-1} and $v_{vc} = 440$ km s^{-1}, respectively, and a density profile of the Virgo complex of $\rho \propto r^{-2}$ (Tammann, Sandage and Yahil 1979). A convenient linear approximation for the expected radial velocity of field galaxies as a function of v_{vc} is given by Schechter (1980, eq. 2). This linear approximation is satisfactory in the present case, because the sample galaxies are sufficiently distant from the Virgo cluster. The angle θ between the sample galaxies and the Virgo cluster are shown in column 3.

An additional correction to the observed velocities concerns the gravitational pull of the Local Group. Its effect can be estimated by adopting a mass ratio between the Local Group, taken as a point mass, and the mass within a sphere centered on the Virgo cluster with the Local Group on its periphery. The total mass within this sphere is a function of v_{vc} and of the density contrast (Yahil, Sandage and Tammann 1980b, eq. 32). The density contrast $\Delta\rho/\rho_0$ is itself a weak function of v_{vc} (cf. Yahil 1980). Taking $v_{vc} = 220$ km s^{-1} and $\Delta\rho/\rho_0 = 3.0$ and $v_{vc} = 440$ km s^{-1} and $\Delta\rho/\rho_0 = 2.75$, as value pairs (cf. Sec. 3 B), one obtains a total mass of 1.5×10^{15} $\mathcal{M}_\odot$ and 4.5×10^{15} $\mathcal{M}_\odot$, respectively. Adopting a Local Group mass of $\sim 3 \times 10^{12}$ $\mathcal{M}_\odot$ (cf. Lynden-Bell 1981) leads us to the conclusion that the gravitational force per unit mass exerted by the Local Group at the distance of the Virgo cluster is 500 or 1500 times smaller than the Virgo cluster effect, i.e. $v_{LG} = 0.44$ and 0.29 km s^{-1} at the distance of Virgo ($r = 21.7$ Mpc). This results in a sizeable infall velocity of 207 km s^{-1} and 137 km s^{-1}, respectively, at a distance of 1 Mpc from the Local Group centroid. The effect of the Virgo cluster and of the Local Group on the observed velocities are illustrated in Figure 22.

It should be noted that the masses adopted above imply different mass-to-light ratios for the large sphere centered on the Virgo cluster and the Local Group. For the former one has $\mathcal{M}/L_B \approx 100$ (with $v_{vc} = 220$ km s^{-1}) or $\mathcal{M}/L_B \approx 300$ (with $v_{vc} = 440$ km s^{-1}), whereas the adopted Local Group mass corresponds to $\mathcal{M}/L_B \approx 25$. Without considerably increasing the mass of the Local Group the mass-to-light ratios cannot be the same, unless our Virgocentric velocity were significantly smaller than 220 km s^{-1}.

Solutions for the distance and velocity data in Table 7 are shown in

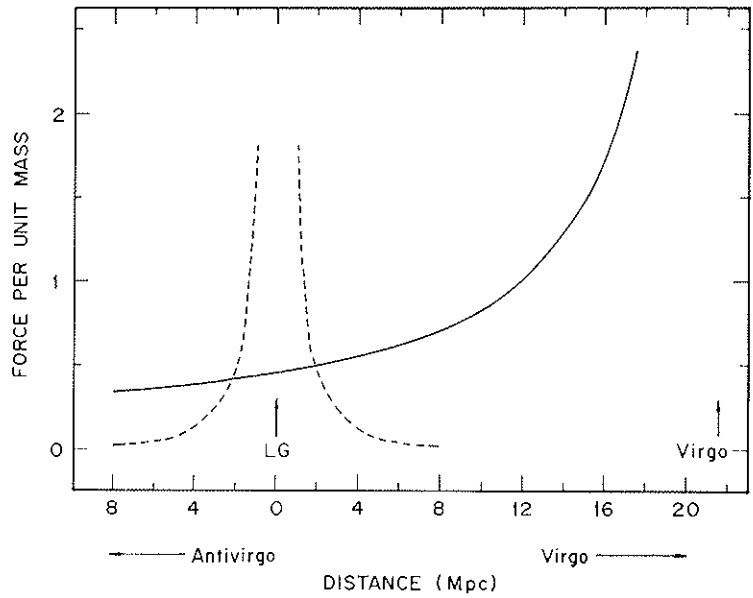


FIG. 22. The gravitational force in arbitrary units exerted by the Virgo cluster (full line) and the Local Group (dashed line) as a function of the distance toward the Virgo cluster and away from it.

Table 8; Figure 23 is a graphic presentation of these data. In order to guard against the effect of random motions of field galaxies on the value of H_0, the corresponding solutions were determined by weighting each galaxy or cluster with its distance r. As it turns out the value of H_0 is quite insensitive to the weighting procedure. Moreover, as can be seen in Table 8,

TABLE 8 - *Solutions for the Local Velocity Field.*

	excl. LG		incl. LG	
	H_0	$< \Delta v^2 >^{1/2}$	H_0	$< \Delta v^2 >^{1/2}$
$v_{vc} = 0$	52	101	—	—
$v_{vc} = 220$	54	91	55	113
$v_{vc} = 440$	56	119	57	115

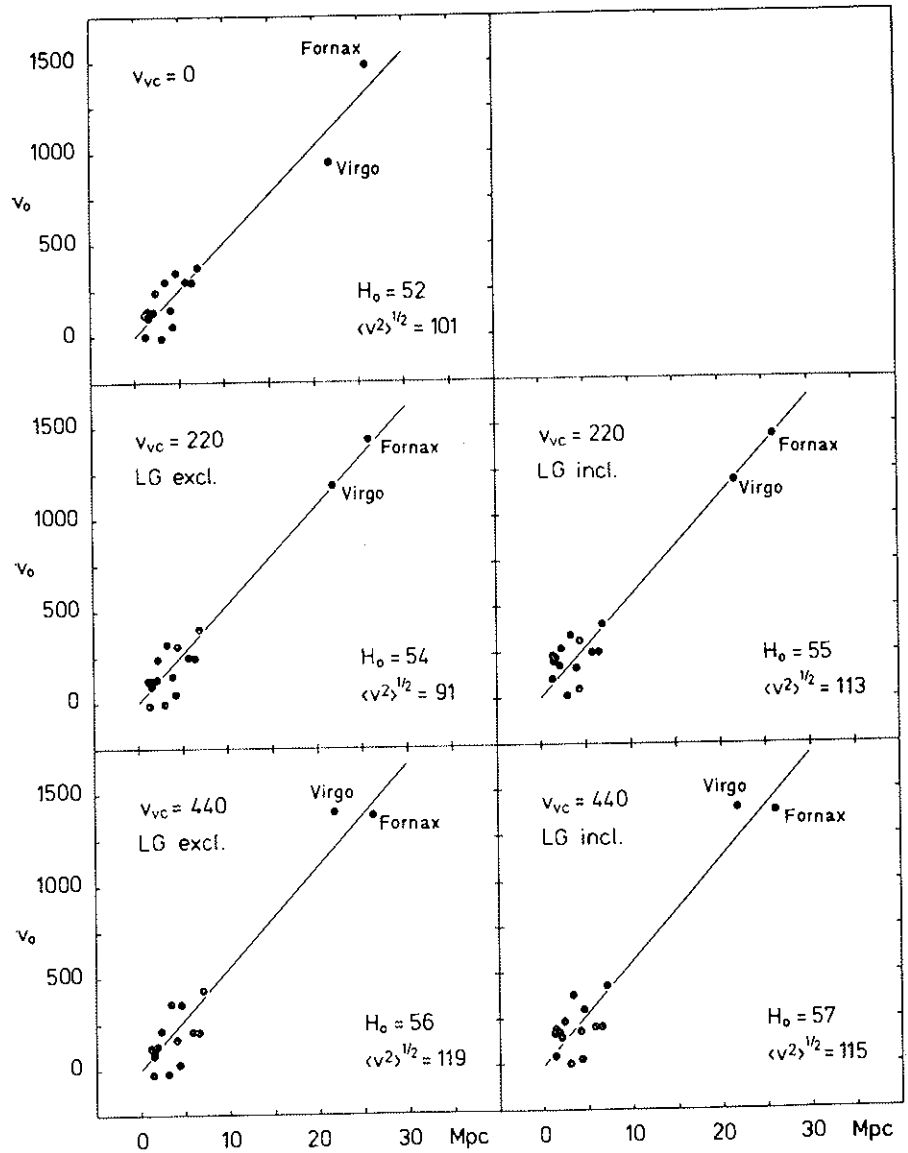


FIG. 23. The very local expansion field from data in Table 7, assuming different values for our Virgocentric velocity v_{vc} and excluding and including the Local Group effect.

H_0 only changes slightly with the adopted value of v_{vc}. All solutions for H_0 (local) agree with H_0 (global) $= 50 \pm 7$ from Sec. 2 D within the statistics.

The random motions $\sigma_v = < \Delta v^2 >^{1/2}$ of the test particles become a minimum for $v_{vc} = 220$ km s^{-1}. This would favor this value of our Virgocentric velocity, but there is no a priori reason why the random motions should take this minimum value. However, independent evidence, presented in Sec. 4 B, does favor a low value of σ_v. In a formal sense it is, of course, unrealistic to attribute tbe full scatter in Figure 23 to random velocities; however, the results are hardly affected by assuming random distance errors even as high as 30%.

The Local Group effect tends to increase the value of H_0 by an insignificant amount and increases somewhat the random velocities. This suggests that the true Local Group effect is smaller, i.e. that we have overestimated the mass ratio of the Local Group to the Virgo complex.

Inspection of Figure 23 shows that the relative position of the Virgo cluster and of the Fornax cluster is optimized by an adopted value of $v_{vc} = 220$ km s^{-1}. From the present data this is the strongest reason why this relatively low Virgocentric infall velocity should he preferred. Independent data supporting this solution will be given in the next Section (Sec. 4 B). The resulting velocity contours in the Virgo complex as seen from the centroid of the Local group is shown in Fig. 24.

It is obvious that future ohservations, e.g. of the red supergiants in the distance range $7 < r < 25$ Mpc with Space Telescope and SN I distances at $v_0 \approx 1000$ km s^{-1} with good sky coverage, will greatly strengthen the results of this Section.

B) *Comparison with Independent Data*

In the previous Section a solution was preferred with a random peculiar radial velocity of very nearby field galaxies of $\sigma_v = 90$ to 100 km s^{-1} and with a Virgocentric motion of the Local Group of $v_{vc} = 220$ km s^{-1}. . These parameters lead to an underlying expansion rate of H_0 (local) ≈ 55, which is in satisfactory agreement with the global value of H_0 (global) $= 50 \pm 7$. Further support for the adopted parameters is discussed in the following paragraphs.

i. *The random peculiar velocity σ_v of field galaxies*

The size of $\sigma_v = 90$ to 100 km s^{-1} for the one-dimensional mean random velocity of nearby field galaxies came as a surprise, after values

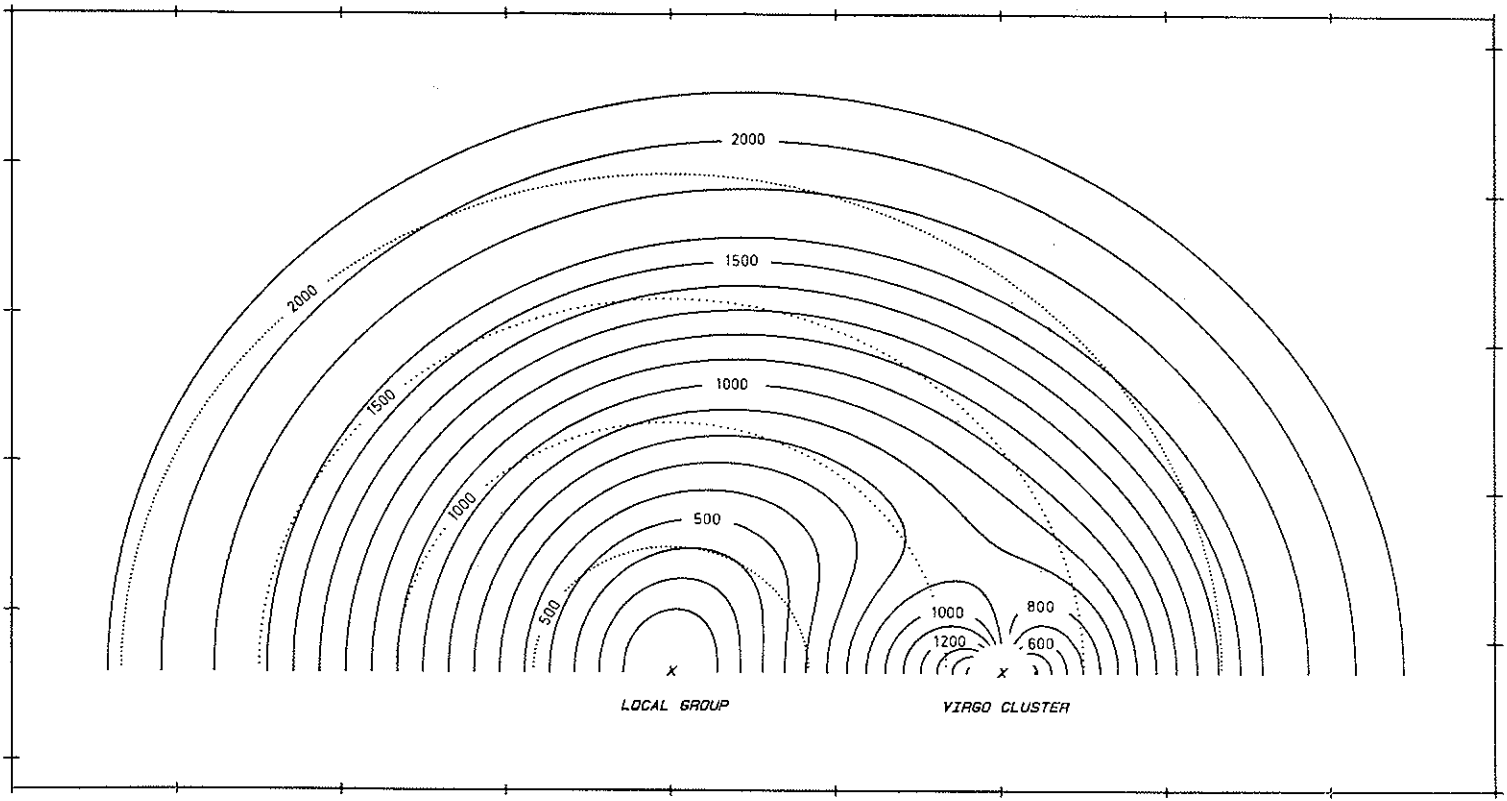


FIG. 24. Velocity contours as seen from the centroid of the Local Group in a plane containing the Local Group and the Virgo cluster, assumed at v_0 (obs.) = 967 km s^{-1}. Lines of equal observed recession velocities are shown for a local Virgocentric velocity of v_{vc} = 220 km s^{-1} (full line) and for the unperturbed Hubble flow (dashed line). (With kind permission of U. Kaeser, Basel).

had earlier been suggested as low as $\sigma_v = 50$ or even 25 km s^{-1} (ST IV; Tammann, Sandage and Yahil 1980). The main reason for the upward revision is the fact that it was previously believed that no field galaxy had an observed recession velocity of less than $v_0 < 75$ km s^{-1}. Now, where it is clear that the galaxies W-L-M and Leo A with negative v_0-values do not belong to the Local Group, and where the low-velocity ($v_0 = 61$ km s^{-1}) Pegasus galaxy lies at a considerable distance of 4.3 Mpc, the low-velocity cutoff of field galaxies cannot be maintained. This leads immediately to an increase of σ_v.

An analysis of the velocity differences of neighboring pairs of Shapley-Ames galaxies gives a mean one-dimensional random velocity of $\sigma_v = 70 \pm 10$ km s^{-1} (Rivolo and Yahil 1981). A sample of nearby spirals with 21 cm observations leads to $\sigma_v = 90 \pm 35$ km s^{-1} (Aaronson *et al.* 1981a). This solution is weighted in favor of very-nearby galaxies and should, therefore, be directly comparable to the present result. It seems therefore that a value of $\sigma_v \approx 90$ km s^{-1} is fairly well supported.

It should be noted, however, that the significance of the present determination of σ_v is not obvious. It applies to a region close to the Local Group, i.e. to the outer regions of the Virgo complex. It could well be that σ_v depends on the mean mass density within the volume studied and that different values apply for different volumes. In addition the galaxies in Table 7 are concentrated toward the Supergalactic plane; therefore, σ_v measures essentially the mean radial component of the random motion within this plane. The corresponding velocities perpendicular to this plane could be quite different.

ii. *Our Virgocentric Velocity* v_{vc}

A simple way to determine our infall motion v_{vc} toward the Virgo cluster is to compare the observed mean Virgo cluster velocity with a predicted value of this velocity. Such a prediction can be made on the basis of the Coma cluster if the distance ratio Coma: Virgo is known (cf. Tammann, Sandage and Yahil 1980). Because the largest peculiar motion the Local Group partakes of, $v_{pec} \approx 600$ km s^{-1}, is given by the dipole anisotropy of the cosmic microwave background, it is reasonable to assume that the same value is also an upper limit for the Coma cluster (cf. Sec. 1 B). This means that the mean velocity of Coma of $< v_0 > = 6890 \pm 68$ (Rood *et al.* 1972) contains at most a 10% peculiar velocity component. The relative distance Coma: Virgo can be estimated from three different standard candles:

	Δ (m-M)	Source
SNe I	4.22±0.33	Tammann 1981
Brightest cluster member	3.37±0.40	Sandage and Hardy 1973
10 brightest cluster members	4.22±0.28	Weedman 1976
	4.18±0.19	

The mean distance modulus difference of 4.18 ± 0.19 translates into a linear distance ratio of 6.85 ± 0.60, and hence v_{Virgo} (expected) = (6890 ± 68)/(6.85 ± 0.60) = 1006 ± 90 km s^{-1}. With v_{Virgo} (observed) = 967 ± 50 km s^{-1} it follows that v_{vc} = 39 ± 103 km s^{-1}.

Independently the Virgo cluster velocity can be predicted from its distance, r = 21.7 ± 2.8 Mpc, and from H_0 (global) = 50 ± 7. With these values one obtains v_{Virgo} (expected) = 1085 ± 207 km s^{-1}, and hence v_{vc} = 118 ± 213 km s^{-1}.

Without reference to any observed or predicted Virgo cluster velocity the value of v_{vc} has been determined relative to the rest frame of the field galaxies in the RSA, i.e. relative to galaxies with $m_B < 13$ mag, excluding the regions of the Virgo and Fornax clusters. From a consistent Virgocentric flow model it then follows (Yahil, Sandage and Tammann 1980b; Yahil 1980) that v_{vc} = 220 ± 75 km s^{-1}.

A similar, but refined solution for spirals with available 21 cm-line widths as relative distance indicators was obtained by Aaronson *et al.* (1981a). Their result is v_{vc} = 250 ± 64 km s^{-1}. The same authors found in addition a significant peculiar velocity of the Local Group with a Virgocentric component of w_{vc} = 81 ± 50 km s^{-1}. This value, however, depends on the adopted correction of the solar motion relative to the centroid of the Local Group. If the correction of Yahil, Tammann and Sandage (1977) is used, as is done throughout this paper, the random Virgocentric component is reduced to the insignificant amount of w_{vc} = 42 ± 50 km s^{-1}. That the total peculiar motion of the Local Group is indeed only of the order of 40 km s^{-1} is independently supported by the near agreement between the solar motion with respect to the centroid of the Local Group (Lynden-Bell and Lin 1977; Yahil, Tammann and Sandage 1977) and with respect to the nearest field galaxies (Lynden-Bell 1982).

Several authors have in addition proposed a rotational velocity component of the Local Group about the center of the Virgo cluster (Aaronson

et al. 1981a, and references therein). This result is difficult to accept because any rotational component in external clusters and cluster halos has so far defied detection.

A formal mean of the above four determinations of our Virgocentric motion yields $v_{vc} = 198 \pm 43$ km s^{-1}. Giving additional weight to the two last-mentioned determinations, one finds $v_{vc} = 220 \pm \sim 50$ km s^{-1}, which is the value used in Sec. 3 B, and which is in agreement with the solution preferred in Sec. 4 A.

Other determinations of v_{vc}, which rely on the assumption that the global parameters of galaxies in different clusters (like the 21 cm-line width of spirals and the dispersion velocities of ellipticals) are directly comparable, and which generally lead to higher values of v_{vc}, are omitted here (cf. Appendix).

5 - Prospects for the future

Recent promising developments related to various distance indicators show that we are not now blocked in the immediate next steps to improve the measurement of H_0, q_0, and any local velocity perturbation.

A) *Prospects for Improving* H_0

An ideal situation would be if type I supernovae proved to be as good standard candles as suggested in Sec. 2. If so, relative distances with high precision could be found to any galaxy that is a parent of SNe I.

The present optical evidence suggests $\sigma (M_B) < 0.4$ mag as the intrinsic dispersion of absolute blue magnitude. But, as discussed in Sec. 2, the true dispersion judging from infrared magnitudes may be even smaller (Elias *et al.* 1981).

If $\sigma (M_{IR})$ is less than 0.2 mag (cf. Elias *et al.* 1981), then relative distances to individual galaxies could be determined at least to $\pm$ 10%. Absolute distances to the few of the nearest galaxies with SNe I must then be found, via red supergiants or even Cepheids (as soon as type I supernovae are seen again in the Local Group).

The calibration of M_V (3) of the brightest red supergiants can be greatly improved from ground-based observations of Cepheids in NGC 3109, IC 5152, NGC 300, NGC 55, and members of the M 81/NGC 2403 group. We can expect eventually to increase the calibrating sample dis-

cussed in Sec. 2 (Figure 5) by at least a factor of 3. This, together with more SNe I in nearby resolved galaxies, should improve the direct calibration of M (max)$_{SNe\ I}$.

The work then consists of three parts, each of which is currently feasible: (1) obtain m_V (3) for red supergiants in more nearby galaxies via Cepheids; (2) discover SNe I in these and other nearby galaxies where red supergiants can be identified; (3) obtain light curves near maximum for type I supernovae in galaxies (preferably E types, but also in spirals if infrared photometry is done) with $v_0 > 3000$ km s^{-1}, so as to avoid any effects of local velocity anisotropies.

The time-scale route to H_0 can also be improved by further developments in the stellar evolution age-dating of globular clusters, but unless there is fundamental change, either in the theory or in the RR Lyrae star absolute magnitudes, this route presently seems relatively secure already to us.

B) *Prospects for Improving q_0*

i. *The Look-Back Approach.* Improvements in our knowledge of the stellar content of E galaxies and how they change with time are expected when data are available on the energy distribution of distant galaxies, measured with spectral resolutions of ~ 10 Å. Of particular promise is the work of Bruzual (1981) where he shows that the $\lambda_0 = 4000$ Å blanketing break changes systematically in size with increasing redshift. This and other such indices (Gunn, Stryker and Tinsley 1981) may eventually define unambiguously the evolutionary correction to magnitudes due to changing stellar content alone, but the problem of cannibalism of first-ranked cluster members (Hausman and Ostriker 1978, Schweizer 1981) remains.

The accuracy with which one must know these corrections is very high. Series expansion of the (m, z, q_0) equation for the Hubble diagram is $m_{bol} = 5 \log z + 1.086\ (1 - q_0)\ z + 0\ (z^2) + \text{const}$. Suppose one can expect many routine observations of (m, z) pairs to be available at $z = 0.5$. Then the accuracy Δm_{bol} with which one must know the evolutionary correction, so as to give an error in q_0 of less than Δq_0, is $\Delta m_{bol} = 1.09\ \Delta q_0\ z$. Hence, to avoid errors in q_0 larger than ± 0.1, requires Δm_{bol} to be known to better than ± 0.06 mag, clearly a formidable task for any theory of the correction.

ii. *Local Tests.* The focus in the next years will undoubtedly remain at finding the ratio of gravitational potential energy to kinetic energy of

the expansion by seeking the velocity perturbations. The prospects seem excellent to improve this method, both for the systematic effect of the Virgo complex (Schechter 1980; Tonry and Davis 1981b), and for a possible random motion test (Saslaw and Aarseth 1981).

One can anticipate that rather precise distances can well be obtained to certain highly resolved galaxies in particular directions, using brightest red supergiants. If so, Figure 23 for the local Hubble diagram should be filled to $\sim$ 7 Mpc by a factor $\sim$ 3 more galaxies with observations from the ground, and perhaps to $\sim$ 25 Mpc using Space Telescope.

Hence, the velocity-distance relation can be obtained for galaxies in the annular zones from Virgo where the systematic velocity effect of infall is near zero. Consequently, direct comparison with the region of maximum effect centered at $\theta = 0°$ and $180°$ should permit one to determine the infall accurately. With this, and with the known value of the density contrast, q_0 can again emerge.

A survey is now in progress at Las Campanas to catalog highly resolved spirals and Sm type galaxies with low surface brightness disks, as candidates for measurement of their brightest red supergiants, both from the ground and from Space Telescope. We anticipate that $\sim$ 50 galaxies, in addition to those already known, can be used for this work.

The goals are: (1) identification of the brightest red supergiants. The ground-based limit is $B \approx 23$, and because the candidate stars must be seen to be very red, $V \approx 21$ only. With $B - V > 2.0$ for $M_V(3) \approx -8$ the ground-based limit is $m - M \approx 29$, or $r = 7$ Mpc; (2) photometry and redshifts (or proper motions) of the candidate stars, with special attention paid to membership. The program is feasible, but requires the best seeing conditions known from the ground.

6 - Appendix: Remarks on various distance scales

Why is it that other extragalactic distance scales have been published up to recently which lead to a pronounced increase of the expansion rate with distance and to values of the global Hubble constant as high as $H_0 = 100$? It is, of course, not possible to give a single clear-cut answer to this question, but certain definite requirements can be set down, against which different distance indicators can be tested. An attempt to do this is presented in the following.

Because purely physical distance determinations, like expansion

parallaxes of SNe, cannot yet carry the full burden of the extragalactic distance scale, the scale still depends on the use of standard candles and standard rods (briefly called standards in the following). To be useful as distance indicators the standards must fulfill at least four principal conditions: (1) they must be demonstrated to be valid standards independent of distance and environment; (2) their variation in intrinsic luminosity or size must have a measurable effect on their apparent properties; (3) their absolute magnitude or linear size must be calibrated, either through direct methods, or through tying them to other distance indicators with known zero-point; (4) the scatter about their mean luminosity or about their mean size must either be small, or else must be well known. This is because such scatter causes *systematic* distance underestimates as one moves to larger distances (Malmquist bias).

Without attempting to be complete, some of the distance indicators used are tested here against these requirements:

(1) The mean absolute magnitude of globular clusters has been used to derive the distance of the Virgo cluster (Hanes 1981; de Vaucouleurs 1977). This is based on the assumption that the luminosity function of globular clusters is universal, i.e. that its Gaussian form, its magnitude M and its dispersion σ (M) are stable. The luminosity function is actually only fairly well observed in our Galaxy, whereas in other galaxies one knows merely its rising branch on the bright side. The basic assumption cannot therefore be tested by observations, and it is in fact possible that it is invalid because the properties of the globular cluster system depend probably on the dynamical conditions during formation and, therefore, on the galaxy type.

If instead of the mean magnitude of the globular clusters, we use the magnitude at which the brightest objects occur, a correlation between the corresponding absolute magnitude and the galaxy luminosity can be established and calibrated using several galaxies with known distances (cf. Hodge 1974). This route fulfills condition (1) and leads consistently to values of $H_0 \approx 50$ (Sandage 1968; de Vaucouleurs 1970; ST V).

A more subtle point concerns distance indicators which can be suspected to depend on the environment. To this class may belong overall parameters of field galaxies and of galaxies in different clusters. There is a pronounced correlation between morphological type and the surrounding galaxy density (Dressler 1980). The (U-V)-luminosity relation of elliptical galaxies (Visvanathan and Sandage 1977) or/and their infrared color-luminosity relation are not the same in different clusters (Aaronson, Persson

and Frogel 1981). Also the 21 cm line width-surface brightness relation seems to be different for spirals in different clusters (Kraan-Korteweg 1982). Any density-dependent dynamical evolution of cluster galaxies could lead to systematic differences from one cluster to the other (Wielen 1979). In that case it seems questionable whether, for instance, the 21 cm line width method is directly applicable to cluster galaxies. This method does not only depend on the morphological type (Rubin, Burstein and Thonnard 1980; Aaronson *et al.* 1981b) and on the surface brightness (Aaronson *et al.* 1981b), but also on the assumption that the following three quantities are constant for the spirals under consideration: the mass-to-light ratio, the ratio of turn-over radius to optical radius, and the ratio of the mass inside the turn-over radius to the total mass (Sandage and Dressler 1981). In view of these problems, the 21 cm distances of cluster galaxies (Tully and Fisher 1977; Aaronson *et al.* 1980, 1981a; Mould, Aaronson and Huchra 1980) may not be secure. For this reason they have not been used in this paper. Similarly the dispersion velocity-luminosity relation for elliptical galaxies in clusters (Tonry and Davis 1981a) yields an improbably high Virgocentric velocity, suggesting that the result is biased hy properties which change with the cluster;

(2) Some standards are disappointingly insensitive to distance effects. For instance the linear size of the largest H II regions in spiral galaxies is coupled with the equally distant-dependent absolute magnitude of the parent galaxy (ST IV; Kennicutt 1979). The situation is improved if, instead of size, the H_α-flux of H II regions (Kennicutt 1981) or their velocity dispersion (Terlevich and Melnick 1981) is used. A similar difficulty exists for the brightest blue stars as discussed in Sec. 2 A. The shape of the brigbt end of the luminosity function of globular clusters is so vulnerable to photometric errors that it becomes a poor distance indicator (Hanes 1979; cf. Tammann, Sandage and Yabil 1979);

(3) Sometimes the difference between absolute and relative distance indicators is not fully acknowledged. The difference is that the former possess a known zero point. The difficulty of setting a distance indicator on an absolute footing is that its nearest representative may already lie outside the narrow range of secure distances. This means that the zero point of essentially all distance scales depends on the distances to the very local galaxies, which are widely agreed upon at the $\sim$ 15% level (Sec. 2 A). One anticipates any given result if, for instance, Virgo cluster galaxies are used for the definition of the zero point. With a given Virgo distance

the value of H_0 (global) is essentially determined either by correcting the observed Virgo velocity for our Virgocentric motion, v_{vc}, or by carrying the distance scale further out with brightest cluster galaxies and SNe I;

(4) Several distance scales depend on a relation of the form

$$M_{Gal} = a\,\Phi + \text{const} \tag{1a}$$

or

$$M_{Gal} = a\,\Phi + b\,\Psi + \ldots + \text{const}\ , \tag{2a}$$

where Φ, Ψ, . . . are observable galaxy parameters which correlate with the luminosity of a galaxy (e.g. morphological type, luminosity class, diameter, color, 21 cm line width, etc.) and a, b, . . . are coefficients which are determined either from members of a cluster or from local calibrators to provide also the constant term, i.e. the absolute zero point. Unfortunately, no dispersionless luminosity indicator is known, and generally the dispersion in absolute magnitude is considerable. If then the method is applied to galaxies which are drawn from a catalog which is limited by apparent magnitude (and all galaxy catalogs are more or less magnitude-limited) a systematic effect is introduced which leads to an underestimate of large distances and, hence, to an apparent increase of H_0 with distance. All suggestions that H_0 increases with distance are, therefore, suspect and probably due to selection bias. Under idealized conditions, where one has Gaussian luminosity scatter, complete samples (or unbiased subsamples thereof) and constant space density, the systematic bias can be analytically corrected for (Malmquist 1920). The correction is, in fact, given by $\Delta M = 1.38\,\sigma_M^2$. However, not only does the determination of σ_M pose a problem (because the unbiased luminosity or size distribution of the standards is generally not known), but also the assumption of constant space density which is untenable within the Virgo complex. In that case the bias can be compensated only by including a velocity (or distance) dependent term into equations 1a and 2a; in linear approximation for instance:

$$(M_{Gal} - k \log v_0) = a\,\Phi + b\,\Psi + \ldots + \text{const}\ . \tag{3a}$$

A self-consistent model calculation shows that E/S0 galaxies of the Shapley-Ames catalog, taken as standard candles, require $k = 3.33$ (Sandage, Tammann and Yahil 1979). Neglecting this term would give an erroneous increase of the Hubble constant from say $H_0 = 50$ at the Virgo cluster to

$H_0 = 150$ at a distance corresponding to $v_0 = 5000$ km s^{-1}! This example is extreme because the E/S0 galaxies used have a wide luminosity function. But in principle any luminosity indicator with non-zero dispersion leads to this kind of error. For instance de Vaucouleurs (1979a, b) has applied a formula of the form of equation 2a to 328 spiral galaxies, and has also found a seeming increase of H_0 with distance. This is expected, because neither Φ (his T parameter along the Hubble sequence) or Ψ (his van den Bergh luminosity class L) depend very strongly on M_{Gal}, and hence a linear combination cannot depend strongly on galaxy luminosity either, and the method fails both as a relative distance indicator, and surely as an absolutely calibrated one.

The selection bias encountered by applying equation 2a for the distance scale can unfortunately not be remedied by simply replacing Φ, Ψ, . . . by other observables, because if they determine M_{Gal} within a similar luminosity scatter σ_M, they necessarily lead also to a similar bias.

In view of these principal difficulties, other differences of the reduction procedure are of minor importance. For instance, with different choices of the period-luminosity-color relation of Cepheids, the local distance scale is not changed by more than 15%. Also the adopted model of the Galactic absorption, i.e. whether the old model with an absorption of ~ 0.25 mag at the poles is used, or whether the poles are assumed to be nearly absorption-free, as indicated by all newer evidence, affects the distance scale by only $\sim 10\%$.

Acknowledgements. One of us (AS) thanks the National Aeronautics and Space Administration for partial support through grant NAGW-118, concerned with the astronomical ground-based preparation for launch of Space Telescope. The other of us (GAT) thanks the Swiss National Science Foundation for similar support.

REFERENCES

Aaronson, M., Mould, J., Huchra, J., Sullivan, W.T., Schommer, R.A. and Bothun, G.D., 1980, *Ap. J.*, **239**, 12.

Aaronson, M., Persson, S.E. and Frogel, J.A., 1981, *Ap. J.*, **245**, 18.

Aaronson, M., Huchra, J., Mould, J., Schechter, P.L. and Tully, R.B., 1981a, preprint.

Aaronson, M., Dawe, J.A., Dickens, R.J., Mould, J.R. and Murray, J.B., 1981b, *M.N.R.A.S.*, **195**, 1 P.

Ardeberg, A. and de Groot, M., 1973, *Astron. Astrophys.*, **28**, 295.

Arnett, W.D., 1981, preprint.

Audouze, J., 1982, this conference.

Baade, W., 1938, *Ap. J.*, **88**, 285.

— 1941, *Novae and White Dwarfs*, Colloque Internationale d'Astrophysique, P. Swings, B. Strömgren and W. Baade, eds., Paris, p. 177.

— 1952, *Trans. I.A.U.*, **8**, Report of Commission 28, p. 398.

Baade, W. and Swope, H.H., 1963, *A. J.*, **68**, 435.

Barbon, R., Capaccioli, M. and Ciatti, F., 1975, *Astron. Astrophys.*, **44**, 267.

Barbon, R., Ciatti, F. and Rosino, L., 1973a, *Astron. Astrophys.*, **25**, 241.

— 1973b, *Mem. Soc. Astron. Ital.*, **44**, 65.

Bohlin, R.C., Savage, B.D. and Drake, J.F., 1978, *Ap. J.*, **224**, 132.

Boughn, S.P., Cheng, E.S. and Wilkinson, D.T., 1981, *Ap. J. Letters*, **843**, L 113.

Branch, D., 1977, *Supernovae*, D.N. Schramm, ed., Dordrecht, Reidel, p. 21.

— 1981, *NATO Conference on Supernovae*, Cambridge, in press.

Branch, D. and Bettis, C., 1978, *A. J.*, **83**, 224.

Branch, D., Falk, S.W., McCall, M.L. Rybski, P., Komoto, A.K. and Wills, B.J., 1981, *Ap. J.*, **244**, 780.

Bruzual, G., 1981, private communication.

Cadonau, R. and Tammann, G.A., 1981, to be published.

Crampton, D., 1979, *Ap. J.*, **230**, 717.

Crampton, D. and Greasley, J., 1981, *Ap. J.*, in press.

David, Y. and Reeves, H., 1980, *Physical Cosmology*, R. Balian, J. Audouze and D.N. Schramm, eds., Amsterdam: North-Holland, p. 443.

Davidson, W., 1959, *M.N.R.A.S.*, **119**, 54.

Davis, M., Tonry, J., Huchra, J. and Latham, D.W., 1980, *Ap. J. Letters*, **238**, L 113.

de Vaucouleurs, G., 1970, *Ap. J.*, **159**, 435.

— 1977, *Nature*, **266**, 126.

— 1978a, *Ap. J.*, **223**, 730.

— 1978b, *Ap. J.*, **224**, 14.

— 1979a, *Ap. J.*, **227**, 380.

— 1979b, *Ap. J.*, **227**, 729.

de Vaucouleurs, G. and Bollinger, G., 1979, *Ap. J.*, **233**, 433.

Dressler, A., 1980, *Ap. J.*, **236**, 351.

Eggen, O.J., 1979, *Ap. J.*, **230**, 786.

Elias, J.H., Frogel, J.A., Hackwell, J.A. and Persson, S.E., 1981, preprint.

Faber, S.M. and Gallagher, J.S., 1979, *Ann. Rev. Astron. Astrophys.*, **17**, 135.

Fall, S.M., 1979, *Rev. Mod. Phys.*, **51**, 21.

Feast, M.W., 1977, Talk presented at the Magellanic Workshop (ESO, Geneva).

Feast, M.W., Thackeray, A.D. and Wesselink, A.J., 1960, *M.N.R.A.S.*, **121**, 337.

Glass, I.S. and Evans, L.T., 1981, *M.N.R.A.S.*, in press.

Graham, J.A., 1973, *Variable Stars in Globular Clusters and in Related Systems*, J.D. Fernie, ed., Dordrecht: Reidel, p. 120.

— 1975, *Publ. A.S.P.*, **87**, 641.

— 1977, *Publ. A.S.P.*, **89**, 425.

— 1981, preprint.

Gunn, J.E., 1982, this conference.

Gunn, J.E., and Oke, J.B., 1975, *Ap. J.*, **195**, 255.

Gunn, J.E., Stryker, L.L. and Tinsley, B.M., 1981, *Ap. J.*, **249**, 48.

Hanes, D.A., 1979, *M.N.R.A.S.*, **188**, 901.

— 1981, preprint.

Hanson, R.B., 1975, *A. J.*, **80**, 379.

Hausman, M.A. and Ostriker, J.P., 1978, *Ap. J.*, **224**, 320.

Heckmann, O., 1942, *Theorien der Kosmologie* (Berlin).

Heiles, C., 1980, *Ap. J.*, **235**, 833.

Hodge, P.W., 1974, *P.A.S.P.*, **86**, 289.

Holmberg, E., 1950, *Lund Medd. Ser.* 2, No. 128.

Hubble, E., 1926, *Ap. J.*, **64**, 321.

— 1929, *Proc. Nat. Acad. Sci.*, **15**, 168.

— 1936a, *Ap. J.*, **84**, 158.

— 1936b, *Ap. J.*, **84**, 270.

— 1938, *The Observational Approach to Cosmology*, Oxford, The Clarendon Press.

— 1953, *M.N.R.A.S.*, **113**, 658.

Hubble, E. and Humason, M.L., 1931, *Ap. J.*, **74**, 43.

Humason, M.L., 1936, *Ap. J.*, **83**, 10.

Humason, M.L., Mayall, N.U. and Sandage, A., 1956, *A. J.*, **61**, 97.

Humphreys, R.M., 1978, *Ap. J. Suppl.*, **38**, 309.

— 1979a, *Ap. J. Suppl.*, **39**, 389.

— 1979b, *Ap. J.*, **231**, 384.

— 1980a, *Ap. J.*, **238**, 65.

— 1980b, *Ap. J.*, **241**, 598.

Humphreys, R.M. and Sandage, A., 1980, *Ap. J. Suppl.*, **44**, 319.

Kayser, S.E., 1967, *A. J.*, **72**, 134.

Kennicutt, R.C., 1979, *Ap. J.*, **228**, 704.

— 1981, *Ap. J.*, **247**, 9.

Kirshner, R.P., Willner, S.P., Becklin, E.E., Neugebauer, G. and Oke, J.B., 1973, *Ap. J. Letters*, **187**, L 97.

Kowal, C.T., 1968, *A. J.*, **73**, 1021.

— 1978, private communication.

Kraan-Korteweg, R.C., 1982, *Astron. Astrophys.*, in press.

Kraan-Korteweg, R.C. and Tammann, G.A., 1979, *Astron. Nachr.*, **300**, 181.

Kristian, J., Sandage, A. and Westphal, J.A., 1978, *Ap. J.*, **221**, 383.

Lee, T.A., Wamsteker, W. and Wisniewski, W.Z., 1972, *Ap. J. Letters*, **177**, L 59.

Lynden-Bell, D., 1981, *Observatory*, **101**, 111.

— 1982, this conference.

Lynden-Bell, D. and Lin, D.N.C., 1977, *M.N.R.A.S.*, **181**, 37.

Malmquist, K.G., 1920, *Medd. Lund Obs. Ser. 2*, No. 22.

Martin, W.L., Warren, P.R. and Feast, M.W., 1979, *M.N.R.A.S.*, **188**, 139.

Mattig, W., 1958, *Astron. Nachr.*, **284**, 109.

— 1959, *Astron. Nachr.*, **285**, 1.

McAlister, H.A., 1977, *A. J.*, **82**, 487.

McVittie, G.C., 1956, *General. Relativity and Cosmology* (London: Chapman and Hall).

Mould, J., Aaronson, M. and Huchra, J., 1980, *Ap. J.*, **238**, 458.

Oke, J.B. and Searle, L., 1974, *Ann. Rev. Astron. Astrophys.*, **12**, 315.

Olson, D.W. and Silk, J., 1978, *Ap. J.*, **226**, 50.

Osmer, P., 1982, *Ap. J.*, in press.

Peebles, P.J.E., 1976, *Ap. J.*, **205**, 318.

— 1979, *A. J.*, **84**, 730.

Pskovskii, Yu. P., 1977, *Astron. Zh.*, **54**, 1188.

Rivolo, A.R. and Yahil, A., 1981, in press.

Robertson, H.P., 1955, *Pub. A.S.P.*, **67**, 82.

Rood, H.J., Page, T.L., Kintner, E.C. and King, I.R., 1972, *Ap. J.*, **175**, 627.

Rubin, V.C., Burstein, D. and Thonnard, N., 1980, *Ap. J.*, **242**, L 149.

Sandage, A., 1961, *Ap. J.*, **133**, 355.

— 1963, *J. Qualt. Spect. Rad. Transfer*, **3**, 541.

— 1968, *Ap. J. Lett.*, **152**, L 149.

— 1972, *Ap. J.*, **178**, 1.

— 1975, *Ap. J.*, **202**, 563.

— 1982, to be published.

Sandage, A. and Carlson, G.A. 1982, *Ap. J.*, in press.

Sandage, A. and Dressler, A., 1981, preprint.

Sandage, A. and Hardy E., 1973, *Ap. J.*, **183**, 743.

Sandage, A. and Katem, B.N., *A. J.*, **81**, 743.

Sandage, A. and Tammann, G.A., 1969, *Ap. J.*, **157**, 683.

— 1971, *Ap. J.*, **167**, 293.

— 1974a, *Ap. J.*, **190**, 525 (ST I).

— 1974b, *Ap. J.*, **191**, 603 (ST II).

— 1974c, *Ap. J.*, **194**, 223 (ST III).

— 1975, *Ap. J.*, **196**, 313 (ST IV).

— 1976, *Ap. J.*, **210**, 7 (ST V).

— 1981, *A Revised Shapley-Ames Catalog.*, Washington: Carnegie Institution of Washington (RSA).

— 1982, *Ap. J.*, in press (ST VI)

Sandage, A., Tammann, G.A. and Hardy E., 1972, *Ap. J.*, **172**, 253.

Saslaw, W.C. and Aarseth, S.J., 1981, preprint.

Schechter, P.L., 1980, *A. J.*, **85**, 801.

Schweizer, F., 1981, *A. J.*, **86**, 662.

Sciama, D., 1982, this conference.

Silk, J., 1974, *Ap. J.*, **193**, 525.

Smoot, G.F. and Lubin, P.M., 1979, *Ap. J. Letters*, **234**, L 83.

Stebbins, J., Whitford, A.E. and Johnson, H.L., 1950, *Ap. J.*, **112**, 469.

Tammann, G.A., 1977, *I.A.U. Coll. 37*, 43.

— 1979, *Astronomical Uses of the Space Telescope*, F. Macchetto, F. Pacini and M. Tarenghi, eds., Geneva: ESA/ESO, p. 329.

— 1981, *NATO Summer School on Supernovae*, Cambridge, in press.

Tammann, G.A. and Kraan, R., 1978, *I.A.U. Symp. 79*, 71.

Tammann, G.A., Sandage, A. and Yahil, A., 1979, *Les Houches Summer School Lectures.*

— 1980, *Physica Scripta*, **21**, 630.

Tammann, G.A., Yahil, A. and Sandage, A., 1979, *Ap. J.*, **234**, 775.

Terlevich, R. and Melnick, J., 1981, *M.N.R.A.S.*, **195**, 839.

Tonry, J.L. and Davis, M., 1981a, *Ap. J.*, **246**, 666.

— 1981b, *Ap. J.*, **246**, 680.

Tully, R.B. and Fisher, J.R., 1977, *Astron. Astrophys.*, **54**, 661.

van Altena, W.F., 1974, *Publ. A.S.P.* **86**, 217.

van Bueren, H.G., 1952, *Bull. Astron. Soc. Netherlands*, **11**, 385.

van den Bergh, S., 1977, *Décalages vers le rouge et expansion de l'univers*, I.A.U. Coll. No. 37, ed. C. Balkowski and B.E. Westerlund, p. 13.

Visvanathan, N. and Sandage, A., 1977, *Ap. J.*, **216**, 214.

Wagoner, R.V., 1973, *Ap. J.*, **179**, 343.

Wayman, P.A., Symms, L.S.T. and Blackwell, K.C., 1965, *Roy. Obs. Bull.* **98**, E 33.

Weedman, D.W., 1976, *Ap. J.*, **203**, 6.

Wielen, R., 1979, *Mitt. Astron. Ges.*, **45**, 16.

Wild, P., 1960, *Publ. A.S.P.*, **72**, 97.

Yahil, A., 1980, *Tenth Texas Symposium on Relativistic Astrophysics*, in press.

Yahil, A. and Beaudet, G., 1976, *Ap. J.*, **206**, 26.

Yahil, A., Sandage, A. and Tammann, G.A., 1980a, *Ap. J.*, **242**, 448.

— 1980b, *Physical Cosmology*, R. Balian, J. Audouze and D.N. Schramm, eds., Amsterdam: North-Holland, p. 127.

— 1980c, *Physica Scripta*, **21**, 635.

Yahil, A. Tammann, G.A. and Sandage, A., 1977, *Ap. J.*, **217**, 903.

Yang, J., Schramm, D.N., Steigman, G. and Rood, R.T., 1979, *Ap. J.*, **227**, 697.

DISCUSSION

OSTRIKER

I would like to ask you or J.E. Gunn about a method of determining q_0 that you discarded, the classical test which uses the Hubble diagram of bright cluster members. It is my impression that both stellar and dynamical evolutionary effects can be determined by observations of colour or galaxy size to sufficient accuracy so that one can correct for evolution. If this is done, what value of q_0 is obtained?

GUNN

I think that dynamical evolution can be very nicely accounted for by techniques that I discuss briefly in my paper, but I fear that stellar evolution still seems to defy all our best efforts. The stellar populations of elliptical galaxies are more complicated upon closer scrutiny than we previously thought (surprise!), and the crucial parameter, the main sequence mass function slope, is very difficult to determine even from very good data. I do not foresee a clear solution very soon.

FABER

Would you care to comment on the rather larger global value of H_0 found by Aaronson *et al.* using the 21 cm technique?

TAMMANN

The large values of H_0 derived so far from 21 cm line widths rest on a zero-point calibration derived from local *field* galaxies and on an application to *cluster* galaxies (the Virgo cluster and more distant clusters). The results depend therefore on the unproven *assumption* that the line width-luminosity relation is the same for field galaxies and galaxies in different clusters. In the Appendix to the written version of our talk we give indications that this assumption may be incorrect.

WOLTJER

Is your error estimate not very optimistic? With $0.^m2$ for the Cepheids, $0.^m3$ for the brightest red supergiants, and $0.^m3$ for the supernova calibration a real uncertainty of 40% would seem more realistic.

TAMMANN

I agree with your quoted error range. Our quoted error of $H_0 = 50 \pm 7$ km s^{-1} Mpc^{-1} was meant to reflect the 1 σ deviation, whereas the assumption that all three zero-point errors work in the same direction corresponds to a ~ 1.7 σ deviation. Your numbers reduced to the 1 σ level then give a 24% error of H_0. The difference between this value and our 14% stems from our somewhat more optimistic estimate of the different errors.

THE GENESIS OF THE LOCAL GROUP

D. LYNDEN-BELL

Institute of Astronomy,
Madingley Road, Cambridge

ABSTRACT

The evidence for heavy halos about the Galaxy and Andromeda is discussed and possible sizes for them are estimated. The transverse velocity of Andromeda is determined from the radial velocities of Local Group members. The result is consistent with the hypothesis of Gott and Thuan that the mutual gravity of these two large galaxies created their spins. However, the sense of the transverse motion adds to the spins so that the Local Group has an angular momentum. If this angular momentum was itself derived from a tidal torque from another group of galaxies, then that galaxy group ought to have given the Local Group a peculiar velocity also. The velocity of the Local Group with respect to nearby galaxies is determined in an attempt to identify such a group.

On the assumption that the Local Group is gravitationally bound and that its members all came from the Big Bang, Newtonian mechanics allows one to deduce both the total mass of the Local Group and the time since its genesis in the Big Bang. Limits are placed on the mass of any possible halo of the Local Group which is not attached to the Galaxy or Andromeda. Possible roles for heavy neutrinos in Local Group dynamics are briefly considered.

1 - HEAVY HALOS OF THE GALAXY AND ANDROMEDA

Hard evidence for the existence of extended heavy halos around galaxies is difficult to obtain. All efforts to detect such halos other than

by their gravitation have failed and the gravitational evidence is somewhat ambiguous as to the total extent and total mass. However, the mass to light ratio of the Coma cluster is some 40 times the mean mass to light ratios seen within the Holmberg radii of bright galaxies. Likewise, the locally evaluated mass to light ratios rise in the outer parts of normal spirals, although it is only in rare cases that good rotation curves extend beyond 20 to 30 kpc from the centre. To make sense of the positions and motions of the 29 members of the Local Group, we need to have a reasonable picture of its mass distribution. To do this for our Galaxy we need a good value for the circular velocity at the Sun. In 1970 Woolley and Savage [1] found from radial velocities that the mean velocity of the very metal deficient RR Lyrae stars (types IIa and IIb) lags 225 ± 24 km/sec behind, V_c, the local circular motion. If these RR Lyrae stars have no mean motion in the Galaxy, then this lag is the local circular velocity. Any forward rotation of the system of RR Lyraes would yield a greater motion of the LSR. Knapp, Tremaine and Gunn [2] reanalysed and remodelled the 21 cm data and obtained $V_c = 220$ km/s with the low formal error of ± 4 km/s. Other recent determinations are listed in Table 1. With these low values the Galaxy's rotation curve, as determined from the 21 cm line, is nearly flat and on that assumption Lynden-Bell and Frenk have shown [6] that the r.m.s. value of the circular velocity determined from the radial velocities of the globular clusters is 212 ± 16 km/sec. Although high circular velocities (> 250) are favoured from extra-galactic methods [7], large errors are always introduced in the disentangling of the Sun's velocity within the Galaxy from the motion of the Galaxy in the Local

TABLE 1 - *Determinations of the Galaxy circular velocity at the Sun.*

Woolley and Savage [1]	79 RR Lyraes (IIa and IIb) Velocities	$V_c > 225 \pm 24$
Knapp, Tremaine and Gunn [2]	21 cm Modelling	$V_c = 220 \pm 4$
Einasto, Haud and Joeveer [3]	Galaxy Models	220 ± 7
Shuter [4]	21 cm and CO	184 ± 9
Frenk and White [5]	Distribution and Motion of Globulars	213 ± 13
Lynden-Bell and Frenk [6]	Globular Cluster Velocities	212 ± 16
Lynden-Bell and Lin [7]	Local Group Kinematics	294 ± 42
Lin and Lynden-Bell [8]	Modelling Magellanic Stream Dynamics	244 ± 20
Adopted		220 ± 10

Group, so all such methods have large errors. They also suffer statistically from the fact that the Local Group of Galaxies has few independent members that are not close satellites of M31 or the Galaxy. In what follows we shall adopt $V_c = 220$ km/s in the belief that this is probably within 10 km/s of the true value and that an error of 20 km/s or more is unlikely.

Figure 1 shows the line of sight velocities v_ℓ of the globular clusters and the satellites of the Galaxy corrected for the solar motion with respect to the LSR and for the motion of 220 km/s of the LSR. Log $3\,v^2_\ell$ is plotted against log r, where r is the distance from the galactic centre. Also shown is the 21 cm rotation curve for the Galaxy log $V^2_c(r)$. The virial theorem applied to a test particle moving in a potential field shows that in time average the mean square velocity will be $< \underline{v}^2 > = - < \underline{r}.\underline{\nabla}\psi >$. Now on the galactic plane $V^2_c(\underline{r}) = - \underline{r}.\underline{\nabla}\psi$ and we may define V_c at other points by the same formula so $< \underline{v}^2 > = < V^2_c >$. Thus if $< v^2_\ell > = 1/3 < \underline{v}^2 >$ the positions of the points should give an approximate picture of the way the circular velocity varies with distance from the Galactic Centre. The two curves superposed are the interpolation between V_c constant and $V_c \propto r^{-1/2}$ given by

$$V^2_c = \frac{V^2_0}{[1+(r/r_h)^2]^{1/2}}$$

and r_h we refer to as the halo radius. The total mass of such a system is $V^2_0\, r_h/G$. If the system were spherically symmetrical its density distribution would be given by

$$4\,\pi G\rho = \frac{V^2_0}{r^2\,(1+r^2/r^2_h)^{3/2}}$$

which behaves like r^{-2} for $r \ll r_h$ and r^{-5} for $r \gg r_h$. The rather weak evidence of the Figure suggests that the halo extends to about 100 kpc. The two curves drawn correspond to $r_h = 75$ kpc and 150 kpc corresponding to total masses of 0.8 and 1.6×10^{12} $M_\odot$.

Figure 2 shows the same thing done for M31 based on the 21 cm rotation curve of Newton and Emerson [9]. All Local Group members with known radial velocities, which lie within 40° of M31, are plotted. Some are plotted twice, once at their apparent separations from Andromeda by a dot and again at their true separations after allowing for different

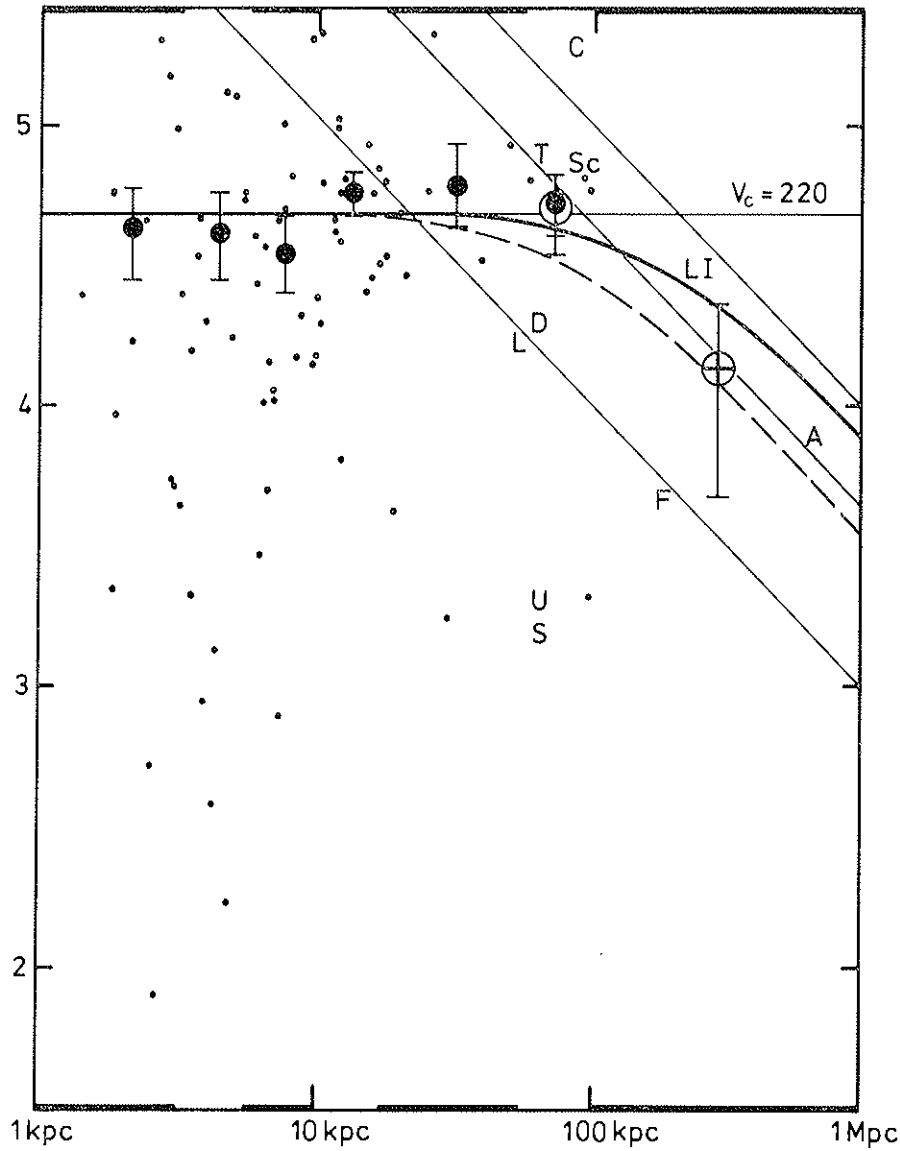


FIG. 1. Log ($3V^2$) versus log r for the line of sight velocities of globular clusters and satellites of the Galaxy. The hard dots denote r.m.s. averages of globular clusters in bins denoted in the margin.
C = Carina, Sc = Sculptor, T = Tip of Magellanic Stream, D = Draco, L = LMC, U = Ursa Minor, S = SMC, LI = Leo I, F = Fornax, A = M31.

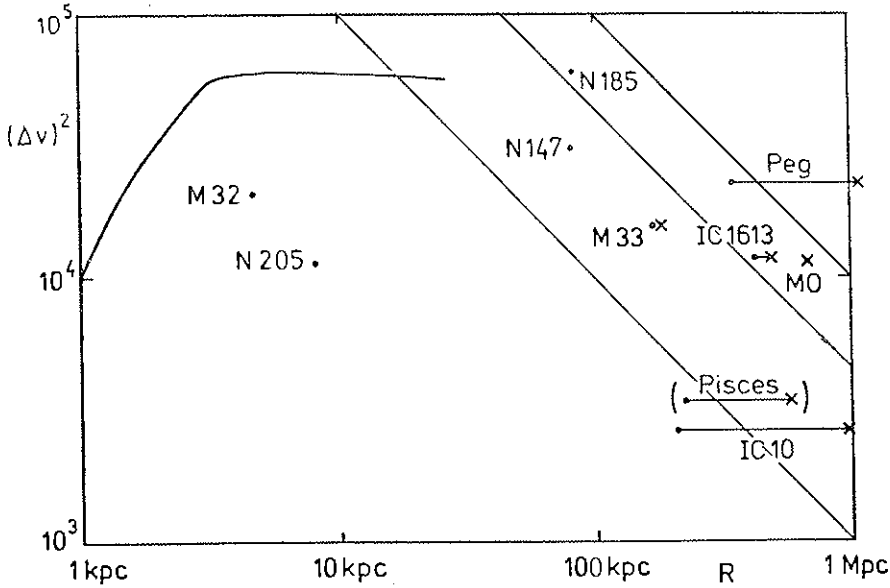


FIG. 2. Possible satellites of Andromeda are plotted with its velocity curve in a log $(\Delta v)^2$ versus log R plot. Δv is the best estimate of the velocity with respect to the centre of M31. The satellites have only one component of velocity measured so their $(\Delta v)^2$ have been multiplied by 3 to compensate. The points · are plotted with R, the projected distance from Andromeda as seen on the sky. The points X are plotted at the estimated true separations from Andromeda. For the Galaxy, MO, the total velocity determined in this paper has been used reduced by a factor $(3/5)^{1/2}$ to correct for a mass ratio that is 2:3 rather than infinitesimal. Notice the fall in $(\Delta v)^2$ beyond 100 kpc. The slanting lines are $(\Delta V)^2$ R/G equal to 2.3×10^{11} $M_\odot$, 1×10^{12} $M_\odot$ and 2.3×10^{12} $M_\odot$.

distances. IC10 and Pegasus are too far from Andromeda to be considered as possible satellites. The distance to Pisces = LGS III is very uncertain.

In each galaxy there is some evidence for a fall in $< \Delta V^2 >$ beyond about 100 kpc. There is also evidence that the number density of small galaxies increases towards the two major ones suggesting that they are the dominant gravitational influence within one or two hundred kpc.

In the above discussion I have left aside data from computations of the Magellanic stream by Murai and Fujimoto [10] and Lin and Lynden-Bell [8], although these works do favour a $\sim 10^{12}$ $M_\odot$ mass for our Galaxy. This is because the deductions are complicated and depend on subsidiary hypothesis, e.g. only gravity is acting, the Magellanic stream was not torn off by the Magellanic Clouds hitting the outermost part of tbe galactic plane, etc. Such calculations can be verified by a direct measurement of

the proper motion of the Magellanic Clouds. We predicted $\mu = .0020''/yr$ due East at the Large Cloud; a proper motion of only $.0015''/yr$ due East would mean that our Galaxy had no heavy halo and a zero or western proper motion would mean that the Magellanic Clouds are going in the opposite sense around their orbit. Davies and Wright [11] advocate this with no halo. Lin and Lynden-Bell [12] originally advocated this sense of motion, but we can only get the right motion for the SMC with the orbit in the sense advocated by Fujimoto and collaborators. Evidence from tidal radii of clusters is two edged. Ostriker originally used it to derive high masses, but a recent preprint from Wakamatsu [13] 'corrects' this for past closer approaches to the galactic centre and, using data from diffuse satellites, finds a low mass for the Galaxy. Any result depends on the mass to light ratio of the objects used. For the outer satellites and diffuse class XII globular clusters this has not been determined; the tidal radii must be cubed and they are none too certain to start witb. Very different results would be obtained if instead of taking M/L constant at tbe globular cluster value, it were assumed to increase at lower surface brightnesses. Petrou [14] has given a law of this type which brings together data in quite a number of galaxies.

Although the data cited above suggest that heavy halos are present, other evidence leaves room for serious doubt. The Ursa Minor dwarf is a short streak in the sky that has such a low star density that Hodge and Michie [15] deduced the Galaxy must be tearing it apart. Indeed its major axis [16] at position angle $50 \pm 4°$ agrees precisely with the direction of the plane of the Magellanic Stream [17] where tbe position angle is $50.6° \pm 2°$. This demonstrates that the orbit of Ursa Minor is in that plane and that it is probably debris from the break-up of the Greater Magellanic Galaxy [18]. Perhaps more surprisingly Draco, which is rounder, is also elongated along the direction of the Magellanic Stream; the stream is at position angle $80° \pm 2°$ when it passes Draco whose elongation is at $84° \pm 3°$. Perhaps Draco is a temporary aggregation in a stellar stream rather tban a properly bound object. I pointed out earlier that Draco is almost exactly opposite to the LMC in the galacto-centric sky and that Ursa Minor is almost exactly opposite to the SMC [18]. The radial velocities of the two pairs (corrected to the Galactic Centre) are -82, $+76$ and -26, $+22$ which are very close to being equal and opposite. Such equal and opposite radial velocities of objects in opposite directions are predicted for objects following the same orbit under the gravity of a point mass. The presence of a heavy halo spoils this property. Lin and

Lynden-Bell [8] find from our latest computation of the Magellanic Stream in a heavy halo that both the debris and the orbital radial velocities are positive rather than negative 180° around the sky from the Magellanic Clouds. This discrepancy highlights the need for more accurate velocities; the error on Draco is ± 35 and on Ursa Minor ± 70, so the apparently equal and opposite radial velocities may not be more than coincidence. Ursa Minor is clearly being torn up now, so it cannot have survived past perigalactic passages. This suggests that it may have been torn out of the Magellanic Clouds at the last close passage some 2.6×10^9 years ago. It is interesting to compare this time with what can be deduced from its unusual H.R. diagram (which has blue stars on the Horizontal Branch). Draco has the red Horizontal Branch typical of dwarf spheroidals. Altbough it shows less sign of current disruption, its larger radial velocity will bring it closer to the Galaxy than Ursa Minor; thus it is unlikely to survive perigalactic passages. Proximity in the sky suggests that these systems may have torn off the Magellanic Clouds together at what is believed to be the event that generated the rest of the Magellanic Stream, but their H.R. diagrams indicate that the stars are of a much greater age, of the order of 10^{10} years. Evidently there is still a mystery here that requires elucidation.

2 - Does the Local Group Spin?

Vortical perturbations become singular as one proceeds back to the Big Bang. Many authors, therefore, prefer adiabatic or isothermal perturbations as the seeds of structure in the universe. There is no observational evidence to back this up. Gott and Thuan [20] have however put forward an ingenious argument based on these ideas that has verifiable consequences. Since the initial perturbations do not spin, the observed angular momenta of galaxies have to arise from somewhere. Hoyle, Peebles, Oort and others have proposed that irregular density perturbations left from the Big Bang generate angular momentum in each other via tidal torques. We are asked to imagine an irregular density distribution expanding with the universe. The quadrupole moment of each proto-galaxy will grow as it expands like the square of its length scale a. The tidal torque due to a given neighbour decreases as its distance ℓ increases, hence $\dot{J} \propto a^2\, \ell^{-3}\, GM_1M_2$. If in the expanding universe ℓ and 'a' both behaved like $t^{2/3}$, then $\int \dot{J}dt$ would behave as $t^{1/3}$. J will grow significantly until the self gravity of our proto-galaxy slows its expansion. When $a^2\, \ell^{-3}\, t$ starts de-

creasing, which is somewhat before 'a' reaches its maximum extent, most of the final angular momentum will have been generated. Gott and Thuan [20] suggest that M31 was the strongest influence in causing the Galaxy to spin and that the Galaxy was the strongest influence in causing M31 to spin. They assume that before any spins were generated the two proto-galaxies were expanding away from each other in a straight line but each had an irregular mass distribution with a quadrupole momentum. Each galaxy responded to the gravity of the other, which may be expanded into monopole, quadrupole, octupole, etc. contributions (no dipole since we expand about the barycentre of each galaxy). The monopole interaction slows the expansion, but produces no spin. Andromeda's monopole field acts on our quadrupole moment to produce angular momentum. Andromeda's quadrupole field does likewise but, since it will make a contribution that is smaller by a^2/ℓ^2, we shall neglect it. Now Andromeda acting as a monopole can produce no torque along the line joining the two galaxies, since the monopole field is symmetrical about that line through us. Hence, the angular momentum produced in us will be perpendicular to the original line joining the galaxies. By symmetry the angular momentum produced in Andromeda by us will likewise be perpendicular to this original line. Now the present angular momenta lie along the spin axes of the two galaxies which are known to better than 1°. In galactic coordinates the unit vectors are $\hat{\omega}_A = (-.419, -.751, -.510)$ and $\hat{\omega}_G = (0, 0, -1)$ so the original line joining us to Andromeda must have been (one way or the other) along $\hat{\omega}_G \times \hat{\omega}_A = (-.751, +.419, 0)$ that is $\ell = 151$, $b = 0$. This can be contrasted with the present direction to Andromeda of $\ell = 121$, $b = -22$ which lies 36° away (or 144° if the other original sense is chosen).

Our attention now turns to possible causes for this offset which is illustrated in Figure 3. When a monopole produces angular momentum in a quadrupole, it obviously experiences an equal back reaction. Thus the same interactions that cause the spins of the perturbations generate a little orbital angular momentum of the galaxies about each other. Since we started with no spins by hypothesis, this orbital angular momentum must be equal and opposite to the total spin angular momentum. Giving the Galaxy μ times as much angular momentum as Andromeda, the total spin angular momentum will lie along $\mu\hat{\omega}_G + \hat{\omega}_A \propto \hat{J}_s$. If no other bodies affected the system, our orbital angular momentum about M31 should lie opposite to this. However, the orbital angular momentum must be perpendicular to the separation and so $\mu = -\dfrac{\hat{\omega}_A \cdot \hat{r}_A}{\hat{\omega}_G \cdot \hat{r}_A} = \dfrac{+.208}{.368} = .57$

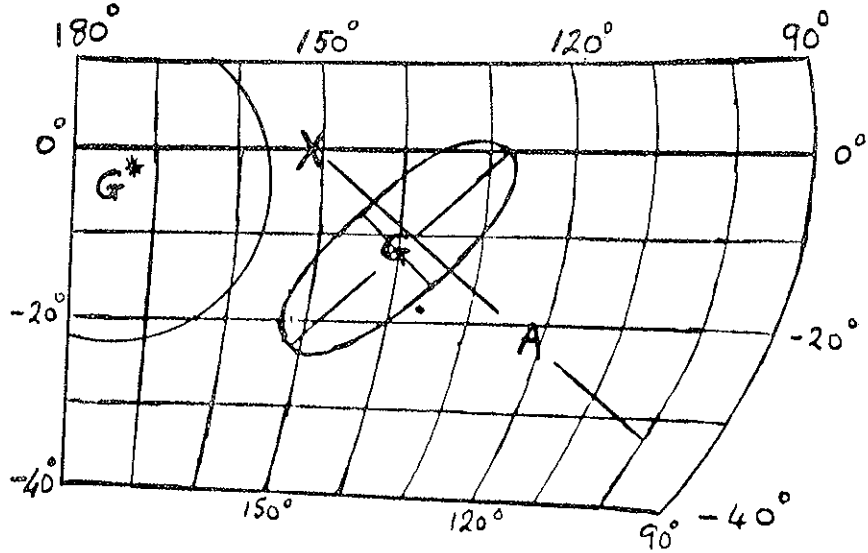


FIG. 3. X is the original direction to Andromeda according to the argument of Gott and Thuan [20]. A is the direction to M31 now. G denotes the direction of the Galaxy's motion through the Local Group according to Einasto and Lynden-Bell [21], including its error ellipse. The dot denotes the determination by Yahil, Sandage and Tammann [23] corrected to $V_c = 220$. G* is the direction of the Galaxy's velocity relative to galaxies.

where $\hat{r}_A$ is the unit vector to Andromeda now. This is not at all an unreasonable value for the ratio of the angular momenta. With $\mu = .57$ the spin angular momentum would lie along $(-.304, -.545, -.781)$ and the orbital would have to be equal and opposite. This is the first difficulty with the zero total angular momentum hypothesis. If the line from Andromeda swings from the old line 36° to the current line, then this sense of motion would add to the spin deduced above.

To save the zero angular momentum hypothesis, Gott and Thuan [20] suggested that the old line to Andromeda should be taken in the other sense to lie at $\ell = 180° + 151°$ rather than 151° and so it has swung tbrough 144° instead of 36°. However, most of such a swing must have occurred when the galaxies were close together at the very time when the spins were being generated. But this theory is based on the idea that during that time the galaxies were separating more or less along their line of centres rather than swinging past each other, so such a solution is incomplete. Nevertheless it makes a definite prediction that the motion of our Galaxy with respect to Andromeda must lie on the extension of the

great circle XA beyond A; only then can the orbital angular momentum counterbalance the spin.

Now it is possible to determine the transverse velocity of the Galaxy about Andromeda. One uses the radial velocities of the stray members of the Group to determine the Sun's velocity through it. One tben subtracts the solar motion about the Galaxy to get the Galaxy's motion with respect to the Group. Since Andromeda is the only other significant body, its direction of motion must be equal and opposite by momentum balance. Indeed, if we are prepared to assume that the mass ratio lies in the range $1 \leq \frac{M_A}{M_G} \leq 3$ the component of solar motion $\underline{U}$ towards Andromeda has to be in the narrow range $240 < \underline{U} < 268$. If the Fisher-Tully relation is assumed correct to within a factor of 1.4 or so $\frac{M_A}{M_G} - \left(\frac{250}{220}\right)^4 = 5/3$ and so $1.2 \leq \frac{M_A}{M_G} < 2.4$ which is still more restrictive. Thus one component of the Sun's motion is effectively determined by momentum balance. We use the stray members of the Local Group, those more than 400 kpc from the Galaxy and M31, to determine the other two components [21]. The 1σ contour for the direction of the Galaxy's velocity is drawn on the sky; while points on the line between X and A are clearly consistent, points on the extrapolation beyond A are excluded at the 90% level. We deduce that the orbital angular momentum is in roughly the same direction as the spin, so the Local Group has a net angular momentum.

The period of spin angular momentum transfer between the two major galaxies should have occurred when separations were comparable to current sizes, until $\ell \sim 100$ kpc or perhaps twice that figure. However, at that stage the quadrupole moment of the dumbell formed by M31 and the Galaxy will still have been growing vigorously. Most of the angular momentum transfer to the Local Group is likely to have occured at something like the present separation of 700 kpc by which time gravity was beginning to slow down the expansion of the Local Group.

The Galaxy and Andromeda may still be responsible for each other's spins but the Local Group may still have acquired its angular momentum from a neighbouring group rather later. If a single neighbouring group dominated it must have been either beyond Andromeda and less than 90° along the great circle XA or behind our heads in the diametrically opposite

quadrant of that great circle . For a given strength of tide, the torques will be greatest between 20° and 70° from X along XA. There is no prominent nearby group in the Northern Sky but it is amusing that the NGC 5128 group lies in the right direction in the Southern Sky.

The hypothesis that a single group was responsible for causing the Local Group to spin is subject to a different observational check. If some group was near enough to have a big tidal effect then it was certainly near enough to produce a sizeable acceleration. If that is the case, then the Local Group ought to be moving (with respect to the mean of very nearby galaxies) in the direction of the group responsihle for its spin. Before investigating this, let us summarise the prediction. We hope to find a prominent nearby group of galaxies on the great circle XA between 20° and 70° from X (in the sense XA) or between 200° and 250°. We hope that it might still have the largest M/ℓ^3 at the Local Group and that the Local Group has a peculiar velocity towards it.

3 - The peculiar velocity of the Local Group with respect to nearby galaxies

The determination of this velocity is beset with difficulties. The distribution of nearby galaxies has a significant density gradient towards Virgo and the distances are poorly known, especially if distances determined from their velocities are to be ignored. One possible approach is to use the Infrared Fisher-Tully relation to get approximate relative distances and then solve for the motion of the Local Group with respect to those galaxies to which the Fisher-Tully method can be applied. J. Mould kindly supplied me with 14 distances determined to large spirals by this method but the resulting motion of the Local Group had so large an uncertainty that the result contains nothing of value: (− 54, − 148, − 20) but the eigenvalues of the error matrix were ± 34, ± 189 and ± 22 km/s.

This method cannot be greatly improved because the number of large spirals at low redshift and suitable inclination is severely limited. To go to larger redshifts inevitably involves us in the Virgo flow, whereas it is our object to look for a more local phenomenon.

To get around these difficulties I developed a new method which uses neither the distances nor the magnitudes of the galaxies, but instead uses the whole mass of galaxy velocities. If the density of galaxies were uniform, then the number N(V) with velocity $<V$ should vary like V^3 out to the

velocity cut-off of the catalogue. Observational selection will change this so that N (V) rises somewhat less steeply. If the distribution is observed from a moving point, then the apparent distribution will have an additional term $\underline{V}_\odot \cdot \hat{\underline{r}} \frac{dN}{dV}$ which varies like a cosine around the sky. If there is not a uniform density of galaxies, but a uniform density gradient so that $\rho = \rho_0 (1 + \underline{k} \cdot \underline{r})$, then N (V) will contain an additional angle-dependent term proportional to V^4. Since both the additional terms vary like the direction cosines they are not distinguishable from their angular dependence. However, since the first term should behave like V^2 and the second like V^4 (temporarily neglecting observational selection) we see that a separation is possible in principle.

From the Kraan-Tammann Catalogue [22] I have used this method to find both the density gradient in the galaxies and the velocity of the Local Group. I first remove from the catalogue the members of the Virgo Cluster which only occur in it because of the large velocity dispersion in the centre of Virgo. I also remove the members of the Local Group. I then correct every galaxy's velocity for the motion of the Sun through the Local Group and plot a histogram F (V) of the numbers of galaxies in different velocity bins. I also plot histograms of the values of $X = \cos \ell \cos b$ against velocity and likewise for $Y = \sin \ell \cos b$ and $Z = \sin b$. A small velocity ΔV of the Local Group should show up in the X histogram as a contribution $(\Delta V)_x \frac{\partial F}{\partial V}$ etc. Likewise a density gradient should show up in the same histogram as a part proportional to $\underline{k}_x N (V)$ where $N = \int F dV$. We thus analyse the X, Y and Z histograms into these two contributions and the coefficients yield the density gradient and the velocity of the Local Group. Unfortunately the velocity of the Sun through the Local Group turns out to be the major source of uncertainty. To avoid this unnecessary source of error, I have added back the value initially assumed so as to get the velocity of the Sun with respect to the distant galaxies. This seems to he quite well determined as the figures indicate.

		$V_{\odot x}$	$V_{\odot y}$	$V_{\odot z}$
All Galaxies	$V_0 < 500$ km/s	− 80	255	25
	$V_0 < 400$ (a)	− 96	254	5
	$V_0 < 400$ (b)	− 59	220	− 15
	$V_0 < 300$	− 97	217	− 5

(a) and (b) are two different reductions of the same data. (a) puts greater emphasis on fitting the low velocity parts of the histogram and (b) the higher velocity parts. (a) gives the better fit overall.

The spread of the above numbers gives dispersions of about 15 km/s and it is probable that the true errors are about 20 km/s in each coordinate. As tbe galaxies with the higher velocities are in a different part of the Virgo flow, we shall use the average of the above numbers; the upper values have more weight but the lower ones are more local.

It is of interest to compare this velocity of the Sun with respect to external galaxies of $(-80, 237, 3) \pm 20$ with its velocity through the Local Group [21] of $(-49{,}280, -9)$ or the Yahil, Sandage, Tammann [23] value of $(-79.1, 295.4, -37.6)$. It seems clear that the errors of the Local Group determinations are so large that no difference with the first solution can be reliably determined. The difficulty in determining the Sun's velocity through the Local Group will probably be with us always, because there are too few outlying members of the system which are not attached to the major galaxies. The only hope would be an attempt at a full integration of all members of the Local Group as they emerged from the Big Bang.

A different but interesting result is obtained by correcting the above figures for the Sun's motion in the Galaxy of $(9, 12, 7) + (0, 220, 0)$. We then obtain the velocity $\underline{G}^*$ of the Galaxy relative to the system of galaxies in the group $V_0 < 300$ km/s.

$$\underline{G}^* = (-89, 5, -4) = 89 \text{ km/s towards } \ell = 175\,, b = -4\,.$$

This direction is better determined than $\underline{G}$ and adds some weight to the idea that the Local Group does spin since the offset of $\underline{G}^*$ from A is in the same sense as the offset of $\underline{G}$ from A. This is to be expected if the velocity of the Local Group is small.

If the Local Group was spun up by a single object then the ratio of the velocity given it to the angular momentum given it will be of the order of $\frac{V}{h} = \frac{\ell}{a^2}$, where '$a$' is the separation of M31 from the Galaxy at the relevant time and ℓ is the distance to the disturbing group. Putting $h = 40$ km/s $\times\, a$ and $a = .5$ Mpc we get $V = 80\ (\ell/\text{Mpc})$ km/s. The velocity of the Local Group seems to be clearly less than this since ℓ can hardly have been less than a megaparsec.

Thus, the Local Group is not moving fast enough to have acquired

its spin by the tidal torque of a single mass concentration. A conceivable way out would be to evoke the Virgo Cluster and a large Virgo-centric flow, but Virgo is not in the right direction to give the observed angular momentum.

There remain the possibilities:

(i) The spin of the Local Group was caused by the combined effect of a number of local groups of galaxies;

(ii) Since the net acceleration weights galaxies like $1/r^2$, while the tidal fields weight them like $1/r^3$, the acceleration will have averaged out over a larger region. Thus, a small velocity is not necessarily in conflict with a significant angular momentum.

In conclusion, the Local Group does spin and the angular momentum is in the direction $\sim (-.304, -.545, -.781) = (\ell = 241\,, b = -51)$.

It is not unlikely that the tidal torque from Andromeda caused us to spin and vice versa, but the origin of the angular momentum of the Local Group as a whole remains an intriguing mystery.

4 - A DYNAMICAL ESTIMATE OF THE TIME SINCE GENESIS [24]

The distance to those Local Group members whose expansion has just been stopped by the gravity of the Local Group yields Mt^2 where M is the mass of the Group and t the time since expansion began. The distance and radial velocity of M31 yield a relationship between Mt^2 and t. Thus M and t may be deduced.

The effect on the Hubble flow of large mass concentrations, such as the Virgo Cluster, has been much discussed in attempts to determine the mass-to-light ratio of the material in the universe and hence q_0. However, interesting results can be obtained from the Local Group itself. In particular, if it is believed that the bodies of the Local Group separated from one another at a time corresponding to $z > 3$ when the scale of the universe was smaller by a factor of 4 or more, then we may assume that Newtonian dynamics has operated since an epoch at which the Local Group was quite small. Kahn and Woltjer [25] have already applied such an argument to the motion of the Galaxy and M31 and have thereby found limits on the mass of the Local Group. Their work gives a relationship between the sum of the masses of the Galaxy and M31 and the time since they were close to one another in the Big Bang. The aim of this paper is to point out

that the small outlying members of the Local Group can give a second relationship between that time and the mass of the Local Group. These relationships may be solved to obtain both the total mass and the time, which we shall call the dynamical age of the Local Group. The assumption made is that the total mass that has decelerated the outlying members of the Local Group is practically the same as the sum of the masses of the Galaxy and M31 which governs our two-body motion.

To introduce the method we start by considering a radial orbit for the Galaxy and Andromeda. If r is the separation between these galaxies then if $M = (M_A + M_G)$, where the subscript A stands for Andromeda, i.e. M31, and the subscript G refers to our Galaxy,

$$1/2\, v^2 - GM/r = -\, 1/2\, GM/a \tag{1}$$

where $v = \dot{r}$ and 2a is the maximum value of r achieved on the orbit. To integrate, we use the substitution

$$r = a\,(1 - \cos 2\,\eta)\ . \tag{2}$$

Notice $r = 0$ when $\eta = 0$. On integration one finds after a little algebra

$$2\,\eta - \sin 2\,\eta = (GM/a^3)^{1/2}\ t\ , \tag{3}$$

where $t = 0$ when $\eta = 0$.
The observables are the values of r and v at the present time, and the unknowns we seek are t and GM. Eliminating a from (2) and (3), we obtain

$$\Omega\, t \equiv (GM/r^3)^{1/2}\ t = 2^{-1/2}\ (\eta - \sin\ \eta\ \cos\ \eta)/\sin^3\ \eta\ . \tag{4}$$

But from (1) we also have

$$\omega \equiv v/r = [GM/r^3)\ (2 - r/a)]^{1/2} = (2\,GM/r^3)^{1/2}\ \cos\ \eta\ \ , \tag{5}$$

so, using (4), we find

$$\omega\, t = (\eta - \sin\ \eta\ \cos\ \eta)\ \cos\ \eta/\sin^3\ \eta\ . \tag{5}$$

Notice that η can be eliminated between (4) and (5) to give Ωt as a function of ωt. We do this graphically by choosing different values of η and plotting expression (4) against expression (5). For $-\,3 < \omega t < +\,0.3$ the graph

(Fig. 4) is rather accurately (within 3 per cent) given by

$$\Omega t + 0.85\,\omega t = 2^{-3/2}\,\pi \cong 1.11\ , \tag{6}$$

which is exact at $\omega = 0$.

So far, we have considered only the binary motion of the Galaxy and M31, but if we now turn to an outlying member, i, of the Local Group at $r_i(t)$ which was always significantly farther from the mass centre than

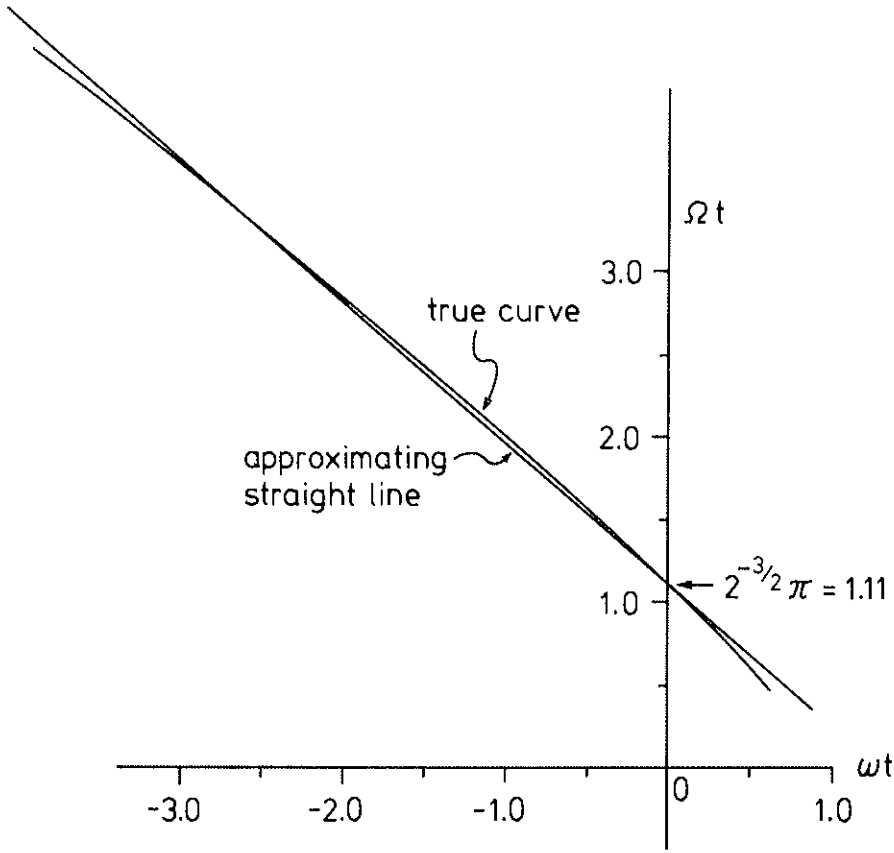


FIG. 4. $\Omega = \left(\frac{GM}{r^3}\right)^{1/2}$. The graph is Ωt against $\frac{vt}{r}$ for straight line motion of two bodies. The straight line is a good approximation in the range $-3 < \omega t < 0.3$.

either of the former, then the equation of motion is

$$\ddot{\underline{r}}_i = -\frac{GM_A(\underline{r}_i - \underline{r}_G)}{|\underline{r}_i - \underline{r}_A|^3} - \frac{GM_G(\underline{r}_i - \underline{r}_A)}{|\underline{r}_i - \underline{r}_G|^3} .$$

As a first rough approximation we put all the mass at the centre:

$$\ddot{\underline{r}}_i = -G(M_A + M_G)\,\underline{r}_i / |\underline{r}_i|^3 ;$$

Assuming that i, too, took part in the Big Bang, we see that r_i obeys the same equations as r. Thus, as in (6)

$$\Omega_i t + 0.85\,\omega_i t = 2^{-3/2}\pi ,$$

where $\Omega_i = (GM/r_i^3)^{1/2}$ and $\omega_i = v_i/r_i$.
We now look at the distant members of the Local Group and determine the distance r_0 (measured from the mass centre) at which the expansion has been momentarily stopped by the gravity of the Group. Evidently $v_0 = \omega_0 r_0 = 0$, and so

$$(GM/r_0^3)^{1/2} t = \Omega_0 t = 2^{-3/2}\pi . \tag{7}$$

Multiplying by $(r_0/r)^{3/2}$, we have $\Omega t = 2^{-3/2}\,\pi\,(r_0/r)^{3/2}$, and substituting in (6) we find

$$t = (0.85)^{-1/2-3/2}\,\pi\,[(r_0/r)^{3/2} - 1] / (-\omega) = 1.31\,[(r_0/r)^{3/2} - 1]\,r / (-v) . \tag{8}$$

Also, from (6),

$$GM = 0.72\,v^2 r\,[1 - (r/r_0)^{3/2}]^{-2} . \tag{9}$$

When dealing with clusters of galaxies it is useful to think with a velocity unit of 100 km/s and a distance unit of 1 Mpc; the time unit is then 10^{10} years. If we further take 2.3×10^{12} $M_\odot$ as our mass unit, we find that $G = 1$.

The observed velocity of Andromeda is -301 ± 2 km/s. Reducing this to the LSR gives -298 km/s. Taking our circular velocity from recent determinations to be 220 km/s we get a correction to the Galactic Centre of 175 km/s which yields the velocity of the Galactic Centre with respect

to M31 of $v = -123$ km/s. From Yahil *et al.* [23] $r = 0.68$ Mpc and, with greater uncertainty, $r_0 = 1.5$ Mpc. Hence, from equations (8) and (9),

$$t = 1.6 \times 10^{10} \text{ years} ;$$

$$M = 3.6 \times 10^{12} M_\odot .$$

This t is the time since the expansion of the Local Group began. Most cosmologists think it should be equal to the age of the universe, which may also be expressed as a function of Hubble's constant H_0 and of q_0. But t is a strict age, not the reciprocal of H_0.

The greatest uncertainty in the above lies in the distance r_0, but there are also uncertainties due to the modelling with purely radial orbits and the approximation of putting both heavy masses at the centre when the third body's orbit was considered. To amend this approximation some 3 body integrations were carried out using coordinates that are co-moving with the two heavy bodies, and a rescaled time.

The equations of motion of N bodies are

$$\ddot{\underline{r}}_i = -G \sum \frac{m_j (\underline{r}_i - \underline{r}_j)}{r_{ij}^3} \quad \text{where } r_{ij} = |\underline{r}_i - \underline{r}_j| \ . \tag{10}$$

Suppose we rescale by writing $\underline{r}_i = R(t)\, \underline{r}'_i$ and we take a rescaled time such that $d\tau = f(t)\, dt$. Then

$$\frac{d^2 \underline{r}'_i}{d\tau^2} + \frac{d}{d\tau} [\log (fR^2)] \frac{d \underline{r}'_i}{d\tau} + \frac{\ddot{R}}{Rf^2} \underline{r}'_i = -\frac{G}{R^3 f^2} \sum_{j \neq i} m_j \frac{\underline{r}'_i - \underline{r}'_j}{\underline{r}'^3_{ij}} . \tag{11}$$

So far both $R(t)$ and $f(t)$ are at our disposal. We now choose $f = \sqrt{GM}\, R^{-3/2}$ so that the right hand side retains its familiar form, but with $\underline{r}'$ replacing $\underline{r}$. We next choose the scale $R(t)$ so that in the rescaled variables Andromeda and the Galaxy have unit separation at all times. In the special case in which these two galaxies are the only important bodies present, we have the following equation for their separation vector $\underline{R}$:

$$\ddot{\underline{R}} = -GM \underline{R}/R^3 \tag{12}$$

where $M = M_A + M_G$. Hence $\underline{R} \times \dot{\underline{R}} = h = \text{constant}$

$$\ddot{R} = + \frac{h^2}{R^3} - \frac{GM}{R^2} \quad . \tag{13}$$

R (t) is the solution of this equation and hence $R^2 \ddot{R} = - GM + h^2/R$. For a motion that comes from the Big Bang we must have $R = 0$ at $t = 0$ and so we have $h = 0$ and straight line orbits. Thus, in our revised units we have

$$\frac{d^2 \underline{r}'_i}{d\tau^2} = - \sum_{j \neq 1} \mu_j \frac{\underline{r}'_i - \underline{r}'_j}{r'^3_{ij}} + \underline{r}'_i - 1/2 \frac{d}{d\tau} (\log R) \cdot \frac{d \underline{r}'_i}{d\tau}$$

where $\mu_i = m_i/M$.

Notice that a "cosmical repulsion" term $\underline{r}'_i$ has appeared and a damping force with "friction" coefficient $1/2 \dfrac{d}{d\tau} (\log R)$. The first two force terms come from a potential. If we now neglect the gravity of all but the two heavyweights and drop the suffix i, then $\underline{r}'$, the position vector of any of the small galaxies, satisfies

$$\frac{d^2 \underline{r}'}{d\tau^2} = - \left[\frac{\mu (\underline{r}' - \underline{r}_G)}{|\underline{r}' - \underline{r}_G|^3} + \frac{(1-\mu)(\underline{r}' - \underline{r}_A)}{|\underline{r}' - \underline{r}_A|^3} - \underline{r}' \right] - 1/2 \frac{d}{d\tau} (\log R) \frac{d\underline{r}'}{d\tau} . \tag{14}$$

The effective potential is $\psi = \left[\dfrac{\mu}{|\underline{r}' - \underline{r}_G|} + \dfrac{(1-\mu)}{|\underline{r}' - \underline{r}_A|} + \dfrac{r'^2}{2} \right]$ (15)

and $\mu = M_G/(M_A + M_G)$ which we take to he 0.4.

To integrate equation (14) one needs an expression for $\dfrac{d (\log R)}{d\tau}$ in terms of τ. This is supplied by integrating (13) to obtain

$$\frac{\dot{R}^2}{2} - \frac{GM}{R} = - \frac{GM}{2a} \tag{16}$$

where a is the maximum separation. Changing variables to τ we have

$$\left(\frac{d \log R}{d\tau} \right)^2 = 2 (1 - R/2a) \; ; \tag{17}$$

putting $R = 2a \sin^2 \eta$ as in (2) we find

$$\frac{d\tau}{d\eta} = \sqrt{2} \operatorname{cosec} \eta \tag{18}$$

and

$$\tau = \sqrt{2} \log [\tan (\eta/2)] \tag{19}$$

where we have chosen τ to be zero at the time of maximum separation ($\eta = \pi/2$, $R = 2a$).

The friction coefficient needed in (14) is

$$1/2 \frac{d \log R}{d\tau} = 1/2 \cos \eta = 1/2 \left(\frac{1 - \tan^2 \eta/2}{1 + \tan^2 \eta/2} \right) = 1/2 \left[1 - \frac{2 \exp \sqrt{2}\tau}{1 + \exp \sqrt{2}\tau} \right] \tag{20}$$

Notice the damping changes to excitation after the time of maximum separation. It is sometimes useful to know t rather than just τ. This is given by using equation (3) for $t(\eta)$ and (19) for $\eta(\tau)$, or (18) is $R(\eta)$ if desired. For the Galaxy and Andromeda the greatest separation, 2a, was of the order of 1 Mpc.

The potential ψ is large both near the heavy bodies and at ∞. There are the usual neutral points of the 3 body problem on the line joining the bodies and at the equilateral triangular points. Figure 5 shows the equipotentials. Bodies starting at (comoving) rest from any point significantly within the rampart that passes through the triangular points fall into one of the heavybodies and are caught due to the damping. Bodies starting significantly outside the rampart rapidly expand away from the Local Group. Thus the only starting positions that lead to Local Group members which are not close satellites of the Galaxy or M31 are those on the rampart or close to the central neutral point. A number of such orbits were computed and are shown in Figure 5. That gives us a rather random selection of orbits starting close to the rampart, but by computing many more orbits of this type we have found ones that are appropriate for IC10, Pegasus, WLM and Aquarius. For these far out members of the Local Group the approximation of putting all the mass at the centre is surprisingly good, considering that they start hardly further away than the distance between the two heavy bodies. Figure 6 shows Ωt plotted against vt/r for these orbits together with the straight line taken from Figure 4. While

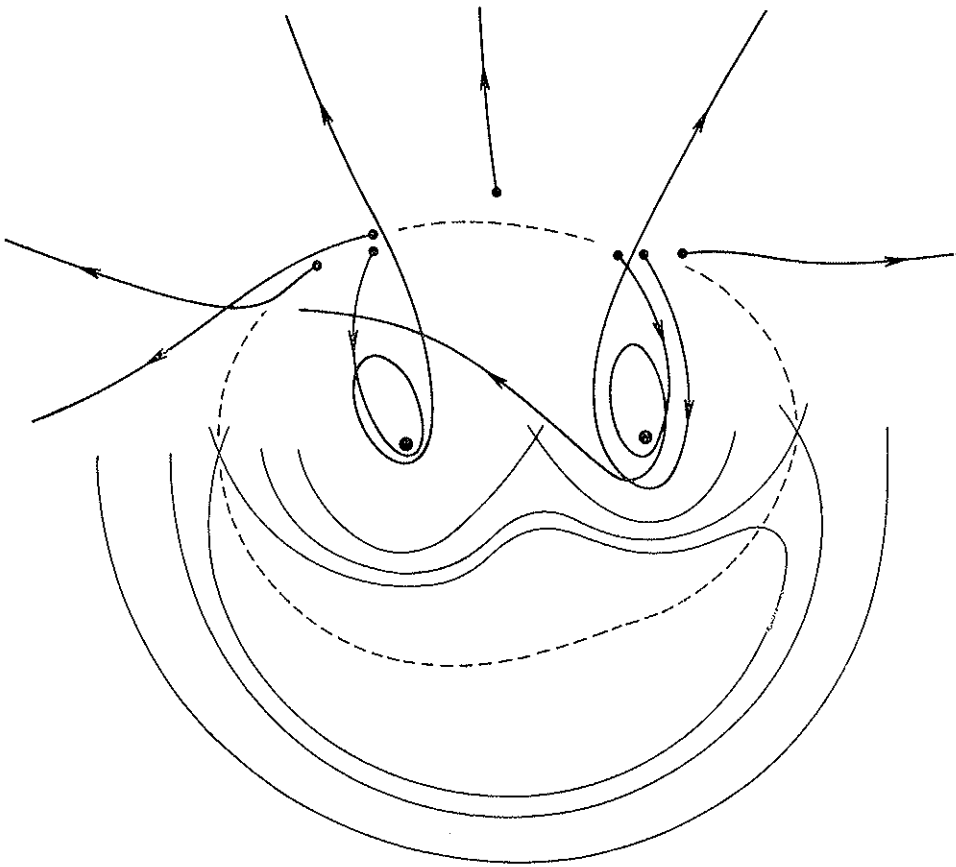


FIG. 5. The lower half of this figure plots the critical equipotentials of $-\psi$, the gravitational potential of the Local Group in comoving coordinates. The rampart or ridge line off which the galaxies will fall is the dotted line. The eyes are M31 on the left and the Galaxy on the right. Several orbits starting near the ridge line are plotted in the upper half. The two that fall in and later climb out do so only after the damping has changed to excitation.

that line is displaced when v is strongly negative the objects concerned are close to $v = 0$ where the displacement is negligible. Thus our 3 body computations so far have vindicated the very rough approximation made to get equations (8) and (9). However, when orbits suitable for the closer bodies IC 1613 and NGC 6822 are available, I expect these 3 body integrations will prove essential.

In conclusion, the dynamical age of the Local Group is in the age range that comes with small values of Hubble's constant, but the mass of

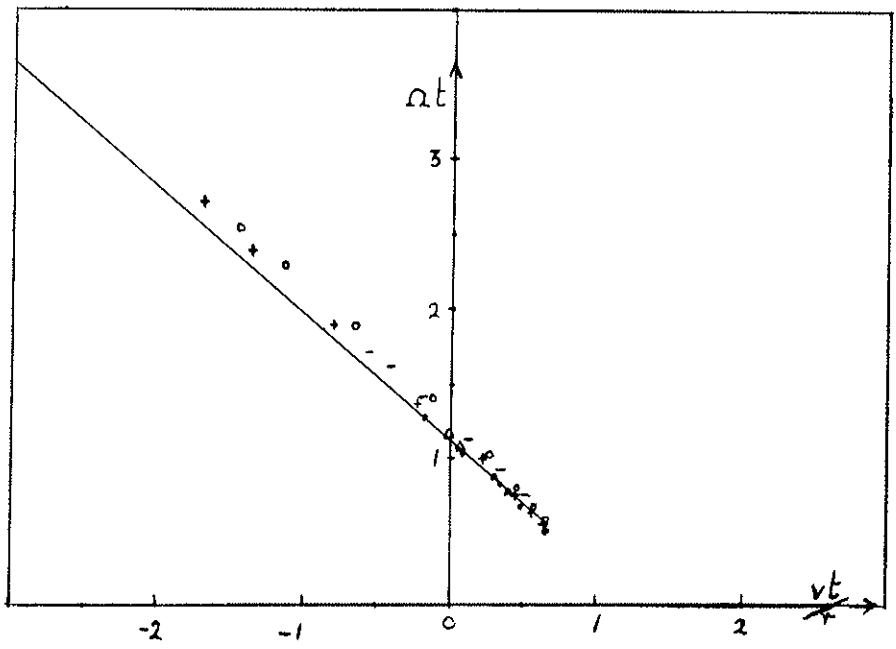


FIG. 6. $\Omega = \left(\frac{GM}{r^3}\right)^{1/2}$. The graph shows Ωt against vt/r for 5 different 3-body orbits that are appropriate for distant Local Group members. The line is the line of Figure 4 which was determined for straight line motion from a single point mass.

the Local Group, although very uncertain [it depends on $(r_0/r_\Lambda)^{-4}$ approximately], seems very large, 3.6×10^{12} $M_\odot$, with an uncertainty which must be at least a factor of 2. If we believe our earlier halo estimates of about 1×10^{12} $M_\odot$ for our Galaxy and 1.5×10^{12} $M_\odot$ for Andromeda, then it seems that the sum of the pieces of the Local Group add up to the total within the rather poor errors. Such a conclusion would leave no room for a halo of the heavy material around the Local Group. However, all we can deduce is that such a superhalo is not yet necessary. If we ask for limits on the mass of any such halo to the Local Group, then all we can conclude is that its total mass is probably not much more than the combined masses of the Galaxy and M31, assuming that their individual halos out to 100 kpc are included. The latter mass may be 2.5×10^{12} $M_\odot$.

All the above must be considered as speculative rather than established. There remains some doubt as to whether heavy halos to the Galaxy and

M31 exist at all. Likewise, it is not inconceivable that Andromeda and the Galaxy are independent objects passing each other in the night. However, if we suspend such doubts, then we are faced with vast amounts of invisible mass. Should we follow the fashion and try to assume that all invisible mass is in the form of massive neutrinos? Although this idea has been much explored and I see no evidence against it, nevertheless I do not see any good confirmatory evidence from astrophysics or from the laboratory. I am however still impressed by the fact that a natural minimum mass of .007 $M_\odot$ follows from fragmentation theory [26]. Theoretically, I do not see why most stellar objects are not of around that superplanet mass. As an astronomer I prefer this astronomical explanation of the missing mass as failed stars or superplanets and, if we believe one speculative theory of light variations in the double quasar, they may have been detected [27].

REFERENCES

[1] Woolley, R. and Savage, A., 1971, *Royal Obs. Bull.*, **170**, 365.

[2] Knapp, G.R., Tremaine, S.D. and Gunn, J.E., 1978, *Astron. J.*, **83**, 1585.

[3] Einasto, J., Haud, U.N. and Joeveer, M., 1979, IAU Symposium Na 84, The Large Scale Characteristics of the Galaxy, p. 231.

[4] Shuter, W.L.H., 1981, *Mon. Not. Roy. Astr. Soc.*, **194**, 851.

[5] Frenk, C.S. and White, S.D.M., 1980, *Mon. Not. Roy. Astr. Soc.*, **193**, 265.

[6] Lynden-Bell, D. and Frenk, C.S., 1981, *Observatory*, **101**, 200.

[7] Lynden-Bell, D. and Lin, D.N.C., 1977, *Mon. Not. Roy. Astr. Soc.*, **181**, 37.

[8] Lin, D.N.C. and Lynden-Bell, D., 1981, *Mon. Not. Roy. Astr. Soc.*, **198**, 707.

[9] Newton, K. and Emerson, D.T., 1977, *Mon. Not. Roy. Astr. Soc.*, **181**, 573.

[10] Murai, T. and Fujimoto, M., 1980, *Pub. Astron. Soc. Japan*, **32**, 581.

[11] Davies, R.D. and Wright, A.E., 1977, *Mon. Not. Roy. Astr. Soc.*, **180**, 71.

[12] Lin, D.N.C. and Lynden-Bell, D., 1977, *Mon. Not. Roy. Astr. Soc.*, **181**, 59.

[13] Wakamatsu, 1982, preprint.

[14] Petrou, M., 1981, *Mon. Not. Roy. Astr. Soc.*, **196**, 933.

[15] Hodge, P.W. and Michie, R., 1969, *Astron. J.*, **74**, 587.

[16] Hodge, P.W., 1964, *Astron. J.*, **69**, 438.

[17] Lynden-Bell, D., 1982, *Observatory*, **102**, 7.

[18] Lynden-Bell, D., 1976, *Mon. Not. Roy. Astr. Soc.*, **174**, 685.

[19] Schommer, R.A., Olszewski, E.W. and Kunkel, W.E., 1978, The H.R. Diagram I.A.U. Symposium 1980, Eds. A.G. Davis Philip and D.S. Hayes.

[20] Gott, J.R. and Thuan, T.X., 1978, *Astrophys. J.*, **223**, 426.

[21] Einasto, J. and Lynden-Bell, D., 1982, *Mon. Not. Roy. Astr. Soc.*, in preparation.

[22] Kraan-Kortweg, R.C. and Tammann, G.A., 1979, *Astr. Nach.*, **300**, 181.

[23] Yahil, A., Sandage, A.R. and Tammann, G.A., 1977, *Astrophys. J.*, **217**, 903.

[24] Lynden-Bell, D., 1981, *Observatory*, **101**, 111.

[25] Kahn, F.D. and Woltjer, L., 1959, *Astrophys. J.*, **130**, 705.

[26] Low, C. and Lynden-Bell, D., 1976, *Mon. Not. Roy. Astr. Soc.*, **176**, 367.

[27] Gott, J.R., 1981, *Astrophys. J.*, **243**, 142.

DISCUSSION

OSTRIKER

I believe that all of the satellites in the Local Group have some old stars in them including the Magellanic Clouds.

LYNDEN-BELL

Yes, but there are still some difficulties in making Ursa Minor and Draco pieces of the Large Magellanic Cloud. Why is Ursa Minor being torn up now as it approaches the Galaxy? If it had approached before it would have been torn up then. If it were torn off the Magellanic Clouds at the last close approach, then that was the time when we believe the gaseous stream was detached. Why is there no gas associated with Ursa Minor and Draco, whereas the rest of the stream is dominated by gas?

OORT

I have been impressed by a recent Japanese article in which it is suggested that the Magellanic Stream originates in a close encounter of the two clouds rather than by a recent passage of the clouds near to the Galaxy. In this picture the Clouds would move in a near-circular orbit. An attractive feature is that it would explain in a natural way why the stream lies so close to a great circle.

LYNDEN-BELL

Provided the two orbits are in the same plane, that is acceptable. The Small Cloud is now in the spin plane of the Large Cloud which is at right angles to the orbital plane of the Large Cloud about the Galaxy. My current belief is that the relative orbit of the two clouds is in that spin plane also, in which case that explanation would not work.

REES

Even if one needs to invoke a third body to explain the motions of us

and Andromeda, do you still think that the *spin* of Andromeda and of our Galaxy can be due simply to their mutual attraction?

LYNDEN-BELL

Yes, I am quite impressed that the direction of our Galaxy's motion adds to the spin angular momentum in about the same direction, so the old position of Andromeda could be that given by Gott and Thuan [20].

SETTI

If a galaxy has a dark halo how do we know the spin of the galaxy?

LYNDEN-BELL

We only know the spin of the visible part. Of course the bending of the galaxy could be caused by the halo having a different spin axis and the Magellanic Stream shows an axis perpendicular to the Galaxy's spin, but I think it likely that our halo has the same axis as the visible part, and likewise for Andromeda.

GUNN

Since the net displacement in angle due to the tidal spinning-up of M31 and the Galaxy seems to be small, a possible test is to verify that the planes of spiral pairs intersect in lines which pass close to their centres. This could be done statistically on those pairs, one member of which is edge-on, for example.

LYNDEN-BELL

I agree that this is an interesting test for objects whose period is greater than their age.

FABER

If the Local Group got its angular momentum from tidal torquing by other nearby groups, you might expect the value of its dimensionless parameter λ to be close to 0.07. Is the observed value close to this theoretical value?

LYNDEN-BELL

For a binary λ of 0.07 corresponds to an eccentricity of 0.995. The sort of orbit we advocate for Andromeda has an eccentricity of about 0.9 which corresponds to a λ of 0.3. However, an eccentricity of 1 is possibly consistent with the data.

WOLTJER

It is worth remarking that all the numbers one finds are extremely dependent on the assumed circular velocity of the Sun, which I think has still a substantial uncertainty.

LYNDEN-BELL

I believe the circular velocity is now stable at 220 ± 15 km sec^{-1}. Recent research on the globular cluster distribution has helped to stabilise this.

SILK

In the adiabatic pancake theory of galaxy formation, groups of galaxies form by a complex process including also fragmentation of larger systems. Tidal interactions between groups at the epoch of pancakes are a natural consequence of this theory and could account for the observed angular momentum of the Local Group.

LYNDEN-BELL

Such a net spin of everything in the Local Group might help us to understand why the orbital and spin angular momenta are all roughly aligned.

THE LARGE SCALE DISTRIBUTION OF GALAXIES

MARC DAVIS
Department of Astronomy
University of California, Berkeley

1 - Introduction

Even the most cursory examination of a map of the observed galaxy distribution convinces one that the distribution of galaxies is inhomogeneous and highly clustered. The study of galaxy clustering has proceeded steadily for over fifty years, and at an accelerated pace this past decade with the computer analysis of statistically complete catalogs. Much of the credit for this burst of activity is due to Jim Peebles, who exhorted his colleagues at Princeton and elsewhere on the virtues of the covariance function $\xi(r)$. These analyses were generally performed on catalogs of galaxies that were magnitude limited, and for which the only information available on each object was its magnitude and position on the sky. Thus to describe the phase space distribution only two out of six parameters for each object were known, with a third parameter, the distance, known only very roughly.

With the advance of technology it has become relatively simple to obtain redshifts of galaxies in considerable quantity, and now several statistically complete redshift surveys have become available. The redshift provides a precise third parameter for the phase space distribution that is a linear combination of the distance, as measured by the Hubble flow expansion, and a radial velocity caused by departures from a smooth Hubble flow. It is important to realize that this non-Hubble, or peculiar velocity component, can significantly distort the redshift space distribution from the true three dimensional distribution, particularly in rich clusters.

Recently Sandage (1978) has completed the redshifts of the Shapley-Ames catalog which is nearly complete to 13.0 m_B over the entire sky and contains some 1200 galaxies (see Sandage and Tammann 1981). This

sample probes to a distance equivalent to $\sim$ 4000 km s^{-1} (40 h^{-1} Mpc with $H_0 = 100$ h km s^{-1} Mpc^{-1}) and gives considerable detail on the local supercluster of Virgo. Fisher and Tully (1981) have completed a 21 cm survey of over 1100 mostly late type galaxies with radial velocities of less than 3000 km s^{-1}, and Tully (1981) has used this data set to make a very detailed study of the group structures within the local supercluster. These catalogs provide invaluable information on our immediate extragalactic neighborhood, but because galaxy clustering is so strong even on scales of 10 to 20 Mpc, we do not know if they probe deep enough to describe the "ensemble" distribution of clustering in the universe.

Several observational groups have obtained statistically complete redshift surveys of galaxies in selected small regions of the sky to limiting magnitude in the range of 15 to 17 m_B. These studies have greatly increased our awareness of the reality of superclusters of sizes up to 50 h^{-1} Mpc, accompanied by equally large holes in space that appear to be quite deficient in galaxies. Nearby rich Abell clusters (e.g. Coma, Perseus, Hercules) have been shown to be much more extended than previously thought (see e.g. Gregory and Thompson 1978; Tarenghi *et al.* 1979, 1980; Gregory, Thompson and Tifft 1981) and holes of up to 60 h^{-1} Mpc in extent have been discovered (Kirshner *et al.* 1981). The universe as seen in these redshift surveys appears extremely lumpy on scales comparable to the sample volumes, and again one cannot be certain that a "fair sample" is being measured.

In this talk I shall concentrate the discussion on a review of the recently completed Harvard-Smithsonian Center for Astrophysics (CfA) survey, which is a magnitude limited redshift survey complete to a Zwicky magnitude of 14.5. The catalog contains some 2400 galaxies in a solid angle of 2.7 steradian. This sample is a compromise in that it has large angular coverage and permits a detailed study of the galaxy distribution to a distance of roughly 80 h^{-1} Mpc, twice the distance of the Shapley-Ames sample but about half the distance of the deeper surveys which in total have surveyed less than 0.3 steradian of the sky. The CfA sample is divided into two pieces; the north galactic polar region $\delta \geqslant 0$, $b \geqslant 40°$; and a slice in the south galactic sky defined by $\delta \gtrsim -2.5°$, $b \leqslant -30°$. Its chief weakness is uncertainty in the Zwicky magnitude system on which it is based. Severe systematic errors with the Zwicky lists will effect our quantitative conclusions but not the qualitative picture of the overall distribution.

In section 2 below we describe the overall distribution in general terms and compare it to existing N-body simulations. Section 3 discusses

several dynamical studies that have been applied to the CfA sample, and Section 4 summarizes the challenge the data pose to theories of galaxy and cluster formation. Further details are given in Davis *et al.* (1982), Davis and Huchra (1982), and Press and Davis (1982).

2a - The galaxy distribution

In Figures 1, 2 and 3 are shown the results of the CfA survey in the northern sky. The circles are curves of constant galactic latitude at $b = 70°$, $50°$ and $30°$. Note the survey boundaries $\delta = 0°$, $b = 40°$. These maps each plot galaxies in a selected velocity range with $0 < v < 3000$, $3000 < v < 6000$, and $6000 < v < 10000$ for Figures 1, 2 and 3 respectively. Very few of our sample galaxies have an observed velocity outside this range.

Here and in all subsequent plots we shall partially volume limit the catalog to a distance of 4000 km s^{-1}, or $M \leqslant -18.5$, which deletes roughly one third of the galaxies in the northern sample, out of a total of 1874. In Figure 1 note that the Virgo cluster is prominent at $12.5^h, +12°$, as is the the Ursa Major cloud at 11.8^h, $+ 55°$ and the Leo cloud at 10.7^h, $+ 15°$. The local supercluster plane which runs between Virgo and Ursa Major is not as prominent, and the galaxy distribution is not symmetric above and below the plane. Details of the clustering and subclustering in the area have been known for some time [c.f. de Vaucouleurs (1975) for a full discussion]. It is apparent that the Virgo complex is not well described as being spherically symmetrical, so that one must exercise caution in using the infall of the galaxy toward Virgo as a dynamical mass probe.

In Figure 2, the galaxies are beyond the domain of the Virgo supercluster, and now one must realize that this sample is mostly magnitude limited so these objects are intrinsically brighter, on the average, than the galaxies in Figure 1. On this slice are a fair number of loose clusters of diameter 4 to 8 h^{-1} Mpc and conspicuous holes of similar size. One hole, for example, is situated directly behind the Virgo cluster core.

On the larger scale, the clusters seem to merge into a filamentary structure that one could connect across the entire sky. However, the observed distribution could also result from a random positioning of loose clouds. There is unquestionably strong clustering in this picture, but no evidence for prominent superclusters or pancakes. Much of the observed

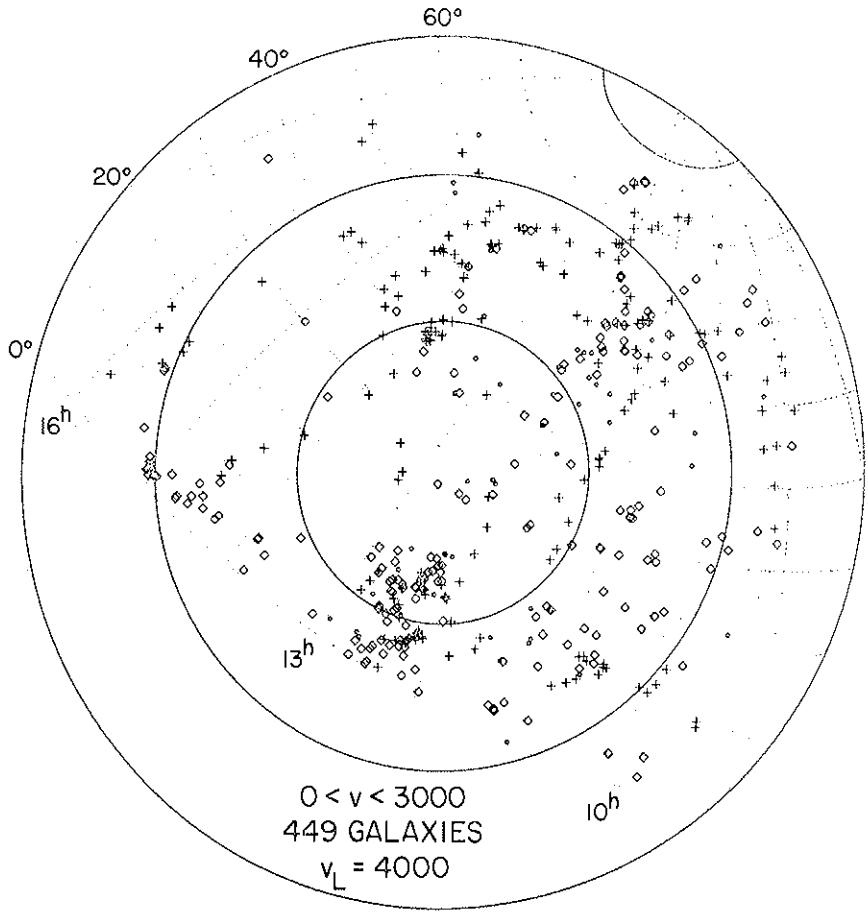


FIG. 1. Galaxies in the 14.5 magnitude limited CfA survey projected on the sky. This window shows galaxies with observed velocity $0 < v < 3000$ km s^{-1} and only galaxies luminous enough to be included to a distance equivalent to 4000 km s^{-1} ($M \leqslant -18.5$) are shown. The pentagons are galaxies with $0 < v < 1000$, diamonds have $1000 < v < 2000$ and pluses have $2000 < v < 3000$. The circles are curves of constant galactic latitude at $b = 70$, 50 & 30°. Lines of constant declination and right ascension are indicated.

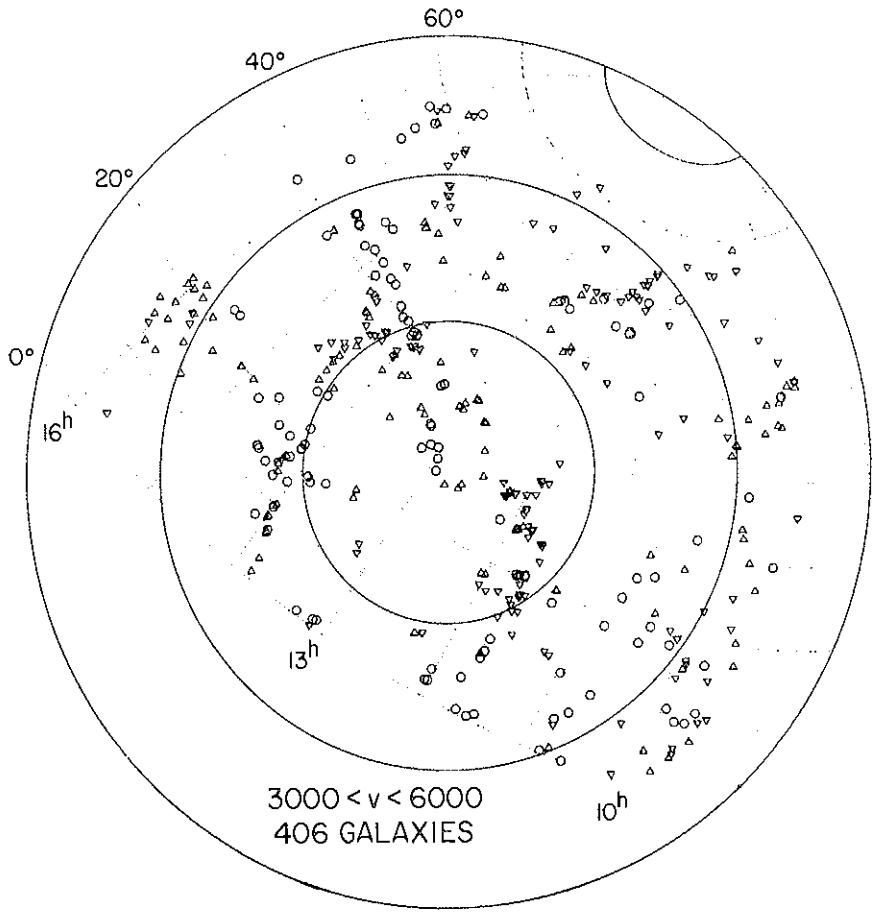


FIG. 2. Galaxies observed in the range $3000 < v < 6000$. The sample is again limited to objects with $M \leqslant -18.5$. The inverted triangles have $3000 < v < 4000$, the triangles have $4000 < v < 5000$, and the circles have $5000 < v < 6000$.

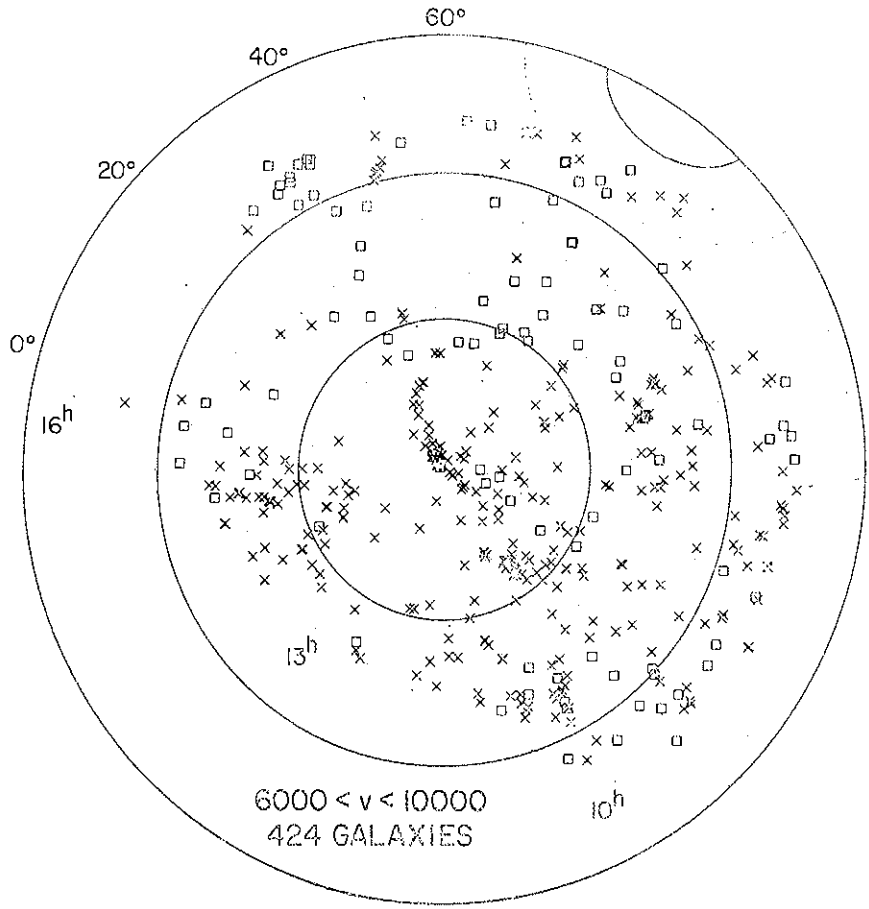


FIG. 3. Galaxies in the range $6000 < v < 10000$, and here the sample is totally magnitude limited. The x's are galaxies with $6000 < v < 8000$, and the boxes have $8000 < v < 10000$.

distribution seems truly random, and a substantial fraction of the galaxies cannot be associated with well defined supercluster structures.

The most distant slice of Figure 3 contains several prominent Abell clusters, such as Coma at 12.9^h, $+ 29°$, and A 1367 at 11.7^h, $+ 21°$. This slice is again magnitude limited and samples only galaxies on the exponentially declining portion of the luminosity function. Details of the clustering are, therefore, sampled more coarsely than in the less distant slices. The well known supercluster connecting Coma and A 1367 is ap-

parent, and extends 50° across the sky, or roughly 60 h^{-1} Mpc, in what appears to be a prolate, filamentary structure. There is one conspicuous hole centered roughly at 14.5^h, + 30°, with a diameter of 30°, or 35 h^{-1} Mpc in which the galaxy density is far below the mean. On the average over the entire solid angle, however, much of the large scale structure is random.

A transverse view of the galaxy distribution in the north is shown in Figures 4, 5 and 6 where right ascension versus observed velocity is plotted for various windows of declination. Here again all maps show a catalog that is volume limited to 4000 km s^{-1} and magnitude limited beyond, so that the falloff in galaxy density beyond 8000 km s^{-1} is a selection effect. The Virgo core is prominent in the foreground of Figures 4 and 5, and the elongation of the cluster in the redshift direction into a cigar shape is caused by virial motion within the cluster. Note how this effect is much less pronounced or even absent in the numerous small groupings. The loose, filamentary nature of the large scale clustering is equally

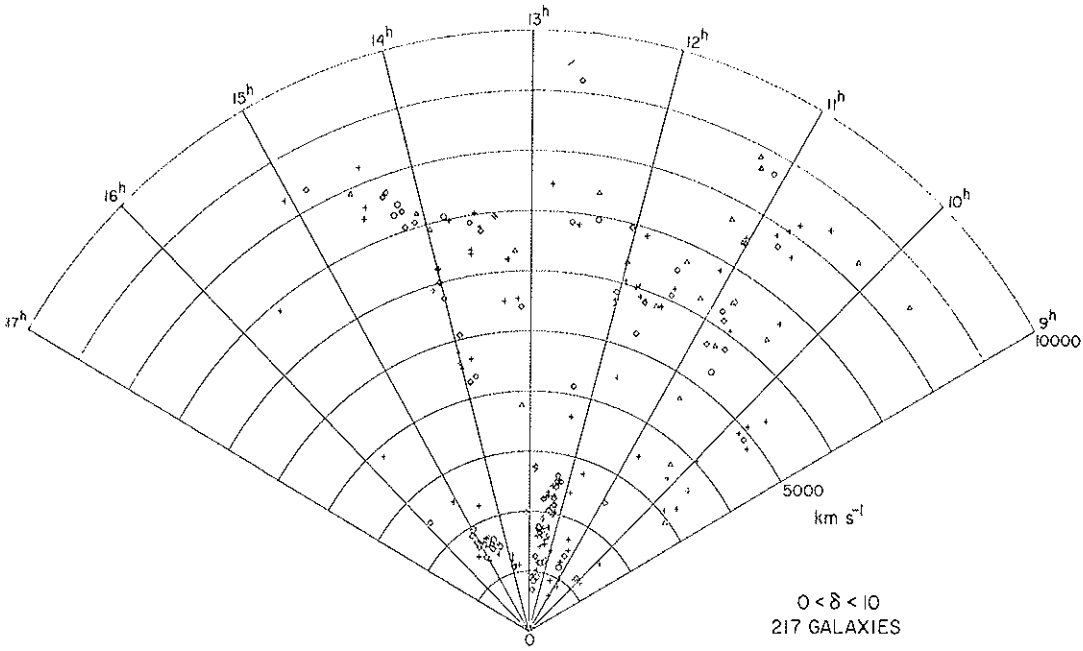


FIG. 4. A transverse view of the redshift space maps showing right ascension versus observed velocity for galaxies with $0 < \delta < 10°$. The pluses are spirals, the diamonds are SO's, the circles are ellipticals, and the triangles are irregulars and other galaxies. The sample is volume limited to 4000 km s^{-1} and magnitude limited beyond.

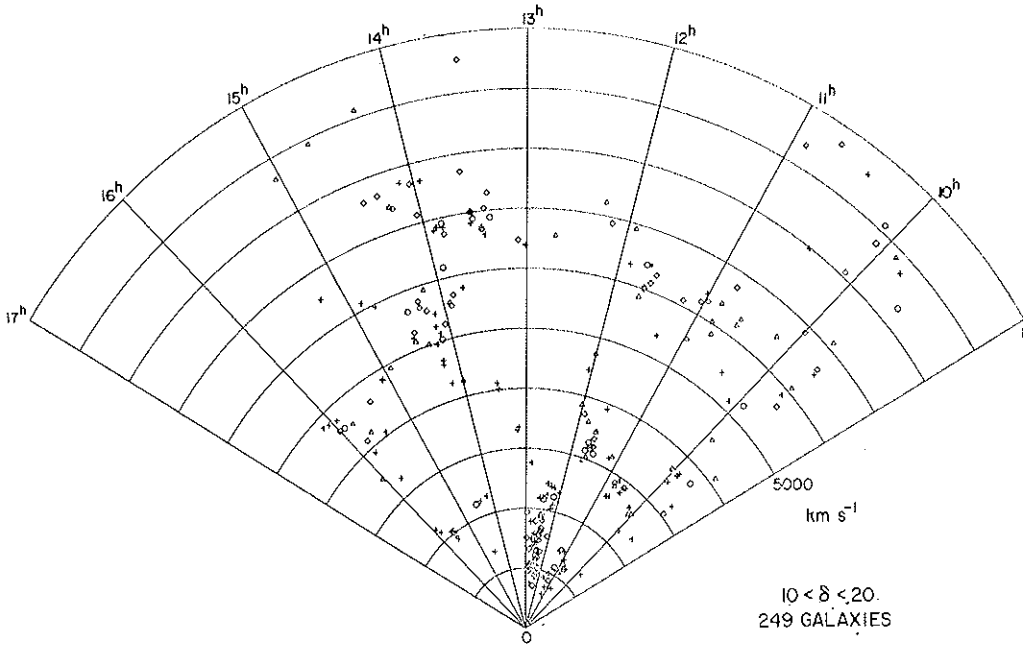


FIG. 5. Right ascension versus velocity for galaxies with $10^{\circ} < \delta < 20^{\circ}$.

pronounced in this series of figures as in the previous, orthogonal series of projections. Tifft (1980) has emphasized that for both orthogonal projections to appear one dimensional the true space distribution itself must be one dimensional. Thus the large scale structure of this sample is describable by the words "prolate", "chain" and "filamentary", and not by the words "oblate" and "pancake". This is not to deny that large scale two dimensional pancakes could not have formed in the past, only that evidence for their existence today is weak. I should hasten to add the word "random" to the list of adjectives for description of the large-scale structure. There is clearly structure but one must be careful not to overinterpret every low contrast fluctuation into a large-scale connection.

Figure 6 shows the Coma-A1367 supercluster quite prominently between 11^h-13^h and 6000-7000 km s^{-1}. The cigar elongation of the Coma core extends $\pm$ 1500 km s^{-1} and is more continuously connected when the survey limit is extended to 15.0 magnitude (Gregory and Thompson 1978). Note on this same figure that the hole at 14.5^h, 7000 km s^{-1} is quite empty, as is a smaller hole directly behind the core of Virgo.

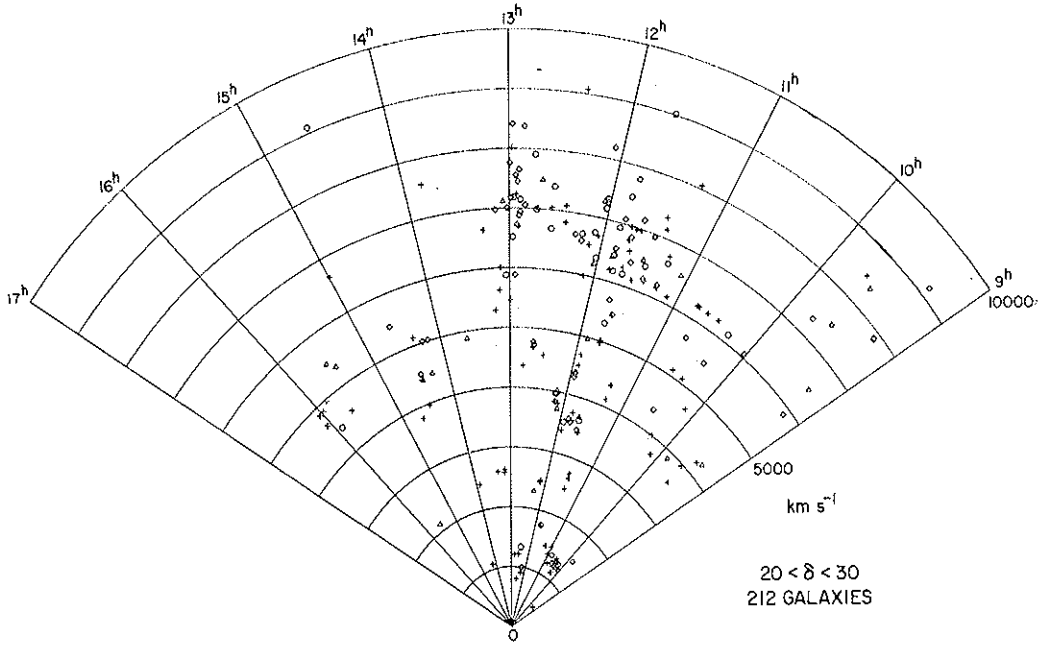


FIG. 6. Right ascension versus velocity for galaxies with $20^o < \delta < 30^o$.

The clustering as seen in the southern slice of the CfA survey shows very similar types of structure, with loose clouds connected in a filamentary fashion. Additional maps are given in Davis *et al.* (1982). There are no clusters in the southern slice as rich as those in the north, so the cigar elongations are even less conspicuous. In this section, the overall redshift distribution is very similar to that found by Gregory, Thompson and Tifft (1981) for the Perseus supercluster, which is outside the boundaries of our survey. There is a strong overdensity at a velocity of 5000 km s^{-1}, and a paucity of galaxies at low and high velocities. Professor Oort will discuss the observed superclusters in further detail in the next talk.

2b - COMPARISON TO SIMULATIONS

As a first stage beyond a qualitative description of the galaxy distribution, one can compare it to distributions resulting from N-body simulations where all the input physics and initial conditions are known. The best

simulations available to me so far have been the 20000 body simulations of Efstathiou and Eastwood (1981) which were run in an expanding cube with periodic boundary conditions and white noise initial conditions. One has the freedom to arbitrarily scale the simulation by a physical length, which was chosen so that the length scale r_o at which the covariance function, $\xi(r) = 1$, matched the observed value of 5 h^{-1} Mpc. Fortunately this resulted in a near match to the observations of an important dimensionless quantity: the number of particles (or galaxies) in one clustering (or coherence) length. This simulation was run for a cosmological density $\Omega = 1$. Existing simulations for $\Omega < 1$ unfortunately do not have sufficient objects per coherence length to compare to the observations.

Each point in the simulation is assigned a luminosity randomly drawn from the luminosity distribution function observed in the northern sample, and only points with apparent $m_B < 14.5$ are retained. The minimum luminosity of a point is $M = -18.5$, so the sample will again be volume limited to a distance of 4000 km s^{-1} and magnitude limited beyond. For the simulations we of course know all six degrees of freedom for each point, so we can generate an observable velocity $v_0 = H_0 \ell + K v_p$, where ℓ is the distance to the object, v_p is the component of random thermal motion of the object projected in the line of sight, and K is a scale factor by which we fudge the cosmological density. To be self consistent in this $\Omega = 1$ simulation, we should use $K = 1$, but this would produce too much elongation of the clusters compared to the observed distribution. We set $K = 1/3$, which corresponds roughly to $\Omega \simeq K^2 = 0.09$. This is not a self-consistent approach but is the best one can do without better simulations.

The resulting catalog when projected onto the sky is shown in Figure 7. Here we plot all objects with $0 < v < 10000$ km s^{-1}. Note how the sample boundaries have been chosen to match the northern CfA survey. Transverse views of the distribution are shown in Figures 8 and 9. Although the space density of objects and correlation length of the clustering is roughly the same in the samples, there are obvious glaring differences between the N-body catalog and the observations.

The simulation is characterized by very prominent dense cluster centers with pronounced elongation in redshift space, but no large-scale connectedness among the clusters. The size and distribution of holes in redshift space is similar in the two samples. If K were further reduced, the elongation of the clusters would shrink but the cluster centers would be too dense.

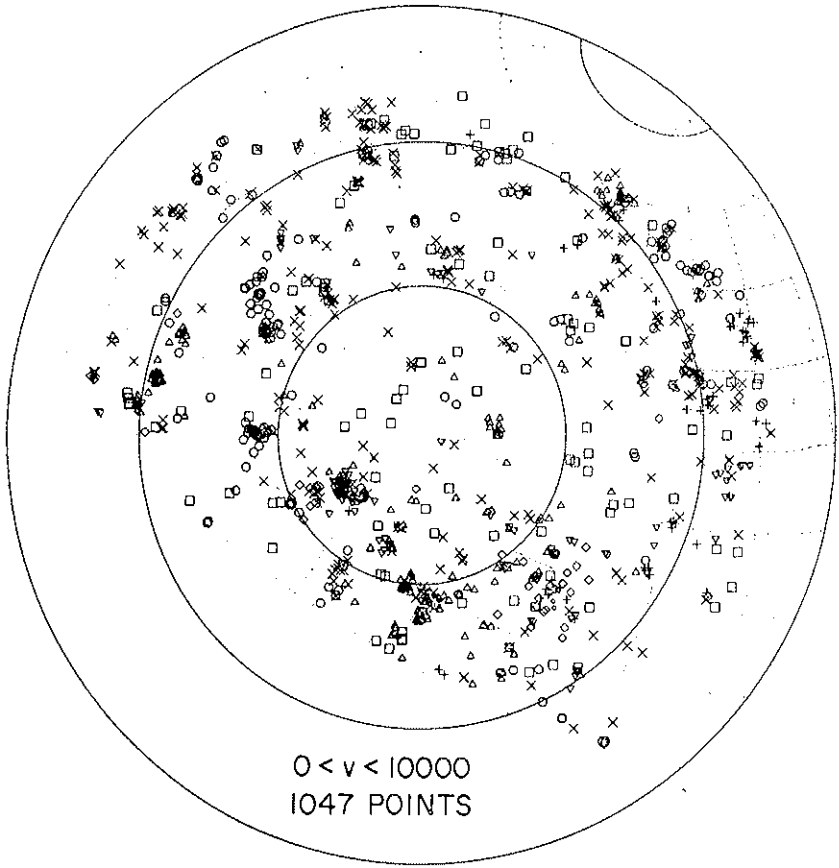


FIG. 7. A sky projection of the N-body simulation of Efstathiou and Eastwood (1981).

The obvious conclusion one could reach here is that this is not the way the universe is constructed. The physics of the simulations includes only dissipationless gravity where clusters form by the standard gravitational instability process. Should some form of dissipation and/or nongravitational physics be included? The points carry all the mass in the simulation, but we know galaxies have extensive halos which dominate the mass density of the universe. We assume the galaxy distribution is a fair tracer of the underlying mass distribution, but what if it is not? Finally, what initial conditions did the universe have? White noise is the easiest to simulate in the computer but nature is not compelled to be so simple.

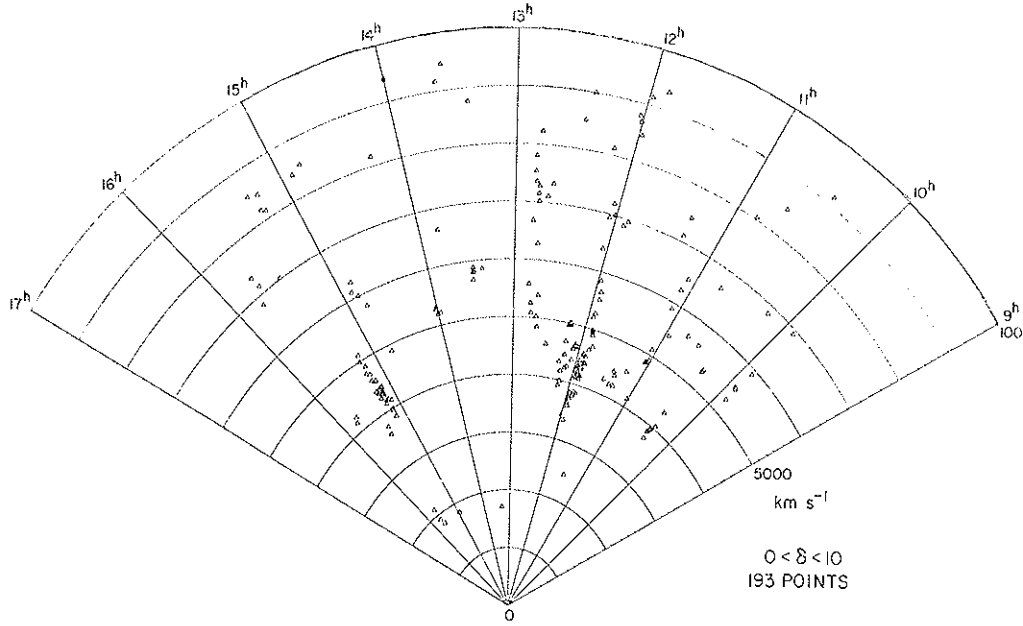


FIG. 8. Right ascension versus velocity for points in the N-body simulation with $0^\circ < \delta < 10^\circ$.

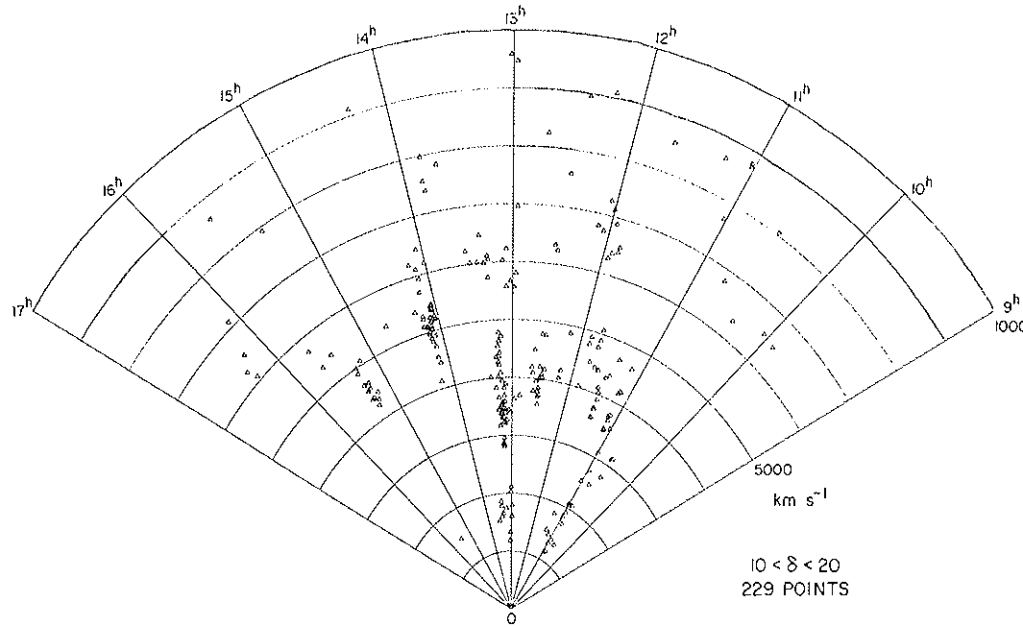


FIG. 9. Right ascension versus velocity for points in the N-body simulation with $10^\circ < \delta < 20^\circ$.

2c - The density field and $\xi(r)$

As an aid to visualizing the large-scale clustering, we can average the galaxy distribution across the sky and plot the observed radial velocity distribution of galaxies. Based on the luminosity distribution function for the catalog, which can be derived in a fashion unbiased by spatial inhomogeneities, one can quite simply predict the expected radial velocity distribution in a homogeneous universe. For the simulation the two curves follow each other quite well, although local clustering produces larger than Poisson fluctuations in the observed radial counts. This is hardly surprising since the scale length of clustering in the simulation is quite small compared to the extent of the volume, and the simulation is spatially homogeneous in the mean.

When similar diagrams are drawn for the observations, the large scale spatial inhomogeneities are strikingly apparent. Figures 10 and 11 show the observed radial distribution of galaxies in the CfA north and south samples, respectively, with the galaxies binned into radial velocity increments of 200 km s^{-1}. The heavy line is the observed distribution and the light smooth curve is the distribution expected in a homogeneous universe with the luminosity function as observed in each sample. The dashed line is the ratio of the observed to expected distribution and represents the density fluctuation at each radius. The horizontal dashed line is the mean density volume averaged over each sample. Note in the north how the Virgo supercluster causes the entire foreground zone to be overdense by about a factor of 3. Note also that a prominent hole of roughly half the mean background density extends from 3000 to 6000 km s^{-1}. In the south the situation is exactly reversed, with a pronounced overdensity at 4000 to 6000 km s^{-1} and holes in the foreground and background. There is thus considerable power in fluctuations of a scale of 30 h^{-1} Mpc. Perhaps surprisingly, however, the mean density level of the north and south CfA samples agree within 10%, which certainly is not the case for shallower samples. This could be evidence that these volumes define a "fair sample" of the universe, although this is by no means certain. The mean background levels are drawn as the horizontal dashed lines in Figures 10 and 11 and serve to normalize each figure.

There is considerable interest in the covariance function $\xi(r)$ and in its large-scale behavior. It is generally observed to have a power law behavior on small scales, $\xi(r) \propto r^{-\gamma}$ with $\gamma \sim 1.8$, but the large-scale behavior has been extremely difficult to measure reliably. Jim Peebles

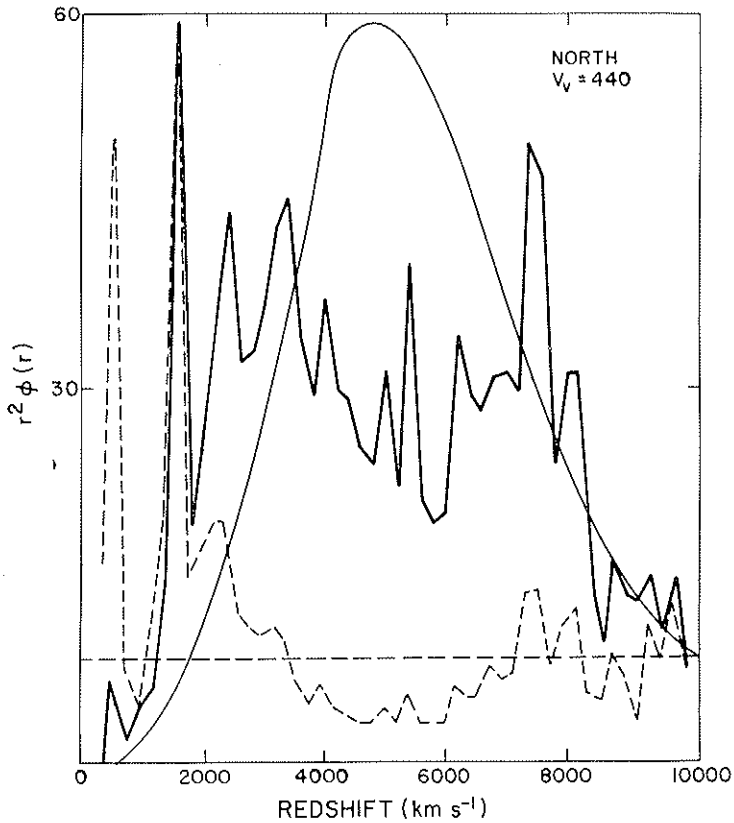


FIG. 10. The observed radial distribution of galaxies in the northern sky. The heavy solid line is the observed distribution and the light solid line is the distribution expected in a homogeneous universe. The dashed curve is the ratio of the two and is a measure of the fluctuation from the mean at a given radius. The horizontal line is the mean background level.

later today will discuss the importance of this measure. It is expected that $\xi(r)$ should break from a power law behavior at $r \sim 10$ to 20 h^{-1} Mpc and could be negative on larger scales. In order to study the power law nature of $\xi(r)$, it is more convenient to examine $r^2\xi(r)$ which is a slowly varying function in the power law domain. Figure 12 is a plot of a determination of $r^2\xi(r)$ for the northern sample. The solid lines are for the northern sample using two subsets of data that are volume limited to 4000 and 6000 km s^{-1} and magnitude limited beyond. The difference in the curves results from an effective

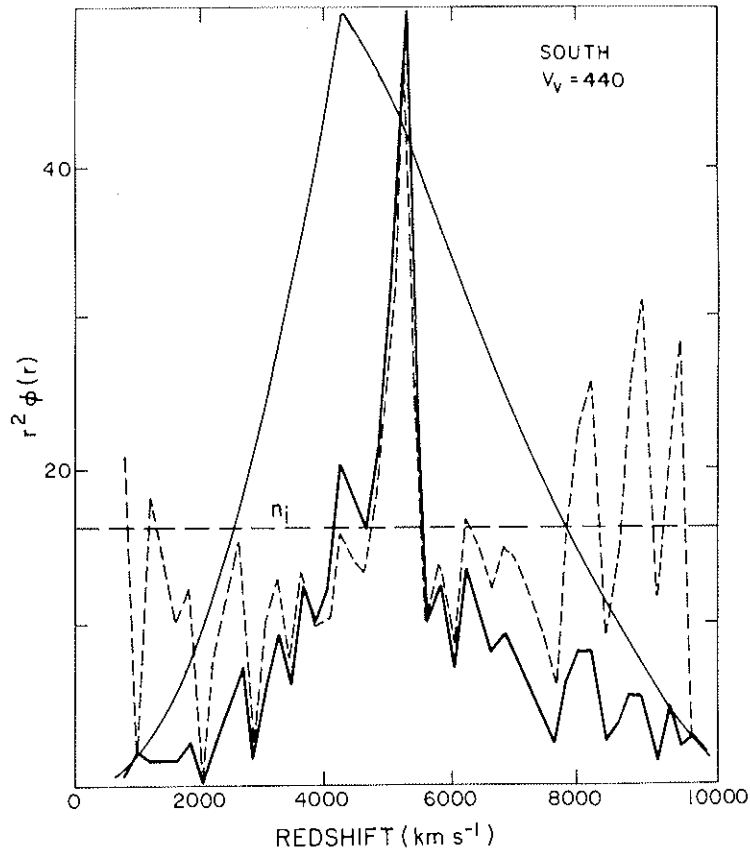


FIG. 11. The same as figure 10, now for the southern CfA sample.

difference in the weighting of nearby versus distant clusters and is a measure of the uncertainty of $\xi(r)$ within this volume. A power law of $r^{0.2}$ is indicated by the dashed line, and we see that $\xi(r)$ does have a power law behavior up to a scale of 5 to 10 h^{-1} Mpc, beyond which it steepens considerahly. Unfortunately the steepening of $\xi(r)$ occurs when $\xi(r) \sim 0.1$ to 0.2, where it is difficult to measure because of uncertainties at the 10% level of the appropriate background level. The break and all larger scale behavior is therefore quite unreliably determined, but it does appear as though $\xi(r)$ is slightly negative on scales of 15 to 40 h^{-1} Mpc. This is not surprising given the pattern of positive and negative fluctuations seen in

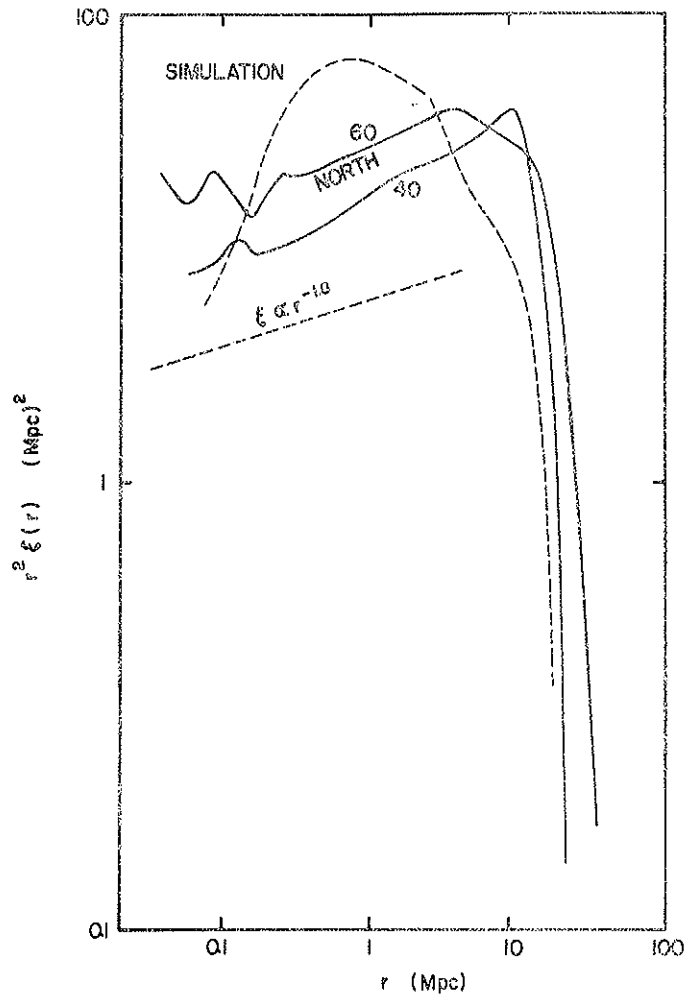


FIG. 12. The solid lines are curves of $r^2 \xi(r)$ in the northern CfA sample when the catalog is volume limited to 40 h^{-1} Mpc and 60 h^{-1} Mpc and magnitude limited beyond. The curves fit a power law for small scales and there is an apparent break at $\sim 10\ h^{-1}$ Mpc. The dashed line shows the N-body covariance, $r^2 \xi(r)$, which never behaves as a power law.

the radial distribution (Figure 10) with a coherence length of $\sim 30\ h^{-1}$ Mpc. I must hasten to add that the expected fluctuation of $\xi(r)$ is substantial on a large scale and therefore we should not be overly concerned if different experiments in separate volumes of space lead to diverging conclusions.

For comparison, the dashed curve of Figure 12 shows $r^2\,\xi(r)$ for the simulation, and here we see that the curve is continuously steepening so that the covariance function never behaves as a power law. Again there is relatively more power in the simulations on a smaller scale than in the observations, and less power on larger scales. This of course is the same impression given by examining the redshift space maps. The details of the covariance studies will be reported elsewhere (Davis and Peebles 1982).

3 - Dynamical studies

The many groups apparent in the redshift space maps are presumably dynamically interacting and at least some of them should be virialized. Press and Davis (1982) have analyzed the CfA sample with a group selection procedure that seeks groups of crossing time less than 1/3 the Hubble time; only for these groups can the virial theorem be expected to apply. Roughly half the galaxies are in clusters that are "weighable" by this technique, but groups of all sizes are measurable. Extensive tests have been performed on the N-body simulations to calibrate this procedure and no serious biases appear.

When the analysis is performed on the CfA sample, we find a linear trend of virial mass per galaxy versus the size of the system, as shown in Figure 13. Each point is the result of one group and the large scatter is expected. It is not much larger than the scatter seen in the N-body calibration for which all points have equal mass by construction. The ragged heavy line is a running median curve of the eight nearest clusters; a median is better than a mean for this purpose because some of the groups could be totally spurious and could easily bias a mean value. The dashed line is a fitted trendline

$$M_* = 3.24 \times 10^{12} \frac{R^{1.01}}{1\ \text{Mpc}}\ M_\odot\ .$$

This trend is totally absent in the N-body simulations. The short horizontal bar in the upper right is the mass per bright galaxy in an $\Omega = 1$ universe, $3.2 \times 10^{13}\ M_\odot$ for the luminosity density observed in the sample.

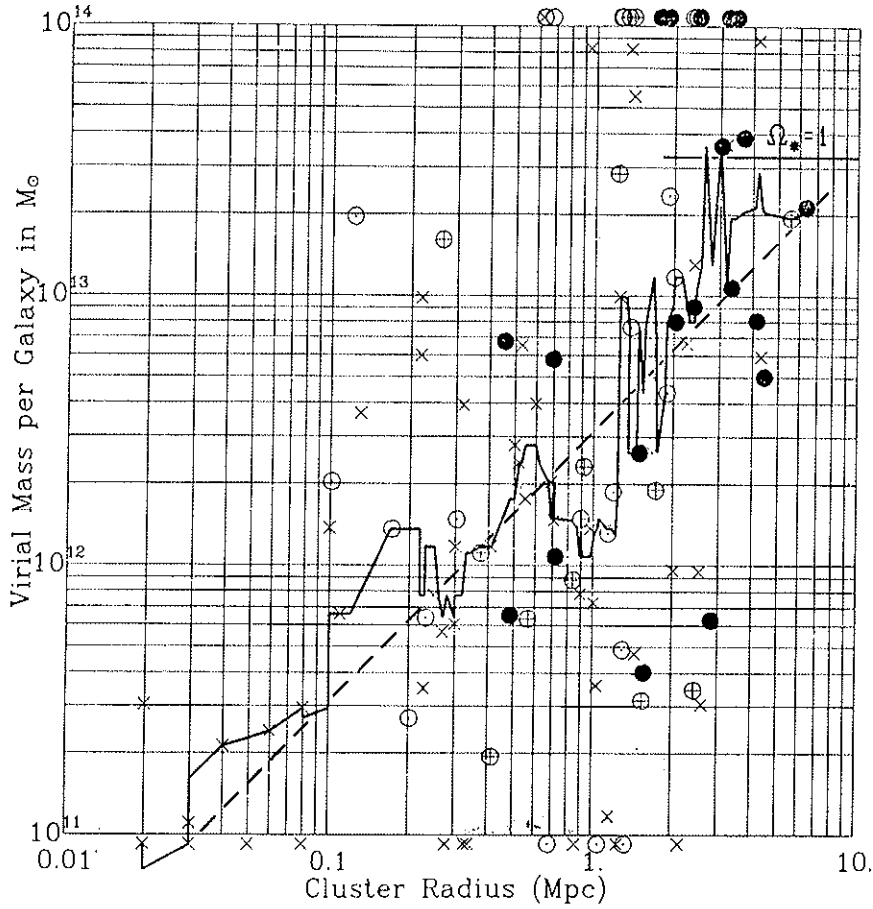


FIG. 13. Mass per bright galaxy versus size of the cluster in which it is contained. The scatter is large but the linear trend and running median appear significant. The $\Omega = 1$ level is indicated.

Our result is certainly not new (see Faber and Gallagher 1979; Einasto, Kaasik and Saar 1974; Ostriker, Peebles and Yahil 1974) but it results from a unified treatment of a uniform data set. Adding up the mass of the weighable galaxies gives a cosmological density $\Omega \simeq 0.07$. If the unweighable galaxies have similar mass, then we have $\Omega = 0.15$. However, before one can claim to have actually measured a proper value for the cosmological density, we must first understand how to interpret the observed trend of mass versus scale size. On a scale of individual halos,

the interpretation is that the dark gravitating matter is not as tightly clustered as the luminous matter. Should we really believe the same effect continues all the way up to scales over $2\ h^{-1}$ Mpc, in which case a "true" cosmological density can be measured only on the largest scales and might be as large as $\Omega = 0.6$? Perhaps, alternatively, there is some effect causing the trend to be spurious and only the smaller groups are reliable.

Fortunately the Virgo supercluster itself provides another dynamical test on the scale of our distance to Virgo, $\sim$ 15 Mpc, a considerably larger scale than can be studied by virial analysis. The Virgo supercluster is unique because it is close enough for astronomers to measure distances and redshifts separately for a large sample of galaxies, and therefore to map the velocity field induced by the overdensity of the cluster. Spherically symmetric models provide a first approximation for the supercluster that is certainly not precise, but do allow an exact treatment of the non-linear problem. These models have been extensively discussed in the literature (Peebles 1976; Silk 1974, 1977; Gunn 1978; Tonry 1980).

The expected velocity contours of the spherical model are shown in Figure 14, where the spherically symmetric infall pattern is superposed on a uniform Hubble expansion. The extent of the spider pattern around Virgo depends on the amplitude of the velocity field, here assumed to be 400 km s^{-1} at the distance of the local group and to fall inversely with Virgocentric distance. Note that for angles close to the Virgo cluster it is possible to observe the same redshift for three different distances, and galaxies in the triple valued zone cannot be unambiguously assigned unless the distance indicator is more precise than any available today. This is also the region where the infall is likely to depart substantially from the spherical model, so in any event one must exercise caution.

To use the Virgo infall as a probe of the cosmological density, it is necessary to measure the infall velocity and mean overdensity, $\bar{\delta}$, within a given radius, for example the local group radius. The large-scale distribution of the CfA redshift provides a reasonably precise estimate of the mean background density, and when compared with the density of galaxies within one Virgocentric radius, we find $\bar{\delta} = 2.0 \pm 0.2$, independent of the Virgocentric infall (Davis and Huchra 1982). This number is slightly lower than the Yahil, Sandage and Tammann (1980) estimate because their background density was determined from the shallower Shapley-Ames catalog in regions now known to be underdense from the mean. Of course we presume that the measured overdensity of galaxy counts or luminosity density corresponds to the same mass density contrast, i.e., that on this scale the

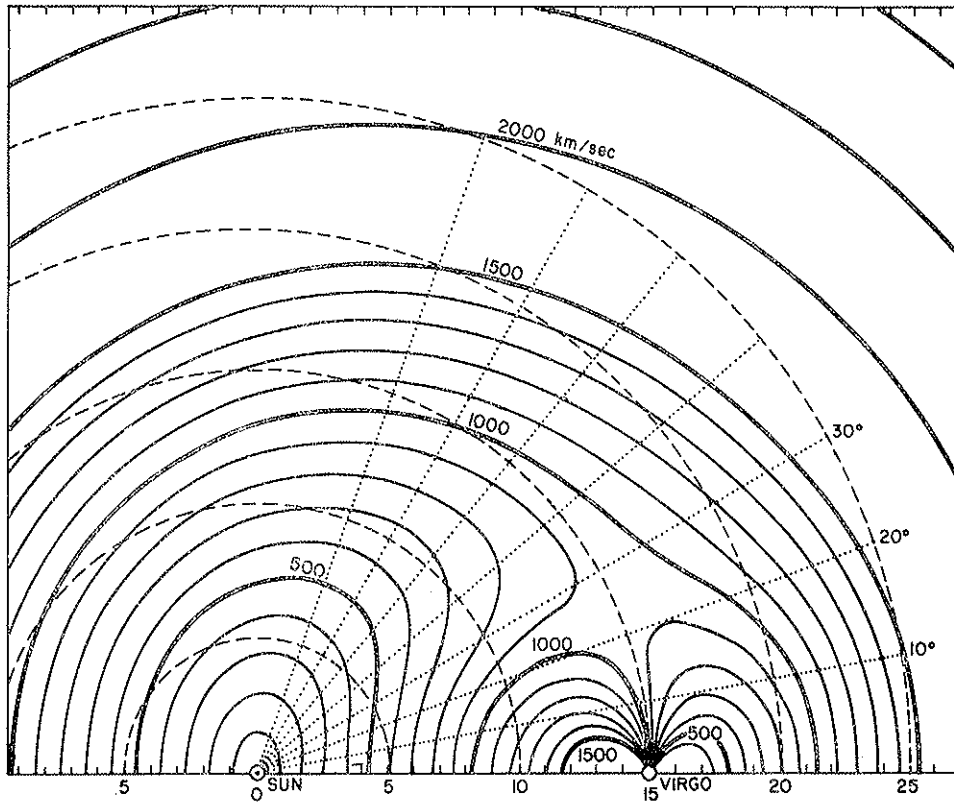


FIG. 14. Curves of constant observed velocity is the presence of a Virgocentric velocity field of amplitude 400 km s^{-1} at our radius. Note the triple valued solutions for angles less than 20° from Virgo.

galaxies are a fair tracer of the mass. This presumption has always failed us on the smaller scales.

The measurement of the infall velocity is considerably more difficult simply because of the usual uncertainties of distance estimation with less than perfect standard candles. Considerable effort has gone into this measurement, and the most recent results are summarized in Table 1. The microwave background anisotropy (Boughn, Cheng and Wilkinson 1981; Smoot and Lubin 1979) is presumably a measure of our motion relative to the comoving frame of the universe, and need not have anything to do with the Virgo cluster. In any event there is a large velocity component trans-

TABLE 1 - *Virgocentric infall velocity* (Local Group Center).

Group	Method	Measured velocity to Virgo (km s^{-1})	Distance of background
Boughn *et al.* (1981)	μ-wave dipole quadrupole	411 ± 36	$z \simeq 1000$
Smoot & Lubin (1979)	μ-wave dipole	373 ± 25	$z \simeq 1000$
de Vaucouleurs & Peters (1981)	optical tertiary indicators	240 ± 60	$\leqslant 30\ h^{-1}$ Mpc (dipole fit only)
de Vaucouleurs *et al.* (1981)	Br-HI correlation	240 ± 40	$\leqslant 30\ h^{-1}$ Mpc (dipole fit only)
Yahil *et al.* (1981)	optical luminosity	230 ± 75	$\leqslant 30\ h^{-1}$ Mpc
Aaronson *et al.* (1981)	IR-HI	313 ± 40	$\leqslant 30\ h^{-1}$ Mpc
Aaronson *et al.* (1980)	IR-HI clusters	520 ± 75	$\leqslant 60\ h^{-1}$ Mpc
Tonry & Davis (1981)	L-σ E-SO galaxies	470 ± 75	$\leqslant 60\ h^{-1}$ Mpc

verse to the Virgo cluster that is quite difficult to explain as being induced by Virgo infall.

The de Vaucouleurs *et al.* (1981) results are derived from a simple dipole fit of our motion relative to a frame of galaxies within roughly 30 h^{-1} Mpc of us, and are based on tertiary distance indicators and the Tully-Fisher correlation using Br magnitudes for ~200 to 300 spiral galaxies. The Yahil, Sandage, and Tammann (1980, 1981) result is derived from analysis of the Shapley-Ames luminosity distribution and fits a full infall model based on the shear field, again within ~ 30 h^{-1} Mpc. The Aaronson *et al.* (1981) result is based on infra-red magnitude 21 cm width correlation for 300 spiral galaxies within 30 h^{-1} Mpc, and again fits a spherically symmetric infall model.

These four determinations all seem consistent within themselves but do not agree with the microwave result or with measurements of the shear field on larger scales. Aaronson *et al.* (1980) found a local group infall of

520 km s^{-1} based on the IR-21 cm Tully-Fisher correlation for Virgo galaxies versus galaxies in four clusters at distances of $\sim$ 5000 km s^{-1}. Further data on these and more clusters, including Hercules at $\sim$ 10000 km s^{-1}, are consistent with their published result (Huchra 1981).

Using the correlation of luminosity and velocity dispersion for elliptical galaxies distributed to distances of 70 h^{-1} Mpc Tonry and Davis (1981) find a Virgo infall of 470 km s^{-1}. As an example of how these tests proceed, Figure 15 from Tonry and Davis (1981) shows the distribution

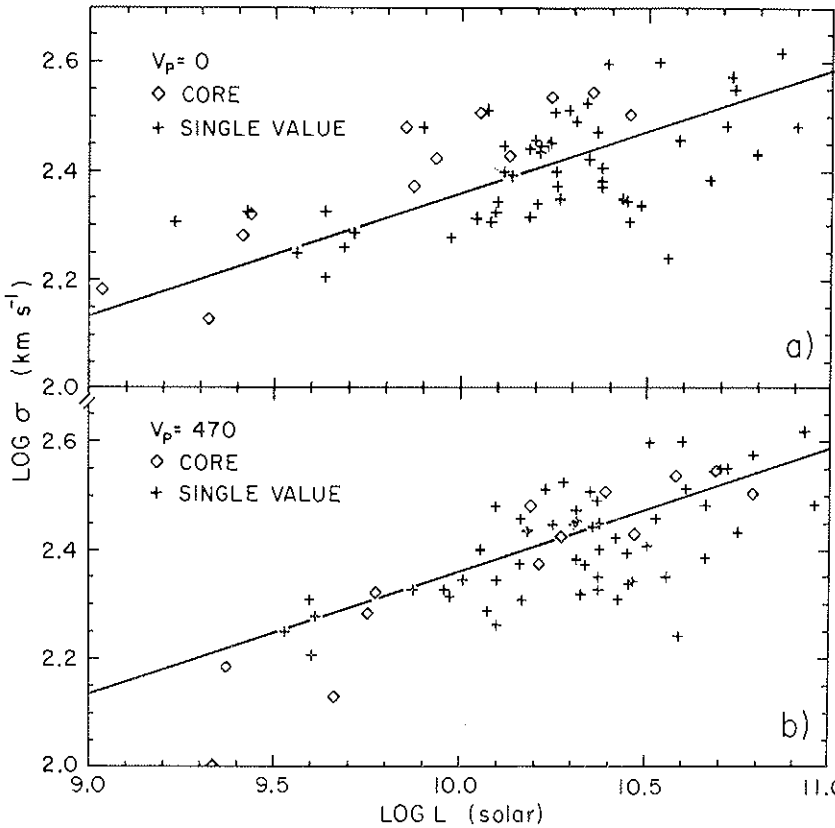


FIG. 15. *a*) Luminosity versus velocity dispersion for elliptical galaxies in the core of Virgo (diamonds) and in the single valued zone (pluses) away from Virgo. No infall velocity is assumed and it is apparent that the Virgo galaxies have too large a dispersion for their luminosity relative to the other ellipticals; *b*) same as *a*), but with a velocity field 470 km s^{-1} at the local group radius.

of luminosity versus dispersion for spherically symmetric models with and without Virgocentric infall. In the absence of infall, all the galaxies in the Virgo core (diamonds) are seen to he underluminous for their dispersion compared to galaxies in the single valued zone (+ signs), but with an infall model adjusted to 470 km s^{-1}, the distributions are fully commensurate.

Thus both early and late type galaxies suggest a higher value of the infall when the velocity shear field is studied on a 60 h^{-1} Mpc scale. These last two tests basically compare galaxies in the Virgo core relative to distant galaxies unaffected by the infall, and are not sensitive to details of the infall model; they only measure our peculiar motion to Virgo. As such they are quite sensitive to the mean redshift of Virgo.

For infall values of 250 to 500 km s^{-1}, the cosmological density in the spherical model for $\bar{\delta} = 2.0$ is in the range $0.2 < \Omega < 0.5$ (Davis *et al.* 1980). Tbis is a very substantial density, considerably higher than the Ω measured in groups, but below the trend of Figure 13. This is an important result and it is important to examine the assumptions and approximations used to derive it. There are at least 4 such major assumptions, none of which are true in detail.

1) Spherical symmetry with no subclustering has been assumed, and this is obviously incorrect as Virgo is highly irregular and subclustered, although its central concentration is sufficient to make the spherical approximation a good initial guess.

2) The background is assumed uniform, which we know is incorrect, although the effects of the substantial holes in the south and beyond Virgo in the north will mostly cancel.

3) The galaxy distribution is assumed to be a tracer of the underlying mass distribution on this scale. If this is incorrect and the mass distribution is not as highly clumped as the galaxy distribution, then the density parameter Ω will be underestimated.

4) The measured infall velocity to Virgo is assumed to be caused by Virgo's excess gravity. Nongravitational origins of peculiar velocities can arise in models such as Ostriker and Cowie's (1981) coordinated supernovae model of galaxy formation and, if this process dominates, then obviously no constraint on Ω will result from the Virgo infall test.

It is possible to relax assumptions (1) and (2), the two weakest approximations, by using linear perturbation theory, where the expected peculiar velocity today is proportional to the peculiar gravity g measured today. Although the Virgo supercluster is a non-linear fluctuation, linear

perturbation theory should apply fairly accurately as long as the system has not turned around and begun to recollapse. With the distribution of galaxies in the CfA survey we can compute the peculiar gravity on a shell by shell basis, each shell weighted equally, out to a distance of $\sim 80\ h^{-1}$ Mpc, beyond which the selection function of galaxies is too small. The Virgo cluster causes most of the effect, but the hole behind Virgo reduces the amplitude slightly. In the southern sky we use the Shapley-Ames catalog, which is complete at high galactic latitudes, but because it is limited at 13.0^m, we can probe only to $\sim 40\ h^{-1}$ Mpc. The south is also underdense and effectively pushes us northward to Virgo. We are assuming that no peculiar gravity is generated by regions outside the solid angle or beyond the distance limits of the survey. Fortunately Virgo is at high galactic latitude where the data is best, and no comparable nearby clusters are known to exist at low galactic latitude. After summing the effect of the shells as described we find that the expected peculiar velocity of the local group is $670\ \Omega^{0.6}$ km s^{-1} and should be directed $\sim 20°$ to the north and west of Virgo (Davis and Huchra 1982). This number should be compared to the values of Table 1 and, in the range of 250 to 500 km s^{-1} for the measured infall, we derive $0.2<\Omega<0.6$. To account for the non-linearity of the Virgo cluster we should increase Ω by roughly 0.1. Thus the linear theory calculation confirms the rather high cosmological density of the spherically symmetric model and yet is not nearly as model dependent. The derived density is seen to be a very sensitive function of the infall velocity, and future work should seek to reconcile the present discrepancies in the derived values. At least the present discrepancy has been reduced to a factor of 2, which is a considerable improvement in the last few years.

4 - Problems for contemplation

The observations of large-scale structure are now of sufficient statistical reproducibility and uniformity to suggest that we have a fair picture of the present appearance of the large-scale universe. These observations provide a severe challenge to all theories of galaxy and cluster formation and will hopefully motivate further theoretical effort on this fundamental problem. Neither the gravitational instability scenario favored by many in the West (Peebles 1980) nor the Zeldovich pancake scenario favored by our Soviet colleagues (Zeldovich 1978; Zeldovich and Novikov 1975) are easily reconciled to the present observations.

The gravitational instability scenario in its usual guise presumes that, beyond the scale length of individual galaxies, only gravity is important for determining the larger-scale structure. Galaxies formed first and clusters formed from individual galaxies, each acting more or less like a point mass. The N-body simulations contain all the relevant physical processes in this scenario, and yet they fail to realistically mimic the observations. A major problem is the difficulty of evacuating large holes in space without creating large velocities in the field and in clusters. The expected velocity V_0, if holes of radius r_0 are typical, is (Davis *et al.* 1982)

$$V_0 \sim H_0 r_0 \Omega^{0.6} \sim 400 \text{ to } 1000 \text{ km s}^{-1}$$

assuming r_0 to be $\sim 10\ h^{-1}$ Mpc. The holes in the N-body simulation are as large as in the observations, and yet the typical velocities are much higher, particularly after removing the fudge factor of 3 applied to the data. This result certainly demands explanation, but seems to be a problem for the gravitational instability scenario for any value of Ω consistent with the Virgo infall results. As an aside we might ask whether the holes are empty of matter, or merely empty of bright galaxies? What theory could explain the latter alternative?

Alternative cosmogonic scenarios involve non-gravitational dissipative processes that could in principle more readily explain large holes accompanied by small velocities. If the mysterious dark matter in halos and binding clusters of galaxies is itself dissipationless, then dissipational formation of galaxies and clusters would provide a natural mechanism for the separation of the luminous galaxies from their dark halos and could partially explain the observed trend of M/L ratios with system scale size.

In the pancake model favored by Zeldovich and collaborators, the first structures to collapse are very massive, in excess of $5 \times 10^{15}\ M_\odot$ if the universe is dominated by massive neutrinos (Doroshkevich *et al.* 1980). Supposedly protoclusters collapse along their shortest axis while still expanding along the other two, but here we run into timing problems. We know galaxies must have formed at an epoch $z > 3$ or else galaxies would he evolving too fast today. Since galaxies in this scenario form after clusters, we require pancake collapse of $> 10^{15}\ M_\odot$ systems at the epoch of galaxy formation. Today we see very little evidence for the existence of pancakes, and systems of $10^{15}\ M_\odot$ (the entire local supercluster) have density contrast of order unity. If our supercluster has already collapsed into a plane, then the Hubble flow transverse to the plane should be totally disrupted, yet there is no evidence that it is. Furthermore there are

numerous clouds of galaxies of mass $\sim 10^{13}$ $M_\odot$ that have virtually no connection with any well defined massive supercluster. Not all gas clouds will collapse at the same time, and so galaxy formation should be an ongoing process. Where are these nearby clusters full of young galaxies? My personal feeling is that the pancake picture has several strong attractions, but I would find it easier to compare to the observations if the pancake mass scale could be reduced to 10^{13} to 10^{14} $M_\odot$.

Alternative theories in which shock hydrodynamics plays a significant role in the formation of holes and structures (Ostriker and Cowie 1981; Ikeuchi 1981) are in a more preliminary state and could possihly explain structure on smaller scales. Whether these processes can explain holes of 60 Mpc extent is unlikely.

A final worry is the observed trend of mass to light ratio versus scale size. The mean M/L grows linearly on small scale, and apparently throughout the virialized clusters seen in the CfA sample, but even the highest mass estimates for the Virgo infall are below the extrapolated linear trend. It is not at all clear, however, that an asymptotic value of Ω has been measured because the mean M/L estimate from the Virgo infall is considerably higher than for clusters of size 1 to 2 Mpc. Why doesn't the Ω derived from the infall test agree with estimates based on rich clusters and other systems which we believe can form without dissipation? Massive neutrinos, hlack holes, or other exotica should fully participate in dissipationless clustering if they are not too hot, which they will not be if they are remnants of the hot Big Bang, and the M/L curve should he flat on that scale and larger. Does this continuing trend imply the existence of a hot massive component of the universe unclustered on observed scales, or can the trend be understood without resorting to such inventions?

As usual, the questions outnumber the answers, but I really do feel progress is being made and will continue in the next few years.

This research was supported in part by NSF grant AST80 - 00876.

REFERENCES

Aaronson, M., Mould, J., Huchra, J., Sullivan, W.T., Schommer, P.H. and Bothun, G.D., 1980, *Ap. J.*, **239**, 12.

Aaronson M., Huchra, J., Mould J., Schechter, P. and Tully, R.B., 1981, preprint.

Boughn, S.P., Cheng, E.S. and Wilkinson, D.T., 1981, *Ap. J. Lett.*, **243**, L 113.

Davis, M., Tonry, J., Huchra, J. and Latham, D., 1980, *Ap. J.*, **238**, L 113.

Davis, M. and Huchra, J., 1982, *Ap. J.*, in press.

Davis, M., Huchra, J., Latham, D.W. and Tonry, J., 1982, *Ap. J.*, **253**.

Davis, M. and Peebles, P.J.E., 1982, in preparation.

de Vaucouleurs, G., 1975, in *Stars and Stellar Systems Vol. IX*, ed. by Sandage, Sandage and Kristian, University of Chicago Press.

de Vaucouleurs, G. & Peters, W.L., 1981, *Ap. J.*, **248**, 395.

de Vaucouleurs, G., Peters, W.L., Bottinelli, L., Gouguenheim, L. and Paturel, G., 1981, *Ap. J.*, **248**, 408.

Doroshkevich, A.G., Khlopov, M.Y., Sunynev, R.A., Szalay, A.S. and Zeldovich, Y.B., 1980, in *Xth Texas Symposium on Relativistic Astrophysics.*

Efstathiou, G. and Eastwood, J.W., 1981, *M.N.R.A.S.*, **194**, 503.

Einasto, J., Kaasik, A., and Saar, E., 1974, *Nature*, **250**, 309.

Faber, S.M. and Gallagher, J.S., 1979, *Annual Rev. of Astron. Astrophys.*, **17**, 135.

Fisher, J.R. and Tully, R.B., 1981, *Ap. J. Suppl.*, **47**, 139.

Gregory, S.A., Thompson, L.A. and Tifft, W.G., 1981, *Ap. J.*, **243**, 411.

Gregory, S.A. and Thompson, L.A., 1978, *Ap. J.*, **222**, 784.

Gunn, J., 1978, in *Observational Cosmology*, 8th course at SAAS-FEE, Geneva.

Huchra, J., 1981, private communication.

Ikeuchi, S., 1981, *Publ. Astron. Soc. Japan*, **33**, 211.

Kirshner, R., Oemler, A., Schechter, P.A. and Shectman, S., 1981, *Ap. J. Lett.*, **248**, L 57.

Ostriker, J.P. and Cowie, L., 1981, *Ap. J. Lett.*, **243**, L 127.

Ostriker, J.P., Peebles, P.J.E. and Yahil, A., 1974, *Ap. J. Lett.*, **193**, L 1.

Peebles, P.J.E., 1976, *Ap. J.*, **205**, 318.

— 1980, *The Large Scale Structure of the Universe*, Princeton Press.

Press, W.H. and Davis, M., 1982, *Ap. J.*, in press.

Sandage, A., 1978, *A. J.*, **83**, 904.

Sandage, A. and Tammann, G.A., 1981, "*A Revised Shapley-Ames Catalog of Bright Galaxies*", Carnegie Institution of Washington.

Silk, J., 1974, *Ap. J.*, **193**, 525.

— 1977, *Astr. Ap.*, **59**, 53.

Smoot, G.F. and Lubin, P.M., 1979, *Ap. J. Lett.*, **234**, L 83.

Tarenghi, M., Chincarini, G., Rood, H.J. and Thompson, L.A., 1980, *Ap. J.*, **235**, 724.

Tarenghi, M., Tifft, W.G., Chincarini, G., Rood, H.J. and Thompson, L.A., 1979, *Ap. J.*, **234**, 793.

Tifft, W.G., 1980, *Ap. J.*, **239**, 445.

Tonry, J. and Davis, M., 1981, *Ap. J.*, **246**, 680.

Tonry, J., 1980, Ph. D. thesis, Harvard University.

Tully, R.B., 1981, preprint.

Yahil, A., Sandage, A. and Tammann, G., 1980, *Ap. J.*,**242**, 448.

— 1981, in preparation.

Zeldovich, Ya. B., 1978, in *The Large Scale Structure of the Universe, IAU Symp.*, 79, p. 409.

Zeldovich, Ya. B. and Novikov, D., 1975, *Structure and Evolution of the Universe*, Moscow.

DISCUSSION

FABER

As I mentioned in my talk yesterday there does not seem to be strong evidence that the ratio of ordinary to total matter is any lower in large clusters, like Coma, compared to small, spiral-dominated groups. This is so despite the much higher mass-to-light ratio of Coma, because it seems that there is a sizeable amount of diffuse gas in Coma that never condensed into galaxies and hence never made stars. Nevertheless, the luminosity function for galaxies in Coma seems to be relatively normal. These two facts together suggest the possibility that the total number of galaxies per unit gas mass made in Coma is lower than in smaller groups. If so the mass per galaxy which you showed to rise strongly with scale size, may be a fragile quantity, owing to the inconstancy of galaxy number per unit mass in different regions. Could you comment?

DAVIS

Yes, I agree that systematic errors of this sort can at least partially explain the observed trend. I worry about your statement of excess diffuse gas in Coma type clusters relative to lesser groups, because the gas is so much harder to observe in the smaller groups. It would be nice to have done the survey based on an R magnitude limit, but unfortunately only B magnitudes are available. None of this, however, affects the results of the Virgo infall test.

REES

I would like to mention a possible origin for "voids" which might avoid the alleged problems of a simple gravitational clustering picture, and also obviate the extravagant energy requirements of a model which involves pushing material around non-gravitationally at thousands of kilometers per second. It could be that the distribution of gas and galaxies is fairly uniform, but that large-scale non-uniformities in the heating of intergalactic gas have inhibited galaxy formation in large volumes. The energy required to heat gas to $\sim 10^6$ K, and thereby (by the raising the Jeans mass above a galactic mass) to inhibit galaxy formation, is only $\sim$ 100 ev per particle. This is a thousand times less than the energy needed to evacuate gas from the volume of the voids in the Hubble time.

DAVIS

This is entirely possible and would then remove the motivation for a non-gravitational origin for our infall into Virgo. It would also have caused us to overestimate the density contrast of the Virgo supercluster relative to the mean background, which would substantially increase the derived density estimate and could suggest $\Omega = 1$.

PEEBLES

I was stimulated, by your problem of forming holes, say 20 Mpc in size, without requiring large peculiar velocities, to study the development of holes in a zero pressure spherical model. The initial central density was depressed 20 per cent from the mean. If that is balanced by an overdense rim one finds that the material piles up in a dense ridge with little peculiar velocity. If there is a net mass deficit, the mass from the center piles up in a ridge that moves out with peculiar velocity one third of the Hubble velocity. This suggests that the peculiar velocity around the rim of the hole with 20 h^{-1} Mpc diameter might be 300 km sec^{-1} or a good deal less than that.

DAVIS

I wonder if this is the generally expected case, or might it be the result of special conditions in your counter-example, such as its spherical symmetry.

OSTRIKER

You remarked that the numerical simulation which was designed to have a similar spatial structure to the real universe had a structure in Hubble space quite different from the real universe. In the simulation there was more evidence for radial structure, "fingers", and less of a tendency towards filaments and holes coherent over large angles. We recall also Tammann's talk and his remarks quoting Zeldovich to the effect that it is remarkable how, on the one hand the Hubble flow is smooth and quiet, and on the other the universe is entirely irregular on scales of less than 100 Mpc. The possible discrepancies here could be understood if some of the structure seen is due to velocities of non-gravitational origin (for example hydrodynamical). We may be deceived in always using gravity to interpret structure.

There is another discrepancy highlighted by your talk. The local dynamical determinations of Ω you quoted ranged between 0·14 and 0·7 and are all much

larger than the best estimate determined by light element formation in the early universe as shown at this conference by Audouze.

DAVIS

I agree entirely. Some hydrodynamical means of evacuating the large holes, such as you have proposed, could be very useful. Regarding the second point, the nucleosynthesis constraints give an upper limit to the baryon density only, which therefore suggests that non-baryonic matter dominates the mass density of the universe.

SILK

Two questions:

1) You used a spherically symmetric linear flow model for the Virgo supercluster to infer that our Galaxy was just turning around towards Virgo, and inferred that the pancake model would have a severe time scale problem. In fact, in a model for the Virgo supercluster flow based on the pancake model, the galaxy distribution is both highly flattened and non-linear, and our Galaxy is not just turning around.

2) You stated that the N-body simulation at $\Omega = 1$ was highly clustered as compared with the observed distribution. Since dissipation can only enhance the clustering, can you therefore rule out a model with $\Omega = 1$?

DAVIS

In response to question 1, the linear model I described was intended specifically to avoid the assumption of spherical symmetry. Only the non-linear model requires the assumption of spherical symmetry, and yet both methods yield very similar values of Ω.

As for the N-body models, I would hate to conclude anything about the cosmological density because they simply are not an adequate match to the observations for any value of Ω. One has the freedom to rescale the simulation to whatever physical length desired, so in practice it should he possible to match the observations of the correlation length even if the physical processes responsible for the clustering are quite different.

THE NATURE OF THE LARGEST STRUCTURES IN THE UNIVERSE (*)

J.H. OORT
Sterrewacht Leiden, The Netherlands

The principal beacons by which we can study the structure of the universe are the galaxies. Already a first glance shows their distribution to be extremely inhomogeneous. They congregate in groups containing from two to many thousand members. There does not appear to be a preferred size. Structures have been found that have dimensions up to ~ 100 Mpc. These very large structures, which are called superclusters, are highly irregular; they appear to be essentially unrelaxed. Nevertheless, as will be evident from the following discussion, they are not chance fluctuations, but well established, discrete structures.

The question naturally arises where the inhomogeneity ends. Can we find a scale above which the universe is sufficiently homogeneous to give meaning to the notions of a universal radius of curvature? That there *is* such a scale is beyond a doubt. We see this for instance from the homogeneity of the cosmic black-body radiation. It is also evident from the isotropy in the distribution of the brighter extragalactic radio sources which shows that the universe is essentially homogeneous when averaged over volumes of 500 Mpc diameter. A similar conclusion can be drawn from the distribution of the rich galaxy clusters contained in Abell's catalogue (cf. Oort 1958). Such volumes are still small compared to the volume within the present horizon of the universe, at roughly 3000 Mpc.

Because superclustering is such an erratic phenomenon the only way I can introduce my subject is by describing a few examples, for which I

(*) In this paper, except where otherwise indicated, the Hubble constant has been assumed to be 75 km s^{-1} Mpc^{-1}.

choose those which have been studied most extensively, viz.: the Virgo or Local Supercluster, the Coma-A1367 supercluster, and the superclusters in Perseus and Hercules.

1 - The Local Supercluster

A good impression of the Local Supercluster may be obtained from the plot of the distribution of approximately 1280 galaxies brighter than the 13th photographic magnitude published by Shapley and Ames (1932) and shown in Figure 1. The most striking feature is the dense Virgo cluster, 15° from the north-galactic pole. It has a recession velocity of 1000 km s^{-1} relative to the Local Group; it is 6° in diameter, which at its distance of ~ 20 Mpc corresponds to ~ 2 Mpc. A tail of about 25° stretches to 50° latitude. An opposite tail, somewhat more broken, reaches across the pole to 50° latitude on the other side, the part between 50° and 80° latitude being called the Ursa Major cluster. This "cluster" and the group close to the galactic pole have approximately the same velocity as the Virgo cluster, and so has the first 15° of the tail on the other side; but between $b = 58°$ and 50° in this tail the velocity gradually increases from 1100 to 2500 km s^{-1}. Judging from the magnitudes of the galaxies concerned, the distance appears to rise in proportion to the velocity. The tail appears to turn sharply away from us. At the very tip of the tail, near $b = 50°$, there is a small group of 5 galaxies with velocities around 4300 km s^{-1}, which looks like a separate group accidentally projected onto the Virgo tail.

Even if we leave out the Virgo cluster and its appendages the north-galactic hemisphere is distinctly more densely populated than the southern. The majority of all the galaxies in this hemisphere may belong to one superstructure. This is distinctly flattened towards an axis identical with that of the dense chain containing the Virgo cluster. The axis has been termed supergalactic equator by de Vaucouleurs (1956) and is shown by the solid line passing near the poles in Figure 1.

Our Galaxy probably belongs to the Virgo supercluster. It is not clear how far the supercluster extends beyond it. Apart from a feature marked B in Figure 1 the galaxies in the south-galactic hemisphere may have *some* tendency to concentrate towards the supergalactic equator. This may be taken as an indication that the supercluster extends considerably beyond the Galaxy. Yahil *et al.* (1980) find that the average space density of galaxies falls off roughly as r^{-2} (r being the distance from the centre of the

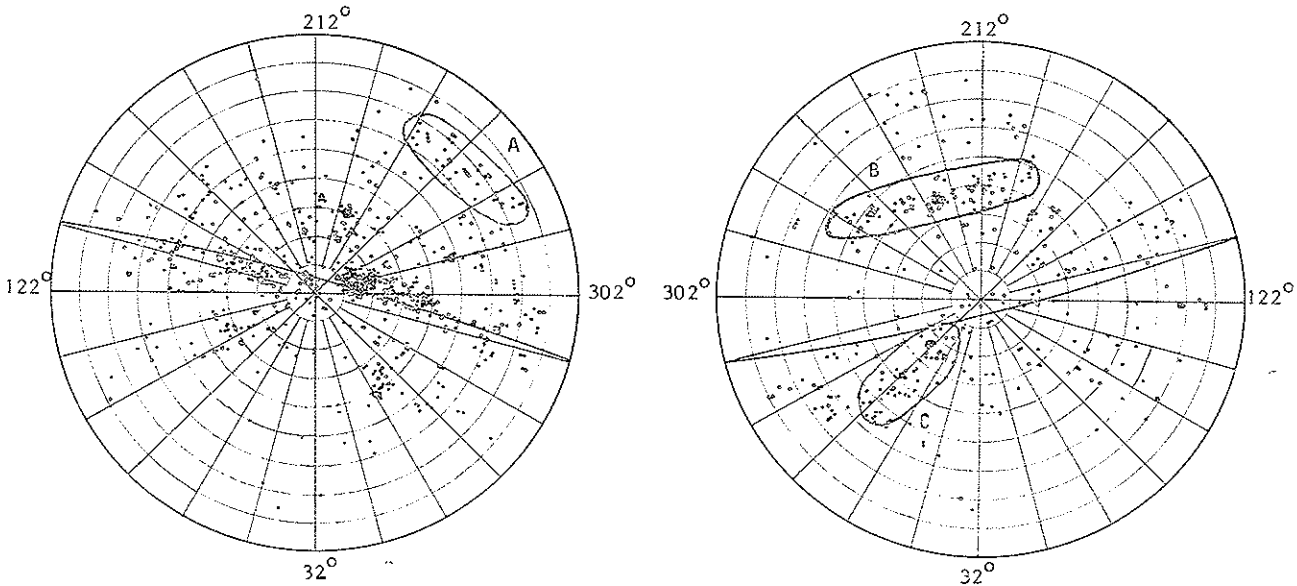


FIG. 1. The distribution of galaxies brighter than the 13th photographic magnitude. The panel on the left shows the north-galactic hemisphere, that on the right the south-galactic hemisphere; the galactic poles are at the centres, the circles at intervals of 10° latitude; ℓ^{II} is shown at the circumference. (Shapley and Ames 1932)

Virgo cluster) and continues to $r = 30$ or 40 Mpc. The total mass may roughly be estimated at 5×10^{14} $M_\odot$.

Within the supercluster the distribution of the galaxies is extremely clumpy, with groups varying from a few to perhaps fifty members; tbe larger ones have a pronounced tendency for strongly elongated shapes. This is most outspoken in the 80°-long ridge through the Virgo cluster and its appendages, but it holds also for the larger outlying groups like those marked A, B and C in Figure 1. Feature A has a velocity of 2300 km s^{-1}, and may be a separate system. Feature B has likewise the appearance of a separate unit; it has an average velocity of 1400 km s^{-1} and a dispersion of only 300 km s^{-1}. We do not know whether these elongated features are disks or bars.

There is a striking difference in composition between the central cluster and all other features; the dense cluster consists exclusively of E and S0 galaxies, while the "tails" have only about 25% E and S0 galaxies.

The supercluster shows a mixture of dynamical evolution. The cluster itself has a velocity dispersion of $\sim$ 650 km s^{-1}. With a corresponding crossing time of about 1.5×10^9 years the cluster might be expected to be largely randomized but a close inspection indicates that its structure is far from regular. It seems instead to be composed of several more or less discrete groups. The lack of strong central concentration likewise shows that it is far from a dynamically steady state. The large supercluster is still expanding, with almost the same velocity as the universe. It may well expand to infinity.

Little is known about the internal dynamics of the various substructures. The larger features, like the appendages of the Virgo cluster, and A, B and C, are presumably expanding, at least in the direction in which they are elongated; in the other direction some sort of collapse may have occurred.

Figure 1 shows a number of fairly large areas in which the density is far *below* the average. Below 20° galactic latitude this shortage is caused by galactic absorption, but the extended near-complete voids above this latitude must be real, and are no less interesting than the concentrations.

2 - The Coma Supercluster

The most striking part of the Coma supercluster is formed by two rich clusters, the Coma cluster and Abell 1367, separated by 17°.1, or about 28 Mpc. From an extensive survey by Gregory and Thompson (1978) of

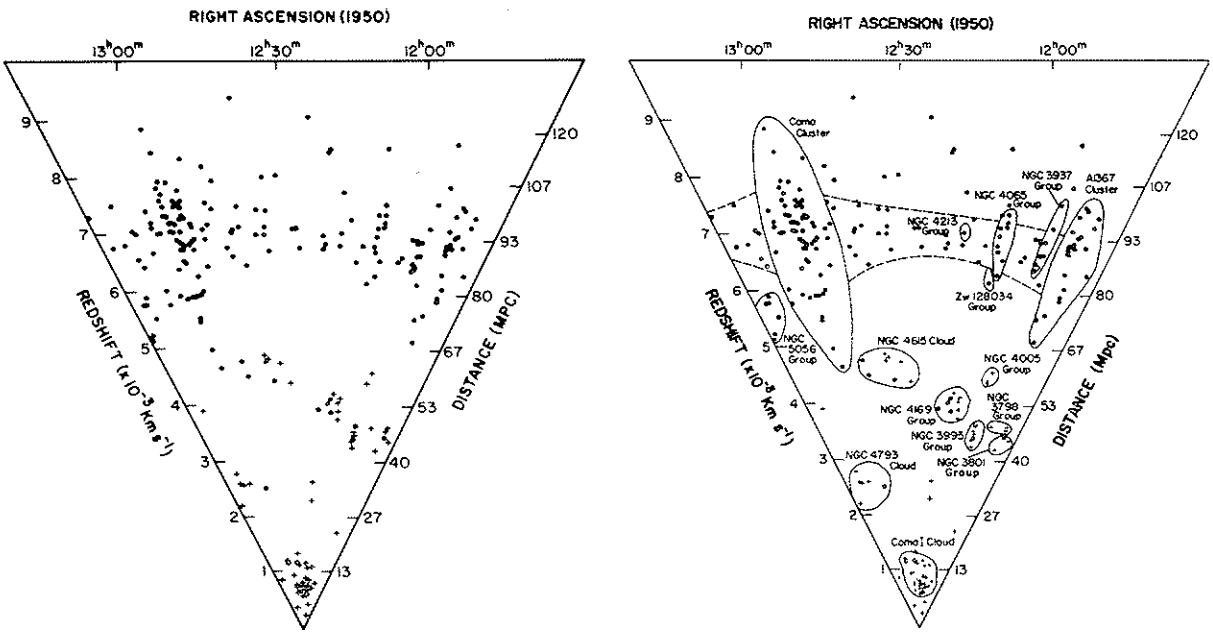


FIG. 2. "Wedge diagram" of the Coma supercluster (Gregory and Thompson 1978). As the supercluster is elongated in the east-west direction, right-ascensions have been chosen as position coordinates; the galaxies lie between + 19° and + 32° declination. The angular size has been magnified about two times compared with the indicated distance scale.

radial velocities in the region surrounding the two clusters a bridge was found connecting the two clusters; the radial velocities of the galaxies in this bridge show a distinct concentration around the practically equal velocities of the two clusters (Figure 2).

An impression of the total extent can be obtained from the recent radial-velocity survey by Davis *et al.* (1982) of all galaxies brighter than $m_B = 14.5$ in the north galactic polar cap down to 40° latitude. Figure 3 shows the distribution on the sky for the galaxies with velocities between 6000 and 10 000 km s^{-1}. The Coma supercluster lies in the range 6000 to 8000 km s^{-1}, which is indicated by small crosses. Its members are evidently concentrated in a strongly elongated region, marked COMA. The three-dimensional structure may be seen in Figure 4 which shows the distribution

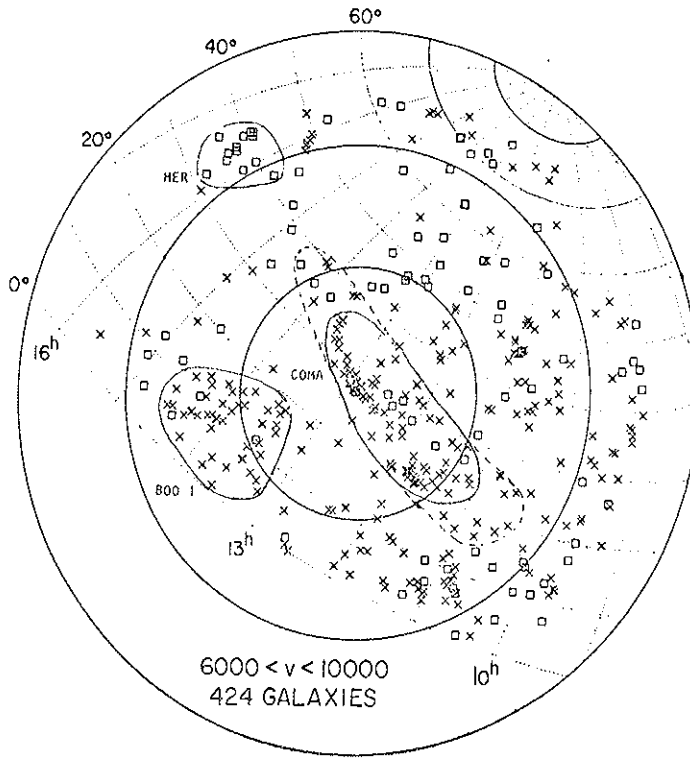


FIG. 3. Distribution of galaxies brighter than Zwicky magnitude 14.5 in the velocity range 6000 to 10 000 km s^{-1} (only galaxies with absolute magnitudes brighter than − 18.5 in the Zwicky scale are included). Crosses indicate velocities between 6000 and 8000, squares those between 8000 and 10 000 km s^{-1}. Reproduced by kind permission from Davis *et al.* (1982).

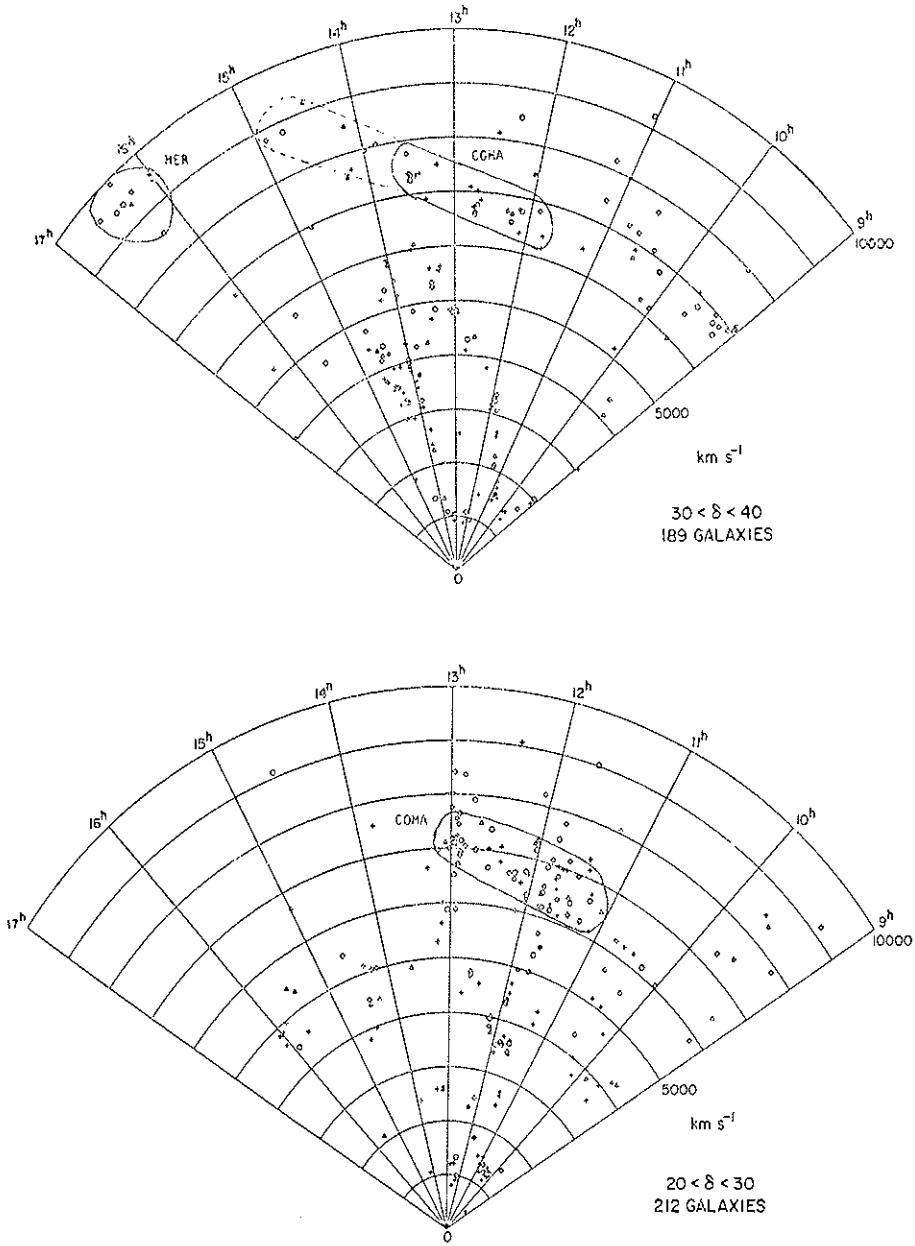


FIG. 4. Right-ascension velocity diagrams in two declination zones for galaxies brighter than 14.^{m}5 in the catalogue of Zwicky *et al.* (1961-1968), and with absolute magnitude brighter than −18.5. Circles are ellipticals, diamonds SO's, plus signs spirals and triangles irregulars. Reproduced by kind permission from Davis *et al.* (1982).

in velocity and right-ascension in the two declination zones over which it extends. In each zone there is a fairly well defined concentration, again indicated by a contour COMA, around the region where the Coma and Abell 1367 clusters are situated. From these data the densest part of the supercluster is found to extend from roughly $11.^h0$, + 20°, 6200 km s^{-1} to $13.^h6$ + 35°, 7400 km s^{-1}, over a length of about 37°, or 68 Mpc. The total feature may well extend considerably farther. The average velocity is + 6800 km s^{-1}. The velocity width is estimated to be between 900 and 1000 km s^{-1} outside the two dense clusters. With normal Hubble expansion this would correspond to a depth of 13 Mpc, but evidently expansion in the supercluster must be entirely different from that in the universe. It may well be zero in the direction of the small axes. The depth in the direction of the line of sight is therefore entirely unknown. The supercluster may be either oblate or prolate, with an axial ratio of at least 5:1.

If the dispersion in the radial velocities is representative of the internal velocity dispersion in general, the galaxies could not have moved over more than a small fraction of the supercluster's length since its birth.

3 - THE PERSEUS SUPERCLUSTER

An idea of the complexity that these superstructures can sometimes have may be obtained from the investigations of Einasto and co-workers (1980a, b) at Tartu. They investigated large-scale structures in the south-galactic hemisphere above − 2° declination, using clusters from the catalogues of Zwicky *et al.* (1961-1968). Distances were derived partly from redshifts and partly from apparent magnitudes. The distribution in two redshift intervals is shown in Figure 5. We see that in the velocity range 3500 to 6500 km s^{-1} there is a concentration in an inclined band indicated by the dashed lines in Figure 6, and a second ridge extending from A 426 in the upper left to A 194 at the bottom of the picture. In the range 6500 to 10 000 km s^{-1} these ridges have disappeared. The two ridges are considered by the authors to be parts of a large supercluster which comprises also the group around Uppsala 487 in the lower right. It covers about 1600 square degrees, or 2000 Mpc^2. The densest part is the chain in the northern part, which starts at the rich Perseus cluster Abell 426 and contains the clusters Abell 347 and 262. A quite remarkable thing is that the Perseus cluster itself contains a narrow chain of galaxies about 0.8 Mpc long in the same direction as the long ridge of 50 Mpc length.

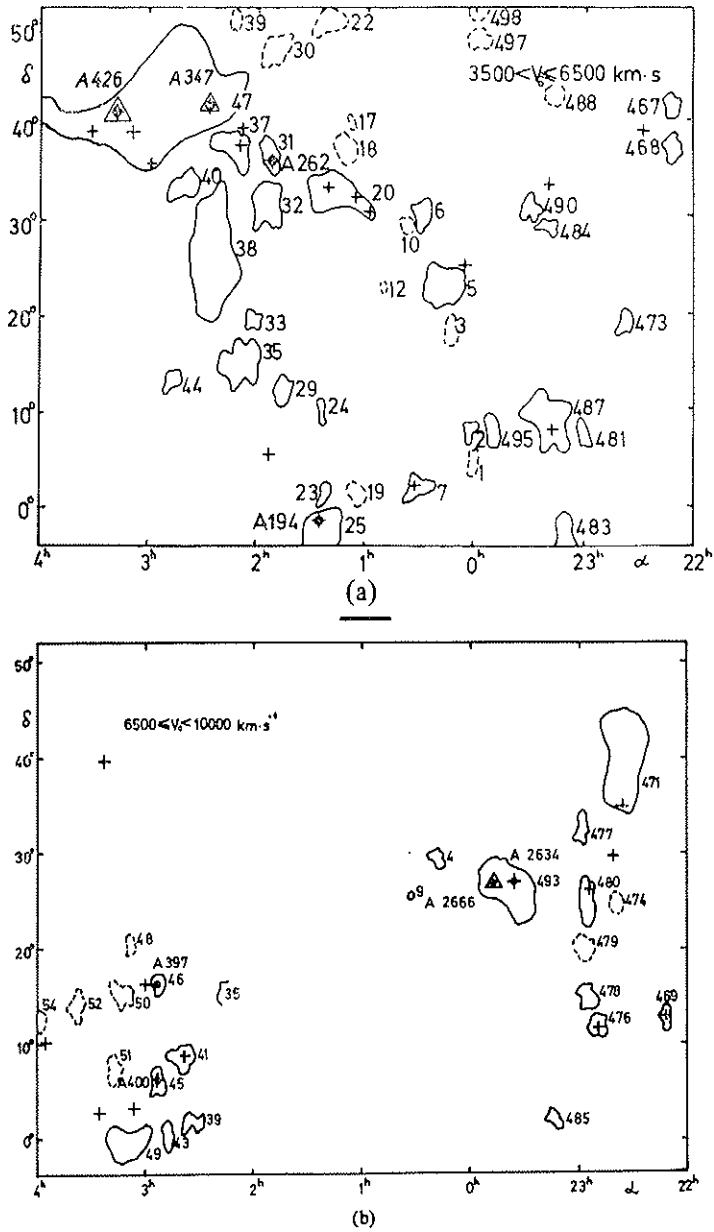


FIG. 5. Perseus supercluster. The distribution of Zwicky clusters with velocities between 3500 and 6500 km s^{-1} (a), and between 6500 and 10 000 km s^{-1} (b), in the south-galactic hemisphere. Full-drawn contours show the clusters with measured redshifts, dotted contours indicate those with distances estimated from magnitudes and cluster diameters. The numbering is from Nilson (1973). Abell clusters are indicated by filled circles and by their numbers in Abell's catalogue (1958); A 426 is the Perseus cluster. (Reprinted by courtesy of Einasto *et al.* 1980 b).

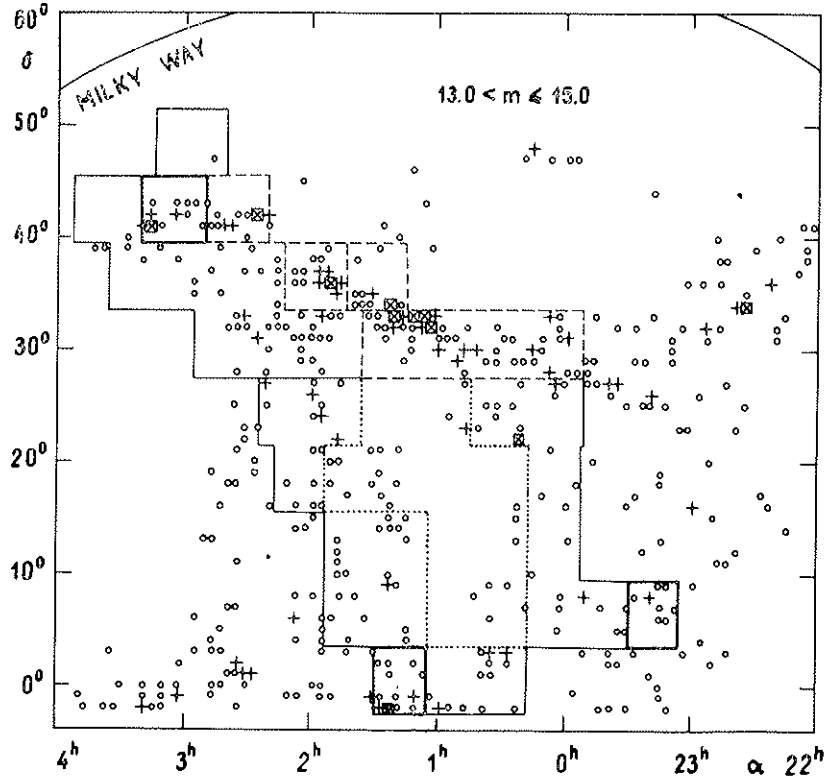


FIG. 6. Perseus supercluster. Distribution of galaxies between 13.^{m}0 and 15.^{m}0 in the same region as that shown in Figure 5 (Einasto *et al.* 1980 b). Symbols indicate numbers per square degree (circles 2-3, crosses 4-7, crosses in square 8-15).

The galaxies between 13.^{m}0 and 15.^{m}0, which in general lie in the same distance range as the Zwicky clusters of the distance class "near" shown in Figure 5, display a distribution closely resembling that of the clusters: a strong concentration along the two ridges and a nearly empty region in between (Figure 6). Practically all galaxies in this magnitude range situated within the limits indicated appear to belong to the Perseus supercluster, as is shown by their velocity distribution. This holds also for the galaxies in the central area of the supercluster (between the ridges) whose velocities are sharply peaked in the interval 4950 to 5050 km s^{-1}, indicating that the galaxies between the chains form a thin stratum. The authors have defined three other main superclusters in the large field

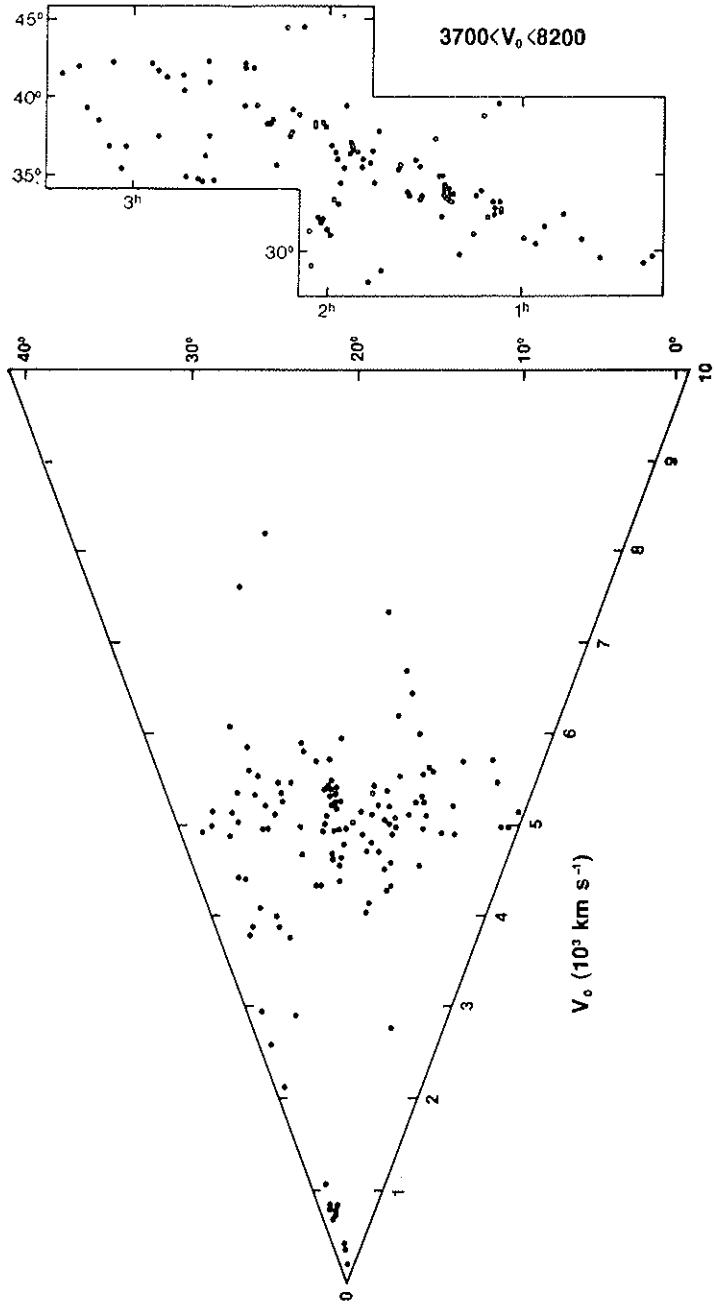


FIG. 7. Perseus supercluster. Above: Distribution of galaxies brighter than $m_p = 14.0$ in the Zwicky catalogue having velocities between 3700 and 8200 km s^{-1}. The boundaries of the region surveyed by Gregory *et al.* (1981) are indicated. The tips of the Perseus supercluster chain lie around $\ell = 150°$, $b = -12°$ (left) and $\ell = 115°$, $b = -34°$ (right). Below: "Wedge" diagram for all galaxies in the survey with velocities less than 10 000 km s^{-1}. The positions are measured along the chain.

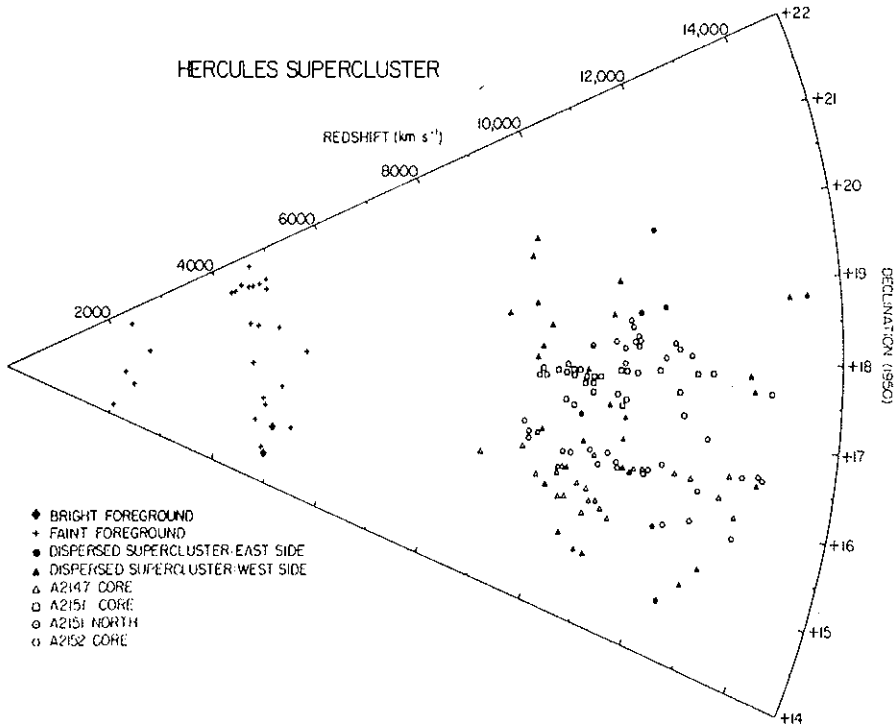


FIG. 8. Southern part of the Hercules supercluster (Tarenghi *et al.* 1979).

investigated. They conclude that all clusters of galaxies and most galaxies are located in superclusters. They believe that neighbouring superclusters are in contact and form walls and ribs of "cells"; "the space inside the cells is void of clusters and almost void of galaxies".

Valuable new data on the principal chain of the Perseus supercluster have been derived by Gregory *et al.* (1981) from a complete set of radial velocities of galaxies with $m_p < 14.0$. Their data, displayed in Figure 7, support the supercluster character and show that it has a "filamentary component" whose depth may be of the same general order as its width.

4 - THE HERCULES SUPERCLUSTER

A third superstructure on which a considerable body of data has been gathered is the Hercules supercluster. It contains two groups of rich

clusters, one near Abell 2151, with a redshift of 10 500 km s^{-1} (Figure 8), the second near Abell 2199 at a redshift of 9000 km s^{-1}. The whole supercluster is shown in Figure 9. It extends over 38° in declination, corresponding to 85 Mpc. Again, galaxies in this area avoid a large interval in velocity, from 5000 to 10 000 km s^{-1} in the region depicted in Figure 8; the corresponding distance interval is 70 Mpc.

5 - Other possible superclusters

Superstructures are also evident in the clumpy distribution of the clusters in Abell's catalogue (1958). Abell himself indicated some 50 what he called "class 2" clusters, but the reality of these groups can only be demonstrated if radial velocities are available. An example is the 1451+22 group of 7 Abell clusters, which in addition contains 10 "groups" and poor clusters. Radial velocities were measured by Ford *et al.* (1981). It has $z = 0.116$, longest diameter $\sim$ 8° or 64 Mpc, true velocity dispersion 372 km s^{-1} which, with a Hubble constant of 75 km s^{-1} Mpc^{-1}, would cor-

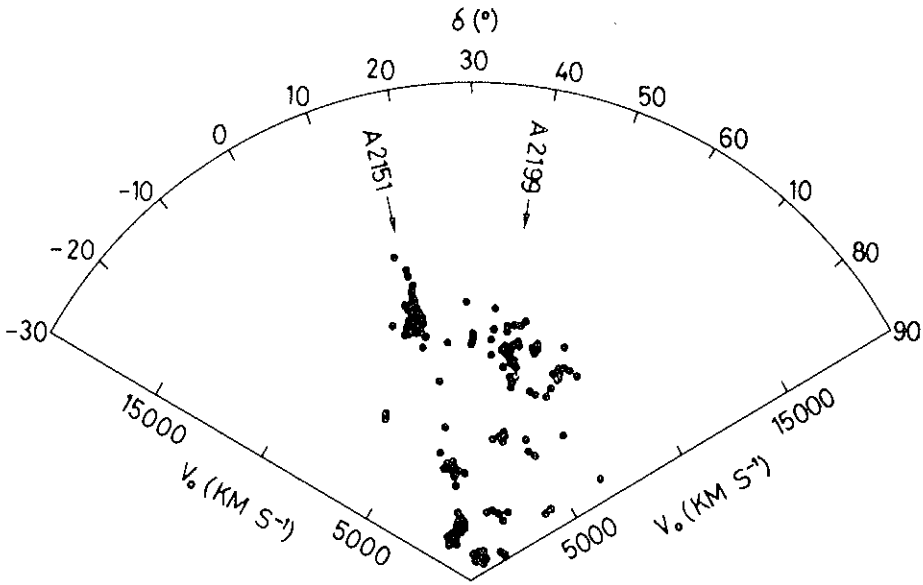


Fig. 9. Hercules supercluster. Cone diagram of velocity versus declination for galaxies between $15^h\,30^m$ and 17^h right ascension (Tarenghi *et al.* 1980).

respond to 5 Mpc. This is notably smaller than the lateral dispersion on the sky, 20 Mpc along the major axis and 12 Mpc along the minor axis, and can either be interpreted as flatness of the supercluster or as a gravitational effect of the supercluster which decreases the effective Hubble constant.

Finally, some still provisional evidence has been found for superclusters indicated by quasars (Oort *et al.* 1981). Quasars are so rare that only a fraction of the superclusters will contain a quasar, and only a few would contain two or more; these latter are the only ones which can be found by means of quasars. One prohable case has been identified at $z = 3.15$, in which a pair of quasars is found with a separation of 12 Mpc and a velocity difference of 1400 km s^{-1}, corresponding to a separation along the line of sight of 4.5 Mpc. Presently data on distant quasars are still scarce. However, it may be hoped that in the future they will provide very valuable information on the evolution of the superstructures.

Interesting data on the redshift distribution of galaxies down to considerably fainter magnitudes than in the surveys considered above are being obtained by Kirshner *et al.* (1981), but as these are still in progress I shall not discuss them here. My description of the distribution of galaxies in space has been somewhat baphazard, and is in any case very incomplete. Even within 100 Mpc there are many more superstructures than the largest ones which were discussed in the preceding pages. A more complete picture, in a restricted part of the sky, can he obtained from the data given by Davis *et al.* (1982).

6 - Correlation analyses

The problem of galaxy clustering may be approached in a different way, viz. by investigating the general correlation between positions of galaxies. First developed by Totsuji and Kihara (1969) it has been applied to a great variety of catalogues by Peebles and co-workers (cf. Peebles 1980). They found that an important property of the universe can be described in an extremely simple manner by the spatial correlation, or "covariance" function $\xi(r)$ which gives the probability $P(r)$ that another galaxy is contained in a unit of volume at a distance r from a given galaxy

$$P(r) = \bar{n}\ (1 + \xi(r))\ ,$$

where $\bar{n}$ is the number density of galaxies averaged over a large volume. The function $\xi(r)$ can be well represented by a power law $(r_0/r)^\gamma$, with

$\gamma = 1.8$ and $r_0 \sim 5.6$ Mpc. The relation holds for a large range of distances, from about 10 kpc to 15 Mpc. It gives a good impression of the clumpiness of the distribution on all scales, except that of the superstructures which I have been discussing. Although a correlation analysis can still be made for radii of the order of 50 Mpc (cf. the discussion of the correlation between the positions of Abell clusters by Hauser and Peebles 1973 and by Peebles 1978), the co-variance function does not give an adequate description of the unrelaxed superclusters.

7 - Summary of some characteristics of the large-scale structure

1. Structures have been found up to dimensions ~ 100 Mpc (for $H_0 = 75$ km s^{-1} Mpc^{-1}) and masses of perbaps 10^{16} $M_\odot$.
2. There exist large "voids" with diameters of the same order as those of superclusters.
3. Some superclusters have a strongly flattened, and possibly elongated, sbape.
4. Along the long axes crossing times are longer than the age of the universe.
5. Complicated internal structure has been found in the Local Supercluster.
6. Within and around the Perseus Supercluster indications have been found that neighbouring superclusters may be interconnected.

8 - Origin of superclusters

Superclusters may have been formed in two ways: by gravitational clustering of pre-existing galaxies, or by local collapse in the gaseous pbase of the universe. Tbe former process is amenable to calculation if we know the epoch at which galaxies were formed. For the second process we must know sizes and amplitudes of the initial density fluctuations, wbich may either date from earlier pbases of the universe or may be caused by explosive phenomena at a later epoch, such as envisaged by Ostriker and Cowie (1981).

The first model, that of the gravitational instability of a universe

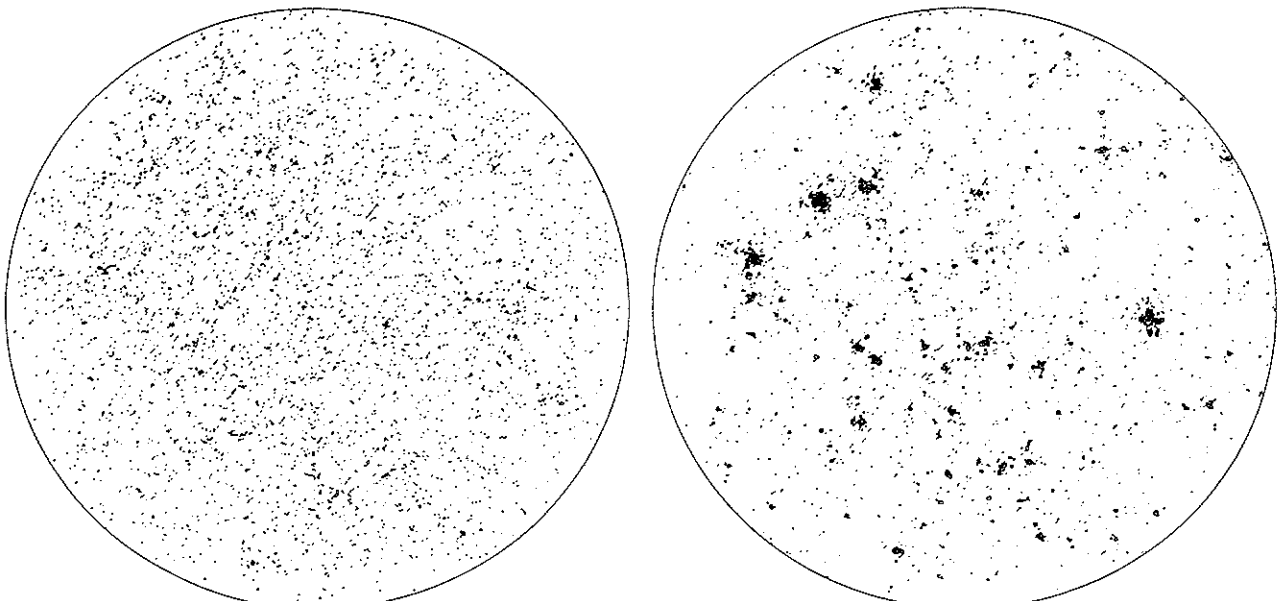


FIG. 10. N-body simulation of galaxy clustering. A sphere of present radius 50 Mpc containing 4000 galaxies in co-moving coordinates. The x and y coordinates of the galaxies are plotted with the z coordinate suppressed. $\Omega = 0.1$, $H_0 = 50$ km s^{-1} Mpc^{-1}. Left: $z = 14.2$, right: $z = 0$ (from Aarseth, Gott and Turner 1979).

consisting of galaxies, distributed at random at the epoch of their birth, has been extremely successful in explaining all the clumpiness and clustering in the universe, except the very largest [cf. the simulations by Aarseth *et al.* (1979) illustrated in Figure 10], and in reproducing the observed co-variance function. It appears probable that this process is responsible for practically all structures. But it is doubtful whether structures of the order of 100 Mpc, with crossing times longer than the age of the universe, can be formed in this way; it is also doubtful whether it could lead to strongly flattened or elongated structures. For the latter anisotropic initial density fluctuations of large mass would seem to be required or possibly a scenario like that proposed by Ostriker and Cowie (1981).

ACKNOWLEDGEMENTS

I am greatly indebted to Drs. Sandage and Tammann for a preprint of the Revised Shapley-Ames Catalogue and to Davis *et al.* (1982) for sending me their valuable data prior to publication.

REFERENCES

Aarseth, S.J., Gott, J.R. III, Turner, E.L., 1979, *Ap. J.* **228**, 664.

Abell, G., 1958, *Ap. J. Suppl.* **3**, 211.

Davis, M., Huchra, J., Latham, D.W. and Tonry, J., 1982, *Ap. J.*, **253**, 423.

de Vaucouleurs, G.H., 1956, *Vistas in Astronomy*, **2**, 1584.

Einasto, J., Jôeveer, M., and Saar, E., 1980a, *Nature* **283**, 47.

— 1980b, *Monthly Not. R. Astr. Soc.* **193**, 353.

Ford, H.C., Harms, R.J., Ciardullo, R., and Bartko, F., 1981, *Ap. J. Letters*, **245**, L 53.

Gregory, S.A., and Thompson, L.A., 1978, *Ap. J.* **222**, 784.

Gregory, S.A., Thompson, L.A., and Tifft, W.G., 1981, *Ap. J.* **243**, 411.

Hauser, M.G., and Peebles, P.J.E., 1973, *Ap. J.* **185**, 757.

Kirshner, R.P., Oemler, A., Schechter, P.L. and Shectman, S.A., 1981, *Ap. J. Letters*, **248**, L 57.

Nilson, P., 1973, *Uppsala Astronomica Obs. Ann.*, Vol. 6.

Oort, J.H., 1958, *Inst. Internat. de Physique Solvay*, 11ième Conseil de Physique, Brussels, 1958.

Oort, J.H., Arp. H.C., and de Ruiter, H., 1981, *Astron. Astrophys.* **95**, 7.

Ostriker, J.P., and Cowie, L.L., 1981, *Ap. J. Letters* **243**, L 127.

Peebles, P.J.E., 1978, in *The Large Scale Structure of the Universe*, p. 217, I.A.U. Symp. No. 79, ed. M.S. Longair and J. Einasto.

— 1980, *The Large-Scale Structure of the Universe*, Princeton Univ. Press.

Shapley, H. and Ames, A., 1932, *Harvard Obs. Ann.* **88**, No. 2.

Tarenghi, M., Tifft, W.G., Chincarini, G., Rood, H.J., and Thompson, L.A., 1979, *Ap. J.* **234**, 793.

Tarenghi, M., Chincarini, G., Rood, H.J., and Thompson, L.A., 1980, *Ap. J.* **235**, 724.

Totsuji, H., and Kihara, T., 1969, *Publ. Astr. Soc. Japan* **21**, 221.

Yahil, A., Sandage, A., and Tammann, G.A., 1980, *Physica Scripta* **21**, 635.

Zwicky, F., Wild, P., Herzog, E., Karpowicz, M., Koval, C., 1961-1968, *Catalogue of Galaxies and Clusters of Galaxies*, 1-6, California Inst. Tech., Pasadena.

DISCUSSION

SILK

There are two approaches to explaining highly asymmetric superclusters. One is to appeal to dissipative processes at a late epoch. The alternative is for the initial fluctuations to be asymmetrical. Shape asymmetry is preserved in the linear growth regime of density fluctuations and will be strongly amplified when fluctuations become non-linear.

REES

In a recent analysis of the Osmer-Smith CTIO quasar sample, Webster at Cambridge has noted that there are 4 quasars with a $z \simeq 0.37$ which could be in a "supercluster" 100 Mpc across. He argues quantitatively that such a grouping of quasars is most unlikely to arise by chance if quasars are randomly distributed.

THE NATURE AND ORIGIN OF LARGE-SCALE DENSITY FLUCTUATIONS

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

1 - Introduction

When the galaxy distribution is averaged through a spherical window of radius 10 h^{-1} Mpc ($H_0 = 100$ h km s^{-1} Mpc^{-1}) the rms fluctuations in the smoothed density are found to be $\delta\rho_g/\rho_g \sim 0.6$ (Peebles 1980, § 59). The number density fluctuations on this and larger scales thus are close to linear, simple to specify and, if gravity is the dominant force, evolve in a simple way. There is no guarantee that simple situations lead to fundamental advances in understanding, but it happens often enough to suggest we may find it profitable to approach from the large-scale end the puzzle of why the universe is strongly clumpy on small scales. An important recent development has been the discovery of large angular scale fluctuations $\delta T/T \sim 1 \times 10^{-4}$ in the brightness of the microwave background. I review here the constraints this new measure may provide on models for the nature and origin of the departures from a homogeneous universe.

It should be noted that the anisotropy of the microwave background was only recently discovered and so many of the wanted measurements of the effect still are in progress. Therefore we can only lay out the implications of each of the possible results of the measurements. I shall concentrate on the case that seems fairly likely to me, that the effect is extragalactic and the autocorrelation function of the background temperature has an angular scale of some tens of degrees.

2 - Large angular scale background temperature fluctuations and the Sachs-Wolfe effect

A convenient measure of the anisotropy of the background radiation is the rms value of the temperature difference measured at angular separation θ,

$$\delta T(\theta) = < (T_1 - T_2)^2 >^{1/2} . \quad (1)$$

The measurements by Melchiorri *et al.* (1981) and Boughn *et al.* (1981) suggest $\delta T/T \sim 3 \times 10^{-5}$ at $\theta = 6^0$ and $\delta T/T \sim 1 \times 10^{-4}$ at $\theta \sim 90^0$.

Now let us consider possible causes of this $\delta T/T$ effect. (For a fuller discussion see Peebles 1981a). If the observed $\delta T/T$ were produced by sources in the Galaxy we would expect to find that $\delta T/T$ varies with the wavelength of the radiation and perhaps also with galactic latitude. Both effects can be checked but the data are not yet in. Next, we could imagine that $\delta T/T$ is the result of emission by the clumpy distribution of sources along the line of sight to the horizon. Since the galaxy two-point correlation function is very small at separations $\gtrsim 30\ h^{-1}$ Mpc we would expect that in the absence of scattering the autocorrelation function of $T(\theta, \phi)$ would be negligibly small when the projected separation at the Hubble distance exceeds $30\ h^{-1}$ Mpc,

$$w(\theta) = < T_1\ T_2 > / < T >^2 - 1 \cong 0 \quad , \quad \theta > \theta_c \lesssim 1^0 . \quad (2)$$

As Hogan (1980, 1981) has discussed, scattering could increase the coherence length θ_c; I estimated that scattering by an intergalactic plasma might make θ_c as large as 10^0 (Peebles 1981b). The temperature fluctuations averaged over scales $\theta > \theta_c$ would vary as shot noise, $\delta T/T \propto \theta^{-1}$, so scaling from the measurement of Boughn *et al.* at $\theta \sim 90^0$ we would expect

$$\delta T/T \sim 10^{-3} \quad , \quad \theta = 10^0 . \quad (3)$$

I have the impression that Melchiorri's group ought to have detected temperature fluctuations that large (but for a contrary opinion see Hogan 1981). If further measurements confirmed that the autocorrelation function $w(\theta)$ is appreciable at $\theta \gtrsim 10^0$ it would appear that this model for the background fluctuation is untenable.

The next possibility is that the temperature fluctuations are primeval, that is, present as initial conditions at high redshift. In the conventional cosmology this is a "primeval isothermal" perturbation (eg. Silk and Wilson

1981), where at high redshift the mass density and space curvature are almost exactly homogeneous and the initial baryon number is distributed in a clumpy fashion (so there are appreciable fluctuations in the baryon mass density after annihilation of baryon-antibaryon pairs). At high redshift the radiation mass density ρ_r dominates and so is almost exactly homogeneous, but where the baryon density ρ_m is high ρ_r is slightly below the mean, so the total density is constant. After the universe expands to the point that $\rho_m > \rho_r$ the situation tends to reverse, ρ_m becoming nearly homogeneous, ρ_r clumpy. At small scales that is prevented by pressure. The pressure gradient force is negligible on scales larger than the Jeans length for the coupled matter and radiation (Peebles 1980, eq. [92.47]),

$$\lambda_J \sim 50\ (\Omega_0 h^2)^{-1}\ (1+z)^{-1}\ \text{Mpc}\ . \tag{4}$$

(The density parameter is $\Omega = <\rho>/\rho_{crit}$). At $\lambda > \lambda_J$ we end up with temperature fluctuations in space

$$\frac{\delta T}{T} = -\frac{1}{3}\left(\frac{\delta\rho}{\rho}\right)_i\ , \tag{5}$$

where $(\delta\rho/\rho)_i$ is the primeval baryon density contrast. Thus we could account for the observed temperature fluctuations by supposing there were primeval fluctuations with

$$\left(\frac{\delta\rho}{\rho}\right)_i \sim 3\times 10^{-4}\ ,\quad \lambda \sim 3000\, h^{-1}\ \text{Mpc}\ , \tag{6}$$

where λ is a comoving length. To make galaxies and clusters of galaxies we would want a roughly flat spectrum of primeval isothermal perturbations with $(\delta\rho/\rho)_i \sim 10^{-3}$ at $\lambda \sim 10$ Mpc. If this flat spectrum were extrapolated to larger scales it would make $(\delta\rho/\rho)_i$ much less than in equation (6), so we would have to postulate a break to power spectrum $\sim k^{-3}$ at large scales. It should also be noted that isothermal perturbations violate the idea discussed at this conference by Weinberg that there is a fixed entropy per baryon number. Nevertheless it is a possibility to be kept in mind.

The second possibility is that the background temperature fluctuations are the result of gravitational potential gradients in a universe with a fixed initial entropy per baryon number. This effect was first analyzed by Sachs

and Wolfe (1967). The feature of the effect that is of particular interest here may be seen by the following heuristic argument (Peebles 1981b). Suppose the mass autocorrelation function is negligibly small at separations $r \gtrsim r_c \sim 30\ h^{-1}$ Mpc. Then the fluctuations in the mass distribution smoothed over the scale $r > r_c$ vary with r as $\delta M \propto r^{3/2}$, which is the usual shot noise effect. The typical gravitational potential difference between points at separation r is

$$\delta\phi \sim G\delta M/r \propto r^{1/2} \ . \tag{7}$$

The microwave background radiation measured at points in the sky separated by angle θ originated at points at spatial separation $r \sim \theta_c H_0^{-1}$. The difference of potential through which the radiation moved causes a like difference in temperature,

$$\delta T/T \sim \delta\phi \propto \theta^{1/2} \ . \tag{8}$$

This produces an autocorrelation function of the background temperature (eq. [2]),

$$w(\theta) \propto \theta_1 - \theta \ , \tag{9}$$

where θ_1 is a constant on the order of the size of the sample in which w is estimated. We assumed no mass clustering on large scales and we have arrived at large-scale fluctuations in the background temperature. It is this feature that seems to be suggested by the results of the Melchiorri and Wilkinson groups.

The expected magnitude of $\delta T/T$ depends on the size of the rms mass fluctuations δM. That is determined by the mass autocorrelation function

$$\xi(r) = <\rho(\mathbf{s})\,\rho(\mathbf{s}+\mathbf{r})> / <\rho>^2 - 1 \ . \tag{10}$$

The mean square fluctuations in the mass found within a randomly placed sphere of radius r is

$$\begin{aligned}(\delta M/M)^2 &= \int_r dV_1\, dV_2 <(\rho_1 - <\rho>)(\rho_2 - <\rho>)> / (<\rho> V)^2 \\ &= \int \xi(\mathbf{r}_1 - \mathbf{r}_2)\, dV_1\, dV_2/V^2 \\ &\cong 4\pi J_3(r)/V \ ,\end{aligned} \tag{11}$$

where

$$J_3(r) = \int_0^r r^2\, dr\, \xi(r) \ . \tag{12}$$

The last of equations (11) assumes $\xi(r)$ is small or varies slowly on the scale r of the sphere. We can estimate J_3 on the assumption that the large-scale fluctuations in mass agree with the fluctuations in the distribution of bright galaxies, $\xi \cong \xi_g$, where ξ_g is the galaxy two-point correlation function. The best estimate of the integral of ξ_g is (Clutton-Brock and Peebles 1981)

$$J_3(x) = (960 \pm 125)\, h^{-3}\, \mathrm{Mpc}^3 \quad , \quad x = 30\, h^{-1}\, \mathrm{Mpc} \ . \tag{13}$$

At $x > 30\, h^{-1}$ Mpc we know $|\xi_g| \ll 1$; but ξ_g could have a small positive tail at large separations that could make J_3 at large x much larger than in equation (13); and it is equally conceivable that ξ_g has a negative tail that makes J_3 at large x quite small.

Let us consider first the possibility that ξ is negligibly small at $x \gtrsim 30\, h^{-1}$ Mpc, so equation (13) is a good approximation to the integral J_3 to very large x. Then in an Einstein-de Sitter cosmological model we get (Peebles 1980, § 93; 1981b)

$$\begin{aligned} (\delta T/T)^2 &= (H_0/c)^3\ J_3\ \sin\ \theta/2 \ , \\ \delta T/T &\cong 1.8 \times 10^{-5}\ \theta^{1/2} \ . \end{aligned} \tag{14}$$

In the second equation θ is expressed in degrees. For an open cosmological model computation of $\delta T/T$ apparently requires a simplifying model for the mass distribution, as in Kaiser's (1981) spherical model, or else a four dimensional numerical integration. Figure 1 shows results of a numerical integration for the Friedmann-Lemaitre model with $\Lambda = 0$, $\Omega = 0.3$ (Peebles 1982). The upper curve is given by the first of equations (14). Both curves scale with the integral of ξ as $\delta T/T \propto J_3^{1/2}$.

The size of $\delta T/T$ at large θ changes by a factor ~ 2 when Ω_0 changes from 0.3 to 1. The more interesting point, however, is that for "reasonable" values of Ω we expect $\delta T/T \sim 10^{-4}$ at large θ, comparable to what has been observed. It is this coincidence that suggests to me there may be something in this model for the anisotropy of the microwave background. However, we must take note of the key assumption that the mass autocorrelation function $\xi(r)$ vanishes at $r \gtrsim r_c \sim 30\, h^{-1}$ Mpc. If large-scale

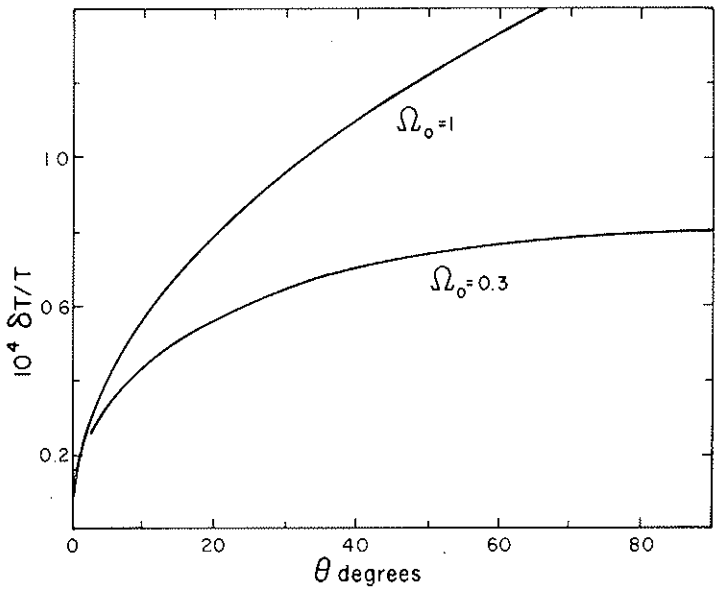


FIG. 1. Sachs-Wolfe effect. In the top curve the density parameter is $\Omega_0 = 1$, in the lower curve, $\Omega_0 = 0.3$. The vertical axis is defined in Equation (1).

clustering made $\xi > 0$ at $r \gg r_c$ it could greatly increase J_3 and hence the expected value of $\delta T/T$. For example, if we added a tail $\xi = 0.1$ extending to $r = 100\ h^{-1}$ Mpc to represent large-scale clustering we would increase the expected value of $\delta T/T$ by a factor of 6, which likely would put it outside the observational limits on $\delta T/T$. Thus if the positive tail of ξ existed we would have to find some way to avoid the Sachs-Wolfe effect. Perhaps the easiest way is to assume the mass distribution is anticorrelated on scales $\sim 100\ h^{-1}$ Mpc, say, so $\xi < 0$ at $r \gtrsim 100\ h^{-1}$ Mpc. That could greatly reduce the integral J_3 and so make $\delta T/T$ smaller than the above estimates. The cases of positive and negative long-range tails of ξ are discussed further in the next two sections.

3 - LARGE-SCALE CLUSTERING?

The estimates of $\delta T/T$ shown in Figure 1 are based on a galaxy two-point correlation function that falls to $\xi_g = 1$ at $r = r_0 = 5\ h^{-1}$ Mpc and is assumed to vanish at $r > r_c \sim 30\ h^{-1}$ Mpc. There has been considerable

recent discussion of the fact that one can find holes and superclusters considerably larger than that (Einasto 1978; Oort 1981). This might lead one to ask whether the adopted ξ_g is too small, which would call into question the coincidence of the observed and expected values of $\delta T/T$. It seems worthwhile therefore to discuss the relation between the actual galaxy distribution and the measure of it by the n-point correlation functions. The theme of this section is that the visual impression of a map of the galaxy distribution can be a misleading indicator of the degree of long-range correlation. In particular, the new redshift samples do not seem to be inconsistent with the two-point correlation function used here.

A prescription for the space distribution of galaxies based on the observed angular correlation functions was proposed by Peebles and Groth (1975) and used by Soneira and Peebles (1978 hereinafter called SP) to make a model map that could be compared to the Lick map of the distribution of galaxies at $m < 19$. In the SP prescription the two-point correlation function is $\xi_g = 1$ at $r = 5\ h^{-1}$ Mpc, and $\xi_g = 0$ at $r > 20\ h^{-1}$ Mpc. We can use this same prescription to make models to compare to the maps from recent redshift surveys. In the following model the limiting magnitude has been adjusted to get about the right number of galaxies, and to reduce fluctuations I have eliminated the rare richest clumps (with 12 levels). These are the only adjustments to the SP prescription.

Kirshner *et al.* (1981 hereinafter KOSS) have found redshift distributions for the galaxies brighter than $R \cong 16.3$ in three fields with areas of two square degrees each. Their data for the three fields are shown in the top three histograms in Figure 2. The next five histograms are model results for five square fields with areas of two square degrees spaced at 5° center-to-center intervals along a line. The bottom histogram is the redshift distribution for the full model field, 10° by 40°, scaled to the mean number in a two square degree field. The results in Figure 2 are based on a consecutive string of numbers starting from the first in the computer's random number generator. The next numbers in the string were used to make the second model shown in Figure 3.

Particularly striking are the large gaps in the redshift histograms. In the models this is the result of the small-scale clustering of galaxies: where there is one galaxy there tend to be others nearby, and since redshifts appear in clumps it must leave gaps between the clumps. It will be noted that the distance between clumps is not directly related to the clustering length. It is a function of the mean space density of clumps. I conclude that the model histograms look reasonably similar to the data. Models with peculiar

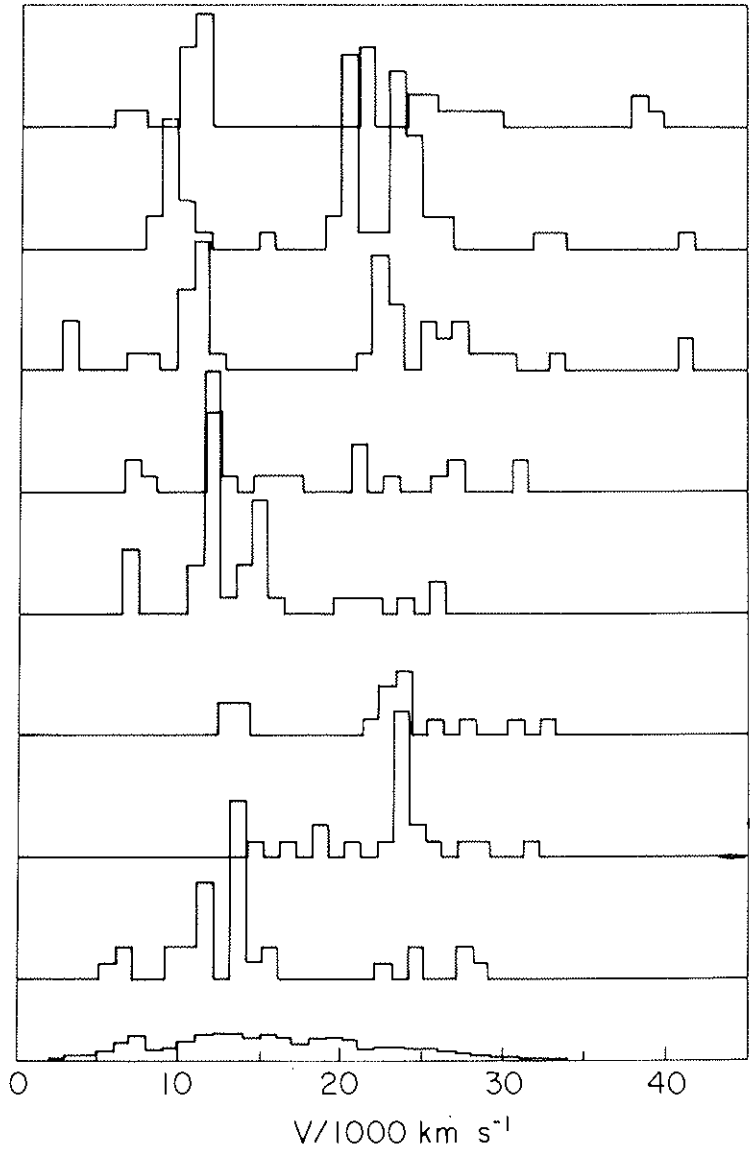


FIG. 2. Redshift histograms at the KOSS (Kirshner *et al.* 1981) depth. The top three histograms are observed redshift distributions in three KOSS fields. The next five histograms are from the model (see text, sec. 3). The bottom histogram is the mean for the 10° by 40° model field and normalized to a two square degree field.

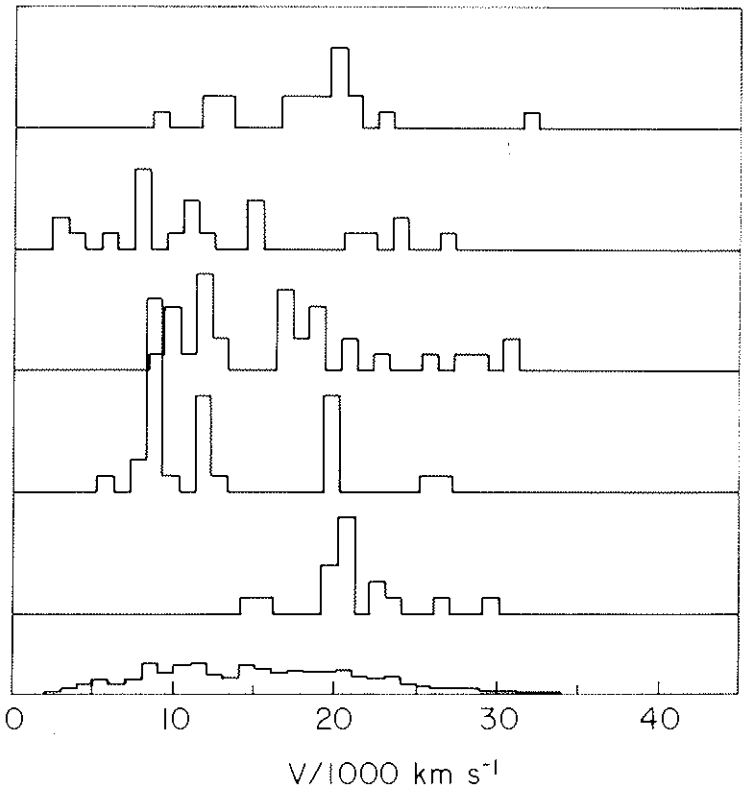


FIG. 3. Redshift histograms at the KOSS (Kirshner *et al.* 1981) depth. This shows the model results for a second set of random numbers.

velocities that can be compared to the Center for Astrophysics redshift maps at $m < 14.5$ will be discussed elsewhere in collaboration with M. Davis and R. Soneira. Here again the model is a not unreasonable approximation to the data although the isolated nature of the clumps is apparent in the model maps and somewhat objectionable.

The space distribution of the galaxies in the model in Figure 2 is shown in more detail in Figure 4. The maps show angular positions and distances of the model galaxies in fields 1° by 40° and hrighter than the model limiting apparent magnitude. The field center lines are 1.5° above and below the line of centers of the square fields in Figure 2. In this model galaxies have been placed in clumps with diameter $\sim 20\ h^{-1}$ Mpc. To make the correlation among galaxies vanish at $r > 20\ h^{-1}$ Mpc we had to place

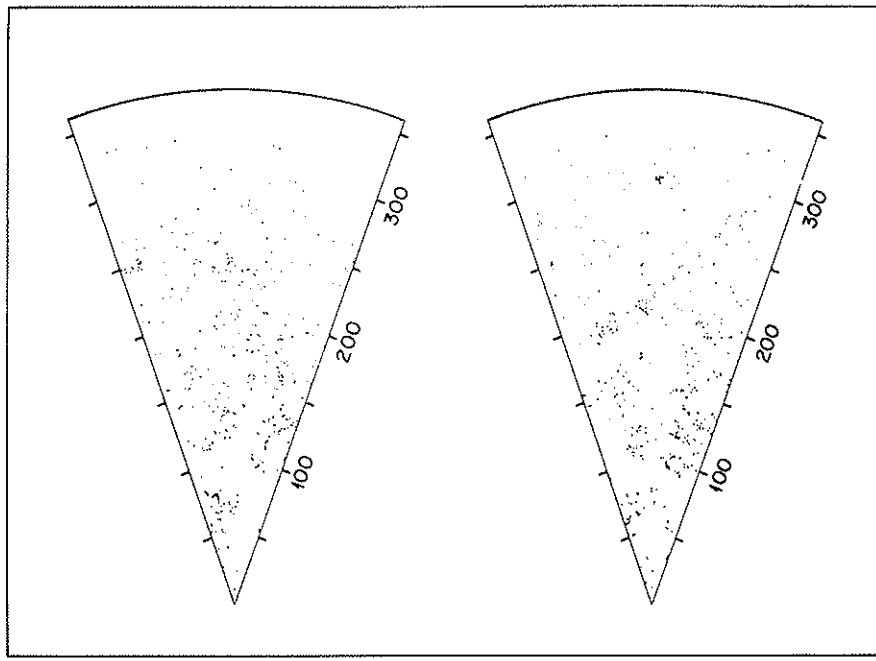


FIG. 4. Space distributions of galaxies in the model. The unit of the radial coordinate is 1 h^{-1} Mpc. The fields are 1° by 40°, and the centerlines of the fields are separated by 3°. This is the same model as in Figure 2.

the clumps at random. That means there must be the usual statistical accidents in how the clumps are placed, and as the eye is very sensitive to patterns one can readily pick out the accidents. Thus in the right hand map in Figure 4 one can make out a band of galaxies some 100 h^{-1} Mpc long running from lower left to upper right. Just below it is a roughly circular hole $\sim$ 50 h^{-1} Mpc across with a tight knot of galaxies just below the center. Neither of these features is apparent in the other map at $\sim$ 10 h^{-1} Mpc distance; they are statistical accidents in a tbin slice.

Configurations like those in Figure 4 may well exist in the real galaxy distribution. The models show that such real configurations need not mean there is large-scale correlation among galaxy positions. Indeed, if such arrangements could not be found it would suggest we were wrong in assuming there is negligible large-scale correlation. It appears therefore that the observations of large chains and holes in the distribution of galaxies do not necessarily tell us ξ_g has been underestimated.

This argument also suggests to me that we should be cautious in interpreting large-scale features in the galaxy distribution. If such features exist then of course they represent physical situations, but that need not mean they are the result of processes operating in a coherent way over large scales; they could be the accidents of a random process, as in the models.

4 - Anti-correlation and models for the origin of clustering

A second possibility to consider is that $\xi(r) < 0$ at $r \gtrsim 30\ h^{-1}$ Mpc so the integral J_3 to $x \sim cH_0^{-1}$ is much less than the value in equation (13). That means each mass concentration is balanced locally by a hole, as in the "swiss cheese" model of Kantowski (1969), so there is negligible large-scale mass fluctuation and thus no source for large-scale fluctuations in the microwave background. This situation arises if some process locally rearranged the originally homogeneous mass distribution. If the cohetence length for the rearrangement is x_c then $\xi < 0$ at $x \sim x_c$ and $J_3(x) \sim 0$ at $x \gg x_c$. We know that if J_3 initially is very small gravity keeps it small (even if the clustering is strongly non-linear on scales $\lesssim x_c$; Peebles and Groth 1976). Thus if we can convince ourselves that the observed large-scale fluctuations in the microwave background are not due to sources in our galaxy or to the primeval "isotbermal" fluctuations mentioned in Section 2 we will have evidence that $J_3(x)$ is about constant for radii $30\ h^{-1}$ Mpc $\lesssim x \lesssim cH_0^{-1}$ and so we will bave evidence against this local rearrangement picture.

As Ostriker describes at this conference he has shown there are some very attractive features in the idea that galaxies were produced by pressure forces. In the scenario of Ostriker and Cowie (1981) mass is redistributed by supernova explosions that pile up matter to make new stars and new supernovae. A somewbat similar scenario was discussed by Doroshkevich, Zel'dovich and Novikov (1967). Either version would predict negligible rearrangement of mass on scales $\gtrsim x_c \sim 30\ h^{-1}$ Mpc and so a negligible Sachs-Wolfe effect. The same is true of Press' idea that density fluctuations on the comoving scale x_c originated at bigh redshift from spontaneous stress fluctuations with coherence length x_c (Press 1981).

To produce tbe observed large-scale temperature fluctuations by the Sachs-Wolfe effect we need initial uncorrelated density perturbations, that is, a roughly flat spectrum. It might be useful to mention one *ad hoc* scheme that yields what is wanted. Suppose the entropy in the hot Big Bang model is produced at epoch z_i in an initially bomogeneous universe,

and suppose the energy (neglecting gravity) is somehow deposited in a strictly thermal, Gibbs distribution. This makes thermal fluctuations (Peebles 1968)

$$\frac{\delta M}{M} = \left(\frac{kT^2 \, c_V}{u^2 \, V} \right)^{1/2} , \tag{15}$$

in the volume V, where u is the energy per unit volume and c_V the heat capacity per unit volume. If u consists of a few species of nearly free relativistic quanta so $u \sim aT^4$ this flat initial spectrum grows to the wanted amplitude of density fluctuations at the present epoch if

$$z_i \sim 10^{22} \quad , \quad kT_i \sim 10^{18} \, eV \quad , \quad t_i \sim 10^{-24} \, \text{secs} \, . \tag{16}$$

In the cold Big Bang model discussed by Layzer and Hively (1973), Rees (1978), Hogan (1981) and others one obtains similar results for the microwave radiation produced at $z \sim 100$ by assuming the baryons are randomly placed at a suitably chosen initial redshift z_i, which again represents strictly thermal fluctuations.

5 - The clustering length problem

If the observed large-scale anisotropy of the microwave background is due to the Sachs-Wolfe effect it suggests the primeval density fluctuations were adiabatic with a roughly flat spectrum. The short wavelength part of an adiabatic spectrum is dissipated during the expansion of the universe. That fixes the coherence length of the mass distribution when the first generation of protogalaxies or protoclusters is forming. There is a problem, that the coherence length tends to be uncomfortably large compared to the present clustering length of galaxies (Press and Vishniac 1980, Peebles 1981c, Davis *et al.* 1981). As will be discussed here it appears that, if the universe is dominated by massive neutrinos, there are enough parameters that the coherence length can be reduced to an acceptable value.

Suppose first the major components of the universe are baryons, zero mass neutrinos (with Fermi energy $\ll kT$), and the primeval fireball radiation. Then if the primeval spectrum is flat the coherence length of the residual mass distribution after decoupling is (Peebles 1981c)

$$\lambda_s \sim 25 \, (\Omega_0 \, h^2)^{-0.65} \, \text{Mpc} \, . \tag{17}$$

This is the comoving length at which the mass autocorrelation function is half its value at zero lag. It is mainly determined by the matter-radiation Jeans length (eq. [4]). Now if the primeval mass distribution approximates a random Gaussian process then, when the rms density fluctuations have grown to $<(\delta\rho)^2>^{1/2} \sim <\rho>$, the first generation of objects is just forming and these objects have sizes comparable to λ_s. Thereafter gravity tends to increase the clustering so we would expect to find that the present mass clustering length is $r_0 > \lambda$. Since $r_0 \sim 5\ h^{-1}$ Mpc this condition is violated even in a dense universe, $\Omega_0 h^2 \sim 1$.

The problem can be stated more quantitatively using the following result. Though the evolution of non-linear density fluctuations in an expanding universe is difficult to deal with, the integral of the mass autocorrelation function satisfies the simple equation

$$J_3(x,t) = \int_0^x x^2\, dx\, \xi(x,t) \propto (D(t))^2 \ ,$$
$$\frac{d^2 D}{dt^2} + \frac{2}{a}\frac{da}{dt}\frac{dD}{dt} = 4\pi G\rho D \ . \tag{18}$$

The variable x is a comoving coordinate, so $x\,a(t)$ is the proper radius. The upper limit in the integral must be large enough that $J_3 \lesssim x^3/3$. For a derivation and tests of the equation in analytic and N-body models see Peebles and Groth (1976) and Efstathiou (1979). The equation is based on the fact that the integral J_3 gives the mean value of the mass in excess of random within distance x of a mass element. Non-linear motions on scales less than x rearrange the mass distribution around each mass element but have little effect on the net mass within x.

We can rewrite equation (18) as

$$J_3(x,t_0) = \xi(0,t)\ C\ [D(t_0)/D(t)]^2 \ ,$$
$$C = \int_0^x x^2\, dx\, \xi(x,t) / \xi(0,t) \ . \tag{19}$$

Here t_0 is the present epoch and ξ is the mass autocorrelation function at comoving separation x and epoch t. If t is chosen early enough that the density fluctuations are small then C is a function of x alone and can be

computed in linear perturbation theory (Peebles 1981c). If the initial spectrum of density fluctuations is flat, one finds

$$C(30\ \mathrm{Mpc}) \sim 5200\ (\Omega_0 h^2)^{-0.15}\ \mathrm{Mpc}^3\ . \tag{20}$$

We have from equations (13), (19) and (20)

$$\xi(0, t)\ [D(t_0)/D(t)]^2 \sim \xi(0, t_0) \lesssim 0.2\ . \tag{21}$$

This says the mean square value of the density fluctuations at the present epoch is $\xi(0, t_0) \lesssim 0.2$. That is, to get agreement with the observed value of J_3 we had to make the initial amplitude so small that bound systems would not yet have formed.

Several ways out might be considered. First, one could imagine the density fluctuations were non-linear prior to decoupling, which would vitiate the computation of $C(r)$. However, that requires that non-linear adiabatic perturbations exist at high redshift where the length scale of the density fluctuations exceeds the horizon. In this case only special initial conditions would prevent the dense parts from forming black holes rather than protoclusters, which seems unsatisfactory. Second, one can adjust the shape of the initial spectrum of density fluctuations, increasing the amplitude at short relative to long wavelengths to counter the effect of the dissipation at short wavelengths. The spectrum $|\delta_k|^2 \propto k^4$ yields a reasonably small coherence length if $\Omega_0 h^2 \gtrsim 0.1$. We would need to join this to a roughly flat spectrum at $\lambda \gtrsim 10$ Mpc if we wanted to account for the large-scale fluctuations in the microwave background by the Sachs-Wolfe effect. To prevent non-linear curvature fluctuations on small scales we would have to introduce another break to $|\delta_k|^2$ roughly flat at $\lambda \lesssim 1$ Mpc. (For further details see Peebles 1981c, § VI). With these adjustments we obtain a model that agrees with known constraints but seems contrived. Third, one could drop the assumption that the initial mass density approximated a random Gaussian process. If the initial power spectrum varies as k^n with $n > 1$ then the acoustic oscillations prior to decoupling mix phases and so make the residual mass distribution approximate a Gaussian process even if it was not Gaussian to begin with. If the initial spectrum is flat the residual distribution is dominated by the part at $\lambda \sim \lambda_J$ (eq. [4]) and so phase correlations would persist. We could therefore imagine that isolated protoclusters with comoving size $\sim \lambda_s$ (eq. [17]) form well before the rms contrast reaches unity. The problem now is converting the matter

between these protoclusters into galaxies; perhaps the explosion process discussed at this meeting by Ostriker would do.

An important recent development has been the recognition that the universe might be dominated by massive neutrinos (or some other weakly interacting massive particle). In this picture the thermal motions of the particles define a smoothing length ct_{nr}, where t_{nr} is the epoch at which the thermal energy becomes nonrelativistic. This has been discussed by several authors (Wasserman 1981 and references therein). As ct_{nr} is comparable to the galaxy clustering length for interesting choices of the parameters in the model, it is useful to get a more accurate value for the coherence length by numerically integrating the evolution of the neutrino distribution function. The results presented here assume the universe is dominated by one type of massive neutrino with $\beta = 2$ spin states, $\varepsilon = 4$ spin states times types of massless neutrinos, and matter plus radiation treated as an ideal fluid. The results are not much changed if β or ε are doubled. The distribution of massive neutrinos in phase space is computed as a function of the magnitude and direction of the momentum. The present value of the density parameter, Ω_0, is supposed to be dominated by the massive neutrinos.

Figure 5 shows the autocorrelation function of the residual mass distribution after the neutrino velocities have become unimportant and before the density fluctuations have grown nonlinear. The primeval density fluctuations are adiabatic and have a flat spectrum. The comoving separation is measured in units of megaparsecs at the present epoch. As indicated, lengths in the computation scale as $(\Omega_0 h^2)^{-1}$.

In the top curve the neutrino temperature is the conventional value

$$T_\nu = T_s = (4/11)^{1/3} T_r \ . \tag{22}$$

It is lower than the radiation temperature T_r because thermal contact between neutrinos and radiation is broken at redshift $z \sim 10^{11}$, and subsequently at $z \sim 10^{10}$ the thermal electron-positron pairs annihilate and add their entropy to the radiation. This model yields a coherence length $\sim 7\ (\Omega_0 h^2)^{-1}$ Mpc. If $\Omega_0 h^2 \sim 1$ this is comparable to the present galaxy clustering length. Also, if $\Omega_0 h^2 = 1$ we get $C\ (30\ \text{Mpc}) = 600$ Mpc^3, which is less than J_3 (eq. [13]). Though little room is left for the growth of clustering after the first generation forms this model might be just acceptable. If $\Omega_0 h^2 = 0.3$ then $C\,(30\ \text{Mpc}) = 5000$ Mpc^3 and we encounter the same problem as for the massless neutrino case (eqs. [20],

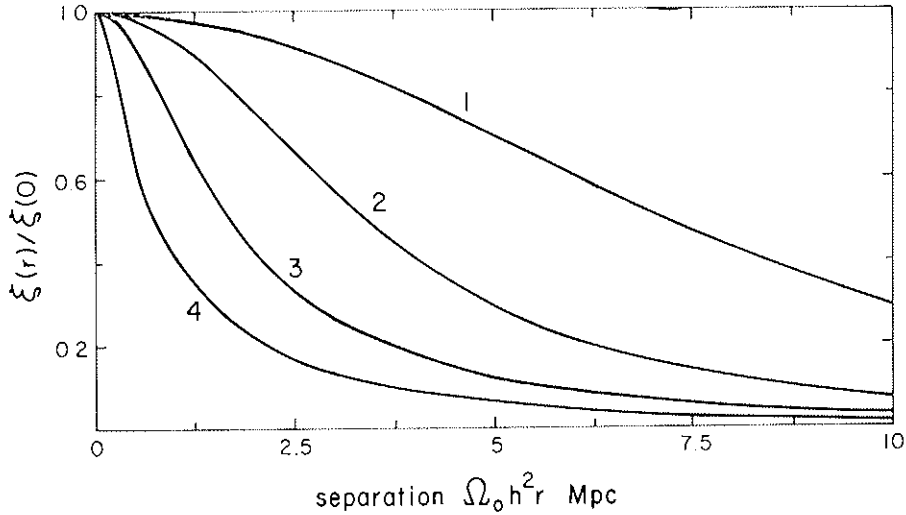


FIG. 5. Autocorrelation function of the residual massive neutrino distribution for a flat initial spectrum. In Curve 1, the neutrino temperature is the standard value T_s in Equation (22); in Curve 2, $T_\nu = T_s/2^{1/2}$; in Curve 3, $T_\nu = T_s/2$, and in Curve 4, $T_\nu = 0$. In all cases it is assumed that the massive neutrinos dominate the present mass of the universe.

[21]). As was discussed above we can reduce the coherence length by increasing the primeval density fluctuations on small scales relative to large, but then we must introduce two breaks in the spectrum, making it close to flat at short wavelengths to avoid diverging curvature fluctuations and also close to flat at long wavelengths to produce the Sachs-Wolfe effect. Another perhaps more attractive way out was suggested by Davis *et al.* (1981). They remarked that the length ct_{nr} is quite sensitive to the neutrino temperature and that T_ν might have been reduced by the decay or annihilation at $z \lesssim 10^{11}$ of possible undiscovered species of particles. The sensitivity of the coherence length to T_ν is shown by Curves 2 and 3 in Figure 5, where T_ν has been reduced from the value in equation (22) by factors $2^{1/2}$ and 2 respectively.

White (1981) pointed out that if $T_\nu \ll T_r$ the evolution of the power spectrum may be determined by another effect. The density contrast $\delta_\nu = \delta_\nu/\rho_\nu$ (defined on hypersurfaces of fixed proper time kept by freely moving observers) grows as $\delta_\nu \propto t$ until the wavelength becomes comparable to the horizon ct. Then pressure makes the radiation distribution oscillate as an acoustic wave, and the neutrino contrast δ_ν stays roughly

contrast until $\rho_\nu \sim \rho_r$ and the universe becomes dominated by the massive neutrinos. Since the epoch at which the radiation leaves δ_ν scales with the comoving wave number k as $t \propto k^{-2}$ the power spectrum is multiplied by k^{-4}, which strongly suppresses the short wavelength part. Curve 4 in Figure 5 shows the autocorrelation function of the final mass distribution for a flat primeval spectrum in the limit $T_\nu = 0$. Here the weakly interacting particles have been treated as a zero pressure ideal fluid; the particles could he very massive neutrinos or any other species that can he arranged to produce present mass density $\rho \sim \rho_{crit}$ and low enough pressure.

In the limit $T_\nu \lesssim T_s/2$ the coherence length of the residual mass distribution is not very sensitive to T_ν, and for a flat spectrum amounts to

$$\lambda_0 \sim (1.3 \pm 0.5)\ (\Omega_0 h^2)^{-1}\ \text{Mpc}\ . \tag{23}$$

If $\Omega_0 h^2 \sim 1$ this is comfortably less than the galaxy clustering length. Also, it is interesting to note that λ_0 is comparable to the mean distance between galaxies. Thus we could imagine λ_0 determined the scale of division of matter into galaxies.

6 - Summary

It is an interesting coincidence that if the mass clustering on scales $\lesssim 30\ h^{-1}$ Mpc were comparable to the observed fluctuations in the galaxy distribution and if there were little contribution to clustering from larger scales then the gravitational potential of the mass distribution would cause a background temperature anisotropy $\delta T/T$ comparable to what is observed. If the anisotropy were a local effect like radiation from our galaxy and the limit on the extragalactic anisotropy could be lowered a factor of 10 or so we would be led to consider several possible implications: mass may be less strongly clustered than galaxies; the mass distribution may be anti-correlated at larger scales so each mass concentration is balanced locally by a hole, or mass fluctuations may have appeared only recently so the radiation along different lines of sight did not pass through significantly different potentials. If there were appreciable clustering of galaxies on scales $\gtrsim 30\ h^{-1}$ Mpc in addition to what is seen on smaller scales then the expected Sachs-Wolfe effect would violate even the present limits on $\delta T/T$ and we would be led to consider the above tbree possible interpretations. However there is no evidence yet available that the galaxy two-point correlation function has been significantly underestimated.

If the microwave background does have irregularities $\delta T/T \sim 10^{-4}$ on angular scales of several tens of degrees two interpretations seem reasonably possible. The effect may be intrinsic, in the initially isothermal perturbations that become temperature fluctuations when $\rho_m \gg \rho_r$. That requires primeval matter density fluctuations $\delta\rho/\rho \sim 3 \times 10^{-5}$ on the comoving scale $\sim 3000\ h^{-1}$ Mpc, which seems high but perhaps could be arranged. The Sachs-Wolfe effect is negligible here because $\xi < 0$ at $r \sim \lambda_J$ (eq. [4], and the coincidence between the observed value of $\delta T/T$ and the computed values in Figure 5 would be only an accident. The second possibility is that the coincidence is significant, the large-scale irregularities in the microwave background being due to the gravitational potential of the shot noise of the mass distribution. It is of particular interest that this is inconsistent with a broad class of models for the origin of the clustering of matter in the universe, those where mass clustering originated from stresses coherent over scales $x_c \ll$ the present horizon. What is wanted is illustrated by the thermal model in equation (15); uncorrelated perturbations to the mass of the universe.

This research was supported in part by the National Science Foundation of the United States.

REFERENCES

Boughn, S.P., Cheng, E.S. and Wilkinson, D.T., 1981, *Ap. J.*, **243**, L 113.

Clutton-Brock, M. and Peebles, P.J.E., 1981, *A. J.*, **86**, 1115.

Davis, M., Lecar, M., Pryor, C. and Witten, E., 1981, *Ap. J.*, **250**, 423.

Doroshkevich, A.G., Zel'dovich, Ya. B. and Novikov, I.D., 1967, *Astron. Zh.*, **44**, 295; *Soviet Astronomy-AJ*, **11**, 233, 1967.

Efstathiou, G., 1979, *M.N.R.A.S.*, **187**, 117.

Einasto, J., 1978, in IAU Symposium No. **79**, *The Large-Scale Structure of the Universe*, eds. M. Longair and J. Einasto, Dordrecht: Reidel.

Hogan, C.J., 1980, *M.N.R.A.S.*, **192**, 891.

— 1981, preprint.

Kaiser, N., 1981, *M.N.R.A.S.*, in the press.

Kantowski, R., 1969, *Ap. J.*, **155**, 89.

Kirshner, R.P., Oemler, A., Schechter, P.L. and Shectman, S.A., 1981, *Ap. J.*, **248**, L 57.

Layzer, D. and Hively, R., 1973, *Ap. J.*, **179**, 361.

Melchiorri, F., Melchiorri, B.O., Ceccarelli, C. and Pietranero, L., 1981, *Ap. J.*, **250**, L 1.

Oort, J.H., 1981, *Astron. Astrophys.*, **94**, 359.

Ostriker, J.P. and Cowie, L.L., 1981, *Ap. J.*, **243**, L 127.

Peebles, P.J.E., 1968, *Nature*, **220**, 237.

— 1980, *The Large-Scale Structure of the Universe*, Princeton: Princeton University Press.

— 1981a, in Proceedings of the Erice Summer School, VII Course on *The Origin and Evolution of Galaxies*, eds. B.J.T. Jones and J. Jones; Dordrecht; Reidel.

— 1981b, *Ap. J.*, **243**, L 119.

— 1981c, *Ap. J.*, **248**, 885.

— 1982, *Ap. J.*, in press.

Peebles, P.J.E. and Groth, E.J., 1975, *Ap. J.*, **196**, 1.

— 1976, *Astron. Astrophys.*, **53**, 131.

Press, W.H., 1981, Paper presented at the Morinod Astrophysics Meeting.

Press, W.H. and Vishniac, E.T., 1980, *Ap. J.*, **236**, 323.

Rees, M.J., 1978, *Nature*, **275**, 35.

Sachs, R.K. and Wolfe, A.M., 1967, *Ap. J.*, **147**, 73.

Silk, J. and Wilson, M.L., 1981, *Ap. J.*, **244**, L 37.

Soneira, R.M. and Peebles, P.J.E., 1978, *Ap. J.*, **83**, 845.

Wasserman, I., 1981, *Ap. J.*, **248**, 1.

White, S.D.M., 1981, private communication.

DISCUSSION

SILK

1) The large scale anisotropy of the cosmic microwave background radiation has been computed recently by M. Wilson in weakly inhomogeneous models with $\Omega < 1$. He finds that while the dipole and quadrupole anisotropies are somewhat reduced if the normalization is unchanged an interesting effect enters associated with the radius of curvature of the universe. This effectively acts as a diverging gravitational lens that results in the generation of higher multipoles.

2) The dipole anisotropy can be appreciable for a flat spectrum of adiabatic fluctuations and sets important constraints.

3) One could argue (on philosophical grounds admittedly!) that if some initial spectrum of density fluctuations led to the formation of galaxies and clusters, nature would have to have been rather perverse to disguise all traces of smaller fluctuations on larger scales.

PEEBLES

1) No comment.

2) I like to separate the dipole moment with the intrinsic part and that due to our motion. I do not think the former is too big in a flat primeval spectrum. The latter is comparable to the expected rms value if $\Omega \lesssim 0.3$.

3) No comment.

WEINBERG

You have told us the expected magnitude of the fluctuation in background temperature over large scales. What do you expect the background temperature to look like as a function of angle when it is accurately measured? In particular, going back to Silk's question, what do you expect to find for the dipole moment, apart from our peculiar motion?

PEEBLES

I expect that the background temperature as a function of angle would

look like a random Gaussian process with spectrum proportional to $\sim k^{-3}$ so that large-scale fluctuations dominate and on them are superimposed smaller angular scale fluctuations with smaller amplitude. The intrinsic dipole moment is expected to be $\sim$ 10 percent of the observed value; the rest presumably is the result of our peculiar motion.

HAWKING

If you have white noise with $\delta m/m$ proportional to $m^{-1/2}$ you can explain the large-scale microwave anisotropy, but you do not have enough fluctuations on shorter length scales to produce galaxies. On the other hand, if you have $\delta m/m$ proportional to m^{-1}, you do not have enough large scale fluctuations to account for the microwave observations. Would an intermediate fluctuation spectrum $\delta m/m \propto m^{-2/3}$ account for both?

PEEBLES

In the spectrum $\delta m/m \propto m^{-2/3}$ the integral over the mass correlation function goes to zero at large r so I would expect this model yields $\delta T/T \ll 10^{-4}$ at $\theta \sim 90°$.

GUNN

You are saying, are you not, that there is *no* satisfactory power-law spectrum, since one needs a rising spectrum at high spectral frequencies which does not vanish at the origin?

PEEBLES

That is the case if the universe is dominated by the usual baryons, electromagnetic radiation and zero-mass neutrinos. If you allow low-pressure weakly interacting matter, I can get by with a power-law power spectrum.

DAVIS

I believe it is difficult to cool the neutrinos below their standard temperature of 1.9 K. If additional entropy is injected, say at $z \sim 10^7$ by the decay of some exotic elementary particles, then the resulting γ-rays will completely destroy the fragile deuterons before they thermalise with the blackbody radiation.

This result was first pointed out by Lindley (*Mon. Not. Roy. Astron. Soc.*, 1979, *188*, 15 P) a few years ago.

PEEBLES

In that case I might fall back on the GeV neutrinos which Lee and Weinberg discussed some years ago.

REES

You have mentioned a specific scheme whereby your preferred spectrum of fluctuations could have been set up at an early epoch. I would like to ask how much credence you would put in this, or any other, scheme which yields a power law. One could alternatively envisage that a "preferred mass" is fed in "ab initio" (after all, if $\Omega < 1$ there is in any case one preferred mass, the mass within the Hubble sphere at epochs corresponding to $z \lesssim \Omega^{-1}$). Also, an initial power law will translate into a post-recombination power law only on mass-scales exceeding the pre-recombination Jeans mass. Does this affect the way you normalise to the galaxy clustering?

PEEBLES

I understand that you mean by "preferred mass" a power spectrum of primeval density fluctuations that vanishes outside some fairly narrow range around co-moving wavelength on the order of 10 Mpc. In that case the power at $\lambda \sim 3000$ Mpc is negligible, so $J_3 \sim 0$, and one expects no Sachs-Wolfe effect.

I do not know of any good argument for a power-law power spectrum; but the $\delta T/T \sim 10^{-4}$ effect does rule out some otherwise attractive possibilities, namely, those where the mass distribution initially is homogeneous and that is broken by the local gathering of material on scales ~ 30 Mpc. The thermal fluctuation model was only meant as an example of what I think may be wanted. It is not meant to be taken at all seriously.

I normalise at $r \sim 30\ h^{-1}$ Mpc; the observed and expected $\delta T/T$ agree if the spectrum is flat longward of that.

GUNN

Since the coincidence you have been discussing depends upon the cutoff of

a divergent integral in a region in which the integrand is poorly, if at all, known, it is difficult to have much appreciation of the likely uncertainty involved. How large do you feel it is?

PEEBLES

The estimate of the integral $J_3(r)$ to $r = 30\ h^{-1}$ Mpc is based on the integral

$$\int_0^{\theta} \omega(\theta)\ \theta\ d\theta$$

where $\omega(\theta)$ is the angular galaxy two-point correlation function in the Lick sample. I think $J_3(30)$ is know to 30 percent. Direct measurements of the galaxy distribution do not provide useful constraints on the contribution to J_3 at $r \sim 3000\ h^{-1}$ Mpc and I am instead using $\delta T/T$ to constrain J_3.

III.

EVOLUTION OF GALAXIES

GALAXY FORMATION VIA HIERARCHICAL CLUSTERING AND DISSIPATION: THE STRUCTURE OF DISK SYSTEMS

S. M. FABER

Lick Observatory, Board of Studies in Astronomy and Astrophysics,
University of California, Santa Cruz

ABSTRACT

The White-Rees theory of galaxy formation via hierarchical clustering and core condensation is used to develop a model for the structure of disk galaxies. Disk parameters such as rotation velocity, radius and surface brightness scale directly in proportion to the corresponding halo parameters, which are in turn determined by hierarchical clustering from primordial density fluctuations. If the original fluctuation spectrum was a power law, scaling relationships among luminosity, velocity, and radius are predicted which look much like the observed Fisher-Tully and radius-luminosity laws. Evidence is presented which suggests that the slope of the fluctuation spectrum must have been flatter than white noise on a scale size from galaxies to clusters of galaxies.

I - INTRODUCTION

This paper and the one following discuss several properties of normal galaxies which appear to be intimately related to the nature of non-luminous halos and, through them, to the global properties of the early universe. In preparing these papers I have come to believe that there exists a direct structural link between the luminous portions of galaxies and their non-luminous halos. Appreciation of this close connection not only aids in understanding the structure and dynamics of present-day galaxies but also sheds

light on the density perturbations from which they came. It is this further connection that is of special relevance to cosmology.

The conceptual framework for this paper is a picture of galaxy formation based on dissipationless gravitational clustering of non-luminous halos, coupled with dissipational settling of gaseous matter within the halo cores. Originally put forward by White and Rees (1978), this picture is able to provide a natural explanation for many of the observed properties of galaxies and clusters (a brief review is given below). The idea of hierarchical clustering increases in power still further when coupled to the notion of angular momentum gain via tidal torques (Hoyle 1953, Peebles 1969, Efstathiou and Jones 1979, Fall and Efstathiou 1980). These concepts collectively serve as a point of departure for the present discussion.

In what follows a carefully defined vocabulary will be helpful. I use the term *halo* to refer to the non-luminous and dissipationless matter surrounding visible galaxies. Halos are assumed to be spherical, or perhaps slightly flattened. The words *luminous* and *ordinary* are used interchangeably to denote the dissipative matter which today constitutes the visible portions of galaxies, that is, the stars, stellar remnants and interstellar medium. This luminous component is further distinguished into two dynamically distinct sub-components, the spheroidal bulge and the disk. I often speak of elliptical galaxies and spheroidal bulges collectively as *spheroids*.

The aim of this and the following paper is to work out what the structure of the luminous portions of galaxies should look like in the context of the hierarchical clustering and dissipation picture. Power-law relations in radius, internal velocity and mass are derived, which look remarkably like the observed relations. These ideas also lead to a tentative picture for the nature of the Hubble sequence. This first paper presents the major concepts in broad outline and then concentrates on the structure of disk galaxies. The following paper is concerned principally with the formation of spheroids.

2 - Properties of present-day galaxies and clusters: A basis for further discussion

It is appropriate to begin by enumerating the most fundamental properties of galaxies. Such a summary serves not only as a framework for further discussion here but also as a platform for cosmological speculation at this conference. The following list of properties is a distillation down

to the bare essentials. It proceeds from well-established fact, through likely supposition, and ends with a speculative but potentially powerful hypothesis.

A. *Galaxies are easily discernible as discrete entities, whereas groups and clusters of galaxies are not.* Put another way, if one were to select a star at random from the universe, it would nearly always be possible to identify its parent galaxy with confidence. The analogous assignment of galaxies to groups and clusters is, in contrast, usually very uncertain.

B. *The luminosity function for galaxies is distinctly non-Gaussian with a long tail extending to low luminosities.* Fall (1980) points out that this form is typical of clustering or coagulation processes in general. It is significant, therefore, that the luminosity function for galaxies bears a familial resemblance to the multiplicity function for groups and clusters. Both of these in turn closely resemble the multiplicity function in N-body clustering simulations (Bhavsar, Gott and Aarseth 1981), which are of course built up in just this way.

C. *There are large peaks in the density distribution of non-luminous matter, and these in turn surround the visible portions of galaxies.* To put this idea quantitatively, the average density of non-luminous matter within groups and clusters is typically only $\sim 10^{-28}$ g cm^{-3}, as judged from virial estimates. In contrast, the density of matter at the optical radius of a spiral galaxy is $\sim 10^{-25}$ g cm^{-3}, most of which is known to be non-luminous. The density of non-luminous matter in the immediate vicinity of galaxies is thus some three orders of magnitude greater than its average density on the scale of groups and clusters.

D. *The specific angular momenta of galaxies correlate closely with optical morphology.* As the data in Table 1 indicate, ellipticals rotate the

TABLE 1 - *Specific angular momenta of galactic components.*[1]

Bare Spheroid (E Galaxy)	240 km s^{-1} kpc
Spheroid in Disk Galaxy	600 km s^{-1} kpc
Disk	4000 km s^{1-} kpc

[1] Values refer to "typical" luminous E, spheroidal bulge, and Sc spiral and are based on data in Burstein *et al.* (1981) and Kormendy and Illingworth (1981). They are mean values determined by integrating over the entire luminous component.

least, spirals the most, with spheroidal bulges somewhere in between. One would very much like to see these data presented for each Hubble type individually. However, to be meaningful, such a comparison would have to be carried out at constant total mass. Defining total mass is such a difficult problem at this stage that I have decided to bypass the issue and present only the very schematic comparison in Table 1. On such a gross level the differences are so well established that mass uncertainties cannot obscure the basic trend.

E. *Morphological types of galaxies are related to the density of galaxies in the immediate neighborhood.* This important fact was discovered by Dressler (1980) from studies of morphological types in 55 clusters.

F. *Radii and internal velocities of galaxies correlate well with galaxy luminosity.* These correlations take the form of power laws. For spirals the well-known Fisher-Tully relation states that $L \propto v^{\alpha}$, with α in the range 4 to 5. Similarly $L \propto R^{\beta}$, where β is close to 2, i.e., surface brightness is nearly constant. Essentially identical relations hold also for ellipticals (with different zero-points). This similarity in behavior hetween two families of galaxies having radically different internal structure is a remarkable fact in need of explanation.

The above assertions are all srrongly supported by the available evidence and may be considered astronomical "fact". Less well established is the following proposition:

G. *The ratio of luminous to non-luminous matter is everywhere constant on scales greater than 100 to 200 kpc.* This statement might at first seem at odds with what is well known about larger mass-to-light ratios seen in great clusters of galaxies (e.g. Faber and Gallagher 1979). This trend in large clusters would naturally lead one to suspect a continually increasing fraction of non-luminous matter on large scales. Part of the trend, however, can be attributed to the difference in mean stellar population in great clusters compared with small groups — the E and S0 galaxies prevalent in large clusters contain mostly old stars, which emit less light per unit mass.

It can further be plausibly argued that the remaining deficit is made up by the larger mass of hot intergalactic gas in great clusters. This material,

TABLE 2 - *The ratio of ordinary matter to total matter.*

	Hubble Constant 50 km s^{-1} Mpc^{-1}	100 km s^{-1} Mpc^{-1}
A. Small Groups		
1) Model M_{lum}/L_B for average spiral-dominated group[1]	2.9 ± 1.0	4.7 ± 1.6
2) Observed M_{tot}/L_B[2]	40^{-10}_{+50}	80^{-20}_{+100}
3) $F = M_{lum}/M_{tot}$	0.07 ± 0.05	0.06 ± 0.05
B. Coma Cluster		
1) Fraction of mass in stars:		
M_{lum}/L_B for old stellar populations[2]	6 ± 1	12 ± 2
M_{tot}/L_B for Coma Cluster[2]	325 ± 50	650 ± 100
M_{lum} (stars)$/M_{tot}$	$0.018 \pm .005$	$0.018 \pm .005$
2) Fraction of mass in IGM:		
Fraction of core mass in gas[3] ($= M_{lum}$ (gas)$/M_{tot}$)	0.07 ± 0.04	0.02 ± 0.01
3) $F = (M_{lum}$ (stars) $+ M_{lum}$ (gas) $)/M_{tot}$	$0.09 \pm .04$	0.04 ± 0.02

[1] M_{lum} includes stars, H I, and H_2. Relative proportions of Hubble types come from Dressler (1980), H I contributions from Roberts (1969). H_2 mass is arbitrarily set to 0.30 M_{HI}. M_{lum}/L_B for E's and S0's from Faber and Gallagher (1979). M_{lum}/L_B for Sd - Irr I is set equal to the solar neighborhood value of 1.5 (Faber and Gallagher 1979). M_{lum}/L_B for Sb - Sc is scaled from solar neighborhood value by addition of bulge with $M_{lum}/L_B = 6.0$. Final values assumed: E (8.0), S0 (6.0), Sb - Sc (2.0), Sd - Irr (1.5).

[2] Faber and Gallagher (1979).

[3] Lea *et al.* (1973).

if formed into luminous galaxies, would add substantially to the total light [I am indebted to a remark by Gunn (1980) that makes this point]. A crude comparison of the ratio of luminous to non-luminous matter in Coma and in small groups bears out this belief (see Table 2). As the notes show, this comparison rests on numerous assumptions and uncertain data. Nevertheless, within the limits of present knowledge, it would seem permissible to assume that the fraction of ordinary matter in large clusters is identical to that in small groups. From this point, it is but a short step to conclude

that the ratio of ordinary to total matter is everywhere constant on scales greater than a few hundred kiloparsecs.[1]

Proposition G is extraordinarily attractive for its essential simplicity and also for the uncomplicated picture it implies about the homogeneous distribution of ordinary and non-luminous matter in the early universe.

Since 100 to 200 kiloparsecs is comparable to the best present estimates for the size of galaxy halos (Ostriker, Peebles and Yahil 1974), Proposition G leads naturally to a further, more daring idea. There admittedly exists at present essentially *no* direct evidence for this further speculation. As with Proposition G, however, the potential power of simplification is so great that I put it forward as a useful working hypothesis:

H. *All protogalaxies initially contained the same ratio of ordinary to total matter.* From the data in Table 2, a reasonable estimate for this ratio, hereafter denoted by F, is roughly 0.07.

As noted by White and Rees (1978), the hypothesis of hierarchical clustering and dissipation (or HCD) is able to account for many of the above properties of galaxies in a completely natural way. The theory is notably successful in explaining the discrete nature of visible galaxies, the extended non-luminous envelopes, the galaxy luminosity function and the excess density of non-luminous matter in the environs of galaxies. When tidal torques are added to the theory, roughly the right order of magnitude is obtained for galaxy angular momenta. The idea of a constant ratio of luminous to non-luminous matter in all protogalaxies is also philosophically quite consistent with the HCD picture and would arise naturally under the simplest assumption of a well-mixed early universe. Six of the eight properties enumerated above are thus nicely subsumed under the theory.

Of the remaining two properties, one — the link between galaxy morphology and environment — is still a subject of ongoing research at present (Aarseth and Fall 1980, Silk and Norman 1981). Adding a new process, such as mergers, to the basic theory may be able to account for this correlation as well.

---

[1] Proposition G might also appear to conflict with Davis' finding (Press and Davis 1981; also Davis, this conference) that the virial mass *per galaxy* is larger in big clusters like Coma. The two views are not necessarily incompatible, however, if the number of galaxies created per unit mass was lower in dense clusters than elsewhere. This might happen, for example, if the formation of low-density galaxies was disrupted by the interaction of protogalaxies. In this view, the copious intergalactic medium in Coma would consist of gas which was not able to remain bound into individual galaxies.

The eighth and final property — structural scaling laws — has, to my knowledge, never been addressed in the context of the HCD picture. It is one of the principal themes of the present work.

3 - Hierarchical clustering and the cooling diagram

Rees and Ostriker (1977) have devised a simple diagram to identify graphically those protogalaxies in which gaseous dissipation seriously affects the dynamical structure. Roughly speaking, these are objects in which the cooling time is less than the dynamical time. Figure 1 is an updated version of their diagram showing the locations of present-day galaxies and clusters. Since just the ordinary matter is dissipative, only its density, n_{lum}, is plotted. For groups and clusters, n_{lum} was assumed to be F times

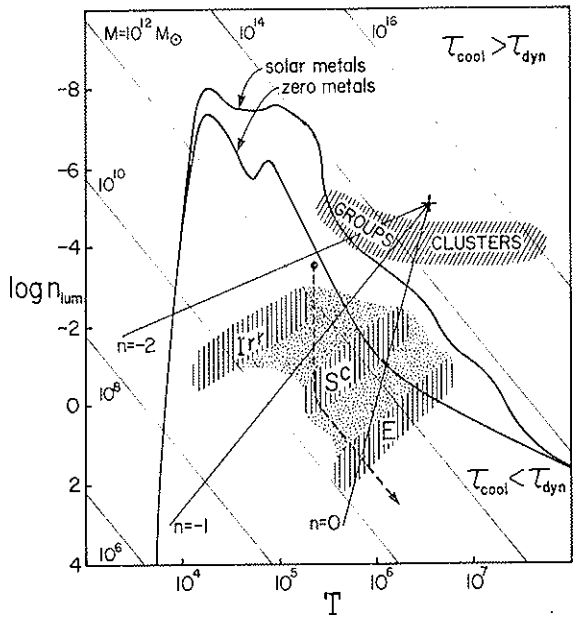


Fig. 1. Ostriker-Rees cooling diagram showing the actual locations of present-day galaxies and clusters. Data on galaxies came from the following sources - E's: σ's from Terlevich *et al.* (1981), M_{lum}/L from Faber and Gallagher (1979); Sc's: v_{rot}'s and radii from Burstein *et al.* (1981), M_{lum}/L_B assumed to be 1.6; Irr's: Thuan and Seitzer (1979). Data on groups and clusters from Rood and Dickel (1978). Cross is mean mass turning around today (White and Rees 1978). Heavy straight lines are clustering loci for various values of n. Light lines are the mass of a self-gravitating body composed purely of ordinary matter. Dashed line is a sample track for rapid dissipation within a heavy halo (see Figure 2).

the virially determined mass density. For galaxies, $n_{\rm lum}$ was derived as explained in the figure caption and refers to the mean density of luminous matter within the optical radii. The necessary data for this plot presently exist only for E's, Sc's, and irregular galaxies of low surface brightness, which occupy the hatched regions. One supposes that the neighboring regions (stippled) are occupied by intermediate Hubble types and will be filled in when the necessary data become available.

The two cooling boundaries refer to self-gravitating structures composed of purely gaseous ordinary matter having zero and solar metallicity. The dynamical temperature used as the abscissa is related to the mean kinetic energy per particle ($T \simeq 0.2\, m_h\, v^2\, k^{-1}$ for a fully ionized gas of pure hydrogen and helium). The masses shown (diagonal lines) are schematic only. They indicate the mass of a self-gravitating body of purely ordinary matter having that density and temperature. In objects where the non-luminous matter contributes significantly to the total mass density, the indicated mass must be multiplied by a correction factor $(n_{\rm lum}/n_{\rm tot})^{1/2}$.

According to Figure 1 all present-day galaxies appear to lie within the region where cooling is rapid relative to the dynamical timescale. The diagram thus quantitatively confirms the notion that dissipation was a major factor in determining the final luminous radii of galaxies and is thus responsible for their appearance as discrete, well-separated entities. Groups and clusters lie well apart from galaxies, above the cooling boundary. If the ordinary matter were by some process to become uniformly distributed within these structures, the cooling time would be long compared to the dynamical time. This is, of course, the minimum requirement needed to account for the long-lived intergalactic medium seen in groups and clusters.

It is worth emphasizing that the clear separation of three orders of magnitude in density between galaxies and groups and clusters is purely a result of focussing on the luminous matter. If $n_{\rm tot}$ were employed as ordinate instead of $n_{\rm lum}$, groups and clusters would move to *higher* densities by a factor of F^{-1}, or ~ 10. Galaxies on the other hand would be represented by their mean halo densities, which are some two orders of magnitude *lower* than the densities within the optical radii. The gap of three orders of magnitude in the present diagram would thus completely disappear, and we would see a smooth, unbroken procession in density and temperature from halos of small galaxies, through large galaxies, to groups and clusters. This continuum is naturally expected in the HCD picture, where galaxy halos and groups and clusters both originate from the same spectrum of initial density fluctuations via dissipationless clustering.

4 - The motion of protogalaxies in the cooling diagram

The cooling diagram becomes even more useful when used to trace out the tracks that clustering masses follow on their way to becoming galaxies and clusters. In the HCD picture clustering starts with small density enhancements somewhere in the lower left-hand region of the diagram. At every epoch there is a characteristic mass just then separating out from the background and turning around. One such point corresponds to the present epoch and has been estimated by White and Rees (1978). It is shown as the cross in the upper right.

Consider now the clustering as evidenced by the dissipationless component, and imagine that it begins at time t_0 with mass M_0 and temperature T_0. If the spectrum of initial density enhancements is a power-law (e.g. Peebles 1974, Gott and Rees 1975), we may write

$$\frac{\delta\rho\,(t_0)}{\rho\,(t_0)} = \left(\frac{M}{M_0}\right)^{-1/2\,-n/6} . \tag{1}$$

A value of $n < 0$ indicates more power on large scales than that expected for white noise.

If this initial power-law spectrum clusters hierarchically and dissipationless in a universe with $\Omega = 1$, it is easy to show (Gott and Rees 1975) that the structure and turn-around times of successively larger aggregates obey the following laws:

$$\rho = \rho_0 \left(\frac{M}{M_0}\right)^{-3/2\,-n/2} , \tag{2a}$$

$$R = R_0 \left(\frac{M}{M_0}\right)^{\frac{5+n}{6}} , \tag{2b}$$

$$v = v_0 \left(\frac{M}{M_0}\right)^{\frac{1-n}{12}} , \tag{2c}$$

$$T = T_0 \left(\frac{M}{M_0}\right)^{\frac{1-n}{6}} , \tag{2d}$$

$$t_{\text{turn}} = t_0 \left(\frac{M}{M_0} \right)^{\frac{3+n}{4}} . \tag{2e}$$

Under these conditions, the mean locus for clustering in Figure 1 is a straight line given by

$$n = n_0 \left(\frac{T}{T_0} \right)^{\frac{-3\,(3+n)}{1-n}} . \tag{3}$$

The exponent in this expression is quite sensitive to n. Slopes for $n = 0$, -1, and -2 are shown in Figure 1.

This locus applies to the non-luminous component only. In the HCD picture, the luminous and non-luminous components are initially well mixed. As long as there is no dissipation in the luminous matter, this matter also clusters along a straight-line locus, parallel to that above but down by a factor of F in density.

The process of dissipation itself consists of two phases. At first, when the gas has not contracted too far relative to the non-luminous component, the gravitational attraction of the non-luminous matter arises principally from the halo. Since the halo is essentially isothermal with constant v (R), the dynamical temperature of the luminous matter stays constant, despite the fact that its radius is shrinking.

After a collapse factor of roughly $F^{-1} \approx 10$, the density of the ordinary matter within its own radius begins to exceed that of the halo. The self-gravity of the ordinary matter now becomes dominant. Within this crossover point, the internal velocity of the luminous matter is given by

$$v_{\text{lum}} \propto \left(\frac{GM_{\text{lum}}}{R_{\text{lum}}} \right)^{1/2} \tag{4}$$

and hence

$$T_{\text{lum}} \propto R_{\text{lum}}^{-1} \propto n_{\text{lum}}^{+1/3} . \tag{5}$$

Dissipation in phase two is thus accompanied by an increase in dynamical temperature.

Figure 2 illustrates this process schematically. An isothermal halo with core radius one-tenth the outer radius is assumed. The luminous component has a mass 0.07 times the total mass. If dissipation is rapid

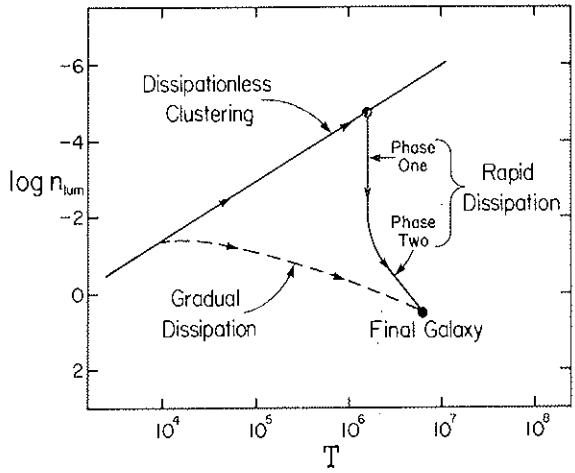


FIG. 2. Two possible dissipation tracks for luminous matter condensing to form a visible galaxy. The straight line is the clustering locus for ordinary matter without dissipation. The solid curved line shows a sample track when dissipation is much more rapid than clustering. The elbow marks the transition between the dominance of the halo potential (phase one) and self-gravitation of the luminous matter (phase two). The dashed line is an illustration of what might happen when dissipation and clustering occur contemporaneously.

relative to the clustering timescale, the halo properties may be considered essentially constant during collapse, and the collapse path is as given by the solid line. The change in slope marks the transition between phase one and phase two. If, as is more likely, the clustering and dissipation timescales are comparable, there will be a more gradual divergence of the luminous and non-luminous matter, and the precise path taken will depend on the dissipation history. One such possible path is shown schematically by the dashed line.

To summarize, the HCD picture in a universe with $\Omega = 1$ and a power-law initial density spectrum predicts that dissipationless clustering will follow a straight-line locus in the cooling diagram, the slope of which depends only on the power-law exponent. Non-luminous halos of galaxies and groups of galaxies should be found along this track. The luminous portions of galaxies depart from this track via dissipative processes, moving downward to higher densities and, ultimately, also to higher dynamical temperatures.

These arguments can be used to place an interesting upper limit to the power-law exponent n. Note that the mean clustering locus must

intersect the location of the mean group collapsing today. Three possible loci are shown in Figure 1. Recall also that the mean clustering locus must lie *above* the region occupied by galaxies. These considerations appear to rule out values of n as low as 0 since the clustering track would then lie below too large a fraction of present-day galaxies.

It is difficult to set an exact upper limit to n with this type of argument owing to an additional complicating factor that the clustering locus has finite width. Such a spread might arise from stastistically random variations in the density of individual fluctuations at fixed dynamical temperature (see Section 5 below). The sizeable width in the area occupied by groups and clusters in Figure 1 suggests that this spread might be considerable.[2]

Even if the clustering locus were imagined to have a width fully equal to that displayed by groups and clusters, however, a value of $n = 0$ would still be incompatible with the location of galaxies; with this value, most irregular galaxies would still lie above the track. A value of $n = -1$, on the other hand, would seem to be marginally compatible with the irregulars while a value $n = -2$ would be fully compatible. These conclusions apply if $\Omega = 1$. If Ω is less than unity, the clustering locus is no longer a straight line but curves upward (White and Rees 1978). For Ω even as small as 0.2, however, this curvature is small and has little effect. Its sense is such as to reduce the above upper limit on n still further.

To summarize, if the power-law hypothesis is valid, a reasonable upper limit to n would seem to be -1. This value implies that, on mass scales from galaxies to groups and clusters, the power-law spectrum of the universe is flatter than white noise. There seems to be growing evidence for such a flat power-spectrum from observations on larger length scales as well. Examples include the large-scale distribution of galaxies (Kirshner *et al.* 1981, Davis, this conference) and the multiplicity function for groups and clusters (Bhavsar *et al.* 1981).

5 - The Hubble sequence: Dissipation or density sequence?

It is only a short step from the preceding discussion to the beginnings of an idea for the formation of the Hubble sequence. Because dissipation

[2] Note, however, that the indicated range for groups and clusters in Figure 1 is illustrative only; the data used to identify this range are in no sense statistically representative.

carries galaxies downward from the clustering locus, it would be natural to conclude that objects lowest in the cooling diagram have dissipated and contracted the most. In this picture, the Hubble sequence would be a sequence in the amount of collapse and condensation of the luminous matter relative to the non-luminous halo.

I shall ultimately embrace this view, but only after much discussion here and in the following paper. The way to this conclusion is not as clear as the simple logic above would imply, for there is a competing explanation for the Hubble sequence which deserves careful consideration. This alternative idea is based on the notion that there must be some intrinsic width to the clustering locus owing to statistical fluctuations. For an *individual* inhomogeneity, for example, Eq. (1) would read

$$\frac{\delta\rho}{\rho} = \eta \left(\frac{M}{M_0} \right)^{-1/2\ -n/6} . \tag{6}$$

The parameter η here expresses the deviation of a particular inhomogeneity from the mean density enhancement of all inhomogeneities of that mass. Eq. (3) now becomes

$$n = n_0\ \eta^{\frac{12}{1-n}} \left(\frac{T}{T_0} \right)^{\frac{-3\ (3+n)}{1-n}} . \tag{7}$$

With n in the range -1 to -2 (see above), the exponent of η lies between 4 and 6. Small random fluctuations in η might thus induce considerable width into the clustering locus.

We now ask whether such an occurrence might not have produced the entire spread in galaxy densities in Figure 1. If so, ellipticals would be dense today simply because they evolved from perturbations whose initial η was greater than average. Under this hypothesis the Hubble sequence would be interpreted primarily as a sequence in η.

For clarity, let us call these two ideas the "dissipation" and "density" pictures, respectively. In my opinion, the available data on the structure of galaxies somewhat favor the dissipation picture, although no clear discrimination between the two processes is really possible at this time. Indeed, both may operate at some level.

The evidence supporting the dissipation scheme will take some time to assemble. Much of it hinges on the structure of spheroids and is there-

fore deferred to the following paper. The remainder of this paper is devoted largely to further exploration of the dissipation hypothesis and to comparison between it and the observational data.

One of the more important features of the dissipation picture is its prediction that the fraction of non-luminous matter within the radius of the ordinary matter should he greater in late-type galaxies than in early types. This behavior should arise from the lower condensation of late-type galaxies within their surrounding halos. As emphasized by Tinsley (1981), observed mass-to-light ratios of galaxies along the Hubble sequence do indeed display such a trend. Table 3 presents quantitative estimates similar to Tinsley's. For practical reasons, the radius of the ordinary matter has been set equal to an isophotal optical radius. This definition is satisfactory for the spirals but considerably underestimates the radii of irregulars, in which a great deal of gas is often present outside the optical image.

The trend of the data in Table 3 is consistent with the dissipation picture, in which the Hubble sequence is a sequence of increasing luminous concentration. The luminous bodies of Sa galaxies appear to be strongly concentrated within their halos. Sc's are transition objects in which the fractions of luminous and non-luminous matter witbin the optical radii are comparable. This conclusion is consistent with the location of Sc's in

TABLE 3 - *Fraction of luminous and non-luminous matter within the optical radius.*

Type	Dynamical M_{tot}/L_B[1]	M_*/L_B[2]	M_{gas}/L_B[3]	$Fraction_{lum}$	$Fraction_{non}$
Sa	6.5:	5.5	0.13	0.86:	0.14:
Sc	(3.5 - 5.0)	1.1	0.52	(0.32 - 0.46)	(0.54 - 0.68)
Irr I	> (2.5 - 3.2)	0.25	0.65	< (0.28 - 0.36)	> (0.64 - 0.72)

[1] Values refer to optical Holmberg radius. Sa: Faber and Gallagher (1979); Sc: larger value is raw from FG, smaller contains additional correction for internal absorption (Burstein, private communication); Irr I: smaller value is mean from Fisher & Tully, quoted by FG, larger value from Thuan and Seitzer (1979). Both values for Irr I's are lower limits owing to the fact that the optical radius may systematically underestimate the radius of the ordinary matter (see text).

[2] Model M_*/L_B from Tinsley (1981), Larson and Tinsley (1978), at appropriate $(B-V)_0$.

[3] M_{HI} from Roberts (1969) for Sa and Sc and from Thuan and Seitzer (1979) for Irr I, increased by 30% for H_2.

Figure 1, near the elbow of the dissipation track. This elbow marks the transition to self-gravitation in the luminous matter, and hence also the point where the densities of luminous and non-luminous matter are equal within the optical radius.

The status of irregulars as diffuse, low-concentration objects would be even clearer if 21-cm rather than optical radii had heen used. The Gaussian line profiles and lack of flattening in irregulars (Thuan and Seitzer 1979) suggest that they are at least partially supported by random motions rather than by rotation. These low-surface-brightness irregulars may be small galaxies whose collapse and dissipation have been delayed by energy input from young stars.

A second prediction of the dissipation picture comes from the rising dynamical temperature of the luminous matter within the self-gravitating regime. As a result of this process, at constant mass we would expect internal velocities in early-type galaxies to be higher than in late types. The observations of rotational velocities in early Hubble types do indeed show such a trend (Rubin, Ford and Thonnard 1980). In the following paper this effect is estimated quantitatively and shown to he consistent with the observational evidence.

6 - The Fisher-Tully relation for Sc spirals

We turn now to the important question of scaling laws in luminosity, radius and internal velocity. Since Sc spirals in the dissipation picture have not contracted very far, the Fisher-Tully relation for them turns out to be particularly simple. We start there and move on to other Hubble types and other scaling laws in the following section.

Attempts to explain the Fisher-Tully relation have concentrated thus far on the behavior of disks themselves. Schechter (quoted by Aaronson, Hucbra and Mould 1979) observed that the relation $L \propto v^4$ is expected for disks of constant surface brightness and constant M/L in dynamical equilibrium. Carrying the argument further, Silk and Norman (1981) invoked cloud-cloud collisions, viscosity, and other gaseous processes to account for the observed constant central surface brightness.

A critical look at Sc galaxies suggests that this focus on the micro-physics of disks may be the wrong approach. We have seen that within the optical radii of Sc's, luminous and non-luminous matter contribute about equally. This conclusion is supported not only by the data in Table 3

but also by stability arguments for disks themselves, which require a massive, stabilizing halo (Ostriker and Peebles 1973). (Since Sc's have only small spheroidal bulges, this halo clearly must consist of non-luminous matter). It therefore seems a mistake to focus solely on the physics of the luminous matter, which is the source of at most only half the total gravitational force.

This argument can be buttressed further. In several late-type spirals, rotation curves can be followed in neutral hydrogen to two or three times the optical radii. Rotation velocities are found to be essentially constant over this range. At these large radii it is prohable that the circular velocity of the halo itself is being sampled. If these galaxies are any guide, then, the rotational velocity at the optical radius of Sc spirals must be essentially equal to *that of the surrounding halo itself.* This result is also expected theoretically if Sc's are transition objects not yet fully in the self-gravitating regime (see the vertical dissipation track in Figure 2).

The import of these arguments is clear: the rotation velocity which figures in the Fisher-Tully relation for Sc spirals must reflect the structure of the halo rather than that of the disk. Eqs. (2) for the structure of the dissipationless halo now apply, and Eq. (2c) becomes

$$M_{tot} \propto (v_{rot})^{\frac{12}{1-n}} . \tag{8}$$

Within a given Hubhle type, the assumption of constant M_{lum}/L may be reasonably accurate. Constant F then implies constant M_{tot}/L. We obtain finally

$$L \propto (v_{rot})^{\frac{12}{1-n}} , \tag{9}$$

a power-law relation which looks like the Fisher-Tully law.

The observed value of the exponent in the Fisher-Tully law is not accurately known. Burstein *et al.* (1981) recently determined a value of 5.7 for Sc galaxies. This is an upper limit, however, owing to the manner in which the residuals were minimized. The true value is probably in the range between 4 and 5. The corresponding limits on n are -2.0 to -1.4, in agreement with the upper limit of -1 set in Section 4. If this explanation for the Fisher-Tully law is correct, we have still further evidence for a relatively flat spectrum of initial density inhomogeneities.

7 - The role of angular momentum

i) *The optical radii and surface brightnesses of spiral disks*

If galaxies of different Hubble types collapse and dissipate by differing amounts, as is required in the dissipation picture, the question naturally arises as to what mechanism controls the extent of the collapse. As suggested by Fall and Efstathiou (1980), the collapse of disk galaxies is most probably halted by their internal angular momentum. In this picture the final radius of the luminous disk is thus determined by the amount of angular momentum absorbed by the protogalaxy.

In the initial stages of protogalaxy condensation, both luminous and non-luminous matter are spun up by the tidal torques of neighboring protogalaxies. In Peebles' (1971) notation, the quantity $\lambda \equiv \mathcal{L}\, E^{1/2}\, G^{-1}\, M^{-5/2}$ is a dimensionless measure of this angular momentum (here $\mathcal{L}$ is the total angular momentum, E the energy and M the mass). Analytic estimates (Peebles 1969), as well as numerical simulations (Peebles 1971, Efstathiou and Jones 1979), suggest that the mean value of λ expected in the HCD picture is ~ 0.07.

For disk galaxies the collapse factor and final disk radius are uniquely determined by λ. Fall and Efstathiou (1980) originally derived this result from numerical models, but it can be obtained easily from simple arguments. Let R_H be the halo radius, R_D the final disk radius, and v_D the circular velocity at the edge of the disk. Assume initially that the luminous mass is so small that its self-gravity can be neglected. The velocity v_D is then also v_H, the circular velocity in the halo. Define a dimensionless radial coordinate $X \equiv R/R_H$ and set $X_0 \equiv R_D/R_H$. The non-luminous mass within radius X can be written $f_H(X)\, M_{tot}$, where f_H is a dimensionless function. We assume also that the halo mass distribution, f_H, is unaffected by the inward redistribution of the luminous matter, i.e. a rigid halo potential.

The final radius of the disk is then such that

$$R_D\, v_D^2 = G\,(M_{lum} + f_H(X_0)\, M_{tot}) \approx G M_{tot}\, f_H(X_0)\ . \qquad (10)$$

If the specific angular momentum of the disk is constant as it falls in, we also have

$$(R_D\, v_D)\, v_D = \frac{\mathcal{L}_D}{M_{lum}}\, v_D = \frac{\mathcal{L}_{tot}}{M_{tot}}\, v_H = 2^{1/2}\, \lambda\, G M_{tot}\ . \qquad (11)$$

Comparison of Eqs. (10) and (11) yields the implicit solution for X_0:

$$f_H(X_0) = 2^{1/2}\ \lambda\ . \tag{12}$$

If λ is large and the collapse factor not too great, in an isothermal halo $f_H(X) \approx R/R_H$. Thus

$$R_D \approx 2^{1/2}\ \lambda\ R_H \tag{13}$$

(see Fall and Efstathiou 1980). Rapidly rotating galaxies therefore fall in only slightly, whereas slowly rotating galaxies fall in a long way.

The above argument requires only slight modification if some of the simplifying assumptions are relaxed. If the self-gravity of the disk cannot be neglected, for example, Eq. (12) becomes

$$\frac{R_D}{R_H} = X_0 = \frac{2\ \lambda^2}{F + f_H(X_0)}\ . \tag{14}$$

Provided f_H is the same in all galaxies of a given mass, we have once again an implicit solution for X_0 which depends only on λ. Write this solution as

$$R_D = g(\lambda)\, R_H\ . \tag{15}$$

It is also straighforward to derive the related result:

$$v_D = h(\lambda)\, v_H\ . \tag{16}$$

(In the previous section we found, for example, that, for Sc galaxies, $h(\lambda) \approx 1$).

These equations are examples of a general rule. For given λ, F and M_{tot}/L, each disk parameter is directly proportional to its corresponding halo parameter. Scaling laws among them may thus be derived directly from Eqs. (2). As pointed out by Burstein and Sarazin (in preparation), the radius-luminosity law for disks is a particularly nice example. To see this, recast Eq. (2b) in the form

$$L \propto R_D^{\frac{6}{5+n}}\ . \tag{17}$$

If n is set equal to $-5/3$ (corresponding to a Fisher-Tully exponent of 4.5), one finds $L \propto R_D^{1.8}$, the observed radius-luminosity law for Sc spirals (Burstein *et al.* 1981).

It is also instructive to consider disk surface brightness, σ_D, as a function of luminosity. We find via Eq. (2c) that

$$\sigma_D \propto L^{\frac{-n-2}{3}} . \tag{18}$$

With $n = -5/3$, $\sigma_D \propto L^{-0.11}$, a very weak dependence. This result is noteworthy because σ_D has figured so prominently in attempts to explain the Fisher-Tully law. In the HCD picture the near-constant surface brightness of disks derives from the near-constant surface mass density of halos rather than from any internal physics of disks themselves.

To summarize, this section suggests that angular momentum is the primary factor controlling the extent of the collapse and dissipation in the luminous component. To reproduce the Fisher-Tully relation we require λ to be constant for a given Hubble type. Sc galaxies in this picture have high angular momenta and did not collapse very far. Ellipticals and spheroidal bulges in contrast are low angular momenta systems which collapsed to high densities. The data on specific angular momenta in Table 2 are consistent with this idea.

It is interesting to note that angular momentum variations seem to fit quite naturally into the dissipation picture for the Hubble sequence, in which differing amounts of central condensation in the luminous matter are required. There is no analogous role for λ variations in the density picture, where all Hubble types in the mean have the same degree of luminous concentration. Indeed, λ variations are an actual hindrance in the density scheme, since they introduce scatter into any predicted structural scaling laws. It is to be counted a small plus for the dissipation picture that it can turn the inevitable angular momentum variations to good advantage.

The basic idea of this section — that the scaling laws reflect the *halo* properties rather than the internal physics of disks — would be true in both pictures, however. The main difference in the density scheme is that the functions $g(\lambda)$ and $h(\lambda)$, which reflect the central condensation of the luminous matter, do not vary systematically with Hubble type.

ii) *The independence of* λ *and* η_i

The parameter η_i was introduced in Eq. (6) to describe the initial overdensity of individual density fluctuations. In this section, a potential

coupling between λ and η is examined. Such a coupling might be anticipated since denser perturbations have shorter collapse times and thus less time for tidal torquing by their neighbors. They also expand less and have smaller quadrupole moments. Both of these factors could tend to reduce angular momentum transfer.

These effects may be estimated by applying Peebles's (1969) analytic estimates of tidal torques to perturbations of differing initial overdensity. Somewhat surprisingly, the analysis indicates that λ is independent of η. Since the results are negative, only a short summary is given in the Appendix.

If this analysis is correct, it would appear that variations in angular momentum could not have arisen from differences in initial overdensity but instead from statistical variations in the quadrupole moments of protogalaxies and in the strength of their nearest-neighbor interactions. Furthermore, because the tidal -torque mechanism provides no coupling between η and λ, the effect of η andλ differences on Hubble type should remain separate and distinct.

8 - Summary

In this paper I have attempted to develop further the theory of galaxy formation via gravitational clustering and dissipation and, thereby, to sharpen comparison between theory and observation. From the work of White and Rees (1978) it is already known that the theory is naturally compatible with many of the hasic properties of galaxies — their discreteness, extended non-luminous halos, angular momenta, and luminosity function. With three new assumptions added to the basic HCD picture, it has been possible to make considerable further progress. These assumptions include the idea that all protogalaxies began with a similar ratio of luminous to non-luminous matter and that λ and M_{tot}/L are, in the mean, constant within a given Hubble type. These ideas in turn imply that the luminosity, radius, internal velocity and angular momentum of visihle galaxies all scale in proportion to the corresponding halo quantities. The visible galaxy can thus be used as a direct probe of the clustering process itself. It is this notion that is of interest to cosmology.

Under the further assumption of a power-law density fluctuation spectrum, it is possible to derive power-law relationships for visible galaxies that closely resemble the ohserved Fisher-Tully and luminosity-

radius laws. To get a good match to the observed power-law exponents, a rather low value of n between -1 and -2 seems to be required, implying excess power above white noise on large scales. Further evidence for a flat power spectrum comes also from the location of galaxies relative to groups and clusters in the cooling diagram (Figure 1). This trend toward more power on large scales seems to agree with other recent results from the multiplicity function and from large-scale surveys of galaxies.

Despite the simplicity and attractiveness of a power-law clustering model, many of the conclusions of this paper are not inextricably wedded to this assumption. For example, Blumenthal and Primack (1981) have attempted to account for the observed Fisher-Tully exponent without recourse to an assumed power law. Their theory is based on gravitational instability in a medium of dissipationless gravitinos. Because their theory yields $M \propto v^5$ directly over a very wide mass range, clustering of their initial density fluctuations would follow a straight-line locus with the right slope in Figure 1. Most of the ideas here would thus carry through equally well with gravitinos, including the derivation of the Fisher-Tully and radius-luminosity scaling laws.

An important question was raised but left unanswered concerning the nature of the Hubble sequence. Two interpretations of the sequence seem possible in the HCD picture — one of dissipation and relative concentration of the luminous matter, the second of initial overdensity η. The choice between these possibilities is intimately bound up with the structure and formation of spheroids. These issues are the principal concerns of the following paper.

I would particularly like to thank P.J.E. Peebles and James Gunn for helpful discussion at the conference, in which several errors in my original ideas came to light. I would also like to thank David Burstein and Craig Sarazin for permission to quote their work prior to publication. Many lengthy discussions with David Burstein added greatly to the present paper. Vera Rubin, Kent Ford, Norbert Thonnard and David Burstein deserve special mention for their extensive observations and incisive analysis of spiral rotation curves. Although not extensively referenced in this final version, their work in fact provided the original inspiration for many of the arguments both here and in the following paper. Finally, I particularly thank my Lick Observatory lunchtime colleagues George Blumenthal, Peter Bodenheimer, William Burke and Douglas Lin for their active assistance in the development of these ideas. Much of this work was completed while the author held an Alfred P. Sloan Foundation Fellowship.

APPENDIX

THE LACK OF DEPENDENCE OF λ ON η

According to Peebles (1969), there are two stages in angular momentum transfer to protogalaxies. The first stage occurs while the perturbation is still growing linearly, the second after the perturbation is well-separated from the surrounding medium and is torqued up according to its quadrupole moment, Q. These two phases are represented by the rising and falling lines in Figure 3. The angular-momentum transfer rate during the transition period presumably looks something like the dotted line. Let the intersection point of the two straight lines be $(t_c, \mathcal{L}_c)$. The total angular momentum transferred is clearly proportional to $\mathcal{L}_c$.

Peebles noted that the time of peak angular momentum transfer ($\sim t_c$) occurs somewhat before the turn-around time for the perturbation, $t_{\rm turn}$, and scales directly with it. To find $\mathcal{L}_c$, one then simply scales the expressions for either stage one or two to a different value of η and evaluates at $t = t_c$. The expression for stage two is the easier one to consider. In stage two, the torque τ is proportional to

$$\tau \propto \frac{Q}{t^2}\,, \tag{A1}$$

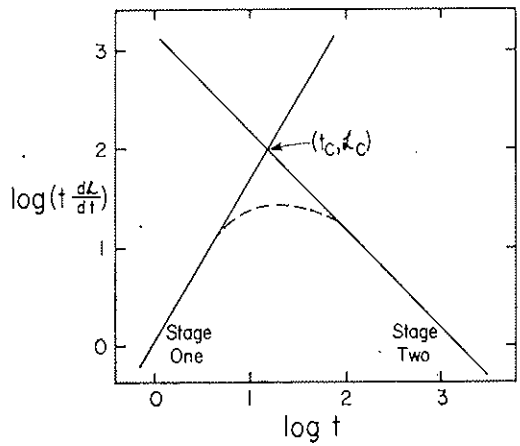


FIG. 3. Angular momentum transfer rate as a function of time in the tidal torque theory.

with $Q \propto MR^2$, R the radius of the protogalaxy. Generalizing Eq. (2b) to include η yields

$$R \propto M^{\frac{5+n}{6}} \eta_i^{-1} . \tag{A2}$$

Hence

$$t \frac{d\mathcal{L}}{dt} \simeq t\tau \propto \frac{Q}{t} \propto M^{\frac{3+n}{3}} \eta_i^{-2} t^{-1} . \tag{A3}$$

Similarly, Eq. (2e) becomes

$$t_{\text{turn}} \propto M^{\frac{3+n}{4}} \eta_i^{-3/2} . \tag{A4}$$

Thus

$$\mathcal{L}_e = t \frac{d\mathcal{L}}{dt}_{t=t_c} \propto M^{\frac{23+n}{12}} \eta_i^{-1/2} . \tag{A5}$$

With $\lambda \equiv \mathcal{L} E^{1/2} G^{-1} M^{-5/2}$, we have

$$E \propto \frac{M^2}{R} \propto M^{\frac{7-n}{6}} \eta , \tag{A6}$$

and

$$\lambda \propto M^0 \eta_i^0 . \tag{A7}$$

The parameter λ is therefore independent of both M and η_i.

REFERENCES

Aaronson, M., Huchra, J. and Mould, J., 1979, *Ap. J.*, **229**, 1.

Aarseth, S.J. and Fall, S.M., 1980, *Ap. J.*, **236**, 43.

Bhavsar, S.P., Gott, J.R. and Aarseth, S.J., 1981, *Ap. J.*, **246**, 656.

Blumenthal, G. and Primack, J., 1981, preprint.

Burstein, D., Rubin, V.C., Thonnard, N. and Ford, W.K., Jr., 1981, preprint.

Dressler, A., 1978, *Ap. J.*, **226**, 55.

— 1980, *Ap. J.*, **236**, 351.

Efstathiou, G. and Jones, B.J.T., 1979, *M.N.R.A.S.*, **186**, 133.

Faber, S.M. and Gallagher, J.S., 1979, *Ann. Rev. Astron. Astrophys.*, **17**, 135.

Fall, S.M., 1980, *Phil. Trans. Roy. Soc. London* (*Ser. A*), **296**, 339.

Fall, S.M. and Efstathiou, G., 1980, *M.N.R.A.S.*, **193**, 189.

Gott, J.R. and Rees, M.J., 1975, *Astron. Astrophys.*, **45**, 365.

Gunn, J.E., 1980, *Phil. Trans. Roy. Soc. London* (*Ser. A*), **296**, 313.

Hoyle, F., 1953, *Ap. J.*, **118**, 513.

Kirshner, R.P., Oemler, A., Jr., Schechter, P.L. and Shectman, S.A., 1981, *Ap. J.* (*Letters*), **248**, L57.

Kormendy, J. and Illingworth, G., 1981, preprint.

Larson, R.B. and Tinsley, B.M., 1978, *Ap. J.*, **219**, 46.

Lea, S.M., Silk, J., Kellog, E. and Murray, S., 1973, *Ap. J.* (*Letters*), **184**, L105.

Ostriker, J.P. and Peebles, P.J.E., 1973, *Ap. J.*, **186**, 467.

Ostriker, J.P., Peebles, P.J.E. and Yahil, A., 1974, *Ap. J.* (*Letters*), **193**, L1.

Peebles, P.J.E., 1969, *Ap. J.*, **155**, 393.

— 1971, *Astron. Astrophys.*, **11**, 377.

— 1974, *Ap. J.* (*Letters*), **189**, L51.

Press, W. and Davis, M., 1981, preprint.

Rees, M.J. and Ostriker, J.P., 1977, *M.N.R.A.S.*, **179**, 451.

Roberts, M.S., 1969, *A. J.*, **74**, 859.

Rood, H.J. and Dickel, J., 1978, *Ap. J.*, **224**, 724.

Rubin, V.C., Ford, W.K., Jr. and Thonnard, N., 1980, *Ap. J.*, **238**, 471.

Silk, J. and Norman, C., 1981, *Ap. J.*, **247**, 59.

Terlevich, R., Davies, R., Faber, S.M. and Burstein, D., 1981, *M.N.R.A.S.*, **196**, 381.

Thuan, T.X. and Seitzer, P.O., 1979, *Ap. J.*, **231**, 680.

Tinsley, B.M., 1981, *M.N.R.A.S.*, **194**, 63.

White, S.M. and Rees, M.J., 1978, *M.N.R.A.S.*, **183**, 341.

DISCUSSION

(Note by editors: In her presentation Dr. Faber discussed the age of the Galaxy as determined from globular cluster ages. This discussion is not included in the paper published here but some of the results are included in Table 1 of the Concluding Remarks by Longair).

SILK

You indicated that the age of 47 Tucanae was uncertain by 11 billion years. This large uncertainty does not inspire confidence in the small uncertainties you quoted for other globular clusters.

FABER

47 Tucanae is a special case, owing to the large disagreement concerning its metal abundance (some say metal-rich, some say metal-poor). I believe that this disagreement will be worked out in the next few years and that 47 Tucanae will no longer stand out so prominently.

DAVIS

Brian Flannery and Carol Johnson (in preparation) have been applying different statistical techniques to some of the globular clusters studies by Bruce Carney and conclude that the same data are consistent with ages as young as 10^{10} years. Have you seen this work?

FABER

No, but I think that Carney's results are based to some extent on new, unpublished data, so I wonder whether all the data are indeed the same. I am not familiar enough personally with this subject to give an informed critique of the two results.

WEINBERG

I wanted to quote the same calculations by Flannery of globular cluster

ages that were mentioned by Davis. I am not enough of an expert in the subject to have a definite opinion of how old globular clusters actually are, but Flannery convinced me that they could be as young as 10^{10} years.

FABER

I likewise do not feel sufficiently familiar with this dispute to give an informed reply.

AUDOUZE

My comment is about the age of the universe deduced from the globular clusters. You know that this age can also he deduced from nucleocosmochronology techniques. I would like to mention the work performed by a group from the Institut de Physique du Globe at Paris (Luck, Birck and Allègre, *Nature*, **283**, 256, 1980). These authors have deduced a "nucleosynthetic age" hy using the $^{187}Os/^{187}Re$ pair in a sample of iron meteorites. They found an age of about 20 hillion years in agreement with the ages that you quote from Carney's analysis.

WOLTJER

Does the luminosity function of galaxies end somewhere at the low luminosity end, or does the mass distribution continue into the globular cluster domain? As a related question, what is the minimum mass galaxy where there is evidence for invisible matter?

FABER

I cannot answer the first part of your question. The faint galaxies in question would lie in and around the questionable areas in the (n_{lum}, T) diagram, where observational selection effects are formidable. With regard to your second question, I would say that the smallest galaxies which show strong evidence for non-luminous matter are the smallest Sc galaxies of Rubin and co-workers, with total masses of about 10^{10} $M_\odot$ within their optical radii.

PEEBLES

Your estimate of the index n from the density-temperature relation compares

galaxies, groups and clusters. When I estimated n from the general clustering of galaxies, I got $n \sim 0$, but the relative position of "groups" and "clusters" in your graph could suggest $n \ll -1$.

FABER

I would not pay any attention to the shape or slope of the region in the diagram denoting groups and clusters. These regions represent different perturbations which collapsed at more or less the same time but with varying amount of initial overdensity. They, therefore, do not constitute a hierarchical evolutionary sequence, which they must do if one is to derive the value of n from them. Rather, one should look at the position of galaxies *as a group*, relative to groups-and-clusters *also considered as a group*. This is a proper point of view, I believe, because it is probably true that galaxies do cluster hierarchically into groups-and-clusters, and so the sequence *is* an evolutionary one in this case.

GALAXY FORMATION VIA HIERARCHICAL CLUSTERING AND DISSIPATION: THE STRUCTURE OF SPHEROIDS

S.M. FABER
Lick Observatory, Board of Studies in Astronomy and Astrophysics
University of California, Santa Cruz

ABSTRACT

It is argued that ellipticals and spheroidal bulges are highly condensed systems in which the ordinary luminous matter falls deep into the central core of the surrounding non-luminous halo. Star formation may be triggered when the density of the luminous matter rises above a threshold level set by the halo, thus preventing further collapse. These ideas are consistent with the observed structural scaling laws for elliptical galaxies and the velocity dispersions of spheroidal bulges in spirals. This view of spheroid structure fits naturally into a picture of the Hubble sequence as a sequence of increasing dissipation and central concentration of the luminous matter relative to the surrounding dissipationless halo.

1 - STAR FORMATION AND THE HALT TO SPHEROID COLLAPSE

The preceding paper advanced the view that the structural properties of visible galaxies relate directly to the non-luminous halos in which they are embedded. This proposition was relatively easy to defend for spiral galaxies. Under the likely assumption that the radial collapse of spirals is halted by angular momentum (Fall and Efstathiou 1980), it was possible to show that the disk radius and rotational velocity are fixed functions of the corresponding halo quantities by an amount which depends only on

the angular momentum parameter λ. Structural scaling laws resulted which looked much like the observed scaling laws for spiral galaxies.

As emphasized in the previous paper, the scaling laws as observed for ellipticals and spirals are essentially identical. This similarity strongly suggests that, for ellipticals, a connection with the halo properties is again the root cause. When one attempts to apply this logic to elliptical galaxies, however, a difficulty is immediately encountered — spheroids, including ellipticals, are not supported by rotational motion. It is therefore clear that angular momentum plays no essential role in halting their collapse. Some other mechanism must operate.

White and Rees (1978) have suggested that radial collapse ceases in spheroidal systems when the luminous matter turns from gas into stars. At that point dissipation ends and further collapse becomes impossible. If true, the final density, radii and velocity dispersions in spheroids are all determined by the star-formation process.

These two ideas — a link to the halo and the role of star formation — seem to point quite clearly towards an interesting conclusion: the onset of star formation in spheroids must be controlled somehow by the properties of the halo.

There exists a plausible theoretical argument for just such a connection. In the beginning of collapse, before dissipation has proceeded very far, the gravitational potential of the background halo is still dominant. Mathews (1972) points out that, under this condition, the ability of the gas to fragment into stars is severely inhibited. Let ρ^0_{lum} and ρ^0_{non} be the initial average densities of luminous and non-luminous matter (note that $\rho^0_{lum}/\rho^0_{non} = F$ in the notation of the previous paper). If dissipation is rapid, the collapse of the gas is essentially pressure-free, a case studied analytically by Mathews. He finds that fragmentation is inhibited until the gas density rises to the point where $\rho^0_{lum}/\rho^0_{non}$ exceeds F^{-2}. With $F = 0.07$ (see previous paper), this criterion predicts that the final density of luminous matter within spheroids should be roughly 200 times greater than the mean matter density in the surrounding halo. Not until the collapse has proceeded this far will efficient star formation be possible.

For several reasons this estimate of 200 ought not to be taken too literally. Collapse would continue for some time beyond this density threshold, during which time the luminous matter density would continue to grow. At the same time, on the other hand, Mathews' criterion ignores the possibility of cloud-cloud collisions and shock fronts, which might raise the local density well above average and thus induce star formation earlier.

Nevertheless, Mathews' calculations do give substance to the idea that passage over a critical density threshold is the key factor in halting spheroid collapse.

2 - The self-gravitation of luminous matter in spheroids

The following sections work out some of the quantitative implications of this idea, which seem to agree well with observations. The structural scaling laws for ellipticals fall out quite directly, for example. Before moving to these points, however, I first want to return to the discussion begun in the previous paper on the origin of the Hubble sequence. Recall that two ideas were developed there, termed the "dissipation" and "density" hypotheses. In the former picture, all non-luminous halos (of a given mass) are imagined to have similar density. The higher luminous density of early-type galaxies is due to their higher degree of dissipation and central condensation relative to their halos. In the density scheme, by contrast, all galaxies dissipate by similar amounts. The higher luminous density of early-type galaxies is seen to reflect a higher-than-average halo density, presumed to have arisen from random variations in the density fluctuation spectrum.

Two small pieces of evidence were given in the previous paper which favored the dissipation picture. The first was the higher proportion of non-luminous matter within the visible boundaries of late-type galaxies (Tinsley 1981), an indicator of their lower luminous concentration. The second point, a weaker one, was the way in which angular momentum variations fit naturally into the dissipation scheme — they become the primary determinant of collapse factors and, through them, of Hubble types.

We now have still a third argument in favor of the dissipation idea — the analytic calculations of Mathews. If these results are at all relevant to spheroid collapse, they imply very strongly that star formation cannot occur *until* the average density of luminous matter far exceeds that of the halo. Since we know that ρ_{lum}/ρ_{non} is near unity in late-type spirals (see previous paper), this reasoning automatically requires that the luminous matter be much more concentrated in spheroidal systems than it is in late-type spirals.

There also exist considerable dynamical data on spheroids to support the dissipation picture. The key idea here is this: the dissipation scheme implies very strongly that the luminous material of spheroids must be fully

self-gravitating out to a sizeable distance from the nucleus. This has to he so in the dissipation picture because, in it, no halo density ever approaches the high densities that are observed in the cores of spheroids. Non-luminous matter is, therefore, always an insignificant component near the core (see Figure 2 below). It is a basic tenet of the density picture, by contrast, tbat the average densities of luminous and non-luminous matter are roughly comparable within the optical radii of all galaxies. The halo matter would therefore strongly influence the gravitational potential deep within the spheroid, and the luminous matter would not be fully self-gravitating.

Velocity dispersion profiles afford a check on these ideas. If the luminous matter dominates the potential, the dispersion should decline with radius since the luminous density profile falls off more steeply than isothermal. If, on the other hand, the luminous matter coexists within a dominant, non-luminous, isothermal component, σ should remain roughly constant with radius.

Davies' (1981) velocity dispersion profiles for eight E galaxies are the most comprehensive body of dispersion data at large radii we now have. In all eight objects, the dispersion falls away from the nucleus by an average of 25% at the last measured point, typically 25 core radii. Three out of four spheroidal bulges studied by Kormendy and Illingworth (1981) show similar behavior. In all of these objects, the dispersion profiles are consistent with self-gravitating King or de Vaucouleurs models, just as the dissipation picture would predict.

Mathews (1978) reached a similar conclusion from models of the X-ray source around the elliptical galaxy M87. To match the X-ray surface brightness, two components to the gravitational potential are clearly required: an inner component where the mass-to-light ratio of old stars is distributed radially like the stellar light; a second component having much larger extent and lower central density — the non-luminous halo. In Mathews' model the gravitational potential is dominated by luminous matter out to a radius of fully 30 Kpc. (This large value for the sphere-of-influence of the luminous component agrees well witb the schematic model shown in Figure 2 below).

Further insight into the structure of spheroids comes from the shape of spiral rotation curves. From their analysis of rotation data in 21 Sc galaxies Burstein *et al.* (1981) conclude that the rotation curves constitute a single-parameter family. Each curve consists of a rising portion near the nucleus and a power-law segment beyond, within which v_{rot} is nearly constant. One possible way to parametrize these curves is based on the length of the

power-law portion of the curve within the optical radius. From the published curves I have measured a suitable indicator of this quantity, called $(\Delta \log R)_p$. In galaxies with large $(\Delta \log R)_p$ the rotation curve rises steeply near the nucleus and then remains flat over most of the visible galaxy. As noted by Burstein *et al.* (1981), galaxies of this type must possess a dense concentration of matter near their nuclei.

Within this sample of 21 objects there exists a close correlation between $(\Delta \log R)_p$ and optical morphology, in the sense that galaxies with large $(\Delta \log R)_p$ have large spheroidal bulges. I have estimated bulge strengths from the direct photographs and spectra published by Burstein *et al.* (1981) and Rubin *et al.* (1980). The resultant correlation (Figure 1) shows a well-defined trend.

Fall and Efstathiou (1980) attempted a similar comparison between rotation-curve shape and optical morphology for a larger but more heterogeneous sample of galaxies. They found no good correlation. The reason for the disagreement with the present results is not known, but their sample does contain objects like NGC 4594 (the Sombrero), in which the measured rotation curve near the nucleus clearly does not represent true circular motion (cf. Faber and Gallagher 1979). Until a careful study of these problems has been carried out, no absolutely firm conclusions can be drawn. Pending such a study, however, I am guardedly optimistic that in these 21 Sc galaxies the gravitational signature of the spheroidal bulge has actually been detected.

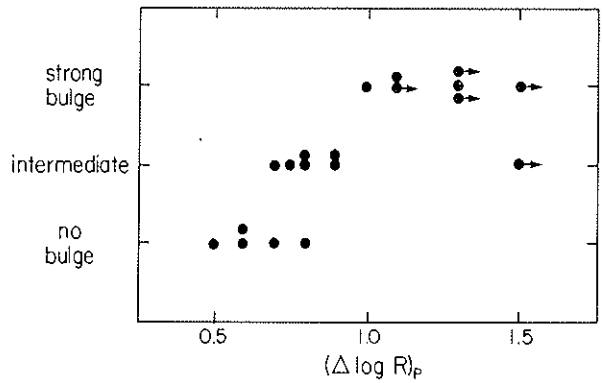


FIG. 1. Length of the power-law section of the rotation curve, $(\Delta \log R)_p$, versus prominence of spheroidal bulge for the Sc galaxies of Burstein *et al.* (1981).

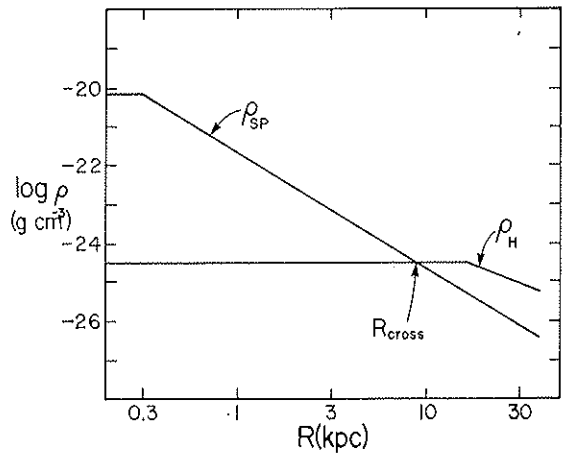


FIG. 2. Schematic comparison of spheroid and halo density profiles in a moderately luminous elliptical galaxy. The central density of the spheroid profile is normalized to measurements of core M/L based on central velocity dispersions and core radii. The halo profile is a "best guess" obtained from scaling halo densities and core radii of small Sc galaxies from Burstein *et al.* (1981), using the scaling laws of the previous paper. Low-luminosity Sc's were employed here since they have no spheroidal bulge. Their halo core radii can thus be estimated directly by making only a relatively small correction for the effect of the disk mass.

If true, spheroids clearly dominate the gravitational potential within their own radii and must be strongly self-gravitating. The Sc rotation curves thus contribute a final piece of evidence in favor of the dissipation picture.

Since there exists considerable if not conclusive support for the dissipation scheme, let us adopt it and see where it leads. It is interesting to ask, for example, at what radius in a moderately luminous elliptical galaxy the mass densities of the luminous and non-luminous components become comparable. A schematic rendering of the two density profiles is attempted in Figure 2 (see caption for details). Although the uncertainties are large, especially for the halo, the two profiles are seen to cross rather far out, at roughly 10 Kpc. This is just slightly smaller than the standard isophotal radii of large elliptical galaxies. If Figure 2 is at all correct, then, it predicts that strong dynamical evidence for non-luminous matter around E galaxies will not appear until velocity measurements are pushed out to radii comparable to the optical size.

3 - Scaling laws for E galaxies

Let us adopt the idea from Mathews' work that passage over a critical density threshold induces star formation in spheroids and thus halts the latter's collapse. Structural scaling laws for elliptical galaxies then arise in a completely natural way. Define ρ_{sp} and ρ_h as the mean densities of matter in the spheroid and halo. If the threshold idea is correct, $\rho_{sp} = Q\,\rho_h$ in all spheroids, where Q is related somehow to the threshold factor. Since, in ellipticals, M_{sp} is approximately FM_h, one finds

$$R_{sp} = \left(\frac{3\,M_{sp}}{4\,\pi\,\bar{\rho}_{sp}} \right)^{1/3} = \left(\frac{3\,F\,M_h}{4\,\pi\,Q\,\bar{\rho}_h} \right)^{1/3} = \left(\frac{F}{Q} \right)^{1/3} R_h \ . \qquad (1)$$

The luminous radius for ellipticals is, therefore, directly proportional to the radius of the surrounding halo, exactly analogous to the situation in spirals (see previous paper).

Since spheroids are presumed to be self-gravitating in the dissipation picture, we also have

$$\sigma_{sp} \propto \left(\frac{M_{sp}}{R_{sp}} \right)^{1/2} \propto \left(\frac{M_h}{R_h} \right)^{1/2} \propto \sigma_h \ , \qquad (2)$$

and thus by Eq. (2c) of the previous paper and constant M_{tot}/L,

$$\sigma_{sp} \propto M_{tot}^{\frac{1-n}{12}} \propto L^{\frac{1-n}{12}} \ , \qquad (3)$$

or

$$L \propto \sigma_{sp}^{\frac{12}{1-n}} \ . \qquad (4)$$

This is just tbe (L, σ) law for elliptical galaxies. As for spirals, the observed value of the exponent between 4 and 5 (Terlevich *et al.* 1981, Tonry and Davis 1981) implies n between -2.0 and -1.4. The near-constant surface brigbtness of E galaxies is also easily derived, just as it was for spirals (see previous paper).

4 - Two quantitative comparisons with observations

i) *Ellipticals Compared to Sc Spirals*

If Eqs. (1) and (2) are derived more carefully, keeping track of the constants of proportionality, it is possible to work out quantitatively the scaling factor between spheroid velocity dispersion and circular velocity in the surrounding halo. This factor can then be compared with the observed ratio of σ to v_{rot} between E galaxies and Sc spirals of the same mass. Let $R_{sp,c}$ and $R_{h,c}$ be the core radii of the luminous and non-luminous matter respectively, $\sigma_{sp,c}$ the nuclear velocity dispersion in the spheroid, and v_{rot} the rotation velocity in the halo. One then obtains the following two equations for F and Q:

$$FQ^{-1} = A \left(\frac{R_{sp,c}}{R_{h,c}} \right)^3 , \tag{5}$$

and

$$FQ^{1/2} = B \left(\frac{\sigma_{sp,c}}{v_{rot}} \right)^3 , \tag{6}$$

where A and B are geometrical factors that depend on the density profiles of halo and spheroid.

These two equations can be used to solve for F and Q from the observational data. The geometrical factors were estimated crudely by modeling the spheroid and density profiles as in Figure 2. Despite the appearance of Eqs. (5) and (6), the important geometrical parameters in the model turn out to be the spheroid core radius, here set equal to 300 pc for a moderately luminous E galaxy (King 1978), and the outer cutoff to the isothermal halo density law. This latter quantity was set equal to the canonical value of 100 Kpc (Ostriker, Peebles and Yahil 1974) but is not well determined.

To obtain σ_{sp}/v_{rot} from the observations we must compare velocities in E's and Sc's of similar mass. The zero-points of the (L, σ) and (L, v_{rot}) relations can be used for this purpose, if adjusted for the different M_{lum}/L_B in E's and Sc's. With this ratio set to 8 for ellipticals and 1.6 for Sc's (see previous paper), the data of Burstein *et al.* (1981) and Tonry and Davis (1981) yield $\sigma_{sp,c}/v_{rot} = .95 \pm .1$.

Eqs. (5) and (6) can now be solved simultaneously for F and Q. One finds Q ~ 190, not far from the theoretical prediction of 200 in Section 1, and F = 0.07, identical to the result in the previous paper. These close agreements are undoubtedly fortuitous, but they do show at least that the self-gravitating model is consistent with other known constraints on the matter distribution.

When an adequate sample of rotation curves for Sa and Sb spirals becomes available, it will be possible to intercompare rotational velocities among the various Hubble types using an analysis similar to the E-Sc comparison above. Owing to the increasing central condensation and self-gravity of the luminous matter along the Hubble sequence, rotational velocities (at constant mass) should increase from Sc to Sa. As noted in the previous paper, this effect may play a role in the observed higher rotational velocities in early-type spirals (Rubin *et al.* 1980). An upper limit to the magnitude of this increase can be estimated by comparing E galaxies with Sc's. The central velocity dispersion in ellipticals is first adjusted downward by ~ 40% [1] to correct to the edge of the spheroid. This dispersion can then be transformed to an approximate circular velocity at that point using the equation of hydrostatic equilibrium and the observed luminosity profile:

$$v^2_{\rm rot} = \sigma^2 \frac{d \ln \rho_{\rm lum}}{d \ln r} . \qquad (7)$$

If $\sigma_{\rm sp,c}/v_{\rm rot} = .95$, one finds that the circular velocity at the luminous radius of an E galaxy should be roughly 1.15 times the circular velocity far out within the halo. The effect of internal heating thus amounts to about 15%. For Sb's and Sa's, intermediate in Hubble type, one therefore anticipates rotational velocities roughly 5 to 15% higher than in Sc's of similar mass.

ii) *Spheroids Compared to their Surrounding Disks*

The preceding section estimated the scaling factor between the central velocity dispersion in elliptical galaxies and the circular velocity in their surrounding isothermal halos. Spheroidal bulges within spiral galaxies invite a similar analysis. The method must be modified slightly since only a

[1] This value was obtained by extrapolating Davies' (1981) curves to the optical radii.

fraction, call it f, of the luminous matter now makes up the spheroid. Hence

$$M_{sp} = f\,F M_{tot} \ . \tag{8}$$

This fraction f can be estimated from the observed bulge-to-disk ratio if M_{sp}/L_{sp} and M_d/L_d are assumed known. Eq. (6), slightly rearranged, now becomes

$$\frac{\sigma_{sp,c}}{v_{rot}} = f^{1/3}\ B^{-1/3}\ F^{1/3}\ Q^{1/6} \equiv C f^{1/3} \ . \tag{9}$$

This equation says that $\sigma_{sp,c}/v_{rot}$ should vary as $f^{1/3}$, where f is the fractional spheroid mass. Since f is unity in elliptical galaxies, the constant C should equal the ratio found in the previous section, namely $\sigma_{sp,c}/v_{rot} = 0.95$.

Whitmore, Kirshner and Schechter (1979) give values of $\sigma_{sp,c}$, bulge-to-disk ratio, and rotational velocity for several spiral galaxies. Unfortunately, most of their galaxies are of fairly early Hubble type (Sa-Sb). Their rotational velocities may therefore somewhat exceed the true halo velocities by 5 to 15% (see above). In a first reconnaissance this potential source of error has been ignored and uncorrected values of $\sigma_{sp,c}/v_{rot}$ versus f are plotted in Figure 3. The solid line is the prediction of Eq. (9), assuming $(M/L_B)_{sp} = 6.5$ and $(M/L_B)_d = 1.6$. The shape of the curve fits the data reasonably well.

The vertical position of the curve was adjusted to optimize the fit.

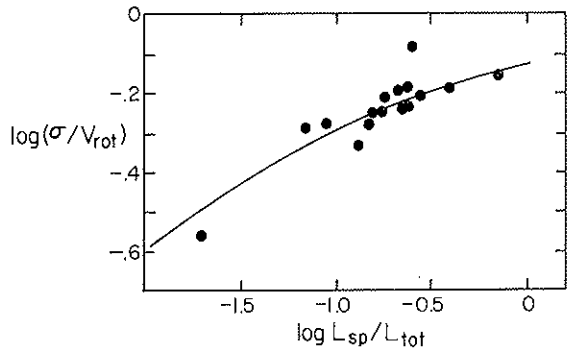


FIG. 3. The data of Whitmore *et al.* (1979) on σ/v_{rot} versus bulge-total luminosity for a variety of spiral galaxies. The solid line is the theoretical prediction of Eq. (9) which has been shifted vertically to match the data.

This is equivalent to determining the constant C in Eq. (9). The value given by these data is 0.75 ± 0.02, somewhat smaller than the value 0.95 from E's versus Sc's. This difference may stem in part from the source of error mentioned above. If the disk rotational velocities were an average of 5 to 15% too high, C would need to be corrected upward by $\sim$ 10 to 15%. The two estimates would then agree within the errors.

This explanation should be checked with careful mass and rotation-curve models for the WKS galaxies to test whether the rotation curve shapes are in fact consistent with an enhancement over halo velocity of 5 to 15%. Until this test is carried out, the numerical results of this section look promising but cannot be considered definitive.

To summarize the results of Section 4, observed internal velocities as a function of Hubble type are generally consistent with the dissipation theory and seem to show the moderate increase towards early Hubble types that the theory would predict. The data also seem to support the idea that spheroid density scales directly with halo density and thus that spheroid collapse is halted by passage over a critical density threshold.

5 - Conclusion

The preferred picture of galaxy formation which emerges in these two papers is one in which galaxies of various Hubble types collapse and dissipate by differing amounts, depending on the angular momenta they absorbed as protogalaxies. Spheroids in this picture consist of extremely low angular momentum material in which an enormous radial collapse is in principle possible before a final equilibrium disk forms. Long before they actually reach this state and, hence, while they are still essentially spherical, the gas density rises above a critical threshold imposed by the surrounding halo and stars are able to form.

The greater angular momentum of late-type spirals prevents collapse by such a large factor. As long as the luminous matter remains spherically distributed its density can never exceed that of the background halo by an amount sufficient to permit star formation. Not until the material settles down into a disk does the density of luminous matter climb to the point where fragmentation into stars becomes possible. The *local* self-gravity of the gas is then sufficient to make stars even though the *global* gravitational potential might still be governed by non-luminous matter. These arguments imply a fundamental difference between star formation in disks and

spheroids — a difference which might well repay closer theoretical scrutiny.

These ideas are consistent with several important properties of present-day galaxies, including the structural scaling laws, the lower angular momentum of spheroids relative to disks and the higher velocities in early Hubble types. Two other properties not mentioned previously can also be added to this list. They are the systematic increase in bulge-to-disk ratio along the Hubble sequence and the apparently higher rotation of spheroidal bulges relative to ellipticals. These, too, fit in quite naturally if angular momentum (λ) varies systematically with Hubble type.

Many important questions remain unanswered in the present picture. Chief among these is a theory for the gaseous dissipation rate. When did dissipation start and how fast did it proceed relative to the clustering rate? These two factors together set the characteristic mass of galaxies and must also have played an important role in determining their final structure. A related need is a detailed picture of the formation of disk and spheroid components. Exactly how and when did the first stars form, how did the level of heavy elements build up, and how were the final density profiles determined? Without answers to all of these questions we cannot claim truly to understand galaxy formation.

I would like to acknowledge once again the help of all those persons named in the previous paper especially James Gunn, who helped greatly to sharpen the comparison between the dissipation and density hypotheses. I also thank G.R. Knapp for important correspondence regarding velocity dispersions in spheroidal bulges. William G. Mathews contributed several valuable insights on the nature of star formation in a background potential and the mass-density implications of the M87 X-ray source. The majority of this work was completed while the author held an Alfred P. Sloan Foundation Fellowship.

REFERENCES

Burstein, David, Rubin, V.C., Thonnard, N. and Ford, W.K., Jr., 1981, preprint.
Davies, R., 1981, *M.N.R.A.S.*, **194**, 879.
Faber, S.M. and Gallagher, J.S., 1979, *Ann. Rev. Astron. Astrophys.*, **17**, 135.
Fall, S.M. and Efstathiou, G., 1980, *M.N.R.A.S.*, **193**, 189.
King, I.R., 1978, *Astrophys. J.*, **222**, 1.
Kormendy, J. and Illingworth, G., 1981, preprint.
Mathews, W.G., 1972, *Astrophys. J.*, **174**, 101.
— 1978, *Astrophys. J.*, **219**, 413.
Ostriker, J.P., Peebles, P.J.E. and Yahil, A., 1974, *Astrophys. J.* (*Letters*), **193**, L1.
Rubin, V.C., Ford, W.K., Jr. and Thonnard, N., 1980, *Astrophys. J.*, **238**, 471.
Terlevich, R., Davies, R., Faber, S.M. and Burstein, D., 1981, *M.N.R.A.S.*, **196**, 381.
Tinsley, B.M., 1981, *M.N.R.A.S.*, **194**, 63.
Tonry, J. and Davis, M., 1981, *Astrophys. J.*, **246**, 666.
White, S.M. and Rees, M.J., 1978, *M.N.R.A.S.*, **183**, 341.
Whitmore, B.C., Kirshner, R.P. and Schechter, P.L., 1979, *Astrophys. J.*, **234**, 68.

DISCUSSION

SILK

i) Infall of gas into a dark potential well and dissipation will increase the specific angular momentum parameter λ. Is this consistent with the initial conditions on λ inferred from simulations so as to result in ellipticals with lower λ than spiral bulges?

ii) The amount of gas in rich galaxy clusters may have been appreciably underestimated because the gas-to-galaxy mass fraction increases with radius. One recent result finds a value of 20 percent for the gas fraction relative to the cluster mass. This suggests that the fraction of ordinary to dark matter may not be a universal constant.

FABER

i) I interpret that you are asking whether the dispersion in λ in the tidal torque theory is large enough to account simultaneously for slow rotating ellipticals and rapidly rotating spirals. I have thought about this, but I frankly do not know. It is an important question.

ii) These estimates are all rather uncertain since the gas cannot be followed out as far as the galaxies, and its total extent is, therefore, unknown. In the face of uncertainty, I find a simple hypothesis preferable to a more complex one.

THE EVOLUTION OF GALAXIES

JAMES E. GUNN
Princeton University Observatory

Visiting Associate
Mount Palomar Observatory
California Institute of Technology

ABSTRACT

The recent observational evidence on the evolution of galaxies is reviewed and related to the framework of current ideas for galaxy formation from primordial density fluctuations. Recent strong evidence for the evolution of the stellar population in ellipticals is presented, as well as evidence that not all ellipticals behave as predicted by any simple theory. The status of counts of faint galaxies and the implications for the evolution of spirals is discussed, together with a discussion of recent work on the redshift distribution of galaxies at faint magnitudes and a spectroscopic investigation of the Butcher-Oemler blue cluster galaxies. Finally a new picture for the formation and evolution of disk galaxies which may explain most of the features of the Hubble sequence is outlined.

1 - PREAMBLE: GALAXY FORMATION BY GRAVITATIONAL COLLAPSE

We will here adopt the picture that galaxies formed by the collapse and subsequent virialization of perturbations present in the matter density in the universe subsequent to the combination of the primeval plasma at about 4000 K, or a redshift of about 1500. This implies a strong prejudice on the part of the author, and perhaps other scenarios fit the scanty observational evidence as well; the "standard" one serves here mostly as a framework for discussion.

Of the parameters of the original perturbation that are of interest currently, probably the most important are the collapse time and, of course, the total mass. The collapse time is related only to the mean amplitude above the critical density at some epoch, and for a roughly spherical perturbation is given by

$$t_c = \frac{\pi}{H_i} (\delta^+)^{-3/2} \tag{1}$$

where δ^+ is the mean fractional amplitude interior to the shell of interest and H_i is the Hubble parameter at that epoch. The maximum expansion factor, i.e. the ratio of the radius of the shell at maximum expansion to its original radius, is just $(\delta^+)^{-1}$ (see, for example, Gunn and Gott 1972). If the structure forms without any dissipation, the maximum radius and collapse time can be obtained from observed dynamical parameters of the present structure, viz.

$$R_{max} = 170 \text{ kpc } M_{12} \sigma_{100}^{-2} , \tag{2}$$

and

$$t_c = 2.2 \times 10^9 \text{ yr } M_{12} \sigma_{100}^{-3} \tag{3}$$

where M_{12} is the mass in units of 10^{12} solar masses and σ_{100} is the one-component velocity dispersion in units of 100 km/sec. For the Galaxy, the "timing" argument based on the orbit of M31 and the Galaxy (see, e.g., Peebles 1971 and the paper by Lynden-Bell at this conference) give a mass for the Local Group of about 3×10^{12} solar masses, of which about a third, or 1×10^{12}, probably helongs to the Galaxy. The circular velocity in the Galaxy is now thought to be about 220 km/sec (Knapp, Tremaine and Gunn 1978, Gunn, Knapp and Tremaine 1979, Lynden-Bell and Frenk 1981) which corresponds to a velocity dispersion in an isothermal halo of $220/\sqrt{2}$, or about 160 km/sec.

The bulge component in the Galaxy falls off in density more rapidly than the r^{-2} halo, and its velocity dispersion should be correspondingly lower, by a factor of $\sqrt{(2/3)}$ if $\rho_{bulge} \propto r^{-3}$. The relation

$$\sigma_{bulge} = v_c/\sqrt{3} \tag{4}$$

in fact seems to be closely satisfied for a large number of spiral galaxies of widely different bulge-to-disk ratios (see, for example, Knapp *et al.* 1981, and the data of Whitmore, Kirshner and Schechter 1979, Illingworth

and Kormendy 1981, and Illingworth and Schechter 1981) for the central velocity dispersion in the bulges. Since it must hold in the region of space where the potential is dominated by the dark $1/r^2$ halo, the implication is that the bulges are roughly isothermal. The case for elliptical galaxies is not so clear, since there is no direct evidence that they have halos at all, and the measurements of the spatial variations of the velocity dispersion in ellipticals is somewhat confusing. We will assume here that the central velocity dispersions are sufficiently representative of the velocity dispersions in the stellar population in elliptical galaxies, and thus assume that the halo velocity dispersions in ellipticals are $\sqrt{(3/2)}$ times larger. (We assume, of course, that ellipticals do have halos, or, at least, once had halos; the missing mass problem in clusters and groups otherwise needs an independent solution. See, for example, Gunn 1980 and Tremaine and Gunn 1979 for a discussion).

With this assumption and the assumption that the halos are in fact dissipation-free we can for any galaxy calculate the collapse time and the maximum expansion radius from presently observed dynamical properties, provided we know the mass. What we can obtain to fair accuracy is the mass in stars, from stellar population models for other galaxies and from direct observation in the Galaxy. It can easily be shown (Gunn *et al.* 1978) that, if the universe contains two non-interacting fluids at the era of decoupling with arbitrarily different density perturbations, they very quickly evolve together on all scales larger than the greater of the two Jeans masses and that, consequently, the ratio of masses in the two components is the same in all such resulting structures. We will here assume that this is the case on the scale of galaxies (which means for the massive neutrino advocates that the mass of the relevant neutrino is at least of the order of 30 eV or so).

For the Galaxy, the mass of the bulge and disk can be estimated dynamically, and is about 7×10^{10} solar masses for a local surface density of 70 $M_\odot/pc^2$, $R_0 = 8$ kpc, and a scale length of 3.5 kpc, with the rotation curve of Gunn, Knapp and Tremaine (1979). For a total mass of 1×10^{12} $M_\odot$, this gives a ratio of "ordinary" to "halo" mass of 0.07. It is interesting that Dr. Faber has arrived at the same figure using completely independent data on other galaxies (see her paper in this collection). We will adopt it in the discussion that follows. For other systems, we use the mass-to-light ratios for the stellar population obtained by Faber and Gallagher (1979), and correct to total mass with the factor obtained above. Velocity dispersions for ellipticals were obtained from the list of Terlevich *et al.* (1981)

and rotation velocities for spirals (these are values from rotation curves, note, not from total HI line widths) from the list of Faber and Gallagher (1979).

The result is shown in Figure 1, where halo velocity dispersion is plotted against total (and luminous) mass. Also shown are lines of constant collapse time. The plotting symbols are different for galaxies of different type and are explained in the caption. Rees and Ostriker (1977) and later White and Rees (1978) in the current context of massive halos have shown that structures larger than about 2×10^{12} $M_\odot$ cannot cool in a dynamical time and, therefore, are unlikely to make single galaxies; however, structures which have a collapse time of less than about 5×10^8 yrs will cool by Compton cooling on the microwave background. It is gratifying that galaxies avoid the forbidden region, and there is a strong suggestion that ellipticals are primarily Compton-cooled objects, as might be the bulges of spirals as well.

The upper envelope of the elliptical galaxy distribution (all of which have collapse times much shorter than the spirals) has a much shallower slope than the constant collapse time loci, and fits very well a line of slope 1/6 which also goes through the Coma Cluster. Since that structure must also be at or near the upper envelope of objects in the mass range around 10^{15} solar masses (less than about 1% of the mass in the universe is in clusters of that mass as tightly bound as Coma), the implication is that the spectrum of primeval perturbations which gave rise to the galaxies and the clusters must have a characteristic run of amplitude with scale such as to give the observed one-sixth power law. This relation implies that the RMS density fluctuation varies as the inverse cube root of the mass ($n = -1$), at least over the range 10^9 to 10^{15} solar masses. This result was obtained by Turner *et al.* (1979) by a comparison of small groups and great clusters with n-body experiments, and our result gives some credance to the view that galaxies and clusters are all part of the same gravitational hierarchy, one which has significantly more power at large scales than that generated by white noise. Faber (this volume) has also arrived at an $n = -1$ hierarchy by quite different considerations concerning the present structure of galaxies.

The curve drawn on the vertical axis in Figure 1 is the expected distribution of velocity dispersions in these coordinates. The shaded curve indicates schematically the effect of the destruction of long-collapse-time objects by more typical ones as the hierarchy develops, and represents something like the expected final distribution. The observed scatter in

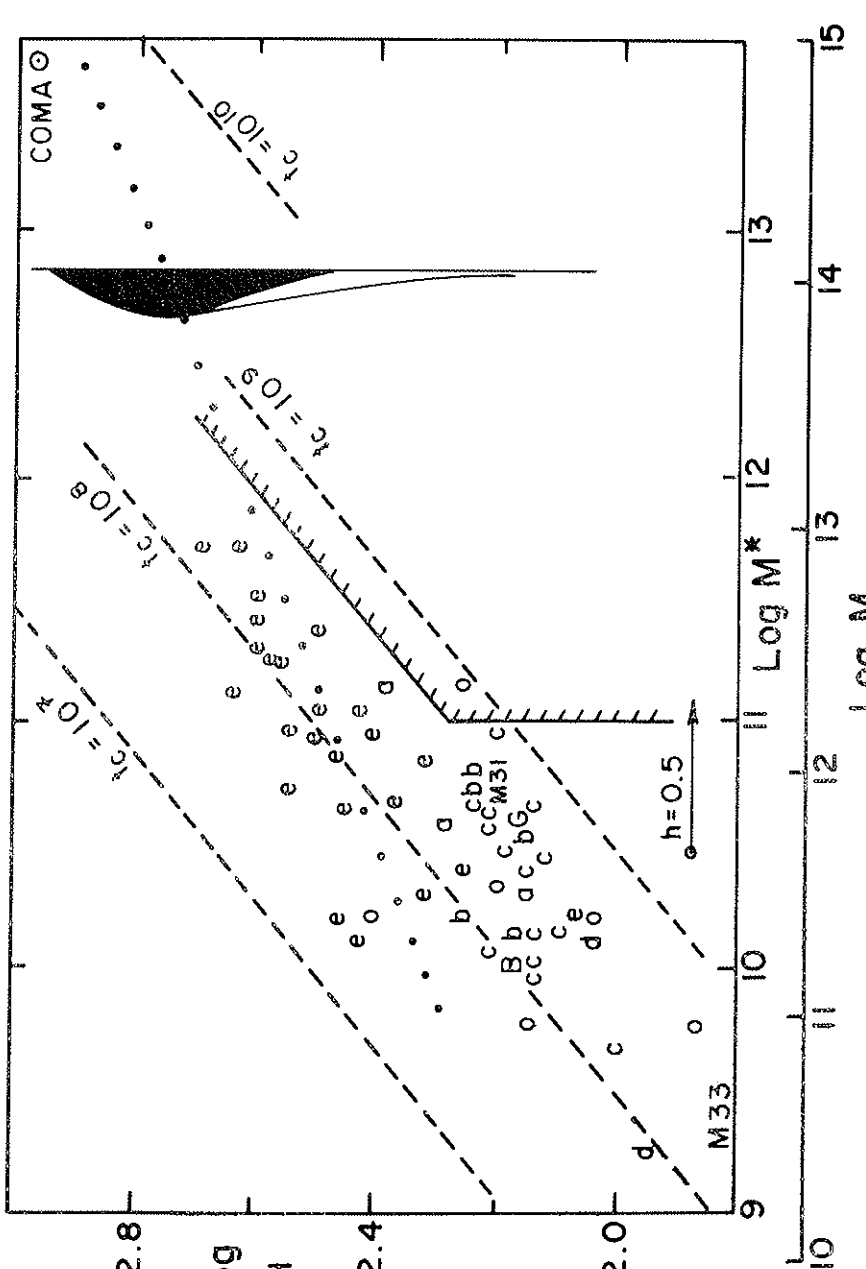


FIG. 1. Plot of the mass of galaxies versus their halo velocity dispersions, derived from observed parameters as described in the text. The plotting symbols are as follows: e are ellipticals, 0 are SO, a are Sa, b are Sab and Sb, c are Sbc and Sc, d are Scd, Sd, and Im. G is the Galaxy, B the bulge, and M31 and M33 are those galaxies. The plot is for H = 100; going to H = 50 moves most of the points except G, B, M31 and M33 to the right by the length of the arrow. The circle labeled COMA is the coma cluster of galaxies. Lines of constant collapse time are plotted; the dotted line is the "ridge line", corresponding to the rms density fluctuation in the $n = -1$ hierarchy which best fits the data. The hatched line represents the approximate boundary of the region to the right of and below which structures cannot cool in their collapse times. The vertical part is set by gas cooling, the sloping part by Compton cooling. The curve is an approximate histogram for the distribution vertically about the hierarchy ridge, as is explained in the text

properties is not too different from the expected one, and lends some support to the hierarchical picture. (We assume, of course, that the distribution of density fluctuations is a Gaussian random process to derive this distribution, a state of affairs guaranteed by the Central Limit Theorem, if the fluctuations have random phase and extend to much smaller scales than the ones of interest here). The assumptions going into Figure 1 doubtless need refining, but a great deal of the physics of galaxy formation is encapsulated in it, and it will guide much of the subsequent discussion.

2 - The Evolution of Elliptical Galaxies

Elliptical galaxies as seen today tempt one to think that they are structurally simple objects, and the fact that they are all, as we have seen, objects with very short collapse times, makes it seem likely that the bulk of their "interesting" evolution occurred very early in the history of the universe. In particular, most of the star formation must have occurred early, by arguments first advanced by Eggen, Lynden-Bell and Sandage (1962) for the bulge of the Galaxy. The stellar population of ellipticals supports this view; the light is dominated by old K giants, and there is no direct evidence for continuing star formation. That the situation is not quite as simple as it first seems has been suspected for some time. The recent detailed spectral synthesis by Gunn, Stryker and Tinsley (1981) makes it clear that there are stars which are indistinguishable from young main-sequence stars in these systems, but that most of the main-sequence light is from very old stars with a turnoff near B-V = 0.8. It is not clear whether these blue objects are "blue stragglers" or are really young, but their presence confuses simple models for the evolution of the colors and luminosities of ellipticals quite badly. The evidence is consistent with the large majority of the mass in ellipticals having formed stars very early, and indeed the fitting of the best current isochrones for the evolution of metal-rich stars in the relevant mass range demands ages for the stars of greater than 10^{10} yr. The evolution of the light and color of ellipticals is of interest, of course, for the classical cosmological tests, which use sizes and luminosities of calibratible standard candles and metric standards to determine the deceleration parameter.

The crucial parameter determining the luminosity evolution is the logarithmic slope of the main sequence, since the luminosity at any time is just proportional to the number of giants present at that time (the

luminosity, temperature, and lifetime of red giants are quite insensitive to the mass of the progenitors over the range of interest) and so is proportional to the rate at which stars are evolving off the main sequence. This rate is just the number of stars per unit mass divided by the rate of change of lifetime with mass. Thus the number of giants is

$$N_g = \bar{t}_g \left(\frac{dt_{ms}}{dm} \right)^{-1} \left(\frac{dN_{ms}}{dm} \right)_{t_0} , \tag{5}$$

and, if $(dN/dm)_{t_0} = Cm^{-(1+x)}$, the rate of change of visual luminosity is

$$\frac{d \text{ Log L}}{d \text{ Log } t} = -1.3 + 0.3\ x \tag{6}$$

(Tinsley 1973). Luminosity evolution is indistinguishable from a change in the deceleration parameter in first order, and if one ignores luminosity evolution or makes an error in it, the error one makes in the deceleration parameter is

$$\Delta q = \frac{2}{2-\tilde{\alpha}} \frac{1}{H_0 t_0} \left(\frac{d \text{ Log L}}{d \text{ Log } t} \right)_{t_0} , \tag{7}$$

where $\tilde{\alpha}$ is a parameter related to the effective aperture correction (Gunn and Oke 1975) which we discuss further below. For $\tilde{\alpha} = 1$ and for $q \sim 0.1$, $\Delta q \sim 0.7\ \Delta x$. Thus x must be obtained to quite exquisite precision, if the deceleration parameter is to be measured well. The models all require x to be smaller than about 1, and the solar neighborhood value is near 0. It seems clear that one cannot learn more about x from the spectra of galaxies at the present epoch, but there is hope that the study of high signal-to-noise spectra of high-redshift galaxies might provide the answer, since the relative contribution of the main sequence to the visual luminosity is much larger at early times.

It is quite clear that one does see the expected stellar evolution in ellipticals as one looks to larger and larger redshift objects. The spectrum of a first-ranked cluster galaxy at a redshift of 0.756 obtained by Hoessel, Oke and the author with our PFUEI CCD camera/spectrograph (Gunn and Westphal 1981) on the Hale telescope is shown in Figure 2, with a typical (but somewhat lower resolution) energy distribution of a small-red-

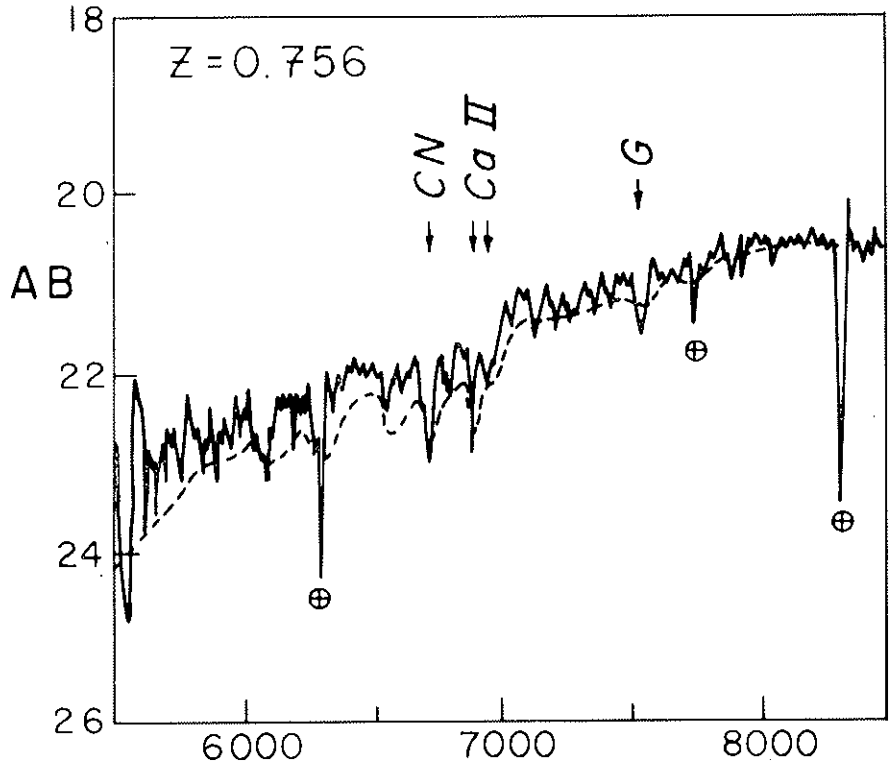


FIG. 2. The PFUEI spectrum for a first-ranked cluster galaxy at a redshift of 0.756 obtained by Oke, Hoessel, and the author. The G band of CH and Fe, the H and K CaII resonance lines, and the violet CN band, the strongest features in galaxy and red giant spectra in this region, are marked, as are several telluric features. The dotted line is the average spectrum of several low-redshift giant cluster ellipticals at somewhat lower resolution shifted in wavelength to agree with the high-redshift galaxy. It is clear that the local objects are considerably redder than the high-redshift one.

shift giant elliptical superposed. The much bluer color and smaller break across the CaII H and K lines is obvious in the high-redshift spectrum. The effect is also quite pronounced in the color-redshift relation for bright ellipticals, as is sbown in Figure 3, also plotted for the data of Hoessel, Oke and Gunn. The colors were derived from high accuracy CCD images of the clusters in our survey for distant clusters of galaxies. The dotted line is the relation expected from the "K-correction" alone; it is clear that the high-redshift objects are much bluer than tbeir nearby counterparts. These effects have also been noted by Kristian, Sandage and Westphal (1978) and

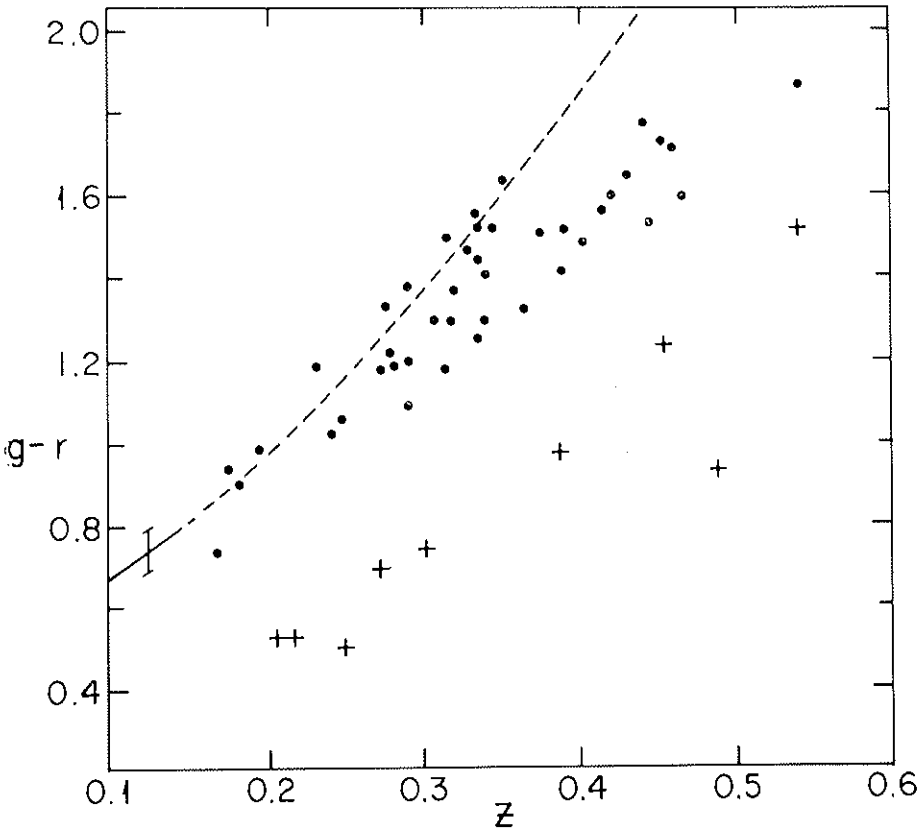


FIG. 3. The g-r colors of the Oke-Hoessel-Gunn sample plotted against redshift. The dotted line is the locus expected for no evolution of the stellar population. The crosses are colors of galaxies with unusual spectra found in the survey. Some are doubtless spirals in front of very distant clusters, but some are probably analogues of NGC 1275.

by Spinrad and collaborators (see, for example, Smith *et al.* 1979). The color evolution is of the order of magnitude predicted by the population models (Tinsley and Gunn 1976) but the colors themselves are uniformly too blue, certainly the result of the small number of "young" stars in the galaxies. There remains an enormous amount of work in this area to be done, and the recent tragic death of Beatrice Tinsley probably means that it will be done more slowly and much less well than it might have been.

It is now generally recognized that there is another evolutionary term which is likely to be as important or more so than the stellar evolutionary

one, viz. dynamical frictional processes whereby the giant central galaxies in clusters consume smaller ones and add to their total mass and light (Ostriker and Tremaine 1975, Gunn and Tinsley 1976, Hausman and Ostriker 1978). If the frictional evolution is roughly homologous (and the data suggest that it is; there is no marked difference in profiles for brightest cluster ellipticals except in their very outer parts, where other processes are likely to dominate) then there are definite relations predicted between the structural parameters of a cannibal galaxy and its total luminosity. In particular, the parameter α introduced by Gunn and Oke in terms of which the aperture correction in cosmology takes a very simple form (cf. equation 7) is related in a simple monotonic way to the total luminosity. α is simply the logarithmic derivative of the luminosity contained within an aperture of some standard radius (in our work 16 kpc for $H = 60$ and $q_0 = 1/2$).

Hoessel (1980) has obtained direct measures of α for the nearby Abell sample of Hoessel, Gunn and Thuan (1980) and those data are shown in Figure 4, along with the predictions of the homology theory (which are

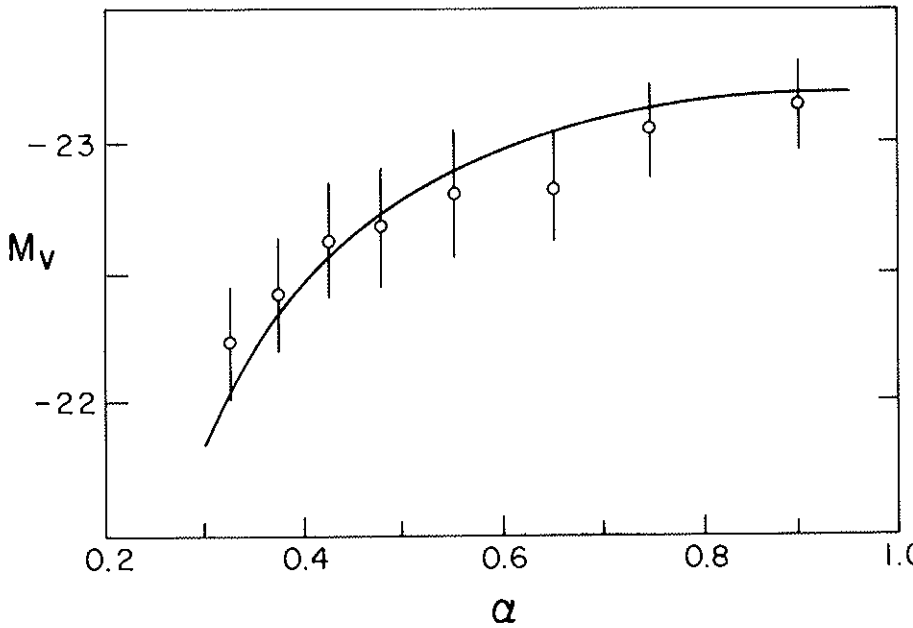


FIG. 4. The absolute magnitude vs alpha relation for the sample of nearby Abell clusters of Hoessel, Gunn and Thuan (1980). The error flags are the sample standard deviations, not the standard errors of the points plotted.

themselves indistinguishable from the Monte Carlo calculations of Hausman and Ostriker). The error flags in this figure are the sample error flags, not the errors of the means (of typically ten points) which are plotted. The fit is quite acceptable, and applying the correction to the luminosity derived from the theoretical curve approximately halves the standard deviation in the luminosity and removes the trends corrected for separately hy Sandage and coworkers (see, for example, Sandage and Hardy 1973) for richness and Bautz-Morgan class, as might be expected if those corrections are in reality corrections themselves for cannibalism. The most interesting facet on the α-correction is that it is in principle epoch-independent, since it measures the cannibal's growth to the epoch the measurement is made. Thus if the cluster galaxies in a sample are individually α-corrected, no further statistical correction need or should be applied for dynamical effects.

Oke, Hoessel and the author have for the past few years been gathering deep CCD imaging data on our distant clusters, and it is evident that with sufficiently good signal-to-noise, α can be measured from the ground for all the galaxies in that sample, in which the largest redshift so far measured is 0.92. At that redshift the standard aperture subtends an angle of about three arcseconds, so the seeing corrections are manageable for good data. Space Telescope, of course, will be able to do the job with exquisite precision (Gunn 1979). One of the unfortunate effects of all this is that the effective $\tilde{\alpha}$ in the stellar evolution correction to q_0 becomes 1.0 instead of the mean α for the sample, which increases the precision required for the stellar part for a given precision for q_0 by about a factor of about 1.4.

It would seem that we find ourselves at a place similar to that occupied just prior to the realization that dynamical effects are likely to be important, i.e., that stellar evolutionary effects are the main uncertainty in using the Hubble diagram for cosmology. The uncertainties are probably smaller then they were at that time and the problems are certainly better understood, but it still seems likely that it will be a long time before the problem is solved. And this in spite of the fact that quite accurate data to very large distances for the Hubble diagram itself — redshifts and magnitudes and even α's — will be available quite soon. It should be remarked that, in principle, the stellar evolution can be solved for by using the surface brightness data contained implicitly in a complete set of m, z, α for a sample of brightest cluster ellipticals, but the statistical efficiency of the scheme is very low and it requires extremely good data. It may in the end prove to be the only workable way to do the problem.

3 - Counts and the Evolution of Spirals

Spiral galaxies account for about 80% of the light and a roughly equal fraction of the nearby counts of galaxies. Since they are on average much bluer than ellipticals and so suffer smaller K-corrections, they must completely dominate the counts of galaxies at moderate redshifts. The count-magnitude relation is independent of the cosmological model through second order in the redshift, so the counts and mean colors of galaxies with magnitude provide a tool which in principle informs us about the evolution of spirals in a reasonably clean fashion.

The use of the faint counts, however, is not completely straightforward, in part because the agreement among various workers is not very good and, with the notable exception of the work of Kron (1980), insufficient detail concerning the techniques of reduction and the various corrections that have been made to the data have been published to allow independent assesments of its quality. The situation with counts at brighter magnitudes, to which the faint counts must be tied, is not much better. The counts of the Zwicky catalog suffer both from the effects of the local supercluster and from less than ideally accurate magnitudes, and the counts in various versions of the Shapley-Ames are completely dominated by the local supercluster. Probably the best compromise is the catalog of Kirshner, Oemler and Schechter (1978) which is complete in several fields to a limiting J magnitude of 15.7. Those authors did not take account of redshift effects in their normalization of the luminosity function, which results in an underestimation of about 30% of that normalization, and there is evidently an additional numerical error which is smaller but in the same direction.

When these are accounted for, models with the resulting luminosity function and normalization and no evolution fit their counts and the counts of Kron (1980) and of Tyson and Jarvis (1979) to a magnitude on their J^+ system of about 23, beyond which Kron's counts are much larger than Tyson and Jarvis', which are, in turn, much larger than those predicted by the model (which by this point is rather inaccurate). The median redshift at that magnitude with no evolution is about 0.35, and there is clearly little room for significant evolution since that epoch. The counts are plotted in Figure 5. Kron's counts suggest that there might be a significant evolutionary contribution beyond that, but it is my opinion that both better models than mine and corroboration of the steep rise in the faint counts are needed before the result is taken too seriously. Indeed, a faint survey with

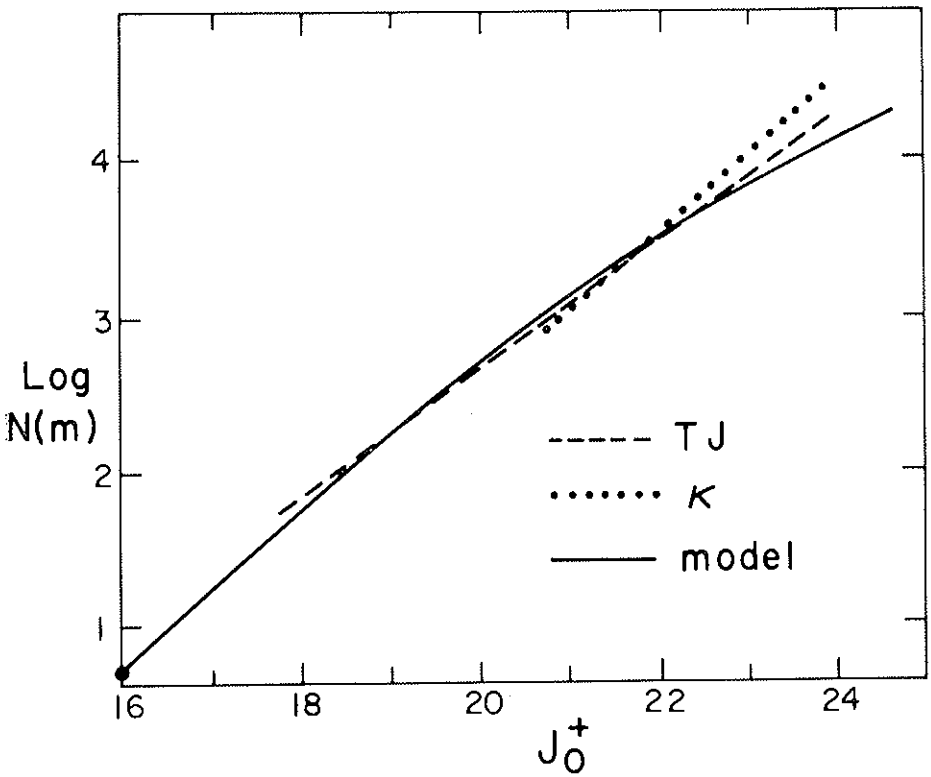


FIG. 5. The faint number counts, in absorption-corrected J^+ magnitudes. The counts of Tyson and Jarvis are schematically represented by the dashed line and those of Kron by the dotted line. A model with no evolution which fits the counts of Krishner, Oemler, and Schechter (the circle at the lower left) predicts the solid line. There is no evidence for evolution at J^+ magnitudes brighter than about 23.

the PFUEI CCD camera by Young, Kristian, and the author fails on preliminary reduction to reproduce Kron's steep slope.

A set of quite detailed models has recently been constructed by Bruzual (1981), but it is not clear in the distributed version of the work how he handled the bright-end normalization, so the model counts themselves are not easily assessed. His calculations for the color and redsbift distributions are not affected by tbose uncertainties, however. He has made models both with no evolution and with various evolutionary schemes, using more complete UV data on both galaxies and stars than has been employed heretofore. The median color as a function of redshift turns out

to be a very poor discriminant, and Bruzual finds that the observations of Kron are adequately fitted by both non-evolving models and by various reasonable evolutionary ones.

The median redshift at a given magnitude is a rather stronger discriminant, although not a cosmology-independent one. The median redshifts for both non-evolving models and a model with strong evolution are plotted in Figure 6, as are Kron's observed colors. The only results available on redshift distributions are from a survey by Turner, Sargent and myself using the SIT spectrograph on a random sample of Kron's galaxies at $J^+ = 20$, the median redshift from which agrees very well with Bruzual's value for no evolution. The distribution of redshifts in the sample of 58 spectra, however, does not agree at all well with that expected in the sense that there is a significant number of galaxies which are much too distant (and therefore much too bright). The situation does not correspond to any of the smooth, simple evolutionary models of Bruzual and Kron (1980) or of Tinsley (1980a, 1980b), and exactly what is going on is not at all clear.

The galaxies in that sample which are responsible for the excess are also systematically blue, and one can speculate whether they are related to the blue cluster galaxies found by Butcher and Oemler (1978). Is there a significant population of blue, bright galaxies in the early universe which disappeared only a short time ago, say at redshift 0.3 or conceivably even a bit later? Investigations of the Butcher-Oemler blue galaxies are a crucial start to the answering of that question.

Dressler and I have begun a spectroscopic investigation of these objects in the 3C295 cluster ($z = 0.46$) and in the very rich cluster $0024 + 16$ ($z = 0.39$). Easily the most enigmatic objects which are appearing in this survey are the bluest, which typically have no features at all in our spectra. Some of the objects have Seyfert spectra and others spectra not unlike late-type spirals. It is fairly clear, however, that the objects are not spirals, or at least not spirals of the sort we see today; their surface brightnesses are much too high. It is likely that these are small systems which are undergoing vigorous star formation in some transient phase; but why such systems are seen never or at least very rarely now is not understood, nor are the processes responsible for their activity. We do not, of course, know the redshifts of the featureless-spectrum blue galaxies, and it is possible that they have nothing to do with the clusters but represent high-redshift objects in some very bright phase. We are currently attempting to obtain very good spectra of some of these galaxies in the hope that some features will materialize.

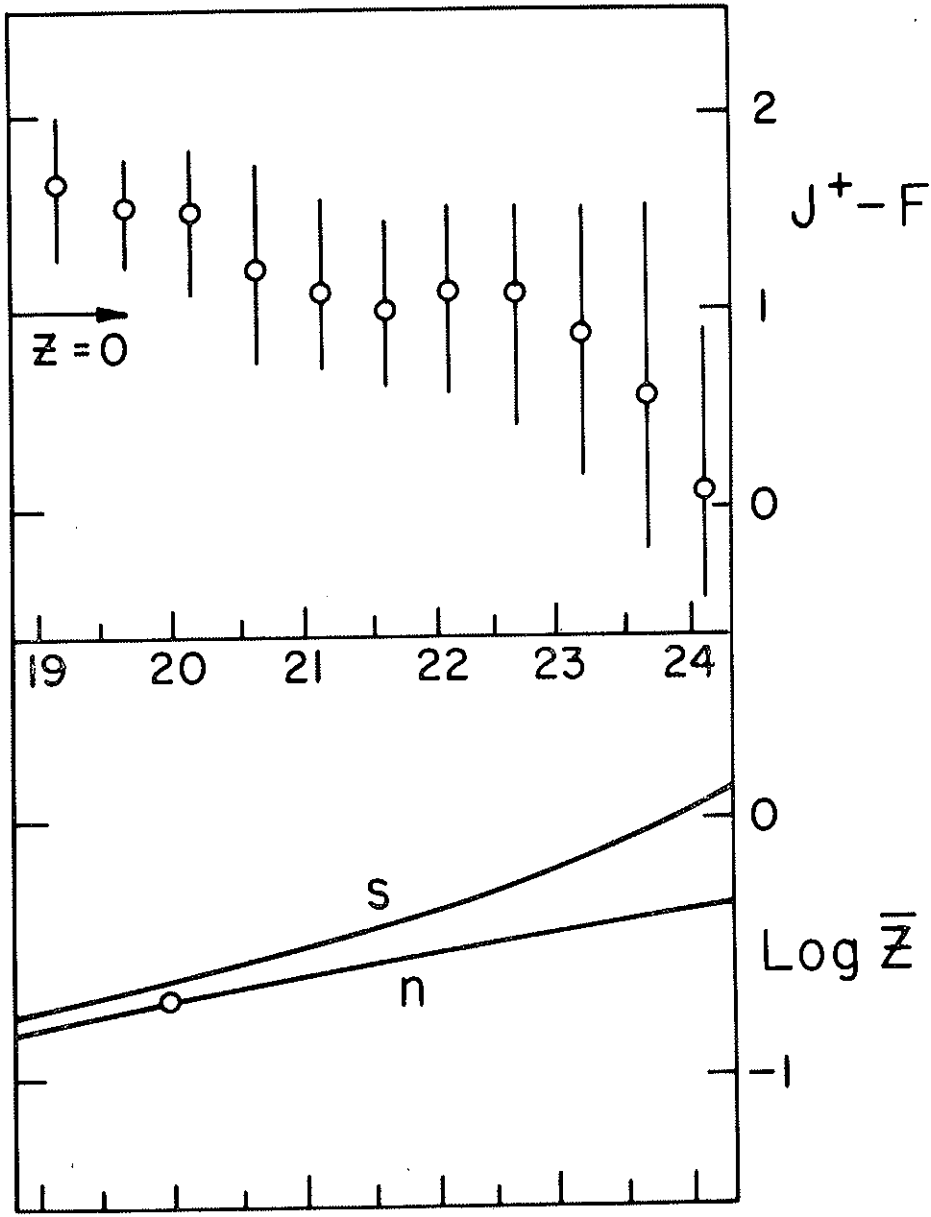


FIG. 6. (Top) The median colors in Kron's counts versus J+ magnitude. The length of the flags represents the range of colors encountered. (Bottom) The median redshift as a function of J+ magnitude for two of the models of Bruzual, the (n)o evolution model, and a model with (s)trong evolution. The circle is the median redshift for the sample of Turner, Sargent and Gunn.

4 - A Model for the Evolution of Disk Galaxies

The data to date suggest that there is no rapid evolution of spiral galaxies, and the evidence has been accumulating for some time that spirals are very likely still accreting significant amounts of mass today (Tinsley 1981, Larson *et al.* 1980, Twarog 1980). This suggests that the disks of spirals are formed relatively slowly from infalling material. We have seen that spirals are formed from perturbations that have long collapse times, and those perturbations inevitably have "tails" which fall in slowly (Gunn 1977). This is emphasized by considering the bulge of the Galaxy as a separate system; it has a collapse time which is short, as short as that of elliptical galaxies of the same mass (see Figure 1), and the recent dynamical results of Frenk and White (1980) demonstrate that, if the globular cluster system of the Galaxy shares the dynamics of the bulge, the bulge has a specific angular momentum only about a third that of the disk.

The latter result is important. It militates against the currently popular idea that galaxies grow from little subsystems, since in that picture the stars in the protogalactic lumps form the bulge and the gas forms the disk, and there can be no way that the two systems can have different specific angular momenta. It suggests rather the somewhat older picture of a rather coherent collapse, the coherence being very likely brought about only by the presence of a single dominant lump in the primeval soup which tells its neighbors to form a galaxy here, not there. The picture in which the galaxies acquire their angular momentum via tidal torques then would predict that the innermost, presumably densest parts, of the initial perturbation would receive the smallest moment. If the initial angular velocity is roughly constant at a time when the perturbation is still roughly homogeneous (this is the expected state of affairs because most of the tidal acceleration occurs before the perturbations go strongly nonlinear; see e.g. Peebles 1969) then one expects the specific angular momentum to be proportional to the 2/3 power of the mass fraction. The bulge is probably about 20% as massive as the disk in the Galaxy, and thus one would expect it to have about a third of the specific angular momentum in this simple picture. Fall and Efstathiou (1980) have shown that the tidal torque picture can nicely account for the total angular momentum of disk galaxies, if one takes into account that the disks as we see them today are imbedded in massive nondissipative halos; without halos one cannot account for the observed angular momentum via this mechanism. The magnitude of the

total angular momentum imparted by tides is described by a dimensionless parameter λ,

$$\lambda = J\ |E|^{1/2}\ M^{-5/2}\ G^{-1} \tag{8}$$

which has a mean value about 0.07 (Peebles 1969) and whose distribution was shown by Efstathiou and Jones (1979) from n-body experiments to be roughly normal with a standard deviation of about half the mean. One would expect, however, that in a given gravitational hierarchy the value of λ would be correlated with deviations from the mean hierarchical behavior, those perturbations with the shortest collapse times at a given mass scale feeling the smallest torques and ending with the smallest values of λ. This may well account for the observed small angular momenta of ellipticals.

Let us carry the angular momentum argument one step farther. Mestel (1963) first noted that the discs of spirals have angular momentum distributions which are not very dissimilar to that of a uniformly rotating uniform sphere, and the same is true for elliptical galaxies (Gunn 1980, Fall and Efstathiou 1980). Mestel investigated the properties of self-gravitating discs which have such angular momentum distributions, the form of which, if $m\,(h)$ is the mass with specific angular momentum h or smaller, is

$$m\,(h) = M\ [\ 1 - (1 - h/H)^{3/2}\ ]\ . \tag{9}$$

He showed that there were at least two solutions, one a uniformly rotating disc with a very flat density distribution and another which is quite centrally concentrated and which is quite strongly differentially rotating. It is quite easy to find the surface density distribution corresponding to a flat rotation curve in the potential of an isothermal halo, and that distribution is plotted in Figure 7 for two values of the inner turnover radius for the rotation curve. The dotted line is an exponential disc with a scale length of a quarter of the outer radius, and it is clear that the two are quite similar.

I would like to suggest that the mystery of the approximately exponential form of observed disks is nothing more complicated than this. If the original angular momentum distribution is approximately conserved as the disk is built up out of the gaseous component of successively infalling shells, the density distribution grows homologously, which means that it is approximately exponential throughout its growth. Its scale radius and central surface density depend on how the rotational velocity changes as the disk grows. If the rotational velocity grows as $M^{1/6}$, i.e., in a fashion

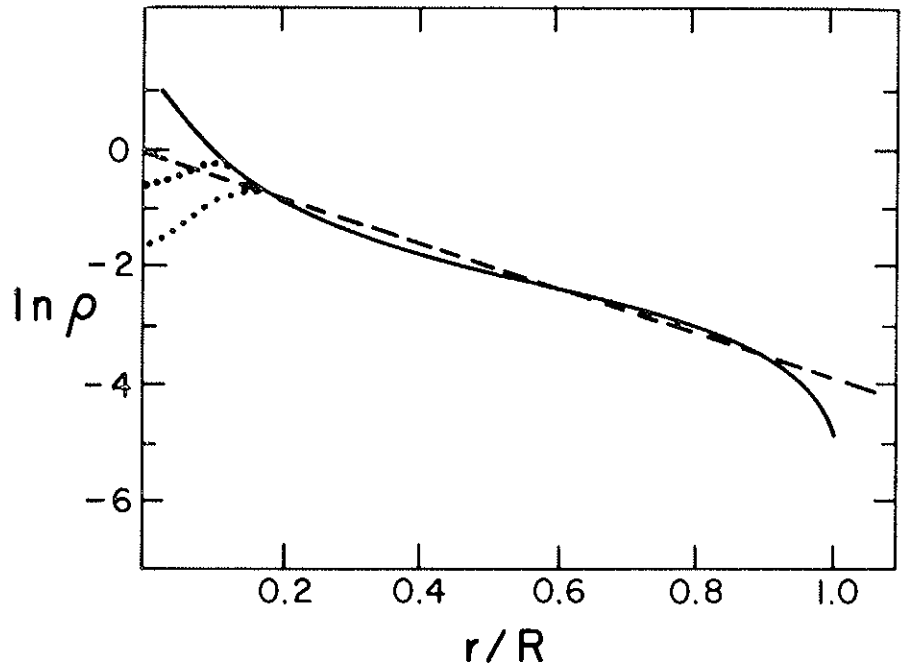


FIG. 7. The surface-density-radius relation for the angular momentum distribution of a uniformly rotating uniform sphere which rotates with constant circular velocity after collapse (solid line) and the same except for a small constant angular velocity core of radii 0.5 and 0.8 scale lengths (dotted lines). The dashed line is an exponential surface density distribution, and the fit is quite good over about three and a half scale lengths.

consistent with the parent hierarchy, then the Lagrangian radius of a given ring (of fixed h) varies as $M^{-1/6}$, the central surface density is constant, and the scale length varies as $M^{1/2}$. Thus the disk grows self-similarly, is always roughly exponential over about 4 scale lengths, and evolves at roughly constant central surface density; and all this happens without appeal to any exotic transport processes.

Let us now discuss the evolution of the vertical structure of a disk which grows in this fashion. The disk material can cool as it falls, and can be expected to be at a temperature just below that for the onset of collisional ionization of hydrogen, about 15000 K, which corresponds to a one-dimensional velocity dispersion σ_D of about 10 km/sec. At first there will be insufficient material for the disk's self-gravity to affect the vertical structure at all, and the characteristic height in the disk will be

$$H \sim \frac{\sigma_D r}{v_c} \tag{10}$$

When sufficient material has accumulated such that the surface density Σ has grown to

$$\Sigma \sim \frac{v_c \sigma_D}{\sqrt{2 \pi G r}} \tag{11}$$

self-gravity becomes important, and the heigbt (as in the previous case the equivalent Gaussian one-standard deviation height) is

$$H \sim \frac{\sigma_D^2}{\sqrt{2 \pi G \Sigma}} \tag{12}$$

The velocity dispersion will still be roughly that of the gas, about 10 km/sec.

The star formation rate at these early epochs might be expected to be small, since the disk is still stably pressure-supported at the gas temperature. (See also Fall and Efstathiou 1980 who made this point about the final configurations of very low-mass galaxies). This point is perhaps debatable since, if the infall is not smooth, cooling shocks can be driven in the disk by the infalling clouds and one might have significant star formation in the absence of instability of the disk structure itself. As the surface density grows still more, however, and reaches the value

$$\Sigma_{crit} \sim \frac{\sqrt{2}\, v_c \sigma_D}{\pi G r} \tag{13}$$

the disk becomes Jeans unstable (Goldreich and Lynden-Bell 1965a, Toomre 1964), and continues to be so as more material falls in. Since the gas cannot with reasonable energy input be heated, tbe instability will persist, driving inhomogeneities which in turn drive star formation. Only wben most of the mass has formed stars can the disk begin to heat and stabilize itself. Since infall continues, however, the resulting structure will always be on the threshold of instability. Once most of the mass is in stars, the velocity distribution will in general be anisotropic, with $\sigma_r = 2\,\sigma_\theta$ for a flat rotation curve. The vertical velocity dispersion is, in principle, decoupled from the other two, but in the presence of large-scale

collective Jeans instability must emerge of the same order as they. We will adopt here a value 1.6 times smaller than σ_r, based on the approximate ratio for old-disk stars in the solar neighborhood. The vertical velocity dispersion then becomes, at the threshold of instability,

$$\sigma_D \sim \frac{1.4\, G\, \Sigma\, r}{v_c} \tag{14}$$

with a corresponding characteristic height of

$$H \sim 0.45\, \frac{\Sigma\, G\, r^2}{v_c^2} \,. \tag{15}$$

Actually, the growth of the surface density itself increases the velocity dispersion through the conservation of the adiabatic invariants. The net effect is still to make the disk unstable for all $\alpha r > 0.75$ (α^{-1} is the scale length of the disk) and for all αr if the flat outer rotation curve makes a transition to an inner solid-body part at $\alpha r = \beta > 0.5$.

The consequences of this simple picture are as follows:

1) Exponential disks are produced in a natural fashion; furthermore, the disks remain exponential throughout their relatively slow growth.

2) The disks of spiral galaxies grow from the inside out, with the scale length growing as the square root of the total disk mass. They grow secularly, since a given shell has an infall time which is typically very much larger than the dynamical time of the disk. In the case of the Galaxy, the material currently being added to the disk (primarily, as we shall see, at radii only a little outside the Sun) has fallen in from radii in excess of 70 kpc.

3) The velocity dispersion is predicted simply as the disk grows; it is always the larger of 10 km/s and the value required for Jeans stability. One expects a dramatic decrease of the gas content of disks at the onset of Jeans instability. It is instructive to look at these relations for the Galaxy, as a function of q, the fraction of the current disk mass present at a given epoch. We take for these calculations the following parameters: the current disk mass, $M_D = 5.3 \times 10^{10}\, M_\odot$, the surface density at R_0, $\Sigma_0 = 70\ M_\odot/pc^2$, the scale length $\alpha^{-1} = 3.5$ kpc, the solar radius $R_0 = 8$ kpc, the circular velocity $v_c = 220$ km/s, and the transition radius

$\beta = 0.8$. The velocity dispersion as a function of q is given for these parameters in Figure 8 as a function of the radius in scale lengths; recall that the scale length varies as $q^{1/2}$, and that the radius of a given parcel of matter in the disk as $q^{-1/6}$. Notice that the Sun, whose position is marked on the curves with a solar symbol, has entered the Jeans-unstable regime

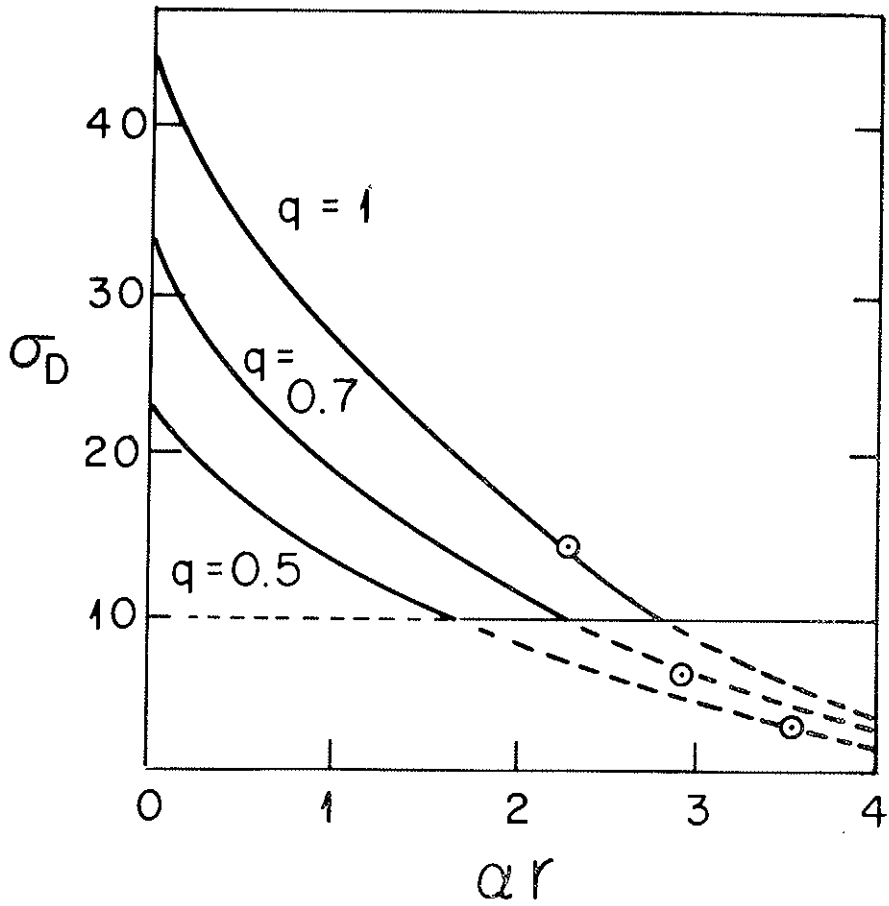


FIG. 8. The one-dimensional velocity dispersion versus radius in scale lengths for the model discussed in the text. The parameter q is the fraction of the present disk mass of the Galaxy accreted at a given epoch. The dotted lines below 10 km/sec are the velocity dispersions corresponding to Jeans instability in the low-density regimes where the real dispersion is set by the gas turbulence at about 10 km/sec. The circles are the Sun's location in scale radii. Note that the Sun has only recently (in terms of q) moved into the Jeans-unstable regime.

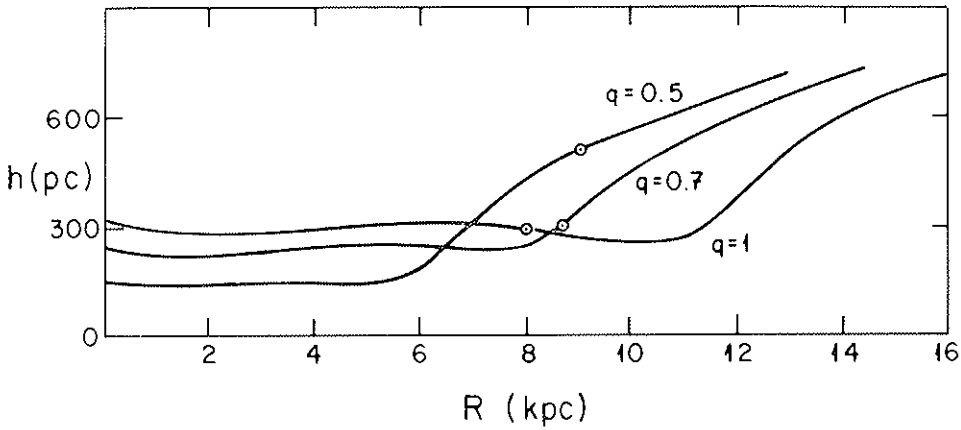


FIG. 9. The scale heights as a function of physical radii for the three cases of Figure 8. Note that the scale heights are almost constant until the disk flares at the point that the velocity dispersion goes to the 10 km/sec floor. About 1 scale length farther out the scale height is essentially that for a non-self-gravitating disk.

only relatively recently in terms of the q parameter. We will return to the question of the timescale for all of this later.

4) The characteristic height of the disk is also predicted, and is plotted, this time in physical radius units, in Figure 9. Notice that over the whole region where the dispersion is determined by Jeans stability the height (and hence the exponential scale height, which is just $1/\sqrt{2}$ times our characteristic height) is almost constant, varying by no more than 10% inside the place where the disk flares out to its non-self-gravitating $H \propto r$ behavior. The latter, incidentally, is in excellent agreement with the observations for the Galaxy (Gunn 1980) and constitutes probably the best direct evidence that the dark mass is in a nearly spherically symmetric halo. Van der Kruit and Searle (1981) have recently determined the stellar scale heights for a few nearby edge-on spirals and find that they are essentially constant, in agreement with these models. They also find an abrupt cutoff in their disks. Whether this is to be identified with the cutoff of star formation at the place where the disk flares or only with the sudden drop in surface brightness due to the flaring itself is not clear, but it would be extremely interesting to know.

5) The infalling material keeps the disk marginally unstable to the Toomre-Goldreich-Lynden-Bell-Jeans modes, and the criterion for that

instability is also the one for the shear amplifier (Goldreich and Lynden-Bell 1965b, Julian and Toomre 1966) to be very efficient in making large-amplitude spiral disturbances out of random ones. If the infall is not absolutely smooth, the random ones are built in to the infall process. Thus the infall probably drives spiral structure, which, without excitation, propagates away (Toomre 1977) in a few revolutions. Furthermore, the form of the spiral pattern should be correlated with the infall rate; when the swing amplifier is driven hard, the spiral pattern will be open, near the angle of peak response, but for weaker driving the effect of differential rotation will be larger and the spiral will be tighter. If these conclusions withstand more quantitative scrutiny there is offered a natural explanation of the Hubble sequence. Since infalling gas must be converted into stars on the infall time scale to retain marginal stability, systems in which the infall rate is high should be blue and dominated by young stars. If the same infall drives an open spiral pattern, one needs nothing else to explain the properties of the late Hubble types; conversely, the galaxies with low current infall rates should be red and tightly wound.

6) Stars formed when the disk was not yet Jeans unstable have large scale heights at formation (see Figure 9). The preservation of their vertical adiabatic invariant requires that their scale height decreases only as the inverse cube root of the local surface density. Thus the increase of scale height with age is a consequence of the formation history of the disk, and does not require secular acceleration.

Thus in this picture the place a galaxy occupies on the Hubble sequence is an accident of its infall rate, which as we shall see is not easily predictable. It is the case, however, that the infall rate at late times will in general be larger for those systems with long collapse times than those with short, and it is certain that no infall at all can occur for galaxies in sufficiently dense clusters, that the cluster itself has collapsed already and heated all the intracluster gas to the virial temperature of the group. Thus there should be no *bona fide* spirals in collapsed clusters, even without stripping. The material responsible for the infall is very loosely bound to the galaxy and is thermally very fragile, and it does not seem unlikely that the very large dependence of type on the density of the environment found by Dressler (1980) can be so explained, as was first proposed by Larson, Tinsley and Caldwell (1980). It seems reasonable that long-collapse-time systems should have, in general, smaller subsystems with short collapse times than short-collapse-time systems should have (!). Thus galaxies with long collapse times should have smaller bulges than ones with shorter

collapse times, if in fact bulges are the parts of galaxies that collapse quickly (and, for example, are Compton-cooled). If late-type systems are generally those with long collapse times (and it appears that that is the case in Figure 1) they should he expected generally to have small bulges; conversely, early-type systems, with short collapse times, will have large hulges. It should also be the case that there are frequent exceptions to these rules, since the present infall rate will depend on more complex things than simply the collapse time, and the size of the short-collapse-time subsystem, if any, depends on the (random) details of the form of the initial density perturbation.

The model does not predict with any ease the evolution in time; indeed, it predicts that one cannot predict that evolution without consideration of the interactions with neighbors, since it is competition for material that typically determines the infall rate at late times (Gunn 1977). At early times, the mass grows approximately like $t^{2/3}$, or $\dot{M}/M = 2/3\ t$. Later, the infall rate depends on the environment, and is typically considerably smaller than the relation for the early rate would predict. This is the case even for an isolated galaxy, if Ω is considerably less than unity, and is small in groups like the local group because of competition.

If we take a rate half the above, $\dot{M}/M = 1/3\ t$, it corresponds to an infall rate of about 2 $M_\odot$/yr for the Galaxy, which is quite reasonable. The fractional growth in surface density is proportional to radius:

$$\dot{\Sigma}/\Sigma = \alpha\ r\ \dot{M}/M$$

so that at any epoch most of the mass is being added to the outside of the disk. If Ω is small, $t = t_0/2$ corresponds to $z = 1$, at which epoch $q = 0.78$, and the disk has about 60% of its present area within the Jeans-unstable region. The star formation rate is roughly the mass infall rate into that region, which is very nearly the same as the present rate. Thus the blue luminosity of the Galaxy should have been much the same at $z = 1$ as it is now, but the surface brightness much higher. If spirals do evolve in this way, Space Telescope will be able to tell with direct imaging; in fact, the "Planetary Camera" will allow the study of spiral structure at redshifts near one with moderate resolution enhancement on images taken near 8000 A, which is near the peak response of its CCD detectors. This band is, of course, near 4000 A in the rest frame, so the images will be directly comparable with blue photographs of nearby spirals (Gunn 1980).

In great clusters like Coma, the growth is arrested quite early, when the mass of a galaxy like M31 or the Galaxy is only of order half the mass

it would have at the present epoch in more benign environments. Since the cutoff in star formation, because of the cutoff of infall and the stripping of the interstellar medium, leads to a population change which increases the mass-to-light ratio by a factor of about two compared to average spirals in the field and the disks of the cluster spirals only grow half as massive for a given initial perturbation (the rest of the gas being incorporated into the intracluster medium), the M/L ratio of great clusters should be about a factor of 4 larger than that of small groups and the field, consistent with the values of about 250 and 70 for those cases found by Faber and Gallagher (1979) on their system ($H = 50$, B_T).

In summary, this picture of slow disk formation accounts in a natural way for most of the systematics observed in disk galaxies and may account for essentially all of them, if the suggested connection between the infall rate and the spiral pattern can be corroborated. Infall models have long been the best candidates to explain the observed aspects of chemical evolution in the solar neighborhood (see, e.g., Tinsley 1981, Twarog 1980); this picture simply extends those ideas to include dynamical consequences.

The author would like to thank Martin Rees and Donald Lynden-Bell for warm hospitality at the Institute of Astronomy in Cambridge, where much of the last part of this work was done, and those gentlemen and Gillian Knapp, Sandra Faber and Jerry Ostriker for illuminating conversations. The work was supported in part by the National Science Foundation and the National Aeronautics and Space Administration of the United States.

REFERENCES

Bruzual, G. and Kron, R.G., 1980, *Ap. J.*, **241**, 25.

Bruzual, G., 1981, thesis, University of California, Berkeley.

Butcher, H. and Oemler, A., 1978, *Ap. J.*, **219**, 18.

Dressler, A., 1980, *Ap. J.*, **236**, 351.

Efstathiou, G. and Jones, B.J.T., 1979, *M.N.R.A.S.*, **186**, 133.

Eggen, O.J., Lynden-Bell, D. and Sandage, A.R. ,1962, *Ap. J.*, **136**, 748.

Faber, S. and Gallagher, J., 1979, *Ann. Rev. Ast. and Ap.*, **17**, 135.

Fall, S.M. and Efstathiou, G., 1980, *M.N.R.A.S.*, **193**, 189.

Frenk, C.S. and White, S.M., 1980, *M.N.R.A.S.*, **193**, 295.

Goldreich, P. and Lynden-Bell, D., 1965a, *M.N.R.A.S.*, **130**, 97.

— 1965b, *ibid*, 125.

Gunn, J.E., 1977, *Ap. J.*, **218**, 592.

— 1979, *IAU Colloq.*, **54**, "Scientific Research with the Space Telescope", p. 383.

— 1980, *Phil. Trans. R. Soc. London A.*, **246**, 313.

Gunn, J.E. and Gott, J.R. III, 1972, *Ap. J.*, **176**, 1.

Gunn, J.E. and Oke, J.B., 1975, *Ap. J.*, **195**, 255.

Gunn, J.E. and Tinsley, B.M., 1976, *Ap. J.*, **210**, 1.

Gunn, J.E. and Westphal, J.A., 1981, in "Solid State Imagers for Astronomy", *Proc. SPIE 290*, 16.

Gunn, J.E., Knapp, G.R. and Tremaine, S.D., 1979, *A. J.*, **84**, 1181.

Gunn, J.E., Lee, B.W., Lerche, I., Schramm, D.N. and Steigman, G., 1978, *Ap. J.*, **223**, 1015.

Gunn, J.E., Stryker, L.L. and Tinsley, B.M., 1981, *Ap. J.*, **249**, 48.

Hausman, M.A. and Ostriker, J.P., 1978, *Ap. J.*, **224**, 320.

Hoessel, J.G., 1980, *Ap. J.*, **241**, 493.

Hoessel, J.G., Gunn, J.E., Thuan, T.X., 1980, *Ap. J.*, **241**, 486.

Illingworth, G., Kormendy, J., 1981, submitted to *Ap. J.*

Illingworth, G., Schechter, P.L., 1981, submitted to *Ap. J.*

Julian, W.H., Toomre, A., 1966, *Ap. J.*, **146**, 810.

Kirshner, R.P., Oemler, A., Schechter, P.L., 1978, *A. J.*, **83**, 1549.

Knapp, G.R., Shane, W.W., Burg, G., van der Bajaja, E., Faber, S.M., Gallagher, J.S., 1981, submitted to *Ap. J.*

Knapp, G.R., Tremaine, S.D., Gunn, J.E., 1978, *A. J.*, **83**, 1585.

Kristian, J., Sandage, A.R., Westphal, J.A., 1978, *Ap. J.*, **221**, 383.

Kron, R.G., 1980, *Ap. J. Suppl.*, **43**, 305.

Larson, R.B., Tinsley, B.M., Caldwell, C.N., 1980, *Ap. J.*, **237**, 692.

Lynden-Bell, D. and Frenk, C.S., 1981, *Observatory*, **101**, 200.

Mestel, L., 1963, *M.N.R.A.S.*, **126**, 553.

Ostriker, J.P., Tremaine, S.D., 1975, *Ap. J. Lett.*, **202**, L 113.

Peebles, P.J.E., 1969, *Ap. J.*, **155**, 393.

— 1971, "Physical Cosmology", Princeton University Press.

Rees, M.J., Ostriker, J.P., 1977, *M.N.R.A.S.*, **179**, 451.

Sandage, A.R., Hardy, E., 1973, *Ap. J.*, **183**, 743.

Smith, H.E., Junkkarinen, V.T., Spinrad, H., Grueff, G., Vigiotti, M., 1979, *Ap. J.*, **231**, 307.

Terlevich, R., Davies, R.L., Faber, S.M., Burstein, D., 1981, *M.N.R.A.S.*, **196**, 381.

Tinsley, B.M., 1973, *Ap. J.*, **186**, 35.

— 1980a, *Fundamentals of Cosmic Physics*, **5**, 287.

— 1980b, *Ap. J.*, **241**, 41.

— 1981, *Ap. J.*, in press.

Tinsley, B.M. and Gunn, J.E., 1976, *Ap. J.*, **203**, 52.

Toomre, A., 1964, *Ap. J.*, **139**, 1217.

— 1977, *Ann. Rev. Ast. and Ap.*, **15**, 437.

Tremaine, S.D., Gunn, J.E., 1979, *Phys. Rev. Letters*, **42**, 467.

Turner, E.L., Aarseth, S.J., Gott, J.R. III, Blanchard, N.T., Mathieu, R.D., 1979, *Ap. J.*, **228**, 684.

Twarog, B.A., 1980, *Ap. J.*, **242**, 242.

Tyson, J.A. and Jarvis, J.F., 1979, *Ap. J. Lett.*, **230**, L 153.

Van der Kruit, P.C. and Searle, L., 1981, *Astron. and Ap.*, **95**, 105.

White, S.D.M. and Rees, M.J., 1978, *M.N.R.A.S.*, **183**, 341.

Whitmore, B.C., Kirshner, R.P., Schechter, P.L., 1979, *Ap. J.*, **234**, 68.

DISCUSSION

SILK

If cloud infall is responsible for the build-up of disks, then collisions between clouds and the resulting viscous transfer of angular momentum may dominate disk evolution.

GUNN

That is perhaps true, but to the extent that the cloud system behaves like a fluid its cloud structure does not matter, and I suspect that, if the initial conditions are not too chaotic, it does not matter even if the cloud mean free path is long.

OSTRIKER

The picture of galaxy formation that you propose is overall very attractive but I think it would be best if the time scale were short and if every little infall were occurring at our epoch. If a system accretes $1M_{\odot}$ per year, it would be expected to dissipate 10^{41} ergs per second emitted as soft X-rays — two orders of magnitude larger than seen by the Einstein satellite.

You pointed out at the beginning of your talk how distinct were the spheroidal bulge and disk components. What physical process predicted their distinction in the model you propose?

GUNN

The argument given by Larson and Tinsley on the processing rate for gas in spirals convinces me that infall is necessary in any case, so the problem is with us independent of this model. I have no pet solution and it is true that many of the possible ways out complicate the simple picture, as, for instance, the infall being viscously modified by a hot wind.

Regarding the second question, I have not given very much thought to the formation of the bulge. It is clear that the requirement of small specific angular momentum demands that the bulge form from the inner region of the

perturbation which is presumably densest. One can argue from the position of ellipticals in the σ-M diagram that the competition between the star formation rate and collapse time must be tipped in the direction of high star formation rate (i.e., formation rate going faster than $\rho^{1/2}$) at high densities. Doubtless one can obtain an empirical form for the necessary physics from the observed distribution, but I have not done that.

SETTI

Concerning the excessive production of soft X-rays due to infall of gas in the Galaxy, it should be borne in mind that this will depend on the precise magneto-hydrodynamics of the flow where galactic magnetic fields and cosmic rays are properly included. It may very well happen that the gas will not be shock-heated to high temperatures, but that instead the greater part of the kinetic energy of the infalling matter will be used up in accelerating the relativistic particles. (Puppi, Setti and Woltjer, 1967, IAU Symposium no. 31, p. 289).

GUNN

Some such process may well work, but still the energy loss rate is high. We would certainly not see it if it went into the cosmic ray gas with high efficiency, but I do not know how efficiently it might be done.

AUDOUZE

In the discussion between J.P. Ostriker and you about the occurrence of infall, among many arguments in favor of such an infall, I would like to recall the very large luminosity in the far infrared wave band of the ring at 5 kpc. This means that there is a very rapid formation of massive stars going on in this region. Infall is needed to replenish quickly the gas in this region.

GUNN

Possibly. I think it more likely that the gas in the ring is mostly deposited there by the action of a bar in the inner part of the Galaxy.

OORT

In connection with the presence of young stars in elliptical galaxies, Sanders

has made the suggestion that star formation in these galaxies might be a "relaxation" process: the gas produced by the evolving stars would accumulate until it is sufficiently dense for star formation to take place. The young stars would compress the gas and blow it away. Such a picture might explain both the virtual absence of gas during a long fraction of the time as well as the presence of some young stars.

GUNN

If the gas builds up gradually half the amount at star formation should be seen in an average object. It is also difficult but doubtless not impossible to prevent the star formation from being very centrally concentrated.

DAVIS

All this scenario occurs in a background of a $1/r^2$ halo. Are you still happy that your model for dissipationless formation of such a halo will work?

GUNN

Yes. I think that recent N-body models have shown that *after* infall is all over, the outer part of the resulting structure has $1/r^3$ profiles, but as far as I know no effort has been made to investigate the profiles in a cosmological setting, *while* they are growing in realistic hierarchical conditions. Dekel has recently attempted such a simulation which seems to show $1/r^2$ structure, but it is of limited dynamic range. Remember also that the "extended" rotation curves probably sample only the inner parts of the real halo distribution.

PEEBLES

The value $n = -1$ you favor for the spectrum $\sim k^n$ of primeval density fluctuations makes the integral J_3 diverge as $J_3 \propto r$; so your power law must be truncated at $r \lesssim 30$ Mpc to avoid excessively large Sachs-Wolfe effect.

GUNN

I fully agree that $n = -1$ cannot extend to very large masses, so one can argue whether one should use a power law at all.

THE COLOURS OF FAINT RADIO GALAXIES

— Finding Ellipticals at $z > 0.5$ —

H. VAN DER LAAN and R.A. WINDHORST
Sterrewacht Leiden, The Netherlands

1 - INTRODUCTION

The automation of deep plate photometry has increased the magnitude range and the statistics of reliably complete, flux-limited galaxy samples enormously in recent years (Kron 1978; Tyson and Jarvis 1979; Ellis 1979). Near the faint end of the apparent magnitude scale it is not possible to distinguish the different galaxy types morphologically. Because the galaxy luminosity distribution is very wide and the colour evolution as a function of cosmic epoch depends strongly on the time-dependent star formation rates, which probably means on galaxy type, it is not possible to relate, even approximately, the apparent magnitude and a single colour of a faint galaxy image to its redshift. If one wishes to study the spectral evolution of a single galaxy type over a substantial redshift range, one has to distinguish members of that type from all other galaxy images on the same plate, mixed in colour and in apparent magnitude with the type to be selected.

As discussed by Gunn in his review at this Study Week, one can attempt to single out giant elliptical galaxies by searching for clusters on deep plates and by taking spectra of the brightest cluster members, to distinguish them from foreground galaxies and from companion spirals in the same clusters. Such a program is very demanding in large-telescope time, even for optimally quantum-efficient spectrometers. We have followed another route to forming substantially large samples of giant elliptical galaxies near the faint end of deep plate magnitude scales. In this contribution we briefly report the method and its current results.

2 - Finding faint luminous elliptical galaxies

Our method uses a combination of deep radio images and optical plates of the same fields. It relies upon the fact, known for twenty years and established with the ever increasing numerical weight of hundreds of measured redshifts, that strong extragalactic radio sources, those with monochromatic powers above the break at $P_{1415\ MHz} = 10^{24.4}$ W Hz^{-1} in the radio luminosity function, if they are not quasars, are uniquely and without exception associated with luminous early type galaxies, ellipticals and S0's. For these radio galaxies the dispersion in absolute magnitude for nearby ($z \leq 0.1$) objects is very small. Given the very large range of radio flux densities ($\sim 10^5$ from the flux limit of the twenty brightest sources in the sky to our faintest survey limits) and apparent magnitude limits of $m_F = 23$, it is possible in principle to find large samples of giant galaxies in the range $z = 0.5$ to 1.0 without time-consuming spectroscopy. Optical colour distributions for such samples can be related to galaxy population evolution models very directly, while a good model fit yields a direct transformation of an apparent magnitude scale into one for redshift. The essential step is to find the giant ellipticals among the tens of thousands of faint images per plate.

Early phases of this radio-optical program were summarized by Katgert, De Ruiter and Van der Laan (1979). Westerbork radio survey sources were identified, on PSS plates and on KPNO 4 m plates. The colour-magnitude diagram of the identified radio galaxies was substantially different from that expected for either *non-evolving* or *passively evolving* giant ellipticals. A weakness of these early results was the lack of accurate photometry: apparent magnitudes were estimated visually. However, the simple fact that nearly all radio galaxies near the red plate limit are still visible on the blue plates, is a strong indication of colour evolution if they are ellipticals.

Several new major surveys have been completed at Westerbork for fields where deep KPNO 4 m plates had been obtained. Elsewhere in this volume this material is described. Radio positional accuracies of the order of one arc second enabled reliable identifications of $\sim$ 50% of all the radio sources in our fields on the Mayall telescope plates. Accurate absolute photometry using a PDS machine and the algorithms described by Kron (1978), yields magnitudes, hence colours. The observational sample presently in hand is a factor of five greater than what can be shown in this interim report, because only for

one hundred or so galaxies is our photometry complete. The radio surveys will be published by Katgert, Katgert-Merkelijn, Robertson, Windhorst and Van der Laan (in preparation). The optical photometry follows tbe methods developed by Kron (1978, 1980). The optical identifications, and a discussion of their reliability, completeness and photometric measures will be published by Windhorst, Koo and Kron (in preparation) for the Leiden-Berkeley Survey and by Windhorst (in preparation) for the Westerbork-Einstein Survey.

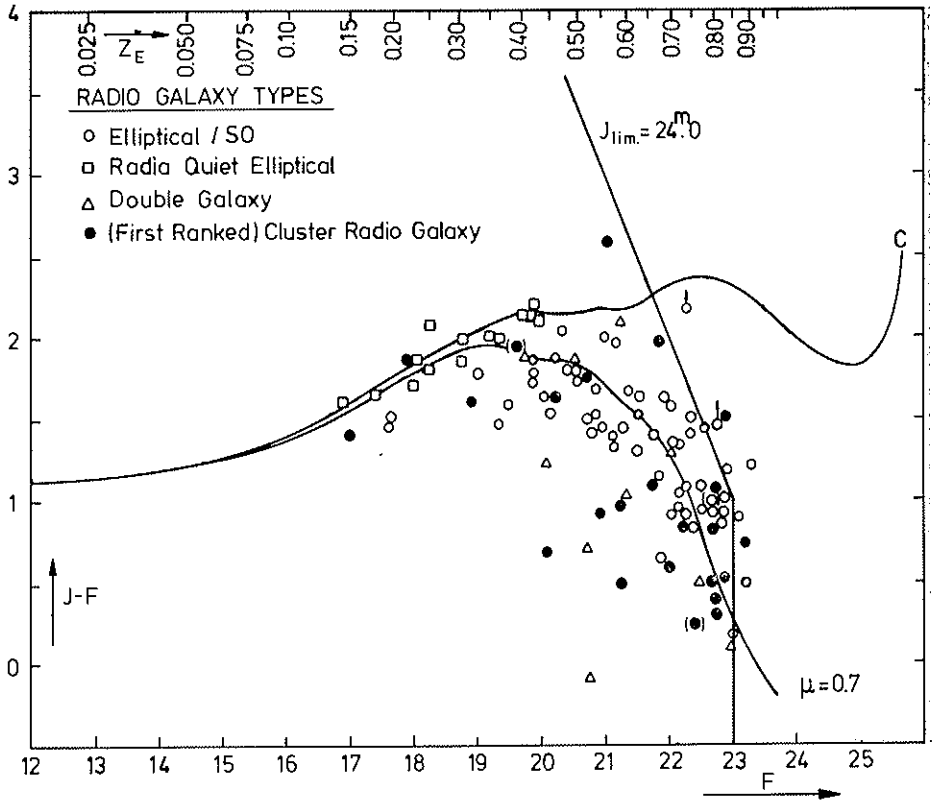


FIG. 1. The colour-magnitude diagram for fourteen radio quiet ellipticals and for one hundred and two radio galaxies. See Kron (1978) for the J (4650 Å) and F (6100 Å) pass band profiles. The redshift scale at the top of the diagram is the z to F match for the $\mu = 0.7$ model in a Friedmann cosmology with $H_0 = 100$, $q_0 = 0$ and $< M_F > = -21.^m5$. The C-curve is a passively evolving elliptical with star formation completed in one Gyr. The $\mu = 0.7$ curve is for continued star formation at an exponentially decreasing rate which had turned 70% of the model galaxy mass into stars after 1 Gyr (Bruzual 1981).

3 - The colours of faint radio galaxies

In Figure 1 the J-F vs. F distribution for about one hundred radio galaxies and for fourteen radio quiet ellipticals is shown. (See Kron 1978 for photometric passbands). Also plotted are two model curves from the work by Bruzual (1981) on evolutionary synthetic models of stellar populations. The curve designated 'C' is a model where all star formation occurs early, at a constant rate and is complete in 1 Gyr, after which the galaxy evolves passively. The curve labelled $\mu = 0.7$ is for a model with continuing but exponentially decreasing star formation rate (SFR), where 70% of the gas has been transformed into stars by the end of the first Gyr.

Photometric errors lead to 1 σ uncertainties of $0.^m1$ at $F = 20$ to $0.^m3$ near the plate limit. The one to two magnitude departure of the actual colours from the C model prediction rules out the latter as a satisfactory representation. The dispersion in J-F is large and the $\mu = 0.7$ curve cannot be called a good fit. The distribution is clearly much closer to the $\mu = 0.7$ model than to the C model: continuing star formation is strongly implied by a comparison of such models and this distribution. Uncertainties concerning the blue light contribution of horizontal branch star evolution inhibit relating our colour distribution to any one single model prediction unequivocally. (Bruzual and Kron 1980; Tinsley 1980; Bruzual 1981). Another caveat concerns our selection: are strong radio galaxies *typical* giant ellipticals or does the radio power peculiarity correlate with differences in stellar population evolution? Nearby radio galaxies are not, generally, distinguishable in broad band colours from non-radio giant ellipticals. It is important to attempt spectroscopic z-determinations for some of these galaxies, near the $\mu = 0.7$ curve, to compare them with the model's redshift to apparent magnitude transformation.

Acknowledgement

We thank Richard Kron (Chicago) and David Koo (Berkeley) for contributing the optical data to our collaborative radio-optical surveys.

REFERENCES

Bruzual, A.G. and Kron, R.G., 1980, *Ap. J.*, **241**, 25.
Bruzual, A.G., 1981, Ph.D. thesis, University of California, Berkeley.
Ellis, R.S., 1979, *ESO Workshop on Two Dimensional Photometry*, Noordwijkerhout, p. 339.
Katgert, P., de Ruiter, H.R. and Van der Laan, H., 1979, *Nature*, **280**, 20.
Kron, R.G., 1978, Ph.D. thesis, University of California, Berkeley.
— 1980, *Phys. Scripta*, **21**, 652.
Tinsley, B.M., 1980, *Ap. J.*, **241**, 41.
Tyson, J.A. and Jarvis, J.F., 1979, *Ap. J. Letters*, **230**, L 153.

DISCUSSION

PEEBLES

Could you fit your data using the evolution model "C" by reducing the time since galaxy formation?

VAN DER LAAN

No. The C-models lose their blueness very quickly. The measured colours cannot possibly be reproduced by any acceptable time shift.

EVIDENCE FOR THE COSMOLOGICAL EVOLUTION OF THE STELLAR CONTENT OF RADIO GALAXIES

S.J. LILLY and M.S. LONGAIR
Department of Astronomy, University of Edinburgh
Royal Observatory, Blackford Hill
Edinburgh

1 - Introduction

It has long been known that normal giant elliptical galaxies observed at large redshifts cease to be optical objects but become infrared sources, the maximum of the energy distribution occurring at about a wavelength 1.3 $(1 + z)$ μm, where z is the redshift. Thus, it is of considerable interest to study these galaxies at large redshifts in the infrared waveband and to compare them with nearby galaxies.

A major problem is to find suitable galaxies which are bright enough to be observed at redshifts of order 1. Simon Lilly and I have been studying this problem using those giant elliptical galaxies associated with strong radio sources (Lilly and Longair 1982a, b). This procedure has the great advantage that we can select very distant giant elliptical galaxies in a systematic way and that, for a number of them, redshifts are available.

Our work has concentrated on radio galaxies from the 3CR catalogue and in our preliminary survey observations of 35 galaxies spanning the range of redshift from 0.03 to 1 have been made with the UK Infrared Telescope (UKIRT) on Mauna Kea, Hawaii. The observations were made in the J (1.2 μm), H (1.65 μm) and K (2.2 μm) wavebands. The apparent optical magnitudes of the galaxies range from 14 to 24. The optical morphologies of these galaxies are similar to those of first ranked elliptical galaxies, including some cD systems and N-galaxies, the latter being associated with broad-line radio galaxies (BLRGs). The current sensitivity

of infrared detectors is such that it has been possible to detect even the faintest of these galaxies in the infrared waveband. In addition, it has been possible to detect some radio galaxies which have no optical counterpart in the optical waveband. We have confirmed the detection of 3C 68.2 reported by Grasdalen (1980) and have obtained a marginal detection of 3C 437.

2 - Colour-redshift diagrams for 3CR radio galaxies

The simplest way of presenting the results is in terms of colour-redshift diagrams, the (H-K) and (J-K) diagrams being shown in Figures 1 and 2.

Considering first the low redshift galaxies ($z < 0.4$), it is clear that, with the exception of the four galaxies classified as BLRG by Grandi and Osterbrock (1977), and represented by crosses on the diagram, the infrared colours of the radio galaxies occupy a well defined locus on the colour-redshift planes. This implies that they may all be represented, with small

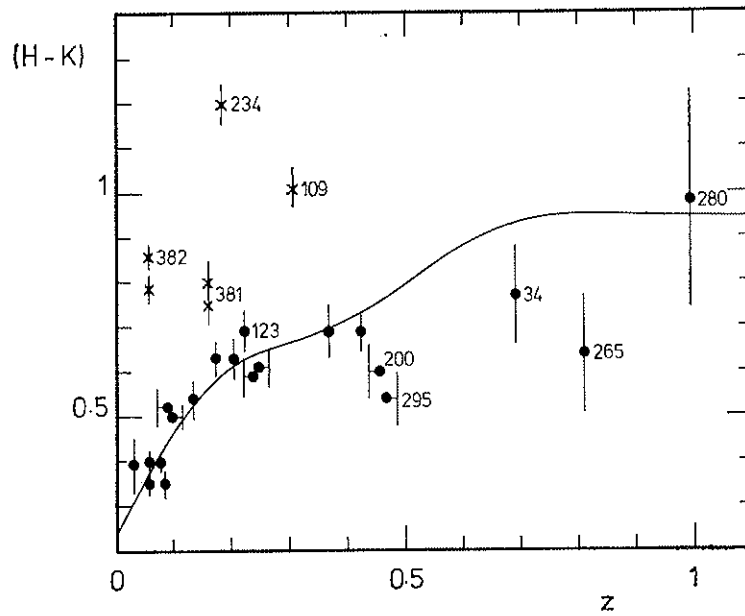


Fig. 1. The (H-K)-redshift relation for 3CR radio galaxies. The crosses are N-galaxies with strong non-thermal components.

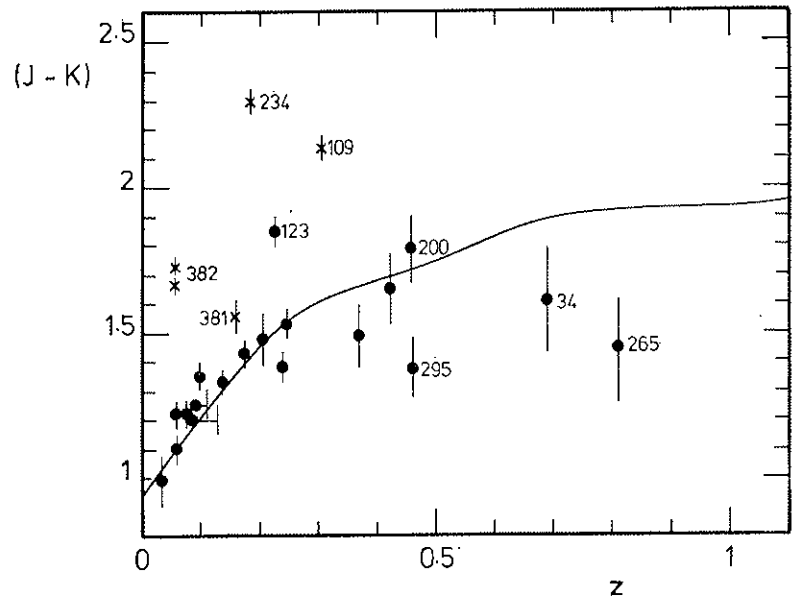


FIG. 2. The (J-K)-redshift relation for 3CR radio galaxies. The crosses are N-galaxies with strong non-thermal components.

cosmic scatter, by a single energy distribution. This was derived using the observed infrared colours of the galaxies with $z < 0.4$ and is shown by the solid line on the diagrams. This mean energy distribution for a giant elliptical galaxy is shown in Figure 3.

The zero redshift intercepts of the colours predicted from this energy distribution are very similar to those found in a large sample of nearby elliptical galaxies by Frogel *et al.* (1978). We can, therefore, conclude that the infrared energy distributions of these galaxies are essentially the same as those of normal elliptical galaxies, and that any additional component associated with the active nucleus must be small. There is no obvious relation between emission line strength and infrared properties amongst the narrow line radio galaxies.

At higher redshifts the galaxies broadly follow the predicted relations computed on the basis of the infrared spectrum constructed earlier and the energy distribution shortward of 1 μm of Coleman *et al.* (1980). There is no strong evidence for colour evolution in the infrared, and this is as predicted by conventional evolutionary models of elliptical galaxies, for example those of Bruzual (1981). Lebofsky (1981) has recently published

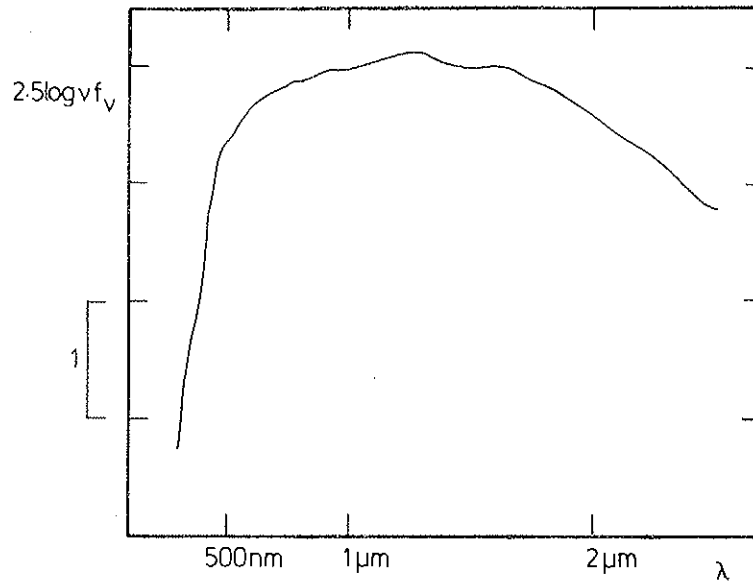


FIG. 3. The standard giant elliptical galaxy spectrum derived from the radio galaxies without strong non-thermal components with redshifts $z < 0.4$ and the elliptical galaxy spectrum of Coleman *et al.* (1980).

(H-K) colours of gE galaxies over a similar redshift range. We remark that the two high z 3CR galaxies considered by her to have anomalous (H-K) colours are not anomalous on our diagram, because of our improved determination of the mean galaxy spectrum, and that we do not find evidence for the large scatter in low z galaxies.

The BLRGs are clearly red in both (J-K) and (H-K), and from the (K, z) relation it is deduced that these are brighter than the typical galaxies in our study at all the wavelengths observed. Sandage (1973a) has shown that these systems may be successfully decomposed into a central nuclear component situated in a normal galaxy, and we have followed a similar procedure for 3C 109, 234 and 382, using our observations of the other galaxies to define the colours and magnitudes of the underlying galaxy. To within the uncertainties of this subtraction procedure, we find that the additional component has a power-law spectrum and that, in the case of 3C382, this extends to 3.5 μm. In addition, the variability of this latter source, which we observed over a 6 months timebase, is compatible with a nuclear component of approximately constant spectral index ($\alpha \approx 1.5$) but of varying intensity. In the small sub-sample for which there exists

good spectrophotometry we find that the BLRGs displaying an infrared excess have more than 10 times the Hβ flux as compared with those other galaxies which do not.

The optical-infrared colours may be similarly constructed from published optical photometry (e.g., Sandage 1972b, 1973b; Kristian *et al.* 1978; Smith *et al.* 1979), although greater uncertainty will be introduced. The (R-K) and (V-K) colours as a function of redshift are plotted in Figures 4 and 5. At high redshift the CCD *r* magnitudes were transformed to the R system using an extension of the colour equation given by Wade *et al.* (1979). In the diagrams the solid lines are the predicted colour-redshift relations based upon the infrared energy distribution derived earlier and the optical spectrum from Coleman *et al.* (1980). The other lines are various evolutionary models from Bruzual (1981). The reddest model assumes no evolution, the others representing different histories of the star formation rate (SFR), being either a constant burst of duration 1 Gyr (the C model) or an exponential decay of the SFR with 0.7 and 0.5, respectively, of the mass of the galaxy in stars at the end of the first 1 Gyr. At high redshifts it is clear on both diagrams that there are large deviations from the relations

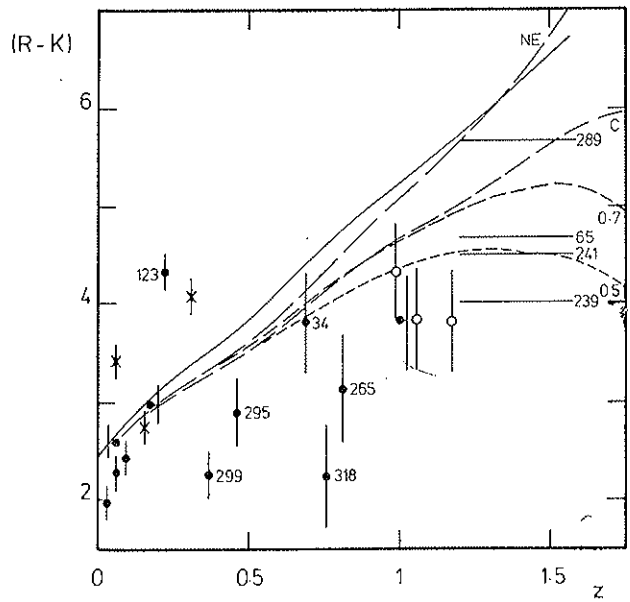


FIG. 4. The (R-K)-redshift relation for radio galaxies. The straight line shows (R-K) values for galaxies of unknown redshift. Open circles are from Lebofsky (1981).

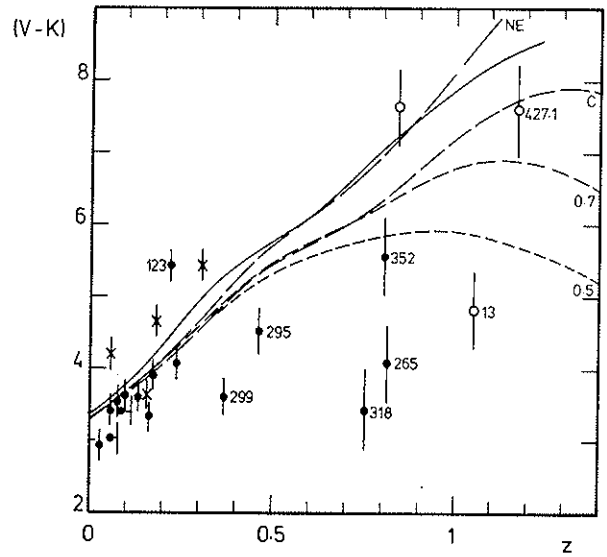


FIG. 5. The (V-K)-redshift relation for radio galaxies. The uppermost lines show the predicted relation of the galaxies which undergo no colour evolution.

predicted by the non-evolving models. We have included the colours derived from the K magnitudes of the high z 3CR galaxies observed by Lebofsky (1981), and these are represented by open circles. We cannot exclude the possibility that part of the blue colours of sources such as 3C 265 may have a non-thermal origin. In this case, Smith *et al.* (1979) have shown that the optical spectrum does not possess stellar absorption features. For the remainder of the galaxies, however, the V-K colours can be accounted for by the evolving galaxy models, with relatively slowly decaying SFR. We have also plotted on the (R-K) diagram the colours of 4 very faint galaxies of unknown redshift, and it may be seen that these galaxies have colours that are consistent with this conclusion. This result has implications for attempts to derive the redshift of distant galaxies from their colours alone.

3 - THE INFRARED HUBBLE DIAGRAM FOR RADIO GALAXIES

Because essentially all the observations were made through the same aperture, we plot on the Hubble diagram, Fig. 6, the magnitudes as observed,

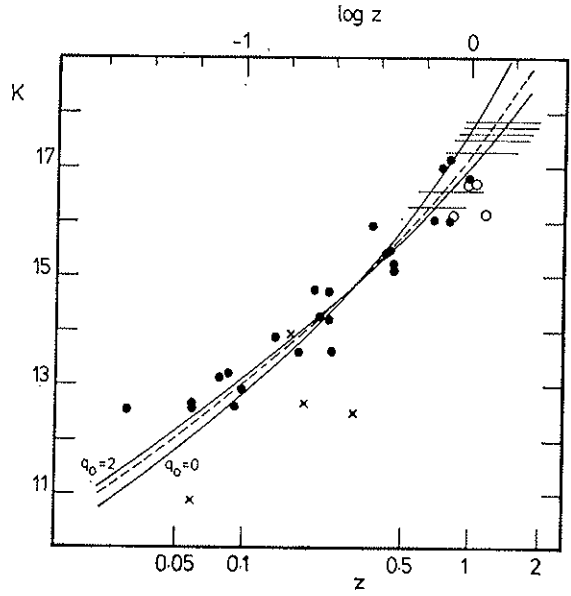


FIG. 6. The redshift-K magnitude relation for radio galaxies. The crosses are N-galaxies and the open circles are data from Lebofsky (1981). The solid lines, labelled $q_0 = 0$ and $q_0 = 2$, show the predicted redshift-magnitude relation for these world models. The dashed line shows the predicted relation in a $q_0 = 0$ model with evolution described by Bruzual's model with $\mu = 0.5$.

and we incorporate the K-corrections (which differ in the evolving and non-evolving models) and aperture corrections (which are different for different world models) into the predicted magnitude-redshift relations for different values of q_0. The aperture corrections were derived from the beam profiles and from the curve of growth given by Sandage (1972a). The best fit model may then be determined by a Chi-squared test, with the absolute magnitude/Hubble constant combination a free parameter in each world model. This method also gives the certainty with which other models may be rejected. Also shown on the diagram are the sources with unknown redshifts and the high z data from Lebofsky (1981), as horizontal lines and open circles, respectively. With the exclusion of the BLRGs, the radio galaxies form a well defined Hubble relation, the cosmic scatter about the best fit models being 0.40 mag, which is very similar to that found at optical wavelengths by Smith (1977). For the unevolving energy distribution, the apparent, uncorrected value of q_0 is considerably in excess of 1. The best fit value occurs at 2.8, and the $q_0 = 1$ model may be rejected with 70% confidence. The

relations for $q_0 = 0$ and 2 models are shown in Figure 6. However, the lack of infrared colour evolution does not imply a lack of luminosity evolution in the infrared. The effect of the $\mu = 0.5$ model discussed in connection with the optical-infrared colour diagrams is to increase the predicted K luminosity at a redshift of 1 by 0.8 mag. Application of the K-corrections appropriate to this model result in a best fit value of $q_0 = 0.5$. Therefore, if lower values of q_0 are preferred for other reasons, then the Hubble diagram provides further evidence of substantial evolution over lookback times of half the Hubble time.

4 - Conclusions

The data indicate the great potential of observations in the infrared waveband for the study of the evolution of the stellar component of giant elliptical galaxies over cosmological time-scales. The present results are preliminary but suggest that the infrared properties of galaxies are less susceptible to strong evolutionary effects than the optical properties. Combining the optical and infrared properties provides additional constraints on the evolutionary effects which influence the magnitude-redshift relation.

REFERENCES

Bruzual, G., 1981, PhD dissertation, University of California, Berkeley.

Coleman, G.D., Wu, C.C. and Weedman, D.W., 1980, *Astrophys. J. Suppl.*, **43**, 393.

Frogel, J.A., Penson, S.E., Aaronson, M. and Matthews, K., 1978, *Astrophys. J.*, **220**, 75.

Grandi, S.A. and Osterbrock, D.E., 1977, *Astrophys. J.*, **195**, 255.

Grasdalen, G.L., 1980, *IAU Symposium No. 92, "Objects with Large Redshift"* (ed. G.O. Abel and P.J.E. Peebles), 269.

Kristian, J., Sandage, A.R. and Westphal, J.A., 1978, *Astrophys. J.*, **221**, 383.

Lebofsky, M., 1981, *Astrophys. J. (Letters)*, **245**, L 59.

Lilly, S.J. and Longair, M.S., 1982a, *IAU Symposium No. 97 "Extragalactic Radio Sources"* (ed. D. Heeschen, in press).

— 1982b, *Mon. Not. R. astr. Soc.*, (in press).

Sandage, A.R., 1972a, *Astrophys. J.*, **173**, 485.

— 1972b, *Astrophys. J.*, **178**, 25.

— 1973a, *Astrophys. J.*, **180**, 687.

— 1973b, *Astrophys. J.*, **183**, 711.

Smith, H.E., 1977, *IAU Symposium No. 74, "Radio Astronomy and Cosmology"*, (ed. D.L. Jauncey), 279.

Smith, H.E., Junkarinen, V.T., Spinrad, H., Grueff, V. and Vigetti, M., 1979, *Astrophys. J.*, **231**, 307.

Wade, R.A., Hoessel, J.G., Elias, J.H. and Huchra, J.P., 1979, *Publ. astr. Soc. Pacific*, **91**, 35.

IV.

EVOLUTION OF QUASARS, RADIO GALAXIES AND THE X-RAY BACKGROUND RADIATION

THE SPACE DISTRIBUTION OF QUASARS

MAARTEN SCHMIDT
Palomar Observatory,
California Institute of Technology

and

RICHARD F. GREEN
Steward Observatory,
University of Arizona

1 - Introduction

The space density of quasars increases steeply with distance out to a redshift of three at least. This phenomenon is a consequence of the variation of the quasar luminosity function with time, which reflects the collective result of the births and subsequent evolution of quasars.

The first evaluation of the variation of the space density was based on quasars that are strong radio sources (Schmidt 1968). Quasars contribute a large fraction (around 25 per cent) of strong extragalactic radio sources, hence early studies concentrated mostly on these. However, the interpretation of the space distribution in terms of the variation of the luminosity function is complicated by the fact that the latter is a function of both optical luminosity and radio luminosity, which may be correlated (Schmidt 1970).

The derivation of the space distribution is, in principle, more straightforward for quasars that are selected purely optically. Until recently, only a small complete sample of optically selected quasars was available. Several surveys have been completed now and we base our discussion of the space distribution on this material.

2 - Optical quasar surveys

We confine ourselves to surveys of spectroscopically confirmed quasars with fairly accurate optical magnitudes. The Braccesi list of quasar candidates (Braccesi, Formiggini, and Gandolfi 1970) contained 175 objects. Spectroscopic work is only complete for objects brighter than $B = 18$, resulting in a complete sample of 19 quasars over an area of 36 square degrees.

Objective-prism surveys have increased dramatically the number of optically selected quasars. Slit spectra are generally required to check the quasar nature and the line identifications. Confirmed lists are now available for a Schmidt survey (Osmer and Smith 1980) and the Tololo 4-meter reflector survey (Hoag and Smith 1977; Osmer 1980). The Schmidt survey is not complete beyond $B = 18$. For the 4-meter survey I assume that the completeness limit is $B = 19.5$. These surveys depend on the detection of Lyman-α emission in the objective prism spectra, hence the redshift should be larger than 1.8.

A large-scale search for bright quasars was started at Palomar in 1972 (Green 1976). Green covered some 10,700 square degrees with 2 color (U, B) exposures with the 18-inch Schmidt telescope and selected several thousand stellar objects with an ultraviolet excess ($U\text{-}B < -0.4$). We have undertaken spectroscopic observations of all objects and have found 108 quasars which constitute a complete sample down to an average B limit of 16.2 mag. (A few low-redshift objects have been added to the list since the Study Week; these have little effect on the results discussed below).

Recently, Kron and Chiu (1981) have obtained a small, deep sample of quasars in Selected Area 57, based on lack of proper motion, grating prism observations, colors and variability. They find 4 spectroscopically confirmed quasars with $B < 21$ in an area of 0.071 square degrees, corresponding to a surface density of around 60 per square degree.

Figure 1 shows the Hubble diagram for all quasars in the above surveys that are brighter than the adopted limiting brightness for completeness. The line corresponds to $M_B = -23$ for a Hubble constant $H_0 = 50$ km s^{-1} Mpc^{-1}. Many of the objects below this line may be classified as nuclei of Seyfert galaxies.

3 - Distance scale of quasars

Between the effective limit of the Palomar Bright Quasar Sample

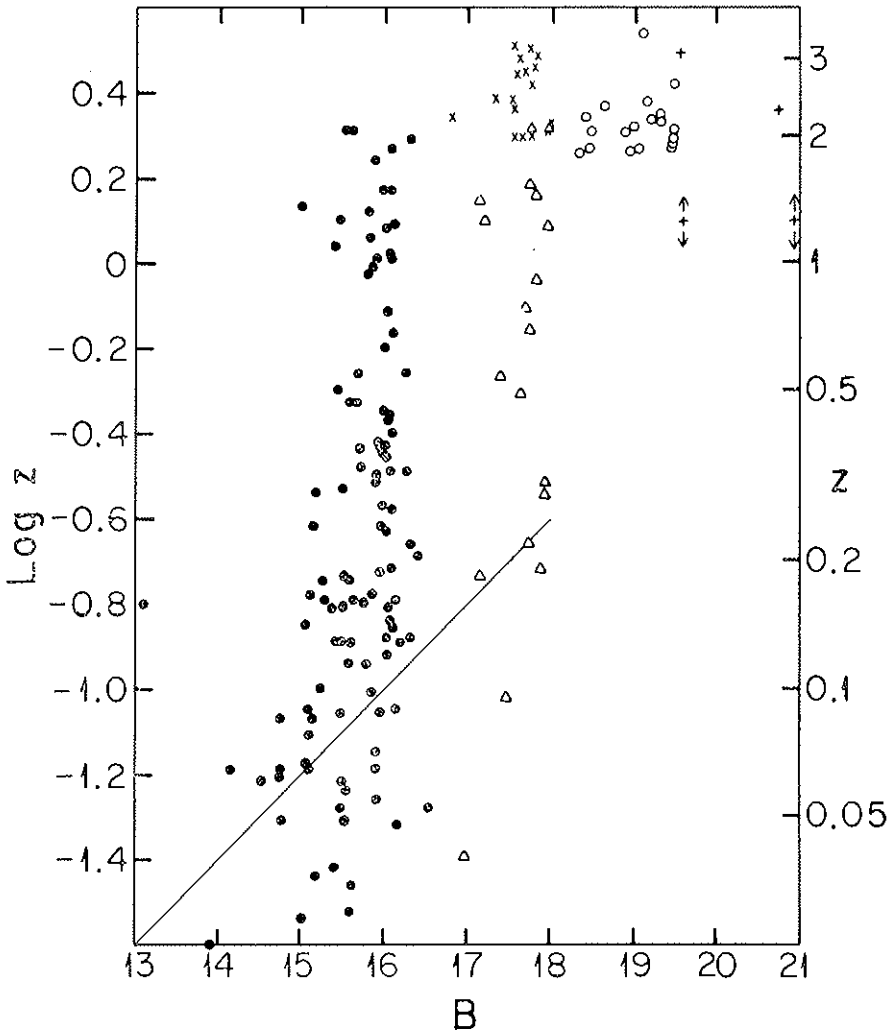


FIG. 1. Hubble diagram for five samples of optically selected quasars. All quasars shown are brighter than the adopted survey limiting magnitude B_{lim}.

● : Palomar Bright Quasar Survey, area 10,714 sq. deg., B_{lim} = 16.2 (average)
Δ : Bracessi sample, area 36 sq. deg., B_{lim} = 18.0
× : Curtis-Schmidt sample, area 340 sq. deg., B_{lim} = 18.0
o : Tololo 4-meter sample, area 5.1 sq. deg., B_{lim} = 19.5
+ : Kron-Chiu sample, area 0.071 sq. deg., B_{lim} = 21 (two undetermined redshifts so indicated).

(B = 16.2) and that of the Braccesi sample (B = 18), cumulative counts of quasars increase by a factor of 8.5 per magnitude, or $d\log N/dB = 0.93 \pm 0.06$. This slope is significantly steeper than the slope of 0.60 that would be observed for objects having a uniform distribution in Euclidean space. In fact, the observed slope would correspond to a space density rising with distance r as $r^{1.6}$. Hence, if quasars were "local", i.e., residing in Euclidean space, we ourselves would be situated in a deep density minimum in the quasar cloud — a conclusion to be rejected on Copernican grounds. This argument only breaks down at quasar distances that are so large that we are looking back in time over a substantial fraction of the age of the universe. Hence, this argument leads to a distance scale for qnasars that is of the same order as that corresponding to the cosmological interpretation of the large observed redshifts.

4 - Statistical evolution

If the Hubble (B, log z) diagram of all quasars were known, then the distribution of the B magnitudes at a given redshift would yield directly the luminosity function at that redshift. The small number of objects, and in particular, the incomplete coverage of the Hubble diagram as shown in Figure 1 makes this procedure unfeasible. Instead, we use interpretation techniques based on the expectation that the luminosity function varies smoothly as a function of luminosity and redshift. In fact, we assume that the co-moving space density of quasars varies as exp ($k\tau$), where τ is the light travel time expressed in the age of the universe and k is a function of the absolute magnitude M_B. We also assume that the continuum energy distribution F(ν) has a spectral index $d\log F/d\log \nu = \alpha$ equal to -0.5, corresponding to typical quasar colors as found by Richstone and Schmidt (1980).

Since the detailed derivation of the statistical evolution of quasars is being prepared for publication elsewhere, we present here some of the main results. These are based on a value of the deceleration parameter $q_0 = 0.5$ (or $\Omega = 1$) and a Hubble parameter $H_0 = 50$ km s^{-1} Mpc^{-1}. We find that the different surveys can be well represented by adopting $k = -3\,(M_B + 21.6)$. The only exception is the Curtis Schmidt survey for which predicted numbers are more than twice the observed numbers. The ap-

parent deficiency of this survey relative to the 4-m Tololo survey has been discussed by Osmer (1980).

According to the $k\,(M_B)$ relation given above the brightest quasars, with $M_B = -30$, will have $k = 25$, i.e., their co-moving space density varies with a characteristic time of only 0.04 of the age of the universe. Quasars of low absolute luminosity, with $M_B = -23$, show little variation of space density with cosmic epoch.

Total numbers of quasars brighter than $M_B = -23$ vary from 28 Gpc^{-3} at $z = 0$, to 1600 Gpc^{-3} at $z = 1$ and 10,000 Gpc^{-3} at $z = 3$. The luminosity function at different redshifts is shown in Figure 2.

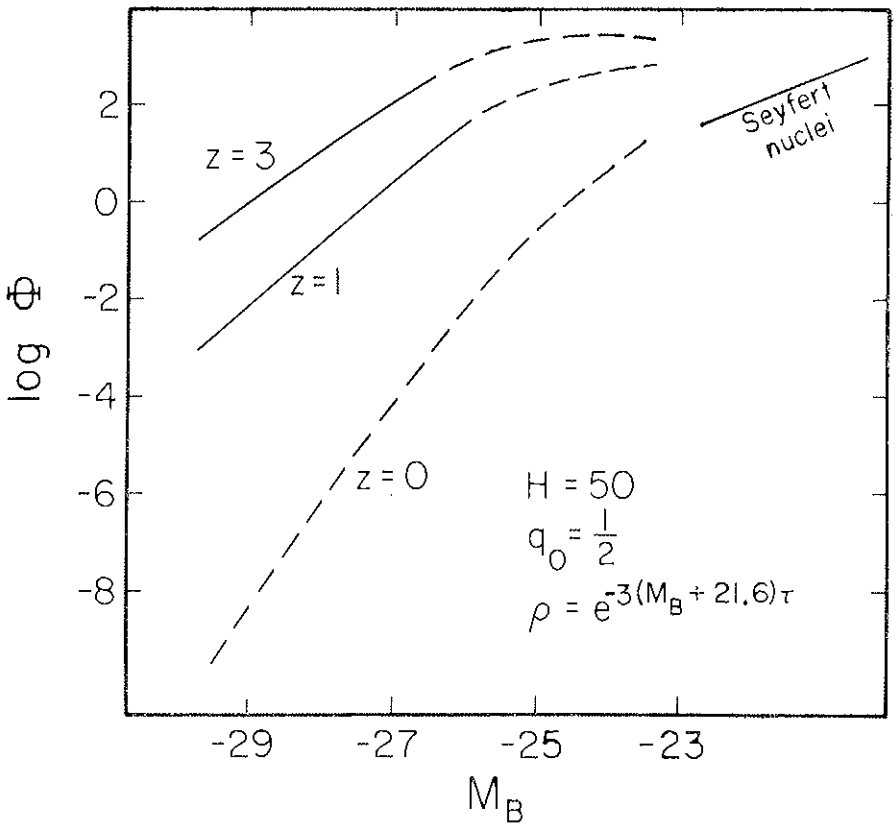


FIG. 2. Luminosity function $\phi\,(M_B)$ of quasars at different redshifts, in units of Gpc^{-3} mag^{-1}. Parts of the luminosity function directly supported by observations are indicated by a solid line. Other parts depend on the adopted space density evolution law.

5 - Quasars at large redshifts

It has long been suspected that the upper limit of the observed redshifts near 3.5 is not due to observational selection but, instead, reflects a real scarcity or absence of objects with larger redshifts. A relevant study was finished recently by Osmer (1982) who searched an area of 5 square degrees by a prism grating technique with the 4-meter Tololo reflector. He should have been able to detect Lyman-α emission from quasars with redshifts in the range 3.7 to 4.7 down to an equivalent B magnitude of 20.5. While he found some 15 objects with emission lines, none of these turned out to be Lyman-α.

The statistical evolution model discussed above allows us, by a small extrapolation, to predict the number of quasars in the redshift range 3.7 to 4.7. We find that Osmer should have seen some 20 quasars, in marked contrast to his failure to detect any. While further searches for quasars of large redshift should be conducted, at present our conclusion can only be that the number of observable quasars beyond a redshift of 3.5 is an order of magnitude lower than that predicted on the basis of a smooth extrapolation from lower redshifts.

The scarcity or lack of quasars at large redshift may reflect their turn-on time at one-tenth of the present age of the universe, or it may be caused by dust from galaxies in formation as proposed by Ostriker and Cowie (1981). In connection with the interpretation of this important phenomenon, it will be important to get more accurate information on the redshift distribution near the cut-off.

6 - Statistics of quasar births and deaths

If the lifetime (or the luminosity history) of individual quasars were known we could, on the basis of the observed population of quasars, derive the number of dead quasars and the birth rate of quasars. Lacking knowledge of quasar lifetimes and considering the large variation in absolute luminosities, we assume that each quasar during its lifetime radiates a fixed amount of energy, $E_{rad.}$

The total space density of all quasars brighter than $M_B = -23$ is derived as the integral of the birth rate from the time corresponding to redshift 3.5 to the present. We find

$$6 \times 10^{-4} \left(\frac{10^{61} \text{ ergs}}{E_{rad}} \right) \text{ Mpc}^{-3} ,$$

which for $E_{rad} = 10^{61}$ ergs is approximately the space density of galaxies brighter that $M_B = -20.5$. Hence, all such galaxies may contain an extinguished quasar, or a fraction of all galaxies above a lower luminosity limit may contain one. It should be remembered that E_{rad} is very uncertain.

For the detection of gravitational waves it is of interest to estimate the present observable birth rate of quasars. With the lifetimes discussed above, the present observable rate for all quasars brighter than $M_B = -23$ out to redshift 3.5 would be

$$0.6 \left(\frac{10^{61} \text{ ergs}}{E_{rad}} \right) \text{ yr}^{-1} .$$

7 - X-RAY BACKGROUND OF QUASARS

X-ray observations with the Einstein satellite have been obtained for more than 60 objects in the Palomar Bright Quasar Sample, in conjunction with H. Tananbaum *et al.* By combining these X-ray observations with the statistical evolution model discussed above, we can derive quasar X-ray counts as well as the integrated background caused by all quasars.

In this derivation we find it particularly useful to write the luminosity function as a sum of inverse density-weighted volumes V_{max}^{-1}, one for each quasar in the Bright Quasar Sample. Then, by "tagging" each of these contributions with the observed X-ray luminosity, we can easily derive total counts and background. This method implicity takes into account any correlation between X-ray and radio luminosities that may exist. It is also unnecessary to derive a mean ratio of X-ray to optical luminosity since the contributions of each X-ray observed quasar is "automatically" entered with the appropriate weight.

The details of the X-ray work on hright quasars will he published elsewhere. A tentative evaluation shows that predicted quasar X-ray counts are close to observed extragalactic X-ray counts, and that the 2 keV background due to quasars is around 47 per cent of the total observed background (Schwartz 1978).

Early high estimates of the X-ray background due to quasars (Setti and Woltjer 1979; Tananhaum *et al.* 1979) were based on a formula for quasar

counts due to Braccesi *et al.* (1980). For B = 21 this formula predicts 200 quasars per square degree, or about three times the number adopted in our present work.

8 - Need for further work

The evaluation of the statistical evolution of quasars suffers from the poor survey coverage of the Hubble diagram as exhibited by Figure 1. Further systematic surveys to fainter magnitudes are much needed. Once these are in hand, the variation of space density with redshift for quasars of given luminosity can be determined directly, so that no assumption about the functional form of the variation will be needed any more.

As mentioned above, further studies aimed at the precise density profile of the quasar space distribution near redshift 3.5 are of major potential interest.

This research was supported in part by the National Science Foundation under grant AST 77-22615A01.

REFERENCES

Braccesi, A., Formiggini, L. and Gandolfi, E., 1970, *Astron. and Astrophys.*, **5**, 264.
Braccesi, A., Zitelli, V., Bonoli, F. and Formiggini, L., 1980, *Astron. and Astrophys.*, **85**, 80.
Green, R.F., 1976, *Pub. Astr. Soc. Pac.*, **88**, 665.
Hoag, A.A. and Smith, M.G., 1977, *Ap. J.*, **217**, 362.
Kron, R.G. and Chiu, L.G., 1981, preprint.
Osmer, P.S., 1980, *Ap. J. Suppl.*, **42**, 523.
— 1982, preprint.
Osmer, P.S. and Smith, M.G., 1980, *Ap. J. Suppl.*, **42**, 333.
Ostriker, J.P. and Cowie, L.L., 1981, *Ap. J. (Letters)*, **243**, L 127.
Richstone, D.O. and Schmidt, M., 1980, *Ap. J.*, **235**, 377.
Schmidt, M., 1968, *Ap. J.*, **151**, 393.
— 1970, *Ap. J.*, **162**, 371.
Schwartz, D.A., 1978, *Proc. IAU/COSPAR Symposium on X-Ray Astronomy*, in press.
Setti, G. and Woltjer, L., 1979, *Astron. and Astrophys.*, **76**, L 1.
Tananbaum, H. *et al.*, 1979, *Ap. J. (Letters)*, **234**, L 9.

DISCUSSION

OSTRIKER

The argument for dust absorption can be made more simply than was done by Ostriker and Cowie. An ordinary galaxy like our own should be opaque at $\sim$ 1200 A to a radius of 15 to 30 kpc. Beyond a redshift of 3.5 the typical line of sight comes within this distance of a galaxy (if they exist at that redshift) and so objects beyond that distance will be obscured.

If you take your evolutionary model of the quasar distribution you can predict the log N - log S distribution at optical and X-ray wavelengths. You can make the same prediction if you supplement your model by the additional assumption that there are no quasars beyond $z = 3.5$. How different are the predicted distributions in these two hypothetical cases from one another?

SCHMIDT

As I mentioned the predicted X-ray counts, based on a redshift cut-off of 3.5, agree well with the observations. Hence a prediction without a cut-off, not yet carried out, will be of interest. For optical counts this is less so, since the counts up to $B = 21^m$ are fitted by construction and those beyond $B = 21^m$ are not yet reliable.

LYNDEN-BELL

Do we need to assume that the quasars disappear beyond $z = 4$ or just that Lα is suppressed as would be the case for neutral hydrogen on the way?

SCHMIDT

Lyman α could be reduced or perhaps suppressed if neutral hydrogen is in the way. In any case, it would be of interest to search for CIV emission at redshifts larger than 3.5, i.e., longwards of 7000 A.

SCIAMA

I would like to suggest that the apparent cut-off in the quasar distribution at a redshift of 3.5 could be due to the intergalactic medium becoming neutral at this redshift. If the IGM is photo-ionised, we would expect it to become transparent at a critical red shift rather suddenly, and this would fit in with the quasar cut-off being rather sudden (if it indeed is).

SCHMIDT

I should reiterate that the objective prism surveys do not agree well among themselves in the observed distribution of redshifts larger than 2.5. Hence it is not clear as to whether the ohserved cut-off is gradual or sudden.

SILK

If the quasar cut-off is due to dust absorption in intervening galaxies, one would expect to find both evidence for reddening of the spectrum and, more importantly, an increasing frequency with redshift of absorption line systems characteristic of cold interstellar gas, including molecular hydrogen.

OSTRIKER

Yes, we did check Osmer's sample of large redshift quasars for reddening. He calculates a curvature observed in each spectrum and we examined the dependence of the curvature found on quasar redshift. We found a strong statistically significant increase consistent with expectations based on normal dust obscuration but the smallness of the sample and difficulty of the measurement makes the result very uncertain.

REES

It is commonly thought that quasars are powered by a large collapsing mass or by a massive collapsed object (black hole), fuelled by accretion or by electromagnetic extraction of its spin energy. In the context of this model, the deficiency of quasars at small redshifts compared with the numbers at $z = 2$ to 3 may be due to the decline in the fuelling rate as the gas density in galaxies decreased.

There seem to be four ways of accounting for the decline in bright quasar numbers beyond $z = \approx 3.5$: (i) intervening absorption (by neutral hydrogen or dust) may be responsible; in this case radio emission and hard X-rays should still get through; (ii) galaxies may not form (and acquire well-defined central potential wells at their centres) until $z \simeq 4$; one would then see no bright discrete objects at larger redshifts; (iii) even if galaxies form much earlier, it may take $\gtrsim 10^9$ years for a black hole to grow to 10^8 to 10^9 $M_\odot$; one might then expect *intrinsically fainter* quasars at large z, but no ultraluminous ones; (iv) there is so much uncondensed gas and dust in young galaxies that quasar activity, even if it occurs, is "smothered"; the energy might then be reprocessed into a diffuse infrared source.

QUASARS IN THE UNIVERSE *

L. WOLTJER
European Southern Observatory
Garsching bei München

and

G. SETTI
Istituto di Radioastronomia, Università di Bologna, CNR

Quasars are relatively rare objects. Down to the 15th magnitude there are thought to be less than a hundred quasars over the whole sky, while there are a few million stars. At the 20th magnitude the situation is more favorable, but there still are ten to a hundred times as many stars as quasars (Figure 1). Hence, if quasars are to be discovered with reasonable completeness, sensitive selection criteria have to be used. The two most important of these are: (a) the (U-B) ultraviolet excess, possibly supplemented by other color criteria; (b) the presence of strong emission lines. In the first technique photographs in U and B light are taken and from these the objects with ultraviolet excess are found. These are then inspected spectroscopically. At bright magnitudes the admixture of hot white dwarfs, subdwarfs, etc. still is substantial, but beyond magnitude 17 the majority of the objects found are quasars. Most quasars with redshifts below $z = 2.2$ to 2.5 may be found this way, but at higher redshifts the strong Ly α line moves into the B hand and makes the color criterion less effective. However, objective prism techniques may then be used to find the emission line objects directly.

---

* Throughout this paper $H_0 = 50$ Km s^{-1} Mpc^{-1} has been adopted.

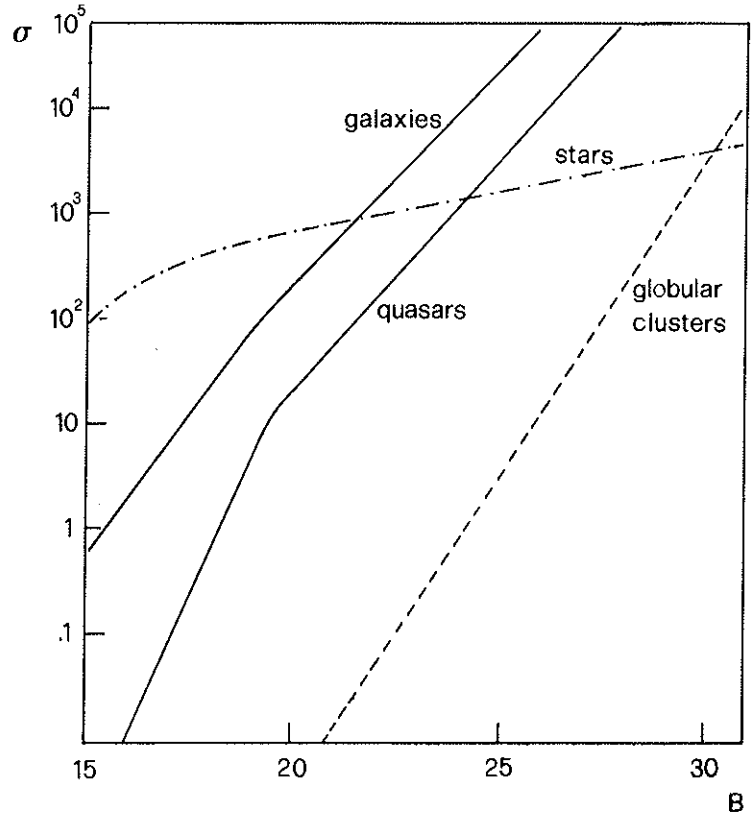


FIG. 1. Approximate integral numbers of stars, galaxies, quasars and globular clusters as a function of limiting B magnitude. The extragalactic globular clusters have been taken from van den Bergh (1979).

1 - CLASSIFICATION

A variety of related objects are generally considered to belong to the overall class here considered, but the criteria by which they are defined are usually imprecise and influenced by the accidents of historical developments. In order to achieve some clarity, at least in our own papers in these proceedings, we define the following classes and subclasses:

I. Quasars

Blue continuum (U-B excess in continuum in rest frame)
Broad emission lines
$M_v < -24$

Subclasses [1]:

QSO: "optical quasars"	Log R < 1.5
QSR: "radio quasars"	Log R > 1.5
subdivided in	
Steep (radio) spectrum	$\alpha < -0.7$
Flat (radio) spectrum	$\alpha > -0.5$

II. Seyferts 1

Blue nuclei of galaxies
Emission lines; total width H lines > 3000 km s^{-1}
$M_v > -24$

III. BL Lac Objects

Moderately blue
No emission lines (or very faint)
Variable optical polarization with $P_{max} > 5$ %
Variable luminosity with typically $\Delta B > 2$ magnitudes

Subclass:
OVV (optically violent variables) quasars: same as BL Lac, but more quasar-like emission lines.

Notes to classification

The quantity R (defined at 500 MHz) was originally introduced by Schmidt (1970). It is not evident that QSO and QSR are really separate classes or whether the transition is fully continuous. Fragmentary data on the luminosity function weakly suggest a bimodal distribution in R.

The separation of steep and flat spectrum sources is not always unambiguous in multicomponent radio sources.

At present, there is no evidence that quasars and Seyferts 1 differ in any important way except in absolute magnitude. Frequently, we shall

---

(1) $R = \dfrac{F(5\ \text{GHz})}{F(2500\ \text{A})}$ (in rest frame), $F_\nu \propto \nu^\alpha$

refer to the whole class somewhat loosely as "quasars". We shall treat all quasars with $M_V > -24$ as Seyfert 1, even when no surrounding galaxy is visible. We have chosen $M_V = -24$ (or sometimes $M_B = -23.8$) as the dividing line, since this was the discovery magnitude of Fairall 9 which seems to represent an appropriate transition case between "fuzzy" quasars and more typical Seyferts. It has been variously classified in the literature as one or the other.

The P_{max} and ΔB for BL Lac objects and OVV depend, of course, on the length of observation. For virtually all objects presently considered to belong to this class a ten year observation period (and in most cases a few years) would suffice to establish these characteristics. The object PHL 5200 which definitely does not belong to this class has $P = 4\ \%$ (presumably due to a scattering envelope), but $P > 1\%$ is very rare in true quasars (Stockman and Angel 1978).

2 - The Number-magnitude relation for quasars

Many searches for quasars have been made, but in the following we shall restrict ourselves to those for which sufficient information is available to estimate the completeness of the samples.

(1) Green and Schmidt have searched an area of 10,000 square degrees at relatively high galactic latitude for U-B excess objects down to about $B = 16$ on the average. Spectroscopy is available for nearly all of these, and a total of 105 quasars were identified. The resulting area density per square degree is $N (< 16) = 0.0105$. For a more complete discussion of this material see the paper by Schmidt in these proceedings.

(2) Braccesi *et al.* (1970) have found 175 U-B excess objects in a field of 37.2 square degrees. In a recent study by Véron and Véron (1982) a list is given of the 21 objects brighter than $b = 18.4$, corresponding to $B = 18.27$ (Setti and Woltjer 1973) which remain after the elimination of two extragalactic H II regions. Redshifts (mainly by Lynds) are available for all but one of these. Three corrections are still to be applied to the area density in this field. By repeating the counts Braccesi *et al.* (1980) found that the survey is slightly incomplete. From their data it follows that for $U\text{-}B < -0.6$ the counts have to be increased by a factor 1.13. The density of U-B objects is lower at the edge of the field than in the central parts. While the general inhomogeneity of the distribution precludes an accurate discussion, we derive a correction factor of 1.22 from a com-

parison of the area densities in the inner $4^{\circ}.5 \times 4^{\circ}.5$ and over the whole plate. This correction is comparable to that found by Usher (1976-77, Annual Report of the Hale Observatories, p. 161) in the Sandage-Luyten field at 8h for which the plates were also obtained with the 48-inch Palomar Schmidt.

A third correction is inferred from the near absence of quasars with U-B > -0.6. Among 3C and 4C radio quasars (at latitudes $> 45^{\circ}$) witb well determined photoelectric colors and with $z < 2.2$ there are 10 objects with $-0.4 >$ U-B > -0.6 against 48 objects with U-B < -0.6 (Figure 2). Taking all the quasars in the Braccesi sample, the numbers are 1 and 28. Although a small portion of the effect may be related to the selection of objects for spectroscopic investigation, and although the identity of the color distributions of the radio and optical quasars is not guaranteed, most of it is probably related to incompleteness in the sample (see also the discussion of Braccesi *et al.* 1980), and we therefore apply a further correction factor of 1.17, bringing the total correction to 1.61. The surface density of quasars in this field then becomes $N(< 18.27) = 0.91$, to which some large redshift objects still are to be added. From the CTIO data (Figure 3) this leads to a total $N(< 18.27) = 0.95$.

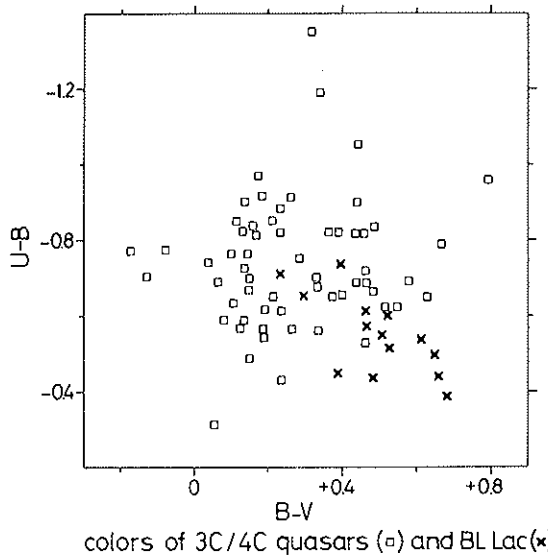


FIG. 2. Photoelectric colors for high latitude quasars from the 3C and 4C catalogues and for BL Lac objects. Only objects at high galactic latitudes are included.

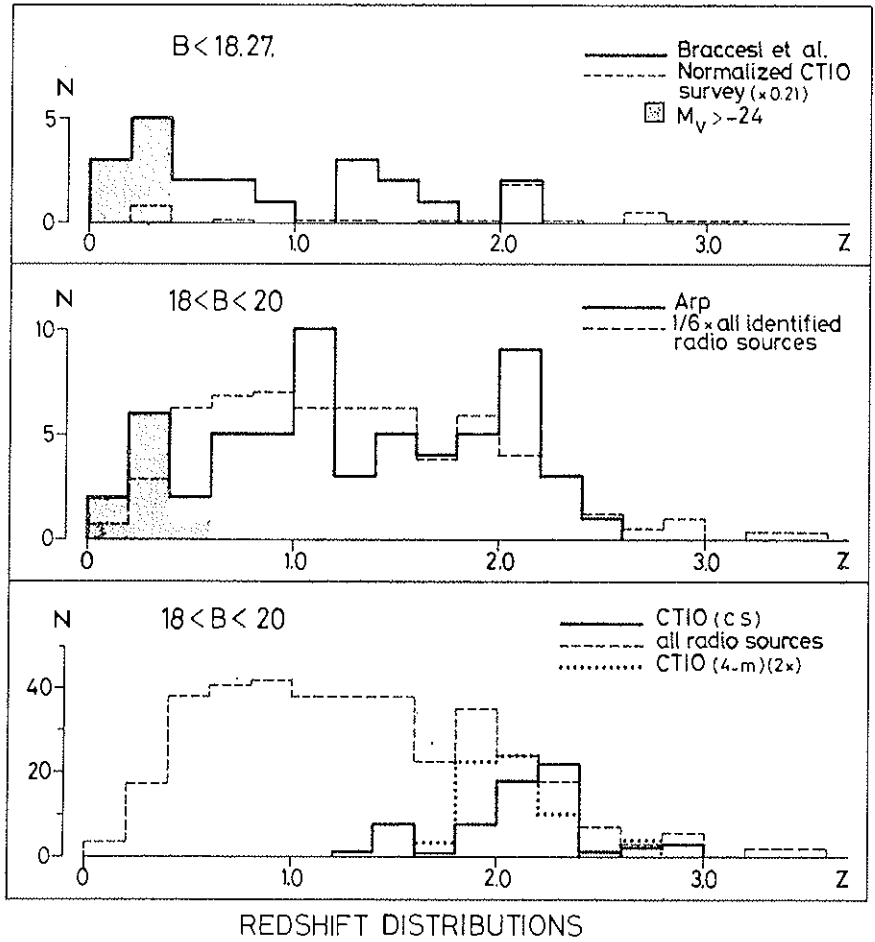


FIG. 3. The redshift distributions of various samples discussed in the text. The dark areas correspond to Seyferts respectively in the Braccesi *et al.* and in the Arp samples.

Some partial confirmation of the Braccesi counts of (U-B) excess objects comes from a study of Steppe, Véron and Véron (1979). In a field of 20.6 square degrees they compiled a catalogue of objects on the basis of surveys by Steppe (1978) and Berger and Fringant (1977). There are 16 stellar objects in this catalogue between B = 17 and 18.27 with U-B < − 0.60. In the Braccesi *et al.* catalogue, there are 27 such objects, corresponding to 15 in the area of Steppe *et al.*, or 21 after the incompleteness corrections have been applied.

(3) The Sandage-Luyten field at 8h has been studied extensively by Sandage and Usher using color and variability criteria. Although the published data are scanty, it is seen from the 1976-77 Annual Report of the Hale Observatories (p. 161) that 39 quasars with $B < 18.5$ have been confirmed spectroscopically in 42.7 square degrees. It is expected statistically that 11 more quasars will be found among the bluer objects for which spectroscopy is not yet available, as well as an unknown number among the somewhat redder ones. In addition, Usher found an incompleteness factor of 1.3 due to edge effects. This leads to a total of 65 objects in the field as a minimum. Adding as before also some high redshift objects on the basis of the CTIO survey, we come to an area density of $N(< 18.5) = 1.56$. For the Sandage-Luyten field at 15h similar results were found. With an increase of about a factor of 7 per magnitude in the integral N (B) relation, this would correspond to $N(< 18.27) = 1.01$. In view of the statistical uncertainties, of the possible addition of some redder quasars and of the sensitivity of the result to the magnitude scales, this is in satisfactory agreement with the result for the Braccesi *et al.* (1970) field.

(4) The reduced Braccesi *et al.* (1980) field consists of the central 1.72 square degrees of the full field. Here in a very extensive search 27 objects were found with $U\text{-}B \approx u\text{-}b < -0.60$ and $18.27 < B < 20$ (Formiggini *et al.* 1980). Since the corrections for missing quasars with $U\text{-}B > -0.60$ and for contamination of the sample by stars should be about equal and opposite, this leads to $N(< 20) \approx 17$, presumably for quasars with redshifts $z < 2.2$. It is to be emphasized that this number depends very strongly on the correctness of the magnitude scales which have been adopted, on the still somewhat uncertain composition of the sample and on its completeness. Correcting for missing quasars with $z > 2.2$ on the basis of the CTIO 4-m survey, we obtain for all quasars $N(< 20) = 19$. Véron and Véron (1982) have shown that, contrary to earlier reports, the quasars in this field have a really "stellar" appearance.

(5) The CTIO spectral surveys provide two samples with complete data for magnitudes and redshifts. The CTIO Curtis-Schmidt (CS) survey has been recently discussed by Osmer (1981) who found that the apparent nonrandomness of the quasar distribution is probably due to one batch of relatively insensitive photographic plates. Deleting these we are left with 190 square degrees of survey. Inspecting the data, we found a further effect which indicates incompleteness. Dividing the plates in declination strips of 0.5 degree width, combining strips at equal distance from the plate center, we found for the total number of objects in each of the five

strips (beginning at the plate center) respectively 25, 26, 13, 5 and 12. This would seem to indicate that in the declination strips at more than 1° from the plate center there is serious incompleteness. Assuming that the central two strips are complete, a correction factor of 1.56 to the quasar density is indicated. No similar effect appears to be present in right ascension strips. The spectra of the quasars in the CS survey have been measured with a vidicon by Osmer and Smith (1980). From these spectra we have determined the B magnitudes of all objects.

The CTIO 4-meter telescope (4-m) survey (Hoag and Smith 1977) covers an area of 5.1 square degrees. Spectra of the objects have been measured by Osmer (1980), from which we have determined B magnitudes. Five bright quasars from the CS survey were used as plate centers in the 4-m survey; they are excluded from the following discussion. However, since only one of the five is in the redshift interval 1.8 to 2.4 on which the final area density will be based, this hardly affects the results.

In view of the fact that Ly α becomes difficult to observe below $z = 1.8$, while few quasars exist with $z > 2.4$, we shall restrict the sample to this interval. In Table 1 are given the counts of such quasars as a function of B for the CS and 4-m surveys and the corresponding surface densities (in the case of the CS survey corrected by the factor 1.56).

From the table it is seen that, although our corrections have diminished the discrepancies (Osmer 1980) between the CS and 4-m surveys, important differences remain. For example, the area density of the CS survey for $B < 18.5$ (the most pessimistic estimate of the magnitude where incompleteness effects begin to increase) is 0.12 ± 0.03, while the corresponding value for $B < 19.5$ in the 4-m survey equals 3.0 ± 0.7.

TABLE 1 - *Numbers and area densities for quasars with* $1.8 < z < 2.4$ *in the CS and 4-m surveys in intervals of B magnitudes.*

B	n_{CS}	n_{4-m}	N_{CS}	N_{4-m}
17 – 17.5	1		.01	
17.5 – 18	2		.02	
18 – 18.5	11	2	.09	0.4
18.5 – 19	20	5	.16	1.0
19 – 19.5	10	8	.08	1.6
19.5 – 20		13		2.5

Although with an increase of a factor of 7 per magnitude these numbers are perhaps not entirely outside the range of statistical possibilities, they suggest that the CS survey is still incomplete by a factor of 2 to 3. Among the reasons for this may be the difficulty of detecting Ly α close to the transmission limit of the CS which may be responsible for the low number of objects with $z = 1.8$ to 2.0, and more generally the difficulty of recognizing fainter lines at the low dispersions and small plate scales used in the Schmidt work. Other Schmidt surveys appear to suffer similar incompleteness, while also the magnitudes may be very uncertain (Véron and Véron 1982).

Incompleteness in the CTIO-CS survey is also suggested by a comparison of the equivalent widths of the lines used in the search with those in the 4-m survey (Osmer 1980). Of course, the question of the completeness of the 4-m survey in this respect also remains to be answered. Spectral measurements of a large unbiased sample of high redshift radio quasars might be useful in this context.

To obtain an estimate of the total quasar density from the CTIO surveys, we proceed as follows. In Figure 3 the redshift distribution of a sample of quasars between B = 18 and 20 obtained (from U-B excess) by Arp (1981) is compared with the redsbift distribution of all radio quasars in the catalogue of Hewitt and Burbidge (1980) in this range. Within the statistical limits, there is no evidence for a difference. Therefore we may, with some confidence, use the distribution of radio quasars (based on more than 300 objects) as the standard redshift distribution also for optical quasars in this range.

In Figure 3 we compare the standard redshift distribution with that for the CTIO objects. To obtain the total area density of quasars, we divide the values in Table 1 by the fraction of the radio objects in the $1.8 < z < 2.4$ interval. This yields from the CS survey $N(< 18.5) = 0.55 \pm 0.15$ and from the 4-m survey $N(< 19.5) = 13 \pm 3.5$ and $N(< 20) = 26 \pm 5$. While the latter values are in excellent agreement with the results from the reduced Braccesi field, the former again suggests that the CS survey is incomplete by a factor of about 3. Of course, the main advantage of the 4-m results in comparison with those of Braccesi *et al.* is that they are based on spectroscopically confirmed quasars with relatively well determined magnitudes.

(6) The deepest survey for quasar-like objects was made by Koo and Kron (1982). In an area of 0.293 square degrees they found 343 stars (and 2037 galaxies) down to about B = 23. On the basis of a three color

classification they find 10 quasars for $20 < B < 21$; 28 quasars for $21 < B < 22$ and 65 (?) for $22 < B < 23$, corresponding to respectively 34, 96 and 222 quasars per square degree. A further search, in which also proper motion criteria were used, was made in part of this field by Kron and Chiu (1981), wbo identified 8 quasars with $B < 21$ (two with $B < 20$) of which six have been spectroscopically confirmed. These identifications support the color criteria used by Koo and Kron.

The quasars found in the various surveys include true quasars ($M_V < -24$), Seyferts ($M_V > -24$) and possibly some BL Lac. The following information may be used to estimate the proportion of Seyferts at various magnitudes. Véron (1979) has studied a complete sample of nearby Seyfert galaxies from the Markarian surveys and determined the magnitudes of the nuclei. In this list there are 3 Seyfert 1 nuclei with $B < 15$ and 16 with $B<16$ in an area of 6000 square degrees. Most have M_V in the range -20 to -22, but a few are brighter. In the Braccesi sample to $B = 18.27$ there are 8 Seyferts out of 21 objects; 7 have M_V between -22 and -24. In the PHL, which covers about the same magnitude range, there are 12 Seyferts out of 40 quasar identifications in the catalog of Hewitt and Burbidge (1980); 10 have M_V between -22 and -24. Even though the PHL sample is subject to some biases (U-V color selection, exclusion of extended objects), this seems to confirm that about 30 to 40% of the objects down to $B = 18.3$ are Sy 1, the great majority of high luminosity. From a comparison of the Véron sample to $B = 16$ and the Braccesi and PHL samples to $B \approx 18$, we might be led to the conclusion that the N-B relation of the Sy 1 is as steep as that for quasars. In view of the relatively low redshifts ($z < 0.35$) for these objects such a strong evolution seems of course surprising.

To give some further information on the bright Seyferts we may make use of the recent HEAO 1 X-ray survey of Piccinotti *et al.* (1982). In 27000 square degrees they found 17 Sy 1 galaxies with an X-ray flux in excess of 3.1×10^{-11} ergs/cm^2/sec in the 2 to 10 kev band. From the average X-ray to optical luminosity ratio, this would correspond to a B magnitude of 14.5 and to a N (< 14.4) equal to 6×10^{-4} per square degree. However, because of the scatter in the ratio of X-ray to optical luminosity combined with the increasing numbers of faint Seyferts, a small correction has to be applied to obtain the effective B magnitude to which this number of Seyferts corresponds. From the expression given in the following paper and from the observed f_X/f_0 values given there, this correction should amount to about 0.2 to 0.3 for a log N—log S with a slope of 1.5 or somewhat larger.

TABLE 2 - *Cumulative numbers derived from various surveys of objects per square degree as a function of B magnitude.*

B	N	Survey
All quasars		
16	0.01	Green and Schmidt
18.27	0.95	Braccesi field
18.5	1.56	Sandage-Luyten field
18.5	(0.55)	CTIO Curtis Schmidt
19.5	13.5	CTIO 4-m
20	23	CTIO 4-m
21	57	Koo and Kron
22	153	Koo and Kron
23	375	Koo and Kron
Seyferts ($M_V < -24$) separately		
14.8	0.0006	(From X-rays) Piccinotti *et al.*
15	0.0005	(3 objects) Véron
16	0.0027	Véron
18.27	0.35	Braccesi field
BL Lac Objects		
15	0.0005	
19.3	< 0.14	

From these data we then conclude that there are 17 Sy 1 galaxies or 6×10^{-4} per square degree brighter than B = 14.8 in the sample of Piccinotti *et al.*

In Table 2 and Figure 4 we show the final N (< B) relation for all quasars and for the Seyferts only. Some of the points below B = 20 may still have to be increased somewhat owing to incompleteness. At fainter magnitudes the results are even more uncertain.

3 - THE NUMBER-MAGNITUDE RELATION FOR THE BL LAC OBJECTS

Few systematic searches for BL Lac objects have been made, and as a consequence only fragmentary data are available. We shall make use of these to obtain an estimate of the N (B) relation for radio emitting BL Lac objects. Virtually all known BL Lac objects are radio sources, although it still is possible that a class of radio quiet BL Lac remains to be discovered. While the results are highly preliminary, they seem to indicate that the

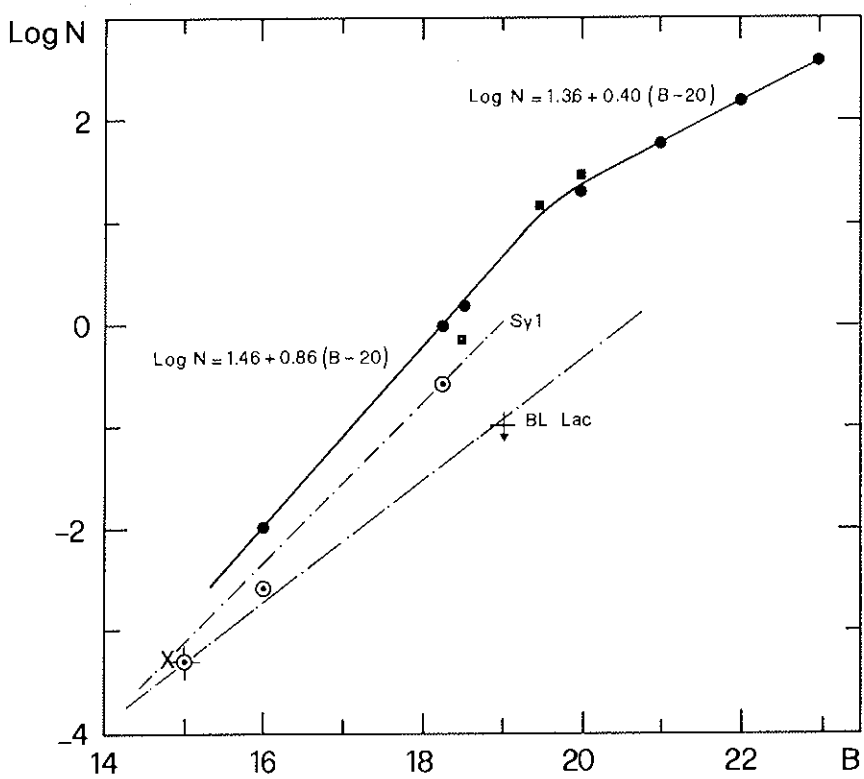


FIG. 4. The integral counts per square degree for quasars of all absolute magnitudes (filled circles from color excess surveys, filled squares from CTIO objective prism surveys), for Seyfert 1 galaxies (open circles) and for BL Lac objects (+). The x is a point derived from X-ray observations of Sy 1 galaxies.

relation is much less steep for these objects than for quasars. A particular problem for the BL Lac objects is also caused by the large light variations.

For most of the brighter BL Lac objects light curves are available which cover with interruptions several years or decades. Taking from these the B_{max} and B_{min} and assuming that they spend equal time in equal magnitude intervals, we may determine the average number of the presently known sample that are brighter than a given magnitude at any particular time. We find 10.5 such objects with $B < 15$. From the fact that these are predominantly northern hemisphere high galactic latitude objects, we infer that an incompleteness correction of at least a factor of two is to be applied which yields for the BL Lac objects an estimate of $N(< 15) =$

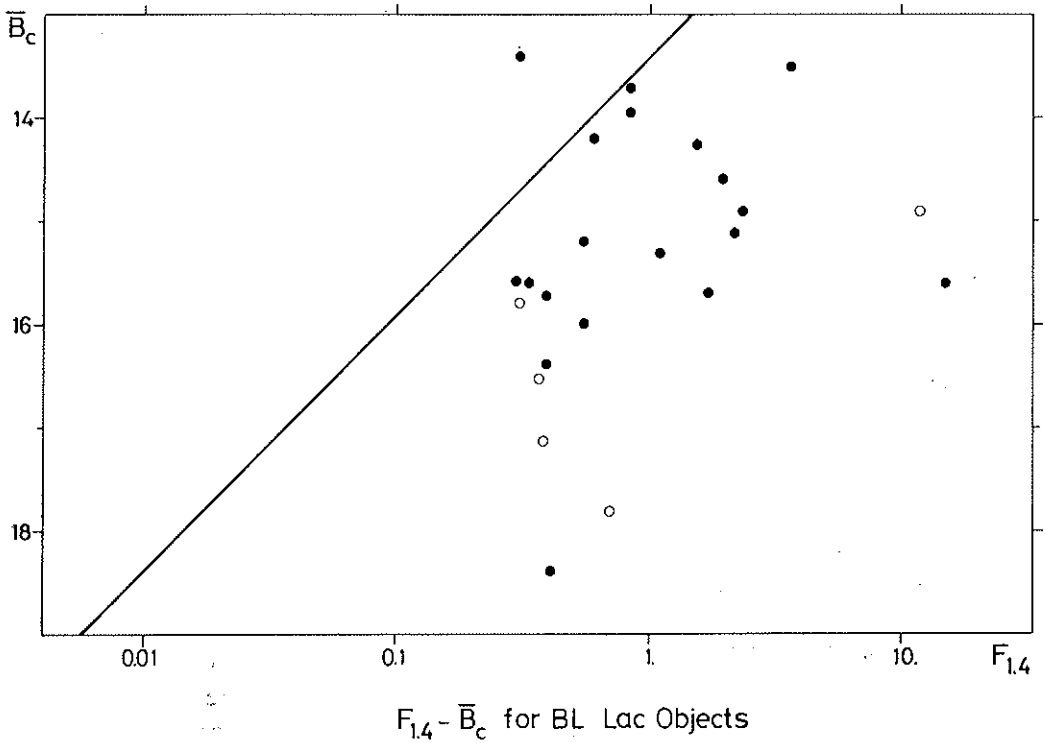


FIG. 5. The relation between the flux (in Jy) at 1.4 GHz and the mean B magnitude for BL Lac objects.

0.5×10^{-3} per square degree. The sometimes spiked character of the light curves may make this figure slightly on the high side, but incompleteness beyond what has been corrected for may make it too low.

In Figure 5 we plot for the known BL Lac objects the $\bar{B}_c = 1/2$ ($B_{max} + B_{min}$) against the radio flux at 1.4 GHz, $F_{1.4}$ in Jy. If these same objects were observed at larger distances, they would move down in this diagram along a 45° line. In Figure 5 we have drawn a line such that all but one of the brighter BL Lac are situated to the right. If then at fainter optical magnitudes we survey the appropriate objects with sufficient radio sensitivity to reach the points on this line, then all the BL Lac objects should be detected as radio sources.

From the sample of 175 objects with $u\text{-}b < -0.35$ in the Braccesi field, 99 have been observed at Westerbork (Katgert *et al.* 1973). Deleting

the objects with B > 19 and the objects for which the uncertainty in the radio flux exceeds 15 mJy at 1.4 GHz, we are left with 49 objects among which there are 25 spectroscopically confirmed quasars and 18 objects wbich in Figure 5 would be situated to the left of the 45° line. Of the remaining objects, two have been detected as radio sources to the right of the line and four have upper limits to the right of the line. The detected sources are AB 82 with $F_{1.4} = 7.2 \pm 4.8$, which is therefore an uncertain case, and AB 114. For the former $u\text{-}b = -0.88$ and $b\text{-}v = 0.12$, both more than 0.1 bluer than the bluest BL Lac objects, and therefore it is almost certainly a quasar. The same is the case for AB 114 with a $u\text{-}b$ color of -0.97. For the four sources with radio upper limits only, AB 50 with $b\text{-}v = 0.11$, AB 58 with $b\text{-}v = 0.03$ and AB 60 with $u\text{-}b = 0.80$, $b\text{-}v = -0.19$ are all bluer than expected for BL Lac objects, while for AB 88 the color situation is unclear. It seems therefore that there is at most one BL Lac object in the sample of the same type as the ones typically seen at brighter magnitudes. Since there are 102 objects with B < 19 in the whole field and taking into account some incompleteness, we conclude that there should be no more than 4 BL Lac in the field, corresponding to an upper limit of 0.1 per square degree. Inspecting Figure 4 it appears that such an upper limit is just compatible with an increase of a factor of four per magnitude. It would be necessary to assume that we have made an error of a factor of 15 in the ratio of the numbers of BL Lac at B = 19 and at B = 15 to obtain the same slope as for the quasars. Apparently the BL Lac have no or only a very weak evolution.

4 - Clustering of quasars

There are about 10^6 quasars down to B = 20. This may be compared with the order of 10^8 for giant galaxies of which there are 10^7 gE, 10^7 clusters and 10^6 superclusters in the universe with $z<3$. As first pointed out by us (Setti and Woltjer 1977), a comparison of the numbers indicates that if quasars are nuclei of galaxies, and in particular if gE, clustering should set in around B = 20.

Oort, Arp and de Ruiter (1981) have studied the evidence for superclustering in available samples on the basis of the distribution on the sky and in redshift. They present lists of pairs of objects which they believe on statistical grounds to be probably or possibly physically related. In their lists there are 14 objects with separations in the interval 10 to 50 Mpc (scaled to

$H_0 = 50$) and 5 with smaller separations, all of which are less than 6.5 Mpc for $q_0 = 0.05$. If we take $q_0 = 0.5$, however, there are 16 separated by 10 to 50 Mpc and 7 with smaller separations, all in fact less than 8 Mpc. Although probably only some of the pairs are real, and although selection effects must be important, it seems tempting nevertheless to suggest tbat the relatively more numerous small separations represent clusters of galaxies with more than one quasar. These close pairs as well as two triples are listed in Table 3. It is seen that also the velocity differences seem to be consistent with what would he expected for clusters or groups of galaxies. Since the redshifts of Osmer (1980) are uncertain by ± 0.01 and those of Vaucher and Weedman (1980) by 1 to 2%, the precise values of most of the velocity differences are not particularly meaningful. In fact, the redshift in the VaucherWeedman pair are based on two lines which, if correctly identified, have measured Δz values of 0.02 (O VI) and 0.11 (Ly α) respectively.

TABLE 3 - *Pairs and triples of quasars which may be located in clusters or groups of galaxies. Linear separations in Mpc (for triples the diameter of the smallest circle containing the triple) have been calculated for $H_0 = 50$ km/sec/Mpc, $q_0 = 0.5$. The magnitudes (sometimes determined from spectral scans) refer to the B band, except for the cases indicated by an asterisk which are closer to V. The velocity differences are given in the frame of the object; for the triples the largest difference is given. The pairs have been taken from the lists of Oort, Arp and de Ruiter (1981), while the data for the "M 82 triple" are from Burbidge* et al. *(1980).*

	mag	Separation	z	Δv (km/s)
Pairs				
Oort *et al.* 5	19.5,18.5*	1.5	1.010,1.011	150
Osmer 46-48	18.9,18.7	2.9	2.03,2.04	(1000)
Vauch/Weed 36-35	19.6,18.9	3.5	2.88,2.82	(4700)
Osmer 14-13	19.7,19.5	7.1	1.92,1.91	(1000)
Oort *et al.* 3	19.7,18.7*	8.0	1.131,1.134	400
Triples				
"M 82"	20.5,20.0,20.5*	1,8	2.048,2.054,2.040	1400
Osmer 54-55-56	18.2,19.7,18.5	4.4	1.83,1.85,1.86	(3200)

While the data are certainly inadequate to establish fully the clustering of quasars, these results seem sufficiently promising to warrant more extensive searches. If the present ideas are correct, at about B = 21 clustering effects should already be more pronounced. Clustering studies should shed light on the evolution of clusters and possibly superclusters as well as on the nature of the quasar phenomenon. Studies of absorption lines in the clustered objects would also be of much interest. It has been suggested on various occasions (Woltjer 1974, Williams and Weymann 1976, Weymann *et al.* 1977) that the absorption systems with redshifts in excess of the emission line redshift could arise in gas clouds or galaxies in the same cluster as a quasar. More extensive observations of absorption lines in clustered quasars could test these ideas.

Direct evidence for the association of single quasars with clusters or groups of galaxies has been looked for in low redshift quasars, but the data are still fragmentary. Of the 57 quasars in the catalog of Hewitt and Burbidge, with redshifts between 0.10 and 0.35 and B below 17.5, there are to our knowledge six associated with clusters or groups in which at least one galaxy has been shown to have about the same redshift as the quasar. However, only a small and unknown fraction of these quasars has heen looked at in detail, and the true fraction of quasars associated with clusters or groups is certainly larger.

5 - The nature of the underlying galaxies

In the case of the Seyferts there is no doubt that the active nucleus is located in the middle of a galaxy, which in several cases appears to be a spiral. In the case of the true quasars the situation is less clear, probably because it is very difficult to see a galaxy around a very luminous nucleus. In fact, Fairall 9 may well be an illustrative transition case. If the nucleus were one or two magnitudes brighter and a factor of two more distant, it would become difficult to see the galaxy.

Some time ago Kristian (1973) made detailed estimates of the conditions under which photographs would reveal an underlying galaxy if quasars were associated with giant ellipticals and showed that the available data were consistent with the expectations. More recently, various authors have tried to observe the "fuzz" around quasars and BL Lac directly and to obtain some information about its nature. The observations are difficult and, as a consequence, the conclusions are still open to some doubt.

Recently several studies of the morphology of quasars have been made.

Twelve quasars with redshifts below 0.35 were observed by Hutchings *et al.* (1981), who find at least ten to be extended. Seven of these could be underlying galaxies but in the others more asymmetrical structures are found. The scale of these structures is in the range of 20 to 80 kpc. Wyckoff *et al.* (1981) have measured photographs of 14 quasars with redshifts below 0.41 and find evidence for underlying structures in all but two cases with sizes of 90 kpc and up. Hawkins (1978) observed seven quasars on electronographic emulsions and found five to be extended. Four of these objects were also observed by Wyckoff *et al.*; unfortunately, the most extended object found by Hawkins is the narrowest of the objects found by Wyckoff *et al.*! The lower limits for the total R magnitude of 24 (under a quasar 7 magnitudes brighter) found by the latter for this case suggests a considerable overinterpretation of the data. Also for the case of 3C 206 the results found by Wyckoff *et al.* (1981) are rather different from the results of Wyckoff *et al.* (1980). Perhaps some of the difficulties relate to the very large range (up to 11 magnitudes) over which photographic photometry was done by Wyckoff *et al.* (1981). In addition, as the authors note, their calibration procedure which relies on the magnitudes of the quasars could lead to errors because of variability. For 11 of their quasars Wyckoff *et al.* (1981) find that the underlying galaxies have M_R magnitudes in the range -23.6 to -20.8 with an average of -22.2 which presumably would correspond to M_V about -21.5. In all these investigations, one may also wonder whether the corrections used could not introduce color dependent effects and thereby systematic differences between quasars and stars.

A variety of other measurements are given in Table 4. Most of these are based on spectroscopic scans and/or photometry with different sized diaphragms. The Wyckoff *et al.* result on 3C 206 is based also on photographic photometry, the result of Vanderriest aud Schneider for Ton 256 on electronography. The latter authors also give data for a number of bright Seyferts. The results of Weistrop *et al.* for 0548-32 and 1 Zw 1727+50 are based on CCD photometry.

It has been found by Sandage (1971) that for the brightest cluster ellipticals on the average $M_V = -23.3$. For both quasars and BL Lac objects the parent galaxies therefore seem to have values that are typically a magnitude fainter.

The nature of the underlying galaxies remains very uncertain. Hutchings *et al.* (1981) emphasize that, with the available resolution, ellipticals and spirals are indistinguishable. From the colors or spectra it has been argued in the case of some quasars and of most of the BL Lac objects that

TABLE 4 - *Inferred Visual Absolute Magnitudes of underlying galaxies in quasars and BL Lac objects.*

	M_V	Reference
Quasars		
3C 206	− 22.8	Wyckoff *et al.* 1980
Ton 256	− 23.3	Silk *et al.* 1973
	− 23.2	Vanderriest and Schneider 1979
BL Lac Objects		
0548-32	− 22.1	Weistrop *et al.* 1979
	(− 22.3)	Fosbury and Disney 1976
MKN 421	− 23.1	Kinman 1978
AP Lib	− 21.8	McGimsey and Miller 1978
	− 21.4	Disney *et al.* 1974
MKN 501	− 21.5	McGimsey and Miller 1978
1 Zw 1727+50	− 21.9	Oke 1978
	− 21.9	Weistrop *et al.* 1981
	− 21.7	Kinman 1978
3C 371	− 22.7	Oke 1978
BL Lac	− 23.0	Oke 1978
	− 22.1	Miller *et al.* 1978

the galaxies are in fact ellipticals. In view of the ambiguity of the spectral data for PHL 1070 and 3C 206 this conclusion remains in doubt for quasars.

The Sy 1 galaxies are generally taken to be mainly spirals (Adams 1977, see also the data by Vanderriest and Schneider 1979). Broad line radio galaxies are generally associated with gE. It was found by Osterbrock (1977) and Phillips (1977) that Sy 1 galaxies always have Fe II emission at 4750 Å, while broad line radio galaxies do not. It was subsequently noted (Setti and Woltjer 1977, see also Miley and Miller 1979) that the Fe II emission is absent also in steep spectrum quasi-stellar radio sources, while in flat spectrum objects it is sometimes present and sometimes not. For the QSO without radio emission the situation is less clear. Peterson *et al.* (1981) find 9 objects all with Fe II emission, but more than half of these we would have classified as Sy 1. Unfortunately, the precise mechanism of the Fe II emission is still uncertain.

A possible interpretation of these data would be that the steep spectrum radio quasars are associated with gE and the objects without radio emission more with spirals. The flat spectrum radio quasars would

be a mixture. However, more data are needed before these considerations can be taken very seriously. It would be particularly valuable to have Fe II data for the same objects for which data about the color and spectrum of the underlying galaxy are obtained. In this respect the Sy 1 object PHL 1070 is instructive. It has Fe II emission, but the spectral data for the underlying object were fitted by Morton *et al.* 1978 (with some ambiguity) to the spectrum of an elliptical galaxy.

6 - Concluding remarks

The adopted curve for the N (B) relation shows strong evolutionary effects to about B = 20 and flattens off soon thereafter. The latter behavior was inferred previously on the basis of the data on the X-ray background (Setti and Woltjer 1979). If the curve were to continue upward, the quasars would collectively produce a background much stronger than observed. As we pointed out previously (Setti and Woltjer 1973b, 1979), the X-ray background argument can also be used to exclude a "local" nature for most quasars. It is nevertheless of importance to see to what extent the observations really exclude a Euclidean relation Log N ($<$ B) = const. + 0.6 B. In fact, in an earlier discussion (Setti and Woltjer 1973a), we showed that such a relation was consistent with all available data at that time. Even at present such a relation with N ($<$ 20) = 16 fits all data points up to B = 22 to better than a factor of two, except for the Green and Schmidt point at B = 16, where the deviation is a factor of about six. Could such an incompleteness exist in their survey?

It was pointed out by us previously (Setti and Woltjer 1973b) that probably the most effective way of discovering bright quasars is through their X-ray emission. In fact, we now may use the X-ray survey of Piccinotti *et al.* (1982) for this purpose. With the f_X/f_0 ratio of about 0.18 found in the following paper, the effective limit of this survey would correspond to B=12.8. With the Euclidean slope of a factor of four per magnitude there should be 8×10^{-4} quasars per square degree or 22 in the X-ray survey. Since only one has been found this seems to provide further evidence that the slope of the N (B) relation at bright magnitudes is really steep.

It is interesting to see that the Seyferts seem to share with the quasars the steep slope in the Log N-B diagram. This is somewhat reminiscent of the behavior of the radio galaxies where, already at relatively low redshifts, evolutionary effects are of importance (see the article by Van der Laan in these proceedings). But why do the BL Lac show such a different behavior?

REFERENCES

Adams, T.F., 1977, *Astrophys. J. Suppl.*, **33**, 19.

Arp, H., 1981, *Astrophys. J.*, **250**, 31.

Berger, J. and Fringant, A.M., 1977, *Astron. Astrophys. Suppl.*, **28**, 123.

Braccesi, A., Formiggini, L. and Gandolfi, E., 1970, *Astron. Astrophys.*, **5**, 264; (erratum *Astron. Astrophys.*, **23**, 159).

Braccesi, A., Zitelli, V., Bonoli, F. and Formiggini, L., 1980, *Astron. Astrophys.*, **85**, 80.

Burbidge, E.M., Junkkarinen, V.T., Koski, A.T., Smith, H.E. and Hoag, A.A., 1980, *Astrophys. J.*, **242**, L55.

Disney, M.J., Peterson, B.A. and Rodgers, A.W., 1974, *Astrophys. J.*, **194**, L79.

Formiggini, L., Zitelli, V., Bonoli, F. and Braccesi, A., 1980, *Astrophys. J. Suppl.*, **39**, 129.

Fosbury, R.A. and Disney, M.J., 1976, *Astrophys. J.*, **207**, L75.

Hawkins, M.R., 1978, *Mon. Not. Roy. Astron. Soc.*, **182**, 361.

Hewitt, A. and Burbidge, G.R., 1980, *Astrophys. J. Suppl.*, **43**, 57.

Hoag, A.A. and Smith, M.G., 1977, *Astrophys. J.*, **217**, 362.

Hutchings, J.B., Crampton, D., Campbell, B. and Pritchet, C., 1981, *Astrophys. J.*, **247**, 743.

Katgert, P., Katgert, J.K., Le Poole, R.S. and Van der Laan, H., 1973, *Astron. Astrophys.*, **23**, 171.

Kinman, T.D., 1978, *Pittsburgh Conf. on BL Lac Objects*, ed. A.M. Wolfe, p. 82.

Koo, D.C. and Kron, R.G., 1982, *Astron. Astrophys.*, **105**, 107.

Kristian, J., 1973, *Astrophys. J.*, **179**, L61.

Kron, R.G. and Chiu, L.-T.G., 1981, *Pub. Astron. Soc. Pacific*, **93**, 397.

McGimsey, B.Q. and Miller, H.R., 1978, *Astrophys. J.*, **219**, 387.

Miley, G.K. and Miller, J.S., 1979, *Astrophys. J.*, **228**, L55.

Miller, J.S., French, H.B. and Hawley, S.A., 1978, *Astrophys. J.*, **219**, L85.

Morton, D.C., Williams, T.B. and Green, R.F., 1978, *Astrophys. J.*, **219**, 381.

Oke, J.B., 1978, *Astrophys. J.*, **219**, L97.

Oort, J.H., Arp. H. and de Ruiter, H., 1981, *Astron. Astrophys.*, **95**, 7.

Osmer, P.S., 1980, *Astrophys. J. Suppl.*, **42**, 523.
— 1981, *Astrophys. J.*, **247**, 762.

Osmer, P.S. and Smith, M.G., 1980, *Astrophys. J. Suppl.*, **42**, 333.

Osterbrock, D.E., 1977, *Astrophys. J.*, **215**, 733.

Peterson, B.M., Foltz, C.B. and Byard, P.L., 1981, *Astrophys. J.*, **251**, 4.

Phillips, M.M., 1977, *Astrophys. J.*, **215**, 746.

Piccinotti, G., Mushotzky, R.F., Boldt, E.A., Holt, S.S., Marshall, F.E. Serlemitsos, P.J. and Shafer, R.A., 1982, *Astrophys. J.*, **253**, 485.

Sandage, A., 1971, *Proceedings Study Week on Nuclei of Galaxies*, Pont. Acad. Scient., Scripta Varia, **35**, 271.

Schmidt, M., 1970, *Astrophys. J.*, **162**, 371.

Setti G. and Woltjer, L., 1973a, *Sixth Texas Symposium on Relativistic Astrophysics*, Ann. New York Acad. Sci., **224**, 8.

— 1973b, *I.A.U. Symposium 55*, 208.

— 1977, *Astrophys. J.*, **218**, L33.

— 1979, *Astron. Astrophys.*, **76**, L1.

Silk, J., Smith, H.E., Spinrad, H. and Field, G.B., 1973, *Astrophys. J.*, **181**, L25.

Steppe, H., 1978, *Astron. Astrophys. Suppl.*, **31**, 209.

Steppe, H., Véron, P. and Véron, M.P., 1979, *Astron. Astrophys.*, **78**, 125.

Stockman, H.S. and Angel, J.R., 1978, *Astrophys. J.*, **220**, L67.

Van den Bergh, S., 1979, *I.A.U. Colloq. 54*, Scientific Research with the Space Telescope, ed. M.S. Longair and J.W. Warner, p. 151.

Vanderriest, Ch. and Schneider, J., 1979, *Astron. Astrophys.*, **76**, 297.

Vaucher, B.G. and Weedman, D.W., 1980, *Astrophys. J.*, **240**, 10.

Véron, P., 1979, *Astron. Astrophys.*, **78**, 46.

Véron, P. and Véron, M.P., 1982, *Astron. Astrophys.*, **105**, 405.

Weistrop, D., Smith, B.A. and Reitsema, H.J., 1979, *Astrophys. J.*, **233**, 504.

Weistrop, D., Shaffer, D.B., Mushotzky, R.F., Reitsema, H.J. and Smith, B.A., 1981, *Astrophys. J.*, **249**, 3.

Weyman, R.J., Williams, R.E., Beaver, E. and Miller, J., 1977, *Astrophys. J.*, **213**, 619.

Williams, R.E. and Weyman, R.J., 1976, *Astrophys. J.*, **207**, L143.

Woltjer, L., 1974, *Sixteenth Solvay Conference Univ. Bruxelles*, 429.

Wyckoff, S., Wehinger, P.A., Spinrad, H. and Boksenberg, A., 1980, *Astrophys. J.*, **240**, 25.

Wyckoff, S., Wehinger, P.A. and Gehren, T., 1981, *Astrophys. J.*, **247**, 750.

DISCUSSION

OORT

You mentioned my suggestion that the Lα lines would be due to superclusters. I have just received a letter and preprint from Sargent in which results are given of the Lα absorption lines in two quasars at $z \sim 2.5$ separated by 3 arcmin, or a little over 1 Mpc. The two sets of lines show no correlation and thus disprove the hypothesis that they may have been formed in a supercluster.

Sargent also sent me a preprint (Young, Sargent and Boksenberg) containing new observations of Lα absorption which indicate the presence of evolution in the number of Lα lines at different redshift. This may become interesting when the range of z can be extended to low values, which may be possible with the Space Telescope.

I believe that the lack of clustering in the Lα lines might be due to the fact that the extended and very thin objects which produce these lines cannot exist in the dense environment of clusters.

SCIAMA

I would like to report very briefly the provisional detection of ionised intergalactic helium by a positive Gunn-Peterson effect in absorption troughs in the spectrum of the quasar Q 2204-408 which has redshift 3.18. My colleagues and I (Gondhalekar *et al.* 1982, in press) have observed the far ultraviolet spectrum of this quasar with IUE and find that we can observe about 100 Å of the continuum beyond the redshifted He II line at 304 Å in the rest frame of the quasar. There is evidence that the continuum is depressed in this region and, if correct, this is the first detection of such an absorption trough, since there is no such trough on the short wavelength sides of atomic hydrogen and helium. The implications of this result are: (a) the quasar is cosmological; (b) there is a general intergalactic medium which has not been previously detected; (c) the gas contains helium; (d) the temperature of the gas must lie in the range $10^4 < T < 10^5$ K. The optical depth of the gas is uncertain but, if scaled to the present epoch, the helium abundance must be greater than $n(z = 0, \mathrm{He}) = 10^{-11}$ cm^{-3}.

THE ORIGIN OF THE X-RAY AND γ-RAY BACKGROUNDS

G. SETTI
Istituto di Radioastronomia, Università di Bologna, CNR
and
L. WOLTJER
European Southern Observatory
Garsching bei München

1 - Introduction

The origin of the isotropic components of the X- and γ-ray backgrounds is thought to be extra-galactic, probably due to events that took place at redshifts $z \gtrsim 1$. Tbe spectrum of the diffuse cosmic radiation for energies $\gtrsim 2$ keV is shown in Fig. 1 for a selected sample of observational results. In the energy interval $3 \div 50$ keV the results obtained with the experiment A-2 on board HEAO-1 are remarkably well fitted by a 40 keV thermal bremsstrahlung with an absolute accuracy of $\sim 10\%$ (Marshall *et. al.* 1980). This is consistent with a power law fit to previous data of the form $7.7\ E_{keV}^{-1.4}$ ph/cm^2 sec ster keV in the energy domain $2 \div 20$ keV and with the existence of a break in the spectrum between 10 and 30 keV (Schwartz 1979).

At higher energies ($20 \div 200$ keV), measurements made with a balloon borne experiment are best fitted with a power law photon spectrum of the form $67\ E_{keV}^{-2.17}$ ph/cm^2 sec ster keV, which smoothly joins with the 40 keV thermal bremsstrahlung spectrum discussed above. The same data may be best fitted by a thermal bremsstrahlung spectrum with a temperature of 97 keV (Kinzer *et al.* 1978). This indicates that the diffuse cosmic X-ray radiation between 2 and 200 keV cannot be obtained by a single isothermal

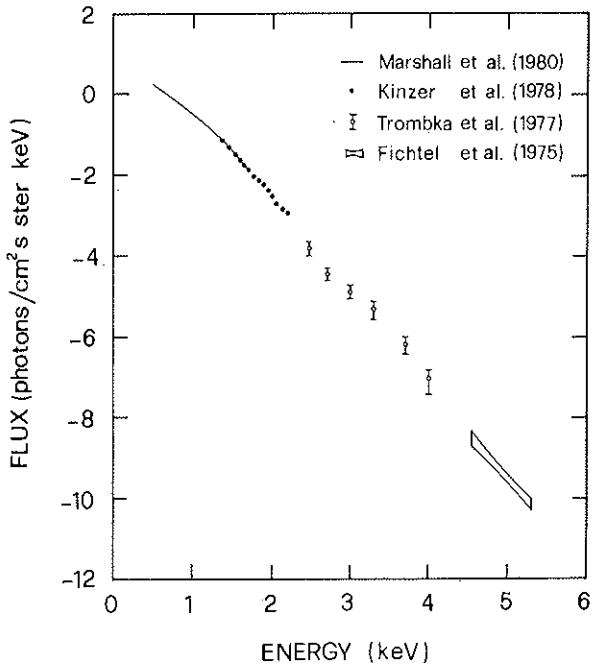


FIG. 1. The observed X- and γ-ray backgrounds based on four representative experiments. The logarithms of the quantities are plotted.

bremsstrahlung spectrum. More recently Matteson *et al.* (1979) have reported the results obtained with the A-4 experiment on board HEAO-1 which indicate a steeper spectrum (3.1×10^{-3} (E/100 keV)$^{-2.9}$ photons/cm^2 sec ster keV) in the energy interval 100 ÷ 400 keV. This is well above the extrapolation of the 40 keV thermal bremsstrahlung fit to the 3 to 50 keV diffuse background. It is well known that above ~ 400 keV numerous effects may contribute unwanted backgrounds which seriously affect the measurements and are difficult to subtract. The reanalysis of the Apollo 15 data in the 0.3 to 10 MeV energy interval (Trombka *et al.* 1977), together with the results from other experiments up to 20 MeV, indicate the presence of a cosmic background radiation at gamma-ray energies which smoothly connects the hard X-ray portion of the cosmic background radiation spectrum with the hard γ-ray region as measured by SAS-2 above 35 MeV (Fichtel *et al.* 1975). Although it has been stated that the available measurements tend to converge and to support the much debated existence of a bump in the spectrum at a few MeV (Schönfelder

et al. 1980), it appears to us that within the errors the photon spectrum above ~ 30 keV can still be fitted by a single power law ($dn/dE \simeq 173\ E_{keV}^{-2.36}$ photons/cm^2 sec ster keV) up to 200 MeV.

The large scale distribution of the diffuse X-ray radiation has been studied by UHURU and by tbe SSI experiment on board Ariel V in the 2 ÷ 10.5 keV and 2 ÷ 18 keV energy ranges respectively (Schwartz 1980, Warwick *et al.* 1980). The most prominent characteristic is a galactic latitude dependence of the intensity which clearly shows the existence of a galactic contribution extending to very high galactic latitudes. The intensity in the direction of the north galactic pole is ~ 2% higher than the intensity in the direction of the south galactic pole, and it increases by several percentage points going toward lower galactic latitudes. The presence of a peak in the surface brightness for $|b| < 20^0$ is clearly shown. More recent results obtained by the A2 experiment on HEAO-1 indicate that the galactic contribution may be separated into two components: one component which is responsible for the high galactic latitude features mentioned above and a disc component probably due to the contribution of a population of unresolved galactic sources distributed in a disc with half-thickness of ~ 250 pc. The total emission of the disc component in the 2 ÷ 10 keV energy interval is only a few percentage points of that of the Galaxy due to resolved sources and it possesses a spectrum which is softer than the spectrum of the diffuse background (Iwan *et al.* 1981, Worrall *et al.* 1981).

Any residual large-scale anisotropy of the diffuse X-ray background (XRB hereafter) appears to be less than 1% (Warwick *et al.* 1980), thus providing a very strong argument in favor of its extra-galactic nature. However, a full exploitation of the XRB as a probe to study the large-scale distribution of matter must necessarily wait for a complete understanding of the galactic effects (see Fabian 1980 for a recent discussion).

The small scale fluctuations of the XRB have been extensively studied by UHURU observers in the 2 ÷ 6 keV energy band. The general conclusion is that the observed fluctuation pattern in sky cells of $10^\circ \times 10^\circ$ may be completely accounted for in terms of the source counts down to the UHURU limit. However, a residual fluctuation level of 1.3% may be allowed and this means that any strongly evolving population of X-ray sources must have a surface density $\gtrsim 6 \times 10^5$ objects ster^{-1} (Schwartz 1980).

Concerning the isotropy of the background radiation at higher energies, one can mention the results obtained by OSO III in the 20 ÷ 200 keV

energy interval which show that there are no large scale fluctuations larger than $\sim 5\%$ (Schwartz 1970).

Concerning the isotropy of the diffuse background radiation at γ-ray energies, very little is known of the large scale isotropy at a few MeV, while for energies > 30 MeV the so-called extra-galactic component appears to be isotropic within thirty percent (30%) or so (Fichtel *et al.* 1978). In this last case it is difficult to separate out exactly the galactic contribution.

The energy density in the XRB radiation flux between 2 keV and a few hundred keV is $\simeq 6 \times 10^{-5}$ eV/cm^3. In itself, this amount of energy is not particularly large as compared to present energy densities found associated with other regions of the extra-galactic background radiation field, but it was immediately realized that the real problem was to understand how so much energy had been channelled into high energy photons. It was also quickly recognized that the interpretation of the XRB is likely to involve important cosmological evolutionary effects (see, e.g., Setti and Rees 1970 and Silk 1973 for earlier reviews of the subject). It should be noted that if the XRB is mostly contributed from sources at large redshifts ($z \simeq z^*$) the power output in X-rays must have been larger by a factor $(1 + z^*)^n$, $n > 2$, to take into account the redshift effects and the shorter expansion time scale (the precise value of n will of course, depend on the adopted model universe). On the other hand, the observational evidence accumulated in the past several decades shows that violent events in galaxies were much more frequent in the past than now, so that it becomes relatively easier to relegate the origin of the XRB to earlier epochs.

The models which have been proposed to explain the bulk of the XRB may be broadly divided into two classes: 1) models which involve diffuse processes taking place in the intergalactic space; 2) discrete source models. In the first class the main processes which have been discussed are the thermal bremsstrahlung emission from a hot diffuse intergalactic gas and the radiation due to Compton scattering of the universal black body radiation by ultra-relativistic electrons around radio sources at large redshifts. The thermal bremsstrahlung model has been recently rediscussed by Field and Perrenod (1977). They concluded that the XRB between 2 and 100 keV is consistent with tbe presence of a hot intergalactic gas, witb a density close to tbe closure density of the universe, which has been heated up at large redshifts by exploding events taking place in active galaxies and then has cooled down adiabatically to the present temperature $T_0 \simeq 4.4 \times 10^8$ K. Because of the relatively inefficient way by wbich energy is transformed into radiation in the thermal bremsstrahlung process, the

energy requirements of this type of model for the sources which must provide the heat input become particularly severe.

In the second class of models one has postulated that the XRB is mainly due to the integrated contribution from discrete sources. From the results of the deep survey done with the Einstein Observatory, one finds that about 25% of the XRB at 2 keV may be accounted for in terms of observed sources (Giacconi *et al.* 1979). Combining the deep survey results with those obtained in previous X-ray surveys, it has also been found that the integral source counts are consistent with the Euclidean slope of -1.5. Setti and Woltjer (1970, 1973) have suggested that quasars, and possibly Seyfert galaxies if subject to cosmological evolution, may provide the main contribution. Indeed, by combining the X-ray emission from a few hright quasars with the optical source counts, Setti and Woltjer (1979) have further shown that the optical counts of quasars must flatten off at a magnitude $m_B \simeq 20$ to 21 to avoid an excessive contribution to the XRB (the so-called X-ray catastrophy). This has been confirmed by the studies of a large sample of quasars done with the Einstein Observatory (Tananbaum *et al.* 1978; Zamorani *et al.* 1981).

Since it appears that discrete sources may provide the bulk of the XRB at a few keV, in the next section we shall concentrate on the estimates of the contribution from various classes of sources leaving aside any further discussion of models which invoke diffuse processes taking place in the intergalactic space.

2 - X-ray Emission Properties of Different Classes of Objects and Their Contribution to the XRB

In what follows we find it convenient to introduce an X-ray magnitude which is defined as:

$$m_X = -2.5 \log F_X(2\,\mathrm{keV}) + 8.38 \qquad (1)$$

where F_X (2 keV) is the observed flux at 2 keV in units of keV/cm^2 sec keV corrected for galactic absorption. This definition is such that for $m_X = m_B$ the X-ray energy flux in a band of several keV is approximately equal to the energy flux in the photometric B band. We will also make use of the X-ray to optical ratio obviously defined as: $\mathrm{Log}\,(f_X/f_o) = -(m_X - m_B)/2.5$. All B magnitudes are corrected for galactic absorption assuming the law $\Delta m_B = -0.24\,(\mathrm{cosec}\, b - 1)$, where b is the galactic

latitude. A Hubble constant, $H_0 = 50$ km/sec/Mpc, and a Friedmann model universe with a density parameter, $\Omega = 0$, have been adopted.

2.1 *Quasars*

The basic set of data is provided by the two samples of Ku *et al.* (1980) and Zamorani *et al.* (1981) which contain respectively 111 and 107 known quasars observed with the Einstein Observatory. We have re-analyzed these data following a procedure somewhat different from those adopted by these authors. We have, of course, obtained a number of results qualitatively similar to those obtained in the previous analysis, but here we shall concentrate mainly on those that are more relevant to the purpose of the present discussion.

Optical data and redshifts have been taken from the catalog of Hewitt and Burbidge (1980) and from the references quoted in the two papers concerning the basic samples. We have omitted the quasars found around the cluster A 2151 for which information on the optical data was not available to us; the quasars UMT 301, 1729+501, and KP 33, for which optical magnitudes and/or redshifts are very uncertain; RS 23 which has apparently been misidentified (Zamorani, private communication); GQ Comae and V 396 Her because they have been discovered out of a search amongst catalogued variable stars (Bond *et al.* 1977) and, therefore, they seem to be rather heterogeneous with respect to the other optically discovered quasars.

Absolute magnitudes have been obtained assuming power law energy spectra of the form $\nu^{-\alpha}$ with $\alpha = 1$, so that the K-correction term may be taken to be identical to zero. Although there is some evidence that the average optical continuum of quasars may be flatter (Richstone and Schmidt 1980), this choice does not affect in any relevant way the conclusions which will he drawn in what follows.

From the point of view of the XRB estimates, the ideal situation would be the one in which one had sufficient information on the X-ray emission down to the faintest optical source counts. That this is not the case is clearly shown in Fig. 2 where we have plotted the observed sample of radio-quiet quasars per interval of apparent B magnitude as a function of the absolute magnitude of the objects. No objects have been detected beyond $m_B = 18.5$. Also, looking at the distributions in the two intervals centered at $m_B = 17$ and $m_B = 18$, one would get the impression that the detection rate decreases much faster than expected with increasing apparent magnitude. However, this conclusion does not appear to he correct, at variance

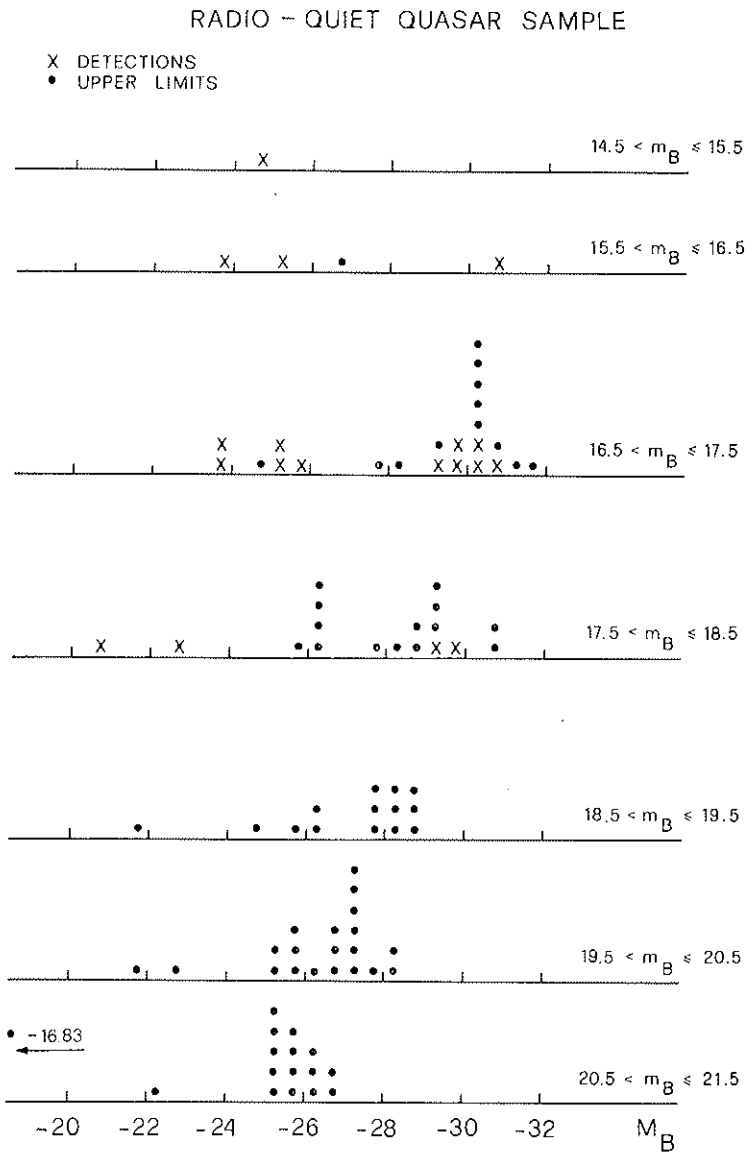


FIG. 2. The combined sample of radio quiet quasars observed in X-rays from Ku *et al.* (1980) and Zamorani *et al.* (1981) subdivided in intervals of apparent and absolute optical B magnitudes.

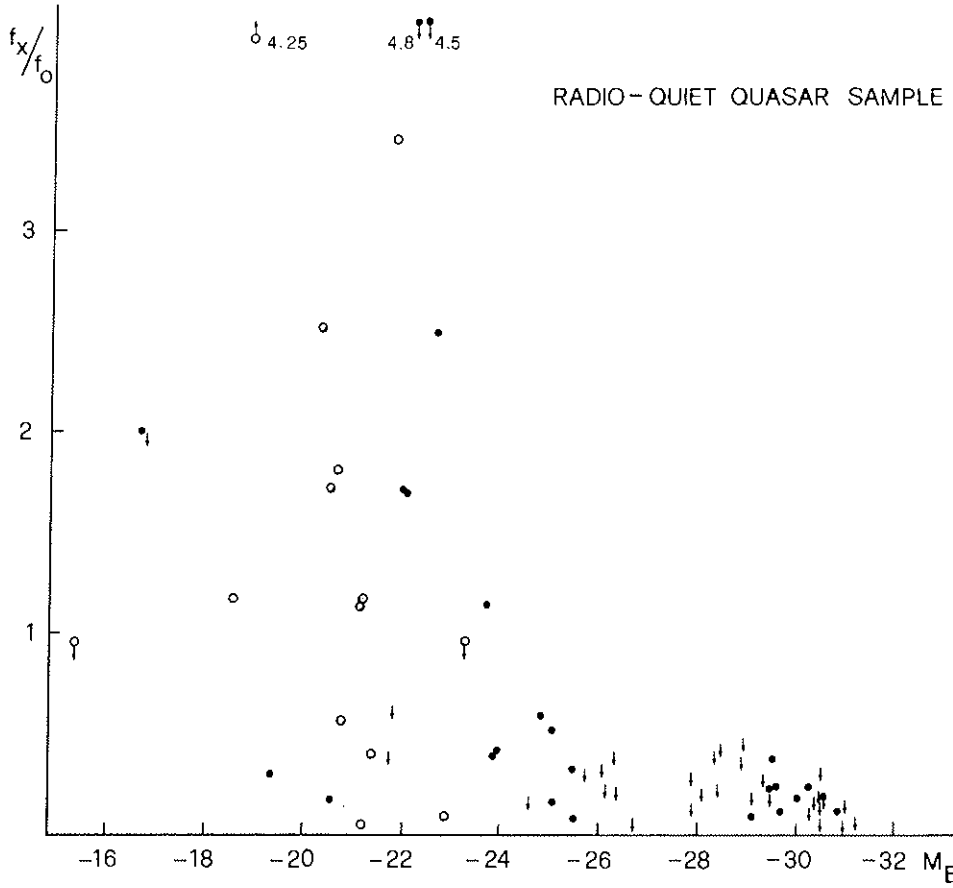


FIG. 3. The distribution of X-ray to optical flux ratios of the combined radio quiet sample of quasars as a function of absolute M_B magnitude. Only relevant upper limits have been included (↓). Sy 1 nuclei from Veron (1979) sample of Table 1 are also shown.

with statements made by Kembhavi and Fabian (1982). In Fig. 3 we have plotted the ratio of the observed X-ray and optical fluxes against the absolute magnitudes, excluding those objects for which the upper limits are too high to be significant. We limit ourselves to those objects with $M_B < -23.8$, which may be considered "true" quasars according to the definition given in Woltjer and Setti (this volume; hereafter referred to as Paper I). It may be noted that no objects have been detected in the interval $-26 < M_B < -29$, which together with the distribution of the

upper limits suggest that, if anything, the detection rate in this interval cannot be higher than that for objects with $M_B < -29$. This is to be contrasted with the detections in the interval $-23.8 < M_B < -26$ which now show a wider range in f_X/f_O and, consequently, a much higher detection rate. Therefore, in comparing the two intervals shown in Fig. 2, we should only consider objects with $M_B < -26$. In the interval $16.5 < M_B < 17.5$, one has detected 5 out of 16 quasars with $f_X/f_O < 0.30$, while the corresponding figure for the interval $17.5 \div 18.5$ is 2 out of 7, as expected. It so happens that the two quasars detected in this last interval have absolute magnitudes $M_B < -29$, which together with the rather peculiar selection of the overall optical sample in turn explains the existence of the gap noted in Fig. 3.

We have taken into account the relevant upper limits of the f_X/f_O ratios, that is to say those that are consistent with the actual detection distributions, by adopting the average f_X/f_O which is the arithmetic mean between the averages obtained by setting the upper limits to zero and those obtained by including the upper limits with their nominal values. As already indicated by Zamorani *et al.* (1981), there exists a very good correlation between the optical and X-ray emissions of quasars. While the intrinsic luminosities span over seven magnitudes, the average f_X/f_O changes less than a factor three.

To estimate the XRB contributed by quasars, we will assume that each interval of absolute magnitude is equally populated at any given apparent magnitude, which leads to an effective ratio:

$$< f_X/f_O >_{eff} = 0.18 \qquad (2)$$

This assumption is not rigorously correct since, for instance, the sample of quasars studied by Arp (1981) around $m_B \simeq 19$ is relatively unpopulated at magnitudes $M_B > -26$, while the existence of a cut-off in the evolution of quasars at $z \sim 3$ would progressively deplete the intrinsically bright objects as one goes to fainter magnitudes, thereby increasing $< f_X/f_O >_{eff}$. In the end, the assumption underlying (2) will lead to a conservative estimate of the contribution of quasars to the XRB.

Since the optical counts contain both radio quiet and radio quasars, before proceeding we have to find out the relative contribution due to this last class of objects. By comparing Fig. 4 with Fig. 2 it is immediately clear that the detection rate of radio quasars is much higher than for radio quiet quasars. This is further evidenced in the plot of the observed f_X/f_O ratios as a function of absolute optical magnitude shown in Fig. 5,

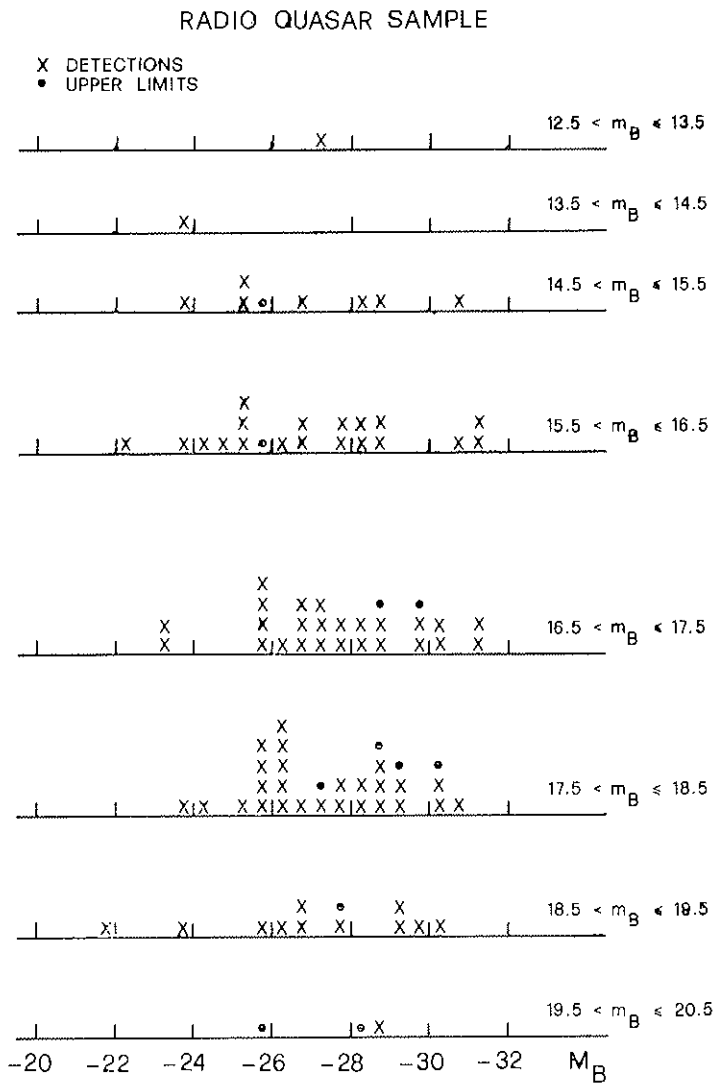


FIG. 4. The combined sample of radio quasars from Ku *et al.* (1980) and Zamorani *et al.* (1981) subdivided in intervals of apparent and absolute B magnitudes. OVV type sources are indicated with crosses.

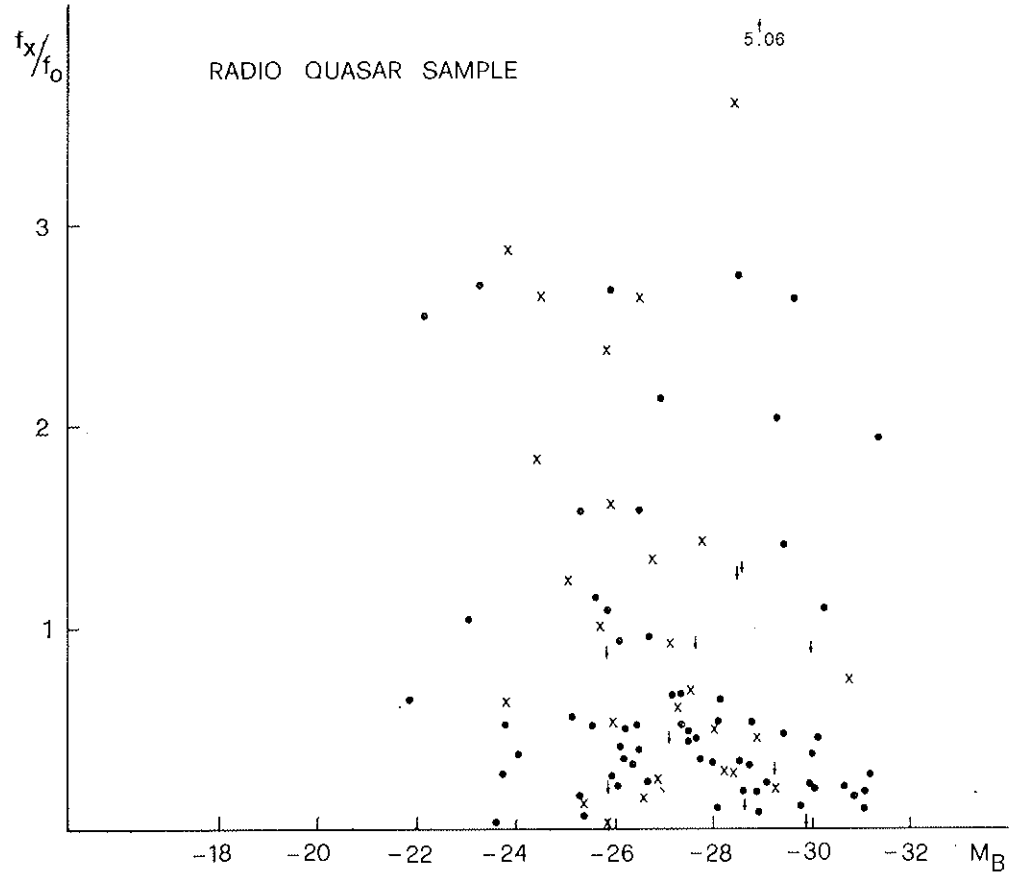


FIG. 5. The distribution of X-ray to optical flux ratios of the combined sample of radio quasars as a function of absolute M_B magnitude.

where one immediately notices a very much larger spread than in the corresponding Fig. 3 for the radio quiet sample. The f_X/f_O ratios for the objects associated with optically strong variable sources (OVV type) have been computed adopting the mean luminosity over the optical variability range, not the mean magnitude, from the data collected in Tanzella-Nitti *et al.* (1982).

As already stated by Ku *et al.* (1980), the strongly variable sources tend on the average to be stronger X-ray emitters than the other radio quasars. Considering only those radio quasars which have $M_B < -23.8$

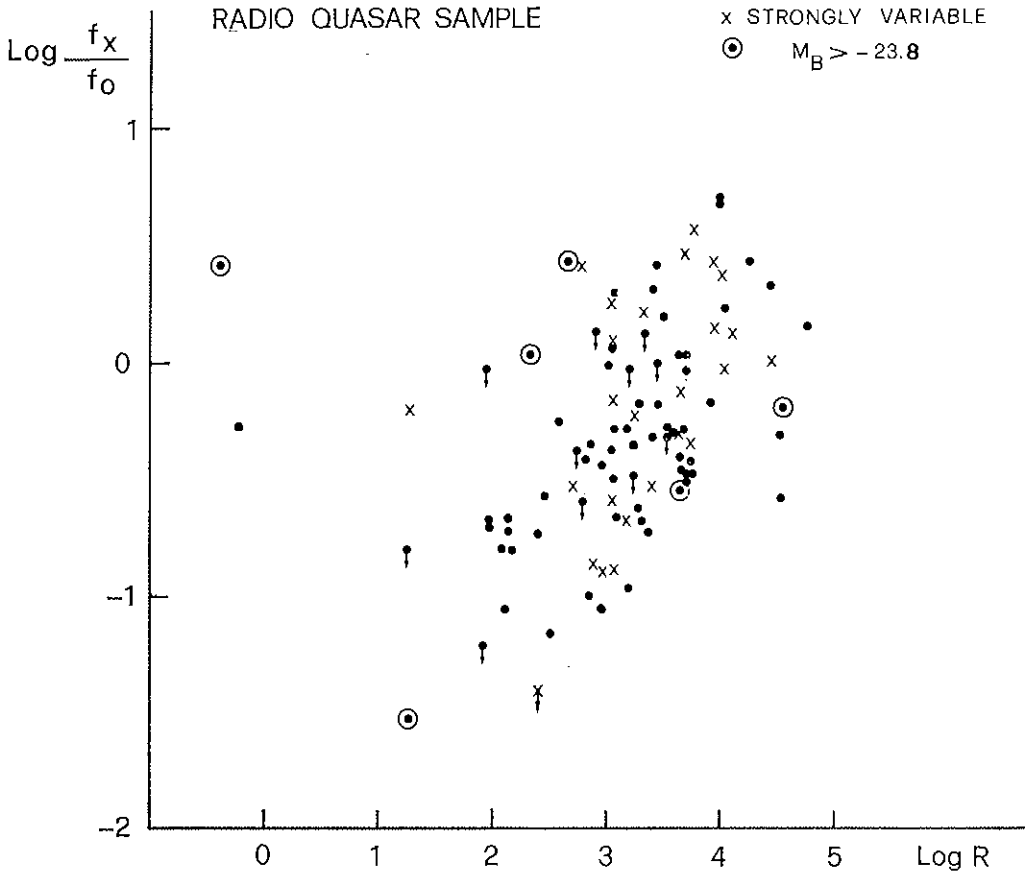


FIG. 6. The distribution of X-ray to optical flux ratio of the combined radio quasars sample against R, the ratio of the radio to optical fluxes in the source rest frame at 5 GHz and 2500 Å respectively.

and located at galactic latitudes $> |30°|$, to avoid uncertain galactic absorption corrections, and excluding OVV type objects, we obtain an average $f_x/f_o = 0.60$ which is about 3 times the value given in (2). A similar result has been obtained by Zamorani *et al.* (1981). The average value for the optically strong variable objects is $f_x/f_o = 1.10$.

In Fig. 6 we plot the f_x/f_o ratios for the overall sample of radio quasars against R, the ratio between the radio and the optical luminosities at 5000 MHz and 2500 Å, respectively, in the rest frame of the sources.

The radio data have been taken from Ku *et al.* (1980) and Zamorani *et al.* (1981). As already noted by Zamorani *et al.* (1981) the existence of a positive correlation between the X-ray and the radio emission is evident. One should note, however, that this correlation does not appear to exist for objects with $M_B > -23.8$ whose points are scattered over the whole diagram. The existence of such a correlation tells us that the contribution of radio quasars to the XRB must be properly weighted over the relative population of this class of objects as a function of R. To this end we have derived the ψ (R) function, first introduced hy Schmidt (1970), for quasars with $M_B < -23.8$. This function, which gives the fraction of all quasars which are radio sources as a function of R, has heen empirically constructed following the procedure outlined by Fanti *et al.* (1977) from the overall sample of optically selected quasars which have been subsequently observed for radio emission (Fig. 7). The data have been collected from Sramek and Weedman (1980) and Condon *et al.* (1981), considering only those objects for which redshifts are well determined. A fuller discussion of this derivation will be given elsewhere, but in passing we note that there is some indication of the appearance of a bimodal distribution. Of course, the significance of this result is still somewhat poor in view of the large statistical errors.

The fraction of radio quasars ($R > 1.8$) is $\simeq 13.1\%$. By convolving the average X-ray emission per interval of Log R of our sample with the ψ (R) function so derived, we compute that radio quasars contribute $\simeq 22\%$ of the X-ray integrated contribution from all quasars, much lower than stated by Ku *et al.* (1980). The inclusion of the optically strong variable quasars would change this figure by only $\sim 2\%$ upward, because, as shown in Fig. 6, their distribution tends to be displaced towards higher values of R and, therefore, falls in the decreasing part of the ψ (R) function. Integrating over the quasars' number count relationship given in Paper I, corrected downwards for the presence of Seyfert type 1 nuclei to be discussed separately in the next section and assuming that the X-ray properties we have discussed apply to the faintest optical apparent magnitudes, we find tbat the contribution of quasars to the XRB is as high as 20.2% down to $m_B = 23$. Approximately 70% of this contribution comes from objects witb $m_B < 21$ and, consequently, the precise knowledge of the counts at ahout this magnitude is of critical importance.

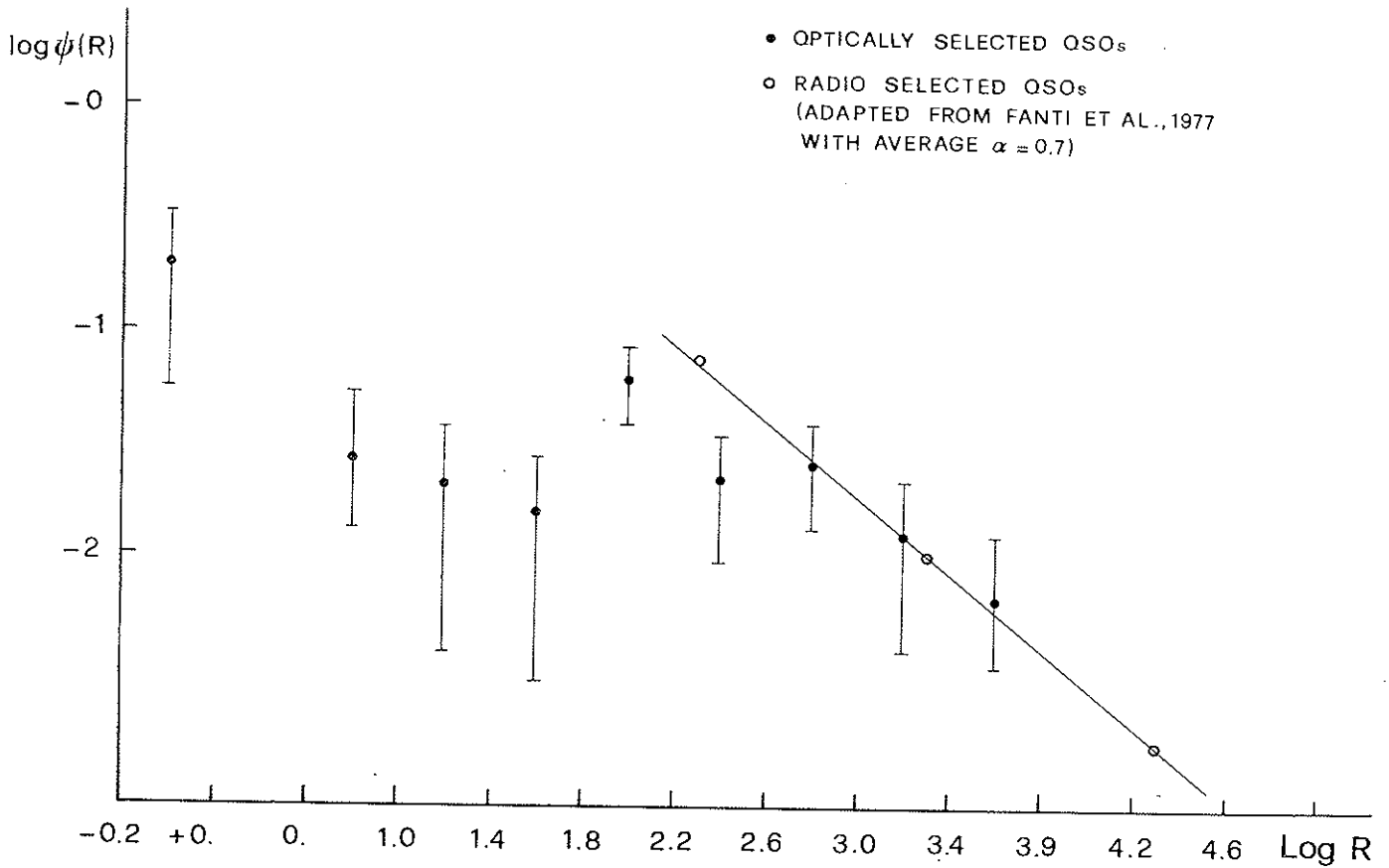


FIG. 7. The morphological $\psi(R)$ function (see the text). The points marked (o) refer to the $\psi(R)$ function derived by the studies of radio quasar samples and have been adapted from Fanti *et al.* (1977) assuming an average radio spectrum of the form $\nu^{-\alpha}$ with $\alpha = 0.7$.

2.2 *Seyfert 1 Nuclei*

From Fig. 3 it is clear that objects with $M_B > -23.8$ have relatively stronger X-ray emission than "true" quasars. This is also confirmed by the f_X/f_O ratio distributions of the serendipitous X-ray source samples obtained by Grindlay *et al.* (1980) and Chanan *et al.* (1981). In addition, we have considered the complete sample of Seyfert 1 nuclei derived by Véron (1979) for which, except for one object, X-ray observations are available. The data are summarized in Table 1 and in Fig. 3, which shows that the f_X/f_O ratios are distributed in the same way as are the objects with $M_B > -23.8$ of the present combined sample. Combining the present sample with the Véron sample (nuclei with $M_B > -18$ are excluded), we find that the X-ray emission from Sy 1 type nuclei may be characterized

TABLE 1 - *Sample of Seyfert nuclei* (Véron 1979).
(complete to $U \leq 15.05$)+

	M_V	U	F_2 (μJ)*	m_X	f_X/f_O
MRK 9	− 21.6	14.54	.38	16.48	0.40
10	− 20.9	14.61	< 1.6	> 14.92	< 1.80
79	− 20.8	14.07	2.5	14.43	1.72
290	− 20.6	15.04	1.5	14.99	2.51
291	− 21.0	14.87	.40	16.43	0.57
304	− 23.1	14.19	.10	17.93	0.08
335	− 21.4	13.74	2.3	14.53	1.16
352	− 19.2	14.79	3.2	14.17	4.25
376	− 22.1	14.82	2.5	14.43	3.44
478	− 23.5	14.16	< 1.3	> 15.15	< 0.96
486	− 21.4	14.73	< 0.03	> 19.24	< 0.04
506	− 21.4	14.93	.75	15.74	1.14
NGC 3516	− 19.5	13.45			
4051	− 15.6	14.42	< 1.0	> 15.43	< 0.95
4151	− 18.8	12.05	11.0	12.83	0.48

+ $U - B = -.95$, $U - V = -.75$

* Adopted X-ray fluxes at 2 keV have been obtained from: Elvis *et al.* (1978), Tananbaum *et al.* (1978), Dower *et al.* (1980) and Kriss *et al.* (1980).

by an average $f_X/f_O \simeq 1.1$. (This figure may be slightly overestimated due to the presence of an object in the Véron sample with a rather large f_X/f_O).

We have argued in the paper by Woltjer and Setti in this volume that the counts provide some evidence that Sy 1 type nuclei may be subject to a considerable cosmological evolution. This together with the enhanced average, f_X/f_O, suggests that their contribution to the XRB may indeed be large. To further constrain this contribution we have made use of the Einstein deep survey result: 19.2 ± 7.9 sources down to a limiting X-ray flux $\simeq 7.9 \times 10^{-6}$ keV/cm^2 sec keV at 2 keV. This is shown in Fig. 8

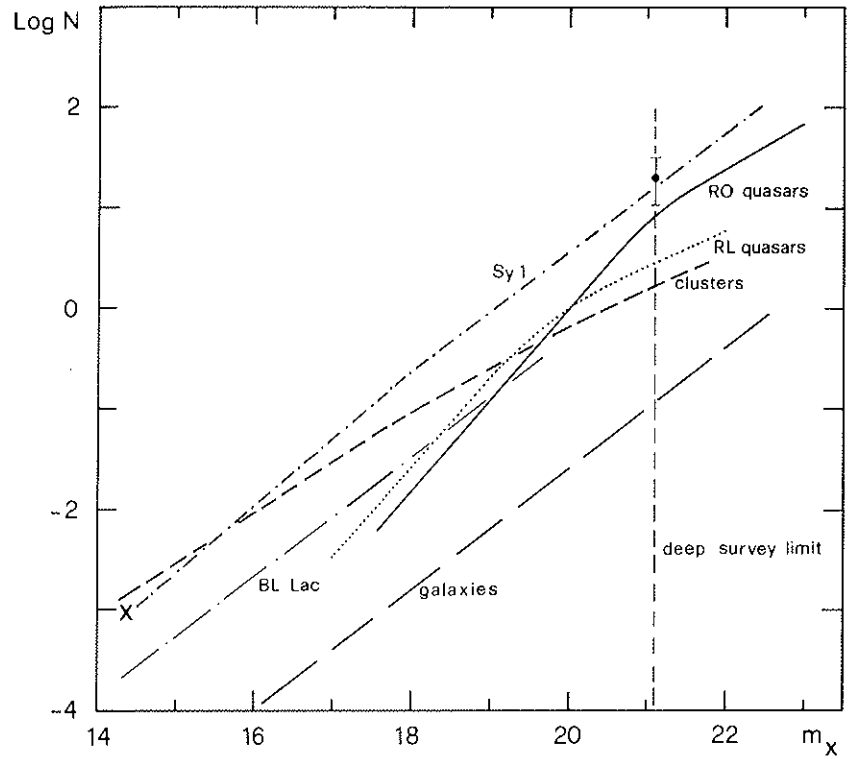


FIG. 8. Integral source counts per square degree. The curve "clusters" is from Maccacaro *et al.* (1982). Other curves are from source counts of Paper I. The RL curve is from quasar optical counts using our (R) function and the distribution of the f_X/f_0 ratios with log R. The point (•) is the deep survey point of Giacconi *et al.* (1979). The point (X) is from the Sy 1 galaxies in the complete X-ray sample of Piccinotti *et al.* (1982). The sum of the various components as a function of m_X is consistent with the Euclidean slope valid down to the deep survey limit.

where we have plotted the Log N (m_X) relationships for different classes of sources. For each class of objects the X-ray source counts have been obtained by convolving the optical counts with the observed distribution of the appropriate f_X/f_O ratios.[1] It is seen that only quasars and possibly Sy 1 type nuclei are the main contributors to the counts at the deep survey limit. The Sy 1 source counts so derived are consistent with the optical data discussed in Paper I and with the X-ray data now available. The contribution of Sy 1 nuclei to the XRB is readily obtained by integration over the source counts and it amounts to $\sim 34\%$ down to $m_B = 23$. At this magnitude the surface density of the Sy 1 nuclei, for which the adopted slope of the integral counts is $\cong -1.5$ for $m_B \gtrsim 19$, begins to be comparable to that of the quasars ($\sim 60\%$).

Classical estimates of the Sy 1 contribution to the XRB, in the 2 to 10 keV energy interval with no cosmological evolution, range from 6 to $\sim 20\%$ depending on the assumptions made about the local luminosity function of these objects, the source variability and the adopted model universe (Elvis *et al.* 1978, Tananbaum *et al.* 1978, Mushotzki *et al.* 1980, Piccinotti *et al.* 1982).

2.3 *BL Lac Type Objects*

These objects are known to be highly variable at all wavelengths and to be relatively strong X-ray sources. Available optical and X-ray data for BL Lac objects are summarized in Table 2, where, in deriving the average X-ray and optical magnitudes, we have assumed that the objects spend the same time in equal magnitude intervals. Since very little is known about the time correlated flux variations at optical and X-ray frequencies, not much meaning can be attached to the f_X/f_O ratios given in Table 2. Disregarding two objects whose f_X/f_O ratios are so high that their contribution would be more important than that from all the remaining objects listed in Table 2, we find an average $f_X/f_O \simeq 1.0$. From the discussion given in the paper by Woltjer and Setti in this volume, we obtain an upper limit of 7% for the contribution of the BL Lac objects to the XRB by assuming

[1] In the simple case of integral source counts, which can be represented by a single power law of slope $-n$, it can be readily shown that the relationship between the X-ray and the optical fluxes is obtained by applying a conversion factor of the form, $[\int x^n F(x)\,dx]^{1/n}$, where $x = f_X/f_O$ and F is the normalized distribution function of the f_X/f_O ratios.

TABLE 2 - *BL Lac sample. Average optical and X-ray magnitudes.**

	$\bar{B}$	$\bar{m}_X$	I_X/I_O
0219 + 42 3C 66A	14.91	15.19	.77
0235 + 16	18.18	17.32	2.21
0521 − 36	15.59	15.12	1.54
0548 − 32	15.79	13.73	6.67
0735 + 17	15.07	17.41	0.13
1101 + 38 MRK 421	13.94	14.16	0.82
1219 + 30	17.15	14.42	12.36
1400 + 16	16.48	17.75	0.31
1514 − 24 AP Lib	14.93	17.57	0.09
1538 + 14	15.73	17.96	0.13
1652 + 39 MRK 501	14.55	13.67	2.25
1727 + 50	15.70	15.49	1.21
2155 − 30	13.44	12.62	2.13
2200 + 42 BL Lac	13.46	< 15.12	< 0.22
2254 + 07	16.40	18.23	0.19
2335 + 03	18.63	> 18.63	< 1.00

* X-ray data from: Schwartz *et al.* (1979), Ku (1980) and Maccagni and Tarenghi (1981). Average B magnitudes ($= 1/2\,(B_{min} + B_{max})$) have been derived from published light curves and corrected for galactic absorption.

that the optical counts for this class of objects may be extrapolated down to $M_B = 23$ with the Euclidean slope.

We note that the BL Lac X-ray source counts drawn in Fig. 8 would predict the existence of 6 BL Lac objects in the complete X-ray sample of Piccinotti *et al.* (1982) and that 4 have been actually found. In view of the difficulty of taking into proper account the source variability, the agreement may be considered very satisfactory.

2.4 *Normal Galaxies*

It has commonly been considered that normal galaxies could contribute only ~ 2% of the XRB, based on the assumption that the same ratios of

X-ray to optical luminosity for our galaxy and for M31 are typical (Setti and Woltjer 1973), unless cosmological evolution effects set in (Silk 1968).

The Einstein Observatory has been used to observe a large sample of normal galaxies which includes a representative mixture of all galaxy types. From the data on 70 galaxies given by Long and Van Speybroeck (1981) we have estimated an average $f_X/f_O \simeq 7.4 \times 10^{-4}$. This value is ~ 4.5 times the ratio found for M31 and, consequently, from our previous estimate we find that normal galaxies may account for $\sim 9\%$ of the XRB. This is quantitatively compatible with the estimate one obtains, following a suggestion made by Long and van Speybroeck (1981), by directly comparing the average f_X/f_O ratio with the intensity of the extragalactic background light given by Dube *et al.* (1977) as equivalent to that of 1 ± 1.2 stars of visual magnitude 10 per square degree at 5100 Å, largely contributed by light in the B band due to redshift effects.

The contribution of normal galaxies to the XRB may become larger if evolution effects such as those proposed by Bookbinder *et al.* (1980) are present. These authors have proposed that young galaxies at a redshift $z \simeq 2$ to 3 may possess an enhanced X-ray emission due to hot galactic winds powered by an increased rate of supernovae explosions and to an increased number of hard X-ray binaries. One nice feature of this model is that it may predict the right kind of spectral shape. However, if young galaxies do indeed contribute the bulk of the XRB, one must be extremely careful not to exceed the γ-ray background. A higher supernovae rate would probably induce a higher production of cosmic rays and, consequently, enhanced γ-ray production via inverse Compton and π^0 decay due to *p-p* interactions. While our galaxy is inconspicuous in X-ray, it is quite visible at γ-ray energies above 35 MeV and, in fact, above 200 MeV the background features can all be explained in terms of a galactic origin (Bignami *et al.* 1979). If the X- to γ-ray flux ratio of our galaxy is preserved, then young galaxies cannot contribute more than $\sim 20\%$ of the XRB radiation flux. We note that the K-correction terms do not seem to play any substantial role in the above argument, since the spectrum of the galactic γ-ray emission is rather flat and, in fact, possesses a slope very close to that of the XRB spectrum at low energies.

2.5 *Clusters of Galaxies*

Extensive observations of Abell clusters have permitted the derivation of the X-ray luminosity function for this class of objects and there is now

fairly good agreement that it is unlikely that they contribute more than 10% of the XRB at 2 keV (McKee *et al.* 1980, Piccinotti *et al.* 1982, Ulmer *et al.* 1981). The contribution becomes of course less important with increasing energy due to the very rapid decrease of the thermal bremsstrahlung spectrum of a typical cluster.

2.6 *Others*

The contribution due to other active galaxies such as Sy 2, Sy 1.8, radio galaxies, etc., is somewhat uncertain. On the basis of a complete X-ray sample, Piccinotti *et al.* (1982) have estimated that the XRB contributed by active galactic nuclei with no cosmological evolution is $\lesssim 20\%$ in the 2 to 10 keV energy interval. This estimate is based on 23 objects, 17 of which are Sy 1 and the remaining are generically classified in a broad class to include Sy 2, narrow-line emission galaxies and N galaxies. However, considering the objects listed in this last category, we note that 3C 445 is an N galaxy with a very broad emission line spectrum (Osterbrock *et al.* 1976), so that it would naturally fall into our definition of Sy 1 nuclei, and that NGC 2992 can be more properly classified as a Sy 1 galaxy (Véron *et al.* 1980).

Therefore, we conclude that the contribution to the XRB due to active galaxies other than Sy 1 probably does not exceed 4%.

3 - Concluding Remarks

From our discussion it follows that known classes of sources are sufficient to explain the XRB intensity at 2 keV. At least 55% of the intensity could be easily contributed by quasars and by Sy 1 nuclei, and even more, if one allows for reasonable extrapolations of the adopted source counts for these objects.[2]

The contribution due to quasars depends critically on the exact form of the optical number count relationship at magnitudes $m_B \simeq 20$ to 21, while the contribution due to Sy 1 nuclei, which in our picture constitute the dominant population (Fig. 8), depends on the assumed source counts down to the faintest magnitudes as constrained by the Einstein Observatory

[2] A similar conclusion has been reached by Zamorani (1981) on the basis of a quite different type of analysis.

deep survey point. In itself this would indicate the presence of a rather strong cosmological evolution. The identification content of intermediate sensitivity X-ray surveys (e.g. Maccacaro *et al.* 1982) down to the deep survey limit may shed some light on the correctness of our assumptions as schematically represented in Fig. 8. Here it should be remarked that although radio quasars constitute only a small fraction of all quasars, due to their enhanced X-ray emission they will show up more numerous than radio quasars in X-ray surveys down to $m_X \sim 20$. Beyond that magnitude the radio quiet quasars will take over because the turn over in the optical counts for these objects, which are intrinsically weaker X-ray emitters, is reflected at fainter X-ray magnitudes. Moreover it should be noted that the approximate Euclidean slope found for the overall X-ray source counts is determined by the different contributions of various classes of sources at different fluxes, the galaxy clusters and the Sy 1 galaxies being the dominant contributors at the bright end, while the cosmological evolution of Sy 1 nuclei and quasars takes over at the fainter end.

In view of the still uncertain knowledge of the optical source counts, and of the fact that we have been forced to assume that the observed X-ray emission properties apply down to the faintest optical magnitudes, our estimate of the contribution of these nuclei to the XRB must be considered still tentative. If quasars and Sy 1 nuclei would have to account for most of the XRB between, say 2 and 50 keV, then their spectra on the average cannot be too different from the XRB spectrum. De Zotti *et al.* (1981) have investigated in some detail the problem of reproducing the XRB spectrum in the interval 3 to 50 keV by means of a superposition of power law sources and they have concluded that the photon spectral indices must possess a rather narrow range centered around 1.4 to 1.5.

At present only two quasars have well measured X-ray spectra: 3C 273 and QSO 0241 + 622. While the low energy X-ray photon spectrum of 3C 273 can be fitted by a power law with a spectral index ~ 1.4 (Primini *et al.* 1979, Worral *et al.* 1979), in excellent agreement with the XRB spectrum in the range 2 to 20 keV, a power law fit to the photon spectrum of QSO 0241 + 622 can be obtained with a spectral index ~ 1.93 (Worral *et al.* 1980). From the observed distribution of the hardness ratio of sources in their sample, Zamorani *et al.* (1981) tentatively suggest that the average photon spectral index may be around 1.4 to 1.5. On the other hand, observations of the X-ray spectra of a local sample of Seyfert 1 galaxies indicate that almost all can be fitted by power laws between 3 and 50 keV with photon spectral indices in the range 1.3 to 2,

the average being ~ 1.65 (Mushotzky *et al.* 1980). This is somewhat at variance with what is required to fit the XRB. However, we believe that the spectral properties of much larger samples of quasars and Sy 1 nuclei spread over the luminosity functions should be determined before one can definitely settle the question of the primary contribution of these sources to the XRB.

It should be noted that, in any case, at least 50% of the XRB at 2 keV are certainly due to the integrated contribution of known classes of sources and this sets a limit to the hypothetical contribution of a diffuse hot intergalactic gas (Giacconi 1980). Even if the relative importance of these sources would decrease with increasing energy, their contribution at lower energies must be subtracted and one would be left with a residual XRB spectrum whose interpretation would probably require the existence of a different and as yet unknown component.

The interpretation of the extra-galactic γ-ray background does not seem to pose such severe problems at the moment, mainly due to the very poor information we have on the emission properties from the various classes of objects. A quasar (3C 273), a Seyfert galaxy (NGC 4151) and a radiogalaxy (Cen A) have been detected up to now in γ-rays, all of them being also well known X-ray sources. One should include our own galaxy also. If these sources are representative members of their classes, one can in fact easily show that under very reasonable assumptions the full intensity of the γ-ray background from 1 to $\sim$ 200 MeV can be satisfactorily explained in terms of unresolved sources (Bignami *et al.* 1979, Woltjer 1980 and references therein). In fact only a small fraction ($\sim$ 10%) of the Sy 1 and of the quasars could emit at the same level as NGC 4151 and 3C 273, otherwise the γ-ray backgrounds at $\sim$ 1 MeV and $\sim$ 100 MeV would be respectively violated. It is clear that many more observations of extra-galactic objects in the γ-ray domain are needed to constrain the various possible models.

ACKNOWLEDGEMENTS

We should like to thank Dr. G. Zamorani for very informative discussions.

REFERENCES

Arp, H., 1981, *Astrophys. J.*, **250**, 31.

Bignami, G.F., Fichtel, C.E., Hartman, R.C. and Thompson, D.J., 1979, *Astrophys. J.*, **232**, 649.

Bond, H.E., Kron, R.G. and Spinrad, H., 1977, *Astrophys. J.*, **213**, 1.

Bookbinder, J., Lowie, L.L., Krolik, J.H., Ostriker, J.P. and Rees, M.J., 1980, *Astrophys. J.*, **237**, 647.

Chanan, G.A., Margon, B. and Downes, R.A., 1981, *Astrophys. J.*, **243**, L5.

Condon, J.J., O'Dell, S.L., Puschell, J.J. and Stein, W.A., 1981, *Astrophys. J.*, **246**, 624.

De Zotti, G., Boldt, E.A., Cavaliere, A., Danese, L., Franceschini, A., Marshall, F.E., Swank, H.J. and Szymkowiak, A.E., 1981, *Astrophys. J.*, **253**, 47.

Dower, R.G., Griffiths, R.E., Bradt, H.V., Doxsey, R.E. and Johnston, M.D., 1980, *Astrophys. J.*, **235**, 355.

Dube, R.R., Wickes, W.C. and Wilkinson, D.T., 1977, *Astrophys. J.*, **215**, L51.

Elvis, M., Maccacaro, T., Wilson, A.S., Ward, M.J., Penston, M.V., Fosbury, R.A.E. and Perola, G.C., 1978, *Mon. Not. Roy. Astr. Soc.*, **183**, 129.

Fabian, A.C., 1980, Tenth Texas Symp. on Relativistic Astrophysics, Baltimore, in press.

Fanti, C., Fanti, R., Lari, C., Padrielli, L., van der Laan, H. and de Ruiter, H., 1977, *Astron. Astrophys.*, **61**, 487.

Fichtel, C.E., Simpson, G.A. and Thompson, D.J., 1978, *Astrophys. J.*, **222**, 833.

Fichtel, C.E., Hartman, R.C., Kniffen, D.A., Thompson, D.J., Bignami, G.F., Ögelman, H., Özel, M.F. and Tümer, T., 1975, *Astrophys. J.*, **198**, 163.

Field, G.B. and Perrenod, S.C., 1977, *Astrophys. J.*, **215**, 717.

Giacconi, R., Bechtold, J., Branduardi, G., Forman, W., Henry, J.P., Jones, C., Kellog, E., van der Laan, H., Liller, W., Marshall, H., Murray, S.S., Pye, J., Schreier, E., Sargent, W.L.W., Seward, F. and Tananbaum, H., 1979, *Astrophys. J.*, **234**, L1.

Giacconi, R., 1980, Proc. of the NATO-ASI on "X-Ray Astronomy", R. Giacconi and G. Setti (eds.), D. Reidel Publ. Co., Dordrecht-Holland, 396.

Grindlay, J.E., Steiner, J.E., Forman, W.R., Canizares, C.R. and McClintock, J.E., 1980, *Astrophys. J.*, **239**, L43.

Hewitt, A. and Burbidge, G., 1980, *Astrophys. J. Supp.*, **43**, 57.

Iwan, D., Boldt, E.A., Marshall, F.E., Mushotzky, R.F., Shafer, R. and Stottlemyer, A., 1981, *Astrophys. J.*, in press.

Kembhavi, A.K. and Fabian, A.C., 1982, *Mon. Not. Roy. Astr. Soc.*, **198**, 921.

Kinzer, R.L., Johnson, W.N. and Kurfers, J.D., 1978, *Astrophys. J.*, **222**, 370.

Kriss, G.A., Canizares, C.R. and Ricker, G.R., 1980, *Astrophys. J.*, **242**, 492.

Ku, W.K.-M., 1980, *Highlights in Astronomy*, **5**, 677.

Ku, W.H.-M., Helfand, D.J. and Lucy, L.B., 1980, *Nature*, **288**, 323.

Long, K.S. and Van Speybroeck, L.P., 1981, preprint.

Maccacaro, T., Feigelson, E.D., Fener, M., Giacconi, R., Gioia, I.M., Girffiths, R.E., Murray, S.S., Zamorani, G., Stocke, J. and Liebert, J., 1982, *Astrophys. J.*, **253**, 504.

Maccagni, D. and Tarenghi, M., 1981, *Astrophys. J.*, **243**, 42.

Marshall, F.E., Boldt, E.A., Holt, S.S., Miller, R.B., Mushotzky, R.F., Rose, L.A., Rothschild, R.E. and Serlemitsos, P.J., 1980, *Astrophys. J.*, **235**, 4.

Matteson, J.L., Gruber, D.E., Nolan, P. and Peterson, L.E., 1979, *B.A.A.S.*, **11**, 653.

McKee, J.D., Mushotzky, R.F., Boldt, E.A., Holt, S.S., Marshall, F.E., Pravdo, S.H. and Serlemitsos, P.J., 1980, *Astrophys. J.*, **242**, 843.

Mushotzky, R.F., Marshall, F.E., Boldt, E.A., Holt, S.S. and Serlemitsos, P.J., 1980, *Astrophys. J.*, **235**, 377.

Osterbrock, D.E., Koski, A.T. and Phillips, M.M., 1976, *Astrophys. J.*, **206**, 898.

Piccinotti, G., Mushotzky, R.F., Boldt, E.A., Holt, S.S., Marshall, F.E., Serlemitsos, P.J. and Shafer, R.A., 1982, *Astrophys. J.*, **253**, 485.

Primini, F.A., Cooke, B.A., Dobson, C.A., Howe, S.K., Scheepmaker, A., Wheaton, W.A., Lewin, W.H.G., Baity, W.A., Gruber, D.E., Matteson, J.L. and Peterson, L.E., 1979, *Nature*, **278**, 234.

Richstone, D.O. and Schmidt, M., 1980, *Astrophys. J.*, **235**, 377.

Schmidt, M., 1970, *Astrophys. J.*, **162**, 371.

Schönfelder, V., Grami, F. and Penningsfeld, F.-P., 1980, *Astrophys. J.*, **240**, 350.

Schwartz, D.A., 1970, *Astrophys. J.*, **234**, 4.

— 1979, *X-Ray Astronomy* (COSPAR), W.A. Baity and L.E. Peterson (eds.), Pergamon Press, Oxford and New York, 453.

— 1980, *Physica Scripta*, **21**, 644.

Schwartz, D.A., Doxsey, R.E., Griffiths, R.E., Johnston, M.D. and Schwarz, J., 1979, *Astrophys. J.*, **229**, L53.

Setti, G. and Rees, M.J., 1970, *I.A.U. Symp. No. 37 on "Non-Solar X- and Gamma Ray Astronomy"*, L. Gratton (ed.), Reidel Publ. Co., Dordrecht, Holland, 352.

Setti, G. and Woltjer, L., 1970, *Astrophys. Sp. Sci.*, **9**, 185.

— 1973, *Proc. I.A.U. Symp. No. 55 on "X- and Gamma Ray Astronomy"*, H. Bradt and R. Giacconi (eds.), Reidel Publ. Co., Dordrecht, Holland, 208.

— 1979, *Astron. Astrophys.*, **76**, L 1.

Silk, J., 1968, *Astrophys. J.*, **151**, 459.

— 1973, *Ann. Rev. Astron. Astrophys.*, **11**, 269.

Sramek, R.A. and Weedman, D.W., 1980, *Astrophys. J.*, **238**, 435.

Tananbaum, H., Peters, G., Forman, W., Giacconi, R., Jones, C. and Avni, Y., 1978, *Astrophys. J.*, **223**, 74.

Tanzella-Nitti, G., Setti, G. and Zamorani, G., 1982, in preparation.

Trombka, J.I., Dyer, C.S., Evans, L.G., Bielefeld, M.J., Seltzer, S.M. and Metzger, A.E., 1977, *Astrophys. J.*, **212**, 925.

Ulmer, M.P., Kowalsky, M.P., Gruddace, R.G., Johnson, M., Meekins, J., Smathers, H., Yentis, D., Wood, K., McNutt, D., Chubb, T., Byram, E.T. and Friedman, H., 1981, *Astrophys. J.*, **243**, 681.

Véron, P., 1979, *Astron. Astrophys.*, **78**, 46.

Véron, P., Lindblad, P.O., Zuiderwijk, E.J., Véron, M.P. and Adam, G., 1980, *Astron. Astrophys.*, **87**, 245.

Warwick, R.S., Pye, J.P. and Fabian, A.C., 1980, *Mon. Not. Roy. Astr. Soc.*, **190**, 243.

Worral, D.M., Mushotzky, R.F., Boldt, E.A., Holt, S.S. and Serlemitsos, P.J., 1979, *Astrophys. J.*, **232**, 683.

Worral, D.M., Boldt, E.A., Holt, S.S. and Serlemitsos, P.J., 1980, *Astrophys. J.*, **240**, 421.

Worral, D.M., Marshall, F.E., Boldt, E.A. and Swank, J.H., 1981, NASA Tech. Mem. 83844.

Woltjer, L., 1980, *Highlights of Astronomy*, P.A. Wayman (ed.), **5**, 753.

Zamorani, G., Henry, J.P., Maccacaro, T., Tananbaum, H., Soltan, A., Avni, Y., Liebert, J., Stocke, J., Strittmatter, P.A., Weymann, R., Smith, M.G. and Condon, J.J., 1981, *Astrophys. J.*, **245**, 357.

Zamorani, G., 1981, Paper presented at the Oxford International Symposium "Progress in Cosmology", in press.

DISCUSSION

DAVIS

Suppose one hypothesised that all sources had the same f_X/f_O. What value is allowed without violating the limits of the optical sky brightness set by Dube *et al.* (1977)?

SETTI

About 25% of the X-ray background.

OSTRIKER

With respect to the contribution of galaxies to the X-ray background, the background optical light is contributed mainly by relatively nearby galaxies due to the redshift corrections affecting distant galaxies. But the distant ones emit, in their own frame, much more optical energy, if any of our present thoughts concerning galactic evolution are correct. If these distant galaxies have the same X-ray to optical emission in their own frames, then they make a major contribution to the X-ray background.

SETTI

I completely agree with you. The use of the Dube *et al.* limit is just a quick way to get an estimate of the possible contribution of normal galaxies down to a redshift of ~ 1. No K-correction terms have been considered. By performing the usual integration over the distribution of the galaxies in the universe, and adopting the same $< f_X/f_O >$ ratio, one would obtain a contribution to the X-ray background of 8 to 10 percent.

FABER

Is the small bump at 1 Mev in the X-ray background the same feature which has inspired interpretation in terms of π^o annihilation at large redshifts? What is your feeling about the status of that interpretation at the present time?

SETTI

Yes. I believe that the history of the interpretations of various astrophysical backgrounds, including the X-rays we are just discussing, may teach us that before resorting to some kind of exotic interpretation it is perhaps wise to completely settle the contributions due to discrete sources.

REES

The ~ 1 MeV bump has been interpreted primarily by people who advocated cosmologies symmetrical between matter and antimatter as photons from π^0 decay at $z \cong 100$.

SCHMIDT

You listed f_X/f_O as a function of M_B for radio quiet quasars and used an average value. I would expect that the intrinsically fainter quasars would dominate in the background, suggesting that a rather higher value of f_X/f_O should be used.

SETTI

In deriving the average X-ray to optical ratio for quasars with $M_B < -24$ I have assumed that at magnitudes around 19 to 20 in the optical counts, which would most contribute to the X-ray background, the various intervals of absolute magnitude are equally represented. This is consistent, for instance, with the evolutionary model derived by Braccesi *et al.* (1980, *Astron. Ap.*, **85**, 80) which in fact at these magnitudes has relatively more objects on the bright side of the luminosity interval. At a redshift of ~ 3 the distance modulus is about 47 and only the brightest objects ($M_B \lesssim -28$) may not be more represented in the counts. Given the fact that only upper limits are present in the X-ray sample under discussion in the absolute magnitude interval -26 to -29, the adopted $< f_X/f_O >$ appears reasonable. It is true that going to fainter magnitudes ($m_B \gtrsim 21.5$) the average f_X/f_O may be underestimated and the average for the -24 to -26 absolute magnitude interval may be more appropriate. Taking this into account, the contribution of the quasars to the X-ray background may increase by a few percent above the estimated values I have presented.

SCHMIDT

You compared the predicted counts of X-ray sources with the deep Einstein

survey of Giacconi *et al.* (1979). How do they compare with the observed counts at brighter fluxes, from HEAO-1 and by Maccacaro *et al.* (1982)?

SETTI

As far as I can see, they are in agreement. As you can judge from the Log N$-m_x$ diagram, summing over all the sources, the slope from the bright end (essentially clusters) to the deep survey point is approximately 1.5.

REES

I would like to ask about the apparently thermal component that seems to dominate the background at 10 to 50 keV. If this is really due to bremsstrahlung, it could be the integrated effect of distant objects containing gas at the right temperature (e.g. quasars, whose X-rays could be Comptonised bremsstrahlung; or supernova-driven winds from young galaxies). This does not involve tbe "inefficiency" and high energy requirements of a hot diffuse IGM. However, are you convinced that the energy requirements for the hot IGM are really all that disturbing?

SETTI

With respect to the first part of your question, I certainly agree that, if the sources which most contribute to the X-ray background do indeed emit thermal bremsstrahlung spectra at the appropriate temperatures, then one can perhaps easily explain the background even at relatively high energies. However, the few examples we have among Seyfert 1 galaxies and quasars do not appear to emit thermal bremsstrahlung spectra. If the sources which appear to contribute most of the X-ray background at 2 keV (quasars and Sy 1 nuclei) are indeed non-thermal, then the possible contribution of a thermal component cannot of course exceed what remains at 2 keV after subtraction of the integrated contribution of these same sources. Concerning the second part of your question, I must say I am not particularly worried about heating a truly diffuse IGG at a very high temperature. However, it seems to be an unnecessary waste of energy unless one is really compelled to do it.

WOLTJER

I think the real problem of the hard X- and γ-ray background is not to find further mechanisms but to explain why it is not much stronger than observed. The few sources for which spectra were available into the Mev range

appear to be flatter than the background, even though this is the same type of sources which are dominant at 2 keV.

GRATTON

In view of the various sources of uncertainty, can you say what is the reliability of your evaluated contributions? May they be wrong by a factor 2 or 3?

SETTI

It seems to me that there is now fairly good evidence that quasars and Sy 1 nuclei must contribute at least 40 to 50 percent of the X-ray background at an energy of a few keV.

REES

Just a comment on the implications of the isotropy of the X-ray background: this isotropy, on angular samples of 10° to 30°, is determined by Ariel V and HEAO-1 data with a precision of a few percent. In any model in which the X-ray background originates at $z \simeq 2$, this translates into an upper limit for $(\delta\rho/\rho)$ on scales of several hundred megaparsecs. This may be a better constraint on the amplitude of possible inhomogeneities on these scales than is obtainable from, for instance, quasar or galaxy counts; it is thus important information complementary to what we get from the microwave background anisotropies.

SILK

Since inverse Compton X-ray emission from the galactic halo amounts to $\gtrsim 1$ to 2 percent of the isotropic X-ray background, it may not be possible to extract much cosmological information by detecting fluctuations in the background.

HAWKING

Is there any evidence for a bump in the spectrum at about 70 MeV due to π^0 production in $p-\bar{p}$ annihilation?

SETTI

I do not think that any such feature has been observed.

THE DISTRIBUTION OF QUASAR REDSHIFTS

L.Z. FANG
Astrophysics Research Division
University of Science and Technology of China

Whether a periodicity exists in the distribution of emission line redshifts for quasars has been studied for a long time. As early as the 1960's Burbidge (1968) discussed the presence of two peaks in the distribution of quasar redshifts $z = 0.06$ and 1.95, and he also pointed out the probable existence of periodicity in that distribution. Subsequent investigations performed by Cowan (1969) and Burbidge and O'Dell (1972) confirmed further Burbidge's results, while Plagemann *et al.* (1969) and Wills and Ricklefs (1976) could not detect any period.

In 1971 Karlsson suggested that the periodicity is with respect to argument $x = 1n\,(1 + z)$, but not to z. The number of quasars included in Karlsson's statistical analysis is only 166. In 1977 this analysis was extended to 574 quasars with available redshift data (Karlsson 1977). In this case the periodicity still persists with the length 0.089 in the variable $x = \log\,(1 + z)$. A similar result of period 0.196 has been obtained by Barnothy and Barnothy (1976).

The purposes of the paper summarized here are twofold: the first is to reinvestigate the confidence level of the periodicity for a much larger number of quasars listed in the new catalogue of Hewitt and Burbidge (1980); the second is to discuss the possible origin of the periodicity and its cosmological implication (Fang *et al.* 1982).

Several authors have discussed the presence of irregularities in the redshift distribution and consider it to be due to various selection effects (cf. Schmidt 1975, Roeder 1971, Roeder and Dyer 1972, Basu 1975). It has been shown (Karlsson 1977) that selection effects might play a

significant role in the formation of certain peaks in the redshift distribution; however, it is very difficult to use them to explain any real periodicity.

The fundamental idea of our model is that the periodicity might be the remains of density wave (sound) perturbations occurring just before the recombination epoch of Big Bang cosmology: some sound-like fluctuations occurring just before the recombination with large wavelength become the "seed" of Jeans instability after that time. The periodic structure of the seeds could then be preserved and amplified hy the clustering at successive times. If the formation of quasars is correlated with density inhomogeneity, the existence of the periodic components in the quasar redshift distribution is a natural outcome.

Several consequences of this model which can be tested by statistical methods are as follows:

1) According to the cosmological principle, it is impossible that the Earth just occupies a preferred position near the center of a spherical wave perturbation. The phases of waves propagating to the Earth from different directions are random. Thus the periodicity should be smeared by superposition of such density wave fluctuations. On the other hand a density perturbation from a certain direction can avoid the effect of random phases. We expect, therefore, that the periodicity should be more remarkable for a set of quasars located near a certain position of the sky than that for quasars in the whole sky.

2) The periodicity is with respect to $x = \mathrm{F}(z, q_0)$, which is a function of redshift and the deceleration parameter, q_0. Karlsson's form $x = 1n\,(1 + z)$ corresponds to $q_0 = 0$. However, support for a value $q_0 > 0.5$ has recently been obtained by investigating the cosmological implication of massive neutrinos and the redshift-magnitude relations of various quasar subsets (cf. Qing *et al.* 1981, Cheng *et al.* 1981). We expect, therefore, that the periodicity with respect to $x = \mathrm{F}(z, q_0 > 0.5)$ should he more remarkable than that with respect to $1\,n\,(1 + z)$.

3) Before recombination, the sound fluctuations were being dissipated by interaction between matter and radiation. The dissipation is lower for longer wavelengths. The strongest "seed" effect is thus given by the sound perturbation with the largest allowed wavelength just hefore recombination. From this we can estimate the length of the period in the quasar redshift distribution.

Our power spectrum analysis shows a peak for $\Delta x = 1/6$ even though the Burbidge-Hewitt list includes quasars over a wide area of sky (whereas

we would expect it to be practically smeared unless we are in a preferred central position). We do *not* however see a significant effect when we consider only the quasars in the fields studied by Savage and Bolton (1979), but this may be due to the small size of this sample. The length-scale associated with this periodicity is, we believe, consistent with the expected Jeans length just before recombination (Fang and Liu 1981); the detailed z-dependence of the quasar redshift distribution supports a high value of q_0.

When a periodicity in the redshift distribution of quasars was first claimed, it was looked upon as evidence for the non-cosmological interpretation of quasar redshift. The model developed here shows how we can in principle explain the possible origin of a periodicity in *standard* cosmology. The main consequences of this model have been preliminarily verified by several statistical tests. The value of q_0 determined by the remarkableness of the periodicity is larger than 0.5. It further strengthens the conclusion which has been supported recently by evidence on massive neutrinos in the universe and by the redshift-magnitude relations of various quasar subsets. More complete samples, especially of quasar sets centred on several small fields in the sky, are needed for further tests on this model.

Finally, we should mention that in this model galaxies and quasars might form with the same basic redshift distribution; the quasars are merely tracers of the distribution at large z. The periodicity might then be present in the galaxy distribution at smaller z.

REFERENCES

Barnothy, J.M. and Barnothy, M.F., 1976, *Publ. Astron. Soc. Pacific*, **88**, 837.
Basu, D., 1975, *Astrophys. Letters*, **16**, 53.
Burbidge, G.R., 1968, *Astrophys. J. Letters*, **154**, L 41.
Burbidge, G.R. and O'Dell, S.L., 1972, *Astrophys. J.*, **178**, 583.
Cheng, F.H., Kiang, T. and Fang, L.Z., 1981, *Acta Astronomica Sinica*, **22**, 357.
Cowan, C.L., 1969, *Nature*, **224**, 655.
Fang, L.Z. and Liu, Y.Z., 1981, *Acta Astrophysica Sinica*, **291**; *Lett. Nuovo Cimento*, **32**, 129.
Fang, L.Z., Chu, Y.Q., Liu, Y. and Cao, Ch., 1982, *Astron. Astrophys.*, **106**, 287.
Hewitt, A. and Burbidge, G.R., 1980, *Astrophys. J. Suppl.*, **43**, No. 1.
Karlsson, K.G., 1971, *Astron. Astrophys.*, **13**, 333.
— 1977, *Astron. Astrophys.*, **58**, 237.
Plagemann, S.H., Feldman, P.A. and Gribbin, J.R., 1969, *Nature*, **224**, 875.
Qing, C.R. *et al.*, 1981, *Acta Astrophysica Sinica*, **1**, 9.
Roeder, R.C., 1971, *Nature Phys. Sci.*, **233**, 74.
Roeder, R.C. and Dyer, C.C., 1972, *Nature Phys. Sci.*, **235**, 3.
Savage, A. and Bolton, J.G., 1979, *Monthly Notices Roy. Astron. Soc.*, **188**, 599.
Schmidt, M., 1975, *Galaxies and the Universe*, eds. A. Sandage, M. Sandage and J. Kristian.
Wills, D. and Ricklefs, L., 1976, *Monthly Notices Roy. Astron. Soc.*, **175**, 81.

EVIDENCE FROM DEEP RADIO SURVEYS FOR COSMOLOGICAL EVOLUTION

H. VAN DER LAAN and R.A. WINDHORST
Sterrewacht Leiden, The Netherlands

1 - INTRODUCTION

The IAU Symposium no. 74 on Radio Astronomy and Cosmology was held in the new Cavendish Laboratory in Cambridge for a week in August 1976. The published volume (Jauncey 1977) contains some forty papers in which under seven main headings the contributions to cosmology using radio astronomy techniques are presented and interpreted. About half the book deals with the radio galaxy population and its evolution, the subject of the present paper. The other half deals with quasars, with the microwave background and with a variety of topics ranging from galaxy formation to intergalactic Faraday rotation and measures of isotropy. These topics reappear, by and large and often interestingly different, in the present Study Week volume. Some aspects that, strictly speaking, belong to our assignment, such as angular size measurements and bright source identifications, are dealt with in the contributions in this volume by Swarup and by Longair. No completeness of literature coverage or of information retrieval spanning the past five years is attempted. Rather, we set forth the chief developments in radio galaxy population studies and our appreciation of them in Sections 2 to 4. Our own work is reported in Section 5.

2 - RADIO GALAXIES HAVE NO RADIO STANDARDS

Even cursory knowledge of radio galaxies leaves one with serious doubts about their suitability to serve as cosmological probes of any sort.

The members of this population are so varied in power, in size, in morphology, in energy content and life expectancy, that aspirations to define rigid rods or standard candles from the assorted radio characteristics are futile. In Fig. 1 projected linear sizes of four well known objects illustrate one such range. The powers among sources of comparable flux density in a complete, flux-limited sample, also span five orders of magnitude. The situation is made even worse by the peculiar relation between power and energy content of a transparent, homogeneous volume element emitting synchrotron radiation: $E_{total} \propto R^{9/7} L^{4/7}$; here E_{total} is the energy content in relativistic particles and magnetic fields, R is the linear size and L is the synchrotron luminosity. Let us compare two sources with these three characteristic parameters: E_{total}, R and L, with one of those parameters equal for both sources and another one just differing by a factor of ten. The following table then shows the wide range that is implied for the value of the third parameter by the energy equipartition assumption.

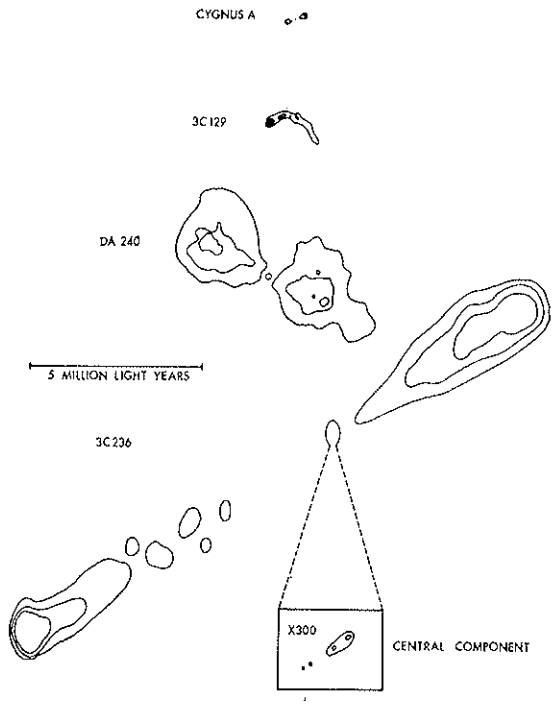


FIG. 1. The typical range of projected linear sizes of radio galaxies.

Parameter	E_{tot}	R	L
	1	10	0.006
Pair ratios	3.7	1	10
	10	6	1

Compact transparent synchrotron sources are much more efficient than diffuse sources. Not only a source engine's integral output but also its spatial distribution determines its luminosity, a factor which may be affected by both environment and the time profile of engine power. Add to this situation the fact that the radio spectra are by and large power laws with minimal information content and it is clear that radio astronomy is unlikely to yield subtle yardsticks usable for the refined art of geometric cosmology. Given the great parameter ranges, large samples with very stable means, medians or percentiles would be required. There is no empirical evidence or astropbysical clue that these exist.

3 - Radio galaxy population characteristics *

3.1 *Intrinsic properties.* Fundamental is tbe radio luminosity function (RLF) at a given epocb. This RLF is the distribution of monochromatic power at a specified radio frequency among all sources per unit volume. It is expressed as $\rho\ (P_\nu, z)$ in units of $Gpc^{-3}\ (\Delta \log P_\nu)^{-1}$. The size distribution among all sources per unit volume is the size function. The spectral index function is the distribution of a two frequency spectral index $\alpha^{\nu 2}_{\nu 1}$ among all sources per unit volume. Note that there may be a correlation between power and/or size and /or spectral index and/or redshift. In principle one wishes to determine the distribution of all sources in a $(P_\nu, R, \alpha^{\nu 2}_{\nu 1}, z)$ parameter space.

3.2 *Population observables.* The first and most famous is the *radio source count* $N(S_\nu)$, the number of sources per unit solid angle per flux density interval ΔS_ν in a sample complete to a specified flux density limit, S_ν^{min}, at a specified observing frequency ν. Another such observable is $n\ (P_\nu, S_\nu)$, the distribution of monochromatic luminosity among the sources in such a flux limited complete sample; similarly there are spectral index distributions, $g^{\nu 2}_{\nu 1}\ (\alpha)$, and angular size distributions, $\phi\ (\theta)$. In order to attempt a cosmic evolution study of source populations, these four dis-

* Unless explicitly stated otherwise, the values $H_0 = 50$, $q_0 = 0$ are used throughout this paper.

tributions can be determined by measurements of observables for all sources in a flux limited complete sample. Then models must be devised for *intrinsic* population properties, the luminosity, spectral index and size functions, at intervals of redshift z. Calculations with these functions and a world model as input can then be performed iteratively until predicted and observed distributions agree within statistical errors.

As starting values for these functions one has sufficient information about bright source samples and some knowledge of the source physics to devise schemes which converge reasonably quickly.

3.3 *Radio source samples.* The most comprehensive source counts currently consist of composite samples each gathered from many surveys. Evidently, given $N(> S_\nu) \propto S^{-3/2}$ and the desire for comparable statistics everywhere on the log S scale, bright source surveys covering $\sim \pi$ steradians can be combined with deeper surveys covering decreasing areas of sky with increasing depth. Careful flux-scale calibrations and corrections for discrimination against large angular sizes have yielded samples at several primary frequencies such as 0.4, 1.4, 2.7 and 5 GHz, consisting of several hundreds to a few thousands of sources that lend themselves for model analysis. These samples are well accounted for and inclusion/exclusion criteria are documented in the literature [e.g. 0.4 GHz (Wall, Pearson and Longair 1980); 1.4 GHz (Katgert 1977; Willis, Oosterbaan and de Ruiter

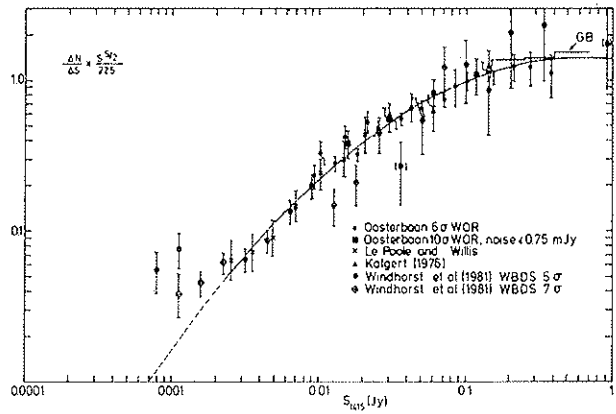


FIG. 2a. The differential 21 cm radio source counts, normalised w.r.t. Euclidean, for various Westerbork Deep Surveys.
Note that the initial steep rise is beyond $S_{1.4} \geq 1$ Jy. (See IAU Symp. No. 74). Strong convergence of the counts is seen in all surveys for 2 mJy $\lesssim S_{1.4} \lesssim$ 100 mJy.

1976; see our fig. 2a); 2.7 GHz (Peacock and Wall 1981); 5 GHz (Kellermann 1980 and references given there)].

From the start the isotropy of the source counts has been striking. It has persisted with all progress in frequency coverage, depth and statistically testable angular scales. Occasional claims to the contrary have always faded as statistics improved. The steep $N(>S) \propto S^{-1.8}$ counts combined with that isotropy enabled Ryle (1959) to decisively reject the so-called Perfect Cosmological Principle on empirical grounds and to claim a large proper density enhancement at z values undetermined at that time, due to the then nearly complete absence of identifications and redshifts.

4 - Modelling methods

4.1 *The basic approach.* One wishes to know the functions defined in 3.1, and particularly their redshift dependence. Ignore for the moment the spectral index and the size function and concentrate on the RLF. Fig. 2b, from a review by Katgert (1980), represents the (P_ν, z) plane and illustrates how the problem can be posed. Radio sources populate this

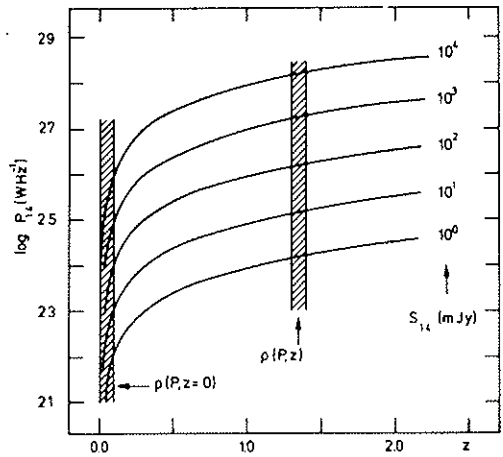


Fig. 2b. The (log P, z) diagram for different values of S_{1412} ($H_o = 100$, $q_0 = 1$), illustrating the relation between radio source counts and epoch dependent luminosity function.
The curved lines show, for a certain flux density level, at which redshift which part of the radio luminosity function is sampled (shown schematically by vertical strips).
For example, because the 21 cm source counts show a maximum excess w.r.t. Euclidean just below 1 Jy and because the evolution turns out be the strongest between about $z \sim 0.3$ and 2 (Fig. 4), this figure shows that the strongest evolution is expected for powers between $\log P_{1.4}/\mathrm{WHz}^{-1} \sim 25.7$ and 27.2.

diagram and it is that population's distribution which is to be established. Given some (P_ν, z) plane population, one can easily obtain the corresponding source counts and luminosity distribution. The differential source counts are obtained by summing over successive curved strips bounded by the appropriate S_ν, and the luminosity distribution by summation of all sources in successive horizontal strips up to the sample's flux limit. The RLF at a given z is now the source distribution in a given vertical strip at that z. Tbe redshift distribution of a complete sample consists simply of the successive sums per vertical strip above the appropriate flux limit. If one uses a standard candle approximation for the absolute magnitude, together with an appropriate SED for the object at issue, then the z-distribution and world model used in the exercise yield a corresponding apparent magnitude distribution to be compared directly with the results of the sample's optical identifications (see Section 5). The world model geometry is of course given in tbe diagram by the trajectories of constant S curves in the (P_ν, z) plane, where with each z a proper volume element is associated, by which each population density is multiplied before summing.

In order to proceed a plausible first-try population of the (P_ν, z) plane must be obtained. This is normally done by deriving a local RLF, $\rho\,(P_\nu, 0)$, from the luminosity distribution, $n\,(P_\nu, S_\nu^0)$, of a bright source sample and its identification content, and by extending this local RLF in tbe z-direction using an evolution function, $E\,(P_\nu, z)$, where

$$\rho\,(P_\nu, z) = \rho\,(P_\nu, 0)\,E\,(P_\nu, z) \tag{1}$$

and $E\,(P_\nu, z)$ now represents the function to be found in some trial and error procedure with appropriate comparison of predicted and empirical distributions in each iteration.

Two reviews by Longair (1971, 1978) and one by Scheuer (1975) show how the method is basically unchanged but increasingly subtle in its practical application. Early developments in source count interpretations led to the distinction of luminosity evolution and density evolution, a terminology which often obscured the issue because the evolution of *individual sources* was entered into the source count discussion prematurely. Strictly speaking, a source count interpretation ought only to refer to *population* parameters, and time scales derived for evolution will only describe the evolution of the radio source population. This is because, except for the optical part of the spectrum, the relevant time scales for an

individual radio galaxy are much smaller than the Hubble time. Pure density evolution then means that $\rho\,(P_\nu, z)/\rho\,(P_\nu, 0)$ is a function of z only and pure luminosity evolution means that the above ratio is a function of both P_ν and z; but:

$$\frac{\int_{P_{\nu,\,\min}}^{\infty} \rho\,(P_\nu, z)\, d\,P_\nu}{\int_{P_{\nu,\,\min}}^{\infty} \rho\,(P_\nu, 0)\, d\,P_\nu} = 1\ .$$

In other words, pure density evolution would show up in Fig. 5 (top) as a vertical displacement of the RLF with z, and pure luminosity evolution as a horizontal shift with P. Both types of evolution could be disentangled if there were a feature in the RLF which one could see moving with z and/or P.

Both conditions are restrictive and physically implausible. It is more likely that the RLF will have a z-dependence in its form and in the integral value over a specified power interval. The evolution is then one in both luminosity and in density, requiring a specification of $E\,(P_\nu, z)$ over the appropriate domains for its satisfactory description.

4.2 *Parametric forms for* $E\,(P_\nu, z)$. Source counts involve only integrals in the (P_ν, z) plane. Hence there is an unlimited number of ways in which to populate that plane and still satisfy the source count constraint. It has been clear since the early sixties that very strong population evolution is required especially for the intrinsically powerful sources, and even the luminosity distribution of an apparently bright source sample will be affected by this population evolution. It is necessary, therefore, in computing the local RLF to take that influence into account (Wall, Pearson and Longair 1978), as well as the geometric factors of the chosen world model. The key relations are:

$$\rho\,(P_\nu, 0) = \frac{n\,(P_\nu, S^0_\nu)}{\int_0^{z\,(P_\nu,\,S^0_\nu)} E\,(P_\nu, z)\cdot dV\,(z)}\ , \tag{2}$$

$$N\,(>S_\nu) = \int_0^{\infty} d\,P_\nu \cdot \int_0^{z\,(S_\nu)} \rho\,(P_\nu, 0)\ \cdot\ E\,(P_\nu, z)\ \cdot\ d\,V\,(z)\ , \tag{3}$$

where $dV\,(z)$ contains all geometric factors appropriate to the chosen world

model; the luminosity distribution $n(P_\nu, S_\nu^0)$ is a data-set used as input, $N(>S_\nu)$ is the result used for comparison to the empirical source counts to provide a goodness-of-fit.

The specification of $E(P_\nu, z)$ is crucial and many forms have been tried, primarily of two mathematically convenient forms, viz.:

$$\left.\begin{array}{l} E(P_\nu, z) \propto (1+z)^{a(P_\nu)} \\ \text{or} \\ \qquad \propto \exp\left[a(P_\nu)\,\tau(z)\right] \\ \qquad = 1 \end{array}\right\} \quad \begin{array}{l} P_\nu > P_\nu^{cr} \\ \\ P_\nu < P_\nu^{cr} \end{array}$$

where $\tau(z)$ is the look-back time, appropriate to the world model. By various specifications of $a(P)$, as well as by the introduction of redshift cut-offs, these basic parametric forms provide a considerable variety of trial functions. Comprehensive attempts at source count fits by these methods were made by Wall, Pearson and Longair (WPL 1978, 1980, 1981) and by Katgert (1977). Some forms exist that reproduce the counts as closely as could be wished within the statistical errors of the data. Even the best fits (e.g. models 4b and 5 in WPL 1977) suffer several defects when their predictions of the apparent magnitude distribution for even bright source samples are compared to empirical data.

The same is true for the amplitude of the evolution function for small ($\sim$ 0.1 to 0.3) z-values or the angular size distributions for modestly weak source samples. As stated in WPL (1980) about parametric representations, "their main role is illustrative and indicates the level of complexity now necessary to explain all the observational data".

4.3 *Robertson's one dimensional free form for* $E(P_\nu, z)$. An important innovation was introduced by Robertson (1978, 1980) with the concept of free-form evolution functions. Here the parametric prescriptions for $E(P_\nu, z)$ are dropped. Using the luminosity distribution as before to compute $\rho(P_\nu, 0)$, the local RLF, and the source counts as the distribution to be precisely represented by the appropriate integral in the (P_ν, z) plane, Robertson computed a piecewise continuous function $E(z)$ by an iterative procedure. The method is only free-form in one dimension, because it prescribes the P-dependence of $E(P_\nu, z)$ using two parameters as shown in Fig. 3.

In his second paper Robertson takes the additional step of substituting expression (2) in equation (3), so that instead of a two step iteration

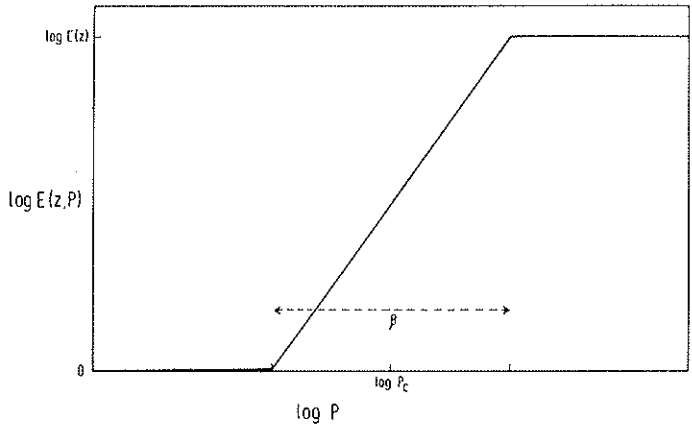


FIG. 3. Robertson's (1980) assumed dependence of the evolution function E (log P, z) on log P. A gradual transition is used around a value of log P_c, at which log E has increased to one half of the value reached at very high powers. The transition luminosity range is taken to be β decades wide (in Robertson's paper β = 1 and log P_c (408 MHz)/WHz^{-1} = 26.3).

scheme where first the local RLF is computed and then the source count fit is tested with the same trial E (P_ν, z), the optimization is done numerically with direct use of all available luminosity distribution information. This avoids several pitfalls of RLF calibration and improves the assessment of the propagation of luminosity distribution errors in the E (z) determinations.

Robertson's best result in fitting the 408 MHz counts, termed the standard model with $P_c = 2.0 \times 10^{26}$ WHz^{-1} and β = 1, is shown in Fig. 4, together with several parametric model fits and with our own direct determination of the population evolution as reported in Section 5. There we also return to the reason for the Robertson standard model's success: a fortuitous guess in choosing the P dependence of the evolution function, shown in Figure 3. Important features of the standard model are the lack of evolution at $z \lesssim 0.25$ and the steepness of the rise in E (z) from 0.3 to 1.0. Beyond $z = 1$ the fit has little statistical significance, except perhaps for quasars.

An additional advantage of Robertson's (1980) method is that the best E (P_ν, z) determined from the luminosity distribution and source count data together can then be used to compute the local RLF, through expression (2).

4.4 *Peacock and Gull's generalized determination of* E (P_ν, z). After

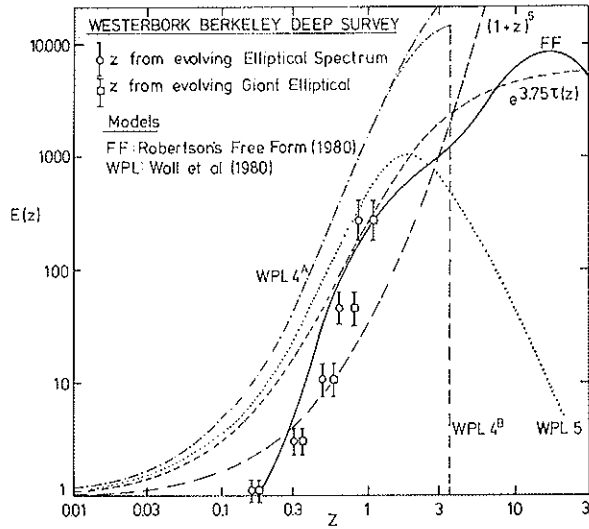


FIG. 4. Various evolution functions E (z) for log $P_{1.4}/WHz^{-1} = 26.2$. Shown are the classical power law and exponential models of Schmidt (1972), the models of Wall, Pearson and Longair (1980), and Robertson's (1980) free-form model. The data points are preliminary points from the Leiden-Berkeley Deep Survey, where evolving standard candles have been used to derive redshifts from magnitudes ($< M_F > = -21.^m5$ for ellipticals, and $< M_F > = -22.^m3$ for evolving giant ellipticals, so the highest two redshift bins could be affected by the Malmquist bias, in the sense that the actual sampled redshift is 0.1 to 0.2 higher than the stated one, which would shift these points to the right).
Power and exponential laws do not provide very good fits, unless perhaps for very steep slopes. The WPL models show enhancement factors too large for $z \lesssim 0.3$. The general shape of Robertson's free-form evolution function is remarkably similar to our Leiden-Berkely Deep Survey data points.

Robertson's innovative steps it is natural to improve his method in two ways. Firstly, to eliminate the (parametric) prescription of the evolution function's P-dependence; secondly, to make better use of the available data, especially by giving due weight to information in spectral index distribution, to source counts and to luminosity distributions at several frequencies. Peacock and Gull (1981) have made good progress in these directions. They utilize source counts at four frequencies (408, 1400, 2700 and 5000 MHz). At the two higher frequencies the counts are available for steep and flat spectrum sources separately. To relate their model (P_ν, z) plane populations to these counts they not only separate steep and flat spectrum sources but also use the spectral index-luminosity correlation of the former. However, it is at the moment not precisely

clear whether a spectral index-luminosity or a possible spectral-index-redshift relation is the basic one (Katgert-Merkeljin *et al.* 1980). Recall that at optical wavelengtbs an exactly analogous problem occurs. In addition some weigbt is given to incomplete luminosity distributions at intermediate flux levels, to $< V/V_{max} >$ values of quasar samples and to identification fractions of a few deep radio samples. Series expansion metbods are used to generate model populations. In that sense Peacock and Gull's method is not truly free-form either, although it allows for variation of $E(P_\nu, z)$ in two dimensions. Their polynomial approximation of $E(P_\nu, z)$ has to replace a truly nonparametric array of delta functions, because of the paucity of data (see Fig. 5), although the data set included is already enormous. The goodness of fit is established by evaluations over the whole sum of various data bins, where in the iteration procedure a certain uniform measure of fitting quality is used to prevent some data of relatively little weight from heing neglected due to optimum fits of larger data sets. Friedmann models with $q_0 = 0.5$ and $q_0 = 0$ are combined with no redshift cut-off and a cut-off at $z = 5$ to produce four models each for steep and for flat spectrum sources respectively. The series expansion used is of the form

$$\log \rho = \sum_{i=0}^{n} \sum_{j=0}^{n-i} A_{ij} (\log P)^i [\log (1 + z)]^j$$

and is developed to quartic order (15 coefficients) and cubic order (10 coefficients) for steep- and flat-spectrum sources respectively. A total of 151 data bins was utilized. The authors use the variations among five satisfactorily fitted models for each of the two q_0 values and each of the two spectral classes to establish the domains in the (P_ν, z) plane where the RLF is well-established. Two results stand out: (i) the RLF is constrained by the data to within a factor ten in only a modest fraction of the (P, z) plane over the ranges $z = 0.01$ to 2 and $\log P_{2.7} = 22$ to 30 (see Figure 5); (ii) the steep spectrum and the flat spectrum source populations behave similarly, i.e. both evolve modestly or not at all at low powers, but similarly strongly at high powers (Fig. 5). The method does not so far allow any definite conclusions about the requirement of a redshift cut off or about world geometry.

4.5 *Conclusions and suggestions.* Peacock and Gull's (1981) use of $< V/V_{max} >$ data in finding a good fit is important. It complements the

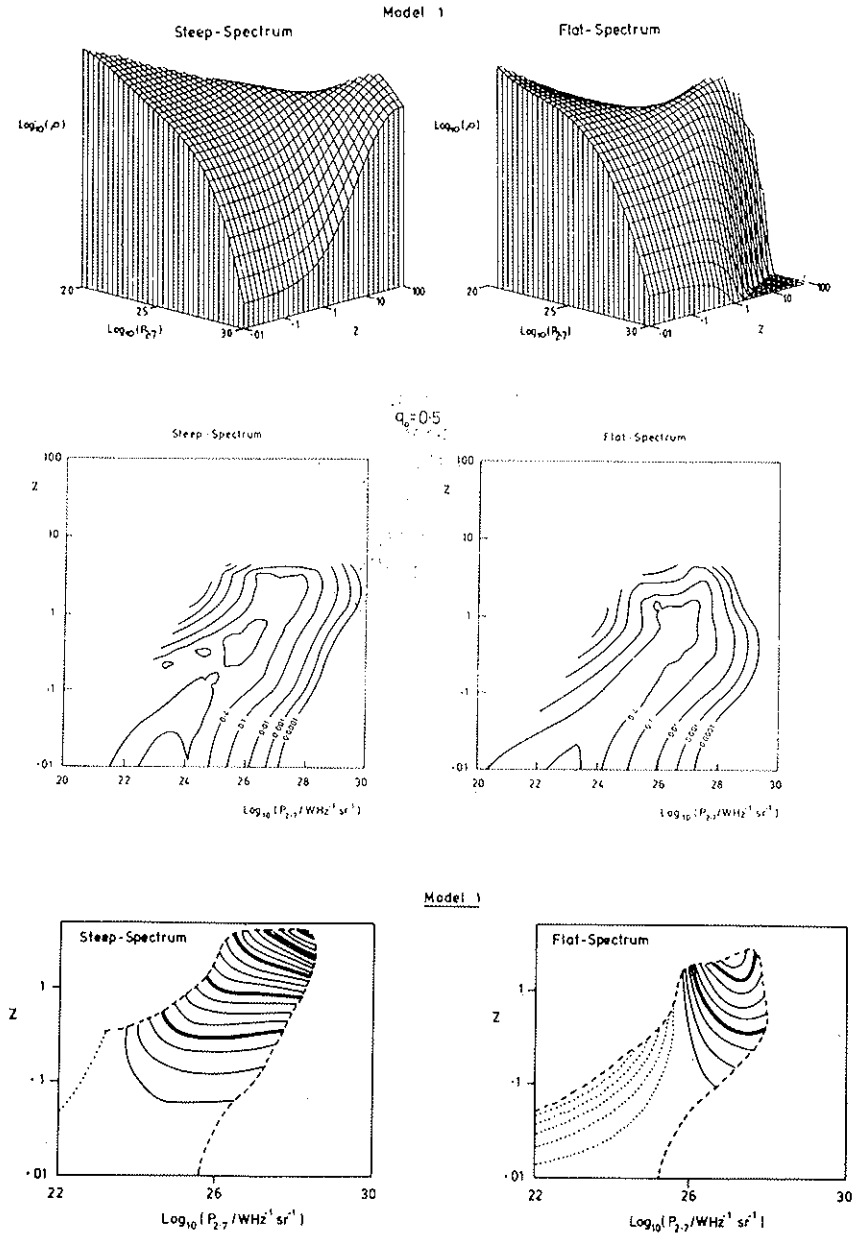


FIG. 5. Peacock and Gull (1981) representation of the evolution functions for steep and flat-spectrum classes respectively, for $H_o = 50$, $q_0 = 0.5$ and no redshift cut-off: (Top.) A three dimensional representation of ρ (log $P_{2.7}$, z). Note that the evolution is similar for both flat and steep spectrum classes, except for $z \geq 1$ at the higher powers. (Middle) Contour plot in the ($P_{2.7}$, z) plane of the reliability parameter, R, which is the ratio of the smallest value of ρ (log P, z) to the largest one, found amongst all models checked. The data do constrain the models only in an acceptable way for R close to unity. Note the evolution functions are called "well defined", if $R > 0.1$, i.e. if the model uncertainties are not larger than a factor of 10. (Bottom) Contour plot of the evolution function E ($P_{2.7}$, z). The enhancement factors are represented by heavy lines for log E = 1, 2, 3, 4 and by light contours for log E = 0.25, 0.50, 0.75, etc. Note the evolution function is only plotted for values of the reliability parameter $R > 0.1$.

source counts, emphasizing the gradients $\frac{\partial E(P_\nu, z)}{\partial z}$ rather than the amplitude of the evolution function. The whole method needs further clarification particularly with respect to the criteria for defining an optimal fit to several and diverse data sets. Finally, the reliable determination of $E(P_\nu, z)$ can be vastly improved over a larger domain in the (P_ν, z) plane by using data sets of maximum dynamic range in both dimensions. These are the very deep combined radio-optical survey data such as are discussed in the next section. It is clearly important to extend such work to higher radio frequencies as well. In our view the method developed by Peacock and Gull is the best modelling method in principle, if the polynomial approximation could be replaced by an array of δ functions. Given improvements of the kind just suggested, it does most justice to data obtained at great effort and expense. The initial results the method has yielded are not significantly better than Robertson's (1980), largely because the P_ν-dependence of the evolution function is not unlike the guess of Figure 3. Both methods demonstrate unequivocally that the radio source population evolves very steeply in proper density at luminosities significantly above 5×10^{26} WHz^{-1} at 408 MHz.

5 - Radio-optical surveys

Since IAU Symposium No. 74 in 1976, radio surveys with the increasingly sensitive Synthesis Radio Telescope at Westerbork have continued. Extensive catalogues at 1412 MHz will be published in *Astronomy and Astrophysics* (Main Journal and Supplements) in 1982 and 1983. These data can, naturally, be used by the methods discussed in the previous sections to contribute towards the determination of the evolution function, providing input and constraints for model fitting procedures. From the radio point of view these surveys will provide the 21 cm source counts down to $\sim$ 600 μJy from the Westerbork Deep Surveys and down to $\sim$ 80 μJy from a recent deep VLA Survey. Furthermore these surveys will provide information about the angular-size flux density relation, $\theta(S)$, measured with resolutions of $\sim$ 10″ to 22″, while their flux density range extends from the above limits up to just below 1 Jy.

From the optical point of view very deep photographic plates in several passbands of these Westerbork Deep Survey fields will yield numerous identifications, whose magnitude distributions, $N(m)$, will yield

equally valuable information as the redshift and luminosity distributions that Peacock and Gull (1981) used in their analysis.

Here we briefly report the progress of a more direct determination of the function, $E(P_\nu, z)$ at $\nu = 1.4$ GHz, using combined radio-optical data. A review of such work done at Leiden in the 1975-1978 period was published by Katgert, de Ruiter and van der Laan (1979). The use of these data for studying the colour evolution of the stellar populations in remote galaxies is discussed in our previous contribution in this volume. Further details can he found in Windhorst *et al.* (1981).

5.1 *The data.* Two programs, one called the Westerbork-Einstein Deep Survey and the other the Leiden-Berkeley Deep Survey, have been recently completed and are being prepared for publication. All the fields of view at issue here have been covered by both Westerbork radio exposures and KPNO 4 m plates. The quality of the data, in terms of sample flux limits, angular resolution, positional accuracy, reliability and completeness are listed in Table 1. So is the quantity of data in terms of numbers of fields, numbers of pass-bands, of sources detected, of sources identified.

5.2 *The z-dependent RLF.* Combined radio-optical surveys yield, after subtle calibration and correction procedures, radio flux density-optical magnitude tables (bivariate flux density tables). These $N(S_\nu, m_{opt})$ distributions may, given a world model, be converted to (P_ν, z) distributions if each identified object in an (S_ν, m_{opt}) bin can be assigned an absolute magnitude and an SED (e.g., Katgert 1977, de Ruiter 1978). As we have argued elsewhere in this volume, there are good reasons to regard the radio galaxies identified in our surveys as evolving luminous ellipticals whose absolute magnitudes have very small dispersion and whose z-dependent colour can be modelled in stellar population evolution scenarios (Bruzual 1981). Using our apparent magnitude and colour data in combination with, as a first order guess, the "standard candle" assumption, $M_F = -22.^m3$, for $q_0 = 0$, $H_0 = 100$, and Bruzual's (1981) $\mu = 0.7$ model SED, every interval Δm_F corresponds to a bin Δz, and thus every flux density, S_ν, transforms to a monochromatic power, P_ν. In other words, given the (radio flux density dependent) survey area, every bin of the bivariate flux density distribution, $N(S_\nu, m_{opt})$, transforms into one point of the epoch dependent radio luminosity function, $\rho(P_\nu, z)$, under the above assumptions. These two assumptions might look less justified than they actually are. In order to check the first assumption, which presumes radio galaxies to be slightly evolving standard candles, we plotted the empirical Hubble diagram in V

TABLE 1 - *Summary of the current Leiden-Berkeley Deep Survey data.*

a. *Quality*	*Radio*	*Optical*
Instrument	WSRT 3 km	KPNO 4^m Mayall
Observing mode	10 MHz/1250 KHz line receiver	Photoelectrically calibrated broadband photographic photometry
Wavelength	21 cm (50 cm in 1982)	3000-9000 Å (1.25-2.2μ UKIRT)
Resolution	21 cm: 12'' in RA 50 cm: 28'' in RA	0.''5 - 1.''0
Positional accuracy	0.''4 - 1.''4	~ 0.''2
Field of view	0.85 deg Ø (QPBW)	0.85 deg Ø
Flux limits	21 cm: 580 μJy (5 σ) 50 cm: 1.76 mJy (5 σ)	0.4 - 3 μJy or: $U_{3600} \lesssim 24.^m0$ $J_{4650} \lesssim 24.^m0$ $F_{6100} \lesssim 23.^m0$ $N_{8000} \lesssim 22.^m0$
b. *Quantity*	*Radio*	*Optical*
Total area	5.90 deg^2	5.90 deg^2
No. of fields (x obs. time)	21 cm: 9 ($\times$ 12^h) 50 cm: 6 ($\times$ 12^h)	9 ($\times$ 8 à 16 plates) ($\times$ 1 à 2^h)
Objects/field	21 cm: 50-70 50 cm: ~ 200-250	~ 15000 (dependent on b^{II}) out of which 10000 are galaxies
Complete sample	297 radio sources at 21 cm, 146 identifications (49%)	
Contamination	5-10% per field, on average 6.5% of all radio sources	
Reliability	83% of all announced identifications are real	
Completeness	92% of all potential identifications are announced	

(Fig. 6) for first ranked cluster galaxies (Sandage 1973) and 3CR radio galaxies defined at 178 MHz (Laing *et al.* 1978, Gunn *et al.* 1981) and 3C and 4C radio galaxies, found at 2.7 GHz (Peacock and Wall 1981). We also plotted eight radio galaxies from the Leiden-Berkeley Deep Survey with spectroscopic redshifts (Kron 1981, Spinrad 1981), which with two exceptions all fall within ~ $0.^m5$ on the above empirical Hubble diagram. One of the two underluminous radio galaxies is a double galaxy and placed between brackets. The drawn curves are for $M_F = -22.^m3$ and $H_0 = 100$,

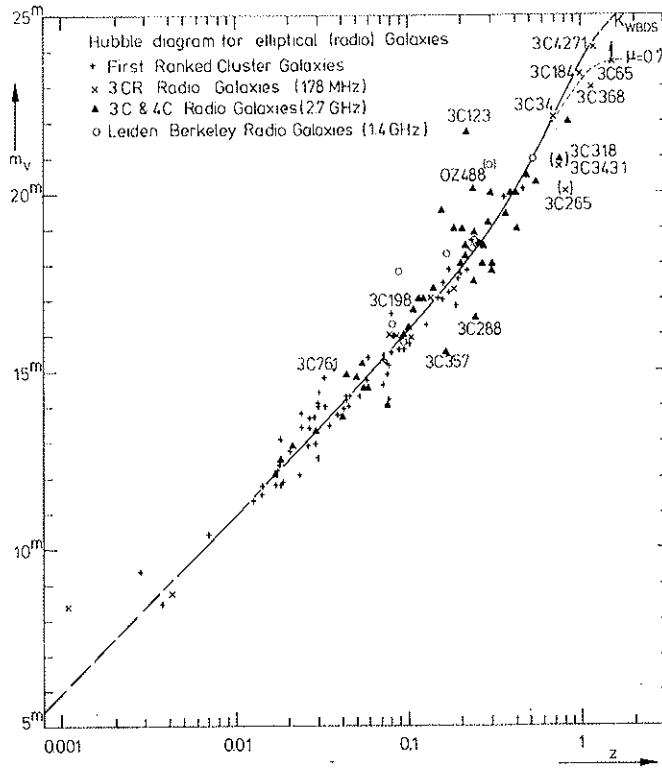


FIG. 6. The Hubble diagram in V for elliptical (radio) galaxies from various (radio) surveys. Shown are FRC galaxy data from Sandage (1973), Laing *et al.* (1978), Gunn *et al.* (1981), Peacock and Wall (1981), Kron (1981), Spinrad (1981).
No corrections have been applied to the V magnitudes, apart from aperture corrections as far as data were available.
The two lines show the Hubble relation as predicted for the V band by the Bruzual (1981) $\mu = 0.7$ model ($H_0 = 100$), and by an earlier empirical determination of the K-correction from Leiden-Berkeley Deep Survey data.
With a few exceptions all radio galaxies are represented well (within $0.^m7$) by the two predictions, which assume $M_V = -21.^m5$. A double galaxy from the LBDS is placed between brackets, as well as a few N galaxies from the 3CR.

$q_0 = 0$. These sparse statistics already show that our mJy radio galaxies have apparently more or less the same narrow absolute magnitude distribution as the brighter radio galaxies. One might argue that the few deviating objects, as well as those from other surveys, are due to an incoming population of radio galaxies that are optically $\sim 1.^m0$ underluminous, as one would indeed expect from the local bivariate radio

luminosity function for elliptical galaxies of Auriemma *et al.* (1977). In the future we will incorporate the structure of the local bivariate LF in determining the epoch dependent RLF. A major worry should be the effects of the Malmquist bias, which becomes significant for $z \gtrsim 0.6$, beyond which our sample is no longer distance limited for an Auriemma *et al.* (1977) type local bivariate LF. However, it should not be a problem to correct for the Malmquist bias provided the functional form or shape (not the amplitude!) of the bivariate LF does not change with epoch.

In order to check the second assumption of known evolving SED, not necessarily the consequence of a valid standard candle assumption, we bootstrapped a model-independent evolving SED from the observed colour magnitude distributions (see Fig. 1 in our previous paper in this volume) and available spectroscopic redshifts. This was done completely independent from and even without the knowledge of the $\mu = 0.7$ model. The resulting non parametric K-correction in V is shown in Fig. 6 as K_{WBDS} and is, within the photometric errors, identical with Bruzual's (1981) $\mu = 0.7$ model, at least for $z \lesssim 0.8$, beyond which both this procedure and Bruzual's models become uncertain due to the lack of calibration. Fig. 6 shows that both evolving SED's with $M_F = -22.^m3$ give, with very few exceptions, a good fit through all available redshifts out to $z \sim 1$. This means that the actual SED evolution of the parent radio galaxy population is not very much unlike the $\mu = 0.7$ model for $H_0 = 100$, $q_0 = 0$. This gives us good confidence that we are indeed dealing with one parent population of evolving optically luminous elliptical radio galaxies, with a narrow ($\lesssim 1.^m0$) absolute magnitude distribution, at least out to the redshifts where the sample becomes magnitude limited. In the above consideration second order effects have been ignored, like a possible relation between optical luminosity and SED (and hence K-corrections; de Goia-Eastwood and Grasdalen 1980).

In Figure 7 the RLF constructed on this basis for six bins in z is plotted, together with the local RLF as determined by Auriemma *et al.* (1977). The plot is based upon 303 radio galaxy identifications and binning in two dimensions means poor statistics per bin.

This rather more direct determination of $\rho(P_\nu, z)$ may be compared with the model fits, particularly of the free-form variety, discussed in the previous sections. Note that the proper density for $P_{1.4} \sim 10^{26}$ WHz^{-1} per z-bin results in the points plotted in Figure 4. The proper density evolves with z for $z \gtrsim 0.34$ and for $P_{1.4} > 10^{24.5}$ (WHz^{-1}). To first order the density enhancement, reaching values ~ 100 at $z \simeq 0.83$, is independent of $P_{1.4}$,

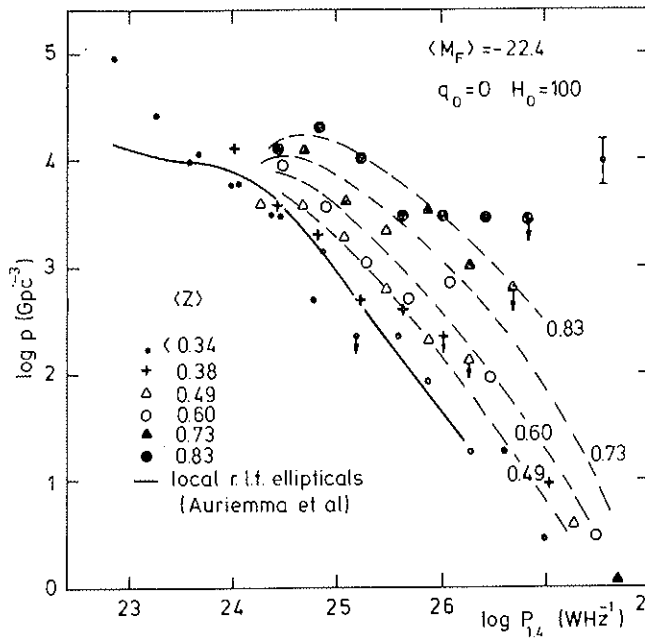


FIG. 7. The epoch dependent 21 cm radio luminosity function ρ (log P, z) for radio galaxies as determined directly from the bivariate flux density distribution N (log S, *m*) of various surveys. Surveys included are the 3RD Westerbork Survey, the Westerbork Background Survey (for references see Katgert *et al.* 1979), the 3CR and parts of the Leiden Berkeley Deep Survey (Windhorst *et al.* 1981). The redshifts are derived from magnitudes, assuming $< M_F > = -22.^m4$, Ho = 100, and the empirical K-correction derived from the (J-F) vs F color magnitude diagram (see also our previous contribution in this volume).
The z estimates of the two highest redshift bins can be affected by the Malmquist bias. Strong evolution is seen for $z \gtrsim 0.4$ everywhere above the break in the LF, more or less independent of radio power.

as presumed by Robertson (1980, cf. Figure 3). Although the data contained hints for a second order P dependence in E (P, z), larger samples are required to explore this possibility. As in the results discussed in the previous sections, based on completely independent data, the population evolution is unmistakable, with a dramatic proper density gradient in z.

5.3 *Prospects and questions.* The first requirement for the further development of this direct determination of the evolution function is confirmation of our distance estimates. To this end direct spectroscopic redshifts must be determined for a number of radio galaxies in our sample. Efforts to that end are being undertaken with the KPNO Cryogenic Camera

and with the IIDS at Lick and at McDonald. This will then serve to calibrate our photometric distance estimates.

Another independent way of checking the distance estimates and the empirically derived SED evolution is the addition of long baseline IR photometry, which was done already for a dozen radio galaxies in H (1.25 μ), J (1.65 μ) and K (2.2 μ) at UKIRT. Because an elliptical SED peaks just beyond 1 μ, the $z \lesssim 1$ radio galaxies should have strongly negative IR K-corrections, which makes them fairly easily detectable for sensitive IR photometers. The colours of a typical $V \sim 23.^m0$ radio galaxy in our sample turn out to be $V\text{-}K \sim 6.^m5$, $H\text{-}K \sim 0.^m75$, $J\text{-}K \sim 1.^m5$, which further supports the hypothesis that the $z \sim 1$ radio galaxies are evolving ellipticals (see also Longair's contribution in this volume). The second requirement is enlargement of the identified sample. We are exploring the population distribution on the (P_ν, z) plane over a considerable range in both parameters and significant statistical weight per bin requires very large samples (cf. Fig. 5). Current Leiden programs will already improve the statistical weight of the data quoted in this progress report by a factor of 3. Thirdly, we require data on non-radio galaxies at the same distance as the identified radio galaxies, particularly their relative colours. Such a program is proceding, using the 4 m-plate material on which this discussion is based (Katgert, private communication). In this manner the question whether the proper density evolution and the colour evolution are merely correlated phenomena among elliptical galaxies or are causally related could be answered. A considerable fraction ($\sim 30\%$) of the identifications lie in clusters where several galaxies of comparable brightness can be measured photometrically. In the longer term, surveys at higher frequencies may be carried out to put greater emphasis on the non-steep spectrum population, i.e. the sources exhibiting current activity, largely quasars. As discussed by Peacock and Gull (1981), the relationship of the steep and flat spectra populations is of particular interest for taking the step from a numerical-spatial interpretation of the source counts to an astrophysical account. (It should be stressed, however, that just like in the optical, the radio K-correction is strongly dependent on α and quite different for the flat and steep spectruun population. Intrinsic population properties should be disentangled from redshift effects and/or evolution). Particularly interesting initial exploratory accounts of the astrophysical mechanism behind the cosmological evolution were published by Grueff and Vigotti (1977) and Rowan-Robinson (1977).

Theoretically the present results pose the question why the present

epoch ($0 < z \lesssim 0.3$) is such a quiet one, while the epoch $0.3 < z \lesssim 1$ was hyperactive. What is the reason for the precipitous proper density decrease since $z \sim 1$? Is this related to cluster evolution and merging rates, as suggested by Roos (1981), or is it determined by the exhaustion of fuel supplies for individual massive black holes in singly evolving galaxies? The spectroscopic calibration of our photometric distance estimates could also show whether the radio galaxy phenomenon at $z > 0.5$ is still uniquely related to the giant elliptical class of stellar systems or whether the parent population is larger, distributed over a greater range of galaxy types and absolute magnitudes.

A new phase in the exploration of active galaxy populations has begun with well matched capabilities in optical and radio astronomy. The spectacular energetics of the individual radio galaxies and quasars and the equally striking evolution in their population characteristics spur vigorous exploration of this fascinating domain in astrophysical cosmology.

REFERENCES

Auriemma, C., Perola, C.G., Ekers, R., Fanti, R., Lari, C., Jaffe, W.J. and Ulrich, M.H., 1977, *Astron. and Astrophys.*, **54**, 41.

Bruzual, G.A., 1981, Ph. D. Thesis, University of California, Berkeley.

de Goia-Eastwood, K. and Grasdalen, G.L., 1980, *Ap. J.*, **239**, L 1.

Grueff, G. and Vigotti, M., 1977, *Astron. and Astrophys.*, **54**, 475.

Gunn, J.E., Hoessel, J.G., Westphal, J.A., Perryman, M.A.C. and Longair, M.S., 1981, *MNRAS*, **194**, 111.

Jauncey, D.L., (Editor), 1977, Radio Astronomy and Cosmology (IAU Symp. nr. 74), Reidel Dordrecht.

Katgert, P., 1977, Ph. D. Thesis, University of Leiden.

— 1980, X-Ray Astronomy, p. 253, R. Giacconi and G. Setti (Eds.), Reidel, Dordrecht.

Katgert, P., de Ruiter, H.R. and van der Laan, H., 1979, *Nature*, **280**, 20.

Katgert-Merkelijn, J., Lari, C. and Padrielli, L., 1980, *Astron. and Astrophys. Suppl. Series*, **40**, 91.

Kellermann, K.I., 1980, *Physica Scripta*, **21**, 664.

Kron, R.G., 1981, private communication.

Laing, R.A., Longair, M.S., Riley, J.M., Kibblewhite, E.J. and Gunn, J.E., 1978, *MNRAS*, **184**, 149.

Longair, M.S., 1971, *Reports on Progress in Physics*, **34**, 1125.

— 1978, in *Observational Cosmology* (Maeder, A., Martinet, L. and Tammann, G. Eds.), Geneva Observatory.

Peacock, J.A. and Wall, J.V., 1981, *MNRAS*, **194**, 33.

Peacock, J.A. and Gull, S.F., 1981, *MNRAS*, **196**, 611.

Robertson, J.G., 1978, *MNRAS*, **182**, 617.

— 1980, *MNRAS*, **190**, 143.

Roos, N., 1981, Ph. D. Thesis, University of Leiden.

Rowan-Robinson, M., 1977, *Ap. J.*, **213**, 635.

Ryle, M., 1959, in *Paris Symposium on Radio Astronomy* (IAU Symp. no. 9, R.N. Bracewell, Ed.), Stanford Univ. Press., p. 523.

de Ruiter, H.R., 1978, Ph. D. Thesis, University of Leiden.

Sandage, A., 1973, *Ap. J.*, **183**, 731.

Scheuer, P.A.G., 1975, in *Galaxies and the Universe*, vol. 9, Stars and Stellar Systems, p. 725 (Sandage, A., Sandage, M. and Kristian, J., Eds.), University of Chicago Press.

Schmidt, M., 1972, *Ap. J.*, **176**, 289 and 303.

Spinrad, H., 1981, private communication.

Wall, J.V., Pearson, T.J. and Longair, M.S., 1978, in *Radio Astronomy and Cosmology* (IAU Symp., no. 74, Jauncey, D.L. Ed.), Reidel, Dordrecht, p. 269.

— 1980, *MNRAS*, **193**, 683.

— 1981, *MNRAS*, **196**, 597.

Willis, A.G., Oosterbaan, C.E. and de Ruiter, H.R., 1976, *Astron. and Astrophys. Suppl.*, **25**, 453.

Windhorst, R.A., Kron, R.G., Koo, D.C., Katgert, P., 1981, in *Extragalactic Radio Sources* (IAU Symp. no. 97, Heeschen, D.S. and Wade, C.M. Eds.), Reidel, Dordrecht.

DISCUSSION

SWARUP

Do the present radio data indicate any significant support for a redshift cut-off at about 4 or 5 in the evolution function of the radio luminosity function?

VAN DER LAAN

Yes they do. At 5 GHz the fraction of flat-spectrum sources decreases rapidly with decreasing flux density. This is likely due to a cut-off in the quasar population at high luminosity and high z. The luminosity function and the spectral index distribution are not well enough known yet to locate the cut-off in the (P, z) plane.

REES

The z-dependence of radio source populations depends, obviously, both on the z-dependence of the central power supply in active nuclei, and also on the way in which the gaseous environment depends on epoch. I wonder if you could comment on this in relation to the different evolution between flat and steep-spectrum sources and differences between the inferred radio evolution and the *optical* evolution that Maarten Schmidt described. Regarding tbe extended sources I would like to emphasize that we do not yet know how the external gas density depends on redshift. If the gas were homogeneous, its density would have gone roughly as $(1 + z)^3$; but the gas has gradually been condensing into cluster and galactic halos, so the relevant density of the gas now (in clusters, etc.) may actually be *higher* than the gas density anywhere in the intergalactic medium at $z = 2$.

VAN DER LAAN

The evolution in proper density is so dramatic that one suspects the engine's *fuelling rate* to be the prime epoch-dependent process. Tbis in turn must depend on the state of galaxy evolution, e.g., merging rates in clusters and groups. For flat-spectrum sources the issue is further obscured hy the relative ease with

which the radio waves are absorbed in the central regions of a galaxy. As for the direct circumgalactic gas density, I agree that it is likely to be much more complex than a simple $(1 + z)^3$ scheme.

LONGAIR

I believe that the answers to Martin Rees' questions will be found by observation rather than by theory. Present understanding of radio galaxies, quasars and active nuclei is not yet at a stage at which any definite statement can be made about the relation between evolution of different types of objects with cosmological epoch. The hope is that, by utilizing all of the observing facilities which are becoming available, we will be able to determine the relation between the optical, radio and X-ray properties. For example, we would like to know the environment for different classes of objects. Is the gas responsible for the confinement of the extended sources intergalactic gas in clusters, as it is associated with the giant elliptical galaxy itself? What are the clustering properties of galaxies about different types of active nuclei? I believe that within the next ten years we will be able to obtain good answers from observation to many such important questions.

SOME ASPECTS OF THE COSMOLOGICAL EVOLUTION OF EXTRAGALACTIC RADIO SOURCES

M.S. LONGAIR
Royal Observatory
Blackford Hill, Edinburgh

I will mention briefly four topics on recent data and interpretation related to the cosmological evolution of extragalactic radio sources which complement the comprehensive review by Professor van der Laan.

1 - The Identification of 3CR Radio Sources

For many years, several of us have been concentrating upon the identification of complete samples of bright radio sources. I will comment in particular on the sources in the 3CR catalogue, which includes the brightest sources in the sky at the low radio frequency of 178 MHz. Much attention has been devoted to a "complete" sub-sample of about 166 sources which lie in regions away from the Galactic plane and at declinations $\delta \geqslant 10^0$. Specifically, the selection criteria were $S_{178} \geqslant 10$ Jy, $|\, b \,| \geqslant 10^0$. Identification for most of these sources have been claimed (e.g. Laing *et al.* 1978, Gunn *et al.* 1981) but there has remained some uncertainty about the reliability of these identifications, because many of them are extended without the presence of a central radio-core coincident with the proposed identification.

This question has been addressed in a recent analysis by Robert Laing, Julia Riley and myself (1982) for the radio sources in the "166" 3CR sample. Part of that analysis involves a discussion of the actual completeness of the sample. As is expected the sample misses some sources of large

angular size and confusion can result in the inclusion of sources which in fact have flux densities less than 10 Jy. In the original definition of the 166 sample by Riley, Jenkins and Longair (see Jenkins, Pooley and Riley 1977) each source had heen scrutinised using data available in 1976. An improved analysis is included in the above paper resulting in a revised sample of 173 sources.

Our analysis of the reliahility of the identifications was performed in three stages. First, all compact sources, $\theta \lesssim 2$ arcsec, can be identified which possess a compact radio-core coincident with an optical object within tions coincide within about 1 arcsec. In addition, all extended sources, which possess a compact raido-core coincident with an optical ohject within 1 arcsec, are certain identifications. Second, we have investigated the reliability of those identifications with radio sources with extended radio structure. Those of Fanaroff-Riley class I sources are all with bright galaxies, $V \leqslant 19$, and are certain because of the presence of a radio-core or because of the presence of strong emission lines in the optical spectrum of the galaxy. Third, we have discussed those identifications with extended double radio structure. As a first step, we have noted the location of the identification with respect to the outer hot spots of the doubles for all those certain identifications with $V < 19$ in which there is a radio-core. This distribution is shown in Figure 1a in which it can be seen that generally the identification does not lie precisely at the centre of the double structure but can be significantly displaced both along and perpendicular to the axis of the source. However, in all but three cases, the identifications lie within a circle of radius 0.2 times the angular separation θ of the outer radio peaks which is shown on the diagram. The same diagram has been plotted for all those double sources with $V < 19$ which do not possess a radio-core (Figure 1b) and it can be seen that all lie within the 0.2 θ circle. Finally, in Figure 1c the same criterion has been applied to all the identifications with faint objects, all of which are faint galaxies with $V > 19$ and which do not possess radio-cores. It can be seen that most of the faint identifications lie within the circle and in only a few cases is the identification likely to be wrong (e.g. 3C68.2,437). A statistical analysis of the probability of galaxies falling at random within the 0.2 θ circle for all sources indicates that at most about 1 identification could be due to chance. Of the 173 sources in the revised sample at most 3.5% are unidentified.

Thus, we have considerable confidence in even the faintest identifications which have been proposed. The importance of this result is that the

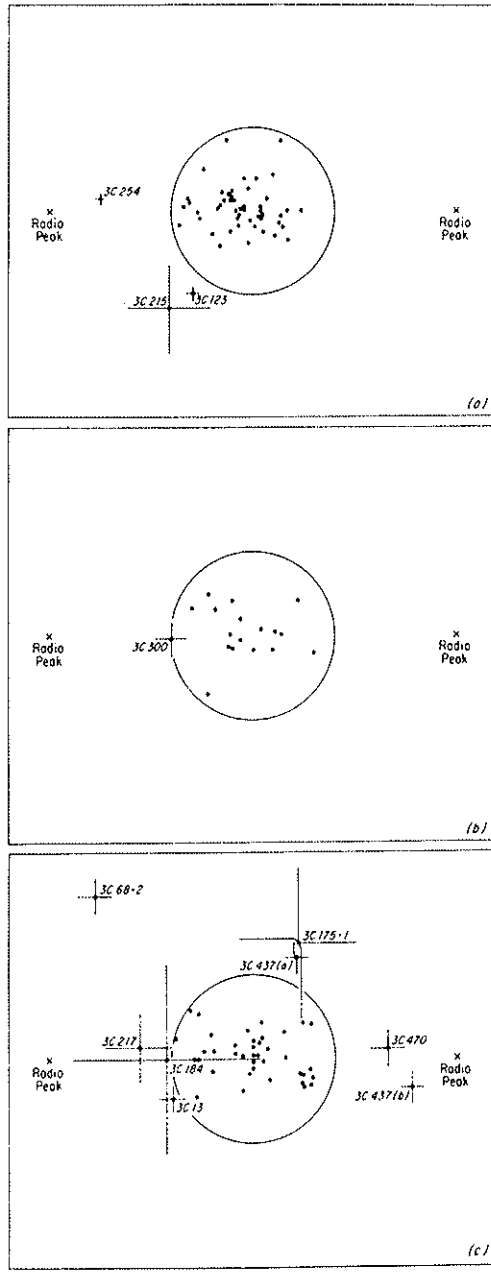


FIG. 1. Diagrams illustrating the location of the optical identification of double radio sources relative to the hot-spots at the leading edge of the double radio structure. The circle shown has radius 0.2 θ where θ is the separation of the hot-spots. Identifications lying outside the circle are named individually with error bars showing the uncertainty in the optical-radio position. (a) All certain identifications having $V < 19$ with radio-cores. (b) All identifications, $V < 19$, which do not possess radio-cores. (c) All faint identifications, $V < 19$, which do not possess radio-cores.

anomalies of the steep source counts, which lead to the inference of strong cosmological evolution, are associated with identified radio sources, about 25.4% of them with quasars and 71.1% with radio galaxies.

2 - The V/V_{max} Test for 3CR Radio Galaxies and Quasars

Using these data, we have performed a standard V/V_{max} analysis of the spatial distribution of radio galaxies and quasars in the 3CR sample. A sample of these results is shown in Table 1. It can be seen that those radio galaxies and quasars which are powerful radio sources ($P_{178} > 10^{26}$ W Hz^{-1} sr^{-1}) exhibit strong cosmological evolution in the sense that they have values of V/V_{max} significantly greater than the mean value of 0.5 expected for a uniform distribution. It is noteworthy that the evolution exhibited by the powerful radio galaxies is as marked as that of the quasars. These data provide important constraints on the cosmological evolution of the radio source population.

3 - Models for the Evolution of the Radio Source Population — A Cut-off at Large Redshifts?

Professor van der Laan has already described the recent computations of Peacock and Gull (1981) which attempt to model the evolution of the radio source population in such a way that all the relevant data are given equal weight in determining the form of the evolution function. The procedure involves an iterative χ^2 procedure in which the evolution function is expressed as a truncated power series in luminosity and redshift for both steep and flat-spectrum sources independently. This procedure is very expensive in computer time but results in best-fit models for the evolution function consistent with all existing data. I find most instructive their Figure 9 which is reproduced here as Figure 2. These diagrams show the evolution function as contours in the radio luminosity-redshift plane (L-Z plane) for steep and flat-spectrum radio sources. If there were no evolution of the radio source population, these diagrams would show the value of unity throughout the L-Z plane. The contours indicate the excess numbers of sources per unit comoving volume necessary to explain a very wide range of data on the source counts at high and low frequencies and identification and redshift data. The four models are for two values of q_0 (0.5 and 0) with and without a cut-off in the distribution of sources at redshift $z = 5$.

TABLE 1 - *Mean values of V/V_{max} for sources having $S_{178} \geqslant 10$ Jy, no optical limit.*

(a) Sources with known redshifts

P_{178} (WHz^{-1} sr^{-1})	Optical type	n	$<V/V_{max}>$	σ	D
$\leqslant 10^{25}$	Galaxy	19	0.508	0.066	0.12
$10^{25} - 10^{26}$	Galaxy	21	0.498	0.063	0.04
$10^{27} - 10^{27}$	Galaxy	32	0.599	0.051	1.9
	Quasar	4	0.682	—	—
$10^{27} - 10^{28}$	Galaxy	47	0.721	0.042	5.3
	Quasar	22	0.631	0.062	2.1
$> 10^{28}$	Galaxy	4	0.610	—	—
	Quasar	18	0.740	0.068	3.5
$< 10^{26}$	Galaxy	40	0.503	0.046	0.06
$> 10^{26}$	Galaxy	83	0.669	0.032	5.3
	Quasar	44	0.680	0.044	4.1
	All Sources	127	0.673	0.026	6.7

(b) All radio galaxies with $P_{178} \geqslant 10^{26}$ W Hz^{-1} sr^{-1} with various assumed redshifts for those faint identifications for which no redshift has been measured.

Assumed redshift	$< V/V_{max} >$	σ	D
0.2	0.656	0.031	5.03
0.5	0.669	0.031	5.45
1.0	0.683	0.031	5.90
2.0	0.696	0.031	6.32

σ is the standard error of the quoted mean values of V/V_{max}, $\sigma = (12\,n)^{-1/2}$, and D is the difference of $< V/V_{max} >$ from the mean value of 0.5 measured in units of σ.

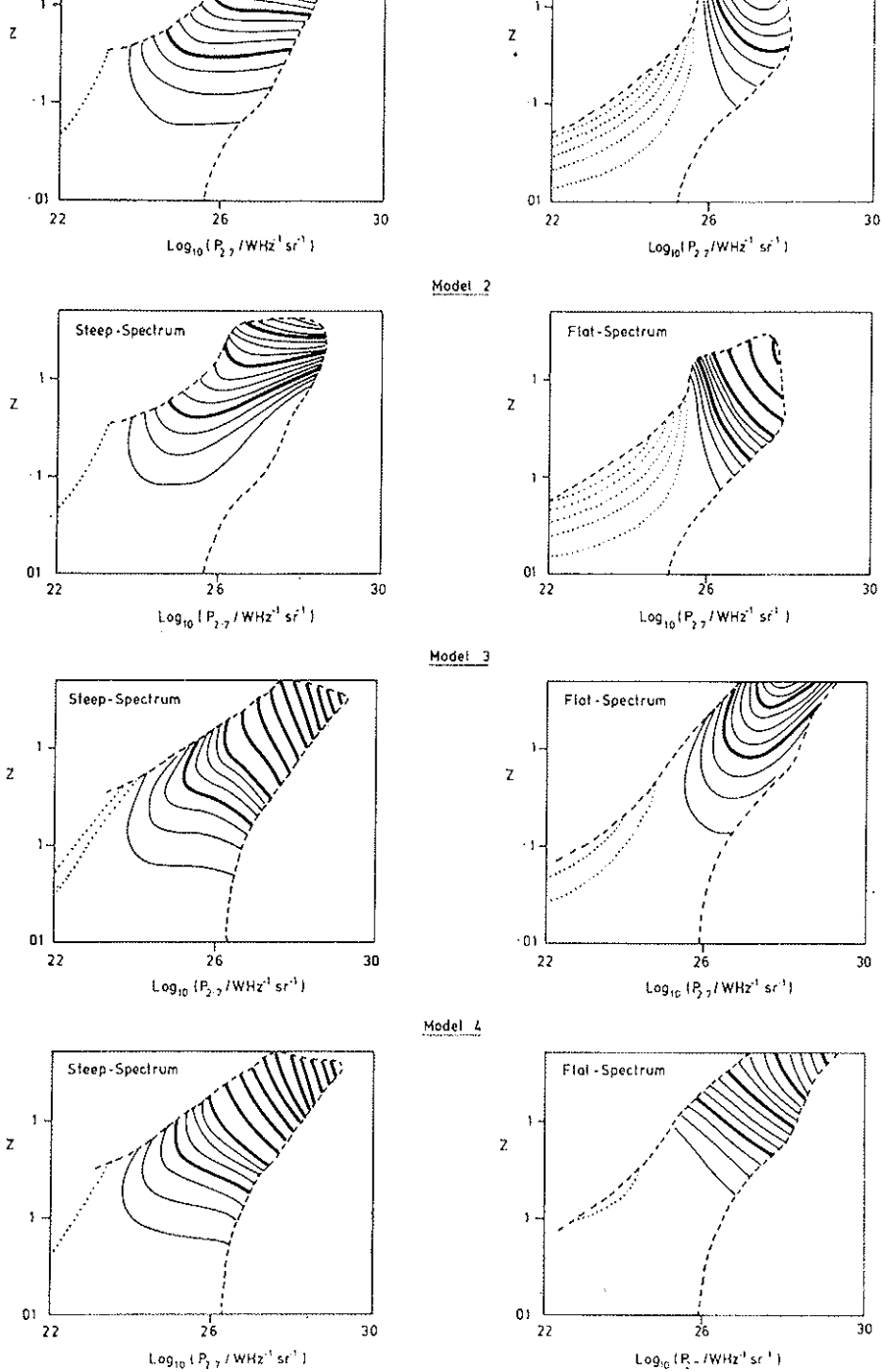


FIG. 2. The evolution functions derived by Peacock and Gull (1981) incorporating all data on source counts, identifications and redshifts for sources at low and high frequencies. Models 1 and 2 are for $q_0 = 0.5$ and models 3 and 4 for $q_0 = 0$. Models 1 and 3 have no redshift cut-off whereas models 2 and 4 have a redshift cut-off at $z = 5$. Evolution functions for steep-spectrum sources are shown in the left hand column and flat-spectrum sources are shown in the right hand column. The dashed contours show the regions within which the functions are reasonably well defined. The solid contours show the evolution function, each

Each model was run with different power series expansions in luminosity and redshift and so, in each case, an estimate could be made of the stability of the best-fit models to statistical uncertainties in the input data. They worked out for each model the contour in the evolution plane at which the range of values of the evolution function varied by a factor of 10. These contours are shown in Figure 2 by dashed lines and outside these regions the models are poorly constrained by the present data.

Within the dashed contours all the models show strong cosmological evolution, the most marked evolutionary effects being associated with the most powerful sources. It is noteworthy that both the steep and flat spectrum sources show similar strong evolution.

One interesting question which can be addressed by such an analysis is whether there is any evidence for a cut-off in the distribution of radio sources at large redshifts. There is some evidence for such a cut-off in the distribution of optically selected quasars (see the presentation by Professor Schmidt) and it is important to find out if such a cut-off is also present for radio selected objects.

At present, no definite answer can be given from the radio data hut there are some suggestive trends in the models shown in Figure 2. It can be seen that, in those models with cut-offs at $z_c = 5$ (models 2 and 4), the contours do not show any evidence of a cut-off, since that has already been built into the model. However, in models 1 and 3, which do not have a cut-off built in, there is some evidence for the distribution of sources flattening out or even decreasing at redshifts greater than 2. This is particularly marked for the flat spectrum sources. This is somewhat suggestive of the necessity of a cut-off in the distribution of radio sources at redshifts $z \sim 3$ to 5.

Further data on the identification and redshift content of larger samples of sources at lower flux densities at all frequencies should help elucidate this problem.

4 - The Effects of Gravitational Lenses on the Observed Distribution of Radio Sources

My last remark concerns the possibility that gravitational lensing might be partly responsible for the anomalies observed in the source counts and V/V_{max} tests. A detailed analysis of these effects in so far as they affect the source counts has been made by Peacock (1982). His results are based upon a realistic model for the distribution of masses in galaxies and clusters

from which the distribution of expected amplification factors for sources at large redshifts due to gravitational lensing can be calculated. Objects at large redshifts can have their flux densities enhanced by 10 to 20% relative to a homogeneous universe but flux conservation implies that de-amplification is as common as amplification. The effects upon the source counts and redshift data were found to be small and certainly not large enough to account for the anomalies of the source counts or V/V_{max} test. The basic point is that the slope of the source counts is not sufficiently large for intrinsically rare lensing events of large amplitude to corrupt the source counts.

REFERENCES

Gunn, J.E., Hoessel, J.G., Westphal, J.A., Perryman, M.A.C. and Longair, M.S., 1981, *Mon. Not. R. astr. Soc.*, **194**, 111.

Jenkins, C.J., Pooley, G.G. and Riley, J.M., 1977, *Mem. R. astr. Soc.*, **84**, 61.

Laing, R.A., Longair, M.S., Riley, J.M., Kibblewhite, E.J. and Gunn, J.E., 1978, *Mon. Not. R. astr. Soc.*, **183**, 547.

Laing, R.A., Riley, J.M. and Longair, M.S., 1982, *Mon. Not. R. astr. Soc.*, (in preparation).

Peacock, J.A., 1982, *Mon. Not. R. astr. Soc.*, (in press).

Peacock, J.A. and Gull, S.F., 1981, *Mon. Not. R. astr. Soc.*, **196**, 611.

DISCUSSION

OSTRIKER

I would like to report on work with a Princeton graduate student, Mario Vietri, concerning the effect of gravitational lenses on the apparent distribution of quasars. We include both minilenses (stars) and extended lenses (galaxies and clusters) and find that the effect of extended lenses is always small, but minilenses can affect the bright end of the luminosity function somewhat if the dark matter is in this form rather than smoothly distributed as it would be if it were in the form of neutrinos.

ON EVOLUTIONARY MODELS OF RADIO SOURCES

G. SWARUP
Tata Institute of Fundamental Research, Radio Astronomy Centre,
Ootacamund

C.R. SUBRAHMANYA and V.K. KAPAHI
Tata Institute of Fundamental Research, Radio Astronomy Centre,
Bangalore

It is well known that evolutionary effects in the properties of extragalactic radio sources dominate the geometrical effects of reasonable world models. We shall comment here on two aspects of the evolution of radio sources, viz. the use of optical identification data to constrain possible changes in the radio luminosity function with epoch and the evidence for evolution in the linear sizes of double radio sources. The physical nature of the evolutionary effects is still very poorly understood and will not be dealt with here.

1 - Percentage Identifications

The observed radio source counts can be explained by a variety of evolutionary models for the change in radio luminosity function (RLF) with epoch (e.g. Wall, Pearson and Longair 1980; henceforth referred to as WPL). The main problem is that little is known about the redshifts of sources at intermediate and low flux densities. In the absence of such information one can use the statistics of optical identifications in order to restrict the range of possible evolutionary schemes. One such attempt was made by WPL by comparing the predicted model redshift distribution with the fraction of sources estimated from deep optical identifications to have $z < 0.6$ in the 5C source surveys. The results indicated that one particular

model (model 5 of WPL, in which the luminosity above which evolution becomes important increases with z) may be untenable because it predicts $\sim 48\%$ of the sources to have $z < 0.6$ compared to the estimated value of ~ 18 percent (*).

The statistics of optical identifications at the limit of the Palomar Sky Survey (PSS) have been investigated recently by Swarup, Subrahmanya and Venkatakrishna (1982) over a wide range in flux density (S) from the available information on the 3CR survey (Smith, Spinrad and Smith 1976), the Bologna samples (Grueff and Vigotti 1972, 1973), the Ooty occultation lists (Joshi and Singal 1980, and references therein) and the 5C samples (Richter 1975, Perryman 1979). The observed percentage identification (PI), restricted to galaxies alone, is plotted in Figure 1, from which it is seen that PI(S) decreases from about 60% at $S_{408} > 15$ Jy to $\sim 15\%$ at the 1 Jy level and to only about 10% at the 5C level of ~ 25 mJy. The predicted PI(S) relations for the models 4a, 4b and 5 of WPL are also shown in the figure. Rather than using a standard absolute magnitude for all radio galaxies, the available information on the local bivariate optical-radio luminosity function (Auriemma *et al.* 1977; Meier *et al.* 1979), has been made use of in calculating the expected PI(S).

Model 5 is clearly seen to be untenable from Figure 1 because its predictions lie well above the observations throughout the range of S considered. Furthermore, the predictions of models 4a and 4b provide a reasonable fit to the observations down to $S_{408} \sim 0.5$ Jy, but lie considerably above the observed points at the 5C flux densities. It is important to note here that, for $S_{408} < 0.1$ Jy, the optically identified sources on PSS ($z < 0.5$) must have $P_{408} < 10^{25}$ W Hz^{-1} sr^{-1}. Since models 4a and 4b of WPL require that no evolution be present for sources of such low luminosity it is clear that the observed PI values on the PSS at the 5C level cannot further constrain the choice of evolutionary schemes. The discrepancy between the observed and predicted values of PI at these flux levels is more likely to arise from an uncertainty in the local luminosity function (LLF) at low values of P.

The observed LLF for ellipticals (Auriemma *et al.* 1977) of $P_{408} < 10^{26}$ W Hz^{-1} sr^{-1} and for spirals and irregulars (Cameron 1971) is

---

* It may be noted that in an earlier paper Wall, Pearson and Longair (1977) estimated that model 5 predicts about 28% of the sources to have $z < 0.6$. However, the predicted z-distribution for $S_{408} > 10$ mJy plotted in their Figure 4 (and reproduced in Longair 1978a, 1978b) appears to be incorrect.

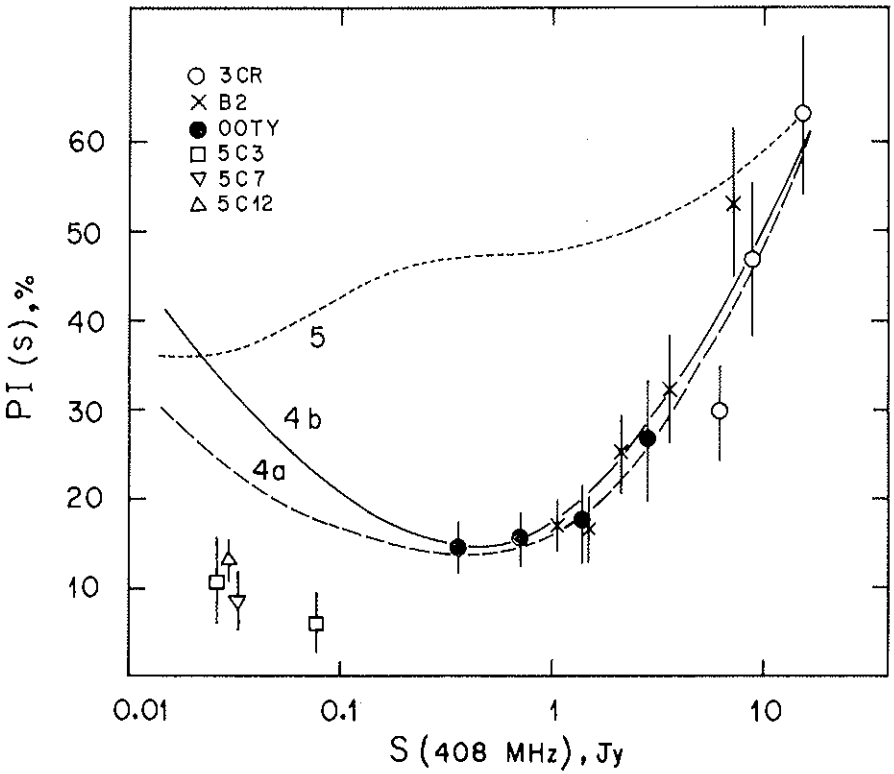


FIG. 1. Percentage of sources identified with galaxies on the Palomar Sky Survey as a function of flux density. The curves show the predictions of WPL models.

shown schematically in Figure 2 together with the total LLF used in the WPL models. There seems to be considerable uncertainty in the LLF in the range of about 10^{23} to 10^{24} W Hz^{-1} sr^{-1} because it is unclear whether or not the spirals make a significant contribution to the total luminosity function in this range of P. At the 5C flux level of $\sim$ 25 mJy a substantial contribution to the optically identified sources comes from the above range of P. If the contribution of spirals at these luminosities is assumed to be small compared to that of ellipticals, unlike the assumption implicit in the total LLF of WPL, the predicted PI at the 5C level can be made to agree with the observations in all the WPL models. A revision in the LLF would of course necessitate a reoptimization of the model parameters in order to

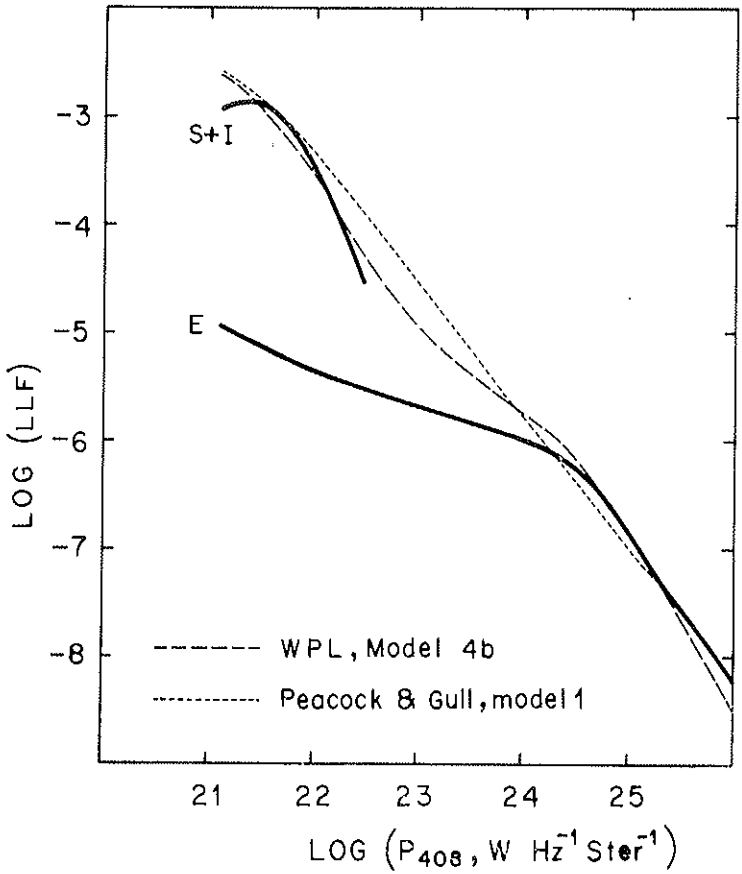


FIG. 2. The local luminosity function (Mpc^{-3}) for $\Delta \log P = 0.5$, $H_0 = 50$ km sec^{-1} Mpc^{-1}.

fit the counts at all flux levels (Swarup *et al.* 1982). A better determination of the LLF at low luminosities is therefore of considerable importance.

Recently, Peacock and Gull (1981; henceforth referred to as PG) have proposed new evolutionary models to explain the source counts at different radio frequencies hy treating the "flat" and "steep-spectrum" sources separately and by incorporating a correlation between luminosity and spectral index for the steep spectrum sources. In these models the RLF is expressed as the exponential of a power series expansion in log P and log $(1 + z)$. The total LLF at 408 MHz in the PG models (dominated by steep spectrum sources), shown in Figure 2, implies an even higher

space density of sources of low luminosity than the WPL models. These models may therefore be expected to predict even larger values of PI at the 5C level. In actual fact they predict approximately the correct PI. The reason for this fit is that the models introduce negative evolution with epoch in the space density of steep spectrum sources with $P_{408} \lesssim 10^{24}$ W Hz^{-1} sr^{-1}. At $P_{408} = 10^{25.5}$ W Hz^{-1} sr^{-1}, for instance, the density of sources in the PG models at $z = 0.25$ is about 2 to 5 times smaller than the local density ($z = 0$), and at $z = 0.5$ it is about 10 to a few hundred times smaller. Little physical significance can, however, be attached to the implied negative evolution in the absence of any observed evidence to this effect. It seems rather to be an artefact of the procedure adopted in the models of expressing the RLF as a free series expansion. The PG models in their present form are, therefore, unlikely to be very useful in predicting the statistical properties of radio sources at very low flux density levels.

2 - Angular size - Flux Density Relation

The observed relation between the median angular sizes (θ_m) of radio sources as a function of flux density has generally been interpreted to mean that the linear sizes of radio sources were statistically smaller at earlier epochs. If the linear sizes are assumed to be proportional to $(1 + z)^{-n}$, values of n between 1 and 2 appear to be required (e.g. Kapahi 1975; Katgert 1977; Subrahmanya 1977). This interpretation has been questioned recently by Downes, Longair and Perryman (1981) who have used the observed properties of sources in the 3CR strong source sample to predict the θ distributions at lower flux densities by taking into account the change in RLF implied by the WPL models. They conclude that no single value of the evolution parameter n in any of the WPL models gives a reasonable fit to the observed θ_m (S) relation over the range of S covered. Similar predictions from the 3CR sample using the WPL models made by Kapahi and Subrahmanya (1982) indicate, however, that the observed θ distributions throughout the observed range of flux density can be adequately explained by the inclusion of size evolution with n in the range of 1 to 1.5 for models 4a and 4b and $n \sim 3.5$ for model 5 (see Figure 3). The required evolution is rather strong in model 5 and may provide additional evidence against this model.

It has also been pointed out by Downes *et al.* (1981) that the observed values of θ_m at the 5C levels could be low because the weaker sources

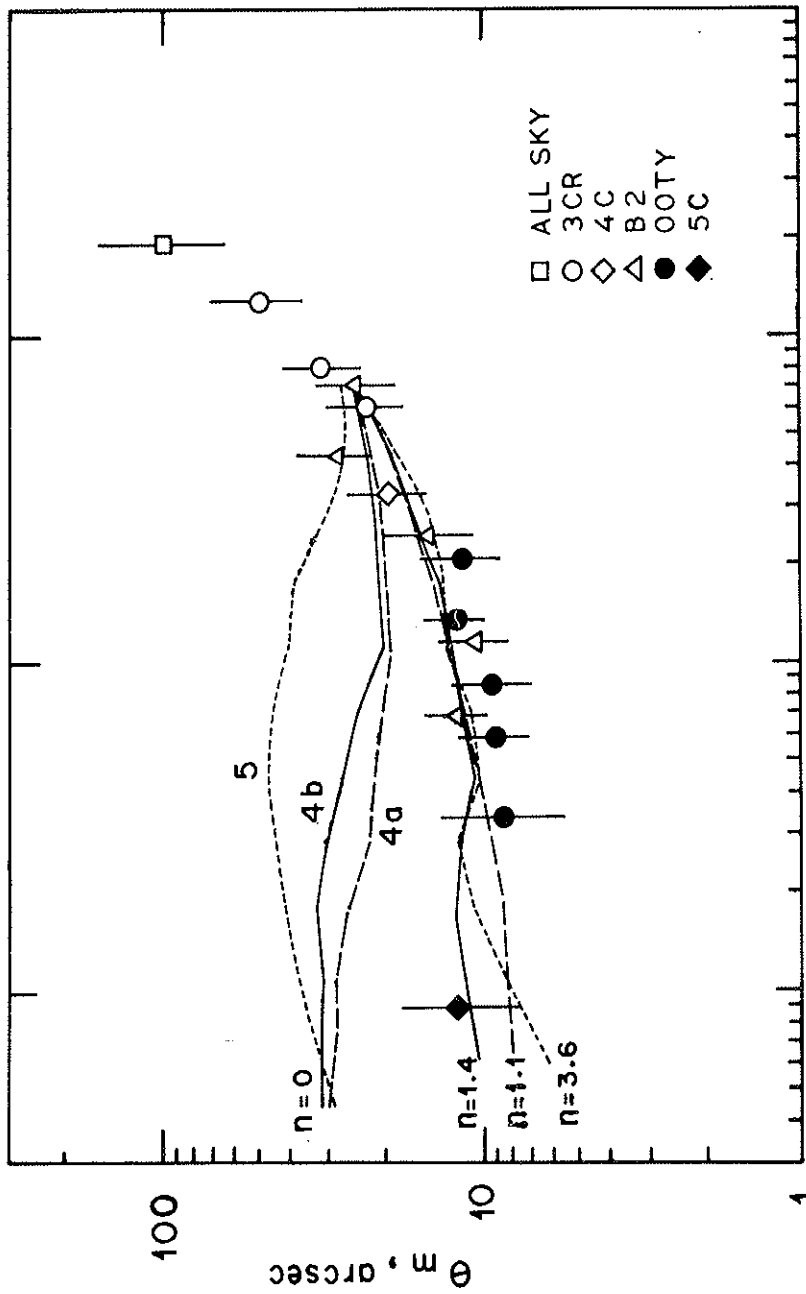


FIG. 3 The observed θ_m - S relation (Kapahi and Subrahmanya 1982) and the predictions based on the 3CR sample ($S_{178} \geq 10$ Jy) for the three evolutionary models of WPL ($q_0 = 0.5$) without size evolution (upper curves) and with size evolution using the indicated values of index n.

may be similar to those observed in high frequency surveys rather than to 3CR sources; an appreciably larger fraction of the steep spectrum sources in high frequency surveys is known to be unresolved as compared to the 3CR sample in which such sources may be underrepresented due to spectral curvature at low frequencies. Recently Downes (1982) has used a parent sample of strong sources selected at 2.7 GHz (Peacock and Wall 1981) and the evolving RLF of PG models to conclude that size evolution may not be required to explain the θ_m at 5C level in a $q_0 = 0$ universe. The predicted values of θ_m in such a scheme appear, however, to be considerably smaller than observed at high flux densities ($S_{408} \gtrsim 3$ Jy). A possible explanation that needs investigation is that extended sources whose spectra steepen at high frequencies may not be adequately represented in the 2.7 GHz sample. A substantial fraction of the extended sources in the 3CR sample are indeed known to show a steepening in their spectra at high frequencies (Laing and Peacock 1980).

REFERENCES

Auriemma, C., Perola, G.C., Ekers, R., Fanti, R., Lari, C., Jaffe, W.J. and Ulrich, M.J., 1977, *Astron. Astrophys.*, **57**, 41.

Cameron, M.J., 1971, *Mon. Not. R. Astron. Soc.*, **152**, 429.

Downes, A.J.B., Longair, M.S. and Perryman, M.A.C., 1981, *Mon. Not. R. Astron. Soc.*, **197**, 593.

Downes, A.J.B., 1982, in "*Extragalactic Radio Sources*" (eds. D.S. Heeschen and C.M. Wade), IAU Symp. 97.

Grueff, G. and Vigotti, M., 1972, *Astron. Astrophys. Suppl.*, **6**, 1.

— 1973, *Astron. Astrophys. Suppl.*, **11**, 41.

Joshi, M.N. and Singal, A.K., 1980, *Mem. Astron. Soc. India*, **1**, 49.

Kapahi, V.K., 1975, *Mon. Not. R. Astron. Soc.*, **172**, 513.

Kapahi, V.K. and Subrahmanya, C.R., 1982, in "*Extragalactic Radio Sources*" (eds. D.S. Heeschen and C.M. Wade), IAU Symp. 97.

Katgert, P., 1977, Ph. D. Thesis, University of Leiden.

Laing, R.A. and Peacock, J.A., 1980, *Mon. Not. R. Astron. Soc.*, **190**, 903.

Longair, M.S., 1978a, *Physica Scripta*, **17**, 361.

— 1978b, in "*Observational Cosmology*" (eds. A. Maeder, L. Martinet and G. Tammann), Geneva Observatory.

Meier, D.L., Ulrich, M.H., Fanti, R., Gioia, I. and Lari, C., 1979, *Astrophys. J.*, **229**, 25.

Peacock, J.A. and Gull, S.F., 1981, *Mon. Not. R. Astr. Soc.*, **196**, 611.

Peacock, J.A. and Wall, J.V., 1981, *Mon. Not. R. Astr. Soc.*, **194**, 331.

Perryman, M.A.C., 1979, *Mon. Not. R. Astron. Soc.*, **187**, 683.

Richter, G.A., 1975, *Astron. Nachr.*, **296**, 65.

Smith, H.E., Spinrad, H. and Smith, E.O., 1976, *Pub. Astron. Soc. Pacific*, **88**, 621.

Subrahmanya, C.R., 1977, Ph. D. Thesis, University of Bombay.

Swarup, G., Suhrahmanya, C.R. and Venkatakrishna, K.L., 1982, *Astron. Astrophys.*, in press.

Wall, J.V., Pearson, T.J. and Longair, M.S., 1977, in "*Radio Astronomy and Cosmology*", (ed. D.L. Jauncey), D. Reidel, Dordrecht, p. 269.

— 1980, *Mon. Not. R. Astron. Soc.*, **193**, 683.

DISCUSSION

LONGAIR

Ann Downes, Mike Perryman, John Fielden, Jeremy Allington-Smith and I have been working on the angular-diameter flux-density relation for radio galaxies and quasars down to the 5C level, i.e., about a factor of 100 fainter than the bright 3CR samples. My own view is that the question of the form of the angular-diameter redshift relation is a bonus which may come out of our studies of the physical nature of the radio sources selected at faint flux densities. For example, we regard it as an essential first step to find out if the types of sources, their identifications and their relative proportions are what one would expect from high flux densities.

When we try to fit the angular-size flux density relation, we find that, using the recent evolutionary models of Peacock and Gull, we may or may not require evolution of the physical sizes of the sources with cosmic epoch. One fact which we have established, and which slightly complicates the picture, is that we are finding numerous compact steep-spectrum sources in our faint low-frequency samples. A strong possibility is that these are really compact high frequency sources which are being redshifted into our sample. This means that it is dangerous to build models for the angular-diameter flux-density relation which do not include tbe high frequency as well as the low frequency populations. My view is that the question is unsettled as to whether or not physical size evolution of radio sources is necessary to explain the observations.

V.

PRIMORDIAL NUCLEOSYNTHESIS AND THE ORIGIN OF GALAXIES

PRIMORDIAL NUCLEOSYNTHESIS AND ITS CONSEQUENCES

JEAN AUDOUZE
Institut d'Astrophysique du CNRS
Paris

ABSTRACT

The purpose of this review is to show that the light element (D, ^{3}He, ^{4}He and ^{7}Li) abundances are among the most powerful tools to decipher the physical conditions of the early universe.

The most recent data especially those concerning the D, ^{4}He and ^{7}Li abundances are first reviewed. Then a brief summary of the nucleosynthetic properties of the Big Bang models, especially the most classical one called also the "canonical" Big Bang model, is presented. The main consequences of this primordial nucleosynthesis are underlined. Tbey concern:

1 - The determination of the present density of the universe and, therefore, its overall evolution (either an everlasting expansion or a possible succession of expansion and contraction phases).

2 - The relative density of baryons with respect to that of photons. The resulting value of this parameter constitutes a very strong argument against a "symmetric" universe (where the antimatter density would be comparable to that of matter) and can be easily explained by theories where the basic interactions (strong or nuclear, electromagnetic and weak interactions) are unified.

3 - The speed of the initial expansion which is mainly related to the number of existing families of neutrinos.

The primordial nucleosynthesis acts as a significant constraint on the Big Bang models and therefore on the cosmology of the early universe. It

constrains also different aspects of particle physics, like the possible non zero mass of the neutrinos, the number of separate lepton families and the lifetime of the neutron.

A few remarks on the use of the light elements to probe some aspects of the chemical evolution of the galaxies are presented before the conclusion which stresses again the importance of these light elements in astrophysics and cosmology.

1 - Introduction

Among the tools which help cosmologists to decipher the early history of the universe the primordial nucleosynthesis and its by-products (D, ^{3}He, ^{4}He and ^{7}Li) are certainly among the simplest and possihly the most effective ones. It is indeed well known that the Big Bang theories explain in a rather natural and straight forward way the formation and the observed abundances of these light elements. The purpose of this contribution is to review the present status of these questions and to outline some of the consequences both for astrophysics and particle physics. The primordial nucleosynthesis and a thorough study of its by-products and their relevant abundances allow us to draw very useful conclusions:

(i) on some cosmological parameters like the present density of the universe and the future of the overall expansion of the universe. Will the universe expand for ever (open universe) or will it contract in the future (closed universe)?

(ii) on the history of galaxies taking into account the evolution with time of the light element abundances;

(iii) on some aspects of particle physics like the number of existing lepton families, the hypothetical mass of neutrinos, etc.

Due to its importance this topic has been reviewed often in the literature. Among the many references which review the primordial nucleosynthesis I draw the attention of the reader to those of Wagoner (1980), Austin (1981) and Audouze (1981) and references therein.

Section 2 reviews the present observations concerning the abundances of the relevant elements (D, ^{3}He, ^{4}He and ^{7}Li) by emphasizing the most recent and intriguing data. Section 3 is devoted to a very brief account of the Big Bang nucleosynthesis. Section 4 analyzes the cosmological and particle physics constraints which can be set from the primordial nucleo-

synthesis. Section 5 contains a few remarks on the implications of this nucleosynthesis on the current models of chemical evolution of galaxies. Finally our conclusions are presented in section 6. At the end of my oral presentation M. Rees inquired about the "heretical" views regarding that type of nucleosynthesis. A few remarks on the "heresy" complete this contribution.

2 - The abundance of the "primordial" elements

2.1 - *The deuterium abundances*

At the beginning of the seventies the only available deuterium abundance was that deduced from the analysis of the deuterated water content in terrestrial and some meteoric samples ($D/H = -1.6\,10^{-4}$ according to Boato 1954).

Today there is a large body of data concerning the D abundances not only in the solar system but also in the interstellar medium throughout the galactic disk. This is due to the efforts of radio astronomers, especially Penzias and his associates, who analyzed the relative abundances of the interstellar deuterated molecules, and tbe guest investigators of the Copernicus missions, including Vidal-Madjar, Laurent, York, Rogerson, Dupree and others, who analyzed thoroughly the D/H abundance for different lines of sight in the nearby interstellar medium. Table 1, which is extracted from a compilation made by Geiss and Reeves (1981), provides a list of the available D abundances in the solar system. Although there are large variations among the various determinations, I would presently agree with the conclusion of Geiss and Reeves (1981) that the large enrichments in the D abundances, especially those observed in some carbonaceous chondrites by Robert *et al.* (1979), are due to the effect of chemical reactions at low temperatures. Therefore I adopt:

$$(D/H)_{\text{sol syst}} = 2 \pm 1\ 10^{-5}$$

Concerning the interstellar medium there are two sets of information:

a) One coming from the observation of deuterated molecules in molecular clouds (Table 2). These observations seem to show that the ratio between deuterated against hydrogenic molecules decreases when

TABLE 1 - *Deuterium abundances in the solar system.*

Earth	$1.6\ 10^{-4}$	Boato 1954
	$2.3 \pm 1.1\ 10^{-5}$	Combes *et al.* 1978
Jupiter	$5.1 \pm 0.7\ 10^{-5}$	Trauger *et al.* 1977
	$2.8\ ^{+1.6}_{-0.5}\ 10^{-5}$	Kunbe *et al.* 1981
Saturn	$5.9 \pm 2.3\ 10^{-5}$	Macy *et al.* 1978
Uranus	$6.6 \pm 1.5\ 10^{-5}$	Macy *et al.* 1978
Meteorites	$1.4\ 10^{-4}$ $8\ 10^{-4}$	Boato 1954 Kolodny *et al.* 1980 Robert *et al.* 1979
Solar winds	$2 \pm 1\ 10^{-5}$	Geiss et Reeves 1972

This compilation is extracted from the paper of Geiss and Reeves (1981).

TABLE 2 - *Relative abundances of deuterated molecules in the Galaxy.*

Source	Radius (kpc)	DCN/HCN	DCO^+/HCO^+	NH_2D/NH_3
Sgr A	0	$1.1 \pm 0.4\ (-3)$		
Sgr B	0.1	$8 \pm 5\ (-4)$	$4 \pm 3\ (-4)$	$1.7\ (-2)$
W 33	5.7	$2.3 \pm 0.4\ (-3)$		
W 51	7.6	$1.8 \pm 0.7\ (-3)$		
M 17	8.0	$1.8 \pm 0.6\ (-3)$	$7 \pm 3\ (-4)$	
DR 21 (OH)	9.9	$2.9 \pm 1.2\ (-3)$		
DR 21	9.9	$1.2 \pm 0.3\ (-3)$	$5.3 \pm 0.5\ (-3)$	
Ori A	10.9	$2.9 \pm 0.6\ (-3)$	$1.7 \pm 0.6\ (-3)$	$5.0\ (-2)$
		$6.8 \pm 0.6\ (-3)$		
NGC 2264	11.1		$1.02 \pm 0.07\ (-2)$	
W3 (OH)	12.2	$4.3 \pm 0.9\ (-3)$		
NGC 7538	12.7	$4.6 \pm 1.0\ (-3)$		

This table comes from the paper of Penzias (1981); the measurements have been performed by Penzias *et al.* (1977), Penzias (1979) and Turner *et al.* (1978).

one goes from external to internal regions of the galaxy. This effect has two consequences:

(i) Such a gradient would be an argument in favour of a pre-galactic origin of D (Ostriker and Tinsley 1975).

(ii) Since in the galactic center much more matter cycles into stars, which should destroy D, the presence of this fragile nuclear species in such an unfavourable environment is quite significant. In section 5 we will attempt to interpret the presence of D in the galactic center (Audouze *et al.* 1976).

b) The observation of the D line in the UV at 918 Å. Fig. 1 (from the work of Laurent *et al.* 1979) displays the data obtained by Laurent, Rogerson, Vidal-Madjar and York on one band and by Dupree and associates on the other hand. There are two points which can be argued from such a diagram:

(i) A value of $(D/H)_{\text{interstellar}} = (1.5 \pm 0.5)\ 10^{-5}$ agrees (within the error bars) with all the data gathered by the Princeton and the Laboratoire de Physique Stellaire et Planétaire groups.

(ii) The uncertainty on D/H obviously grows when one adds the two determinations coming from the Harvard group. We devoted two works (Vidal-Madjar *et al.* 1978 and Bruston *et al.* 1981) in an attempt to propose reasonable physical processes explaining such a dispersion between the D/H abundance observed in the direction of α Aur and α Cen A. Two processes have been advocated: first that the radiation pressure mechanism acts on D and not on H because of the saturation of the Lyman H lines; and secondly that there exist some molecular fractionation effects due to an easier dissociation of the DH molecule compared to the H_2 one. A combination of these two effects might explain some depletion of the D/H ratio observed in the galactic center direction and the corresponding enrichment in the α Aur line of sight.

To conclude this survey of the present situation regarding D, we would still favour a solar system $D/H = (2 \pm 1)\ 10^{-5}$ and an interstellar (in the solar neighbourhood) $D/H = (1.5 \pm 0.5)\ 10^{-5}$. Nevertheless the reader should note the very large dispersions both for the solar system and the interstellar abundances. To complete this review, one should note the interesting upper limit on the D/H abundance in the Canopus (a giant star) atmosphere which is said not to have suffered much D depletion. As determined by Peimbert *et al.* (1982) $D/H \leqslant 9\ 10^{-6}$.

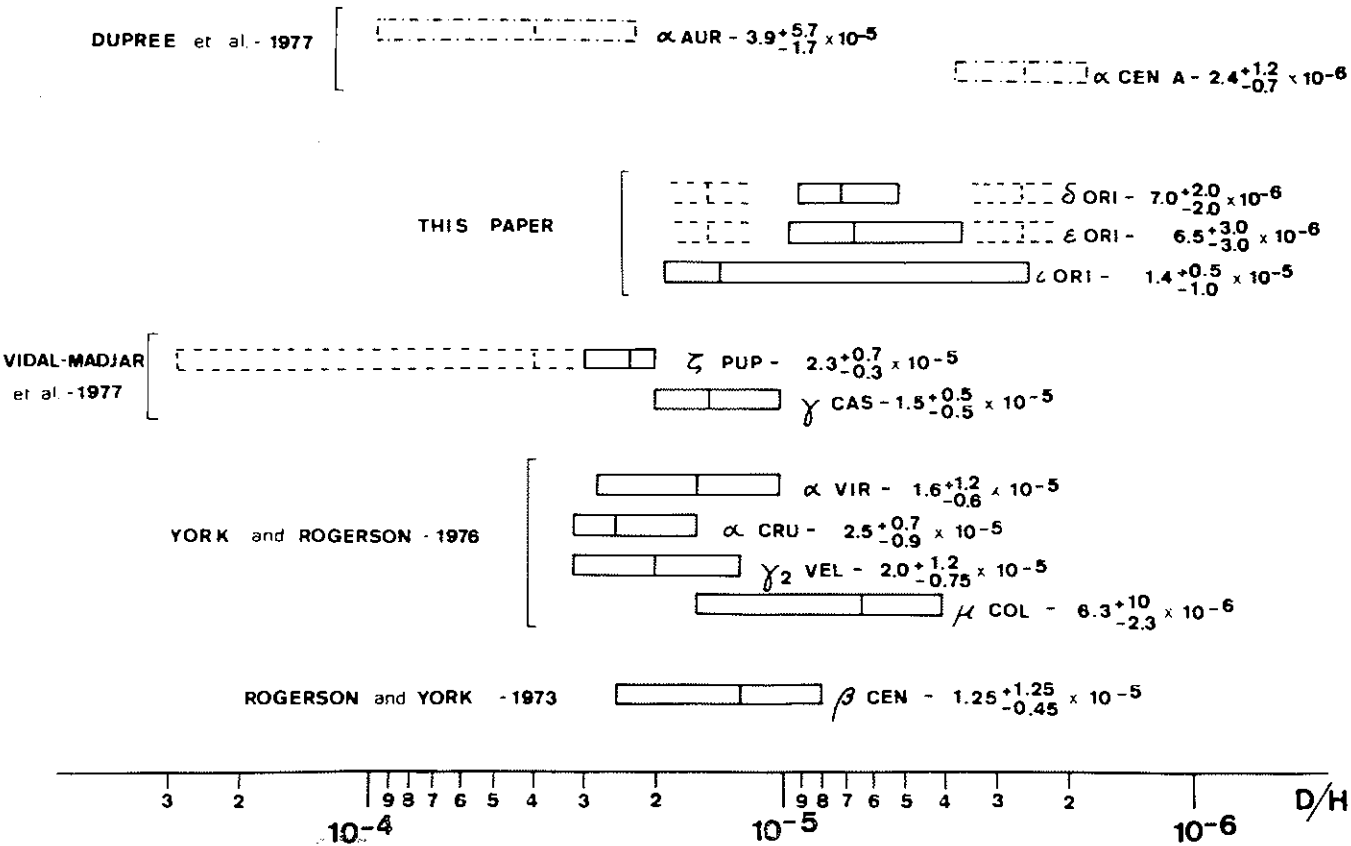


FIG. 1. Summary of the available determinations of the D abundance in the nearby interstellar medium (from Laurent, Vidal-Madjar and York 1979). One can notice the fairly large variation from one line of sight to another. The biggest variations come from the α Aur and α Cen A lines of sight investigated by Dupree *et al.* (1977). They are discussed by Vidal-Madjar *et al.* (1978) and Bruston *et al.* (1981). Despite these large variations this overall set of data seems to be consistent with D/H ~ 10^{-5}.

2.2 - *The ^{3}He abundances*

The situation is much simpler because of the scarce data available. I will repeat only what I wrote in Audouze (1981): A combination of the solar wind data (Geiss and Reeves 1972), the ^{3}He determination in gas rich meteorites (Black 1972) and the present upper limit for the interstellar ^{3}He (Rood *et al.* 1979) lead to $(^3He/H)_{primordial} = (1.5 \pm 0.5)\ 10^{-5}$.

2.3 - *The ^{7}Li abundances*

In the spring of 1981 when I wrote my previous review on the topic the situation regarding this nuclear species was quite simple. At that time the primordial ^{7}Li/H abundance was assumed to be about 10^{-9}. This conclusion was reached by using the solar system and interstellar medium data (Table 3). Very recently during this summer M. Spite and F. Spite (1981) released a quite exciting new set of Li abundances determined for five population II dwarf stars:

TABLE 3 - *"Primordial" abundances of the light elements.*

	"Canonical value"	Remarks
D/H	$(2 \pm 0.5)\ 10^{-5}$	(1) $3\ 10^{-6} < D/H < 3\ 10^{-5}$ if Dupree *et al.* (1977) can be applied to deduce the present interstellar D/H ratio. (2) a $(D/H)_{prim}$ value of $\sim 5\ 10^{-6}$ might be favoured by Vidal-Madjar *et al.* (1981) and Vidal-Madjar (1981).
^{3}He/H	$1 - 2\ 10^{-5}$	$\sim 5\ 10^{-5}$
^{4}He/H	$0.22 < Y < 0.245$	
^{7}Li/H	10^{-9}	$(^7Li/H)_{prim}$ can be $\sim 10^{-10}$ if one shows that the population II stars studied by Spite and Spite (1981) have not destroyed their surface Li

HD 19445 $(Li/H \sim 10^{-10})$
HD 76932 $(Li/H \sim 9\ 10^{-11})$
HD 134169 $(Li/H \sim 1.6\ 10^{-10})$
HD 140283 $(Li/H \sim 1.1\ 10^{-10})$
HD 201891 $(Li/H \sim 8\ 10^{-11})$.

These authors have used the coudé spectrograph of the CFH telescope with a RETICON detector. Their analysis shows that contrary to old disk stars, these halo dwarf stars do not exhibit any significant Li/H variation over a large range of effective temperatures. Moreover their Li/H abundance ($Li/H \sim 10^{-10}$) is comparable to that of the F type dwarf stars (ten times more than that of the Sun). It is of course necessary to check the hypothesis that the superficial Li observed in these stars has not been depleted by thermonuclear reactions occurring in inner regions of these stars. It would be tempting at present to guess from this exciting data that this hypothesis is correct and therefore that $(Li/H)_{primordial} \sim 10^{-10}$ while the (Li/H) abundance at the formation of the solar system and in the present nearby interstellar medium is about 10^{-9}. The consequences of such a guess are discussed in section 5.

2.4 - *The ^{4}He abundances*

A tremendous amount of observational and theoretical work has been devoted to an attempt to determine the ^{4}He abundance in many different astrophysical sites (Sun and solar system, old stars, ionized regions of the interstellar medium, other galaxies, etc.). In this contribution it is not possible to mention all of them. The reader is referred to Kunth (1981) for a recent and complete account of this subject.

2.4.1. - *^{4}He abundances in the Sun and the solar system*

Because of its well known volatility the ^{4}He abundance cannot be measured in the telluric planets and in the meteorites. In the case of the Sun, this abundance can be determined in three different ways:

(i) the use of theoretical solar models;

(ii) the direct measurements of the He content in the solar wind and the solar cosmic rays;

(iii) the analysis of the He emission lines.

The two first techniques are not free from difficulties arising from the low values of the solar neutrino fluxes and the occurrence of acceleration processes which affect the solar wind and cosmic ray composition. From the emission lines, $Y_{sun} = 0.21 \pm 0.03$.

Using the infrared spectrograph IRIS launched by the Voyager mission, Gautier *et al.* (1981) have determined the helium content in the Jupiter atmosphere. From a fit of the infrared spectrogram between 280 and 600 cm^{-1} and a comparison between a thermal profile and the IR spectra these authors found that $0.15 \leqslant Y_{Jupiter} \leqslant 0.25$. A similar study of the He abundance at the surface of Saturn is presently in progress.

2.4.2 - *^{4}He abundances in stars*

There is a large spread in the stellar He abundance: $0.24 \leqslant Y \leqslant 0.30$ at the surface of population I stars. For population II stars, I mention the work of Carney (1979) who found $Y_{pop\ II} = 0.19 \pm 0.04$ (see Kunth 1981 for more references and discussion).

2.4.3 - *The helium abundance in the galactic interstellar medium*

Among many determinations and/or analyses of the helium content in the interstellar medium one can quote the upper limit coming from the planetary nebulae (Thum *et al.* 1980) and the He abundance deduced from the radio recombination lines H 109 α and He 109 α : $Y = 0.22 \pm 0.02$ (Thum 1980).

2.4.4 - *The blue compact or "lazy" galaxies*

A class of galaxies which are blue, have low luminosity and mass and are compact has been thoroughly studied by many authors including Lequeux *et al.* (1979), Kunth and Sargent (1979, 1982) and Kunth (1981). Because these galaxies have a large gas content and are blue, it is assumed that they are just beginning to process gas into stars. These authors argue that one can deduce from their analysis the primordial helium content by extrapolating the Y (Z) relationship up to Y_p ($Z = 0$).

From their sample Lequeux *et al.* (1979) found:

$$Y = 0.233 \pm 0.005 + (1.73 \pm 0.90)\, Z$$

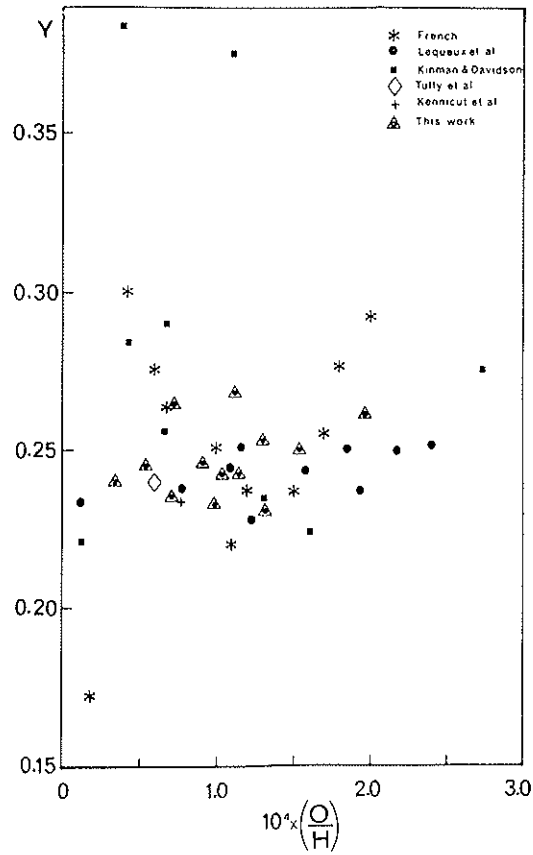


FIG. 2. Compendium of the available Y determinations versus the Oxygen/Hydrogen ratio determined in different blue compact galaxies and drawn up by Kunth (1981). This figure clearly shows the lack of correlation between Y and the O/H ratio.

The most extensive analysis is the one of Kunth (1981) who found that $Y_p \leqslant 0.245$ without any significant correlation between Y and Z (see Figure 2) contrary to previous investigations such as Peimbert (1974) and Lequeux *et al.* (1979). Moreover this work concludes that no galaxy may have a He content lower than Y = 0.20. This contradicts the analysis of I Zw 18 published recently by French (1980) which is not supported by other He determinations for the same galaxy. My own view on the Kunth (1981) analysis is that:

$$0.22 \leqslant Y_{\text{prim}} \leqslant 0.245 .$$

2.5 - *Conclusion*

The available abundances which are relevant to the analysis of the primordial nucleosynthesis and which, therefore, can be used as probes of any cosmological model are presented in Table 3. It is hoped that further work will fix more accurately the primordial abundance of D, ^{3}He and ^{7}Li.

3 - The Big Bang nucleosynthesis

This question has already been reviewed in several papers (Wagoner 1980 and Audouze 1981). I mention first assumptions which are made in the Big Bang models and then I review the principal features of the primordial nucleosynthesis.

If the birth of the universe is described by a Big Bang model one assumes that during its early phases the universe was very hot ($T > 10^{11}$ K) and dense (above 10^{13} g cm^{-3}). One adopts also the equivalence principle according to which the physical laws are invariant with time and with location.

The current models of primordial nucleosynthesis such as those of Wagoner (1973) have been constructed by using specific Big Bang frameworks often called the "canonical" or "standard" Big Bang models.

In these models, one assumes:

1) That the early universe was homogeneous and isotropic. This assumption known as the Cosmological Principle implies that the universe can be described by the Robertson-Walker metric.

2) The expansion of the universe is correctly described by Einstein's theory of relativity. The expansion rate is then $(1/V)\ (dV/dt) = (3/R)\ (dR/dt) = \sqrt{24\,\pi\,G\,\rho}$ where V represents a volume element, R (t) is the scale factor of the universe, G is the gravitational constant and ρ is the total density of the universe (including not only nuclei but also other forms of matter like neutrinos). In the following discussion the actual expansion, which could have been slower or quicker than the "free fall" rate written above, can be expressed as $(1/V)\ (dV/dt) = \xi\ \sqrt{24\,\pi\,G\,\rho}$ where $\xi > 1$ for a rapid expansion and $\xi < 1$ for a slow expansion.

3) The universe is asymmetric, i.e. the amount of antimatter present in the universe is negligible in comparison with the amount of matter. This is a consequence of the observed ratio between the number of photons

n_γ and the number of baryons n_B observed in the universe $n_\gamma/n_B = 10^9$ (a symmetric universe leads to much larger $n_\gamma/n_b \gg 10^{12}$).

4) The lepton number is the quantity:
$L = n_{e^-} - n_{e^+} + n_{\nu_e} - n_{\bar{\nu}_e} + n_{\mu^-} - n_{\mu^+} + n_{\nu_\mu} - n_{\bar{\nu}_\mu} + \ldots$ which expresses the differences between the electron and positron densities, the negative and the positive muon densities and the corresponding neutrino and antineutrino densities. This sum might have other terms if other types of leptons are existing in nature. In the canonical Big Bang this lepton number is assumed to be much lower than the photon density.

The primordial nucleosynthesis which forms D, ^{3}He, ^{4}He and ^{7}Li starts only when the temperature is low enough so that deuterium will not be destroyed by photodisintegration reactions, i.e. $\gamma + D \rightarrow p + n$. This temperature is, therefore, around 10^9 K ($kT \sim 0.1$ MeV). At such temperatures the thermonuclear reaction cycle shown in Figure 3 can proceed starting from the $n + p$ absorption.

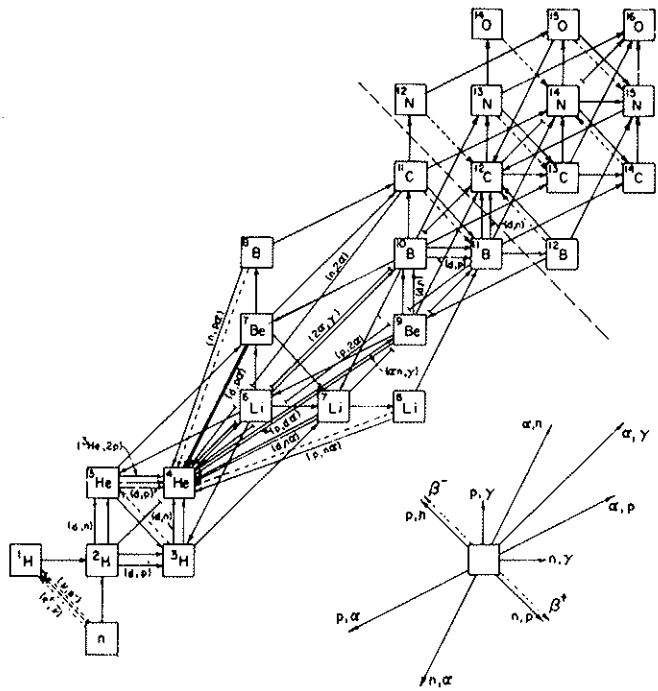


FIG. 3. The nuclear reaction network used and analyzed by Wagoner (1973) in his Big Bang models.

The rate of this absorption reaction depends mainly on the relative densities of neutrinos and protons at the time when $kT < 0.1$ MeV, i.e. when the equilibrium (due to the $n + \nu_e \rightleftharpoons e + p$ and $n + e^+ \rightleftharpoons p + \bar{\nu}_e$ reactions) between protons and neutrinos does not hold anymore because of the neutron decay. These relative densities between protons and neutrons are given by:

$$X_n/X_p = \exp - (M_n - M_p) / kT_f \ ,$$

where $M_n - M_p$ is the difference between the masses of the proton and the neutron and T_f is the freezing temperature at which the equilibrium recalled above stops. From the above relations one sees that if the expansion is rapid the equilibrium between p and n stops at higher temperatures leading to higher X_n/X_p ratios.

The baryon density during the nucleosynthesis period governs also the reaction rates and then the relative abundances hetween the synthesized light elements. The Big Bang models have been built up under the assumption that the expansion of the universe is basically adiabatic. The consequence is that the parameter h (called the baryon density parameter) $= \rho_B/T_9^3$ where ρ_B is the baryon density of the universe and T_9 the temperature in 10^9 K units can be supposed to remain constant during the whole expansion. The classical Figures 4 and 5 extracted from Wagoner 1973 show respectively the evolution of the relevant abundances with time in the frame of a canonical Big Bang where h has been taken equal to $1.15\ 10^{-5}$ ($\rho_{\text{present}} = 2.8\ 10^{-31}$ g cm^{-3}) and the final (at the end of the Big Bang period) abundances with respect to the present density of the universe, ρ_B (which is proportional to h). Figure 4 shows the synthesis of D, ^{3}He, ^{4}He and ^{7}Li from the absorption of n and the subsequent thermonuclear reaction starting from (D , p) ^{3}He, D (D , n) ^{3}He and D (D , p) T. Figure 5 shows that the ^{4}He abundance does not depend significantly on the present density of the universe. As has been recalled by Yang *et al.* (1979)

$$X(^4\text{He}) \simeq \frac{2\ X_n/X_p}{1 + (X_n/X_p)} \ .$$

By contrast the abundances of D, ^{3}He and ^{7}Li depend strikingly on the parameter h or the present density of the universe.

To conclude one can see from these figures that the quite simple Big Bang model is able to explain the production of the light elements D, ^{3}He, ^{4}He and ^{7}Li with their observed abundances (for $\rho_{B\ \text{present}} = 3\ 10^{-31}$ g cm^{-3}).

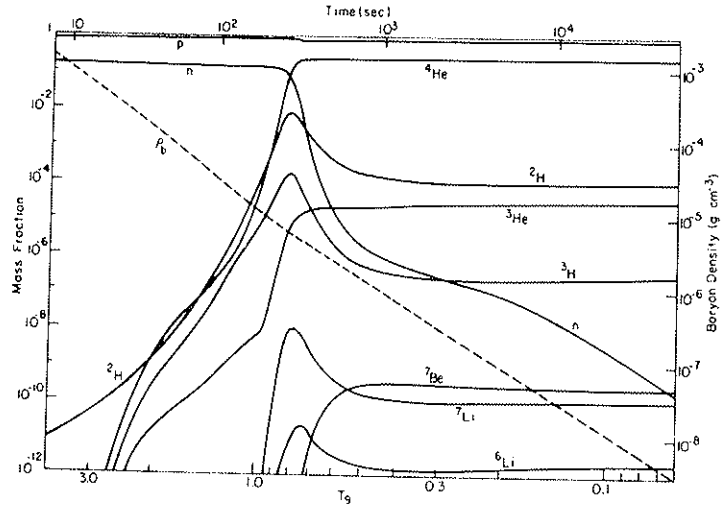


FIG. 4. The evolution with time (and temperature) of the primordial abundances and the baryon density ρ_b (dashed line) calculated by Wagoner (1973) in the context of the canonical Big Bang model in which the parameter h (see text) has been taken equal to $h = 1.15\ 10^{-5}$. One notices the decrease of the neutron density at $T_9 \sim 1$ ($t \sim 3$ minutes) coinciding with the synthesis of D, T, ^{3}He, ^{4}He, Hi and ^{7}Be which is completed after $t \sim 15$ minutes.

Moreover one can distinguish already between ^{4}He which depends mainly on the neutron/proton equilibrium conditions and the other nuclear species which depend on the present density conditions.

4 - SOME CONSEQUENCES OF THE PRIMORDIAL NUCLEOSYNTHESIS

Before discussing in more detail the implications of the dependence between primordial nucleosynthesis and some cosmologically related parameters one can already see that this light element formation process provides significant information on the occurrence of the Big Bang. The simplest models are also those which account satisfactorily for the observed abundances. Moreover these results confirm the asymmetric nature (prevalence of the matter over antimatter) of the universe.

4.1 - *Primordial nucleosynthesis and baryon density of the universe*

It has been known for some time that the D abundance restricts the

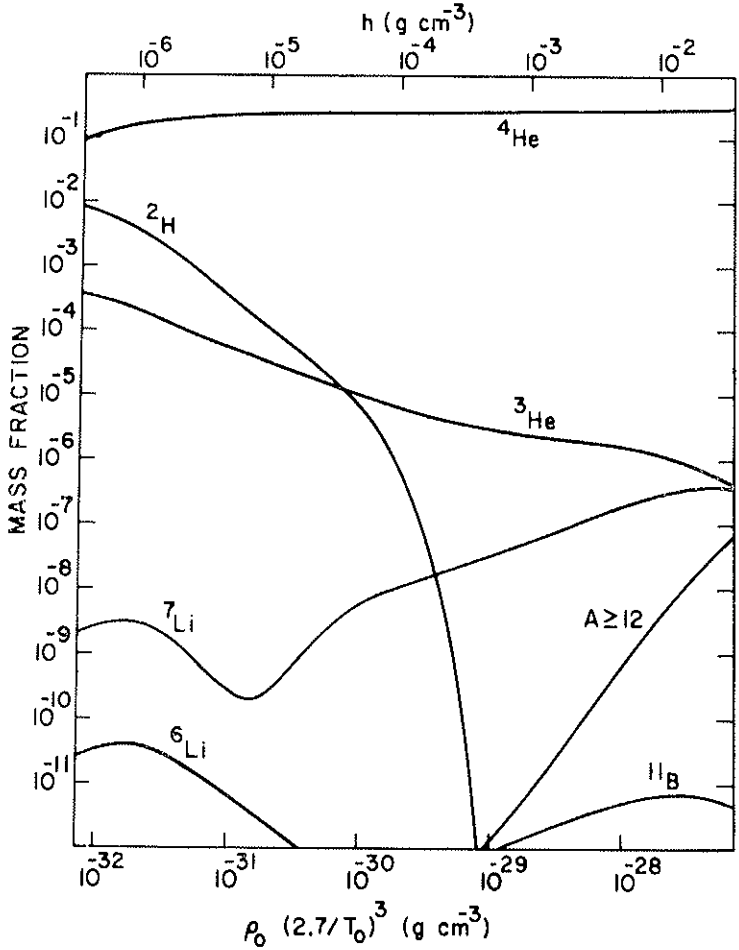


FIG. 5. Abundances of the light elements at the end of the primordial nucleosynthesis plotted by Wagoner (1973) against the present density of the universe (in the context of the canonical Big Bang model). From this very well known diagram one can see that the calculated abundances are quite in agreement with the observations for $\rho_{\text{present}} \sim 3\ 10^{-31}$ g cm^{-3} corresponding to an open universe (cosmological parameter $\Omega = 0.04$).

range of the possible present baryon density of the universe.[1] Figure 5 shows that in the frame of the canonical Big Bang model the baryon density should be lower than $\rho_b = 10^{-30}$ g cm^{-3} in order to have $X_D \geqslant 10^{-5}$. This value of the density is much lower than the critical density of the universe, $\rho_c = 6\ 10^{-30}\ (H_0/55)^2$ g cm^{-3} (where H_0 is the Hubble constant expressed in units of km s^{-1} Mpc^{-1}), above which the universe is closed. As discussed by many authors including Reeves *et al.* (1973) and Gott *et al.* (1974), the observed D abundance constitutes one of the strongest arguments in favour of an open universe expanding for ever.

Two important remarks have been proposed afterwards which strengthen this point by using the ^{3}He and ^{7}Li abundances. As shown for instance by Schramm (1981) the upper limit can he used to set up a lower limit of the present density of the universe which cannot be lower than 10^{-31} g cm^{-3}. A few years ago Austin and King (1977) made the very interesting proposal that the ^{7}Li abundance which can be built up by other nucleosynthesis processes allows one to fix an upper limit of $9 \pm 4\ 10^{-31}$ g cm^{-3} for the present density of the universe (if one uses a primordial ^{7}Li/H abundance of 10^{-9}). The same type of argument can be used again if $(^7\text{Li/H})_{\text{primordial}} = 10^{-10}$. The upper limit is then $(3 \pm 1)\ 10^{-31}$ g cm^{-3}.

4.2 - *The implications of the ^{4}He nucleosynthesis*

The now classical calculations of Wagoner (1973) show clearly (Figures 6a and 6b) the effect of the speed with which the expansion occurred on the nucleosynthesis by-products. Figure 6b shows how large the effect is on the ^{4}He abundance. Yang *et al.* (1979) have proposed the following relation:

$$X(^4\text{He}) \simeq 0.33 + 0.02 \log h + 0.38 \log \xi$$

where the influence of the expansion speed parametrized by ξ is obvious.

These authors have discussed several ways by which the parameter ξ can be affected. There may be a change of gravity theory which is very unlikely. There may be a departure from homogeneity and isotropy, if the

(1) In order to avoid any possible confusion one should recall that the thermonuclear reaction rates depend only on the baryon density while the expansion rate depends on the total energy density.

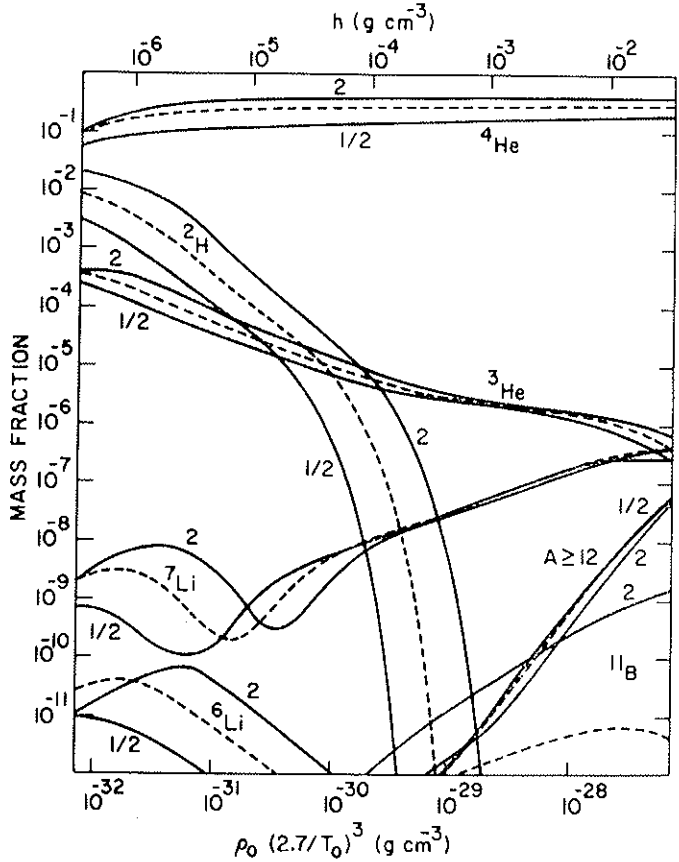


FIG. 6a. Final abundances of the light elements with respect to the present density of the universe calculated in the frame of the canonical Big Bang by Wagoner (1973) for different values of the parameter ξ governing the speed of the expansion: $\xi = 2$ (rapid expansion); $\xi = 1$ (dashed lines, free fall expansion) and $\xi = 0.5$ (slow expansion).

cosmological principle does not apply. In the directions where the matter density was larger the expansion leads most rapidly to bigger ^{4}He abundances. Most interesting is the possibility of increasing ξ through the total energy density by considering the possibility of existing new lepton categories, i.e. an increase of the leptonic number L.

Several contributions deserve to be discussed in this respect. One is by David and Reeves (1980) where these authors have attempted to find out conditions where an increase of the baryonic density is still compatible

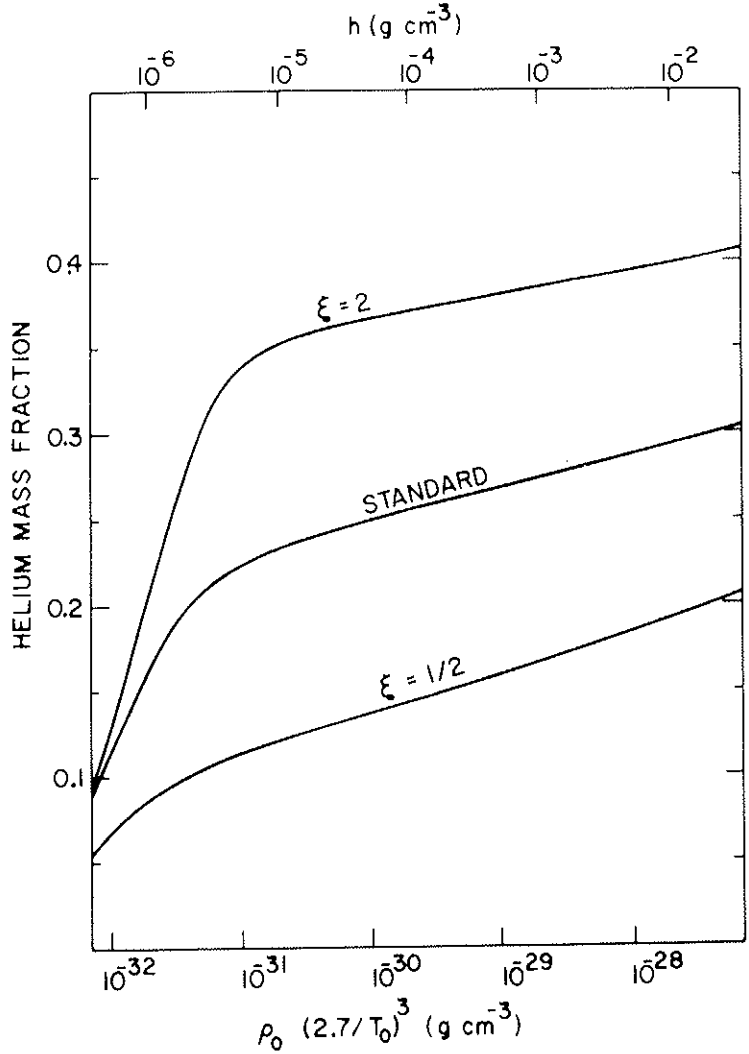


FIG. 6b. Enlargement of Figure 6a to show the dependence of Y with ξ (from Wagoner 1973).

with the observed abundances. We have already quoted the work of Yang *et al.* (1979) where the ^{4}He abundance is used to set up a limit to the maximum number of lepton (and therefore neutrino) families. This work has been challenged by Stecker (1980) and supplemented by Olive *et al.* (1981) who have recently summarized the relationship between the ^{4}He abundance, the number of lepton families, the actual neutron lifetime and the ratio between the photon and the baryon density.

The contribution of David and Reeves (1980) is to show that if one assumes very large numbers of lepton families the nucleosynthesis of the light elements becomes compatible with a universe where the baryon density can be much larger than in the canonical Big Bang model where the present density of the universe is about $3\ 10^{-31}$ g cm^{-3} corresponding to a cosmological parameter $\Omega \sim 0.04$. Figure 7 is a summary of their conclusions where the values of ζ_e and ζ_μ parameters [such that the ζ_e and ζ_μ parameters are respectively the ratios of the chemical potential (μ_e) and (μ_μ) of the electronic neutrinos and all the other neutrinos (muonic, tauic, etc.) to their thermal energy], compatible with the observed ^{4}He abun-

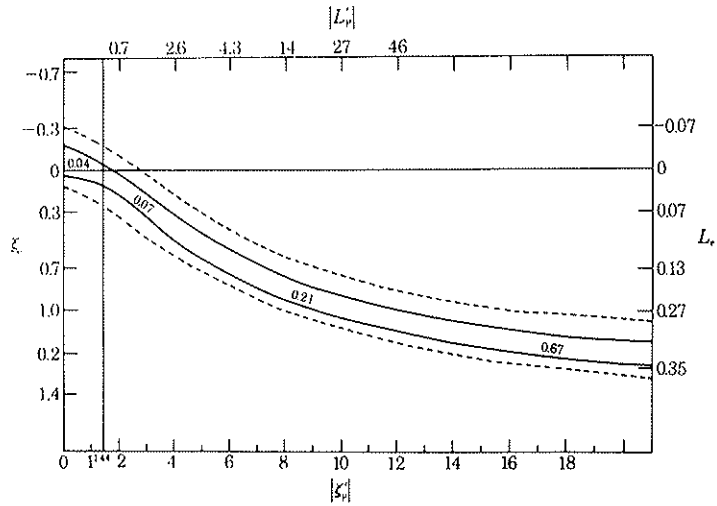


FIG. 7. Sketch of the compatibility regions in the plane ζ_e, ζ_μ (see text) for the cosmological parameter Ω ranging from 0.04 up to 0.7. These compatibility regions are defined by comparison with the observed abundances. One can notice that if ζ_μ increases (because of a very large number of different lepton families) the cosmological parameter should be larger corresponding to bigger values of the present density of the universe, but not yet to the point of closing the universe (from David and Reeves 1980).

dances, are exhibited with respect to the relevant values of the cosmological parameter Ω. If the number of different families of neutrinos is 2, $\zeta_\mu = 0 - 3$, $\zeta_\mu = 1.4 - 10$, $\zeta_\mu = 3.7 - 50$, $\zeta_\mu = 6$ and 5000 $\zeta_\mu = 22$. For such a "pathological" number of different neutrino families, the present density of the universe, compatible witb the observed abundances of the light elements, could be as large as 2/3 of the critical density.

This relationship between cosmology (defined for instance by the present density of the universe) and particle physics (the number of different lepton families) has also been establisbed by Yang *et al.* (1979). These authors have calculated again the abundance of the light elements in the context of Big Bang models where they can modify the possible number of lepton and neutrino families. An examination of Figure 8 which exhibits tbeir calculations shows that the maximum number of lepton-neutrino

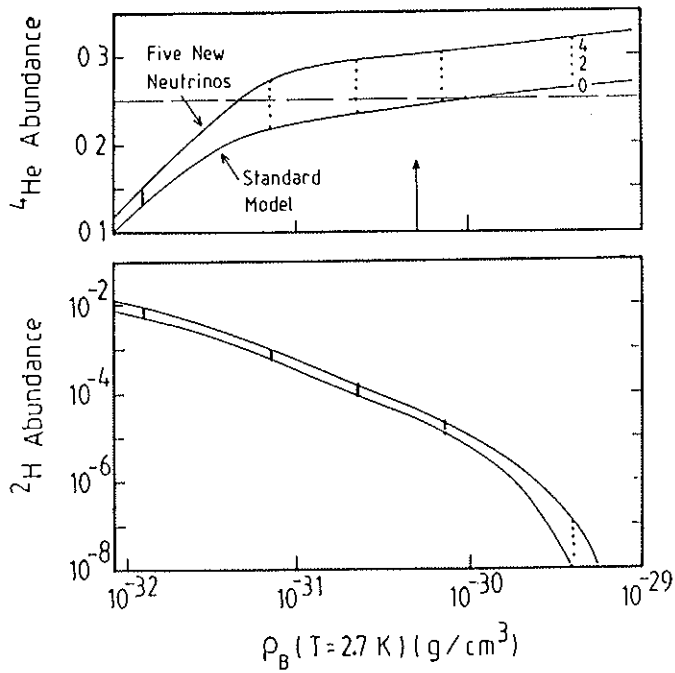


FIG. 8. Primordial abundances of ^{4}He and D sketched by Yang *et al.* (1979) with respect to the present density of the universe by assuming only two types of neutrinos (electronic and muonic neutrinos, standard models: solid lines) up to 5 new neutrino families. This figure clearly shows that there is only room for one (the tau neutrino) or two more new neutrinos in the context of the classical Big Bang models.

families is 3 or 4 depending on the actual value of the D/H abundance and, therefore, the corresponding value of the present density of the universe. Since the discovery of the third family of leptons (the tau lepton), this does not provide much room for still unknown leptons.

At this point we mention briefly the paper by Stecker (1980) who criticized the Yang *et al.* (1979) contribution on the grounds that the helium abundances make the standard Big Bang model hardly valid. The recent work of Kunth (1981) who showed that $0.23 \leqslant Y_p \leqslant 0.25$ contradicts somewhat the argument presented by Stecker who took as a "conservative" value $Y_p = 0.228$.

Finally let us summarize the interesting conclusions reached by Olive *et al.* (1981) concerning the relations between the helium abundance Y, the neutrino lifetime τ_n, the ratio between the baryon density and the photon density η and the number of neutrino (lepton) families.

The value of Y has already been discussed thoroughly. There is some uncertainty on the actual value of τ_n. The available data are those of Christensen *et al.* (1972) where $\tau_n = 10.61 \pm 0.16$ min. More recently Bondarenko *et al.* (1978) found a quite lower value (10.13 ± 0.09). This value is contradicted by those found respectively by Kugler *et al.* (1978) and by Byrne *et al.* (1980), 10.62 min and 10.82 ± 0.20 min. It seems that the higher value $\tau_n = 10.6$ min has more proponents than the lower value.

Olive *et al.* (1981) discuss at length the possible values for the parameter η. These values would come from the mass to light ratios determined in different astrophysical systems (galaxies and clusters of galaxies) by the dynamical techniques discussed by many contributors at this conference. In their analysis Olive *et al.* (1981) favour a lower limit for $\eta > 2\ 10^{-10}$ coming from the dynamics of binary galaxies and small clusters of galaxies. The most stringent lower limit is $\eta > 10^{-10}$ below which the corresponding present density of the universe would lead to D and ^{3}He abundances much larger than those quoted from the observations.

Figure 9 extracted from their work clearly shows that, to fulfil the constraints $Y \leqslant 0.25$, $\tau_n = 10.6$ min and $\eta > 2\ 10^{-10}$, the maximum number of neutrino families is 3, meaning that one should not observe any new lepton.

If the neutrinos have a mass, Olive *et al.* (1981) argue that there is still a lower limit on the baryonic density coming from the D and ^{3}He nucleosynthesis. In these conditions there is room only for about eight

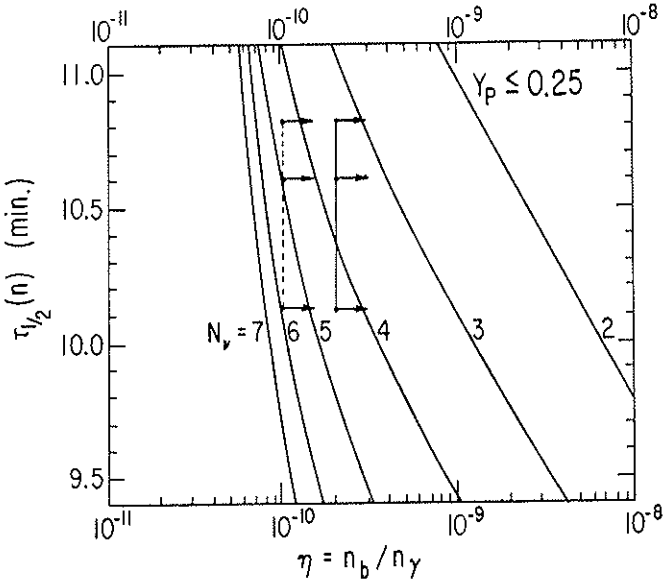


FIG. 9. Sketch drawn by Olive *et al.* (1981) providing the maximum number of neutrino families for $Y_p \leqslant 0.25$ [consistent with the Kunth (1981) conclusions] and with respect to η (the ratio between the number of baryons and photons) and the half-life of the neutron (against its beta decay). The two sets of arrows correspond to $\eta > 2\ 10^{-10}$ (dynamics of binary galaxies and small groups of galaxies $\Omega \sim 0.04$, solid lines and $\Omega > 10^{-10}$, dashed lines, lower limit of ρ_B compatible with the D and ^{3}He nucleosynthesis).

different neutrinos. In this case the limit on the number of neutrinos is not as stringent as in the case where they have no mass.

To conclude this section one must realize how many answers can be given from the simple nucleosynthesis of the light elements. We are concerned with the density of the universe, its dynamics, the presence of new families of leptons and also the hypothetical mass of neutrinos.

5 - LIGHT ELEMENT ABUNDANCES AND EVOLUTION OF GALAXIES

I shall be brief on this topic because it is a bit outside the scope of the conference which is concerned almost exclusively with cosmological problems. Nevertheless I should mention that the light elements play a very important role in the building of models which attempt to account

for the evolution of the galaxies.[2] For instance deuterium (contrary to other elements) is destroyed and not synthesized inside stars. This is at least partly the case for ^{7}Li which can be produced only by some specific objects like novae or some red giants. Moreover it has been generally claimed, until the recent observations of Kunth (1981), that there is a correlation between the helium and the heavier element enrichment.

The specific role of these light elements in the problem of the evolution of galaxies was first analyzed by Audouze and Tinsley (1974) and then more recently along similar lines by Reeves and Meyer (1978).

The infall of external gas into regions whose evolution is being discussed helps to alleviate the difficulties presented by the so-called "simple models" (see e.g. Audouze and Tinsley 1976). In those models where the rate of star formation is assumed to be constant with time and proportional to the amount of available interstellar gas, where the primordial metallicity and the stellar content is assumed to be zero and where the stellar lifetime is neglected in comparison with the galactic evolution time scales, the evolution of metallicity with time is much too steep in comparison with infall models which smooth this evolution so as to better fit the observations. The main argument presented in the Audouze and Tinsley (1974) work is that infall of external gas is able to replenish the interstellar medium with the observed deuterium nuclei. This is especially true in the case of the galactic center for which Audouze *et al.* (1976) show that it is necessary to advocate either such an infall or the possible effect of strong cosmic ray fluxes to account for the presence of this very fragile nuclear species in a region where the rate of matter cycling into stars is extremely high.

However, I would like to call the attention of the reader to the fact that, if the primordial ^{7}Li/H is as low as 10^{-10}, there is no more need of infall processes in order to reconcile D/H $\sim 10^{-5}$ and ^{7}Li/H $\sim 10^{-9}$ as stated by Reeves and Meyer (1978). In this last case (low primordial ^{7}Li/H) it is necessary to explain the bulk of the ^{7}Li/H $= 10^{-9}$ observed both in the interstellar medium, the youngest stars and the early solar system by sources like novae explosions (as proposed for instance hy Starrfield *et al.* 1978 and Vigroux and Arnould 1979). This important point deserves to be reinvestigated again in the light of a possibly low primordial ^{7}Li/H abundance.

---

(2) The reader is referred to Audouze and Tinsley (1976) for a review of the current problems related to the chemical evolution of galaxies.

Finally the lack of correlation between Y and Z means that tbe effect of mass loss during normal stellar evolution should not he overemphasized. This gives weight to models like those worked out very recently by Maeder (1981) who improves significantly upon earlier attempts presented by Chiosi (1979).

6 - Conclusion

This contribution attempts to stress the importance of primordial nucleosynthesis as a tool to select among different cosmological scenarios. M. Longair in the conclusion of the Study Week has recalled again the importance of the synthesis of D and ^{4}He as arguments in favour of a universe which started its evolution by the Big Bang phases and which should expand forever.

One realizes that the problem of determining the abundances of D, ^{3}He, ^{4}He and ^{7}Li is not yet fully solved. One should be able to increase the accuracy of the determinations of the primordial abundance of D. After the recent observations of Spite and Spite (1981) concerning ^{7}Li one should analyse the possible depletion of ^{7}Li in the halo stars studied by the authors. A depletion of ^{7}Li in these stars would confirm that $(^7\mathrm{Li/H})_{\mathrm{prim}} = 10^{-9}$ (this situation has been already thoroughly analyzed in the past). If, as suggested by the lack of correlation between the ^{7}Li abundance and the effective temperature, there is no noticeable depletion it means that $(^7\mathrm{Li/H})_{\mathrm{prim}} \sim 10^{-10}$. These findings would have two consequences:

(i) they would reinforce the open character of the universe;[3]

(ii) they would imply that the bulk of ^{7}Li observed in the solar system and the interstellar medium has been synthesized in novae (or alternatively in some red giant stars). Finally one might expect some improvements in the ^{3}He and ^{4}He abundance determinations, although the case of ^{3}He is especially difficult event for very careful and talented observers, and the case of ^{4}He can not be easily improved upon after the thorough analysis of Kunth (1981).

From the above survey of the Big Bang nucleosynthesis it appears that the simplest (also called canonical) Big Bang models are the most

---

(3) Assuming a universe filled up with zero mass neutrinos and a finite number of neutrino (lepton) families.

successful ones to account for the synthesis of these elements. Within these models, we are convinced of the predominance of matter (compared to antimatter). If the neutrinos are not massive and if there is not a very large number of lepton (neutrino) families, the universe should be open. From the Olive *et al.* (1981) analysis one has a correlation between the helium abundance $Y \leqslant 0.25$, the number of neutrino flavors < 4, the lifetime of the neutron ($\sim$ 10.6 minutes) and the baryon density/photon density ratio ($> 2\ 10^{-10}$).

The primordial nucleosynthesis is one of the strongest arguments in favour of the Big Bang cosmology. It has some very interesting consequences for some aspects of particle physics (neutrino flavors and mass; Grand Unification schemes, etc.). Finally it has some interesting consequences for the chemical evolution of galaxies, including the effective role of the infall of external unprocessed material and the actual sources of ^{3}He and ^{7}Li during the galactic evolution, if the primordial ^{7}Li is found to have an abundance as small as $^7\mathrm{Li/H} \sim 10^{-10}$.

This concludes the discussion of the "canonical" views regarding the primordial nucleosynthesis and its consequences for cosmology and for particle physics. At the end of the oral presentation of this review I was questioned by M. Rees on possible heresies, namely the possibility of synthesizing the light elements by other means than during the Big Bang. In my opinion there might be some contrived mechanisms such as population III stars, by which ^{4}He could be formed after the occurrence of the radiative phase of the evolution of the universe. By contrast it seems to me very difficult at present to work out other mechanisms than the primordial nucleosynthesis to account for the formation of deuterium.

REFERENCES

Audouze, J., 1981, *Cosmology and Particles*, Ed. J. Audouze et al., Editions Frontières, p. 231.

— 1981, in *Nuclear Astrophysics*, Ed. D. Wilkinson, Pergamon Press.

Audouze, J. and Tinsley, B.M., 1974, *Ap. J.*, **192**, 487.

— 1976, *Ann. Rev. Astron. Astrophys.*, **14**, 43.

Audouze, J., Lequeux, J., Reeves, H. and Vigroux, L., 1976, *Ap. J. Letters*, **208**, L 51.

Austin, S.M., 1981, *The creation of the light element cosmic rays and cosmology*, (to appear in Progress in Particle and Nuclear Physics).

Austin, S.M. and King, C.H., 1977, *Nature*, **269**, 782.

Black, D.C., 1972, *Geochim. Cosmochim. Acta*, **36**, 347.

Boato, G., 1954, *Geochim. Cosmochim. Acta*, **6**, 209.

Bondarenko, L.N., Kurguzov, V.V., Prokof'ev, Yu. A., Rogov, E.V. and Spirak, P.E., 1978, *JETP Letters*, **28**, 303.

Bruston, P., Audouze, J., Vidal-Madjar, A. and Laurent, C., 1981, *Ap. J.*, **243**, 161.

Byrne, J., Morse, J., Smith, K.F., Shaikh, F., Green, K. and Greene, G.L., 1980, *Phys. Lett. B.*, **92**, 274.

Carney, B.W., 1979, *Ap. J.*, **233**, 877.

Chiosi, C., 1979, *Astron. Astrophys.*, **80**, 252.

Combes, M., Encrenaz, T. and Owen, T., 1978, *Ap. J.*, **221**, 378.

Christensen, C.T., Nielsen, A., Bahnsen, A., Brown, W.K. and Rustad, B.M., 1972, *Phys. Rev. D.*, **5**, 1628.

David, Y. and Reeves, H., 1980, in *Physical Cosmology*, Eds. R. Balian, J. Audouze and D.N. Schramm, North Holland, p. 443.

Dupree, A.K., Baliunas, S.L. and Shipman, H.L., 1977, *Ap. J.*, **218**, 361.

French, H.B., 1980, *Ap. J.*, **240**, 41.

Gautier, D., Conrath, B., Flaser, M., Hanel, R., Kunde, V., Chedin, A. and Scott, N., 1981, *J.G.R.*, in press.

Geiss, J. and Reeves, H., 1972, *Astron. Astrophys.*, **18**, 126.

— 1981, *Astron. Astrophys.*, **93**, 189.

Gott, J.R., Gunn, J.E., Schramm, D.N. and Tinsley, B.M., 1974, *Ap. J.*, **194**, 543.

Kolodny, Y., Kerridge, J.F. and Kaplan, I.R., 1980, *Earth and Plan. Sci. Letters*, **46**, 149.

Kugler, K.J., Paul, W. and Trinks, U., 1978, *Phys. Lett. B.*, **72**, 422.

Kunbe, V., Honel, R., Maguire, W., Gautier, D., Baluteau, J.P., Marten, A., Chedin, A., Husson, N. and Scott, N., 1981, *Bull. A.A.S.*, **13**, 3, 735.

Kunth, D., 1981, *These de doctorat d'Etat des Sciences Physiques*, Univ. Paris 7.

Kunth, D. and Sargent, W.L.W., 1979, *Ap. J. Supp.*, **36**, 259.

— 1982, in preparation.

Laurent, C., Vidal-Madjar, A. and York, D.G., 1979, *Ap. J.*, **229**, 923.

Lequeux, J., Peimbert, M., Rayo, J.F., Serrano, A. and Torres-Peimbert, S., 1979, *Astron. Astrophys.*, **80**, 155.

Luck, J.M., Birck, L., Allègre, C.J., 1980, *Nature*, **283**, 256.

Macy Jr., W. and Smith, W.H., 1978, *Ap. J. Letters*, **222**, L 73.

Maeder, A., 1981, *Astron. Astrophys.*, in press.

Olive, K.A., Schramm, D.N., Steigman, G., Turner, M.S. and Yang, J., 1981, *Ap. J.*, **246**, 557.

Ostriker, J.P. and Tinsley, B.M., 1975, *Ap. J. Letters*, **201**, L 51.

Peimbert, M. and Torres-Peimbert, S., 1974, *Ap. J.*, **193**, 327.

Peimbert, M., Wallerstein, G., Pilachowski, C., 1982, *Astron. Astrophys.* to be published.

Penzias, A.A., 1979, *Ap. J.*, **228**, 430.

— 1981, *Nuclear Processing and Isotopes in the Galaxy* (preprint).

Penzias, A.A., Wannier, P.G., Wilson, R.W. and Linke, R.A., 1977, *Ap. J.*, **211**, 108.

Reeves, H., Audouze, J., Fowler, W.A. and Schramm, D.N., 1973, *Ap. J.*, **179**, 909.

Reeves, H. and Meyer, J.P., 1978, *Ap. J.*, **226**, 613.

Robert, F., Merlivat, L. and Javoy, M., 1979, *Nature*, **282**, 785.

Rood, R.T., Wilson, T.L. and Steigman, G., 1979, *Ap. J. Letters*, **227**, L 97.

Schramm, D.N., 1981, in *Cosmology and Particles*, Ed. J. Audouze *et al.*, Editions Frontières.

Spite, M. and Spite, F., 1981, in *IAU Colloquium n. 68, Astrophysical Parameters for Globular Clusters*, Schenectady, NY, October 1981 (in press).

Starrfield, S.G., Truran, J.W., Sparks, W.N. and Arnould, M., 1978, *Ap. J.*, **222**, 600.

Stecker, F.W., 1980, *Physical Rev. Letters*, **44**, 1237.

Thum, C., 1980, *Vistas in Astronomy*, **24**, 355.

Thum, C., Mezger, P.G. and Pankonin, V., 1980, *Astron. Astrophys.*, **87**, 269.

Trauger, J.T., Roesler, F.L. and Michelson, M.E., 1977, *Bull A.A.S.*, **9**, 516.

Turner, B.E., Zuckerman, B., Morris, M. and Palmer, P., 1978, *Ap. J. Letters*, **219**, L 43.

Vidal-Madjar, A., Laurent, C., Bruston, P. and Audouze, J., 1978, *Ap. J.*, **223**, 589.

Vidal-Madjar, A., Ferlet, R., Laurent, C. and York, D.G., 1981, *Astron. Astrophys.*, submitted.

Vigroux, L. and Arnould, M., 1979, in *Les Eléments et leurs isotopes dans l'Univers*, Université de Liège, p. 47.

Wagoner, R.V., 1973, *Ap. J.*, **179**, 343.

— 1980, in *Physical Cosmology*, Ed. R. Balian *et al.*, North Holland, p. 398.

Yang, J., Schramm, D.N., Steigman, G. and Rood, R.T., 1979, *Ap. J.*, **227**, 697.

DISCUSSION

REES

If there are any "heretics" here, they would be asking whether the ^{4}He and D could be non-primordial (but still pre-galactic), and produced by a generation of "Population III" objects at a redshift perhaps as large as 100. How would you reply to these "heretics"?

AUDOUZE

My "heretical" views are different regarding D and ^{4}He. Concerning D it seems to me impossible to avoid Big Bang nucleosynthesis for this very fragile (but abundant) element if the overall D/H ratio is above 3×10^{-6}, as it is today. For ^{4}He the situation might be more favorable for heresy if we understood how some Population III stars manage to produce ^{4}He at very early epochs without releasing any significant amount of heavier elements.

REES

If one supposes that the M/L of "baryonic matter" is < 10 solar units, corresponding to $\Omega_b \leqslant 0.005$, then the predicted ^{4}He fraction stays below 25 percent even if several extra neutrino species are added. However, the amount of (D + ^{3}He) produced in the Big Bang would then be unacceptably high.

WEINBERG

I may be the only person here who missed this point, but could you explain how it is that having massive neutrinos removes the constraint on the number of neutrino types?

AUDOUZE

This point, raised by Olive *et al.* (1981), is the following: assume that the baryonic density is the one calculated from the solar neighborhood mass distribution, i.e. $\Omega = 1.4 \times 10^{-3}$ corresponding to a baryonic number of

$\eta > 1.4 \times 10^{-11}$, the rest of the mass being due to massive neutrinos; then there would be no constraint on the number of different types of neutrinos. But this is too extreme a lower limit for η, since the ^{3}He and D abundances provide a lower limit for the present density of the universe of about 10^{-31} g cm^{-3} corresponding to $\eta > 1.6 \times 10^{-10}$. Tben the maximum number of different types of neutrinos would be 5 if $\tau_{1/2} = 10.6$ minutes and $Y_p > 0.23$. (Note from editors: There issued a confused discussion which ended with the following summary remarks by Dr. Weinberg).

WEINBERG

As I understand it, it had been suggested that if the dynamically determined masses of galaxies, etc., are all dominated by massive neutrinos, then the baryonic number density might be very small, so small that the reactions which build up He^4 have trouble going to completion, so that we could tolerate a faster expansion rate and a larger n/p ratio at the time of nucleosynthesis. This would normally be expected to produce too much of the light elements, because the reactions which convert H^2 and He^3 to He^4 could not go to completion; but it was supposed that the excess H^2 might have been destroyed in stars. I gather from our discussion that this idea has now been given up because it has been realised that the destruction of this much H^2 would create too much He^3. Also there seems to be a consensus here that, even apart from problems of nucleosynthesis, Ω (baryonic) cannot be lower than 0.01 (or perhaps 0.005?) just on the basis of the stars we see.

WOLTJER

The role of the light element abundances in determining cosmological parameters is, at the moment, so important that any possibility one has to get around it should be explored. I want to consider the possibility that, in the early universe at redshifts of 10^3, 10^4 or even larger, a substantial cosmic ray flux was accelerated as a result of processes that may be associated with the formation of black holes or similar mechanisms. Suppose the cosmic rays consist entirely of hydrogen and helium. Then one can look at interactions between this flux and the cold gas which also consists only of hydrogen and helium. In the course of these interactions, γ-rays, ^{3}He and D will be produced and eventually the high particles lose energy by electrostatic interactions (i.e. ionisation losses) with the surrounding medium.

This whole scenario only works if deuterium formation takes place before

any heavy elements have been produced because, if it happens later, excessive abundances of the light elements would be produced, which we do not want. Typically, you need a number of Gev's of cosmic ray energy for every deuteron you form and, if you require an abundance of 10^{-5}, you have to have a cosmic ray energy of about 10 keV per present proton in the universe. This is not different from the energy you need to explain the X-ray background radiation. The epochs at which this process would work are constrained by the requirement that the fragmentation mean free path should not be too long. This means red shifts greater than 10^3. A similar restriction comes from the fact that one does not want to produce an excessive flux of γ-rays. The γ-rays must, therefore, be absorbed, which again requires redshifts of 10^3 or greater. The process cannot take place much earlier, else there will be much pair production and the particles are killed very rapidly. I am indebted to Rees for pointing out that similar scenarios have been considered by Epstein (*Ap. J.*, **212**, 595, 1977).

One may wonder how probable such processes are but, nevertheless, before coming to absolutely definite conclusions about the canonical model for the production of D and ^{3}He, every alternative ought to be explored.

SILK

Did you mention the alpha-alpha problem?

WOLTJER

There are alphas in the cosmic rays and that is why you have to make them sufficiently energetic to ensure that you get fragmentation in virtually all interactions and that you do not get the production of lithium.

AUDOUZE

You might have some problems with your interesting proposal before your process occurs. The argument goes the following way:

(i) If the universe has a low present density, there is a lot of D available produced by the standard Big Bang nucleosynthesis and there is no reason to worry about subsequent D production by "cosmological" cosmic rays.

(ii) If the present density is higher, then the standard Big Bang produces too much ^{7}Li [especially if the $\left(\frac{^7\text{Li}}{\text{H}}\right)_{\text{primordial}} = 10^{-10}$ is the actual value].

And then you encounter the difficulty with an overproduction of ^7Li before the occurrence of your proposed process.

WOLTJER

Nevertheless an additional mechanism for D + He^3 production relaxes the Big Bang conditions to some extent. In addition it may facilitate other scenarios where the He^4 does not come entirely from the Big Bang either.

OSTRIKER

It is interesting to note that the energy per particle which you require is not far from the cosmic ray energy produced by a supernova in our Galaxy divided by the number of particles in the ejected mass.

WOLTJER

Of course you have to be careful then not to produce heavier elements at the same time, which could cause difficulties for fragmenting into lighter elements.

FUNDAMENTAL TESTS OF GALAXY FORMATION THEORY

JOSEPH SILK
Department of Astronomy
University of California, Berkeley

ABSTRACT

Galaxy formation is a highly complex process, involving both gravitational and dissipational interactions. Consequently, it is difficult to propose observational criteria that can clearly distinguish between rival theories. While the gravitational aspects of galaxy formation theory clearly differ between the isothermal and adiabatic fluctuation scenarios, the non-linear processes of hierarchical clustering, on the one hand, and gaseous pancake fragmentation, on the other, result in models for protogalaxy evolution that are remarkably similar. Both gaseous and gravitational processes play important roles, as is indeed essential if many of the characteristic properties of galaxies are to be explained. Only by examining the earliest stages of galaxy formation and clustering, where gravitational effects predominate, can one hope to assess the validity of a particular theory. This is especially apparent since it is the initial conditions in the early universe that determine much of the subsequent evolution, and distinguishing between theories is best achieved by attempting to infer the initial conditions.

Three tests are reviewed here that utilize the large-scale structure of the universe as an environment where traces of the seed fluctuations from which galaxies formed may be sought. Out to a scale of about 20 Mpc one can study the dynamics and structure of the Local Supercluster of galaxies, the density contrast of which is barely into the non-linear regime. On larger scales, out to about 100 Mpc, one can examine the large-scale matter distribution, which appears to contain numerous filamentary and shell-like structures and large holes. Finally, up to and even beyond the dimensions

of our horizon, the large-scale angular anisotropy of the cosmic microwave background radiation can probe the spectrum of density fluctuations in the matter distribution both now and at a very early epoch in the universe. At present, an assessment of these tests also depends on certain key aspects of particle physics: the role of grand unified theories in suppressing primordial entropy fluctuations and the rest-mass of the neutrino.

1 - Introduction

The theory of galaxy formation necessarily spans the entire evolution of the universe, from the singularity at $t = 0$ to the present epoch. Such a wide range is unavoidable because of the secular nature of the growth rate for gravitational instability in an expanding system. The initial conditions play an important role in the outcome of the instability, and any self-consistent theory must therefore take account of the evolution of fluctuations after the Planck time at $t = 10^{-43}$ s and even during the quantum gravity era.

Needless to say, while cosmologists are beginning to grope towards a theory for the initiation of fluctuations at these very early epochs, it is difficult to expose such ideas to any critical observational test. The emphasis here will be on observational tests of fundamental aspects of galaxy formation theory. Therefore, we begin by reviewing the evolution of the density fluctuation modes that lead to the eventual formation of matter inhomogeneities (Sec. 2). We discuss how the resulting clumps develop into galaxies and clusters of galaxies, acquiring characteristic masses, velocity dispersions, and metallicities (Sec. 3). To distinguish between rival galaxy formation theories, one has to probe the earliest accessible stages of evolution. Cosmologists are ingenious, and the highly non-linear phases where both dissipative and gravitational processes play important roles offer ample opportunity for explanations of most characteristic properties of galaxies. In Sec. 4 tests are described that utilize the large-scale structure of the universe, including the dynamics of the Local Supercluster, the large-scale matter distribution, and the anisotropy of the cosmic background radiation. Finally, the role of particle physics is described with regard to its observable implications for galaxy formation. One of the most important effects is that of a substantial neutrino rest-mass, causing (if confirmed) a drastic revision in the observational implications of the gravitational instability theories of galaxy formation.

2 - Gravitational aspects

Density perturbations of the early radiation-dominated Friedmann universe comprise three modes, two growing in proper time elapsed since the singularity and one that is approximately constant. While this statement presumes a specific choice of coordinate gauge (the synchronous gauge will be adopted in what follows), no observable aspects of the theory should be gauge-dependent. The growing modes involve metric perturbations while the constant mode does not (as $t \to 0$). One usually refers to the growing modes as curvature or adiabatic perturbations (since the total energy density is perturbed), whereas the constant mode is an entropy or isothermal perturbation (since only the baryon density is perturbed). At late times in the evolution of the universe only the most rapidly growing adiabatic mode and the isothermal mode are of interest.

Other types of perturbations of the Friedmann universe can be expressed in terms of vector or tensor eigenfunctions of the 3-space Laplacian operator that determines the spatial structure of the perturbed metric coefficients, and correspond to either vorticity or gravitational wave modes. Of these, primordial vorticity fluctuations have been advocated as seeds for galaxy formation by Ozernoi and co-workers. The isotropy of the cosmic microwave background radiation on small angular scales severely constrains such theories, which require rotational velocities that are a substantial fraction of c. The ensuing discussion therefore considers only the scalar mode of density perturbations.

It is helpful to discuss separately the isothermal and adiabatic perturbations. One might expect the two modes to both be generated *ab initio* given sufficiently general initial conditions, although baryosynthesis may suppress the isothermal mode (Sec. 5). Insofar as the linear theory is concerned, there is no loss of generality that may arise in neglecting any coupling between the modes. Throughout the ensuing discussion, Ω denotes the ratio of mean cosmological density of matter to the critical value for closure, $3\,H_0^2/(8\pi\,G)$. The Hubble constant adopted is $H_0 = 100\ \mathrm{kms}^{-1}\ \mathrm{Mpc}^{-1}$; scaling to other values of H_0 will occasionally be indicated by the scaling parameter $h \equiv H_0/100\ \mathrm{km}^{-1}\ \mathrm{s}^{-1}\ \mathrm{Mpc}^{-1}$.

A. *Isothermal Fluctuations*

These fluctuations commence to grow after the decoupling epoch at redshift $z \approx 1000$ on scales somewhat in excess of the Jeans mass at that

time, $10^{5.8}\ \Omega^{-1/2}\ M_\odot$. Once the fluctuations enter the non-linear regime and form self-gravitating clouds, they cluster together hierarchically to eventually build up into systems of galactic mass. These clouds must avoid catastrophic cooling and gravitational collapse and survive for a period of at least 10^8 yr, the minimum dynamical time-scale for forming a galaxy. Internal star formation provides the most likely means of stabilizing these clouds. Only if the clouds are predominantly gaseous at the epoch of galaxy formation can one account for many of the observed features of galaxies (Sec. 3).

An independent inference of the mass of the first clouds that underwent hierarchical clustering may be obtained from N-body simulations of galaxy clustering, from which correlation functions have been derived. Comparison with the observed galaxy covariance function restricts the initial conditions at tbe decoupling epoch, if a power-law spectrum of fluctuations is adopted, to be of the form

$$\delta\rho/\rho \approx (M/M_0)^{-1/2\ -n/6}\ ,\ 0 \gtrsim n \gtrsim -1\ , \tag{1}$$

with $M_0 = 10^8$ to $10^9\ M_\odot$ (Gott, Aarseth and Turner 1979; Efstathiou and Eastwood 1981). Such clouds could have been made primarily of dark matter, perhaps formed by fragmentation, and clustered dissipationlessly to form the potential wells in which gas infall and dissipative galaxy formation subsequently occurred (White and Rees 1978).

Once the fluctuations become non-linear and cluster hierarchically, one obtains an approximately constant potential energy per unit mass for the gravitationally clustering matter only if $0 \lesssim n \lesssim 2$, since over scale R

$$GM/R \propto R^{(1-n)\ (n+5)}\ . \tag{2}$$

In fact, a recent analysis of the dynamical masses involved in galaxy pairs, groups and clusters, identified by a statistical analysis of a sample of galaxies with redshifts that is complete (at least for the luminous galaxies) out to $\sim$ 100 Mpc, indicates that the virial mass per galaxy increases approximately linearly with scale (Press and Davis 1981). This result would require $n \approx -3$ if self-similar hierarchical clustering is responsible for the distribution of dark matter on scales from 0.1 to 10 Mpc and the luminous matter distribution has approximately constant surface density over systems of galaxies in this range of scales. Dissipationless clustering also fixes the epoch of galaxy formation at $z \sim 10$ to 100 in order to obtain the observed mean densities and potential wells characteristic of galaxies.

B. *Adiabatic Fluctuations*

Surviving adiabatic fluctuations are limited by radiative diffusion and viscosity prior to decoupling to bave masses in excess of $\sim 3 \times 10^{13}$ $\Omega^{-5/4}$ $M_\odot$. Prior to the decoupling epoch adiabatic density fluctuations have amplified on scales above the Jeans length, which is close to the horizon size in the radiation-dominated era. One can assume that fluctuations on smaller scales will have acquired random phases. Consequently, an initial power-law spectrum of fluctuations with index n,

$$|\delta_k|^2 \propto k^n \ , \tag{3}$$

where

$$\delta_k = \int (\delta\rho/\rho)\ e^{i\underset{\sim}{k}\,\underset{\sim}{r}}\ d^3k \tag{4}$$

and k/a is the comoving wavenumber with a (t) the cosmological scale factor, will have flattened after decoupling on scales less than the horizon size at this epoch ($\sim 10^{18}$ $M_\odot$) to $n' = n - 4$. The residual spectrum after decoupling has the approximate form

$$|\delta_k|^2 \propto k^{n-4} \exp\ (-k/k_d) \ , \tag{5}$$

where the damping comoving wavelength scale is given by

$$2\pi/k_d \approx 6\ \Omega^{-3/4}\ \mathrm{Mpc} \ . \tag{6}$$

The results of a calculation of $|\delta_k|$, involving simultaneous solutions of the linearized and coupled Boltzman photon transport and gravitational field equations (Wilson 1981), are shown in Figure 1. The adopted primordial spectrum is a power-law with $n = 4$: this gives tbe minimal level of large-scale structure expected (Sec. 4).

Fluctuations on such large scales cannot have gone non-linear until a relatively late epoch: $z \lesssim 10$. The masses of adiabatic fluctuations are many Jeans masses, pressure forces being completely negligible prior to the collapse. Consequently, even small deviations from spherical symmetry will amplify once the collapse goes non-linear, although they will have been preserved in the linear stages (Barrow and Silk 1981). Tidal torques lead to generation of anisotropy when $\delta\rho/\rho \sim 1$ on galactic scales, but are of lesser importance on very large scales. Zel'dovich (1970) has argued that the generic type of collapse will be one-dimensional, initially leading to formation of a thin pancake. This is similar to the well-known flattening

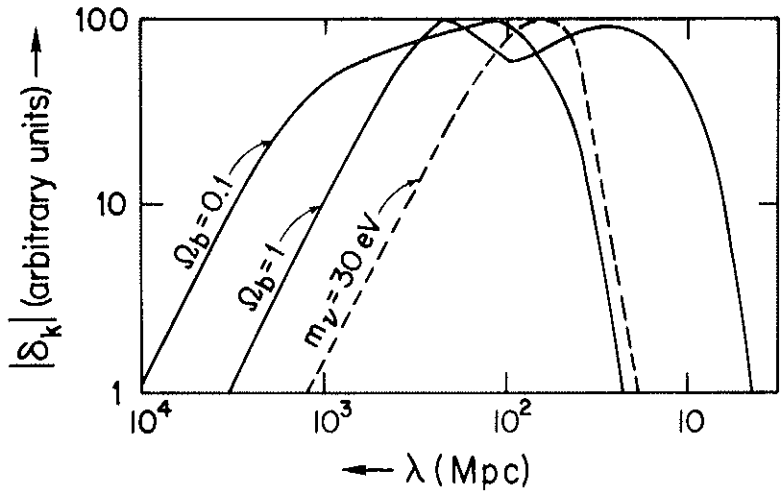


FIG. 1. Density fluctuation spectrum $|\delta_k|$ as a function of $\lambda = 2\pi a/k$ for the $\Omega_b = 1$ and 0.1 (subscript b denotes baryonic) adiabatic fluctuation models with $n = 4$. Also shown is the neutrino adiabatic fluctuation spectrum ($m_\nu = 30$ eV) with $n = 4$ as computed by Bond and Szalay (1981). Normalization is arbitrary and $H_0 = 50$ km s^{-1} Mpc^{-1}; the baryonic models scale with $\Omega H_0^2 = \text{constant}$.

instability exhibited by the collapse of a weakly spheroidal pressure-free uniform cloud.

As infalling material accumulates in the midplane, cooling must eventually occur and the pancakes will be unstable to fragmentation. There is some controversy over whether any purely gravitational fragmentation occurs prior to pancake formation (Doroshkevich and Zel'dovich 1975; Juskiewicz 1981). We shall ignore this possible complication, which arises from non-linear mode-mode coupling and is unlikely to decrease the minimum fragmentation length scale hy more than a factor of two relative to the damping scale.

The pancake structure is determined as follows. The trajectory of a fluid element initially at comoving coordinate ζ measured perpendicular to the pancake plane is given, for a one-dimensional pressure-free collapse, by

$$x(t, \zeta) = a(t) \left\{ \zeta - k_d^{-1} \frac{a(t)}{a_p} \sin(k_d \zeta) \right\}. \tag{7}$$

This non-linear solution combines the Hubble flow (in the first term) with the influence of a linearized plane wave density perturbation of the Hubble flow (in the second term) and can be shown to be consistent both with the equations of motion and Poisson's equation (Zel'dovich 1970). Mass conservation then yields

$$dM = \rho_0 \, a \, d\zeta = \rho(z, t) \, dz$$

where ρ_0 is the background density. Consequently

$$\rho = \rho_0 \, a \, |\partial z/\partial \zeta|^{-1} = \rho_0 \left(1 - \frac{a}{a_p} \cos k_d \zeta\right)^{-1}, \tag{8}$$

indicating that infinite density is attained in the midplane ($\zeta = 0$) at a redshift given by $a(t) = a_p$. This signifies tbe formation of a caustic surface which arises from the intersection of particle trajectories. In the vicinity of the caustic surface,

$$\rho = \rho_0 (k_d \zeta)^{-2} \, . \tag{9}$$

The range of ζ is $0 \leqslant k_d \zeta \leqslant \pi$.

At sufficiently high density, the gas cools and forms a thin dense layer at $T \sim 10^4$ K. A stand-off shock on either side of this layer separates it from the infalling gas. The velocity with which the infalling material enters the shock is

$$\dot{z} \approx \dot{a}_p \zeta = (2/3 \, t_p) \, a_p \zeta = (GM/r_0)^{1/2} \, (2 \, k_d \zeta/\pi) \, , \tag{10}$$

where

$$r_0 = 2\pi \, a/k_d \, , \; M = (\pi\rho_0/6) \, (k_d/a_p)^3 \tag{11}$$

are the initial comoving radius and mass of the pancaking cloud and ρ_0 is the background density at t_p. The mass-fraction ζ_c that constitutes the cold layer is inferred from the condition that the cooling time scale of the post-shock gas be less than the dynamical time-scale:

$$t_c \equiv 3 \, n \, k \, T/L_{ff} \, n^2 \, T^{1/2} < \frac{1}{4} z/\dot{z} \, , \tag{12}$$

where n is the post-shock density (equal to four times the pre-shock density), L_{ff} is the usual free-free cooling coefficient for gas of primordial composi-

tion, and the temperature of the shocked gas satisfies

$$kT/\mu = \frac{3}{16}\dot{z}^2 \ . \tag{13}$$

Using (9) and (10), this yields $k\,\zeta_c/\pi \approx 0.25\,(5\ \mathrm{Mpc}/r_0)^{1/3}$, indicating that as much as 25 percent of the initial mass may have cooled and been compressed into a thin layer.

Since only the cooled gas fragments and forms galaxies, this leads to a clear prediction of the pancake model: there should be at least a factor four more hot gas in galaxy clusters than baryonic mass in galaxies.

Observationally, one finds M/L for rich clusters to be of order 400 ($H_0/100$ km s^{-1} Mpc^{-1}), with $\sim$ 10 to 20 percent of this accounted for by hot X-ray emitting gas. Since the luminous regions of galaxies have M/L $\sim$ 5 to 10 (Faber and Gallagher 1979), the pancake theory prediction will be satisfied provided that both galaxy halos and the dark matter in galaxy clusters are non-baryonic. One would not be able to reconcile this theory with observation if halos and clusters were predominantly composed of dissipative material, such as compact stellar remnants, which had not undergone collapse prior to pancake formation.

The minimum fragment mass turns out to be relatively small. The cool gas layer is ram-pressure confined, and tbe minimum mass that is gravitationally unstable is given by the Bonnor-Ebert criterion:

$$M_{min} = 1.18\ v_s^4\ G^{-3/2}\ p^{-1/2} \ ,$$

where v_s is the sound velocity and p is the ram pressure. Now

$$p = \frac{3}{16}\rho\,\dot{z}^2$$

and is independent of ζ (for $k_d\,\zeta \ll 1$). Consequently, one obtains

$$M_{min} = 10^7\,(T/10^4\ K)^2\ (5\ Mpc/r_0)\ (10/z_p)^3\ \Omega^{-1}\ M_\odot \ .$$

It seems unlikely that one could avoid fragmentation down to M_{min}. Galaxy formation will presumahly occur by the aggregation of many smaller clouds

of mass 10^7 to 10^9 $M_\odot$, much as in the primordial isothermal fluctuation model. The principal difference now is that the clouds originate by fragmentation at the relatively late epoch of pancaking, and so perhaps are more likely to be predominantly gaseous as galaxies form.

3 - Dissipational aspects

Both of the extreme viewpoints described above, involving either primordial isothermal or adiabatic fluctuations, result in the formation of clouds of characteristic mass 10^6 to 10^9 $M_\odot$. These are the precursors of the galaxies. The idea that a protogalaxy consists of a collection of interacting gas clouds is not new: it has formed a central role in such models as those of Peebles and Dicke (1968), Larson (1974) and Gott and Thuan (1976). Earlier discussions have however glossed over a simple point, namely the physical mechanism by which such primordial clouds are supported against collapse and ensuing fragmentation and star formation.

To illustrate this problem, let us note that the free-fall time of a self-gravitating spherical cloud of mass $M_6 \equiv M/10^6$ $M_\odot$ and virial temperature $T_4 \equiv T/10^4$ K (including both thermal and turbulent contributions) is 9×10^6 $M_6 T_4^{-3/2}$ yr, whereas the time-scale over which cloud aggregation occurs during galaxy formation is $\gtrsim 10^8$ yr, the dynamical time-scale for a typical galaxy. In order for the clouds to remain gaseous, avoiding cooling and collapse, an internal energy source is required. A similar problem is encountered for interstellar molecular clouds, whose lifetimes significantly exceed collapse time-scales. Ongoing low-mass formation apparently provides a substantial momentum and energy input via protostellar winds (Norman and Silk 1980), and it is logical to assume that a similar phenomenon can energize the primordial clouds.

It is, of course, necessary to demonstrate that primordial star formation is not disruptive and, moreover, can be self-regulating. One can show that even the coherent action of successive supernovae need not destroy clouds of mass 10^6 to 10^7 $M_\odot$ within 10^9 yr, if the supernova rate is taken to be proportional to the star formation rate and constant in time (Silk and Norman 1981). In principle, one might expect the star formation rate to be self-regulating; as the energizing pre-main-sequence phase ends for the most massive stars, the cloud can cool, begin to collapse, and make more massive stars, which in turn inhibit further collapse for some

10^6 to 10^7 yr. In this manner several generations of stars can form and evolve.

The characteristic mass of the first generation of stars must be several solar masses in order to avoid an excess of low metallicity stars at present. The primordial clouds will inevitably become enriched, since whatever the slope of the initial mass function, some extremely massive stars will have evolved, although the amount of enrichment is very uncertain and possibly variable from cloud to cloud. This suggests that galaxies form from an aggregation of preenriched gas clouds containing a considerable number of stellar remnants as well as surviving stars of lower mass. Theoretical arguments suggest that, once the enrichment of a cloud exceeds that of extreme Population II, stars will form with a more or less conventional mass function (Silk 1980). After substantial numbers of low mass stars have formed, the gas reservoir will rapidly be depleted. The end product of the evolution of such an isolated cloud may resemble a dwarf spheroidal galaxy or maybe a globular star cluster.

In the inner regions of a protogalaxy, frequent collisions between clouds will inevitably occur. These will have interesting consequences. One can consider two regimes. First, in the core of the galaxy, direct collisions between clouds result in coalescence of the gas if the collision occurs at sufficiently low velocity. Any embedded stars will be released, however. The shocked gas will be stimulated to form additional stars, and one expects that a cloud cannot survive more than one or two collisions before the gas supply is exhausted. Outside the core, dynamical interactions between clouds are likely to be more important than direct collisions, and the cloud distribution will undergo violent relaxation following the initial collapse of the system. In regions of high galaxy density, mergers between protogalaxies will refuel the cores with a fresh supply of clouds.

In isolation, a protogalaxy gradually dissipates its kinetic energy via cloud-cloud collisions. Angular momentum is conserved and a disk forms, having collapsed by a factor $\sim \lambda^{-1}$ in the presence of a preexisting dark halo, where the dimensionless parameter $\lambda \equiv JE^{1/2}\, G^{-1}\, M^{-5/2} \approx 0.07$ due to tidal interactions between protogalaxies of mean mass M, total energy E and angular momentum J (Fall and Efstathiou 1980). The cloud collision-dominated core, defined by the condition $t_{\rm coll} < t_{\rm dyn}$ where $t_{\rm coll}$ is the cloud-cloud collision time-scale and $t_{\rm dyn}$ the local dynamical time, and formed by violent relaxation of the cloud distribution, will rapidly form stars as collisions release existing stars and form new stars. This inner region constitutes the central bulge. The newly formed stars whose formation

is triggered by cloud collisions will preferentially be produced towards the inner bulge regions because of the loss of kinetic energy in the collisions: consequently a metallicity gradient develops.

One suspects that the few surviving clouds will form the globular star clusters. Those clouds whose orbits take them into the cloud-collision-dominated core will acquire greater metallicity as a result of the first collisions with other clouds which enhance their internal star formation rate. In this way, one expects the orbital parameters of globular clusters to correlate with metallicity, in addition to the trend (with wide dispersion) in metallicity enhancement expected as a function of decreasing galactocentric radius. Note that halo stars (which are stars from disrupted clouds) should generally possess an inverse correlation between eccentricity and metallicity, whereas globulars should display a positive correlation.

Mergers between protogalaxies in the initial stages of clustering (before large virial velocities have been acquired) lead to S0 and eventually to elliptical galaxy formation. After each merger fewer and fewer clouds may be expected to survive that can potentially form a disk. Only two or three mergers are required to make an elliptical. This may be inferred from the dependence of morphological type on local density and the fact that the elliptical luminosity function characteristically peaks at about one magnitude higher luminosity than the spheroidal bulges of spirals (Dressler 1980).

One obvious weakness arises in this model. Galaxy groups are spiral-rich yet, since in the hierarchical clustering model clusters formed as groups merged together, one would expect a predominance of ellipticals and S0's in groups. A possible resolution is that groups are dynamically young systems, forming from galaxies that have already evolved beyond the environmentally vulnerable protogalaxy stage. Similarly, spiral-rich clusters would be envisaged as intrinsically younger systems than spiral-poor clusters, which in fact tend to be richer and more centrally concentrated. The pancake fragmentation model provides an alternative viewpoint: here, the great clusters form directly in denser regions undergoing a greater protogalaxy merger rate than the groups which form in less dense regions of the pancake.

Two other notable consequences for ellipticals are worthy of mention. Once the self-similar clustering hierarchy is broken as bound galaxy clusters develop, the specific angular momentum generally decreases whenever mergers occur from near parabolic orbits with very low impact parameter, as is expected during cluster formation. Bulges and low luminosity el-

lipticals are expected to be indistinguishable with respect to their dynamical properties. The core parameters for ellipticals are similarly defined by $t_{coll} < t_{dyn}$: this naturally leads to

$$M_{core} \sim (\sigma^4 G^2 \mu) (t_{coll}/t_{dyn}) \sim \mu R^2 (t_{dyn}/t_{coll}) , \quad (13)$$

where μ is the mean surface density of the colliding clouds and R is the radius of the core. One infers characteristic (M, σ, R) relations, more or less in accord with observations (cf. Faber and Jackson 1976).

Dissipation in the core because of cloud collisions, with a fresh supply of clouds being injected as a consequence of the post-merger violent relaxation, results in the formation of the luminous cores and metallicity gradients that are characteristic of most luminous ellipticals. At the same time the final structure retains a considerable velocity anisotropy as a consequence of the discrete nature of the cloud collisions. It is this hybrid nature of a cloud collision model for protogalaxy formation that enables it to be so successful in at least crudely accounting for many of the observed characteristics of galaxies. Metallicity gradients, correlations involving metallicity, luminosity and velocity dispersion, and rotational properties: all of these characteristic properties are produced during the protogalaxy stage. Parenthetically, one may remark that it is difficult to see how mergers between conventional, relatively gas-poor spirals could form a luminous elliptical. However, the successes of the protogalaxy model cannot be used to distinguish between rival theories of galaxy formation, since many aspects of the model are common to these theories. Indeed this is inevitable, since many of the observed properties of galaxies point to a hybrid dynamical and dissipative model of galaxy formation.

4 - Tests of galaxy formation theory

One often cited discriminant of galaxy formation theory is the discovery of protogalaxies. According to the traditional argument, the isothermal fluctuation theory predicts protogalaxies at $z \sim 100$, whereas in the adiabatic fluctuation theory, protogalaxies form at $z \sim 10$. Unfortunately, this argument is too simplistic. The environmental influences described in the previous section compel us to associate the bulk of elliptical galaxy formation with the epoch of galaxy clustering. The continuous range of properties between bulge components and ellipticals suggests that these too must have mostly formed at a similar epoch. The principal modification required of the isothermal scenario for this delay in galaxy

formation to occur is one recognized at the outset: in order for globular clusters to display a metallicity gradient, they must be predominantly gaseous at the epoch of galaxy formation. Similarly, in order for galaxies to acquire such properties as systematic dependence of morphological type on local density and correlations between luminosity, velocity dispersions, and metallicities, they must also be extremely gas-rich at the epoch of galaxy clustering. In other words, protogalaxy formation must inevitably occur at $z \lesssim 10$. Of course, some protogalaxies could form earlier: detection of even a single protogalaxy at $z \gg 10$ would be strong evidence for the isothermal theory.

There is some evidence that formation of low mass galaxies is occurring at the present epoch. Compact galaxies such as I Zw 18 are presently undergoing a vigorous burst of star formation, yet have only acquired an extremely low metallicity (Lequeux and Viallefond 1980). It appears that certain of these gas-rich extragalactic HII regions are presently undergoing their first star formation burst, otherwise excessive metallicity would have been produced. A related phenomenon may be the young globular clusters around the Large Magellanic Cloud. Indistinguishable in appearance from Milky Way globulars, these systems are very blue, and some have ages $\lesssim 10^8$ yr.

None of this can he said to favor any particular galaxy formation theory. Formation of isolated low mass galaxies at the present epoch is expected in both the isothermal and adiabatic fluctuation models. A small number of the primordial clouds produced either by Jeans instability at $z \sim 1000$ or by pancake fragmentation may be expected to survive against internal star formation and external triggering until a very late epoch. To obtain a more fundamental probe it is necessary to consider structure on much larger scales, where non-linear processes have not yet had time to mask traces of the initial conditions. Three possible regimes may be examined, where the predictions of hierarchical clustering and gaseous fragmentation theory could reasonably be expected to differ.

A. *The Local Supercluster*

The dynamics and structure of the Virgo Supercluster out to some ~ 20 Mpc from Virgo affords a possible test of galaxy formation theory in a regime where the linear or mildly non-linear theory might be expected to be applicable. Certainly for the isothermal clustering theory the mean flow field has been modelled (Silk 1974; Peebles 1976) and found to be

in reasonable accord with data from redshift surveys (Tonry and Davis 1981; Yahil, Sandage and Tammann 1980).

This conclusion is based on a spherical model for the Virgo Supercluster. Despite the considerable concentration of galaxies towards the supergalactic plane, the flattened distribution does not significantly modify the Virgocentric flow unless the collapse is extremely one-dimensional, as in the adiabatic pancake theory. In this case, the infall velocity relative to Virgo tends to be underestimated because of the neglect of certain nonlinear terms when a spherical model is adopted (Szalay and Silk 1981). In a less extreme situation, modelled by the collapse of a uniform oblate spheroid (Barrow and Silk 1981), the density contrast at turn-around of the shortest axis is found to be reduced by up to a factor of three relative to the usual value of $(3\pi/4)^2$ that applies in the spherical case. A similar model was used by White and Silk (1979) to reconcile our motion relative to the cosmic microwave background radiation with a Virgocentric flow model: a weakly triaxial collapse with a modest amount of shear was indicated.

The Hubble flow must remain relatively uniform in the supergalactic plane while collapse has occurred along the perpendicular axis. This is required by the redshift observations of nearby galaxies which are observed to show little deviation from a uniform Hubble flow. In order for a pancake model to be viable for the Virgo Supercluster, our galaxy must have formed in the initial collapse phase, while the Hubble flow has largely been preserved in the supergalactic plane. The plausibility of this requirement has been recently investigated by means of N-body simulations of pancake collapse (Dekel 1982, Dekel and Szalay 1982). These show that the virialization time can greatly exceed the initial minor axis collapse time. A uniform Hubble flow in the midplane is preserved for several initial collapse times, consistent with pancake collapse and the principal epoch of galaxy formation occurring at $z > 5$. Our galaxy itself remains in the Supergalactic plane because the matter from which it formed was centrally located (along with the bulk of the matter) and initially acquired only a modest velocity during the collapse. Part of this may have been dissipated; in any event, the orbit of our galaxy stays predominantly coplanar for a considerable time. Evidently, the galaxy flow field in the Virgo Supercluster is consistent with models based on either the primordial isothermal or adiabatic fluctuation scenarios.

There are several additional difficulties that arise in the pancake model applied to the Virgo Supercluster. First, the predicted flattening is

appreciable, even for a dissipationless model. It is considerably greater after even two collapse times have elapsed than that inferred for the distribution of RSA galaxies in the Local Supercluster, for which the mean flattening is 2 or 3 to 1 (Yahil, Sandage and Tammann 1980). However, the study by Tully (1981) of the nearby spiral galaxy distribution indicates an interesting new aspect that may be consistent with a slight modification of the pancake model. Tully finds that the Local Supercluster consists of a thin disk (flattening 6:1) in the supergalactic plane together with a number of elongated clouds of galaxies above the supergalactic plane. The disk component can be identified with the pancake model prediction: one would then have to argue that the supergalactic halo clouds of galaxies are the remnants of smaller pancakes.

This latter result provides a means of understanding a second objection to the pancake model, namely the small deviations from a Hubble flow observed even well above the supergalactic plane, since the galaxy clouds, originally noted by de Vaucouleurs (1975), themselves participate in the Hubble flow. However a final difficulty is now even more prominent. This concerns the pancake mass. With $\Omega \approx 0.1$, the minimum pancake mass is $\sim 10^{15}\ M_\odot$, and the mass distribution is fairly flat up to $\sim 10^{16}\ M_\odot$. Yet the Virgo Supercluster mass is $\sim 10^{14}\ M_\odot$, and the masses of the de Vaucouleurs-Tully clouds are only $\sim 10^{13}\ M_\odot$. One can only rescue the situation if $\Omega \sim 1$. Alternatively, pancakes may have fragmented by a factor $\gtrsim 10$ in mass due to mode-mode coupling induced fluctuations, or the neutrino may have a rest mass in excess of 10 eV. The pancake theory undergoes a very considerable modification in a neutrino-dominated cosmology. One possible outcome, perhaps demanded by the Local Supercluster studies, is the formation of much lower mass pancakes.

B. *The Large-Scale Matter Distribution*

Recent redshift surveys reveal the presence of large filamentary superclusters and correspondingly large holes in the galaxy distribution, over scales of 40 to 100 Mpc (Davis *et al.* 1982; Tarenghi *et al.* 1980; Gregory and Thompson 1978; Einasto *et al.* 1980; Kirshner *et al.* 1981). While it is difficult to quantify this effect, it does seem apparent that the N-body clustering simulations fail to reproduce much of the fine-structure revealed by the data in redshift space (Efstathiou and Eastwood 1981). The two-point galaxy correlation function for the simulations likewise differs from that for the observations, not being well fitted by a single power-law in distance scale. The galaxy correlation function inferred from large-scale

studies is well described by a power-law of slope -1.8 over scales $\lesssim 10\ h^{-1}$ Mpc.

The pancake fragmentation theory, at least qualitatively, can reproduce the observed filamentary and cell-like structure of the galaxy distribution. Existing simulations are, however, only two-dimensional (Doroshkevich *et al.* 1980a), and it is not yet possible to compare the model predictions in any detail with the data. Dissipation certainly improves the appearance of the simulations, and provides a means of obtaining large holes without simultaneously requiring excessively large random velocities for the galaxies. On the other hand, one has to bear in mind that initial conditions can play an especially important role in the isothermal theory, where large-scale fluctuations have not acquired random phases and, moreover, exert little feedback on the space-time curvature. Consequently, one could plausibly argue that suitably chosen initial density fluctuation profiles can lead to the formation of holes of any desired size (Olson and Silk 1979), and large-scale shape anisotropies, including holes, could arise from a non-Gaussian distribution of initial fluctuations (c.f. Barrow and Silk 1981); in this case, arbitrarily large holes would be expected as larger and larger volumes of space are surveyed.

Perhaps the best quantitative means at present of comparing the observed large-scale distribution with theory is by examining the linear regime. Here, the theoretical predictions are straightforward, and while the evidence for any positive or negative features in tbe correlation function on scales $\gtrsim 10$ Mpc is unclear, there nevertheless is a reasonably well-defined upper envelope with which one can compare the model predictions.

To commence, let us represent the distribution of primordial density fluctuations by a power-law Fourier power spectrum. A power-law dependence on wave number is chosen in order to avoid introducing any preferred scale, although in a cosmological model with spatial curvature, for example, there is indeed a natural scale associated with the curvature radius. However, this is sufficiently large that it only affects predictions of the large-scale anisotropy of the cosmic background radiation. Restricting the present discussion to a spatially flat cosmological model, we infer from equation (1), which expresses the fluctuation spectrum as a power-law in the comoving mass-scale, that the case of constant gravitational potential (or equivalently, constant metric or curvature) fluctuations corresponds to setting $n = 1$, and we recover $\delta\rho/\rho \propto M^{-2/3}$.

This spectrum is considered likely to emerge naturally from quantum gravity considerations. On the other hand, the choice $n = 0$ corresponds

to white noise. This is believed by some cosmologists to be a natural value that might arise from an unspecified mechanism for spontaneously generating density fluctuations (such as a first-order phase transition in the very early universe). Any such process would have to be causal, and it is of interest to extrapolate back to very early epochs, when the horizon scale contained less mass than that of any structure of interest in the present universe. For example, the mass of a galaxy was first encompassed by the horizon at $z \approx 10^8$. Now density fluctuations grow in amplitude on scales greater than the Jeans length and, in the radiation-dominated era, this is effectively the horizon scale. The growth rate in this regime is $\delta\rho/\rho \propto t$ for the fastest growing mode, where t denotes proper time. Strictly speaking, this result holds for curvature fluctuations in a specified coordinate gauge, such as the synchronous gauge. Since the scale-factor $a \propto t^{1/2}$, at a given comoving wavenumber k/a, the spectrum of density fluctuations must flatten as the specified mass-scale enters the horizon, where growth is suppressed due to radiation pressure. One can readily see that the spectrum in fact flattens by the 2/3 power in mass. Consequently, a primordial density fluctuation spectrum $\delta\rho/\rho \propto M^{-7/6}$ on scales larger than the horizon will be associated with a white noise spectrum within the horizon. One can show that such a spectrum, corresponding to $n = 4$, also represents the minimal fluctuation level expected to arise from any causal rearrangement of the matter distribution into non-linear clumps (Peebles 1980). Finally, a minimum value of n arises from requiring that the r.m.s. density fluctuations do not diverge on large scales. This convergence criterion yields $n > -3$.

The only "natural" spectrum to emerge from this discussion was the one that initiated it: $n = 1$. The gravitational potential (metric) fluctuations diverge strongly if $n \neq 1$: on large scales if $n < 1$ and on small scales if $n > 1$, and in either case, the constant index power-law assumption necessarily becomes invalid. What one can hope to do eventually is to directly measure n in different regimes, bearing in mind that even if the seed fluctuations are arranged at some epoch t_1 into discrete non-linear clumps, one expects that a power-law tail to the fluctuation spectrum with $n = 4$ will have been generated on scales $\gg ct_1$. Of course, in the absence of any fluctuation generating mechanism, there is little justification for the choice of a power-law fluctuation spectrum. Perhaps the most extreme alternative would be if the universe were to form discrete clumps of some specified mass. If the clumps are randomly distributed, as is likely

for a phase transition, then on scales larger than the clump mass but less than the horizon size, the $n = 0$ power spectrum describes the distribution of fluctuations in the gravitational potential. In any event, a spectrum that initially is a simple power-law is far from a power-law after the recombination epoch.

In the linear regime, the galaxy correlation function is defined by assuming random phases for the different Fourier components,

$$\xi(r) = \frac{32\pi^4}{V} \int_0^\infty k^2\, dk\, |\delta_k|^2 \sin(kr) / (kr) \propto r^{-n-3} \quad (kr \ll 1), \qquad (14)$$

where the integral approximates a Fourier series that is summed over a finite 3-torus of circumference $2L > ct$ much larger than the present horizon and $V = (2L)^3$. To normalize the amplitude of the density fluctuations, $\xi(r)$ is required to be unity at r_0, where one has from the recent redshift surveys (Davis 1981) that

$$\xi(r) = (r/r_0)^{-1.8}\ ,\ r_0 = 5\, h^{-1}\ \text{Mpc}$$

on scales $r \lesssim 10\ h^{-1}$ Mpc. Over larger scales, $\xi(r)$ appears to drop more rapidly than the extrapolated power-law. A similar but less reliable result is inferred from inversion of the angular correlation function (Peebles 1980).

The correlation function for primeval adiabatic density fluctuations has been evaluated for a range of values of n, and is compared with the observed 2-point correlation function for a sample derived from the Revised Shapley-Ames redshift catalogue that extends beyond the Local Supercluster (Rivolo and Yahil 1981) in Figure 2. About all one can say is that values of $n \lesssim 2$ are excluded. In particular, the spectrum $n = 1$ appropriate to constant curvature fluctuations gives excessive power on scales > 20 Mpc, far in excess of the ohserved $\xi(r)$. One needs $n = 3$ to 4 for consistency with the data (Wilson and Silk 1981; Peebles 1981). It is not possible to say anything about possible structure on large scales, since the plotted $\xi(r)$ is probably dominated by effects associated with the finite volume sampled at $r \gtrsim 50$ Mpc, other than that the apparent hole beyond the Virgo supercluster shows up clearly. Of course, with $n > 1$ the potential fluctuations diverge on small scales, and the assumption of a constant power-law index for the initial spectrum becomes invalid. For $n = 3$, one requires n to decrease to a value less than unity below a comoving scale of ~ 1 Mpc.

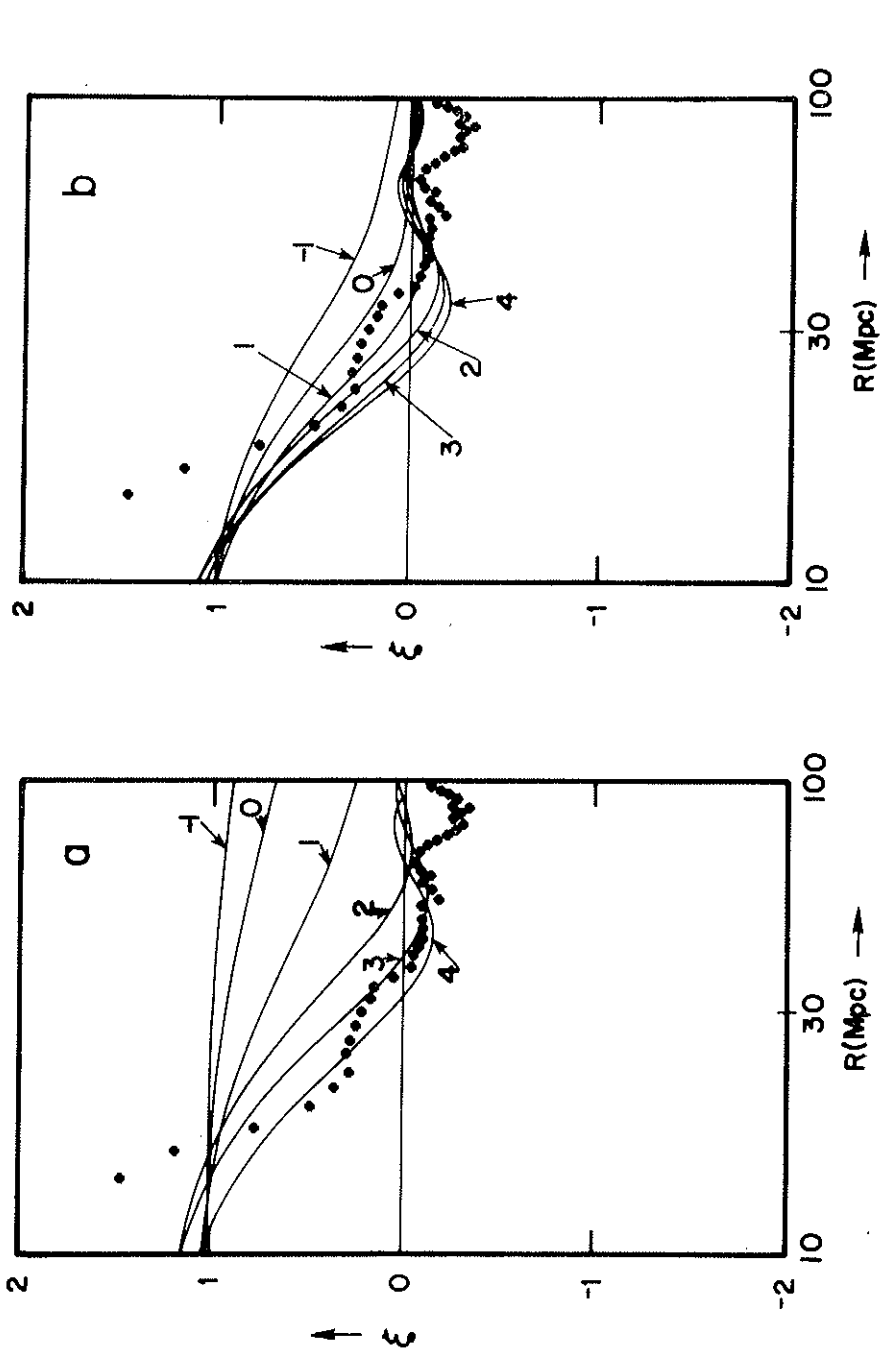


FIG. 2. Comparison of 2-point galaxy correlation function predicted by linear theory of adiabatic density fluctuations (Wilson and Silk 1981) with observed correlation function for the RSA sample. Data points are from Rivolo and Yahil (1981); the sample volume extends only to a depth of 80 Mpc. Two sets of predictions are illustrated: (a) $\Omega = 0.1$ in the standard model and (b) $\Omega_\nu = 0.98$, $\Omega_b = 0.02$ for a neutrino-dominated universe for various values of the initial power-law spectral index.

C. *Large-Scale Anisotropy of the Cosmic Microwave Background Radiation*

Over scales in excess of $\sim$ 100 Mpc the cosmic microwave background radiation provides a powerful probe of the matter distribution. In principle, so also does the cosmic X-ray background which originates at $z \sim 1$; however, a possibly insuperable problem with this enters in subtracting the locally inhomogeneous contribution at the ~ 1 percent level from our galactic halo. No such difficulty arises with the microwave background radiation, which provides an unimpeded view of the early universe to $z \gtrsim 10$, and most probably to $z \approx 1000$, the epoch of matter-radiation decoupling.

Anisotropy in the microwave background radiation that was present on the surface of last scattering of the radiation at $z=10$ to 1000 could not have been erased on angular scales above the corresponding horizon size at this epoch by any causal process. This angular scale is approximately $(\Omega/(1+z))^{1/2}$ radian or $2\,\Omega^{1/2}$ degrees at $z \approx 1000$. Of course, anisotropy is also inevitable at some level on smaller angular scales. Even if the intergalactic medium were reionized at $z \gtrsim 10$, and primordial structure erased, the ionization sources must have been inhomogeneously distributed and themselves subject to gravitational clustering: they would therefore have left an imprint on the radiation field. This does, however, provide a means of reducing the predicted amplitude of the small-scale radiation anisotropy, especially for primordial isothermal fluctuations, which would have become nonlinear at $z \gg 10$. A pure adiabatic fluctuation model is probably immune to this effect, however, since galaxy formation and the generation of potential sources of ionization occurred at $z \lesssim 10$ when the universe already possessed insufficient optical depth to rescatter the radiation.

The small angular scale structure of the cosmic microwave radiation has been the subject of many experiments. Only upper limits have hitherto been reported. The upper limits are sufficiently low that the adiabatic theory is in considerable difficulty, however. Over large angular scales positive detections have been reported of dipole and quadrupole anisotropy. These observations provide a fundamental means of evaluating the predictions of galaxy formation theory.

Very approximate estimates of the predicted anisotropy can be given as follows. There are three distinct contributions. First, the adiabatic coupling between matter and radiation generates temperature fluctuations

at the epoch of last scattering t_s, of order

$$(\delta T/T)_{t_s} = \frac{1}{3}(\delta\rho/\rho)_{t_s} \propto \theta^{-3/2-n/2} . \tag{15}$$

The angular dependence assumes a power-law density fluctuation spectrum of the form (1). This result assumes that the decoupling of matter and radiation is instantaneous, whereas in fact the residual ionization ($n_e/n_H \sim 10^{-5}$) after decoupling occurs is sufficient to give considerable scattering of small-scale fluctuations. In effect, the surface of last scattering is blurred, becoming much thicker for smaller-scale fluctuations. Only fluctuations on a comoving mass-scale $\gtrsim 10^{15}$ $M_\odot$, corresponding to angular scale

$$\theta = 10' (M/10^{15}\ M_\odot)^{1/3} \tag{16}$$

approach the adiabatic limit (15). Fluctuations of this type are smoothed on smaller scales. However, smaller angular scale temperature fluctuations are generated by the motions of the scattering inhomogeneities. These gravitationally induced motions result in secondary fluctuations for both primordial adiabatic and isothermal inhomogeneities of order

$$(\delta T/T)_{t_s} \sim (v/c) \sim (\delta\rho/\rho)_{t_s} (\ell/ct_s) \propto \theta^{-1/2-n/2} , \tag{17}$$

where ℓ is the fluctuation wavelength (and subtends an angle θ at the observer). A third contribution, dominant on very large angular scales, arises from the perturbed gravitational potential due to the presence of fluctuations on the surface of last scattering and at the present epoch. The corresponding temperature fluctuations are of order

$$(\delta T/T)_{t_s} \sim G\delta\rho\,\ell^2 \sim (\delta\rho/\rho)_{t_s} (\ell/ct_s)^2 \propto \theta^{-1/2-n/2} , \tag{18}$$

and are the dominant source of large angular scale anisotropy (on angular scales $\gg 2^\circ$, corresponding to the horizon scale at decoupling). The angular dependences given here assume a small-angle approximation ($\theta \ll 1$ radian).

In refining these estimates, several complications arise. Even if the adiabatic fluctuation spectrum is a simple power law as $t \to 0$, it will no longer be so on scales comparable to or less than the horizon after decoupling. Damping occurs on scales helow $\sim 3 \times 10^{13}\ \Omega^{-5/4}\ M_\odot$. Mass scales above the Jeans mass prior to decoupling ($\sim 10^{17}\ M_\odot$) have grown continuously,

whereas smaller scales have undergone acoustic oscillations. This results in a flattening of the initial spectrum. The abrupt drop in sound velocity at decoupling results in velocity overshoot and enhancement of density fluctuations as the potential motions on large scales can become highly supersonic. The residual ionization results in a Compton drag force that tends to nullify this effect (Press and Vishniac 1980; Silk and Wilson 1980). One can see from Figure 1 that this overshoot effect does amplify the fluctuation spectrum on large scales in the $\Omega = 0.1$ model, but damping rapidly takes over. For $\Omega = 1$, Compton drag dominates. The amplitude of the resulting temperature fluctuations, if these complicating factors are ignored, is

$$\delta T/T \sim 3 \times 10^{-4}\ \Omega^{-1}$$

over the largest scales to have attained density contrast unity by the present epoch, corresponding to about 10′ to 20′.

In a more sophisticated treatment, the linearized Boltzmann equation, coupled with the linearized gravitational field equations and conservation equations, is used to describe the evolution of the radiation and density fluctuations through the decoupling epoch (Peebles and Yu 1970; Silk and Wilson 1980). To compare with observations, typically performed using a beam-switching technique, the correlation function is evaluated for radiation fluctuations in directions separated by an angle θ,

$$(\delta T/T)^2(\theta) = < \delta T/T\,(\underline{x}, \underline{\gamma}_1)\, \delta T/T\,(\underline{x}, \underline{\gamma}_2) > \ ,$$

where $\underline{\gamma}$ is a unit vector describing the radiation direction, and $\underline{x}$ is the location of the observer. The average is taken over all space, and is therefore appropriate for a random point in space. One has to convolve $\delta T/T$ with the antenna beam response function to obtain a mean temperature fluctuation that is smeared by the beam on small angular scales. This yields $\delta T/T \propto \theta$ for $\theta \ll \sigma$, where σ is the beamwidth (assumed to be Gaussian). For $60^\circ > \theta \gg 2^\circ$, the angular dependence approaches $\delta T/T \propto \theta^{1/2\ -n/2}$.

A representative upper limit is (Partridge 1980)

$$\delta T/T < 2 \times 10^{-4}\ ,\ \theta = 9'\ ,\ \sigma = 1.'8\ .$$

A comparison of the predicted fluctuations with this upper limit has been given by Wilson and Silk (1981). The predicted anisotropy scales as Ω^{-1}. Provided that there has been no appreciable rescattering of the radiation since the epoch of decoupling, it was found that adiabatic fluctuations can

only be reconciled with the observational limit if $\Omega \sim 1$. For $\Omega \approx 0.1$, the adiabatic theory is untenable for any value of the power-law index n, whereas isothermal fluctuations can be reconciled, and indeed are predicted to be on the verge of detectability, if $\Omega \sim 0.1$ and $n < 0$. However fluctuations on an angular scale less than a few degrees can be smoothed by reionization of the intergalactic medium at $z_D \gg 10$.

To estimate the significance of this effect we may suppose that L is the comoving damping length associated with the secondary smearing, so that $L(1 + z_D)^{-1}$ is the depth of the last scattering surface. This surface subtends an angular scale θ_D given by

$$\sin \theta_D/2 = \frac{\sinh \{(-K)^{1/2} r/2\}}{\sinh \{(-K)^{1/2} \int_{t_D}^{t_0} dt\,(1+z)^{-1} H_0\}}$$

where $K = -1$ if $\Omega < 1$. Now Hogan (1981) shows that the smearing effect can be estimated if an effective antenna beamwidth $(\theta^2 + \theta_D^2)^{1/2}$ is used. In the most extreme case for $\Omega = 0.1$, if tbe universe is reionized at $z < 100$, the effective surface of last scattering is at $z_D \approx 40$ and $\theta_D \approx 5°$. Thus we can make use of Figure 3 to estimate how the constraint of Partridge's upper limit at 9′ is weakened in the $\Omega = 0.1$ model by reionization: evidently all adiabatic fluctuation models are now consistent with the small-scale anisotropy limits.

However, on intermediate angular scales this secondary smearing by reionization becomes relatively umimportant. An especially significant constraint comes from the observations of Fabbri *et al.* (1979) and Melchiorri *et al.* (1981), who report $\delta T/T = 3\ (\pm 0.7) \times 10^{-5}$ with $\sigma = 3°$ over $\theta = 6°$. If we conservatively take tbis as an upper limit, a comparison of the predicted fluctuations (Figure 3) implies that $n > 1$ for adiabatic fluctuations or $n > 0$ for isothermal fluctuations, even if $\Omega = 1$.

Reionization of the intergalactic medium at $z > 10$ smoothes out the structure in the radiation distribution induced at decoupling, but results in the generation of new fluctuations on the surface of last scattering. Suppose that these fluctuations have some preferred scale, which presumably is that of the HII regions around the ionizing sources. One expects the secondary fluctuations to be of order

$$\delta T/T \sim N^{-1/2 - n/6}\ (v/c)\ ,$$

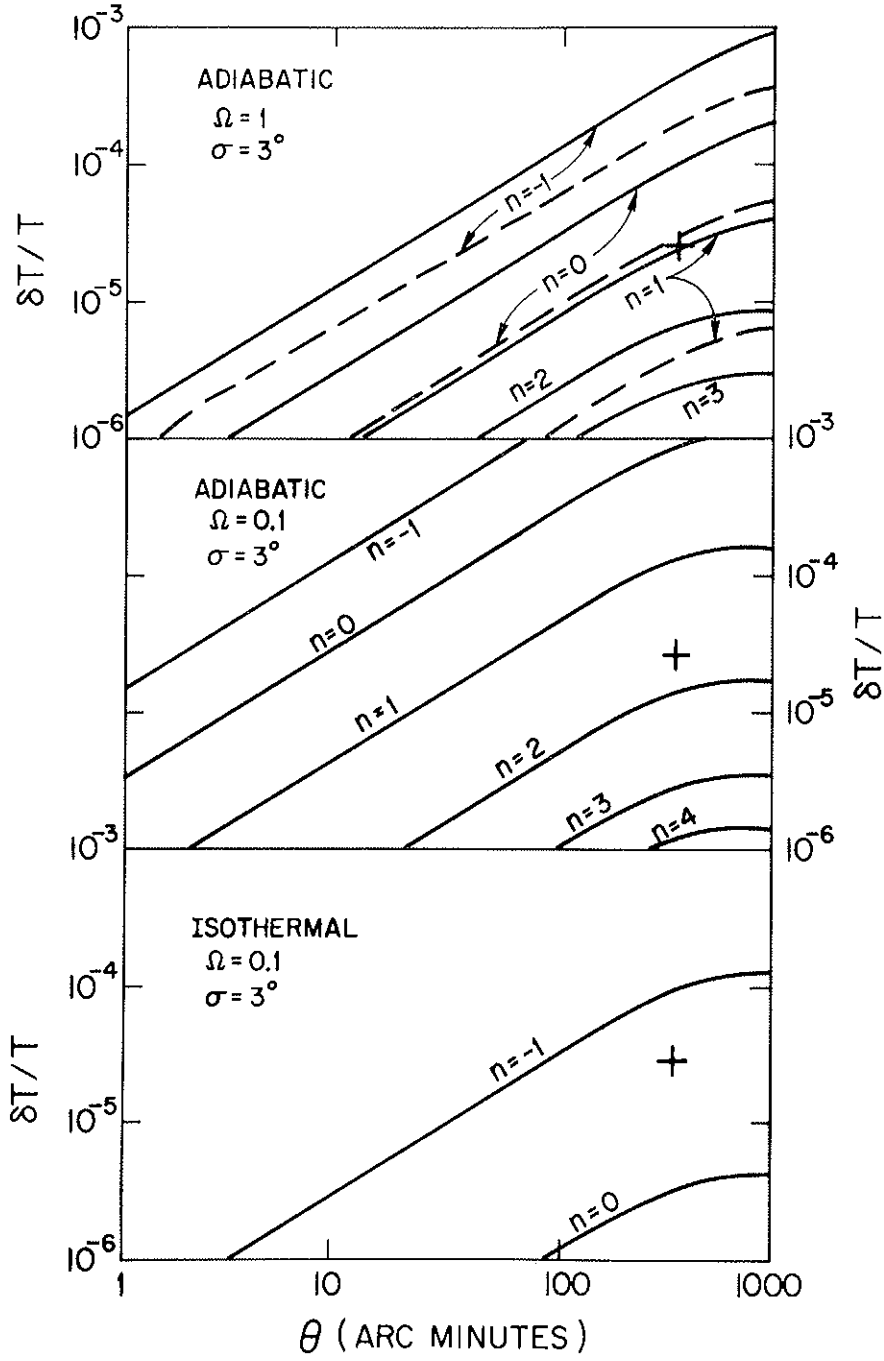


FIG. 3. Predicted temperature fluctuations in the cosmic microwave background radiation for a 3° beam, compared with the observational limit of Fabbri *et al.* (1981) at an angular scale $\theta = 6°$. Predictions by Wilson and Silk (1981) are shown for adiabatic fluctuations ($\Omega = 1, 0.1$) and for isothermal fluctuations ($\Omega = 0.1$). Dashed lines are predictions for adiabatic fluctuations in a neutrino-dominated universe ($\Omega_\nu = 0.98$, $\Omega_b = 0.02$).

if the ionization sources are associated with galaxies that have formed from a density fluctuation spectrum with power-law index n, and N is the number of scattering fluctuations along a line of sight. This results in the reduction of the small-scale temperature fluctuations by about an order of magnitude. This conclusion is somewhat sensitive to the adopted value of n. If most of the power is in large wavelength fluctuations as with $n \lesssim -1$, then the primordial fluctuations mostly survive even on small angular scales.

In general, constraints on the small-scale anisotropy below a few degrees are not likely to be conclusive, because reionization could conceivably occur in a highly uniform way. Moreover, there may be no intrinsic large-scale structure in the ionizing medium other than that associated with an extremely non-linear reshuffling over small scales in order to produce galaxies. The minimal level of fluctuation on large scales corresponds to a power spectrum with n equal to 4 in this case, and the small-angular scale anisotropy would be almost completely smoothed by reionization at $z > 10$.

The large angular scale anisotropy is immune from this effect, and consequently provides the most fundamental test of the gravitational instability theory. It is simplest to analyze the angular anisotropy in a multipole expansion. The gravitational potential fluctuations are the almost exclusive contributors to the lowest multipoles. One finds that, to lowest order, the dipole and quadrupole anisotropies are the most important. Over scales from 100 Mpc to the horizon scale it is the dipole anisotropy that makes the major contribution. This is due to the peculiar velocity (both of our local frame and that of the radiation) associated with large-scale potential fluctuations. Over scales comparable to and larger than the present horizon, the quadrupole anisotropy is dominant. In addition, the higher order multipoles are predicted to be present, but the amplitudes of individual spherical harmonic components are reduced by a factor $\sim N$ for the Nth multipole (since the rms amplitudes of the various multipole matrices are comparable). The Fourier power coefficient of the dipole anisotropy is

$$\sim 2/3\, \delta_k \cos\theta\, (kct_0)^{-1} + 0\,[\,(kct_0)^{-2}\,]\ ,$$

neglecting terms of order $(kct_0)^{-2}$, and for the quadrupole anisotropy is

$$\sim \delta_k \cos^2\theta + 0\,[kct_0]\ ,$$

neglecting terms of order kct_0, where

$$\theta = \cos^{-1} \underline{k} \cdot \underline{\gamma}$$

is the angle of photon arrival relative to some specified direction. The relative contribution of various wavenumbers to $\delta T/T$ for the dipole and quadrupole anisotropy is illustrated in Figure 4. It is apparent that a very steep spectrum (large n) is required to suppress the quadrupole anisotropy, but the dipole anisotropy will not decrease as much.

The large dipole contribution has profound implications. Because it always dominates the quadrupole anisotropy, it means that confirmation of a quadrupole anisotropy requires a substantial contribution to the dipole anisotropy from potential fluctuations on scales greatly exceeding that of the Virgo Supercluster. This conclusion is largely independent

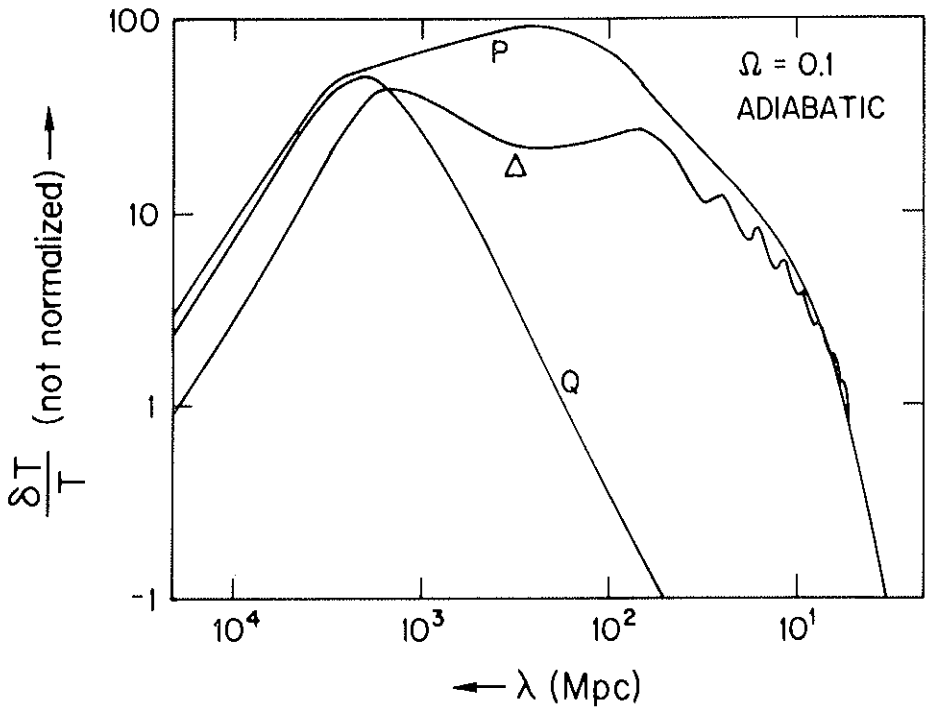


FIG. 4. The radiation perturbation spectrum $\delta T/T \cong k^{3/2} |\delta_k^{\rm rad}|$ for the $\Omega = 0.1$ adiabatic fluctuation model. Shown separately are the contributions to the dipole anisotropy P, the quadrupole anisotropy Q, and the small angular scale anisotropy Δ. From Wilson (1981).

of the assumption of a initial power-law spectrum, since potential fluctuations on such different scales are responsible for the quadrupole and dipole anisotropies, and these potential fluctuations are primarily due to the small-scale density inhomogeneities (unless $n < 0$) which are present in order to account for structure on cluster and supercluster scales.

The situation with respect to our peculiar velocity relative to the Virgo cluster is presently unclear, with analyses of different samples of galaxies resulting in infall velocities of the Local Group of between 200 and 500 km s^{-1}. The direction of the vector is uncertain, because the transverse component of motion has only been measured with respect to one of the samples, for which it is 74 ($\pm$ 71) km s^{-1} (Aaronson *et al.* 1981). The motion of the Local Group relative to the microwave background radiation is 540 ($\pm$ 60) km s^{-1} towards a direction about 45° away from the Virgo cluster; the component of the motion towards Virgo is 411 ($\pm$ 30) km s^{-1} (Boughn *et al.* 1981). This may change somewhat when complete sky coverage is attained and the higher order anisotropies are measured. At present, especially if the low value for our infall velocity relative to Virgo is adopted, there is clear evidence for a substantial contribution to our peculiar motion from matter well beyond the Virgo Supercluster. The apparent deviation of the motion from Virgo only weakly argues for this, because a substantial shearing motion could have been acquired if the Virgo Supercluster has undergone any recollapse, as in the pancake model.

The Rubin-Ford effect (Rubin *et al.* 1976) also suggests that a purely Virgocentric flow model is likely to be inadequate. It provides evidence for a possible motion of our galaxy of 600 ($\pm$ 125) km s^{-1} relative to a distant sample of Sc galaxies (at $cz \approx 5000$ km s^{-1}) in a direction some 90° away from Virgo. To explain the effect, one has to invoke either a large-scale weak shearing motion ($\delta H/H \sim 0.1$) or the presence of large-scale inhomogeneities ($\gtrsim 100$ Mpc in extent) that bias the luminosity function by appearing near the edge of the region studied. Either explanation is tantamount to requiring a component of the gravitational acceleration of our local rest frame beyond the Virgo Supercluster that would inevitably contribute to the dipole anisotropy.

While the dipole anisotropy may or may not provide evidence of large-scale inhomogeneities, the recently detected quadrupole anisotropy is reasonably unambiguous. Two measurements have been reported of the quadrupole anisotropy. The most complete sky coverage was obtained by Bougbn *et al.* (1981), who measured positive values for two of the

five spherical harmonic coefficients that describe the quadrupole moment. The amplitude (Q) of the quadrupole matrix (essentially the square root of the sum of the squares of the individual components) is 1.1 ($\pm$ 0.3) $\times$ 10^{-3} T^{-1}, about 30 percent of the amplitude (D) of the dipole anisotropy, 3.3 ($\pm$ 0.6) $\times$ 10^{-3} T^{-1}, where T $\approx$ 2.9 K.

Comparison of the observed values of the dipole and quadrupole anisotropies with the theoretical predictions (Figure 5) leads to some interesting conclusions. One should first bear in mind that the predicted anisotropies are averaged over the observer's position: the dipole and quadrupole anisotropies are effectively independent averages, and actual values could differ by a factor of 2 or 3 in either direction. Other uncertainties come from the incomplete sky coverage, especially for the

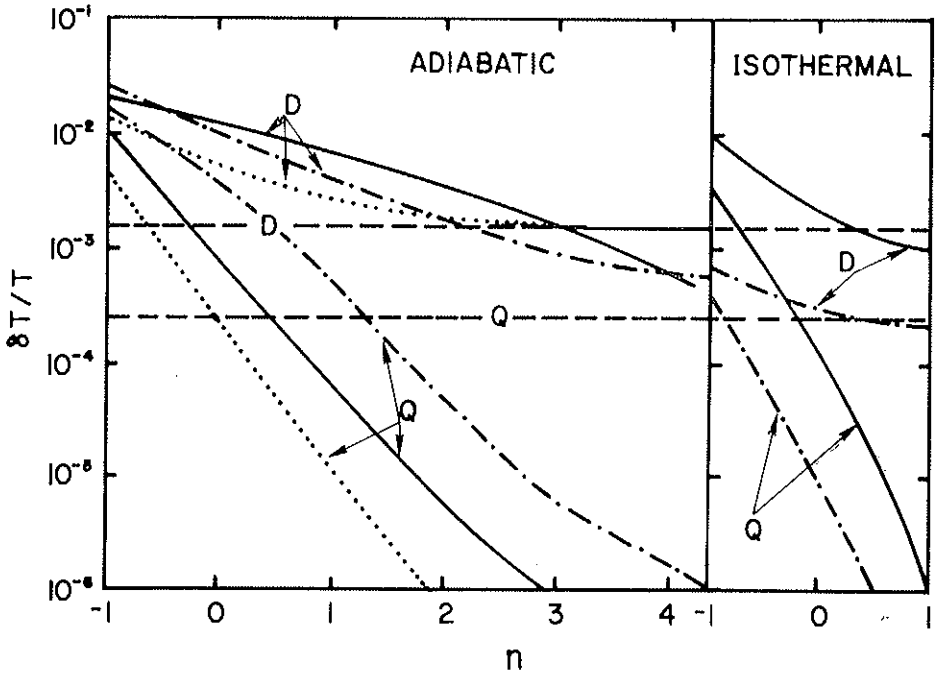


FIG. 5. Predictions of dipole and quadrupole temperature anisotropy, as a function of the power-law index of the fluctuation spectrum, compared with observed values. The dashed lines denote the observed amplitudes (D and Q are the rms values of the dipole and quadrupole anisotropies). Predictions by Silk and Wilson (1981) are for the standard $\Omega = 1$ (continuous curves) and $\Omega = 0.1$ (dash-dot curves) models ($m_\nu = 0$) for adiabatic and isothermal fluctuations and for a neutrino-dominated cosmological model with $\Omega_\nu = 0.98$, $\Omega_b = 0.02$ (dotted curves).

quadrupole anisotropy, which may substantially bias the reported value, and the normalization of $\delta\rho/\rho$ to the galaxy correlation function in the linear regime. Especially for $n \gtrsim 1$, $\xi(r)$ displays oscillations for adiabatic fluctuations, due to phase differences that arise during the damping prior to decoupling that results in the short-wavelength truncation of the fluctuation spectrum. This must also introduce an uncertainty amounting to a factor of 2 or 3. With due regard to these caveats, inspection of Figure 5 shows that isothermal fluctuations with $n \approx 0$, corresponding to white noise initial conditions, can simultaneously account for both the quadrupole and much of tbe dipole anisotropy. Comparison with the measurement at 6° shows that the $\delta T/T \propto \theta^{1/2}$ interpolation for $n = 0$ provides a reasonable fit to the quadrupole anisotropy.

The adiabatic fluctuation theory fails, however, to account for tbe quadrupole anisotropy. Independently of any value of n, a fit to the quadrupole anisotropy results in excessive dipole anisotropy. Are there any loopholes in this conclusion? A spatially flat cosmology has hitherto been assumed. One expects that if $\Omega < 1$, the dipole anisotropy will be reduced, as is the case for peculiar motions relative to the expansion, roughly by a factor $\Omega^{1/2}$. Detailed computations (Wilson 1981) are consistent with this scaling. Since a minimum value for Ω is now believed to be 0.1, this would not seem to be a sufficiently large effect to rescue the adiabatic theory. Dropping the assumption of an initial power-law spectrum would obviously make a major change in this conclusion. For example, the presence of a large lump just on the edge of our present horizon, at $z \sim 1$ say, would boost the quadrupole anisotropy considerably relative to the dipole anisotropy. Alternatively one could choose a primordial power-spectrum with variable index n, small on large scales $\gtrsim 100$ Mpc ($n \sim 0$) but large on small scales $\lesssim 100$ Mpc ($n \sim 3$). To avoid divergent gravitational potential fluctuations, the spectrum must flatten out ($n < 1$) on scales $\lesssim 1$ Mpc. One might hope that observations of the diffuse X-ray background, believed to originate primarily from quasars at $z \lesssim 4$, will eventually probe the presence of such large lumps. For now, we do not consider this possibility any further, as it seems a relatively *ad hoc* model that in any event contradicts underlying assumptions of the adiabatic fluctuation theory.

One concludes that in the standard model with the simplest assumption of a constant power-law spectrum for the primordial density fluctuations, primordial adiabatic fluctuations are untenable for any single power-law index as the seeds for galaxy formation, whereas isothermal

fluctuations with a white noise spectral power distribution satisfy all observational constraints. Indeed, no monotonic fluctuation spectrum will be capable of saving the adiabatic theory. One final note of caution: this sweeping conclusion is based on the reality of the quadrupole anisotropy with amplitude about one-third that of the dipole anisotropy. Conceivably this could seriously overestimate any true anisotropy if there were significant contamination, for example, by galactic thermal or synchrotron emission. Experiments currently in progress are expected to resolve this question, by increasing both the sensitivity and the amount of sky coverage.

5 - Particle physics and cosmology

Considerations from particle physics can have a major impact on much of the preceding discussion. In fact, the predictions of the amplitudes of the largest scale structures observed in the universe, the dipole and quadrupole anisotropies of the cosmic microwave background radiation, may depend sensitively on the most microscopic scales of the ultimate theory of elementary particles. This apparently paradoxical situation is not really so surprising, given that our theory of large-scale structure depends on the specified initial conditions at an epoch when causally connected regions encompassed only a small number of elementary particles.

A recent development in cosmology has centered on the application of grand unified theories (GUTs) of the strong, weak and electromagnetic interactions to the very early universe. Indeed, the extreme energies attained as $t \rightarrow 0$ provide one of the few environments for studying the implications of GUTs. One of the few fundamental numbers needed to characterize the Friedmann cosmological model is the dimensionless entropy per baryon, equal to $2.5 \times 10^{10} (\Omega_B h^2/0.01)$ for photons and three neutrino flavours. Perhaps the most spectacular success of GUTs has been to provide a natural explanation (at least in order of magnitude) for this number in terms of the baryon-non-conserving processes that occur in these theories. While the entropy-per-comoving-volume remains approximately constant, the baryon number, an excess of particles over antiparticles, which is initially zero, is generated as the symmetry of the grand unified era is broken soon after the Planck time. Since a universal baryon number is generated, any pre-existing isothermal perturbations, which are equivalent to baryon number or specific entropy variations, cannot survive the baryosynthesis era in the standard model. Only adiabatic perturbations should be present to form galaxies, according to this result.

There is at least one noteworthy means of resurrecting isothermal fluctuations. Spatial variations in the expansion rate on scales larger than the horizon during baryosynthesis will give rise to entropy fluctuations. To understand how this arises, consider first the role of the X-bosons, which unify the electromagnetic, weak and strong interactions, and consequently mediate proton decay. Only because the baryon non-conserving reactions go out of equilibrium as the universe expands does baryon number build up, X-bosons being destroyed more often than they are created. Reactions like $X \rightarrow qq$ dominate over $X \rightarrow \bar{q}\bar{q}$ by an amount that depends on how much CP violation there is in X-decay. This is a free parameter, although measurements for the kaon system indicate a value $\varepsilon \sim 3 \times 10^{-3}$ for the fractional amount of CP violation. As the temperature drops to $kT \ll m_x \sim 10^{15}$ GeV, both the baryon formation and destruction rates become small relative to the expansion rate, and a net baryon number is frozen out. For a uniformly expanding Friedmann cosmological model, the only free parameter is ε: a value of order 10^{-6} is required in order to attain the observed entropy per baryon (Kolb and Wolfram 1980; Frye, Olive and Turner 1980).

Now consider the fate of energy density inhomogeneities, which can be decomposed into two independent linear fluctuation modes corresponding to constant and decaying curvature perturbations. While any primordial entropy fluctuations are "cooked" away, the so-called growing mode of primordial adiabatic (equivalent to constant curvature fluctuations) can in principle perturb the expansion rate at baryosynthesis, except that its effect must be small at this epoch in order to avoid resulting in unacceptably large energy density inhomogeneities in the very early universe. The decaying curvature mode can lead to a potentially large perturbation of the expansion rate as $t \rightarrow 0$, provided that the amplitude of the growing mode is initially suppressed. This could happen naturally if one required, at say the Planck epoch, that there was equipartition of energy between the modes (Kompaneets, Lukash and Novikov 1981).

There is an additional type of fluctuation that can have a dramatic effect on baryosynthesis, namely spatial variations in the expansion rate due to shear or vorticity perturbations. Shear inhomogeneity includes a decaying mode that could have been large near the Planck time, $t_{pl} \sim (8\pi\, Gh/c^5)^{1/2} \sim 10^{-43}$ s. One might conjecture that at this epoch, the threshold of cosmology, all possible decaying modes were present, only asymptotically resulting in a nearly uniform Friedmann cosmology. Provided the grand unification era is sufficiently close to the Planck era at $kT \sim 10^{19}$

GeV, and supersymmetric unified theories indeed advocate a grand unification energy $m_x \sim 10^{17}$ GeV ,the decaying modes should still be significant at baryosynthesis. For example, if collisionless particles dominate the energy density, the decaying shear mode $\Sigma = \Sigma_{\rm initial}\ (m_x/m_{\rm Planck})^{1/2} \sim 0.1\ \Sigma_{\rm initial}$ at this epoch, where Σ is the ratio of shear to expansion rates. Analysis of the effect of shear on baryosynthesis shows that isothermal fluctuations $\delta\rho_b/\rho_b \sim (\delta\Sigma\ (m_x)\ )^2 \sim 0.01\ \delta\Sigma_{\rm initial}$ are generated with $m_x \approx 0.01\ m_{\rm Planck}$ (Bond, Kolb and Silk 1982). These are more than adequate for the isothermal fluctuation theory of galaxy formation if $\delta\Sigma_{\rm initial}$ is present on comoving scales well above 10^6 $M_\odot$. Moreover, the initial shear will have rapidly decayed, and be completely negligible even by the epoch of nucleosynthesis. One can also imagine a universe which initially is dominated by inhomogeneous shear with $\Sigma\ (m_{\rm Planck}) \gg 1$: in this case $\delta\rho_b/\rho_b \sim \delta\Sigma/\Sigma$.

Another far-reaching implication of GUTs is that neutrinos should possess a non-zero rest mass. While the GUTs only indicate a minimal lower bound ($m_\nu \gtrsim 10^{-6}$ eV), there are tentative (and presently unconfirmed) experimental indications that m_ν could be as large as 30 eV from a tritium decay experiment (Lyubimov *et al.* 1980) or $\gtrsim 1$ eV from a neutrino oscillation experiment (Reines *et al.* 1980). These preliminary results have motivated cosmologists to reconsider the implications of a neutrino rest mass. The mass density of cosmological background neutrinos exceeds the luminous matter density if the neutrino rest mass m_ν (assumed to be in one flavour) $\gtrsim 1$ eV, and closes the universe if $m_\nu \gtrsim 100\ h^2$ eV, where $h \equiv H_0/100$ km s^{-1} Mpc^{-1}. A neutrino mass in the range 10 to 30 eV also has a dramatic effect on galaxy formation theory, as will now be described.

With primordial adiabatic fluctuations there are associated fluctuations in the neutrino density. Secondary neutrino fluctuations are also generated by isothermal or stress perturbations once they cross the horizon in the matter-dominated era. Neutrino perturbations undergo collisionless damping on suh-horizon scales: this defines a characteristic comoving mass-scale $M_{\nu m} \approx 4 \times 10^{15}\ m_{30}^{-2}\ M_\odot$, to which corresponds a comoving length-scale $\lambda_{\nu m} = 40\ m_{30}^{-1}$ Mpc, where m_{30} is the neutrino mass (in one flavour) in 30 eV units. This is analogous to the Jeans length in a collisional fluid, except that density fluctuations cannot be sustained at all on smaller scales while growing via gravitational instability on larger scales.

Two effects are responsible for the damping. While the neuttrinos are relativistic, phase mixing occurs between peaks and troughs of adjacent

waves. When the neutrinos are non-relativistic, the fluctuations disperse by Landau damping, the faster neutrinos overtaking the slower neutrinos. In either case, the neutrino velocity dispersion v_s determines the effective Jeans length. While the neutrinos are relativistic, the instantaneous neutrino Jeans length $\lambda_\nu \sim v_s t$ increases as t, and when the neutrinos become non-relativistic, $v_s \propto a^{-1}$ and $\lambda_\nu \propto t^{1/2}$. Consequently, the Jeans mass $\rho_\nu \lambda_\nu^3 \propto \lambda_\nu^3 t^{-2}$ attains a peak value $M_{\nu m}$ at redshift $1 + z_m = 35000\ m_{30}$ (Bond, Efstathiou and Silk 1980; Doroshkevich *et al.* 1980b). The neutrinos actually become non-relativistic at a somewhat larger redshift $57300\ m_{30}$ No primordial neutrino fluctuations can be sustained on scales below $M_{\nu m}$, although primordial isothermal fluctuations do generate secondary adiabatic fluctuations on scales above λ_ν, but below $\lambda_{\nu m}$. This occurs because of the associated stress perturbation induced because $\delta p \neq \delta\varepsilon/c^2$ for an isothermal fluctuation, whereas $p = \varepsilon/3$ is the background equation of state at early epochs. This results in the generation of energy density fluctuations once the appropriate wavelength first enters the horizon. This effect is only significant once the predominant constituent of the universe is non-relativistic. In principle, one could fine-tune the initial conditions to remove this secondary adiabatic component by subtracting a small adiabatic fluctuation from the primordial isothermal perturbation, but this may require a very specific choice of initial conditions.

Damping of neutrino fluctuations is extremely severe on scales below $\lambda_{\nu m}$. Detailed computations of the damping of fluctuations in a neutrino universe involve a simultaneous solution of the Boltzmann transport equation for the collisionless neutrinos and the perturbed Einstein equations for the evolution of density perturbations (Bond and Szalay 1981). The neutrinos are still semi-relativistic when much of the damping occurs, and the system of equations must be evaluated numerically. Computations of the evolution of several Fourier components of the density fluctuation spectrum show that a wave with $\lambda = 1/3\ \lambda_{\nu m}$ suffers extreme damping by a factor ~ 200 before its growth phase begins after the neutrino velocity dispersion has dropped sufficiently, whereas one with $\lambda = 2\ \lambda_{\nu m}$ suffers very little damping. The damping of a given component is first effective when about one-half wavelength of the Fourier component has entered the horizon, for $\lambda \lesssim \lambda_{\nu m}$. The resulting fluctuation spectrum for an initial adiabatic power spectrum with $n = 4$ is shown in Figure 1.

The baryonic component of adiabatic fluctuations also suffers damping from radiative viscosity and diffusion on scales below $M_d \sim 3 \times 10^{13}$

$\Omega_b^{-1/2}\,\Omega_\nu^{-3/4}\,h^{-5/2} \sim 10^{15}\,M_\odot$, coincidentally within an order of magnitude of the baryonic mass associated with $M_{\nu m}$, namely $M_{bm} \sim 1.2 \times 10^{14}\,m_{30}^{-3}\,h^2\,M_\odot$. Baryonic fluctuations on scales above M_d and below the baryonic Jeans mass ($\sim 10^{17}\,M_\odot$ just prior to decoupling) do not grow significantly in amplitude until after the decoupling epoch. After recombination occurs, baryonic fluctuations (even if not present initially) rapidly grow to respond to the neutrino perturbation on scales $\lambda > \lambda_{\nu m}$. The first fluctuations to become nonlinear have masses of order $M_{\nu m}$, and these undergo aspherical collapse preferentially along one axis, as envisaged in the original Zel'dovich pancake theory. Because of the sharp cut-off of structure at short wavelengths, a caustic surface forms as trajectories intersect. The baryonic component develops a radiative shock and undergoes fragmentation. The neutrinos freely penetrate the caustics and separate from the baryons. Multiple streaming motions develop in the midplane, due to the infalling neutrinos and those at low velocity initially near the midplane. It is likely that such a configuration is unstable in the presence of the inhomogeneous baryonic component, the gravitational two-stream and Jeans instabilities being effective as the streams interpenetrate. The baryonic fragments perturb the local gravitational potential in such a way as to generate sheet-like density fluctuations. A similar phenomenon appears to be present in N-body collapses of spheroidal distributions of particles (Miller and Smith 1979). According to one-dimensional N-body simulations of the pancake collapse, streaming progressively develops in phase space as neutrinos that have already passed through the plane are turned back, seeing a deeper gravitational well as the infall continues. One ends up with a sandwich of bound neutrinos surrounding a fragmented baryonic pancake.

A substantial fraction of the neutrinos retain a low velocity until several collapse times (in a direction perpendicular to the plane of symmetry) have elapsed, according to one-dimensional simulations (Doroshkevich *et al.* 1980a; Melott 1981), and these cold neutrinos become bound to the baryonic fragments. Little dilution of phase space density occurs for these neutrinos. A similar effect is well known from earlier studies of one-dimensional collapse (e.g. Janin 1971). One expects that the mass fraction of neutrinos should increase progressively with the depth of the potential well of the baryonic core. Consequently, neutrino infall can at least qualitatively account for the dark matter in galaxy halos and also produce an increase in mass-to-luminosity ratio with increasing scale. One of the successes of the original pancake model is preserved, namely the

large halos and filamentary nature of the galaxy distribution, since the baryonic component still undergoes extreme dissipation. The dominant mass constituent of the universe, the neutrinos, should possess a much more uniform distribution. Fragmentation of the neutrino sandwich can also occur directly by Jeans instability provided that fluctuations on scales $< \lambda_{\nu m}$ are present in the infalling neutrinos. This could result in the formation of loosely bound dark structures which may be largely devoid of any associated baryonic component.

Massive neutrinos rescue the adiabatic fluctuation model from two otherwise nearly fatal difficulties. Since most of the power in the density fluctuation spectrum is at a mass-scale $M_{\nu m}$, as opposed to being spread over the range between the damping mass and the baryonic Jeans mass prior to decoupling, the correlation function is greatly reduced at large scales if $m_\nu \sim 30$ eV. The observations are perfectly consistent with a primordial adiabatic density fluctuation power spectrum $\delta\rho/\rho \propto k^n$ with either $n = 0$ (white noise) or $n = 1$, the constant curvature value that offers such theoretical appeal (Figure 2). Because the neutrino fluctuations grow between z_{nr} and decoupling, whereas the radiation fluctuations do not, the strong coupling after the recombination epoch between radiation, baryons and neutrinos results in radiation temperature fluctuations that are reduced by a factor $\sim z_d/z_{nr}$. This applies on small and intermediate angular scales, and suffices to reconcile the observational upper limits with the anisotropy required to form galaxies in the adiabatic model (Figure 3). The large angular-scale anisotropy presents more of a problem, since here the gravitational potential fluctuations are relatively independent of m_ν and cause large-scale structure in the microwave background radiation. Because the power in fluctuations on the horizon scale prior to decoupling is reduced, however, neutrinos do improve matters. The removal of the broad peak in the matter fluctuation spectrum suppresses the dipole anisotropy (to which it contributes an appreciable amount) more strongly than the quadrupole anisotropy, which is largely due to potential fluctuations on our present horizon scale associated with scales $\sim M_{\nu m}$. The quadrupole anisotropy can now be accounted for without producing excessive dipole anisotropy if $0 \lesssim n \lesssim 1/2$, although $n = 1$ apparently results in too small a quadrupole anisotropy (if we adopt the result of the Princeton group as a firm value, see Figure 5).

The isothermal fluctuation theory also undergoes a major revision if neutrinos are massive: in fact, the first structures to form may be pan-

cakes of total mass $M_\nu(z_d) \sim 3 \times 10^{13}\, m_{30}^{-7/2}\, M_\odot$. This apparently paradoxical result arises because baryonic fluctuations are immersed in a hot neutrino fluid at decoupling. Growth can only commence after the decoupling epoch on a mass-scale above the instantaneous neutrino Jeans mass. The amplitude of the growing mode is suppressed by a factor Ω_b/Ω_ν. Scales above $M_\nu(z_d)$ grow from decoupling to the present, while scales below $M_\nu(z_d)$ experience delayed growth. The results from the linear theory of fluctuation growth are schematically summarized in Figure 6. The inhibition of growth results in a flattening of the fluctuation spectrum below $M_\nu(z_d)$ by $M^{2/3}$. If $n < 1$, fluctuations on scale $M_\nu(z_d)$ are the first to go non-linear. Since baryonic pressure is completely insignificant on this mass-scale, one again ends up with pancake formation, but now on mass-scale $M_\nu(z_d)$.

An intriguing difference between adiabatic and isothermal pancakes is that the former exhibit a sharp cut-off in internal structure at $\lambda < \lambda_{vm}$, whereas the isothermal fluctuation spectrum continues below $\lambda_\nu(z_d)$. The minimum scale can be at least as small as the present neutrino Jeans mass, $M_\nu(0) = 10^9\, m_{30}^{-7/2}\, M_\odot$, although $\delta\rho_b/\rho_b \sim \Omega_\nu/\Omega_b \gg 1$ would be required on this scale for it to be presently non-linear and undergo collapse. It seems more plausible that small-scale structure is formed not directly from primordial fluctuations in a uniform background but by the fragmentation which will occur readily during the non-linear stages of isothermal pancake collapse. The lack of a well-defined short wavelength cut-off ensures that the baryonic fluctuations within a given pancake are large, and these suffice to result in fragmentation of the collapsing neutrino cloud. The non-linear background of the collapsing pancake gives a huge boost to the growth rate of substructure, provided only that small fluctuations are present initially (Sato and Takahara 1981). Within the neutrino fragments that extend in mass down to $M_\nu(0)$, baryonic dissipation will occur to presumably form the luminous regions of galaxies.

The amount of radiation anisotropy for this isothermal neutrino pancake model can be estimated from the required amplitude on scale M_ν at decoupling. To obtain $\delta\rho_b/\rho_b \sim 1$ on scale r_0 at present, where the galaxy correlation function

$$\xi(r) = (r/r_0)^{5/3} \text{ and } r_0 = 5\ h^{-1}\ \text{Mpc}\ ,$$

we require that at decoupling

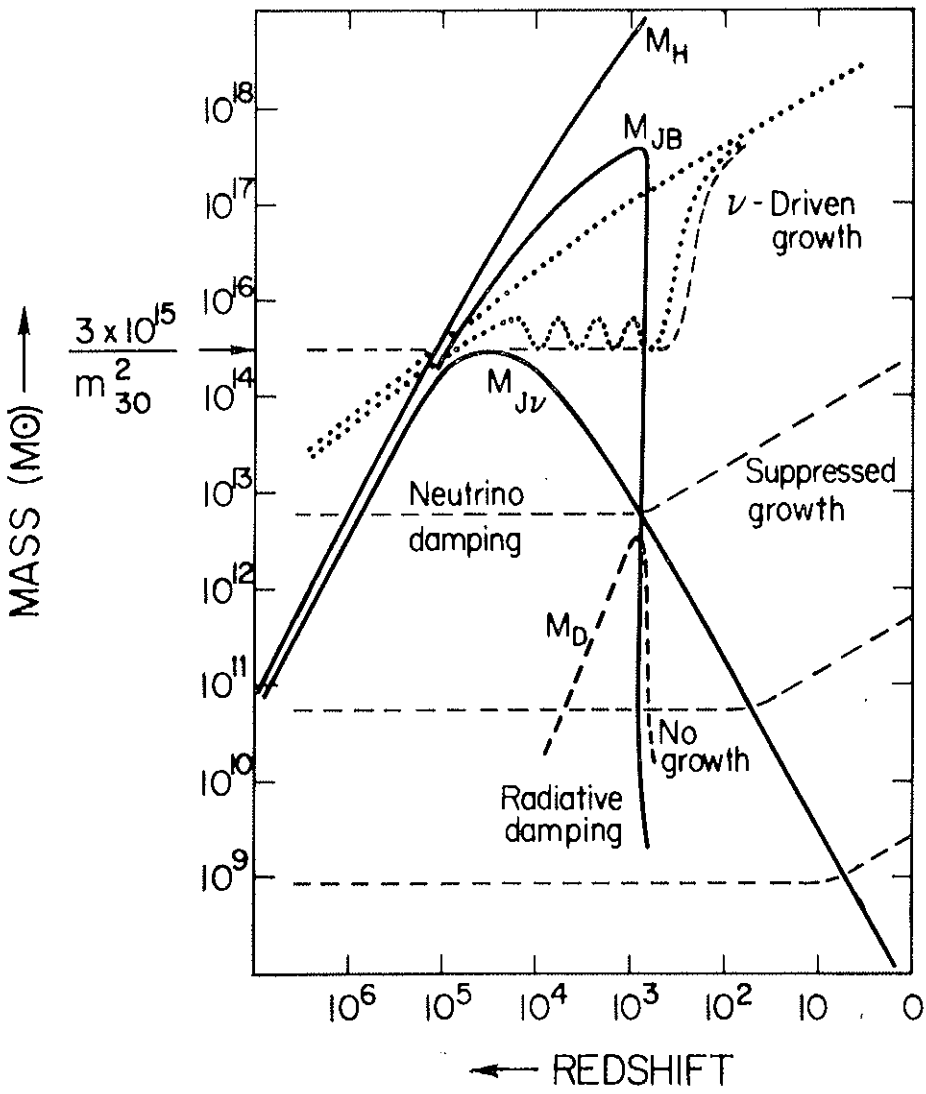


FIG. 6. Fluctuation growth in a neutrino-dominated universe. Dotted lines indicate schematic evolution with redshift of adiabatic density fluctuations and broken lines the evolution of isothermal fluctuations for different mass scales that correspond to the scale on the ordinate. Heavy lines are characteristic mass-scales: M_H is horizon mass, M_{JB} = baryonic Jeans mass, $M_{J\nu}$ = neutrino Jeans mass (assuming $m_\nu = 30$ eV), and M_D is baryonic damping mass.

$$\delta\rho_b/\rho_b \approx 0.01\, m_{30}^{-2}\ (r_0/r)^{(n+3)/2}\ .$$

Now extrapolation of the linear regime of the correlation function back to decoupling yields the effective normalization used previously for evaluating isothermal radiation fluctuations, namely

$$\delta\rho_b/\rho_b = [r_0\,(1 + z_d)^{-3/5}/r]^{(n+3)/2}\ .$$

The radiation fluctuations associated with the isothermal density fluctuations in the presence of massive neutrinos are, therefore, increased by a factor $\sim 0.01\ (1 + z_d)^{3\,(n+3)/10}\ m_{30}^{-2} \sim 10\ m_{30}^{-2}$.

These estimates all assume that all the neutrino mass resides in one flavor, but they can readily be generalized. It appears as though isothermal fluctuations with $m_\nu < 30$ eV would be untenable, by producing excessive radiation anisotropy on an angular scale of a few degrees, for a white noise initial spectrum. However, $n = 1$ would be consistent with this and other upper limits. A more severe difficulty is tbat the predicted dipole anisotropy now exceeds the measured value for $n < 1$. However, with $n = 1$, the predicted quadrupole anisotropy is only 10^{-5} for $m_{30} \approx 1$. This may barely suffice to account for the observed anisotropy.

There is an additional effect that may boost the fluctuation growth rate by a large factor. If the isothermal fluctuations are strictly entropy fluctuations with $\delta\rho_b/\rho_b$ = constant, then $\delta\rho_r/\rho_r = \delta\rho_\nu/\rho_\nu = 0$ at some initial epoch when they are produced, presumably at the end of baryosynthesis (Bond *et al.* 1982). In this case, the fluctuations contain a small amount of curvature, associated with stress perturbations of order ρ_b/ρ_r. This has no significant effect until a fluctuation first enters the horizon after the neutrinos have gone non-relativistic. At this stage, adiabatic fluctuations of amplitude $\delta\rho_b/\rho_b\ (\Omega_b/\Omega_\nu)$ are generated. The neutrino component of these secondary adiahatic fluctuations grows in amplitude on mass-scales greater than $M_{\nu m}$. After decoupling, these scales have a far larger amplitude than the primary isothermal fluctuations on scale $M_\nu\,(z_d)$, having been boosted by a factor $\sim (1 + z_{nr})/(1 + z_d)\ [M_\nu\,(z_d)/M_{\nu m}]^{1/2+n/6}$.

One now obtains pancakes just as in the adiabatic theory, except that there is considerable substructure. The growth rate of the isothermal fluctuations is boosted in the pancake collapse on scale $M_{\nu m}$, resulting in efficient fragmentation that continues down to scales of or even below $\sim 10^6\ M_\odot$. This is because the primordial fluctuation isothermal spectrum may extend down to scales below the post-decoupling baryonic Jeans mass.

If $n < 1$, the fragment spectrum peaks at $M_\nu(z_d)$, corresponding to a baryonic mass of $\sim 10^{12}\ M_\odot$, since it must reflect the amplitudes of density fluctuations present at the onset of collapse. Since an isothermal fluctuation spectrum with $n > 1$ fails to result in any appreciable quadrupole anisotropy, one may consider that fragmentation will first occur on scales $\sim M_\nu(z_d)$. This, therefore, provides an intriguing possibility for generating large-scale structure, including the quadrupole anisotropy and large-scale features in the galaxy distribution, from primordial isothermal fluctuations.

6 - Conclusions

If the neutrino has a rest-mass $\gtrsim 30$ eV, both the adiabatic and isothermal fluctuation theories provide acceptible interpretations of the quadrupole anisotropy in the cosmic microwave background radiation. While an excessive dipole anisotropy is not generated with a near flat ($n \approx 0$) fluctuation spectrum, there is nevertheless likely to be a substantial contribution to the observed dipole component. Indeed, it seems unavoidable if there is a cosmological quadrupole anisotropy. This means that one should not expect the motion of the Local Group with respect to the Virgo Supercluster to coincide in amplitude or in direction with that inferred relative to the hackground radiation.

Galaxy formation now occurs via pancake formation. This is practically inevitable in both adiabatic and isothermal theories. In the latter case, secondary adiabatic fluctuations also lead to pancakes of mass $M_{\nu m}$. The initial conditions required are either curvature fluctuations, $K \approx 10^{-4}$, or entropy fluctuations, $\delta\rho_b/\rho_b \approx 10^{-4}$, on scales $M_{\nu m}$ with possible (but not necessary) interpolations $K \propto M^{1/6}$ and $\delta\rho_b/\rho_b \propto M^{-1/2}$ to smaller mass-scales. The interpolation of the fluctuation mass spectrum to larger scales also is not crucial, since, for $n=0$, the covariance vanishes on very large scales, and the potential fluctuations due to the $M_{\nu m}$ lumps are adequate to account for the quadrupole anisotropy.

Even if isothermal initial conditions are precisely adjusted to avoid the secondary adiabatic fluctuations, scales $M_\nu(z_r)$ are the first to go non-linear because of the constraint $n \approx 0$ imposed by the interpretation of the quadrupole anisotropy. Only in this case, large initial amplitudes are required, with $\delta\rho_b/\rho_b$ of order unity on galactic mass-scales in order to overcome the Ω_b/Ω_ν suppression factor. In general, isothermal fluctuations first become non-linear on mass-scales that are essentially pressure-free,

since the baryonic Jeans mass is below 10^6 $M_\odot$, and should therefore collapse anisotropically to form pancakes. These pancake masses are typically those of the most massive galaxies. Pancake collapse results in efficient fragmentation when isothermal fluctuations are present: the substructure aids galaxy formation, since galactic cores can now presumably form halos more easily out of the collapsing neutrinos.

Now suppose that experiments eventually demonstrate that $m_\nu \ll 30$ eV. In the standard model in which baryonic matter dominates the mass density of the universe, a primordial adiabatic density fluctuation spectrum that accounts for galaxy formation is untenable. This conclusion only presumes that the reported quadrupole anisotropy of the cosmological background radiation is genuinely of cosmological origin. One cannot account for this by gravitational potential fluctuations without producing an excessive dipole anisotropy, unless an exceedingly *ad hoc* initial fluctuation spectrum is introduced.

One is left with primeval isothermal density fluctuations, which cluster hierarchically after becoming non-linear at $z < z_a$ to eventually form galaxies and clusters. Now preexisting isothermal fluctuations cannot survive the grand unified era if baryosynthesis occurs, unless one appeals to some additional process that can affect the baryon production. The most plausible possibility is if decaying inhomogeneous curvature or shear modes exist as residues of more chaotic conditions in the quantum gravity era. These only yield significant fluctuations if the baryosynthesis era is sufficiently close to the Planck epoch. Incorporating the decay rate of acceptable shear modes, one evidently requires the grand unification mass to be $> 10^{15}$GeV, if inhomogeneously decaying shear is to play any significant role in affecting the expansion rate of the universe, and yield $\delta\rho_b/\rho_b > 10^{-4}$.

Clearly, the higher the value of the grand unification mass the more likely one is to have residual decaying modes from the quantum gravity era perturbing baryosynthesis. A high value of the grand unification mass may arise naturally in a supersymmetric grand unification scheme. This places fermions and bosons on an equal footing. One consequence is that in such supersymmetric theories, fermion equivalents of identical spin to bosons arise naturally. In particular, photinos are produced in equal number to the photons, and the photinos can have a non-zero rest mass. According to a recent calculation (Weinberg 1981), a possible photino rest mass is $\sim$ 100 eV. Thus, one can now repeat all calculations involving massive neutrinos, simply replacing the neutrino by the (unobservable) photino.

This conclusion suggests that a pancake formation theory for galaxies may be inevitable. The quadrupole anisotropy of the microwave background radiation, if interpreted in terms of potential fluctuations associated with large-scale structure, inevitably leads to a scenario in which large scales collapse first. Galaxies must form by fragmentation of scales of order 10^{15} to 10^{16} $M_\odot$ (for primeval isothermal fluctuations) if the neutrino/photino mass is sufficient (10 to 100 eV) to account for the dark matter in the universe. In the absence of such massive non-baryonic elementary particles, the only acceptable theory is that involving primeval isothermal fluctuations. These violate GUTs unless something like a supersymmetric theory is adopted that allows a large grand unification mass-scale and thereby enhances the plausability of decaying shear fluctuations as a residue to the quantum gravity era. Yet the presence of massive photinos is likely in such a theory. In addition, of course, one notes that the primordial nucleosynthesis of the light elements (including D, He^3 and possibly even He^4) is acceptable only if non-baryonic matter dominates the present mass density.

It seems that the large-scale structure of the universe is intimately related to its microscopic structure on elementary particle scales. This is perhaps not so surprising if one recalls that it is the initial seed fluctuations at the Planck epoch that are likely to determine the asymptotic growth of irregularities in the expanding universe. What perhaps is remarkable is that one can now begin to make assertions about the relevance of grand unified theories to the early universe on the basis of the cosmological origin of the quadrupole anisotropy.

Both isothermal and adiabatic theories make similar predictions about the evolution of protogalaxies. While one might expect some galaxy formation to occur at large redshift in the isothermal theory, it seems inevitable from the constraints imposed by the observed characteristics of galaxies that the bulk of galaxy formation occurs at a similar epoch and in a similar manner in both theories. The structure of the Local Supercluster appears to be compatible with both theories. The covariance function of the galaxy distribution over large scales only seems capable of setting constraints on the primeval fluctuation spectrum in either theory. The microwave background is unique in that the dipole, quadrupole and, ultimately, higher multipole anisotropies can actually discriminate between the adiabatic and isothermal theories. This is because even the assumption of an initial power-law fluctuation spectrum leads to considerable large-scale structure in the matter spectrum after the recombination epoch. Since the quadrupole anisotropy arises from potential fluctuations on the horizon

scale and the dipole and higher order anisotropies primarily probe smaller scales, one has a sensitive measure of the large-scale power in the fluctuation spectrum.

With adequate sky coverage of the multipole structure of the radiation background, one could hope to distinguish between the hierarchical clustering and pancake fragmentation theories of galaxy formation, unfold the role of massive neutrinos or photinos in modifying the primordial fluctuation spectrum, and even measure the curvature radius of the universe. All of these enigmas are encoded in the large-scale anisotropy. Deciphering it promises to be one of cosmology's greatest and most rewarding challenges.

I thank M.L. Wilson and C. Norman for helpful discussions. The research described here has been supported in part by NASA and the National Science Foundation.

REFERENCES

Aaronson, M., Huchra, J., Mould, J., Schechter, P. and Tully, R.B., 1981, preprint.

Barrow, J. and Silk, J., 1981, *Ap. J.*, **250**, 432.

Bond, J.R., Efstathiou, G. and Silk, J., 1980, *Phys. Rev. Lett.*, **45**, 1980.

Bond, J.R., Kolb, E. and Silk, J., 1982, *Ap. J.* (in press).

Bond, J.R. and Szalay, A., 1981, Proc. Neutrino '81, ed. R. Cence, E. Ma and A. Roberts (Honolulu: University of Hawaii), p. 59.

Boughn, S.P., Cheng, E.S. and Wilkinson, D.T., 1981, *Ap. J. Letters*, **243**, L 113.

Davis, M., 1981, private communication.

Davis, M., Huchra, J., Latham, D.W. and Tonry, J., 1982, *Ap. J.* (in press).

Dekel, A., 1982, *Ap. J.* (in press).

Dekel, A. and Szalay, A., 1982, in preparation.

de Vaucouleurs, G., 1975, in *Stars and Stellar System*, **9**, ed. A. and M. Sandage and J. Kristian (Chicago: University of Chicago Press), p. 557.

Doroshkevich, A.G., Lukosh, V.N. and Novikov, I.D., 1975, *Sov. Astron.*, **18**, 554.

Doroshkevich, A.G., Kotok, E.V., Novikov, I.D., Polyudov, A.N., Shandarin, S.F. and Sigov, Y.S., 1980a, *M.N.A.R.S.*, **142**, 321.

Doroshkevich, A.G. and Zel'dovich, Ya. B., 1975, *Ap. Sp. Sci.*, **35**, 391.

Doroshkevich, A.G., Zel'dovich, Ya. B., Sunyaev, R.A. and Khlopov, M. Yu., 1980b, *Sov. Astron. Letters*, **6**, 457.

Dressler, A., 1980, *Ap. J.*, **236**, 351.

Efstathiou, G. and Eastwood, J.W., 1981, *M.N.R.A.S.*, **194**, 503.

Einasto, J., Joever, M. and Saar, E., 1980, *Nature*, **283**, 47.

Fabbri, R., Guidi, I., Melchiorri, F. and Natale, V., 1980, *Phys. Rev. Letters*, **44**, 1563.

— 1979, *Proc. 2nd Marcel Grossman Meeting* (Trieste).

Faber, S. and Gallagher, J., 1979, *Ann. Revs. Astron. Ap.*, **17**, 135.

Faber, S.M. and Jackson, R.G., 1976, *Ap. J.*, **204**, 668.

Fall, S.M. and Efstathiou, G., 1980, *M.N.R.A.S.*, **193**, 189.

Frye, J.N., Olive, K.A. and Turner, M.S., 1980, *Phys. Rev.*, **D22**, 2953.

Gott, J.R. and Thuan, T.X., 1976, *Ap. J.*, **204**, 609.

Gott, J.R., Aarseth, S. and Turner, E., 1979, *Ap. J.*, **234**, 13.

Gregory, S.A. and Thompson, L.A., 1978, *Ap. J.*, **222**, 784.

Hogan, C., 1981, preprint.

Janin, G., 1971, *Astron. and Ap.*, **11**, 188.

Juskiewicz, R., 1981, *M.N.R.A.S.*, **197**, 931.

Kirshner, R.D., Oemler, A., Schechter, P.L. and Shectman, S.A., 1981, *Ap. J. Letters*, **248**, L57.

Kolb, E. and Wolfram, S., 1980, *Nucl. Phys.*, **B172**, 224.

Kompaneets, D.A., Lukash, V.N. and Novikov, I.D., 1981, preprint.

Larson, R.B., 1974, *M.N.R.A.S.*, **166**, 585.

Lequeux, J. and Viallefond, F., 1980, *Astron. and Ap.*, **91**, 269.

Lyubimov, V.A., Novikov, E.G., Nozik, V.Z., Tretyakov, E.F. and Kozik, V.F., 1980, *Phys. Lett.*, **94B**, 266.

Melchiorri, F., Melchiorri, B.O., Ceccarelli, C. and Pietranera, L., 1981, *Ap. J. Letters*, **350**, L1.

Melott, A., 1981, preprint.

Miller, R.H. and Smith, B.F., 1979, *Ap. J.*, **227**, 407.

Norman, C.A. and Silk, J., 1980, *Ap. J.*, **238**, 158.

Olson, D.W. and Silk, J., 1979, *Ap. J.*, **233**, 395.

Partridge, R.B., 1980, *Ap. J.*, **235**, 681.

Peebles, P.J.E., 1976, *Ap. J.*, **205**, 318.

— 1980, *The Large-Scale Structure of the Universe* (Princeton: Princeton University Press).

— 1981, *Ap. J.*, **248**, 885.

Peebles, P.J.E. and Dicke, R., 1968, *Ap. J.*, **154**, 891.

Peebles, P.J.E. and Yu, J.T., 1970, *Ap. J.*, **162**, 815.

Press, W. and Davis, M., 1981, *Ap. J.* (in press).

Press, W. and Vishniac, E., 1980, *Ap. J.*, **236**, 323.

Reines, F., Sobel, H. and Pasierb, E., 1980, *Phys. Rev. Lett.*, **65**, 1307.

Rivolo, R. and Yahil, A., 1981, private communication.

Rubin, V.C., Thonnard, N., Ford, W.K. and Roberts, M.S., 1976, *A. J.*, **81**, 719.

Sato, H. and Takahara, F., 1981, *Prog. Theor. Phys.*, **66**, 508.

Silk, J., 1974, *Ap. J.*, **193**, 575.

— 1980, in *Star Formation*, 10th Saas Fee Course, I. Appenzeller, J. Lequeux and J. Silk (Geneva: Geneva Observatory), p. 133.

Silk, J. and Norman, C.A., 1981, *Ap. J.*, **247**, 59.

Silk, J. and Wilson, M.L., 1980, *Physica Scripta*, **21**, 708.

— 1981, *Ap. J. Letters*, **244**, L 37.

Szalay, A. and Silk, J., 1981, in preparation.

Tarenghi, M., Chincarini, G., Rood, H.J. and Thompson, L.A., 1980, *Ap. J.*, **235**, 724.

Tonry, J. and Davis, M., 1981, *Ap. J.*, **246**, 680.

Tully, B., 1981, preprint.

Weinberg, S., 1981, private communication.

White, S.D.M. and Rees, M.J., 1978, *M.N.R.A.S.*, **183**, 341.

White, S.D.M. and Silk, J., 1979, *Ap. J.*, **231**, 1.

Wilson, M.L., 1981, in preparation.

— 1982, *Ap. J. Letters*, **253**, L 53.

Wilson, M.L. and Silk, J., 1981, *Ap. J.*, **243**, 14.

Yahil, A., Sandage, A. and Tamman, G., 1980, *Physica Scripta*, **21**, 635.

Zel'dovich, Ya. B., 1970, *Astron. and Ap.*, **5**, 86.

DISCUSSION

AUDOUZE

At the beginning of your presentation you call for the existence of gaseous clouds to form galaxies. It seems to me that they are presently well observed. There is the case, for instance, of I Zwicky 18 or so-called "lazy" galaxies which are found to be a collection of gaseous clouds of about $10^9\,M_\odot$ just starting now to form stars (see e.g. the recent work of Lequeux and Viallefond 1980, *Astron. Astroph.*, 91, 269, on this topic).

PEEBLES

I do not agree that the isothermal massive neutrino scenario leads to pancakes or that in this scenario the first generation of hound objects is quite massive. Suppose that at high redshift some baryons are in tight lumps. I choose initial conditions that are orthogonal to adiabatic perturbations by assuming that there are no curvature perturbations. That means the mass excess in the baryon clump is balanced by a like deficit in radiation and neutrinos. Radiation and neutrinos would have the same initial distribution because they were in thermal equilibrium. As the baryon lump comes within the horizon it tries to smooth itself out but that is resisted by radiation pressure, so we get the original clump somewhat expanded plus acoustic waves in the baryon-radiation mixture. The latter are strongly dissipated by photon diffusion if the clump radius is small. When the neutrinos become non-relativistic they accrete around the clump. A simple scaling argument shows the resulting density to be $\rho \propto r^{-1/2}$. To get the desired velocity dispersion at $r \approx 10$ kpc the baryon clump mass would have to be $\sim 10^9\,M_\odot$. The neutrinos near the center do not collapse far, so they do not create much entropy. Thus the core radius is $\lesssim 3\,h^{-1}$ Mpc, which is a comfortably small value and an advantage over the pancake picture.

SILK

If the baryonic fluctuations are linear, the growth of small-scale fluctuations after the decoupling is suppressed relative to larger scales. This alone suffices to make first bound objects rather massive in the isothermal neutrino scenario. The effect becomes even more compelling if the isothermal fluctuations consist of baryon density inhomogeneities at very large redshift. Associated small

curvature perturbations result in the growth of large-scale density fluctuations once the expansion rate of the universe is dominated by nonrelativistic particles. While these secondary curvature fluctuations could in principle be subtracted off with an appropriately chosen admixture of the adiabatic mode, I see no compelling reason to do so.

SETTI

1) I would like to know what is the energy involved in the collisions of the 10^7 $M_\odot$ clouds you mentioned in your talk and what is the energy density involved at the appropriate redshift.

2) If we did not know that there were galaxies, would we be able to predict their existence from current theory?

SILK

1) Collision velocities are initially a few km/s and build up as the clustering of clouds develops. The limiting velocity is about 500 km/s; above this velocity the shocked clouds cannot cool and will become unbound. The energy density involved is comparable to the total binding energy of galaxies, since this provides a measure of how much dissipation must have occurred during the cloud collision-dominated galaxy formation process.

2) No.

HAWKING

You assume that the microwave background has not been scattered since a redshift of about 1000, but is it not possible that there may be intergalactic matter which has scattered it at $z \sim 10$? If so, would you then be able to rule out adiabatic fluctuations?

SILK

Rescattering at $z \gtrsim 10$ would smear out any preexisting anisotropy in the microwave background at an angular scale below about 20°. This corresponds to the angle subtended by a proper distance corresponding to the horizon scale at $z = 10$, the maximum scale over which causal effects could operate. Predictions of large scale anisotropy, in particular of the quadrupole or dipole anisotropies, are unaffected by rescattering.

GALAXY FORMATION

J.P. OSTRIKER
Princeton University Observatory

1 - Introduction

The universe as we see it in our own neighborhood is about as inhomogeneous as it could be while remaining consistent with our own existence and the known laws of physics. On every scale, from the solar system outwards to the local supercluster (10^{13} cm to 10^{26} cm), we find that most of the mass is in objects (planets, stars, galaxies, clusters) which occupy a small fraction of the volume. On these same scales we see very significant anisotropies: if we imagine spheres centered on the earth split by our equatorial plane, then the difference in mass contained in the northern and southern hemispheres would be a significant fraction of the total mass for each sphere on all scales from 10^{13} to 10^{26} cm.

Yet, when we survey the universe on the largest scale possible, studying the microwave background emitted at a redshift of $z \approx 1000$ and a distance from us of $\sim 10^{28}$ cm, we find an extraordinary degree of isotropy with large scale fluctuations (dipole component) of the order of 2×10^{-3} and no fluctuations seen on galaxy or cluster mass scales which are as large as 10^{-4} (cf. the papers by Silk and Peebles in this volume for a review of the observations). How did so much structure grow from such smooth beginnings? This is possibly the most important question in cosmology today and a more or less standard approach to solving it has developed. One postulates a spectrum of perturbations of unknown origin in the early universe which must be small enough in amplitude at the decoupling of radiation from matter so that fluctuations in the microwave background be no larger than are observed. Then, depending on whether one hypothesizes adiabatic or isothermal perturbations and depending on the initial spectrum assumed, one integrates forwards in time allowing for growth,

decay and mode-mode coupling, until some of the perturbations grow to unit amplitude. This era when, metaphorically speaking, the waves break, should be the (wavelength dependent) epoch at which self-gravitating systems separate out of the background. These, depending on one's theoretical beliefs concerning the spectrum of perturbations, will be globular clusters (Peebles and Dicke 1968), galaxies (cf. review by Gott 1977) or clusters of galaxy-sized "pancakes" (cf. Zeldovich 1978).

Logically, this picture is attractive. The theoretician follows the same course as nature, i.e., forward in time from $(1 + z) \approx 1000$ to $(1 + z) \approx 3$ to 10, the era of galaxy formation. In practice there are great difficulties, since we are extremely uncertain about the initial conditions. We do not even know, for instance, whether baryons or neutrinos carry most of the mass and we are not really confident that we know all the relevant physical processes which affect the growth and decay of various modes.

Another, more low-brow, approach suggests itself. We can study the part of the universe we know best, our galaxy and those galaxies within the local supercluster, to learn two things: (1) what are the general properties of galaxies that must be "explained" by a theory of galaxy formation?; (2) what physical processes are important locally in the formation of self-gravitating systems (stars and star clusters) which might also be important on the cosmological scale (galaxies and galaxy clusters)? After we gain confidence studying these better, but still imperfectly, understood systems, we can then extrapolate backwards from presently observed galaxies at $(1 + z) = 1$ to 2 to the era of galaxy formation at $(1 + z) =$ 3 to 10. While the philosophical difficulties with this approach are clear (how, for example, does one tie the model of galaxy formation derived by this backward looking approach to conditions in the early universe?), it has the advantage that one is laying out a proposed map over the relatively recent epochs of cosmic time which are now or will soon be accessible to direct observations; i.e., it is, in principle, testable.

The important facts that we learn from our survey of the local universe are two: (1) galaxies show a surprising degree of uniformity requiring very few parameters to predict most of their properties and, furthermore, the distribution of these properties (mass, radius, etc.) show characteristic scales, implying that other than purely gravitational processes were important at the birth of galaxies; (2) non-conservative energy processes are very important locally in the formation of self-gravitating systems with both the radiative losses from cooling gas and the energy input from

massive stars essential ingredients. Complex feedback processes are evident, wherein star formation begets (or can limit) more star formation and, in any case, all memory of initial conditions is quite lost in the ongoing process, except for constraints imposed by overall conservation laws relating to mass, angular momentum, etc.

I will discuss these two topics briefly in Sections 2 and 3, turning in Section 4 to a specific theory of galaxy formation, and presenting summary remarks in Section 5.

2 - Properties of galaxies

A) *Definition of Components*

Baade (1944a, b) is usually credited with the recognition that the distinct stellar components observed in our galaxy in the vicinity of the Sun have parallels in other systems, so that galaxies may be considered to contain a mixture of more or less distinct components. A specific classification scheme of galactic components, developed on the basis of earlier work, received widespread adherence following Blaauw's (1965) summary of the 1957 Vatican Conference on Population Types. According to present views this classification scheme remains largely valid and is seriously incomplete only insofar as it omitted mention of the dark halo, now believed to be a dynamically important component (although its composition remains quite unknown). However, at the time of the 1957 Vatican Conference the properties of these components were known only locally with any security. We now have a fairly accurate picture of their spatial distribution and of the correspondence between the locally observed components and constituents of other galaxies. A specification of four components suffices to provide a moderately complete characterization of most galaxies:

a. *The Spheroid*: The approximately spherical distribution of old stars and globular clusters with radial dependence well approximated by the de Vaucouleurs law (1959) seems fundamental to most stellar systems. Locally, stars in the spheroid are identified with the metal poor "Population II" and in the center of the galaxy with the more metal rich "Galactic Bulge". Similar spheroidal distributions are seen in most spiral galaxies and the basic, elliptical components of elliptical galaxies appear to be identical except for scale factors.

b. *The Disc*: The flattened distribution of relatively metal rich, middle aged and old stars with radial dependence well approximated by an exponential in surface density (de Vaucouleurs 1959, Freeman 1970) that may or may not contain a central "hole". The Sun belongs to this component. It is seen in other spirals, often shows weak spiral structure (cf. Schweizer 1976), and dominates the light in some S0 systems.

c. *Population I*: The very flattened distribution of gas and young stars with surface density distribution like that of the disc (exponential), but also showing, typically, prominent spiral arm patterns. As to mass, this component is always insignificant (except possibly in very low mass systems), but it is most important in the chemical evolution of galaxies and dominates the total light in late Hubble type systems.

d. *The Dark Halo*: The approximately spherical distribution of some collisionless fluid (low mass stars, neutrinos, black holes...?) which contains, in our galaxy, perhaps 1/4 of the mass interior to the Sun's orbit, but 9/10 of the mass interior to 100 kpc, with density dependence approximately $\rho \propto r^{-2}$ in its outer parts.

A more complete and quantitative description of these components is given by Caldwell and Ostriker (1981) and Bahcall and Soneira (1980). A historical review and a relatively detailed and up-to-date presentation is given by Mihalas and Binney (1981).

B. *Components: Statistical Properties*

The basic spheroidal components show an extraordinary degree of uniformity with most astronomical examples, from the 10^{11} $L_\odot$ spheroid of M 87 to the 10^8 $L_\odot$ of tbe local group dwarf NGC 205, showing such family resemblances that they all appear to belong to a single one-parameter family. Sandage (1972) noted the color-luminosity relation for ellipticals which, it now appears, can be extended to the spheroidal component of spiral galaxies. Faber (1973) found a metallicity-color relation, Faber and Jackson (1976) a velocity dispersion vs. luminosity relation and Kormendy (1977) and others a core radius vs. luminosity relation (this last showing considerable scatter).

Thus, the knowledge of one parameter, *the luminosity*, allows one to make a prediction of almost any other measurable property of a spheroidal system to an accuracy of better than 20% of the dispersion of that quantity

in normal galaxies. Furthermore, following the extremely plausible arguments presented by Eggen, Lynden-Bell and Sandage (1962), we believe that this subsystem of our Galaxy and, by extension of other galaxies, was formed first and in the relatively short time frame of a few free-fall times ($< 10^9$ yrs). It is remarkable that the basic component of galactic systems can be specified by one number. Reasoning from first principles one would have expected that, at the minimum two parameters, mass scale (M) and radius (R), would have been required (or possibly three, M, R and angular momentum J). Alternatively, one could choose the two parameters to be the epoch of formation and mass. The observations indicate clearly that either most galaxies formed at an essentially fixed epoch, or that there was a tight relationship between epoch of formation and mass of system formed (with most massive systems formed last).

To a surprising degree, specification of the luminosity of the spheroidal component ($L_{sp,\,B}$) allows a good prediction of the other components as well. For example, if $L_{sp,\,B} > 10^{10.5}\,L_\odot$, then the other observed components are usually relatively faint; the system is classified as elliptical. If $10^{10.5}\,L_\odot > L_{sp,\,B} > 10^{9.5}\,L_\odot$; then the system disc population makes a significant contribution but Population I is relatively unimportant and the galaxies are classified as S0 or Sa. If $L_{sp,B} < 10^{9.5}\,L_\odot$, Population I typically dominates the light (a result possibly affected by selection). However, at the extreme low luminosity end of tbe sequence, dwarf spheroidals are known with essentially no Population I or disc components. The fact that environmental influences are also clearly important (cf., for example, Melnick and Sargent 1977 or Dressler 1980) in determining the ratios of different galactic types to one another depending on galaxy density, may or may not require an extra parameter. It is possible that low luminosity spheroids are relatively more rare in regions of a high galaxian density, thus accounting for the relatively small number of spiral systems; or, alternatively, they may be present in the same relative numbers as in the field, but in dense systems of galaxies they do not occur within disc galaxies. Analysis of existing observations should allow one to answer this question.

Thus, galaxies can be thought of as a one-parameter set of objects.[1] The sequence is, of course, very similar to the familiar Hubble-Sandage sequence.

[1] It is at first surprising that angular momentum is as unimportant a separate parameter for galaxies as it is for stars. Although it may be important in individual systems it appears that its value can be predicted from the value of other quantities so that it need not be treated as an independent parameter.

The important new fact is that, quantitatively, most properties along this sequence can be predicted from the luminosity of the spheroidal component.[2] As an example we examine how the Holmberg radii (R_H) of spiral galaxies might be correlated with L_{sp}. We know that the velocity dispersion in the bulge satisfied a Faber-Jackson relation $v_{B,\,rms} \propto L_{sp}^{1/4}$ (Whitmore, Kirshner and Schechter 1979) and that the rotation velocity of the Population I component satisfies $v_{rot} \approx \sqrt{3}\, v_{disp}$ (cf., e.g. Gunn's paper in this volume) and that v_{rot} is well correlated with the Holmberg radius (Rubin *et al.* 1980). This implies that R_H should correlate well with L_{sp}.

The extreme simplicity, statistically speaking, of galaxy properties is an important clue to the processes of formation to which we shall return. It would seem that a fairly deterministic theory of galaxy formation, in which stocbastic effects were relatively unimportant, would have the best chance of explaining the observations.

C) *Physical Scales*

It is a very significant fact that definite scales are associated with galaxies. As an example, consider the physical scale length of the luminous parts. We do not, for example, know of many objects with characteristic scales $\lesssim 1$ kpc or $\gtrsim 100$ kpc and there is no obvious selection that would prevent us from finding such objects among nearby galaxies. This fact is not simply produced by our definitions — wbat we *call* galaxies — since there is not a large amount of light emitted by objects near the edges of the observed distribution of size; and the optical cosmic background (Shecbtman 1974) is well enough accounted for by existing objects to argue that it is impossible for *most* of the light to come from either very low or very high surface brightness galaxies. Quantitatively, it is perhaps easiest to make the point by considering luminosity as the variable rather than size. A convenient way of characterizing the galaxy luminosity function was found by Schechter (1976) who approximated the distribution with $N(\ell)\, d\ell = \ell^{-\gamma} e^{-\ell}\, d\ell$, where ℓ is tbe luminosity of the object in units of some fiducial luminosity $\ell \equiv L/L_*$ and γ is typically found to be in the range 1.0 to 1.5. The Schechter luminosity function has a "break" with few galaxies brighter than L_* and a characteristic luminosity of the typical[3]

[2] For a possible interpretation of this fact see Ostriker (1977).

[3] Typical in the sense that most of the light is emitted by galaxies having luminosity within one magnitude of L_*.

galaxy of order L_*. Others such as Abell (1964) have found other forms of the luminosity function to be more convenient but all proposed distributions show a break at some characteristic luminosity. As another example consider the "multiplicity function" of Gott and Turner (1979) defined so that $G(\ell)\, d\ell$ is equal to the light emitted per unit volume by groups of galaxies having luminosity in the range $\ell \to \ell + d\ell$, where ℓ is the group luminosity in the units of some fiducial luminosity. The function $G(\ell)$ is found to have the same shape for clusters as Abell found for galaxies — two power laws connected at a break — so that there is a characteristically well-defined typical luminosity of a cluster.

Returning to the distribution of properties of individual galaxies, we can couple the distribution of luminosity shown by spheroidal components with a power-law-like relation between the luminosity of the spheroid and other statistical properties such as velocity dispersion, to determine that there must exist characteristic values for these other properties as well.

One may immediately ask whence such physical scales might arise. Gravity itself has no scale and the phenomena thought to arise from purely gravitational effects also show no scale, as found, for example, by Peebles (1980) in the analysis of galaxy counts or by Aarseth, Gott and Turner 1979 in the analysis of gravitational n-body experiments.

Detailed modelling of galaxy formation inevitably includes gas dynamics and, more importantly, radiative cooling of gas clouds, since it seems impossible to form the observed bound, centrally condensed systems of stars without significant energy dissipation. The requirement that cosmical gas clouds be able to cool on the relevant time scale will, as noted by several authors (Ostriker 1974, Binney 1977, Rees and Ostriker 1977, Silk 1977), singles out certain scales as possible for self-gravitating entities. It is reassuring that the simplest arguments based on this principle lead to proposed values for the preferred scales close to the range observed. A similarity to stars is suggested.

In the following respects stars are like galaxies: a) for most (main sequence) stars the observed parameters are well correlated with one another and a one-parameter (mass or luminosity) sequence is a good approximation; b) most of the mass in stars is in objects within a factor of 10 of the solar mass, with characteristic values of quantities such as mass and radius determined, we think, during the process of star formation by the (atomic) physical processes of radiation generation and transfer.

It appears that galaxy formation must involve, in a critical way, radiative dissipation and other physical processes, and that a purely gravi-

tational approach to the prohlem is likely to fail. Also the morphological analogies to stellar properties lead one to believe that analogous processes may occur in star formation and galaxy formation.

3 - Star formation in the interstellar medium

Since the model for galaxy formation proposed will be conceptually very close to the present picture of star formation, it is perhaps best to review very briefly current ideas on that subject. The interstellar medium (ISM) is quite complex with a description of several components required to model tbe observations or understand the known physical processes.

Most of the volume in the solar neighborhood, it is generally agreed, is filled with hot low density coronal gas having temperatures $\sim 10^{5.7}$ K, density $\sim 10^{-2.5}$ cm^{-3} and cooling time $\sim 10^{6.6}$ yrs. There are theoretical reasons for believing that this region is not very atypical and that, by volume, the coronal gas dominates in most of the galaxy. Imbedded in this hot gas are warm ($T \sim 10^4$ K) clouds which in some cases have cool ($T \sim 10^2$ K) cores, the "diffuse interstellar clouds", which are in pressure equilihrium with the hot medium. The most massive of these clouds ($M \gtrsim 10^5 M_\odot$) are colder still ($T \sim 10^1$ K) and primarily molecular in composition. The mass of the ISM is primarily in the cold clouds and particularly in the regions where there is a high concentration of massive molecular clouds.

The extremely inhomogeneous nature of the ISM is now well understood and the explanation of its origin probably applies to the intergalactic medium (IGM) as well. There are two complementary reasons; one related to heating, the other to cooling. First, consider heating. The energy input to the medium comes primarily in two forms. Photoionization produced by the 13.6 to 40 eV photons emitted by hot stars (10^{-25} ergs cm^{-3} s^{-1} on the average) will characteristically heat low density gas to 10^4 K and higher density gas to lower temperatures. A comparable amount of energy is injected into the medium by shock heating arising from explosions and winds of all kinds, with standard type I and II supernovae probably making the largest input. The shock velocities are typically in the range 10^2 to 10^3 km/s and tbey naturally heat gas to temperatures in the range $10^{5.5}$ to $10^{7.5}$ K. Furthermore, shock heating promotes inhomogeneity insofar as shocks tend to propagate in the low density component of a multicomponent medium hy passing and compressing (isothermally) higher density imbedded clouds. The second reason is related to the thermal instability of cooling gas (cf.

Field 1965 [4]) and has its origin in the cooling peak at 10^4 to 10^5 K, due to resonance line radiation of cosmically abundant elements. Shock heated gas will, on a cooling time scale, condense into new or onto previously existing cool clouds. Estimated values for the IGM show that analogous components ("hot" and "warm") probably exist in similar ranges of temperature and density.

The spectrum of cloud sizes in the ISM is limited on the upper end by gravitational collapse and on the lower by evaporation; both of these processes are also likely to occur for clouds in the IGM. Thus the heating of both the principal components, even the existence of hot and warm components in the ISM, is dependent on the continual energy input from young stars with lifetimes typically $\ll 10^8$ yr and which, therefore, must be steadily replenished over the lifetime of the galaxy. Processes leading to the birth of these stars are very imperfectly understood at present but several elements can be isolated as generally agreed upon: (a) star formation occurs primarily (or entirely) in gas rich regions of the highest density and lowest temperature; (b) gravitational collapse of clouds (perhaps initiated by an ambient overpressure) is likely to be important; (c) there is some observational evidence for "chain reactions" (Elmegreen and Lada 1977, Elmegreen and Moran 1979), the interpretation being that hydrodynamical energy released by young stars (in expanding H II regions, winds or supernovae) compresses by shocks adjacent cold clouds leading to isothermal collapse of these regions and more star formation.

Nowhere in this schematic picture of star formation is there any discussion of the initial spectrum of irregularities that existed when the galaxy was formed. Presumably this is not because it is thought that the galaxy was a perfectly spherical, uniform density ball of gas at that time (it probably would have collapsed to a black hole if it had been!) but because any memory of the initial spectrum of perturbations has been erased by the ongoing physical processes in the ISM. A typical time scale is the heating or cooling time of $10^{6.6}$ years (also approximately the supernova remnant overlap time) which is much shorter than the age of the galaxy. It is important to ask if the same situation applies in the IGM. Formation of some galaxies will clearly affect the surrounding IGM. Will effects from the various centers of activity overlap on time scales short compared to the age of the universe so that we can consider the evolution of an IGM largely decoupled from its initial conditions?

---

[4] A specific application to galaxy formation is included in this paper.

4 - Galaxy formation

It is not difficult to estimate the energy output by a standard ($L_{i_*} = 10^{10.3}\ L_\odot$) galaxy in the form of shock waves from young stars. Estimates arrived at by several different means give 10^{61} ergs/L_* galaxy, most of it, of course, emitted during a brief early epoch coincident with the formation of the spheroidal component. Given a present mean luminosity density of $j_0 = 10^{8.3}\ L_\odot/Mpc^2$ in the universe, this corresponds to an energy input of 10^{59} ergs per current Mpc or

$$\bar{u}_0 = 10^{-14.5}\ \text{erg/cc} \ . \tag{1}$$

If we equate this energy density to that of a monotonic gas for which $\tilde{P} = nT = 2\ u/3\ k$, then $\tilde{P} = 10^{1.2}$. A comparable energy was injected by QSOs and active galaxies (cf. Ikeuchi 1981, Blandford 1981), and it is quite possible that the two phenomena are really the same (or closely related) and distinguished by primarily observational limitations. If this energy had been injected at an epoch $z = z_i$ then, due to adiabatic losses, the energy density at some epoch $z < z_i$ in the IGM would be

$$u\,(z) = \bar{u}_0\ \frac{(1+z)^5}{(1+z_i)^2} \ , \tag{2}$$

so that, for example, if $z_i = 3$ then the current value of u would be $10^{-15.7}$ erg/cc and the current value of $\tilde{P}$ would be 10^0. Can disturbances propagate between galaxies? Ignoring (conservatively) the two-particle correlation function which makes the typical distance between galaxies considerably less than $n^{-1/3}$, tbe mean distance between typical galaxies is approximately

$$\lambda = (j_0/L_*)^{-1/3} = 10^{0.7}\,(1+z)^{-1}\ \text{Mpc} \ . \tag{3}$$

If we express the local density of the IGM in units of the critical density $n_{gas} = 10^{-7}\ \Omega_{gas,\,-2}$ atoms/cm^3, then the local speed of sound in the IGM is

$$C_0 = 10^{2.5}\,(P_0/\Omega_{0,\,gas,\,-2})^{+1/2}\ \text{km/s} \tag{4}$$

and scales as

$$C = C_0\,(1+z)^{+1} \ . \tag{5}$$

Thus the time required for a disturbance to propagate half way to the nearest

galaxy is $\Delta t = 10^{9.9}\ (P_0\Omega/_{0,\ \mathrm{gas}\ -2})^{-1/2}\ \ (1+z)^{-2}$ years. Expressing this in terms of the age of the universe at $z = z_i$ (and assuming an open universe) we have

$$\left(\frac{\Delta t}{t}\right)_i = 10^{-8.0}\,(\Omega_{0,\ \mathrm{gas},\ -2})^{+1/2}\ [u_0\,(\mathrm{erg/cc})]^{-1/2} \approx 0.2 \qquad (6)$$

independent of the initial epoch z_i. Thus there is ample time for communication between adjacent centers of activity if our estimates for energy input are even approximately correct. The *lower* the density in the IGM, the better is this approximation. This conclusion was arrived at by Schwarz *et al.* (1975) by an analogous but somewhat different route. Thus a model for galaxy formation analogous to star formation is at least causally possible. Is it however quantitatively reasonable?

I will discuss a specific model put forward by Ostriker and Cowie (1981) which is very similar to that of Ikeuchi (1981) based on quasars. We do not discuss the initiation of galaxy formation and it seems entirely possible that either of the standard scenarios that do investigate this, especially the Zeldovich pancake approach, could be modified to include many of the elements proposed here.

We imagine that a seed, a self-gravitating, stellar system (e.g. globular cluster or small galaxy) of mass M, forms with a normal complement of massive stars as determined by the initial mass function. Those stars over 10 $M_\odot$ have a short lifetime and are expected to die explosively as supernovae, injecting of the order of 10^{51} ergs of blast wave energy into the gaseous medium surrounding them. The total blast energy of all such supernovae E is expected to be approximately proportional to the mass of the system and can be expressed in terms of a dimensionless efficiency ε

$$E \equiv \varepsilon\, mc^2 \equiv 10^{-4}\,\varepsilon_{-4}\, mc^2\ , \qquad (7a)$$

numerically

$$E_{61} = 1.8\,\varepsilon_{-4}\, M_{11}\ , \qquad (7b)$$

where our ignorance of E has been rephased as uncertainty in the parameter ε_{-4}. M_{11} is the mass of the seed in units of 10^{11} $M_\odot$ and E_{61} the energy in units of 10^{61} ergs. Various independent arguments (cf. Ostriker and Cowie 1981, Bookbinder *et al.* 1980), based primarily on the observed metallicity of galaxies or the locally observed initial mass functions, indicate that $\varepsilon_{-4} \sim 0.3$ is a reasonable estimate for the efficiency. Much of the combined energy from these early supernovae will propagate into the IGM as an

adiabatic blast wave, initially similar to the standard Sedov-Taylor solution. Then under certain circumstances a dense, cool shell will form that may be gravitationally unstable to fragmentation and lead to the formation of new bound stellar systems. If all of the required conditions are satisfied and the mass in new systems is larger than the mass of the original seed, then there exists an amplifier which can cause a chain reaction of very rapidly growing rates of galaxy formation. Two further conditions must be met if the process is to modify significantly the rate of galaxy formation: the ε-folding time of the process must be short compared to the age of the universe and blast waves from different centers must not overlap before cooling and fragmenting. For certain epochs and proposed galaxy masses all of the required conditions are met as outlined in Ostriker and Cowie (1981) and spelled out in greater detail in Vishniac, Ostriker and Bertschinger (1982). Here let me focus for definiteness at a particular era, $z = z_i = 5$, and further assume that the cosmic density of primordial gas at the time corresponds to a current value of $(\rho/\rho_{crit})_0 = 0.2$ with an assumed value of the Hubble constant equal to 100 km/s/Mpc. Then, from an explosion of energy E_{61} a dense shell would have formed after a time of $9 \times 10^8\ E_{61}{}^{0.2}$ years with a radius of $R_{cool} = 0.67\ E_{61}{}^{0.3}$ Mpc, a velocity of $300\ E_{61}{}^{0.05}$ km/s containing a mass of $9 \times 10^{12}\ E_{61}{}^{0.88}\ M_\odot$. Since the cosmic age at that time was $\sim 20 \times 10^8$ years, explosions more energetic than 10^{62} ergs would not have had time to cool. Thus we might imagine an explosion originating in a $2 \times 10^{11}\ M_\odot$ galaxy ($E_{61} = 1.0$); it would lead to amplification by a factor of 90 producing a system with present-day properties like that of a typical group of galaxies: mass $\approx 10^{13}\ M_\odot$ and size ≈ 1 Mpc. One definite prediction of this theory is that galaxies should be formed on sheet-like surfaces, a characteristic common to it and the Zeldovich pancake picture. An approximate stability theory indicates that the mode of maximum instability in the sheet corresponds to a mass of $1.5 \times 10^9\ M_\odot$, comparable to the relatively common Magellanic irregular type systems. This is near the lower bound of the instability strip which predicts that galaxies can be made in the mass range, $10^9 \lesssim M_g/M_\odot \lesssim 10^{12}$, in rough agreement with observational facts. Since galaxies form on an essentially two-dimensional surface, they will automatically satisfy a Faber-Jackson type relation between mass and velocity dispersion (cf. Ostriker 1982). Also in agreement with observations is the prediction that galaxies should typically be made in bound groups of mass $\sim 10^{12}$ to $10^{13}\ M_\odot$ and crossing times of $\sim 10^9$ years (cf. Gott and Turner 1977). The pancake model, I believe, would tend to predict larger masses and smaller crossing times.

However, much more work must be done before the theoretical distributions of either the luminosity function or the multiplicity function can be predicted with any confidence.

The rapidly growing instability will be limited, in any given volume of space, when adjacent blast waves overlap. This will leave the remaining gas in a two-phase medium similar to the ISM, with "clouds" of 10^4 K (photoionized) gas occupying a small fraction of the volume and being pressure confined by the hotter and lower density, shock-heated intercloud medium. Analysis of the QSO absorption lines, associated by Sargent *et al.* (1980) with intergalactic Lα absorbing clouds, indicates properties for the intercloud IGM similar to those predicted. The required pressure of the IGM, $\tilde{P} \equiv nT \geq 10^0$, is an important specific test of the model.

It is interesting to ask about the fate of explosions that might occur at an earlier epoch. Then, in addition to ordinary radiative gas cooling, inverse Compton cooling due to collisions with blackbody photons can be quite important. This process will allow quite massive and energetic explosions to cool with the ratio of cooling to the age of the universe being

$$\frac{t_{\rm cool}}{t} = \left(\frac{100}{(1+z)^{5/2}} \right) \quad , \quad (z \gtrsim 6) \tag{8}$$

independent of E. However, the cooling is so efficient that very large explosions will cool before they have swept up more mass than was in the seed, so no amplification will result. The maximum energy that will amplify is

$$E_{61,\,\rm max} = 4 \times 10^{13}\, \epsilon_{-4}^{2.5}\, (1+z)^{-9} \quad , \quad (z \gtrsim 6) \ , \tag{9}$$

which restricts the epochs of interest to $z \leq 30$. In the allowed interval, $6 \lesssim z \lesssim 30$, if the instability begins, it will likely not lead to objects like ordinary galaxies. Cooling is so efficient that the Jeans mass in the fragmenting shell is quite small and, for example, at $z = 25$, the mode of maximum instability in the shell has a mass of $2 \times 10^4\, M_\odot \times E_{61}^{-0.2}$. Thus, it is possihle that very massive stars rather than galaxies would he produced at this epoch; these have short lifetimes and leave no optically visible remnants. Thus, volumes of space suhject to the instability at an early epoch might now seem "empty". What volume would be enclosed by such apparently empty regions? So long as the explosions continue to amplify one can think of the hierarchy proceeding as a detonation wave outwards from the original seed turning matter to energy with an efficiency

ε. A detonation wave in an $\Omega = 1$ universe (appropriate for such early times) expands at the rate

$$R_s = 1.46\, \varepsilon^{1/2}\, ct \tag{10a}$$

or

$$E = 13.04\, \varepsilon^{3/2}\, c^3\, t^3\, \rho = 8.8 \times 10^{67}\, \varepsilon_{-4}{}^{2.5}\, (1+z)^{-3/2}\ , \tag{10b}$$

(Ostriker *et al.* 1982). Then equating (9) and (10b) gives $z_{max} = 6.7$ and $E_{61,\,max} = 4.1 \times 10^5\, \varepsilon_{-4}{}^{2.5}$. This gives a maximum radius of $5\, \varepsilon_{-4}{}^{1/2}$ Mpc at $z = 6.7$ and a current ($z = 0$) radius of $\sim 50\, \varepsilon_{-4}{}^{1/2}$ Mpc, comparable in size to the void found by Kirshner *et al.* (1981) from galaxy redshift surveys.

5 - Summary

a) Locally observed galaxies can be quantitatively considered to follow a one-parameter sequence, the defining parameter being the optical luminosity of the spheroidal component; this indicates either that most galaxies were formed during a relatively narrow interval of cosmic time, or that the mass of the formed galaxy was tightly correlated with the epoch of formation;

b) All quantifiable properties of individual galaxies show characteristic scales which indicate that gravity was not the sole force operating during formation; existing theoretical work leads one to believe that gas cooling processes set the scales for the observed systems;

c) A process exists whereby explosions originating in one forming galaxy will drive a shock into the surrounding medium, with gas cooling ultimately causing the formation of a dense, cool, gravitationally unstable shell, out of which a new generation of galaxies can emerge. The properties of the galaxies and groups so formed have the characteristic scales of observed systems.

Many further details of this model remain to be worked out. It should be thought of as complementary to, rather than a replacement of, existing theories since it does not address the question of the initiation of galaxy formation. In the usual theories a spectrum of fluctuations is followed in cosmic time until nonlinear density fluctuations form. The explosive processes described here can be thought of as an *amplifier*, multiplying by large factors certain parts of the spectrum at certain cosmic epochs. Its

virtue appears to be that the scales amplified and the epochs preferred agree with those indicated by our very imperfect empirical knowledge of galaxy formation.

I would like to thank many of my colleagues, especially L.L. Cowie, J.E. Gunn, C.F. McKee, P.J.E. Peebles, M. Rees and E. Vishniac, for helpful readings and suggestions that have contributed to this ongoing work.

REFERENCES

Abell, G., 1964, *Ap. J.*, **140**, 1624.

Aarseth, S.J., Gott, J.R. and Turner, E.L., 1979, *Ap. J.*, **228**, 664.

Baade, W., 1944a, b, *Ap. J.*, **100**, 137; 147.

Bahcall, J. and Soneira, R., 1980, *Ap. J. Suppl.*, **44**, 73.

Binney, J.J., 1977, *Ap. J.*, **215**, 483.

Blaauw, A., 1965, Galactic Structure in *Stars and Stellar Systems*, Vol. 5, ed. A. Blaauw and M. Schmidt (Chicago: Univ. of Chicago Press), p. 435.

Blandford, R., 1981, private communication.

Bookbinder, J., Cowie, L.L., Krolik, J.H., Ostriker, J.P. and Rees, M., 1980, *Ap. J.*, **237**, 647.

Caldwell, J.A.R. and Ostriker, J.P., 1981, *Ap. J.*, **251**, 61.

de Vaucouleurs, G., 1959, *Handbuch der Physik*, ed. S. Flugge (Berlin: Springer-Verlag), **53**, 311.

Dressler, A., 1980, *Ap. J.*, **236**, 351.

Eggen, O., Lynden-Bell, D. and Sandage, A., 1962, *Ap. J.*, **136**, 748.

Elmegreen, B. and Lada, J.C., 1977, *Ap. J.*, **214**, 725.

Elmegreen, B. and Moran, J.M., 1979, *Ap. J. (Lett.)*, **227**, L93.

Faber, S.M., 1973, *Ap. J.*, **179**, 731.

Faber, S.M. and Jackson, R.E., 1976, *Ap. J.*, **204**, 668.

Field, G.B., 1965, *Ap. J.*, **142**, 531.

Freeman, K.C., 1970, *Ap. J.*, **160**, 811.

Gott, J.R., 1977, *Ann. Rev. Astr. and Astroph.*, **15**, 235.

Gott, J.R. and Turner, E.L., 1977, *Ap. J.*, **213**, 309.

— 1979, *Ap. J.*, **216**, 357.

Ikeuchi, S., 1981, P.A.S.J., **33**.

Kirshner, R.P., Oemler, A.G., Schechter, S. and Shechtman, S.A., 1981, *Ap. J. (Lett.)*, **248**, L57.

Kormendy, J., 1977, *Ap. J.*, **218**, 333.

Melnick, J. and Sargent, W.L.W., 1977, *Ap. J.*, **215**, 401.

Mihalas, D. and Binney, J., 1981, *Galactic Astronomy*, 2nd ed., (San Francisco: W.H. Freeman & Co.), Sect. 4.5.

Ostriker, J.P., 1974, Lecture on "Galaxy Formation" at the 7th Texas Conference.

— 1977, *Proc. Nat. Acad. of Sci.*, **74**, 1767.

— 1982, Supernovae and the Formation of Galaxies, in *Proc. Cambridge NATO-ASI Conf. on Supernovae*, July 1981 (Cambridge Univ. Press), in press.

Ostriker, J.P. and Cowie, L.L., 1981, *Ap. J. (Lett.)*, **243**, L127.

Ostriker, J.P., McKee, C.F. and Cowie, L.L., 1982, in preparation.

Peebles, P.J.E., 1980, *The Large Scale Structure of the Universe* (Princeton, NJ: Princeton Univ. Press).

Peebles, P.J.E. and Dicke, R.H., 1968, *Ap. J.*, **154**, 891.

Rees, M.J. and Ostriker, J.P., 1977, *M.N.R.A.S.*, **179**, 451.

Rubin, V.C., Ford, W.K. and Thonnard, N., 1980, *Ap. J.*, **238**, 471.

Sandage, A., 1972, *Ap. J.*, **176**, 21.

Sargent, W.L.W., Young, P.J., Boksenberg, A. and Tytler, D., 1980, *Ap. J. Suppl.*, **42**, 41.

Schechter, P., 1976, *Ap. J.*, **203**, 297.

Schwarz, J., Ostriker, J.P. and Yahil, A., 1975, *Ap. J.*, **202**, 1.

Schweizer, F., 1976, *Ap. J. Suppl.*, **37**, 313.

Shechtman, S.A., 1974, *Ap. J.*, **188**, 233.

Silk, J., 1977, *Ap. J.*, **211**, 638.

Vishniac, E., Ostriker, J.P. and Bertschinger, 1982,

Whitmore, B.C., Kirshner, R.P. and Schechter, P.L., 1979, *Ap. J.*, **234**, 68.

Zeldovich, Y.B., 1978, The Large Scale Structure of the Universe, in *Proc. IAU Symposium No. 79*, ed. M.S. Longair and J. Einasto, p. 409.

DISCUSSION

REES

I wonder to what extent one can adopt some of your ideas on an "à la carte" basis, even if one feels uneasy about swallowing them completely. In particular would not your calculations of fragmenting cool sheets apply also in some versions of the "pancake" picture, where gas is squeezed in one dimension but may be expanding in the other two?

OSTRIKER

Some aspects of the model could presumably be carried over to the pancake theory. However, the magic value for the surface mass density will occur in gas cooling from a high temperature blast wave but will not necessarily occur for gas with increasing temperature in an accreting pancake.

SILK

Primeval galaxy searches have been unsuccessful out to $z \lesssim 5$. One possible solution may be that the initial protogalaxy collapse occurs slowly, lasting up to $\sim 10^9$ years.

OSTRIKER

Perhaps primeval galaxies are not observed due to dust obscuration associated with intervening galaxies.

HAWKING

How many galaxies are made from each seed galaxy, and so how many galaxies would you need?

OSTRIKER

Approximately one seed per group of several galaxies (10^{12} to 10^{13} $M_\odot$) is required and the seed may be as small as a globular cluster.

OORT

I would feel very happy with your attractive theory if it were not for the seeds. The seeds show that galaxies can be made in other ways. Have you given thought to the possibility of a direct confirmation of your model, for instance, a test in which you could directly observe the shells? And can you make superclusters?

OSTRIKER

Direct observations of galaxies (and gas) on two dimensional expanding surfaces would provide direct but not unique evidence for the theory. The point I wished to make is that, regardless of how galaxy formation begins, the hydrodynamical amplifier should take over quickly, since its time scale is much shorter than the Hubble time. Thus, almost all of the galaxies will have been made through the chain reaction, and small mass objects such as globular clusters might be remnant seeds.

DAVIS

How do your models compare with the observational constraints set by the X-ray background and optical background?

OSTRIKER

The Bookbinder *et al.* (1980) paper discussed thermal bremsstrahlung of the gas leaving galaxies and concludes that a few percent of the background could easily be produced by this process. I have not yet thought about non-thermal emission but it may be significant.

SETTI

As Davis just said, I am also worried about the contribution of these galaxies to the X-ray background. In fact, if the 10^{10} or so supernovae which explode emit X-rays as they are known to do, I have just estimated that the integrated contribution would exceed the X-ray background in the 2 to 10 kev range by at least a factor of 10. Moreover if the same supernovae do produce cosmic rays with the same efficiency which they are thought to have in our Galaxy, then it seems to me one has to be careful that they do not produce too many γ-rays via the $p-p$ interaction and Compton X-rays from the electron

component of the cosmic rays. This would again violate the relative backgrounds.

VAN DER LAAN

A second or third generation event finds itself in an environment with an extreme range of density contrasts. The explosion energy will then be dissipated in the very tenuous, very hot volume elements. How will this offset the efficiency of producing next generation galaxies?

OSTRIKER

Yes, there is a phase when the IGM will resemble the present ISM and blast waves propagate in the low density medium. They pass and crush or evaporate "clouds". Those clouds which collapse produce more galaxies and the process continues until we are left with a spectrum of stable or evaporating clouds. The parameters agree well with those deduced by Sargent *et al.* (1980).

VAN DER LAAN

The triggers should be strong to moderate radio sources. Their luminosities may be comparable to that of the usual radio galaxies but their radio morphology would be altogether different. The dumb-bell configuration will not be there. It is of interest to do the sums to see if current VLA sensitivity/resolution combinations warrant a search.

OSTRIKER

Yes, VLA searches would be valuable.

REES

If there are no genuinely primordial fluctuations in the gravitational potential, can your model account for gravitationally bound systems like the Coma cluster? If $\Omega \leq 1$, any region is initially gravitationally unbound and the production of peculiar motions would tend to make it even more so.

OSTRIKER

Galaxies will be made in small bound groups having a size of the order of

the present two-particle correlation length (2 Mpc compared to 5 Mpc). After this the same boosting process that occurs in the N-body simulations (even in the case of low values of Ω) will tend to propagate fluctuations to longer wavelengths. Superclusters cannot be made by the process unless explosions occur at some earlier epoch.

REMARKS ON A POSSIBLE PREGALACTIC « POPULATION III »

MARTIN J. REES
Institute of Astronomy
Madinglay Road, Cambridge

If the inhomogeneities in the early universe involve entropy fluctuations, rather than being purely adiabatic, then they will not be destroyed by radiative damping before recombination. If these inhomogeneities have a smooth spectrum with the amplitude decreasing towards larger scales, as is indicated by the data on galaxy clustering, then the amplitude at recombination may be $\gtrsim 1$ on scales up to 10^6 to 10^9 $M_\odot$. (This number is uncertain because of the unknown slope of the fluctuation spectrum). Since the Jeans mass just after recombination is $\sim 10^6\,\Omega^{-1/2}\,M_\odot$, objects of mass $\gtrsim 10^6$ $M_\odot$ would therefore be expected to condense gravitationally at $z \simeq 1000$. Before one can formulate a real theory of galaxy formation based on initial entropy fluctuations, it is essential to have some idea of what happens to these post-recombination condensations of sub-galactic mass. Do they turn into supermassive black holes or do they fragment down to stellar masses? Does the energy input from these objects affect the residual gas by heating it up, by generating "secondary" fluctuations via bulk motions, or by injecting heavy elements? What fraction of the initial gas gets trapped in these "population III" objects or their remnants? Can these objects, after they have undergone hierarchical clustering, provide the "hidden mass"? I should like briefly to report some work on this subject which I have been doing with A. Kashlinsky.

Two interesting papers written in the 1960s addressed this question, but reached quite different conclusions. Peebles and Dicke (1968) considered the fate of clouds of just above the Jeans mass. They argued that the clouds would fragment into stars, yielding systems resembling globular clusters. In contrast, Doroshkevich *et al.* (1967) envisaged that each cloud formed a single supermassive object. The energy thereby released would

heat the intergalactic medium, exert a negative feedback on the formation of further supermassive objects, and would generate "secondary" inhomogeneities on larger mass-scales from which galaxies would later condense (an idea reminiscent of that developed by Ostriker in his paper at this meeting).

Kashlinsky and I have tried to consider the evolution of clouds substantially above the post-recombination Jeans mass (i.e. of 10^7 to 10^8 $M_\odot$). We have noted two physical effects which could be important:

(i) Each cloud would have acquired some angular momentum via tidal interaction with its neighbours before it started to collapse. The amount of angular momentum can be conveniently parametrised in terms of binding energy E and angular momentum J by the quantity $\lambda = \frac{J\,|E|^{1/2}}{GM^{5/2}} \simeq \frac{V_{rot}}{V_{free\ fall}}$. Numerical simulations (Aarseth and Fall 1980, Efstathiou and Jones 1979) show that clouds should acquire a range of angular momenta, the quartile points for λ being 0.03 and 0.09. The cloud would thus become rotationally supported after collapsing by a factor $\sim \lambda^2$, if it had not already fragmented before that stage;

(ii) If we are concerned with collapse soon after recombination, the background temperature will be 1000 K. This means, obviously, that no part of the cloud can get cooler than this. Also, even if only a small fraction of the material gets re-ionized, Compton drag may become important.

A recent discussion by Tohline (1980) suggests that fragmentation may not occur until the cloud has collapsed by a large factor. The possibility of Compton drag would inhibit fragmentation still more in this particular context. For this reason, we consider it plausible that the gas may increase its density by a factor λ^{-6} to λ^{-8} before fragments have separated out. (The factor λ^{-6} is reached when the central part of the cloud becomes affected by rotation, the full factor λ^{-8} would be attained if the entire cloud developed into a thin rotation-supported disc). If enough H_2 forms to permit cooling below 10^4 K, these densities are enhanced further by a factor, $(T_{gas}/10^4\,K)^{-1}$.

The naively estimated Jeans mass at the end of such a collapse is below ~ 1 $M_\odot$; and we see no reason why, in principle, the gas could not all end up in low-mass stars. On the other hand, most of the material in the original cloud might end up in a single supermassive object. This could happen either for the general reasons outlined by Larson (1978), or

because, when the outer part of the cloud attempts to settle into a disc, viscous effects (including those arising from the gravitational instabilities themselves) heat and thicken the disk, causing the gas to drain instead onto a central object.

It would of course be unrealistic to expect to predict the masses of Population III objects with any precision. After all we are still unclear about what determines the IMF in the sites of present-day star formation in our Galaxy, and we are even further from understanding the fragmentation process at early epochs where physical conditions were very different. However, our considerations have made us even more sceptical and open minded than earlier workers have been. Fragmentation depends critically on very uncertain parameters: the degree of inhomogeneity in the primordial clouds on mass-scales below 10^6 $M_\odot$, the amount of viscosity in massive discs, etc. Until these parameters can be quantified, it seems worthwhile to explore the consequences for galaxy formation of two extreme alternative hypotheses: (i) the hidden mass is in a population of low-mass stars; (ii) there was a pregalactic population of supermassive objects. We note also that the fate of a 10^6 to 10^8 $M_\odot$ cloud may be sensitive to the value of its angular momentum parameter λ. Since there will be a factor of ~ 3 spread in λ from cloud to cloud, there may be two different modes of evolution (corresponding to different ranges of λ) for a coeval population of clouds.

REFERENCES

Aarseth, S.J. and Fall, S.M., 1980, *Astrophys. J.*, **236**, 43.
Doroshkevich, A.G., Zeldovich, Y.B. and Novikov, I.D., 1967, *Sov. Astron.*, **41**, 233.
Efstathiou, G. and Jones, B.J.T., 1979, *M.N.R.A.S.*, **186**, 133.
Larson, R.B., 1978, *M.N.R.A.S.*, **184**, 69.
Peebles, P.J.E. and Dicke, R.H., 1968, *Astrophys. J.*, **154**, 891.
Tohline, J.E., 1980, *Astrophys. J.*, **239**, 417.

DISCUSSION

PEEBLES

If the gas cloud mass were near the Jeans limit and if the gas were appreciably ionised, it would expand rather than contract.

REES

Tbe model I have discussed applies to clouds which would still be collapsing even if they were at 10^4 K (rather than only at the recombination temperature of $\sim 3 \times 10^3$ K). The relevant masses are therefore 10^7 to 10^8 $M_\odot$. Note that the mass scale on which $\delta\rho/\rho \approx 1$ at the recombination temperature (for isothermal fluctuations) may well be substantially higher than the post-recombination Jeans mass.

OSTRIKER

I had the impression that cold self-gravitating discs were always unstable to the bar mode on the time scale of one rotation period. How would this affect your scenario?

REES

Yes; the disks I have discussed would of course be unstable to all modes from the bar mode down to the Jeans length. The bifurcation would happen on a dynamical timescale and the growth rate for Jeans length instabilities would be even faster. But the crucial question is whether this *does* happen; or whether, contrariwise, the initiation of these instabilities so greatly enhances tbe effective viscosity and internal dissipation that the disc thickens and becomes centrally condensed to a sufficient degree that the negative feedback prevents the fragments from separating out.

SILK

I would like to point out one effect which suggests that primordial clouds are capable of fragmentation down to stellar-mass scales during their isothermal

collapse phase. A spherical, uniform, pressure-free cloud that collapses from rest and contains density fluctuations of amplitude δ will undergo fragmentation after the mean density has increased by a factor of $\sim \delta^{-2}$. The value of the minimum fragment mass must, therefore, exceed the minimum Jeans mass in order for the fragmentation to have occured before the collapse becomes adiabatic by a factor δ^{-1}.

Tbis could be considerable and may inhibit fragmentation, especially if pressure forces delay tbe growth of fluctuations during the initial isothermal collapse phase. However, relatively modest deviations from sphericity qualitatively modify this result. For example, analysis of the fragmentation of uniform oblate spheroids collapsing from rest indicates that fragmentation occurs after a density increase by a factor of δ^{-1} for initial flattenings (ratio of sbort to long axis) of $\lesssim 0.7$. Since we expect any realistic collapse situation to involve appreciable deviations from spherical symmetry, we expect the minimum fragmentation mass in a primordial cloud to be of the order of $\delta^{-1/2}\ M_\odot$. This implies that a primordial collapsing cloud could fragment into stars with masses 10 to 100 $M_\odot$, if there are initial density fluctuations of the order 1 to 10 percent.

VI.

THE VERY EARLY UNIVERSE AND PARTICLE PHYSICS

ELEMENTARY PARTICLE PHYSICS IN THE VERY EARLY UNIVERSE

STEVEN WEINBERG
Lyman Laboratory of Physics
Harvard University

and

Smithsonian Center for Astrophysics
Cambridge, Massachusetts

and

Department of Physics
University of Texas
Austin, Texas

Several conference participants have expressed to me their wish to get some feeling for how confident elementary particle physicists are in the remarkable things they have been saying about the very early universe. I will try to show here that we particle physicists are almost as cautious, modest and prudent as you professional cosmologists. To do this I will depart from the usual format for this sort of talk — a breathless tour through the first 0.01 seconds — and instead organize the talk around some of the major developments in theoretical physics of the past few years, emphasizing the physics background of the recent applications of particle physics to cosmology, so that you can judge for yourselves what fraction of this work should be taken seriously.

1 - Spontaneous Symmetry Breaking

The first topic is spontaneous symmetry breaking. One of the most important developments in physics in recent years has been the discovery that nature is far more simple and symmetric than it appears at first sight;

there are a number of symmetries that govern the underlying equations of particle physics hut are not apparent in the phenomena that are ohserved. Such broken symmetries are often restored hy raising the temperature, so we can conclude that at an earlier time in the history of the universe these symmetries were manifest. It appears that the universe has gone through a sequence of phase transitions, at which as the temperature dropped various symmetries became successively broken.

The first of the spontaneously broken symmetries to have an impact on elementary particle physics (hut the last to get hroken in the history of the universe) was the "chiral" symmetry of the strong interactions. This hroken symmetry is a development of the early and mid 1960's. To explain it, I will rely on a well-known analogy between strong interactions and magnetism. Just as the equations that govern a magnet are invariant under three-dimensional spatial rotations, the modern theory of strong interactions known as quantum chromodynamics has a so-called chiral symmetry, which can be thought of as consisting of four-dimensional rotations on the four-vector

$$V_1 = \bar{u}\,\gamma_5\, d + \bar{d}\,\gamma_5\, u \qquad V_2 = -i\,\bar{u}\,\gamma_5\, d + i\,\bar{d}\,\gamma_5\, u$$

$$V_3 = \bar{u}\,\gamma_5\, u - \bar{d}\,\gamma_5\, d \qquad V_4 = \bar{u}\, u + \bar{d}\, d$$

where u and d are the Dirac fields of the up and down quarks, and γ_5 is the usual Dirac matrix. This symmetry if unbroken would require that the expectation value of the four-vector V must vanish, just as an unbroken ordinary rotational symmetry would rule out an expectation value for the magnetic field. But spontaneous magnetisation is possible, and in the same way the spontaneous breakdown of the chiral strong-interaction symmetry is signalled by the spontaneous appearance of a non-vanishing vacuum expectation value of V_4 (leaving unbroken the isospin group, consisting of three-dimensional rotations of V_1, V_2, V_3). This is expected to have occurred as the temperature of the universe dropped to a value comparahle with the characteristic energy scale Λ of quantum chromodynamics, say about 300 to 400 MeV.

The analogy can be carried further. Even apart from spontaneous symmetry breaking the rotational symmetry in a magnet is not perfect, being broken for instance by the weak external magnetic field due to the earth. Similarly, the four-dimensional symmetry of the strong interactions is not perfect, being intrinsically broken by the small bare masses of the up and

down quarks. At low temperatures the effective mass of the quarks arises mostly not from their bare masses, but from the spontaneous breakdown of the four-dimensional chiral symmetry, and hence is of the order of the characteristic energy scale Λ of quantum chromodynamics, about 300 to 400 MeV, but at temperatures above Λ the effective masses of the quarks (and of the proton etc.) drop to their bare values, a few MeV. One consequence of this intrinsic symmetry breaking is that the spontaneous breakdown of the chiral symmetry as the temperature is lowered does not take place in a sharp phase transition at one definite temperature, but over a small range of temperatures, in what is called a "rounded" phase transition. A great deal of work has been done in calculating the behaviour of solid-state systems in such rounded transitions, and it is easy to adapt these results to see that the spread of temperatures over which the strong-interaction phase transition takes place goes as the 0.538 power of the small bare quark masses.

Even though these symmetries are spontaneously broken at zero temperature, they are not without consequences: for each broken symmetry there is a zero mass excitation, i.e., one for which the frequency vanishes as the wave number goes to zero. For magnetism these are the spin waves; for the chiral symmetry they are the pions. Because the chiral symmetry is not exact, the pions are not really massless, only much lighter than other hadrons. At temperatures above 300 to 400 MeV there are no massless pions, only various chiral multiplets (4-vectors, 4-spinors, etc.) of nearly degenerate massive or nearly massless hadrons; at the phase transition one of these multiplets of spin zero hadrons becomes nearly massless; and at lower temperatures it is only the pions that remain very light.

By the way, you will sometimes hear references to a quark-liberation phase transition, a temperature above which quarks are untrapped. It is widely believed that at low temperatures quarks are trapped; they can never get isolated because the quark-antiquark forces get stronger and stronger as we try to pull pairs apart. It is also believed that quarks behave as essentially free particles at very high temperature because the strong interaction coupling constant decreases toward zero with increasing energy. However, this does not imply that there is a sharp quark liberation phase transition, a specific temperature at which quarks suddenly become liberated. After all, if you heat a gas like air, the proportion of nitrogen or oxygen atoms bound in molecules gradually decreases, but there is no phase transition at which the air suddenly goes from diatomic to monatomic. There may be a quark-liberation phase transition, but this is not clear. We

can be much more confident about the phase transition at which the strong interaction chiral symmetry became spontaneously broken: a symmetry is either broken or not, and we know that chiral symmetry is unbroken at high temperature (because the QCD forces become too weak to bind quarks and antiquarks to form pions) and is broken at low temperature (because the observed proton mass is about 1 GeV, not a few MeV), so there must be an intervening phase transition.

The second example of spontaneous symmetry breaking that has become important in elementary particle physics is the breakdown of the "electroweak" symmetry [known as $SU(2) \times U(1)$] that connects the weak and electromagnetic interactions and also connects the neutrino with the electron and the charge $2/3$ quarks with the charge $-1/3$ quarks. Although it is not ruled out that this symmetry breakdown is also produced by strong ("extra-strong" or "technicolor") forces which bind the Goldstone bosons, it is more generally assumed that the spontaneous breakdown of the electroweak symmetry is due to the appearance of vacuum expectation values for one or more weakly coupled elementary scalar fields. If one adopts the simplest hypothesis, that there is just one doublet $\{\phi^+, \phi^0\}$ of these scalars, then from the observed value of the Fermi coupling constant of beta decay we can infer that at the present time the vacuum expectation value of ϕ^0 is

$$< \phi^0 > = 2^{-1/4} G_F^{-1/2} = 247 \text{ GeV} .$$

Another difference between the electroweak and the QCD phase transition is that the electroweak symmetry is a gauge symmetry, like electromagnetic gauge invariance, rather than an ordinary global symmetry, like chirality. In consequence, there are no Goldstone bosons; instead the spin-one sihlings of the photon, the $W^{\pm}$ and Z^0, acquire helicity zero components and thereby become massive, with $m_W = 40 \text{ GeV}/\sin\theta$ and $m_Z = 80 \text{ GeV}/\sin 2\theta$, where θ is the $Z^0-\gamma$ mixing angle, which must be taken from experiment. In the simplest case of one scalar doublet the only remaining physical spin zero particle is the "Higgs hoson" associated with the neutral scalar field $\text{Re}(\phi^0)$. This particle would not have been seen experimentally yet, but may be found in the next decade.

Another interesting thing about gauge symmetries (which has been understood for a long time in solid state physics) is that, for a range of parameters of the theory, the phase transition in which the symmetry becomes broken is not of second order, but at least weakly of first order. A second order phase transition is one in which the order parameter (the

quantity whose appearance signals the breakdown of the symmetry) rises suddenly but continuously from zero as the temperature drops below the critical value for the phase transition. This is the case for instance for the magnetisation and for the expectation value of V_4 in the examples of broken global (that is, not gauge) symmetries discussed earlier. A first order phase transition is one in which the order-parameter (e.g. $<\phi^0>$) rises discontinuously from zero to a finite value at the critical temperature, a small or a large jump for a transition that is weakly or strongly of first order. The reason for this behavior in gauge theories is that the curve of free-energy versus order parameter has two local minima, one with vanishing and the other with non-zero order parameter. As the temperature drops, the minimum with non-zero order parameter becomes deeper, and at the temperature where it becomes lower than the other minimum the order parameter jumps from zero to the location of the deeper minimum, perhaps with some delay or "supercooling".

In the simplest case of one scalar doublet, the parameters of the first-order electroweak phase transition can be calculated in terms of a single unknown: the mass of the Higgs boson, or equivalently the curvature of the potential at its minimum at zero temperature. There are two interesting cases:

(1) The Higgs boson mass, m_H, may be much larger than its minimum value, about 7 GeV. In this case the electroweak phase transition is at a critical temperature

$$kT_c = \frac{350\ \text{GeV}}{\left[1 + \left(\frac{100\ \text{GeV}}{M_{\text{Higgs}}}\right)^2\right]^{1/2}} .$$

The phase transition is only weakly of first-order, and occurs with very little delay. (This sort of phase transition was first considered by Kirzhnits and Linde; the methods for showing that it occurs and for calculating quantities like T_c in general theories were then worked out by Dolan and Jackiw and by me. There are interesting complications that enter when one tries to carry these calculations to temperatures very close to T_c, discussed recently by Ginsparg).

(2) The Higgs boson mass may be comparable with its minimum value, say of order 10 GeV. In this case the critical temperature is also of order 10 GeV, and the phase transition is strongly first-order and ap-

preciably delayed. This case has been considered by Witten and Linde. In a case like this, it is possible for the rate of penetration through the barrier that separates the two minima to be so slow that the world stays for a very long time in the vacuum state in which it started, that of unbroken symmetry, even though as the temperature changes this becomes no longer the deepest minimum of the free energy. Specifically, this happens for a vanishing scalar mass term in the Lagrangian, corresponding to a Higgs boson mass very close to $\sqrt{2}$ times its minimum value, or 9.4 GeV. (This is the case originally studied at zero temperature by Coleman and Weinberg, Erick Weinberg, not me. There are reasons discussed by Gildener and myself for suspecting that the Higgs boson mass may have this value, but this is something that will have to be settled by experiment. Unfortunately, a Higgs boson of mass 9 GeV, although not too heavy to be produced by modern accelerators, has decay modes which do not seem to lead to any easily identifiable signature). The transition from unbroken to broken symmetry occurs through the formation of buhbles of broken-symmetry vacuum, which appear at random in the unbroken symmetry vacuum, at a rate shown by Callan and Coleman to be of order (using units with $\hbar = c = 1$)

$$\Gamma \sim (kT)^4 \, e^{-A} \text{ bubbles cm}^{-3} \text{ s}^{-1}$$

where A is the barrier penetration factor, calculated in this case by Witten to be (for $A \gg 1$)

$$A = \frac{32\pi^2 \sin^2\theta}{3e^3} \quad \frac{[2+\sec^2\theta]^{1/2}}{[2+\sec^4\theta]} \quad \frac{18.897}{\ell n\,(m_W/kT)} \simeq 3900/\ell n\,(m_W/kT)\,.$$

(Related calculations have been done by Steinhardt; Guth and E. Weinberg; and Sher). In order to have at least one bubble per Hubble volume per Hubble time, it is necessary to have $\Gamma \gtrsim H^4$, where H is the Hubble expansion rate, which in the broken symmetry vacuum is of order $(Gm_W^4)^{1/2}$. This does not happen, or at least not until kT drops to the derisory value of $e^{-1000} m_W$, where A becomes comparable with unity and the WKB approximation breaks down. However, as pointed out by Witten, long before this the phase transition does occur, not by a tunneling through the barrier between the two minima, but by the disappearance of the barrier itself. At temperatures of the order of $\Lambda \approx 300$ MeV, the ordinary strong interactions produce a breakdown not only of the chiral four-dimensional symmetry but also of the electroweak symmetry, leading to terms in the free

energy which allow an immediate decay of the unbroken-symmetry vacuum into the deeper broken-symmetry vacuum. Still, even though the pbase transition does not wait for barrier penetration, it is striking that the universe stays in a metastable supercooled state of unbroken electroweak symmetry over a time long enough for the temperature to drop and the cosmic scale factor to increase by a factor of 1000.

Tbis sort of supercooling would have an effect of great cosmological importance. Taking the free energy density of the present vacuum state to be zero, the free-energy density in the unbroken-symmetry vacuum at temperatures $\ll$ 300 GeV must on dimensional grounds be of order $m_{\rm H}^4$ ($m_{\rm H} \approx 10$ GeV). After the phase transition this energy becomes thermalized so the temperature rises up to about $m_{\rm H}$ and the entropy density is about $m_{\rm H}^3$. Witten's more careful estimate gives $(24\ {\rm GeV})^3$. But just before the phase transition the temperature was only about 300 MeV, corresponding to an entropy density of about $(300\ {\rm MeV})^3$. That is, the non-equilibrium supercooling increases the entropy and the entropy-baryon ratio by a factor of 10^6!

It is not clear whether this result should be welcomed or deplored. About 15 years ago, when we first learned from the microwave background temperature that the entropy-baryon ratio is about 10^8 to 10^{10}, it seemed to many cosmologists that unity would be a far more plausible initial value. I tried hard to think of something like bulk viscosity which could produce a great deal of entropy, but nothing came remotely close to giving a 10^8 to 10^{10} fold increase. At that time I would have been delighted to think of something like supercooling which could increase the entropy by a factor of 10^6. However, as discussed below we now think that the entropy-baryon ratio was perhaps fixed at very early times at some large value. If this is the case, then a further increase by a factor of 10^6 may not be tolerable. But there is no problem if $m_{\rm H} \gg 10$ GeV.

2 - Baryon and Lepton Nonconservation

The second of our major topics is baryon and lepton nonconservation. Our thinking about this possibility has been powerfully affected by the discovery of the gauge symmetries that underly the strong and electroweak interactions. There is a theorem to the effect that, if you write down the most general renormalizable theory that is consistent with these symmetries and involves just the familiar quarks and leptons and gauge and Higgs

bosons, then there is simply no way to make such a theory complicated enough to violate the conservation of baryon and lepton number. Hence baryon and lepton conservation do not have to be assumed as fundamental symmetries in order to explain why they are conserved in ordinary strong and electroweak interactions, and this presents us with the possibility that they are *not* conserved in other interactions.

To be a little more specific, we can imagine that there may be "exotic" particles with masses given by some very large mass scale M, exotic in the sense that they have unfamiliar assignments of the quantum numbers of the strong and electroweak gauge symmetries. We would not observe these exotic particles in experiments at ordinary ($\ll$ M) energies, but their exchange could produce effective nonrenormalizable as well as renormalizable interactions at ordinary energies. By "nonrenormalizable" or "renormalizable" interactions we mean in effect interactions whose coupling constant is either a negative power of mass or dimensionless, respectively. I have already said that there are no renormalizable baryon or lepton nonconserving interactions allowed by the strong and electroweak gauge symmetries, so to encounter baryon or lepton non-conservation we must look to the nonrenormalizable interactions. Their coupling constants are then negative powers of some mass, and what mass could it be but the characteristic mass scale M of the superheavy particles? Thus baryon and lepton non-conservation may occur, but would naturally be expected to be suppressed by one or more powers of very large masses.

For instance, the least suppressed interaction which could lead to baryon nonconservation is a Fermi interaction among three quark and one lepton fields, whose coupling constant (like the Fermi coupling of beta decay) is the -2 power of a mass. The proton lifetime is inversely proportional to the square of this coupling, and hence proportional to the fourth power of this mass; in order for the lifetime to be greater than the experimental lower bound of $\sim 3 \times 10^{30}$ years, this mass must be greater than a few times 10^{14} GeV.

There are in fact at least three reasons for suspecting that superheavy exotic particles may actually exist. Tbe first is that as realized by Max Planck the combination of quantum theory with gravitation introduces a natural mass scale $m_{\rm PL}=1.2\times10^{19}$ GeV, and something new must enter in physics at or below the Planck mass scale to cure the problems of quantum gravity. The second reason has to do with grand unification, which I will come to later, and the third is connected with the scenario for creating baryons in the early universe, to which I will return shortly. All three argu-

ments point to the existence of a new mass scale greater than about 10^{14} GeV, and if there are exotic particles not much heavier than this then we may expect baryon nonconserving processes like proton decay at a rate we might be able to observe.

There are a number of experiments going on now in Europe, Asia, and the U.S. which look for proton decay. The two largest ones are in a salt mine in Ohio and a silver mine in Utah. In these experiments the decaying sample of matter consists respectively of 10,000 or 1,000 tons of water, and the decays are monitored by watching for the odd flash of Cerenkov light which would be produced by relativistic charged particles emitted in the decay of a proton or bound neutron. There is no news to report from these experiments as neither of them has yet filled their tanks. I asked how things were going in these experiments just before coming here. As I understand it, right now the price of both salt and silver are so high that the mining companies are too busy mining to allow the physicists to get their equipment down the elevators and fill their tanks. So we will have to wait for a little while and hope for a drop in the price of salt and/or silver.

If baryon number is not conserved then maybe one can explain the cosmic excess of baryons over antibaryons through physical processes in the very early universe. An early detailed study of this possibility was given by Sakharov, and more recently by Yoshimura; Dimopoulos and Susskind; Ellis, Gaillard, and Nanopoulos; and others. A number of theorists have worked to develop a realistic scenario for how a baryon number might actually have been generated. There is one scenario developed by Toussaint, Treiman, Wilczek, and Zee and by me that seems to me pretty plausible. It goes as follows.

Suppose that there is an exotic particle of very large-mass — we will call it an X particle — and suppose it decays in such a way as to violate baryon conservation. The usual sort of particle one thinks of in this context is a color-triplet boson, because such a particle could decay (by renormalizable interactions) into channels with different baryon number: either a quark (also a color triplet) and a lepton (color singlet), or into two antiquarks (two antitriplets can make a triplet). Suppose that such a boson, X, decays into these two channels, $q\ell$ and $\bar{q}\bar{q}$, with branching ratios r and $1-r$ respectively. Its antiparticle $\bar{X}$ will then decay via the anti-channels $\bar{X} \to \bar{q}\bar{\ell}$ and $\bar{X} \to qq$, with the same total rate (according to the TCP theorem) but in general with different branching ratios, say $\bar{r}$ and $1-\bar{r}$.

Quarks have baryon number 1/3, so the net baryon number produced when an X and $\bar{X}$ both decay is then

$$\Delta B = \frac{1}{3} r - \frac{2}{3}(1-r) - \frac{1}{3}\bar{r} + \frac{2}{3}(1-\bar{r}) = r - \bar{r} .$$

This demonstrates one point that is obvious anyway — no net baryon number is produced unless CP as well as baryon number is nonconserved, because CP would require that the branching ratios for corresponding X and $\bar{X}$ decays all must be equal. But that is not the end of the story. Even if there is an exotic color triplet X boson with branching ratios $r \neq \bar{r}$, you will not get any baryon production if you start with a state of thermal equilibrium, because then inverse decay processes $qq \to X$, $\bar{q}\ell \to X$, etc. will simply use up whatever baryon excess is produced in the X and $\bar{X}$ decays. What one needs is that the X and $\bar{X}$ boson decays occur at a time when the temperature is below the mass of the X bosons, so that a Boltzmann factor $\exp(-m_X/kT)$ can inhibit the inverse decays, and yet have X and $\bar{X}$ bosons still abundant, because the universe has been expanding too fast for their number to have had time to drop to the level expected in thermal equilibrium at these temperatures.

It is easy to work out the consequences of these requirements. The lifetime of the X or $\bar{X}$ bosons will be of the order of

$$\text{mean life} \approx 1/g_X^2\, m_X \qquad (\hbar = c = 1)$$

where g_X^2 is the coupling constant for the decay interactions, say of order 10^{-2} for gauge bosons but perhaps as small as 10^{-5} for scalar Higgs bosons. On the other hand the age of the universe at early times is of the order

$$\text{age} \approx \frac{1}{H} \approx \sqrt{\frac{3}{8\pi G\rho}} \approx \frac{m_{PL}}{(kT)^2} ,$$

where m_{PL} is the famous Planck mass, equal to $1/\sqrt{G}$ in units with $\hbar = c = 1$, or 1.22×10^{19} GeV. The X and $\bar{X}$ bosons will decay when this age becomes comparable with their lifetime, i.e., at a temperature

$$kT \approx (g_X^2\, m_{PL}\, m_X)^{1/2} .$$

But to avoid the inverse decays we want $kT \lesssim m_X$ so we want

$$m_X \gtrsim g_X^2 \, m_{PL} = \begin{cases} \approx 10^{17} \text{ GeV} & \text{gauge bosons} \\ > 10^{14} \text{ GeV} & \text{scalar bosons} \end{cases}$$

These are very large masses, but it is clear that cosmological baryon production will not work unless there are very heavy particles. This is one of the reasons I referred to earlier why we want to consider particles of extremely high mass.

How much baryon number is actually produced in this way? We must consider two cases, either that m_X is greater than $g_X^2 \, m_{PL}$, or that it is somewhat smaller.

If $m_X > g_X^2 \, m_{PL}$, then our analysis is very simple. The baryon-entropy ratio produced at the time of X boson decay will be just equal to the net baryon excess ΔB produced per X or $\bar{X}$ decay, which as we have seen for color triplet bosons is the branching ratio difference, $r - \bar{r}$, times the ratio of the X boson number density and the entropy density, which is of order unity. A more careful estimate gives a baryon to entropy ratio of order

$$k n_B / s \approx 0.01 \ \Delta B \ .$$

It is important that, even though we cannot calculate the baryon-entropy ratio with precision, this analysis shows that it is a pure number which is determined by purely microscopic considerations, and therefore must be the same everywhere in the universe, provided that $m_X \gg kT$ at the time of X and $\bar{X}$ boson decay everywhere in the universe. Under this assumption, even if the early universe is highly chaotic, the perturbations will have a uniform value for the baryon-entropy ratio, and hence will be of the type called adiabatic.

If m_X is somewhat less than $g_X^2 \, m_{PL}$, then the universe is not far out of equilibrium at the time of X boson decay, and much of the baryon number is lost in subsequent thermalization processes like inverse decay. A great deal of work has been done on this case in the last two years, especially by Kolb and Wolfram and by Fry, Olive and Turner. They carry out computer solutions for the complicated Boltzmann equations governing various reactions that would have occurred as the temperature dropped from where X bosons began to decay down to values well below m_X. These calculations show that the baryon production is indeed suppressed, but

only by a power of the small number $m_X/g_X^2\, m_{PL}$; it is found that the baryon-entropy ratio is of order

$$kn_B/s \simeq 0.01\ \Delta B\ (m_X/g_X^2\, m_{PL})^{1.3}\ .$$

More important than the precise numerical value is the fact that in an anisotropic universe, where the local expansion rate can differ from $(kT)^2/m_{PL}$, the baryon-entropy ratio can vary from point to point. For instance, if in some region anisotropies make the expansion rate much larger than $(kT)^2/m_{PL}$, then the effective Planck mass in the formula for kn_B/s is smaller than the nominal value, and the baryon-entropy ratio is increased, possibly up to the maximum value of 0.01 ΔB. In this way one gets non-adiabatic perturbations, which have been studied in detail by Bond, Kolb and Silk and by Barrow and Turner. As you know, such perturbations survive much better than the adiabatic perturbations down to the time of recombination.

What value do we expect for the crucial quantity ΔB? The most important point here is that ΔB is always small, because it vanishes in the lowest-order, or "Born", approximation. This can easily be seen by an amusing little argument. There is a famous symmetry in physics, TCP, the invariance under a combined time-reversal invariance, charge-conjugation, and parity inversion. This by itself does not require that $r = \bar{r}$ or $\Delta B = 0$; for this one would need C or CP invariance. But as everyone learns in statistical mechanics courses, in Born approximation all cross sections and transition rates satisfy a principle of "detailed balancing"; they effectively respect time reversal invariance. But TCP and T imply CP, which implies $\Delta B = 0$, so lowest order graphs never yield a net baryon production.

Nanopoulos and I have tried to calculate ΔB as accurately as we could. (Similar calculations have been done by Ellis, Gaillard and Nanopoulos and by Barr, Segre and Weldon). We found that in sufficiently complicated theories ΔB can get contributions from the interference of zero and one-loop graphs, but even so the possible range of values is enormous, from 10^{-6} to about 10^{-11}, corresponding to a maximum baryon-to-entropy rate of about 10^{-8} to 10^{-13}. This overlaps the observed range of possible values for this ratio, but the agreement here hardly constitutes a triumph of theoretical physics.

I should mention an interesting remark made in this connection by Ellis, Gaillard, Nanopoulos and Rudaz. They point out that the sort of CP violation which must occur when baryon number is produced cosmo-

logically would also show up at low energies in one other sort of CP violation, the electric dipole moment of the neutron. "Barring cancellations", they find that a given value for the baryon-photon ratio n_B/n_γ (essentially the same thing as the baryon-entropy ratio kn_B/s) would require a neutron dipole moment

$$d_n > 3 \times 10^{-16} \frac{n_B}{n_\gamma} \text{ e cm} .$$

Right now the upper bound on the neutron dipole moment is a few times 10^{-25} e cm, implying an upper bound on the baryon-photon ratio of about 10^{-9}. This is not yet a real conflict. However, if the experimental limit on the neutron electric dipole moment continues to decrease, we may have a real problem here.

I have emphasized baryon nonconservation here (even though lepton as well as baryon number would be violated in processes like $X \to qq$ and $\bar{q}\ell$ or $p \to e^+ \pi^0$) because we have evidence that the universe is not baryonically neutral, while no one has any idea whether there is a tiny imbalance in the population of black-body neutrinos and antineutrinos which would give the universe a non-vanishing lepton number comparable to its baryon number. Nevertheless lepton nonconservation is interesting physically, because lepton number is, along with baryon number, one of the two supposedly conserved quantities which may in fact not be conserved, and it may be interesting cosmologically as well.

Just as there are no baryon number nonconserving interactions among ordinary particles that are renormalizable (i.e., with dimensionless coupling constants) there are also no renormalizable interactions that violate lepton number. Thus lepton number nonconservation at ordinary energies will be suppressed by one or more powers of superheavy masses. However, there is an effective interaction which violates lepton (though not baryon) number and has a dimensionality of only 5, so that its coupling is suppressed by only one power of a superheavy mass M. It is of the form

$$g (\nu \phi^0 - e^- \phi^+)^2$$

with g of order α/M. The appearance of a vacuum expectation value $< \phi^0 > \simeq 300$ GeV will then produce a Majorana mass term $m_\nu \nu\nu$ for the neutrino, with

$$m_\nu = g < \phi^0 >^2 \approx \alpha \, (300 \text{ GeV})^2/M .$$

If M is as large as 10^{15} GeV, then m_ν will be about 10^{-3} eV. This is large enough to make neutrino oscillations important in solar neutrino experiments, but much too small to have any importance in other astrophysical or cosmological phenomena. Witten has shown that in some SO(10) theories the relevant M might be as small as 10^{11} GeV, giving a neutrino mass as large as 10 eV, which would be of great interest in understanding the dynamics of galaxies. On the other hand it is possible that baryon number minus lepton number is strictly conserved, as it is in the simplest SU(5) model, in which case the neutrino mass would be strictly zero. Unfortunately terrestrial experiments have not yet definitely settled the question of the neutrino mass.

3 - Grand Unification

The third of my major topics is grand unification. The idea here is that the strong and electroweak interactions and the quarks and leptons form part of larger families of interactions and particles which are connected by some big group G of gauge symmetries. If the spontaneous breakdown of this big group is characterized by a larger scale of vacuum expectation values than the 300 GeV associated with the breakdown of the electroweak gauge symmetry, then the other vector bosons besides the $W^{\pm}$, Z^0, γ and gluons will be very heavy and the forces they produce might have escaped detection. Much work has been done to develop such theories, by Pati and Salam; Georgi and Glashow; Georgi; Fritsch and Minkowski; and many others. I will try here to emphasize those aspects of grand unified models that do not depend on the details of specific models.

There is an obvious problem with any sort of unification of strong with electroweak interactions: the strong interactions are observed to be much stronger then the electroweak interactions. In 1974 Georgi, Quinn and I pointed out that the coupling strengths would only be expected to be equal if measured at an energy scale M comparable to (or greater than) the scale of the vacuum expectation values which break the big group G. The strong interaction coupling constant is believed to decrease slowly with energy, and the electroweak couplings change even more slowly, so it is possible for the couplings to come together at some scale M, but only if M is very large indeed. We found that the two electroweak and the one strong coupling constant would all become equal at some energy, but only

if the parameter $\sin^2\theta$ which describes Z^0–γ mixing were adjusted to have the value

$$\sin^2\theta = \frac{1}{6} + \frac{5\ e^2}{9\ g_s^2(m_W)} \simeq 0.2\ ,$$

where $g_s(E)$ is the strong coupling constant as measured at energy E. Experiment did not agree with this in 1974, but it does now, to within a few percent. With $\sin^2\theta$ given this value, the mass M at which the couplings come together is

$$M \simeq \Lambda \exp\left[\frac{4\pi^2}{11}\left(\frac{1}{e^2} - \frac{8}{3\ g_s^2(m_W)}\right)\right] \simeq 10^{16}\ \text{GeV}\ .$$

This result, and the above formula for $\sin^2\theta$, apply to a wide variety of proposed grand unified models. There are many corrections that have been developed, by Buras, Ellis, Gaillard and Nanopoulos; Goldman and Ross; Marciano and Sirlin; Hall; and others. The most important numerically is that the e here should be replaced with the electric charge $e(m_W)$, measured at frequencies comparable with the W mass, with $e^2(m_W)/4\pi$ about 1/128 instead of 1/137, reducing M to about 5×10^{14} GeV. Grand unified theories are generally found to contain vector or scalar bosons whose exchange can produce baryon nonconservation, and with their mass of this order, the proton lifetime should be of order $M^4/m_p^5\,\alpha^2 \approx 10^{30\text{–}32}$ years.

Of course, we are not really certain that there is no new physics from the few hundred GeV accessible to accelerators all the way up to 10^{14} GeV. It may be that there is a sequence of symmetry breaking scales filling up this whole range. One can only record a number of constraints that must be satisfied by the physics of these intermediate scales. First, particles with masses $\ll 10^{14}$ GeV must not have interactions that could produce "normal" proton decay processes (e.g. $p \to e^+\pi^0$, $n \to e^+\pi^-$) because then these processes would be so fast that they would have been already discovered. Also, whatever baryon violation is produced by particles of these intermediate masses must not allow the baryon number to equilibriate back to zero. Finally, the change in the dependence of coupling constants with energy at intermediate scales must not mess up the agreement between theory and experiment for $\sin^2\theta$.

If there really is a spontaneous symmetry breakdown produced by

vacuum expectation values of order 10^{15} GeV, then it is to be expected that the universe experienced a phase transition at comparable temperatures: the big group G was unbroken (and all its vector bosons massless) at temperatures above 10^{15} GeV, but then the symmetry became spontaneously broken down to the symmetry group $SU(3)\times SU(2)\times U(1)$ characterizing the observed strong and electroweak interactions. But was this a first or a second-order phase transition?

One of the most interesting applications of elementary particle theory to cosmology of the last year or so has been the exploration of the consequences that follow if the spontaneous breakdown of the big group G to $SU(3) \times SU(2) \times U(1)$ was a first order phase transition. That is, at a temperature of order 10^{15} GeV the G-noninvariant vacuum state became of lower energy than the G-invariant vacuum, but the G-invariant vacuum remained metastable at least temporarily. In such a case the universe may be trapped for a while in the G-invariant or "false" vacuum, and only make the transition to the true vacuum at a later time and lower temperature through barrier penetrations or thermal fluctuations, just as water vapor may not turn into liquid water until the temperature drops well below 100 °C. This sort of delayed phase transition has been studied recently by Guth; Linde; Cook and Mahanthappa; Tamvakis and Vayonakis; Sato; Steinhardt; and Lapchinsky, Rubakov, and Veryaskin. Guth has emphasized that such a scenario might help to solve three outstanding problems of modern cosmology: the entropy (or flatness) problem, the horizon problem, and the problem of monopoles. Let us look in turn at each of these problems, and see how a delayed first-order phase transition might help to solve them.

Entropy. The total entropy S within the radius of curvature $R(t_0)$ (for $k = \pm 1$) of the universe is at least as large as that contained within the present horizon radius $\approx H_0^{-1}$, so

$$S > (kT_\gamma)^3 H_0^{-3} \simeq 10^{87} .$$

Why is this so big? This problem is related to the flatness problem: the universe is some 10^{61} Planck times old, and all this time the deceleration parameter, q, has been increasing or decreasing away from $q = 1/2$, so why is q still within a factor of 10 of $1/2$? That is, why is the universe still so flat? To see that these are the same problems, note that the equations of the radiation-dominated Friedmann model tell us that the universe will start with $q = 1/2$, and will remain essentially flat for a period of

$S^{2/3}$ Planck times, so its present flatness is understandable if $S \geq 10^{92}$. At least in my opinion, the mystery is simply why S is so large. Of course, one possible answer is that S is infinite, because we are in a flat ($k = 0$) Robertson-Walker universe, with infinite radius of curvature. Then q remains 1/2 forever. Guth argues that this assumption of perfect flatness makes no sense in a universe which is manifestly not perfectly homogeneous or isotropic. I am not sure. At any rate, he and I agree that after 10^{61} Planck times it is implausible that we should just now be emerging from a period of essential flatness, so it seems likely that the universe will remain nearly flat for a while longer, in which case q_0 must be now close to 1/2.

If the universe underwent a transition from the false to the true vacuum at a temperature $kT_1 \ll M$, and if the energy density difference ($\approx M^4$) of the two vacua then subsequently thermalizes, the entropy density will become of order M^3 instead of $(kT_1)^3$, so the entropy of the universe will increase by a factor of order $(M/kT_1)^3$. In order to get an entropy $S \approx 10^{87}$ from an initial value $S \approx 1$, it is necessary for the universe to remain in the metastable false vacuum long enough for the temperature to drop from M to $10^{-29}\,M \simeq 10^{-5}$ eV, i.e., long enough for the Robertson-Walker scale factor to increase by a factor 10^{29}. The production of baryon number would have to take place after the universe was reheated by the transition to the true vacuum, back up to temperatures nearly of order M.

Horizon. It is well known that points in opposite directions in the sky have nearly equal black-body background temperatures. This suggests that they must have been in causal contact with some common influence. However, since we are midway between the sources of microwave background in opposite directions, their distance from us in co-moving coordinates must be less than their coordinate distance from this common influence. The coordinate distance travelled by light from time $t = 0$ to the time t_R when the radiation was emitted must therefore be greater than the coordinate distance travelled from time t_R to the present time t_0. In terms of the Robertson-Walker scale factor $R(t)$, this requires that

$$\int_0^{t_R} \frac{dt}{R(t)} > \int_{t_R}^{t_0} \frac{dt}{R(t)} .$$

If $R(t)$ behaved as in a matter-dominated Friedmann universe, with $R(t) \propto t^{2/3}$, then this condition could be satisfied only if $t_R > t_0/8$, or in

other words if the radiation redshift z_R is less than 3. However we have every reason to believe that $z_R \approx 1000$, in which case the $R(t)$ of the Friedmann model could not possibly satisfy the above inequality.

On the other hand, if the universe went through a long period of supercooling in the false vacuum, then during this period its energy density was of order M^4 (supposing it to be zero now), and during this time

$$R(t) = R(0)\, e^{\alpha t}$$

$$\alpha = \sqrt{\frac{8\pi\rho G}{3}} \approx \sqrt{GM^4} .$$

After the transition to the true vacuum and the subsequent thermalization of the released energy, the energy density became proportional to $T^4 \propto R^{-4}$, so $R(t) \propto (t + \text{const})^{1/2}$. Imposing a condition of continuity on $R(t)$ and $\dot{R}(t)$ at the time t_1 of this transition, we have for $t > t_1$

$$R(t) = R(0)\; e^{\alpha t_1}\; (1 + 2\alpha\,(t - t_1))^{1/2} .$$

(For simplicity we ignore here the subsequent crossover to a matter dominated universe with $R(t) \propto t^{2/3}$). Then for $t_0 \gg t_1 \gg 1/\alpha$, the above inequality becomes

$$\frac{1}{\alpha} \gtrsim e^{-\alpha t_1}\; \sqrt{2\, t_0/\alpha}$$

or inother words (recalling that $S_{min}^{2/3} \approx t_0\, m_{PL}$)

$$M/T_1 \gtrsim \left(\frac{M}{m_{PL}}\right) S_{min}^{1/3} .$$

For $M/m_{PL} \simeq 10^{-4}$ and $S_{min} \simeq 10^{87}$, this gives a supercooling expansion factor $M/T_1 \gtrsim 10^{25}$, a condition only slightly less stringent than that imposed by the flatness problem.

Monopoles. It was discovered mathematically by 't Hooft and Polyakov some years ago that gauge theories like the electroweak theory are likely to involve heavy magnetic monopoles. These are not present in theories based on semi-simple groups, as for most grand unified theories, but they appear when the vacuum makes its transition from a state which

is invariant under the big semi-simple group G to one which only respects the strong and electroweak groups SU (3) and SU (2) × U (1). (As explained by Kibble, monopoles appear as the knots left over when the domains in which G is broken in different directions coalesce). Their mass will be of the order of M/α, where M as before is the scale of the vacuum expectation values which break G, and α is some appropriate fine structure constant of order 10^{-2}. This suggests a mass of order 10^{17} GeV, so in order that the mass density of these particles should not produce too large a cosmic deceleration, their present number density should be less than about 10^{-25} of the present entropy density. However, as pointed out by Preskill and Zeldovich and Khlopov, this condition is not easy to satisfy in grand unified theories. They point out that the monopole and antimonopole densities are likely to be one per horizon volume at the time of the phase transition from the G-invariant to G-noninvariant vacuum, and only a small fraction of monopoles and antimonopoles will annihilate. (However, Dicus and Teplitz have raised the possibility that the monopole-entropy ratio might be reduced to below 10^{-25} by the increase in entropy caused by monopole-antimonopole annihilation). The horizon volume at a temperature T_1 is of order

$$V_{\text{horizon}} \approx H^{-3} \approx [G(kT_1)^4]^{-3/2} \approx m_{PL}{}^3/(kT_1)^6 ,$$

while the entropy density produced in the phase transition is of order M^3, so the monopole-entropy ratio produced at temperature T_1 is of order

$$\text{monopoles/entropy} \approx (kT_1)^6/m_{PL}{}^3M^3 .$$

If the phase transition is not delayed, so that $kT_1 \approx M \approx 10^{15}$ GeV, this gives a monopole-entropy ratio of about 10^{-12}, too large by about 13 orders of magnitude. In order to reduce the monopole-entropy ratio to something like 10^{-25}, one would want the phase transition to be delayed until the temperature drops by a factor of about $10^{-13/6}$, to about 10^{13} GeV. The production of monopoles in these phase transitions has been studied in detail by Guth and Tye; Einhorn, Stein, and Toussaint; Einhorn and Sato; Lazarides and Shafi; Langacker and Pi; and Fry and Schramm. Unfortunately one cannot be confident regarding the number that would actually be produced or the fraction that would then annihilate.

As we have seen, the entropy, horizon, and monopole problems suggest a period of supercooling in which the universe remains in the "false" vacuum while it expands by a factor of at least 10^{29}, 10^{24}, or 10^{2} respectively.

What sort of expansion factor would we actually expect for the period of supercooling? The answer depends on what we think brings the period of supercooling to an end.

One possibility is that the supercooling ends by a quantum-mechanical "tunneling" of the vacuum, from the higher energy metastable G-invariant false vacuum to the lower-energy stable SU (3) × SU (2) × U (1)—invariant true vacuum. This has been studied by Guth and E. Weinberg, using methods of Coleman and Callan and Coleman. They find that the temperature T_1 at which the tunneling occurs is sensitive to the detailed parameters of tbe theory, and could be almost anything. A severe problem with this approach is that the tunneling process is random: bubbles of true vacuum appear here and there in an uncorrelated fashion, but the universe is expanding so fast that they can never coalesce and thermalize the energy in their walls.

Another possibility is that the transition from the false to the true vacuum occurs not by a tunneling from a metastable to a stable state but by the disappearance of the barrier between them. This loss of metastability would occur everywhere in the universe at the same temperature and hence at the same time, so there would be no problem of bubble coalescence. The problem in understanding how this is possible is that we want the supercooling to last while the temperature drops by many orders of magnitude, and if metastability lasts this long, why at some relatively tiny temperature should it ever end?

One way that metastability might be lost is through the slow logarithmic increase of gauge coupling constants, finally producing strong forces. (This ends the metastability only if the zero-temperature theory is of the type originally studied by Coleman and E. Weinberg). Now, the bigger the group, the faster the couplings increase with decreasing energy, so one might expect the coupling constant of the vector bosons associated with the big group G to increase in the false vacuum more rapidly than the ordinary strong coupling constant of QCD increases in the true vacuum. Sher estimates that the G-forces would become strong in the false vacuum when the temperature drops to about 10^8 GeV. These strong forces may produce a dynamical breakdown of the big group G, but then again, they may not.

A related possibility that does not seem to have been studied in the literature is that it is not the increase in strength of the coupling of the hig group G that ends the metastability of the false vacuum, but rather it is the increase in strength of the coupling associated with some sub-

group of G, that eliminates the potential barrier along directions in the space of scalar fields for which this subgroup is unbroken, and thus opens up a channel along which the false vacuum can decay. (This is similar to Witten's picture of $SU(2) \times U(1)$ breaking, discussed earlier). The attractive thing about this possibility is that the coupling of a small SU (2) subgroup gets strong with decreasing energy much more slowly than the coupling of G or QCD, so that there is a chance that the metastability might survive down to really low temperatures.

Another possibility, studied by Press, is that although the universe stays in the false vacuum during many e-foldings of expansion, the temperature does not drop very much during this period, being kept high by Hawking radiation from the horizon. It is then quite plausible that a moderate change in temperature would eventually change the dependence of free energy on vacuum expectation values in such a way as to end the metastability of the false vacuum.

It is also possible that gravitation has something to do with the loss of metastability. In any reasonable theory of gravity, we would expect a scalar field ϕ to interact with the Ricci scalar R through a term $b\,\phi^2\,R$, with b of order unity. (Such a term is needed in any case as a counterterm to ultraviolet divergences). Abbott has pointed out that this will become more important than the term $e^2\,(kT)^2\,\phi^2$ produced by thermal fluctuations when $e^2\,(kT)^2$ falls below $R \approx \alpha^2 \approx GM^4$, i.e., when kT drops below $M^2/em_{Pl.} \approx 10^{11}$ GeV. He finds that metastability is lost at this temperature if b is positive, while if b is negative it lasts forever. Similar conclusions are reached by Fujii and by Hut and Klinkhamer.

In summary, it is not difficult to see how the transition out of the false vacuum could be delayed until the universe cools by a very large factor, large enough apparently to solve the monopole problem; but it is difficult to see how this transition could be delayed until the temperature drops to the level of 10^{-5} eV or so that is required to solve the entropy and horizon problems.

4 - Supersymmetry

My last major topic is supersymmetry. This is a symmetry, first proposed in its modern form in 1974 by Wess and Zumino, which puts particles of different spin into the same symmetry multiplets. In this respect it is unique, which may be why it has been so actively studied despite the lack of any experimental support.

Of course, we do not see supermultiplets of particles in nature with equal mass and different spin, so supersymmetry, if a valid symmetry at all, must be spontaneously broken. Its implications depend critically on the scale of the vacuum expectation values which break it. If these vacuum expectation values are of order 10^{15} GeV, then supersymmetry does not offer much in the way of direct implications for physics at ordinary energy. On the other hand if the vacuum expectation values which break supersymmetry are of the same order as those that break the electroweak gauge symmetry, say of order 300 GeV, then the particles of different spin in supermultiplets are split in mass by amounts of order m_W, or in some cases much less. This is the case I will consider here.

If the gauge bosons such as the $W^{\pm}$, Z^0, γ and gluon have superpartners of spin 1/2, with masses that are not different by more than 100 GeV or so, then these superpartners must be taken into account in calculating the dependence of coupling constants on energy. Generally speaking, the main effect is simply to reduce the rate at which all squared coupling constants vary with the logarithm of the energy by a factor 9/11. As pointed out by Dimopoulos, Raby, and Wilczek, this means that the successful prediction for $\sin^2\theta$ is unaffected, but the mass M where the strong and electroweak couplings come together is increased, to about 10^{17} GeV. "Ordinary" proton decays like $p \to \pi^0 e^+$ would still occur, but at a rate decreased by a factor of 10^8, and hence would be quite unobservable. On the other hand the increase in M is actually helpful from the point of view of cosmological baryon production, because everything happens earlier, when the universe is further from thermal equilibrium.

The quarks also have superpartners, scalar bosons which like the quarks are color triplets. I mentioned earlier that color triplet bosons are just what are needed to produce baryon nonconserving processes like proton decay, and the slow rate of proton decay requires that the color triplet bosons which mediate the decay must be heavier than about 10^{14} GeV. How then can we tolerate superpartners of the quarks whose mass is no more than 100 GeV or so? One possible answer that I have been studying lately is that there is an additional gauge symmetry at ordinary energies, in addition to the SU (3) and SU (2) $\times$ U (1) of the strong and electroweak interactions, which together with supersymmetry rules out any renormalizable baryon nonconserving interactions of the scalar superpartners of the quarks. (Such an addition to the low energy gauge group had already appeared in various models of Fayet, which aim at explaining how the scalar superpartners of the quarks and leptons can be so much heavier than the quarks and leptons

themselves). Such a symmetry would also have the effect of ruling out a neutrino mass at any astrophysically relevant level.

However, if supersymmetry takes away the neutrino mass, it replaces it with other odds and ends that may play an important role in astrophysics. One of these is the goldstino, a massless particle of spin 1/2 which has the same relation to broken supersymmetry that the pion and the spin waves have to the broken chiral and rotation symmetries in strong interactions and ferromagnets. Goldstinos have coupling constants comparable to e, but they connect ordinary particles like quarks, leptons, etc. to their heavy superpartners, so at low energies goldstino interactions are suppressed by large masses in internal lines, and hence are similar to neutrino interactions. They would count along with the neutrinos in determining the expansion rate of the universe during the period just prior to helium synthesis, and hence tend to increase the cosmic helium abundance.

There is a fascinating complication regarding goldstinos: when gravity is taken into account the goldstino disappears from the theory, being replaced by the helicity $\pm$ 1/2 components of the spin 3/2 superpartners of the graviton, the gravitino. Through this "super-Higgs" mechanism the gravitino acquires a mass, calculated by Ferrara and Zumino to be

$$m_{\text{gravitino}} = \sqrt{\frac{4\pi}{3}} \; \frac{F}{m_{PL}}$$

where F is the constant with dimensions of (mass)^2 that characterizes the strength of the supersymmetry breaking. If F = (300 GeV)^2, the gravitino mass is 1.5×10^{-5} eV. For a gravitino that is this light, the helicity $\pm$ 1/2 states behave for all practical purposes just like the massless goldstino which would be present in the absence of gravitation. On the other hand, if F is of order 10^{17} GeV, then the gravitino mass is of order 10^{15} GeV. Such a particle would of course play no role at the time of nucleosynthesis, much less at present, but its relatively slow decay might generate large amounts of entropy at very early times. (The astrophysical implications of a 1 keV gravitino have been considered by Pagels and Primack).

Another interesting oddity is the photino, the spin 1/2 superpartner of the photon. Its mass vanishes in the lowest order of perturbation theory and is, therefore, suppressed relative to m_W by at least a factor $e^2/8\pi^2$. Farrar and I have tried to estimate the photino mass, but we find that it depends critically on the numbers of supermultiplets of different kinds and

on the pattern of symmetry breaking; depending on what one assumes, the photino mass could be a few MeV, or a fraction of a keV, or 10^{-27} eV, or zero. Photinos of mass less than an MeV would contribute like neutrinos and goldstinos to the cosmic energy density just before nucleosynthesis, and hence would further increase the helium abundance. It is possible that photinos have a mass of the order 10 eV to 100 eV which would make them an important contribution to the galactic and cosmic mass densities, but this does not seem particularly likely. Photinos decay into goldstinos plus photons at a rate calculated by Cabibbo, Farrar, and Maiani to be of order

$$\Gamma_{\text{photino}} = 8\pi\, m^5_{\text{photino}}/F^2 \ .$$

For $F = (300\ \text{GeV})^2$, this gives a photino lifetime longer than 10^{10} years for photino masses below about 800 eV, so these photinos would still be abundant in the present universe. Even though their decay is slow, it is much faster than neutrinos of comparable mass, so UV photons from galactic photinos might be detectable. Heavier photinos would be very rare, but they still might he of astrophysical interest. For instance, photinos with masses of a few MeV could be produced in supernova cores and then deposit their decay photons in the outer layers of the supernova or in the space outside.

As this talk has progressed, we have been moving steadily into areas of elementary particle physics about which we feel less and less certainty. This sense of uncertainty will be heightened if I remind you that in all this work on supercooling, gravitino masses, etc., one is forced on observational grounds to assume that the present energy density of the vacuum (minus whatever bare cosmological constant may be present in the Lagrangian) is zero, or at any rate less than about 10^{-44} GeV per cubic fermi, an assumption for which at present there is not the slightest rationale in terms of fundamental physical principles. But even if particle physicists and cosmologists are not yet able to offer definite answers to each others' problems, there does seem to be an increasing area of common concern in the physics of the very early universe, about which it is worthwhile for us to keep in touch.

I am very grateful to Alan Guth and Edward Witten for their help in correcting errors in an earlier version of this paper.

DISCUSSION

WEISSKOPF

It is very hard to start out commenting in any general way after this splendid talk. The first impression I received during this exciting report is that particle physics seems to have become even more uncertain and hypothetical than cosmology. This is by no means a reproach; on the contrary it is wonderful how much one can expand ideas that have been conceived in connection with experiment. Let me point out how far we actually are from experiments. In view of what we have just heard, the theory of electro-weak interaction is old hat and already should be considered almost like Maxwell's equations. By the way, Steve Weinberg is one of its founders. This theory, of course, is based upon the existence of the intermediate bosons, just like Maxwell's equations are based on the existence of light waves. You know, it took about 20 years after Maxwell, if I am not mistaken, until Hertz actually produced such waves. If he had not succeeded, Maxwell's equations would have been forgotten. Now the bosons have not yet been seen. I think they will be seen soon, if they *can* be seen, because the CERN proton-anti-proton experiment is prepared for that purpose and goes extremely well. A telegram to Steve from Geneva may already be written, but it may not come. Now I mention this in order to emphasize that, although the electroweak theory contains tremendous generalizations of quantum electrodynamics, there are a good number of experimental observations that support it.

But what about the Grand Unification Theory (GUT)? Steve said himself that at present we have just one experimental support, a quite impressive one, namely, the fact that GUT, in almost all its versions, can predict what he modestly bas refused to call the Weinberg angle. Important as it is, it is only one number. I am very worried that, in order to accept this very elegant theory, one has to assume that up to 10^{14} GeV there will be no new particle appearing. So far we always have found new substructures whenever we went 1, 2, 3 or 4 orders of magnitude up in energies. One cannot but ask oneself whether the existence of five quarks (probably they are 6) and of 3 electrons is not already a sign of substructure. Are we supposed to assume that this is not so and tbat we already know the groups that include all truly elementary particles, the ones that are known at

present, with a few small exceptions? That idea leaves one with an extreme uncertain feeling. But we are accustomed to such feelings especially in the framework of cosmology. So far cosmology neither supports or disproves GUT. So far that theory does not give any definite answers to those questions that are worrying us, such as the proton surplus, the famous 10^{-9}, the horizon problem, the entropy problem. None of these questions are really answered by GUT, although it is not yet excluded that GUT contains some of the answers, in particular to the proton surplus problem. As to the others there are, I would say, "romantic" possibilities how one *could* solve them, by introducing phase transition. Maybe particle physics develops too many new unproven ideas per unit time. Well, these are my personal impressions that are probably influenced by having been to long in this field.

SILK

An attractive feature of the supersymmetric theory may be the large value of the grand unification mass. This is sufficiently close to the Planck mass that decaying inhomogeneities, shear modes, and other possible remnants of the quantum gravity era could still have sufficient amplitude ($\gtrsim 10^{-4}$) to perturb the expansion rate during baryosynthesis and generate baryon number fluctuations.

HAWKING

If you have a Higgs mass of 10 Mev in the electroweak theory, you say that the electroweak phase-transition would be strongly first order. In this case, work we have done at Cambridge indicates that one would produce an unacceptably large number of black holes of about the mass of the earth.

MASSIVE NEUTRINOS IN COSMOLOGY AND GALACTIC ASTRONOMY

D.W. SCIAMA
Department of Astrophysics, Oxford University
and
Department of Physics, University of Texas at Austin

ABSTRACT

This article gives a brief qualitative discussion of the consequences for cosmology and galactic astronomy of a neutrino type possessing a rest-mass of order tens of electron volts. Emphasis is laid on the possible detectability of ultra-violet photons which may be emitted by massive neutrinos dominating both the universe and individual galaxies including the Milky Way. One speculative possibility is that ionisation observed both in the intergalactic medium and in the halo of our Galaxy may be produced by such photons. This would require a neutrino mass $\sim$ 100 ev and a radiative lifetime $\sim 10^{27}$ seconds, values which are compatible with both cosmological constraints and considerations from elementary particle physics.

1 - INTRODUCTION

This Study Week is taking place at a tantalising moment. There are suggestions from both theoretical and experimental particle physics that neutrinos may have a non-zero rest mass of order up to tens of electron volts. If this turns out to be so and if the hot Big Bang picture of the origin of the universe is at least approximately correct, then the implications for cosmology are so far-reaching that it is reasonable to talk in terms of a breakthrough as significant as the discovery of the 3 K cosmic background radiation.

This follows in part from the remark of Cowsik and McClelland (1972) that finite mass neutrinos left over from the hot Big Bang might today be nearly as numerous as 3 K photons. This would be the case, for example, for left-handed Majorana neutrinos with a mass less than 1 MeV. The contribution of such neutrinos to the present value of the density parameter Ω (ρ/ρ_{crit}) is given to a good approximation by

$$\Omega = \sum_i m_i / 100\, h_0^2 ,$$

where the sum is over the different neutrino types, m_i is in electron volts, and h_0 is related to the Hubble constant H_0 by $H_0 = 100\, h_0$ km sec^{-1} Mpc^{-1}.

We see immediately from this formula that with $0.5 \leqslant h_0 \leqslant 1$ and $m_i \sim$ tens of electron volts, Ω can easily reach unity. This could mean that neutrinos contribute more, perhaps far more, to the density of the universe than does anything else, including baryons, and that they could even close the universe and bring about its eventual re-collapse (without deuterium being destroyed primordially). This was Cowsik and McClelland's (1972) main original point.

There is the additional possibility (Cowsik and McClelland 1973) that these neutrinos could provide the missing matter in clusters of galaxies and in individual galaxies including the Milky Way. In this manner one could account for the large velocity dispersions observed in clusters and for the extended flat rotation curves observed in galaxies. Again in these cases the total mass in the form of neutrinos would then exceed that in baryons by a factor of order ten or more. This could turn out to be a particularly fruitful hypothesis, since the structures of clusters, and especially of individual galaxies, can be studied in detail, and these details could then be related to the neutrino distribution.

A further fruitful interaction between astronomy and particle physics would arise if the neutrinos dominating both the universe and localised condensations were to decay into lighter neutrinos, emitting photons in the process. It has been pointed out by de Rujula and Glashow (1980) that, with neutrino masses lying in the tens of electron volt range, the photons would probably lie in the ultra-violet and so might be detected by the methods of ultra-violet astronomy.

Two detection methods have already been discussed in the literature. The first involves searching for the integrated ultra-violet background arising from neutrinos decaying both cosmologically and in our Galaxy and

for an ultraviolet signal from the Virgo and Coma clusters. In these cases the absence of any observable effect attributable to neutrinos leads to a lower limit on the radiative lifetime of the neutrino (Stecker 1980; Kimble, Bowyer and Jakobsen 1981; Henry and Feldman 1981; Shipman and Cowsik 1981).

The second method involves searching for ionisation effects produced by the decay photons on the assumption that these photons possess sufficient energy. The presence of high velocity clouds of atomic hydrogen in regions remote from the plane of the Galaxy again provides negative evidence which leads to a further lower limit on the neutrino lifetime (Melott and Sciama 1981).

Positive evidence of ionisation effects which could be attributed to photons emitted by neutrinos would, of course, be more useful. Two tentative possibilities have been suggested so far. In one (Sciama 1982a) the ionisation of the intergalactic medium, which is usually invoked to account for the absence of absorption troughs in quasar spectra (Gunn-Peterson effect), is attributed to photons emitted by a cosmological distribution of decaying neutrinos rather than by the quasars themselves. Such a hypothesis is very radical but it does have the merit of leading to fairly definite values for the mass and radiative lifetime of the neutrino type involved. In round terms one requires a mass in the range 50 to 100 ev and a lifetime of $\sim 10^{27}$ seconds.

The other possibility (Sciama and Melott 1982) involves the highly ionised carbon (CIV) and silicon (Si IV) found high above the galactic plane by observations with the International Ultra-Violet Explorer (IUE). The existence of these highly ionised atomic species may well have a conventional explanation in terms either of collisional effects in a hot gas ($T \sim 10^5$ K) or of ionisation by suitably hard photons emitted by hot stars or the sources of the soft X-ray background. However, these explanations are not certain and an alternative, speculative, possibility is that the photons involved arise from decaying neutrinos which dominate the galactic halo. This again leads to a determination of the mass and the lifetime of these decaying neutrinos, with results remarkably similar to those derived from our intergalactic hypothesis.

The derivation of these numerical results and the extent to which they are reasonable are discussed in the final section of this paper. First, however, we describe briefly the theoretical and experimental evidence from particle physics that neutrinos may have a non-zero mass. This is followed by an equally brief account of the consequences of such a non-

zero mass for the hot Big Bang and for the present density of the universe. We then analyse the possibility that these neutrinos may dominate galaxy clusters and individual galaxies. Finally we turn to the ultra-violet implications if the neutrinos emit photons at a significant rate.

2 - Massive Neutrinos in Particle Physics

It has been known since 1934 that "the rest mass of the neutrino is either zero or at any rate very small with respect to the mass of the electron". Fermi (1934) deduced this from the shape of the electron spectrum in beta decay near its high energy end point. Until recently the most precise experiment of this type (Bergkvist 1972) led to the limit $m_\nu < 60$ ev at the 90% confidence level.

It was in the same year that Cowsik and McClelland (1972) pointed out the fundamental importance for cosmology of a non-zero rest-mass for the neutrino if the hot Big Bang picture of the origin of the universe is at least approximately correct. However, at that time such a non-zero mass was not popular amongst particle physicists and the consequences of Cowsik and McClelland's ideas were not widely investigated. Since then the climate of opinion has changed, thanks mainly, on the theoretical side, to the development of gauge theories, and especially of grand unified theories (GUTS). With one exception, the minimal SU (5) scheme, these theories do lead to a mass term in the neutrino part of the Lagrangian. Weinberg (1979) has shown that, under very general conditions,

$$m_\nu \sim \frac{f < \phi_0 >^2}{M} ,$$

where f is a dimensionless coupling constant, $< \phi_0 >$ is the vacuum expectation value of the scalar field responsible for intermediate vector boson masses in the theory, and M is the grand unification mass scale, which is related both to the proton lifetime and the energy at which all coupling constants in the theory become equal.

In these theories one expects that $< \phi_0 > \sim 300$ Gev and $M \sim 10^{15}$ GeV. The value of f depends on the model and can vary roughly between α^2 and α^{-2}, where α is the fine structure constant. In this way one finds masses for the neutrino ranging from 10^{-5} to 10^2 ev. For a more detailed

account and fuller references to the literature we refer the reader to the surveys of Langacker (1981) and de Rujula (1981).

A related property of importance for us is the number of neutrino types or flavours. Electron and muon neutrinos have been directly observed and are known to be different particles. The tau neutrino has not yet been observed but is presumed to exist. Experiments to search for it are described by Morrison (1980). Further flavours could also exist and their number is not specified by existing GUTS. A limit on this number has been derived from arguments involving light element synthesis in the Big Bang, since their existence would speed up the early expansion rate of the universe. However, this limit is compatible with the existence of several neutrino types beyond those already established.

In addition to the mass of each neutrino flavour playing a role in our arguments, there is the further question of neutrinos oscillating from one flavour to another as they propagate. This possibility was introduced by Pontecorvo (1968) in connection with the solar neutrino problem. The idea is that the neutrinos produced by the weak interactions are not in pure eigenstates of mass but in a coherent mixture. The components of this mixture propagate with different phase velocities because they have different masses and, if a wave packet is considered, one finds that its constitution expressed in terms of its states as defined by the weak interactions (that is, flavours) changes with time. Astrophysicists will recognise the analogy with Faraday rotation, wherein a linearly polarised wave, since it is a mixture of oppositely circularly polarised waves which propagate through a magnetised medium with different phase velocities, will rotate as it propagates through such a medium. In the neutrino case the effect is quantum mechanical and its relation to the Uncertainty Principle is clearly described by Kayser (1981).

The situation is that neutrino oscillations require at least one of the neutrino flavours involved to have a non-zero mass, but the converse is not true, that is, the existence of a non-zero mass does not guarantee that oscillations occur. If they do occur, then for two flavours one has a mixing matrix which describes the effect: e.g.

$$\begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}$$

Weak interaction eigen states — Mixing Matrix — Mass eigenstates

The mixing matrix is orthogonal and so defines a mixing angle α. The probability P for a certain neutrino, say ν_e, of energy E to turn into ν_μ after propagating a distance L is then given by

$$P = \sin^2 2\alpha \sin^2 \frac{(m^2_1 - m^2_2) L}{4 E} .$$

Thus a detector sensitive only to ν_e could measure $m^2_1 - m^2_2$, the difference of the squared masses of the propagating eigenstates.

If several flavours are involved the mixing matrix is larger and the situation more complicated. Moreover in this more general case the simple distinction hetween Majorana (2 component spinor) and Dirac (4 component spinor) neutrinos breaks down (Wolfenstein 1982).

Our final theoretical remark, which will be needed in the last section, is that the more massive neutrino, say ν_1, can decay into the lighter one, emitting a variety of other particles including photons. For example, the decaying neutrino can first turn virtually into a charged lepton and an intermediate charged vector boson W. The W then emits a real photon and recombines with the virtual charged lepton to form the less massive residual neutrino. In this case we would have overall

$$\nu_1 \rightarrow \nu_2 + \gamma .$$

If ν_1 is at rest, then conservation of energy and momentum in the decay process leads to

$$E_\gamma = \frac{m^2_1 - m^2_2}{2 m_1} .$$

If $m_2 \ll m_1$ we can simplify this to

$$E_\gamma \sim \frac{m_1}{2} .$$

The lifetime for this process depends critically on the number of neutrino flavours assumed, since this decides whether a powerful interference effect called GIM suppression is operating (GIM = Glashow, Iliopoulos, Maiani; see de Rujula and Glashow 1980). With only three

flavours present theoretical values for the radiative lifetimes for neutrinos of flavour 1 range from

$$\frac{10^{36}}{\mathrm{Sin}^2 2\beta}\left(\frac{30\ \mathrm{ev}}{m_1}\right)^5 \mathrm{sec}\ ,$$

to ten times this quantity (Pal and Wolfenstein 1982). Here β is the mixing angle between ν_1 and ν_2, which in some models is large (Goldman and Stephenson 1981). With wider assumptions about the number of neutrino flavours and associated Higgs types one obtains (de Rujula and Glashow 1980) a lifetime of

$$\frac{1.5 \times 10^{30}}{\mathrm{Sin}^2 2\beta_1}\left(\frac{30\ \mathrm{ev}}{m_1}\right)^5 \mathrm{sec}\ ,$$

or possibly less (Pal and Wolfenstein 1982), with β_1 the mixing angle between the fourth neutrino and ν_1.

We now consider briefly the experimental situation. Two recent claims have excited great interest, although they have not been generally accepted. The first involves the beta decay of tritium whose energy spectrum has been studied by Lyubimov *et al.* (1980). They infer from its shape at the high energy end that the mass m_ν of the electron (anti-)neutrino lies in the range

$$14 \leqslant m_\nu \leqslant 46\ \mathrm{ev}$$

at the 99% confidence level. For criticism of this experiment see de Rujula (1981).

The second experiment claims to have discovered neutrino oscillations, thereby implying the presence of a non-zero neutrino mass. It was performed by Reines, Sobel and Pasierb (1980) who used reactor neutrinos to study charged current reactions in hydrogen and charged and neutral current reactions in deuterium. The ratio of the charged to neutral cross-sections was found to be less than expected theoretically, and this discrepancy was attributed to the presence of neutrino oscillations. For criticisms of this experiment see Morrison (1980). Morrison discusses also a number of other on-going oscillation experiments. Present results are negative and are usually presented in the form of a diagram relating squared mass differences and mixing angles. The negative results then rule out certain regions in this diagram.

To summarise, there seems to be no reliable experimental evidence yet that any neutrino flavour has a non-zero rest-mass. Fortunately several experiments are now being performed which could, within a year from now, determine either a direct mass or a difference of mass squares. Similar remarks apply to the proton decay experiments. If they turn out to be positive this would strengthen our confidence in GUTS, which by implication would give some theoretical support to the hypothesis of a non-zero rest-mass for neutrinos.

3 - Cosmological Implications of Massive Neutrinos

In the early hot stages of the universe, when kT exceeded m_ν, neutrino-anti neutrino pairs could be produced by thermal excitation. Reactions of the type

$$e^+ + e^- \leftrightarrow \nu + \bar{\nu}$$

would then keep the neutrinos in thermal equilibrium with other particles so long as the reaction rates involved were shorter than the expansion rate of the universe. These reaction rates are determined by the weak interaction coupling constant and thermal equilibrium for left-handed neutrinos (Shapiro *et al.* 1980) would have been maintained until kT dropped to about 1 Mev (Weinberg 1972). Thereafter the neutrino distribution became frozen out. So long as $m_\nu \ll 1$ Mev one can continue to treat the massive neutrinos as a relativistic gas down to this freezing-out point. Their number-density, n_ν, would then have been of the same order as the photon number-density, n_γ, but because of the difference between Fermi and Bose statistics one would have (Weinberg 1972)

$$n_\nu = 3/4 \; n_\gamma \qquad (\mathrm{kT} \gtrsim 1 \text{ Mev}) \; .$$

When the temperature drops to about 1/2 Mev the electron-pairs recombine permanently giving up their energy to the photon gas. Since the neutrinos are now decoupled from the photons, they do not receive much of this energy and this further increases the difference between n_ν and n_γ. The new relationship is maintained indefinitely thereafter, unless, of course, other dissipation processes intervene to increase n_γ. This final relationship is (Weinberg 1972)

$$n_\nu \sim 1/4 \; n_\gamma \qquad (\mathrm{kT} < 1/2 \text{ Mev}) \; .$$

After the freezing-out process the neutrinos go out of thermal equilibrium when they become partially relativistic ($kT \sim m_\nu$). This is a general property of partially relativistic particles which are non-interacting but in a state of expansion or contraction. The momentum, p, of each particle varies inversely as the scale-factor, $R(t)$, but the energy of each particle, being given by $(p^2 + m^2_0)^{1/2}$, varies in a complicated manner and there is no way the temperature can vary with the scale-factor, R, so as to keep the energy distribution in an invariant form. This is, however, possible in the extreme limits of relativistic particles (when $E = p$ and $T \propto 1/R$) and completely non-relativistic particles (when $E = p^2/2m$ and $T \propto 1/R^2$). Thus the neutrino distribution today would be out of thermal equilibrium and, if a temperature is specified for it, one must be careful to state whether one is referring to a momentum distribution of the Fermi type for zero rest-mass particles or a mean energy property $E = 3/2\ kT$.

If the 3 K photons observed today all come from the hot Big Bang (we shall question this assumption later) then the present value of n_ν is about one-quarter of n_γ, which for a 3 K radiation-field gives ~ 400 photons cm^{-3}. We would thus have

$$n_\nu \sim 100\ cm^{-3}$$

for each neutrino flavour satisfying the assumptions we have been making. (For the more complicated calculation of n_ν when $m_\nu \gtrsim 1$ Mev see Lee and Weinberg 1977). Accordingly, the contribution of each flavour to the density parameter $\Omega = \rho/\rho_{crit}$ would be given by

$$\Omega_i \sim \frac{m_i}{100\ h_0^2}\ ,$$

where m_i is in ev and h_0 is related to the Hubble constant by

$$H_0 = 100\ h_0\ km\ sec^{-1}\ Mpc^{-1}\ .$$

This result, and its possible implications for a high density universe, have already been discussed in the introduction.

By contrast, the present typical velocity v_0 of a neutrino is far less than that of the photons. One finds

$$v_0 \sim \frac{50}{m_i} \text{ km sec}^{-1} ,$$

with m_i again measured in ev.

The question now arises, what constraints on m_i or $\sum_i m_i$ are implied by our formulae (other than $m_i \ll 1$ Mev)? Two types of constraints arise. The first is purely cosmological and involves the observational upper limit on Ω stemming from estimates of the age of the universe. The second involves the further hypothesis that massive neutrinos dominate clusters of galaxies or individual galaxies. We consider this second type of argument in the next section, and discuss here only the age of the universe.

In general the age of a relativistic model of the universe has to be computed numerically. This bas been done by various authors and is given in convenient form by Stabell and Refsdal (1966) in their table II, and also by Bernstein and Feinberg (1981) in connexion with the massive neutrino problem. To see what is involved we note that if $h_0 = 0.5$ (see Tammann's article in this volume, then $\sum_i m_i = 100$ would lead to an age of 9.5×10^9 years, while $\sum_i m_i = 150$ would yield 8.3×10^9 years. For $h_0 = 0.7$ and $\sum_i m_i = 100$ one obtains 8.2×10^9 years.

These ages are marginally less than the lowest ages usually quoted for globular clusters ($\sim 13 \times 10^9$ years, Demarque 1980). It is not clear how seriously one should take these differences. If they are taken seriously we are not necessarily forced to reduce $\sum_i m_i$ below the 100 to 150 range, although such reduced upper limits are often quoted in the literature. There are two possible alternative paths we can take.

(i) We can assume that the cosmical constant is non-zero. This would have implications for the observed value of the deceleration-parameter q_0 (which is discussed by Tammann in this volume), but at present it is an ill-determined quantity. In this way, the limit on $\sum_i m_i$ could be raised to a value of 200.

(ii) We could attribute a substantial fraction of the 3 K photons to processes occurring at intermediate red shifts, such as $z \sim 100$, thereby reducing (*) the concentration of neutrinos left over from the hot Big Bang

(*) I am grateful to Dr. A.C. Edwards for drawing my attention to this possibility.

and so decreasing Ω. The so-called population III sources involved could be a pregalactic generation of supermassive stars or, alternatively, accreting black holes. A recent discussion of these sources has been given by Carr (1981) together with references to earlier treatments.

Two astrophysical points are worth making here. The first, in favour of the supermassive stars, is that one needs in any case a mechanism for generating pre-galactic heavy elements in order to account for the shortage of stars with low metal abundance in the Galaxy. Heavy elements produced by population III stars could solve this problem, and also, by forming into molecules and dust particles, help to thermalise the emitted radiation in a manner which conforms with the presently observed spectrum of the 3 K background.

The second point concerns the question whether this mechanism could account naturally for the photon-baryon ratio of $\sim 10^{10}$ involved in the 3 K background. This has been discussed especially by Rees (1978). Indeed Rees claims that all the 3 K background could be due to population III sources, thereby doing away with the hot Big Bang altogether. The helium and deuterium normally associated with "the first three minutes" would then have to be attributed either to the Population III sources themselves, or to some unknown galactic processes occurring later. His main point is that it is quite natural for about 0.01 per cent of the rest-mass of a proton to be on average converted into radiation in a Population III source. If this occurred at a redshift ~ 100, one would end up with a photon-baryon ratio $\sim 10^{10}$.

Of course if we followed Rees all the way, there would be no cosmological neutrinos to worry about! However, we may recall that GUTS are having some success in accounting for the baryon asymmetry of the universe in a hot Big Bang model (see Weinberg's article in this volume). In principle, this theory also leads to a definite value for the photon-baryon ratio, which in some models is of the general order of 10^{10}. I do not regard it, therefore, as unduly artificial to contemplate a mixed model in which Population III sources provide a substantial fraction, but not all, of the photons in the 3 K background. This would again permit a considerable relaxation of the mass constraint on the neutrinos arising from their implications for the age of the universe.

4 - Neutrino Domination of Galaxy Clusters and Galaxies

As discussed elsewhere in this volume, it is now generally agreed that both galaxy clusters and many individual galaxies, including the Milky Way, contain some form of dark matter whose mass exceeds that of the visible matter by a factor of the general order of ten. The form this matter takes is presently unknown; it could conceivably be suitably faint stars, brick-sized objects or black holes of appropriate mass. Several people have recently added massive neutrinos to this list. The suggestion is attractive, since if neutrinos do have non-zero rest-mass and the hot Big Bang picture is correct, then a supply of neutrinos is in principle available. However, several conditions have to be satisfied if this suggestion is to be taken seriously.

These conditions arise from two types of consideration. The first type concerns the galaxy formation process itself. Can galaxies and clusters of galaxies form which are dominated by massive neutrinos? A number of recent papers have appeared discussing this question with markedly differing conclusions. Since galaxy formation is the subject of other articles in this volume and since it is a tricky subject which is difficult to summarise briefly, we prefer not to enter into it here.

The second type of consideration does admit of a brief discussion. Let us assume that neutrino dominated condensations have somehow managed to form. What constraints on the neutrinos are implied by observations of the present structure of these condensations?

The first of these constraints arises from a consideration of the velocity-dispersion of the neutrinos in the gravitational condensation. Since the neutrino distribution has to be spatially extended to account for the observed flat rotation curves of galaxies, one would expect that its velocity-dispersion would be in virial equilibrium with the gravitational potential of the condensation. (The possible role of degeneracy pressure is touched on in a later paragraph). This leads to a root mean square neutrino velocity of order 200 km sec^{-1} for the Milky Way.

One now requires that, at the time of galaxy formation, the cosmological velocity of the neutrinos be less than this on the general grounds that, however a galaxy forms, this velocity is likely to increase in the process. In other words, the cosmological neutrinos must not be so hot that they cannot form or join localised condensations. If the galaxy formed at a red shift z_f, a typical cosmological velocity at the corresponding epoch

would be given by

$$v_e \sim \frac{50\,(1 + z_f)}{m_\nu} \text{ km sec}^{-1} .$$

We thus require that

$$m_\nu > \frac{1 + z_f}{4} \text{ ev} .$$

If, for example, we take $z_f \sim 11$, then

$$m_\nu > 3 \text{ ev} . \tag{1}$$

This simple argument, therefore, leads us into a mass-range for the neutrinos which is already of interest to us from theoretical and experimental particle physics and from cosmological considerations of the mass-density and age of the universe.

A more stringent limit on the neutrino mass can be derived from the opposite type of argument, that is, by considering the role of the Pauli Exclusion Principle to which the neutrinos must conform. The basic idea is that if the neutrino mass were very low we would need a large volume concentration of them to produce the required gravitational field. In addition the momentum space concentration would be large. Too large a concentration would, however, lead to a violation of the Exclusion Principle. Thus this argument leads, like the previous one, to a lower limit on the mass of the neutrino.

A similar result is obtained if one assumes that galaxies are formed after the neutrinos have become collision-free. In this case their distribution function in phase-space satisfies the Liouville equation and so is a constant of the motion. One knows this distribution function at the outset of the formation process from the cosmological discussion and so one can again limit the resulting concentration of neutrinos. In an interesting extension of this idea Tremaine and Gunn (1979) have applied Lynden-Bell's (1967) ideas on violent relaxation to this problem. Such relaxation would lead to phase mixing. The resulting macroscopic or coarse-grained phase space density can only decrease, however, so that one still obtains a lower limit to the neutrino mass.

We now consider these arguments in more detail. For purposes of orientation it is illuminating, following Cowsik and McClelland (1973),

Melott (1980) and Gao and Ruffini (1980), to consider the extreme case of a galaxy as a gigantic "white dwarf" or "neutron star", that is, as a self-gravitating spherical assemhly of neutrinos in their ground state. We shall see that such a model is not altogether unrealistic, although it does need modification. Given the total mass, M, of such an assembly, assuming that it is significantly less than the critical mass at which the assembly becomes gravitationally unstable, one obtains a unique radius, R, which for non-interacting neutrinos would be given by

$$MR^3 = \frac{6\,\hbar^6}{G^3\, m_\nu^8} \,. \tag{2}$$

The critical mass M_c is given by

$$M_c \sim \frac{6.4 \times 10^{17}\, M_\odot}{m_\nu^2} \,.$$

It is amusing to note that when $m_\nu \sim 30$ ev, $M_c \sim 10^{15}\, M_\odot$, which is comparable to the mass of a supercluster of galaxies. This may or may not be significant. In any case the mass of a cluster of galaxies is less than critical for the range of neutrino masses of interest to us. We note in addition that the Jeans mass, M_J, is given by

$$M_J \sim 10^{17}\ M_\odot / m_\nu^2 \,.$$

Returning to the relation (2) for the ground state assembly we note, following Melott, that if we substitute $10^{12}\ M_\odot$ for M, and 100 kpc for R, values appropriate to an extended galaxy, we obtain

$$m_\nu \sim 30\ \text{ev} \,. \tag{3}$$

Thus this simple computation again leads to a neutrino mass of tens of electron volts. This result is suggestive but cannot be taken literally because the resulting density distribution deviates significantly from the $1/r^2$ law which, for a spherical distribution, would be required to account for a flat rotation curve.

It follows that the neutrinos dominating a galaxy cannot be completely degenerate. The intermediate case of partial degeneracy has been discussed

by Melott (1982), but for simplicity we consider here the other extreme of a non-degenerate assembly with the pressure being provided by a classical velocity dispersion. The flat rotation curve now suggests using an isothermal distribution, the parameters of which have been evaluated by Peebles (1980). Here we give a simplified discussion.

We begin from a dynamical analysis of the structure of the Milky Way in terms of the distribution and motion of known material and of an unknown spherical halo. Several models of this kind have been computed recently (Shuter 1981; Monet, Richstone and Schecbter 1981; and Caldwell and Ostriker 1981), of which the most detailed is that of Caldwell and Ostriker. One obtains from these analyses a density for the halo material which, at the Sun's position in the Galaxy, can be great as 10^{-24} gm cm^{-3} and is determined to within about a factor 2. If this halo material consists of neutrinos of rest-mass, m_ν, one then has for their concentration, n_ν, at the sun's position

$$n_\nu \sim 10^7 \left(\frac{100 \text{ ev}}{m_\nu} \right) \text{ cm}^{-3} .$$

This represents an enhancement over the cosmological concentration of $\sim 10^5$ for $m_\nu \sim 100$ ev.

Now the phase-space density f of these neutrinos has the general form

$$f \sim \frac{1}{(m_\nu \bar{v})^3} n_\nu .$$

But f must be less than h^{-3} in order to satisfy the Exclusion Principle. We thus require that

$$m_\nu^4 \gtrsim 10^{-24} h^3 v^{-3}$$

with m_ν, h and v now expressed in c.g.s. units. With $v \sim 200$ km sec^{-1} we obtain

$$m_\nu \gtrsim 30 \text{ ev} \qquad (4)$$

which is a more stringent limit than (1). Clearly the similarity of (3) and (4) suggests that the case of partial degeneracy should be kept in mind, as discussed by Melott (1981).

When this type of analysis is applied to galaxy clusters one has a lower density and a larger velocity dispersion to deal with. Thus the resulting

limit on m_ν is less severe ($\sim$ 10 ev). By contrast, in a star the density is very much higher and so the implied limit on m_ν is much greater than any value of interest to us here. There is no question, therefore, of neutrino domination of a star! The conclusion of this discussion is then that, if we leave aside the formation problem, galaxies can be dominated by massive neutrinos if m_ν lies between about 30 ev and 200 ev, and galaxy clusters can be so dominated if m_ν lies between about 10 ev and 200 ev.

There remains then the question, could such neutrinos be detected by any other property than the gravitational field which they exert? To this question we now turn.

5 - Massive Neutrinos and Ultra-Violet Astronomy

The most promising possibility for detecting galactic neutrinos appears to involve the ultra-violet photons which they may emit (de Rujula and Glashow 1980). These photons might be observed directly or from their ionising effects if they are sufficiently energetic. We have discussed their energy and the lifetime for the process in a previous section of this article. Here we consider the astronomical aspects.

We note first the coincidence that the photon flux at the Earth coming from galactic neutrinos has the same order of magnitude as that coming from cosmological neutrinos (ignoring absorption effects for the moment). This follows from the fact that the galactic enhancement in the neutrino concentration for $m_\nu \sim 100$ ev is of the same order ($\sim 10^5$) as the ratio of the radius of the universe to the scale-height of the galactic halo ($\sim$ 30 kpc). Tbe main difference is that the galactic flux is nearly monochromatic (since the velocity-dispersion of the galactic neutrinos $\sim$ 200 km sec^{-1} $\ll c$), whereas the cosmological flux would be drawn out into a continuous spectrum by the differential red shift associated with the expansion of the universe.

In fact this spectrum would have the form

$$\underset{(\lambda \geqslant \lambda_0)}{I_\lambda} = \frac{c^{n_\nu}\,(z=0)}{H_0\, 4\pi\tau} \frac{\lambda_0^{3/2}}{\lambda^{5/2}} \left(1 + (2q_0 - 1)\left(1 - \frac{\lambda_0}{\lambda}\right)\right)^{-1/2},$$

where λ_0 is the rest wavelength of the decay photon, λ the observed wavelength, τ the neutrino lifetime and q_0 the deceleration parameter. Any absorption effects would have to be added to this relation. [There is a

numerical error of $\sim 10^4$ in equation (7c) of de Rujula and Glashow (1980) which has been corrected by Kimble, Bowyer and Jakobsen (1981)].

It has been pointed out by Stecker (1980) and Kimble, Bowyer and Jakobsen (1981) that one can use this spectrum to obtain lower limits on τ from the observed u-v background even if λ_0 corresponds to a wavelength at which the Galaxy is opaque (e.g. $\lambda_0 < 912$ Å). Since the observed background is probably due to other sources, it has been used to limit τ rather than to determine it. From an observed background ~ 200 to 300 photons cm^{-2} sec^{-1} $ster^{-1}$ $Å^{-1}$, they deduce that

$$\tau > 10^{22} - 10^{23} \text{ sec} \quad , \quad 10 \text{ ev} < \frac{hc}{\lambda_0} < 50 \text{ ev}$$

if the intergalactic medium is transparent out to the largest redshifts involved ($z \sim 6$). We return to the question of this transparency later.

Stecker (1980) also discussed the possibility that a reported increase in the background at 1700 Å might represent a photon flux from galactic neutrinos and he suggested that $\tau \sim 3 \times 10^{24}$ sec. According to Kimble, Bowyer and Jakobsen (1981), this increase has not been confirmed and is best treated as an upper limit to the actual intensity.

These authors also derived a lower limit on τ for galactic neutrinos from the general observed background in the 30 to 50 ev range (which corresponds to the mass range in which neutrinos of cosmological origin could dominate the Galaxy). Their limit depends on the uncertain opacity of the Galaxy at these photon energies and on the photon energy itself, but lies in the range

$$\tau > 10^{20} - 10^{22} \text{ sec} \quad , \quad 30 \text{ ev} < E_\gamma < 50 \text{ ev} \; .$$

More stringent limits have been derived by Shipman and Cowsik (1981) and Henry and Feldman (1981) from optical and ultra-violet observations of the Virgo and Coma clusters of galaxies. If these clusters are dominated by neutrinos of appropriate mass the derived limits are

$$\tau > 10^{23} - 10^{25} \text{ sec} \quad , \quad 1 \text{ ev} < E_\gamma < 10 \text{ ev} \; .$$

Shipman and Cowsik consider that with existing or proposed instruments one could improve these limits up to the range 10^{26} to 10^{27} sec.

All these limits are derived from direct observations of photon fluxes. One can also use arguments derived from the ionising effects of the photons,

as pointed out by Melott and Sciama (1981). These differ from the previous arguments in that they can lead to limits on (or evaluation of) photon fluxes at positions distant from the galactic plane, where the effects of absorption by neutral hydrogen or dust will be different (and in general less). For example, Melott and Sciama (1981) demanded that these photon fluxes should not completely ionise the high velocity clouds. These clouds are neutral hydrogen features observed at 21 cm to be predominantly approaching us with velocities of a few hundred kilometres per second. Various arguments suggest that some clouds lie at least a kiloparsec above the galactic plane, while a recent estimate puts them at tens of kiloparsecs away. In fact the limits obtained for τ are valid so long as the clouds lie within the Local Group of galaxies, but not within the galactic plane. One finds in this way that

$$\tau > 10^{24} \text{ sec} \qquad E_\gamma > 13.6 \text{ ev} .$$

There is no suggestion in any of these arguments (with the exception of Stecker 1980) that one is close to observing a real effect which could be attributed to photons from neutrino decay. Moreover, the theoretically expected value of τ is several orders of magnitude greater than even the largest of the lower limits so far quoted. The question therefore arises whether a more sensitive ionisation process can he discovered which would lead either to an actual effect being identified, or at least to a much more stringent limit on τ in the relevant photon energy range.

One possible such process is the ionisation of the intergalactic medium (Sciama 1982a). This ionisation is usually invoked to account for the absence of absorption troughs in quasar spectra (Gunn-Peterson effect), which leads to extremely low upper limits ($\sim 10^{-12}$ cm^{-3}) for the intergalactic density of atomic hydrogen and helium. Recently tentative positive observational evidence has been obtained (Gondhalekar *et al.* 1982) for an absorption trough due to singly ionised helium. This evidence was derived from a spectrum of the quasar 2240-408, whose red shift is 3.18, obtained using the International Ultra-Violet Explorer. If confirmed, this would show that: (a) an IGM actually exists; (b) it contains helium; and (c) the helium is singly ionised but not all doubly ionised.

The conventional view is that the IGM is probably ionised by photons emitted by the quasars themselves (Sherman 1981). Such a mechanism is plausible but is hard to quantify, especially because of uncertainties in the emissivity of quasars in the hard ultra-violet and in their spatial distribution at large red shifts. It is, therefore, of interest to examine the alternative

hypothesis that the ionising photons are emitted by a cosmological distribution of decaying neutrinos.

This hypothesis leads to stringent restrictions on the possible mass m_ν and radiative lifetime τ of the neutrinos involved. Since helium in the IGM is singly ionised we require $m_\nu \geqslant 50$ ev. This would be compatible with the upper limit on m_ν derived from lower limits on the age of the universe. If helium in the IGM is indeed not doubly ionised we require $m_\nu \leqslant 110$ ev. Thus the relevant neutrino type must have a mass lying in the range

$$50 \text{ ev} \lesssim m_\nu \lesssim 110 \text{ ev} .$$

We can place limits on the radiative lifetime, τ, by requiring that we satisfy all the observational constraints implied by the data relating to absorption troughs. This question is discussed elsewhere (Sciama 1982a). Since the IGM is ionised at least out to a red shift ~ 3.5, we find that we require

$$n_{\mathrm{H}}(0)\,\tau \leqslant 3 \times 10^{18}\,\mathrm{cm}^{-3}\,\mathrm{sec} ,$$

where $n_{\mathrm{H}}(0)$ is the present density of ionised hydrogen in the IGM.

It is interesting to make the subsidiary hypothesis that the IGM is neutral at red shifts greater than 3.5 and that this explains the red shift cut-off in the quasar distribution (Osmer 1982) which is discussed by Schmidt in this volume. In that case

$$n_{\mathrm{H}}(0)\,\tau \sim 3 \times 10^{18}\,\mathrm{cm}^{-3}\,\mathrm{sec} .$$

This condition ensures that there are, roughly speaking, enough photons to ionise every hydrogen atom in the IGM. We still have to ensure that this ionisation is complete enough to be compatible with low upper limits on atomic hydrogen and helium in the IGM. Calculation of the ionisation balance shows that, for example, at $z \sim 3.5$,

$$n_{\mathrm{HI}} \sim 10^{-4}\, n_{\mathrm{H}}(3.5) .$$

The observations would then be satisfied if

$$n_{\mathrm{H}}(3.5) \lesssim 5 \times 10^{-7}\,\mathrm{cm}^{-3}$$

so that

$$n_{\mathrm{H}}(0) \lesssim 5 \times 10^{-9}\,\mathrm{cm}^{-3} .$$

The upper limit on $n_H(0)$ is a few per cent of the baryon density arising from galaxies, if all "missing matter" is of non-baryonic form. The efficiency of galaxy formation is unknown, but an efficiency of a few per cent is not unreasonable. With this limit on $n_H(0)$ we arrive at the following limit for τ (assuming that the critical ionisation red shift ~ 3.5):

$$\tau \gtrsim 10^{27} \text{ sec} .$$

Another possible ionisation process due to decaying neutrinos has been proposed by Sciama and Melott (1982). It should be treated with great caution, since more conventional explanations for the ionisation involved are quite plausible. Nevertheless these conventional explanations have not yet been established with certainty and, since the parameters required by the neutrino hypothesis are similar to those derived from our intergalactic discussion, we touch on it here.

It has recently been discovered by the International Ultraviolet Explorer (IUE) that measurable column-densities of highly ionised carbon (CIV) and silicon (Si IV) exist in the galactic halo several kiloparsecs from the plane [Savage and de Boer (1979, 1981); Ulrich *et al.* (1980); Bromage, Gabriel and Sciama (1980)]. The existence of sucb high ionisation stages can plausibly be attributed either to collisional effects in a hot gas ($T \sim 10^5$ K) or to photo-ionisation due to hot stars or the soft X-ray background.

The alternative explanation is that the photons arise from the decay of galactic neutrinos. The photon energy would have to be at least 47.9 ev, the ionisation potential of CIII, so that $m_\nu \geqslant 95.8$ ev. This would satisfy the neutrino domination constraint (4), the cosmological constraint based on the age of the universe, and our intergalactic constraint. Indeed, if both our hypotheses are correct, the neutrino mass is confined to lie in the narrow range 95.8 ev $\leqslant m_\nu \leqslant$ 110 ev.

Our second hypothesis also leads to an actual estimate of the lifetime, τ. To evaluate τ we first determine the photon flux required to account for the observations. These observations lead to a CIV number density $\sim 10^{-9}$ cm^{-3}, and to densities of SiII, SiIV, CII, CIV, and NV roughly in the ratio 10 : 1 : 120 : 4 : < 1 (unobserved), with localised variations. In a photoionised gas these ratios are set by charge exchange with HI, HII and HeII, as well as by direct photoionisation and recombination with free electrons. The rates of the processes which we need to know have all been given in the literature and their application to this

prohlem has heen discussed hy Sciama (1982b). It is straightforward to determine whether a monochromatic flux of 50 ev photons can reproduce both the observed ionisation ratios and the absolute values of the ion densities. We find that with an assumed ambient gas density of $\sim 10^{-4}$ cm^{-3} and a normal cosmic abundance of He, C, Si and N, we can satisfy all the observations with a 50 ev photon flux of ~ 600 cm^{-2} sec^{-1}. With this value of the photon flux the steady state ionisation balance which we are assuming to hold would be estahlished in a time less than the age of the Galaxy.

The photon flux produced by halo neutrinos has been estimated by de Rujula and Glashow (1980). Since the flux from extragalactic neutrinos could be comparable we must consider the opacity of halo material for 50 ev photons. The mean free path, ℓ, for such a photon is given by

$$\ell \sim \frac{1}{n\,(\mathrm{HI})\,\sigma} .$$

If the halo is opaque the column-density of HI within one mean free path is

$$n\,(\mathrm{HI})\,\ell \sim 10^{19}\ \mathrm{cm}^{-2} .$$

If $n\,(\mathrm{H}) \sim 10^{-4}$ cm^{-3} and $n\,(\mathrm{HI}) \sim 1/3\,n\,(\mathrm{H})$, we have $\ell \sim 100$ kpc, which is of the same order as the size of the halo. The observed column density of HI at high galactic latitudes, including the disk contribution, is ahout 2×10^{20} cm^2 (Dickey *et al.* 1978), which is comfortably larger than the value we are taking for the halo alone. We conclude that it is likely that at a height of several kiloparsecs the extragalactic flux is shielded out, but that in computing the internal flux the opacity is not an important factor. Using the value for n_ν given earlier, we obtain for the photon flux I_γ

$$I_\gamma \sim \frac{6 \times 10^{29}}{\tau\ \mathrm{sec}}\ \mathrm{cm}^{-2}\ \mathrm{sec}^{-1} .$$

Since the flux required to account for the observations is ~ 600 cm^{-2} sec^{-1}, we conclude that our hypothesis requires that

$$\tau \sim 10^{27}\ \mathrm{sec}\quad ,\quad E_\gamma \sim 50\ \mathrm{ev} .$$

This is comfortably longer than the previous lower limits and remarkably similar to the estimate derived from our intergalactic hypothesis.

It is, perhaps, worth considering the relation of the actual value of τ derived here to estimates from theoretical particle physics. These estimates were discussed in an earlier section. We note that agreement can be obtained for $m_\nu \sim 100$ ev if one assumes that there is a fourth neutrino flavour and that the relevant mixing angle is not very small. In view of the sensitive dependence of τ on m_ν ($\propto m_\nu^{-5}$), this agreement is not entirely trivial. We therefore pursue here briefly the significance for particle physics of our requirement that $m_\nu \sim 100$ ev (Sciama 1981).

Both our cosmological considerations and the dependence of τ on m_ν imply that 100 ev is the mass of the most massive neutrino which is less than ~ 1 Mev. Now there is one particular recent particle physics model which also leads to a value for the mass $m_{\nu,\,\max}$ of the most massive neutrino. This is a left-right symmetric Grand Unified Theory based on the gauge group SO (10). According to Fukugita, Yanagida and Yoshimura (1981) this theory leads to a value for $m_{\nu,\,\max}$ which depends on the baryon asymmetry parameter, on the assumption that this parameter is determined by GUTS in the manner discussed in Weinberg (1979). These authors find that for the simplest feasible Higgs-fermion coupling.

$$m_{\nu,\,\max} \sim 100 \left(\frac{m_t}{200\ \text{Gev}} \right)^{1.8} \left(\frac{3 \times 10^{10}\ \text{Gev}}{m_x} \right)^{0.4} \left(\frac{n_\gamma / n_b}{10^{11}} \right) \text{ev} \ ,$$

where m_t is the mass of the top quark, m_x is the mass of the Higgs boson and n_γ/n_b is the photon-baryon ratio for the 3 K background. This result is independent of the number of fermion generations if $m_t > m_b$ in the heaviest generation.

It is a striking consequence of this formula that $m_{\nu,\,\max}$ is about 100 ev, the value we require, only if all the parameters contained within it take on their most extreme values as allowed by either theoretical or observational considerations. Thus ~ 200 Gev is the maximum value allowed for m_t from vacuum stability considerations (Hung 1979, Politzer and Wolfram 1979, Cabibbo, Maiani, Parisi and Petronzio 1979). A recent observational analysis (Barger, Long, Ma and Pramudita 1982) concludes that either $m_t = 50 \pm 25$ Gev or $m_t > 100$ Gev. Moreover a theoretical attempt to determine m_t from renormalisation group arguments, on the assumption

that there is an infra-red stable quasi-fixed point (Pendleton and Ross 1981, Hill 1981), leads to $m_t \sim 250$ Gev.

Secondly, nuclear stability requires that m_x cannot be less than 3×10^{10} Gev. This limit is based only on the charm state. The admixture given by the quark mixing matrix $| U_{im} |$ is unknown, but must be less than 1/50 for this limit to apply with $m_t \sim 200$ Gev. Otherwise m_x would appreciably exceed 3×10^{10} Gev.

Finally we note that the maximum permitted value of n_γ/n_b is of order 10^{11}. This limit is derived from the supposition that only known types of stars contribute to n_b through their baryon content. Any greater mass determination derived dynamically would have to involve non-baryonic matter. Of course in the spirit of this paper we would attribute all such excess material to massive neutrinos.

We conclude that our hypothesis for the ionisation mechanism producing CIV in the galactic halo leads to very stringent requirements on the age and baryon density of the universe and on m_t and m_x, at least in this particular Grand Unified Theory. One could take the view that this makes our hypothesis rather implausible. However, we prefer to hold on to it until it has been definitely refuted. In this connection we note in particular that, if as in some models (e.g. Witten 1980) m_ν is proportional to the squared mass of the associated lepton, then our hypothesis would be compatible with an unmeasurably small e neutrino mass, and unmeasurable oscillations between the e, μ and τ neutrinos. Accordingly, negative results from the current round of experiments need not rule out the cosmological breakthrough on the neutrino domination of the universe and of galaxies, which has been the subject of our speculations.

I am grateful to my student A.L. Melott for convincing me of the potential importance of galactic neutrinos and to G.E. Bromage, A.C. Edwards, G. Feinberg and L. Wolfenstein for helpful discussions and correspondence.

REFERENCES

Barger, V., Long, W.F., Ma, E. and Pramudita A., 1982, to be published.

Bergkvist, K.E., 1972, *Nucl. Phys.* **B39**, 317.

Bernstein, J. and Feinberg, G., 1981, *Phys. Lett.* **101B**, 39; **103B**, 470.

Bromage, G.E., Gabriel, A.H. and Sciama, D.W., 1980, *Proc. 2nd European IUE Conf.* (ESA SP-157, April 1980).

Cabibbo, N., Maiani, L., Parisi, G. and Petronzio, R., 1979, *Nucl. Phys.* **B158**, 295.

Caldwell, J.A.R. and Ostriker, J.P., 1981, *Ap. J.* **251**, 61.

Carr, B.J., 1981, *Mon. Not. Roy. Astr. Soc.* **195**, 669.

Cowsik, R. and McClelland, J., 1972, *Phys. Rev. Lett.* **29**, 669.

— 1973, *Ap. J.*, **180**, 7.

Demarque, P., 1980, *I.A.U. Symposium No. 85, Star Clusters*, ed. J.E. Hesser (Reidel, Dordrecht 1980), p. 281.

Dickey, J.M., Salpeter, E.E. and Terzian, Y., 1978, *Ap. J. Suppl.* **36**, 77.

Fermi, E., 1934, *Nuovo Cim.* **11**, 1.

Fukugita, M., Yanagida, T. and Yoshimura, M., 1981, *Phys. Lett.* **106B**, 183.

Gao, J.G. and Ruffini, R., 1980, *Phys. Lett.* **97B**, 388.

Goldman, S. and Stephenson, T., 1981, *Phys. Rev. D.* **24**, 236.

Gondhalekar, P.M., Malin, D.F., Boggess, A., Wilson, R. and Wu, C-C, 1982, to be published.

Henry, R.C. and Feldman, P.D., 1981, *Phys. Rev. Lett.* **47**, 618.

Hill, C.T., 1981, *Phys. Rev.* **24**, 691.

Hung, P.Q., 1979, *Phys. Rev. Lett.* **42**, 873.

Kayser, B., 1981, *Phys. Rev.* **24**, 110.

Kimble, R., Bowyer, S. and Jakobsen, P., 1981, *Phys. Rev. Lett.* **46**, 80.

Langacker, P., 1981, *Phys. Rep.* **71**.

Lee, B.W. and Weinberg, S., 1977, *Phys. Rev. Lett.* **39**, 165.

Lynden-Bell, D., 1967, *Mon. Not. Roy. Astr. Soc.* **136**, 101.

Lyubimov, V.A., Novikov, E.G., Nozik, V.Z., Tretyakov, E.F. and Kosik, V.S., 1980, *Phys. Lett.* **94B**, 266.

Melott, A., 1980, unpublished manuscript.

— 1982, to be published.

Melott, A. and Sciama D.W., 1981, *Phys. Rev. Lett.* **46**, 1369.

Monet, D.G., Richstone, D.O. and Schechter, P.L., 1981, *Ap. J.* **245**, 454.

Morrison, D.R.O., 1980, *CERN/EP* 80-190.

Osmer, P., 1982, to be published.

Pal, P.B. and Wolfenstein, L., 1982, *Phys. Rev. D.* **25**, 766.

Peebles, P.J.E., 1980, in *Physical Cosmology*, eds. R. Balian, J. Audouze and D.N. Schramm (North-Holland Amsterdam), p. 265.

Pendleton, B. and Ross, G.G., 1981, *Phys. Lett.* **98B**, 291.
Politzer, H.D. and Wolfram, S., 1979, *Phys. Lett.* **82B**, 242; **83B**, 421.
Pontecorvo, B., 1968, *Sov. Phys. JETP* **26**, 984.
Rees, M.J., 1978, *Nature* **275**, 35.
Reines, F., Sobel, H.W. and Pasierb, E., 1980, *Phys. Rev. Lett.* **45**, 1307.
Rujula, A. de and Glashow, S.L., 1980, *Phys. Rev. Lett.* **45**, 942.
Rujula, A. de, 1981, *T.H. 3045-CERN.*
Savage, B.D. and de Boer, K., 1979, *Ap. J. Lett.* **230**, L 77.
— 1981, *Ap. J.* **243**, 460.
Sciama, D.W., 1981, *ESO Inauguration Symposium, Evolution in the Universe.*
— 1982a, *Mon. Not. Roy. Astr. Soc.* **198**, 1 P.
— 1982b, *Oxford International Symposium, Progress in Cosmology* (Reidel 1982).
Sciama, D.W. and Melott, A.L., 1982, *Phys. Rev. D.* **25**, April 15.
Shapiro, S.L., Teukolsky, S.A. and Wasserman, I., 1980, *Phys. Rev. Lett.* **45**, 669.
Sherman, R.D., 1981, *Ap. J.* **246**, 365.
Shipman, H.L. and Cowsik, R., 1981, *Ap. J. Lett.* **247**, L 111.
Shuter, W.J.H., 1981, *Mon. Not. Roy. Astr. Soc.* **194**, 851.
Stabell, R. and Refsdal, S., 1966, *Mon. Not. Roy. Astr. Soc.* **132**, 379.
Stecker, F.W., 1980, *Phys. Rev. Lett.* **45**, 1460.
Tremaine, S. and Gunn, J.E., 1979, *Phys. Rev. Lett.* **42**, 407.
Ulrich, M.H. et al., 1980, *Mon. Not. Roy. Astr.. Soc.* **192**, 561.
Weinberg, S., 1972, *Gravitation and Cosmology* (Wiley, New York).
— 1979, *Phys. Rev. Lett.* **43**, 1566.
Witten, E., 1980, *Phys. Lett.* **91B**, 81.
Wolfenstein, L., 1982, to be published.

DISCUSSION

OSTRIKER

The density you report is $n = 10^{-8.5}$ cm^{-3} now. If the pressure parameter is $p/k \equiv \tilde{p}$ then the temperature now is $T = 10^{8.5}\,\tilde{p}$ now and $\geq 10^{9.7}\,\tilde{p}$ at $z = 3$, since T increases at least as fast as $(1 + z)^2$ in the past. Since you say that n_{HI} is obscured at $z = 3$ the temperature cannot be higher than $10^{5.5}$ i.e. $T < 10^{5.5}$ at $z = 3$. Combining these inequalities we see that you require the present intergalactic pressure to be less than $10^{-4.2}$ K = P. This is a bit lower than one would think on other grounds.

SCIAMA

I do not tbink that we have any unambiguous evidence for the value of the pressure in the IGM.

DAVIS

The dynamical problem associated with massive neutrinos is an extremely serious one in my opinion. This morning Joe Silk showed that the damping mass of neutrino fluctuations is larger than 10^{15} $M_\odot$, so that galaxy formation is made even more difficult than it is in the standard scenarios. If pancakes of the damping mass form as the first collapsed structure in the universe, I don't see how the neutrinos can be stuffed back into a small structure as a halo with $1/r^2$ and with the proper phase space density.

To avoid this disaster a group of us (myself and Ed Witten, Mike Lecar, and Tad Pryor) have invented several fanciful schemes that preserve galaxy sized fluctuations in a neutrino dominated universe. None of these schemes are likely to be correct, but they exhibit the lengths to which one must go to preserve galaxies in a neutrino dominated universe (Davis, M. *et al.* 1981, *Ap. J.*, **250**, 423).

Three schemes were considered. First, if there exist isothermal seeds such as black holes or large amplitude baryon fluctuations of mass 10^8 $M_\odot$, then galaxies of 10^{12} $M_\odot$ can form hy accretion in one Hubble time. Peebles has considered the latter possibility in some detail. The scenario is not ruled out

but seems ad hoc because we simply are replacing one mystery for another. Why should the seeds exist?

A second scenario is to realize that the damping mass of neutrino perturbations varies as the present momentum temperature of the neutrino as T_ν^{12}, so that if T_ν can be reduced from its canonical value of 1.9 K to under 1.0 K, neutrino fluctuations on the size of galaxies will be preserved. This can be accomplished by the generation of additional entropy after the epoch of neutrino decoupling, such as might result by the radiative decay of a massive neutrino. However it is difficult to thermalize the new photons without first destroying all the fragile deuterons, an effect first pointed out by Lindley. For the moment this scenario is therefore moribund.

A third scenario would result if a massive neutrino, perhaps the tau neutrino, decayed non radiatively into three lighter neutrinos. The massive neutrino must decay only after the lighter neutrinos become non relativistic and have clustered in the potential wells defined by the heavier neutrinos. For a wide range of parameters, galaxy sized perturbations can be retained. The massive neutrino must decay non radiatively to avoid the constraint on the visible and near IR photons background, but if the decay resulted in additional massive, but lighter neutrinos, these will today be too hot to cluster even on supercluster scales and will comprise a nearly uniform background of perhaps half the mass density of the universe. This could explain why measured M/L estimates continue to increase even beyond the size scale of rich clusters, which presumably form in a dissipationless fashion. The chief weakness of this model is that the non radiative decay of the heavy neutrino is unknown but is likely to be so long that the universe today would still be radiation dominated from its decay products.

WEISSKOPF

Are you now inventing new physics?

DAVIS

Yes, these scenarios will not work for all grand unified models, but models that forbid electromagnetic decay of a heavy neutrino have been considered by de Rugula and Glashow, among others.

WEISSKOPF

It is most improbable that a particle theory gives a heavy neutrino that

decays only in light neutrinos and does not emit photons. The theory would have to be even more artificial than the present G.U.T. theories.

FABER

Could you tell us what the status is of heavier neutrinos, around a few Gev, in cosmology? Is there any reason known why the non-luminous matter might not consist of this type of neutrino?

SCIAMA

Gev neutrinos would be non-relativistic at freeze-out, and so their number density would have been suppressed by annihilation. They could not now play an important role in cosmology or galactic structure.

GUNN

The suppression due to non-relativistic decoupling and decay does not set in seriously until masses of the order of several Mev. The result is a forbidden (by $\Omega \leq 1$) range of neutrino masses between 100 ev and 2 Gev.

SOME REMARKS ON PHASE-DENSITY CONSTRAINTS ON THE MASSES OF MASSIVE NEUTRINOS

JAMES E. GUNN
Princeton University Observatory

Some time ago Tremaine and I (1979) published an often misquoted paper which pointed out that there are severe constraints on the masses of light "massive" (a few tens of eV) neutrinos, if one wishes to retain the simple, and to us compelling, hypothesis that all the dark matter in the universe is the same stuff and is, in fact, made out of those same neutrinos. We concluded, in fact, that either that hypothesis must be given up, or the missing mass is not made of such particles. The problem comes basically from the halos of galaxies, where the densities are high and the velocity dispersions relatively low, which makes the phase-density rather high. The phase-density can only decrease during violent relaxation, as pointed out by Lynden-Bell (1967), and the phase-density in the early universe is given uniquely by the neutrino mass by considerations of thermal equilibrium.

If the neutrinos are to make a structure of one-dimensional velocity dispersion σ and core radius a, the mass of the particles must obey the following inequality in order for the phase density not to have increased during the formation of that structure:

$$m_\nu^4 \geqslant \frac{9\,\hbar^3}{4\,(2\pi)^{5/2}\,g\,G\,\sigma\,a^2} \tag{1}$$

or

$$m_\nu > 101 \left(\frac{100\ \text{km/s}}{\sigma}\right)^{1/4} \left(\frac{1\ \text{kpc}}{a}\right)^{1/2} g^{-1/4}\ \text{eV} \tag{2}$$

Here g is the statistical weight of the particle state, 1 for Majorana neutrinos and 2 for Dirac ones. In the Galaxy the core radius of the halo must be quite large to explain the rotation curve (see, for example, the models of Ostriker and Caldwell 1979), but in any case cannot be much larger than the radius at the solar circle, about 8 kpc. The velocity dispersion in the halo material must be about 160 km/sec (see my paper on galactic evolution in this volume). The corresponding lower limit on the mass is

$$m_\nu > 32\, g^{-1/4}\ \text{eV} \tag{3}$$

and the corresponding limit on Ω is

$$\Omega > 0.33\, g^{3/4} \left(\frac{100\ \text{kms/s / Mpc}}{H_0} \right)^2 \tag{4}$$

The mass below which phase damping occurs is about $10^{14}\ M_\odot$.

Tremaine and I argued that such large values of Ω which would certainly be reflected in the measured mass-to-light ratios of groups, and especially large clusters whose masses exceed the damping mass, disagreed with the observations. This is still true, but by a smaller margin than was the case two years ago, especially if recent trends towards a larger Hubble constant turn out to be correct (Aaronson *et al.* 1980). The luminosity density in the universe is almost certainly larger than the value we used: the one derived from the catalog of Kirshner, Oemler, and Schechter (1979), with the corrections discussed in my evolution paper, is

$$\mathcal{L} = 3.1 \times 10^8\ L_{\odot B}\ \text{Mpc}^{-3}\ (H_0/100)\ . \tag{5}$$

The closure mass-to-light ratio is

$$M/L_{BT})_{crit} = 822\ (H_0/100) \tag{6}$$

and the mass-to-light for the field population as derived from the compilation of Faber and Gallagher (1979) is about

$$M/L)_{field} = 140\ (H_0/100) \tag{7}$$

which yields a value of Ω of 0.17, only a factor of 2 away from that of equation (4), if $H = 100$. If, however, H is as small as 50, the discrepancy is a factor of 8. Given the severe timescale (age) problems associated with large values of the Hubble constant, especially for such relatively large values of Ω, it is clear that either the original problem or worse ones still persist.

REFERENCES

Aaronson, M., Mould, J., Huchra, J., Sullivan, W., Schommer, R., Bothun, G., 1980, *Ap. J.* **239**, 12.

Faber, S., Gallagher, J., 1979, *Ann. Rev. Ast. and Ap.* **17**, 135.

Kirshner, R., Oemler, A., Schechter, P., 1979, *A. J.* **84**, 951.

Lynden-Bell, D., 1967, *M.N.R.A.S.* **136**, 101.

Ostriker, J., Caldwell, J., 1979, in "*The Large-Scale Characteristics of the Galaxy*", IAU Symposium No. 84, p. 441.

Tremaine, S., Gunn, J., 1979, *Phys. Rev. Letters* **42**, 467.

DISCUSSION

SCIAMA

My student Melott is computing numerically a collapse problem with massive neutrinos to see, amongst other things, how easily the inequality of m_ν is saturated.

LYNDEN-BELL

A number of numerical experiments were originally done on violent relaxation. Many of those experiments show a little pip in the middle that remained degenerate, although this could be avoided in the most violent collapses.

FABER

It might be worth noting that we just manage to squeeze the neutrinos into our galaxy, which is pretty sizeable as galaxies go. Since we know that there are galaxies with smaller characteristic velocities and since it is plausible that such galaxies will have smaller scale lengths, the lower limit on the neutrino mass would be increased in these objects. It is clearly important to study the rotation curves of small galaxies closely to determine whether, in fact, they are surrounded by non-luminous halos and, if so, to see what the scale length of these halos are.

GUNN

Indeed.

FABER

Since it is important for the neutrino question to know what is the smallest galaxy containing non-luminous material, I wonder whether there are many dwarf spirals and irregulars for which Westerbork can make 21-cm maps and rotation curves.

VAN DER LAAN

We can in principle measure rotation curves (or limits on their shape) at Westerbork for hydrogen-rich dwarf galaxies bigger than $\sim$ 20 arcseconds. In fact, some such information is probably already at hand or in the reduction process.

THE BOUNDARY CONDITIONS OF THE UNIVERSE

S.W. HAWKING
University of Cambridge, Cambridge

This paper considers the questions of what are the boundary conditions of the universe and where should they be imposed. It is difficult to define boundary conditions at the initial singularity and, even if one could, they would be insufficient to determine the evolution of the universe. In order to overcome this problem it is suggested that one should adopt the Euclidean approach and evaluate the path integral for quantum gravity over positive definite metrics. If one took these metrics to be compact, one would avoid the need to specify any boundary conditions for the universe. This approach might explain why the apparent cosmological constant is zero, why the universe is spatially flat, and why it was in thermal equilibrium at early times.

The aim of physics is to provide a mathematical model of the universe which will agree with all observations that have been made so far and which will predict the results of further observations. Our present models would consist of two parts:

i) a set of differential equations that govern the variables in the theory. These are normally derived from an action principle;

ii) boundary conditions for the differential equations, or for the fields that are considered in the action principle.

One could conceive of models which did not have this division into field equations and boundary conditions but the separation of the two has been very valuable in enabling us to make progress by considering only a

local region of the universe instead of having to try to devise a model which would account for the whole universe at one go. We have made a lot of progress on the first part of the problem in recent years and it now seems possible that we might find a fully unified field theory within the not-too-distant future. However, we shall not have a complete model of the universe until we can say more about the boundary conditions than that they must be whatever would produce what we observe.

In Newtonian theory, the gravitational potential obeys an elliptical equation in three-dimensional space at a given time. However, since the discovery of special relativity, we have believed that the fundamental fields all obey hyperbolic equations with the same characteristics, the null cone. This means that the boundary values can be taken to be data on a space-like surface. In the General Theory of Relativity the metric that defines the null cone is itself a dynamical field. Thus the causal structure of space-time may be very different from that of flat space. There may be closed or almost closed time-like curves such as shown in Figure 1. In this case the data on a space-like surface would have to be such that it reproduced itself exactly after a certain interval. This would be a stringent requirement which might have only one or a small number of solutions. I shall return to this kind of idea as a way of getting rid of boundary conditions.

Even if there are no closed time-like curves tbe universe might not

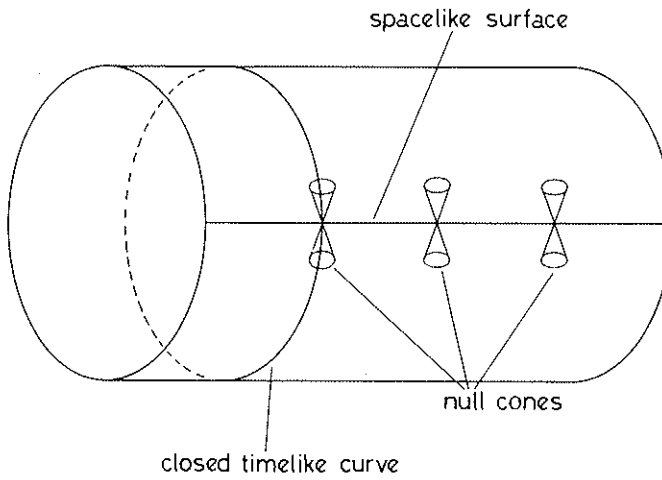


FIG. 1. An illustration of closed or almost closed time-like curves.

be globally hyperbolic, i.e. there might not be any space-like surface such that data on that surface would determine the universe everywhere. A simple example is provided by anti-de Sitter space which is conformal to a time-like strip of Minkowski space. The conformal factor goes to infinity on the edges of the strip which are at infinity. As illustrated in Figure 2, data on a space-like surface determines the solution of the field equations only in a limited region, the Cauchy development, which is bounded by a null surface called the Cauchy horizon. Points beyond the Cauchy horizon have time-like curves through them which can go off to spatial infinity with-

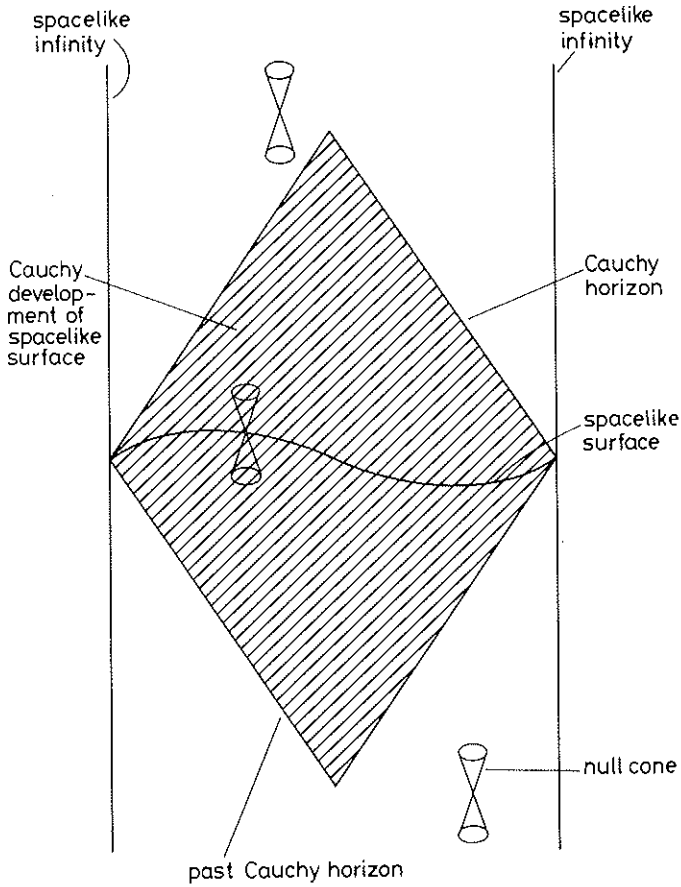


FIG. 2. An illustration of how data on a space-like surface determines the solution of field equations only within the Cauchy development.

out intersecting the space-like surface. They can thus be influenced by data coming in from infinity.

Although there are a number of exact solutions such as anti-de Sitter space, the Reissner-Nordstrom and Kerr solutions which have Cauchy horizons, it seems likely that they are all unstable in the sense that a small perturbation of the initial data on the space-like surface will convert the Cauchy horizon into an unpassable singularity. This leads to the very plausible (though unproved) strong version of tbe Cosmic Censorship Hypothesis: generic initial data on a complete space-like surface produces a Cauchy development which is inextendible, i.e. it cannot be embedded in a larger non-singular spacetime. In other words the Cauchy development is either infinite in extent or it is bounded by singularities. In the case of a solution of the Einstein equations which was a model of our universe, we know from observation that there would be enough matter to cause it to be bounded in the past by a singularity, the Big Bang. It would seem natural to take this initial singularity as a place at which the boundary conditions of the universe were imposed. We would then have to explain why the universe started off:

a) spatially homogeneous and isotropic;

b) in thermal equilibrium;

c) spatially flat, i.e., with the kinetic energy of expansion exactly balanced by the negative gravitational potential energy;

d) notwithstanding *a*), *b*) and *c*), with small initial perturbations which gave rise to galaxies and the other inhomogeneities we observe in the universe today.

Properties *a*) and *b*) might seem sensible on the grounds of simplicity but property *c*) would seem to require extremely fine tuning and we do not have any good explanation for property *d*).

The discussion so far has treated gravity and the causal structure of spacetime that is determined purely classically. However, the prediction of spacetime singularities indicates that the classical theory will break down and that quantum gravitational effects will be important. Some people have suggested that these quantum effects would eliminate all the singularities. There are two possible scenarios:

i) the universe expanded for an infinite time as a steady state (de Sitter) solution and then, for some ill-defined reason, it went unstable and converted itself into the present day observed form;

ii) there was a previous contracting phase followed by a non-singular "bounce".

I do not believe that quantum effects can remove all the singularities for the following reason. As shown in Figure 3 consider a star of a few solar masses containing about 10^{60} baryons which undergoes gravitational collapse to produce a black hole. Because of quantum effects the black hole will create and emit particles like a body with a temperature of about 10^{-7} K. At this low temperature the emission would consist almost entirely of zero rest-mass particles, such as neutrinos, photons and gravitons. These would carry away energy so that the mass of the black hole would slowly decrease. After a very long time the temperature of the black hole would become high enough for baryons and anti-baryons to be emitted in significant numbers. However by then the mass of the black hole would have fallen to about 10^{-20} of the original mass. Thus there would not be sufficient energy left to emit the original number of baryons. The semi-classical approximation used to derive the black hole emission will break down when the black hole gets down to near the Planck mass. We do not know how to treat this stage properly. There seem to be three possibilities:

i) the black hole turns into a naked singularity with a negative mass;

ii) the black hole stops emitting at about the Planck mass and remains as a stable object thereafter;

iii) the black hole disappears completely.

Possibilities i) and ii) would imply that quantum gravity was inherently asymmetric in time. One would expect the formation rate of black holes of mass M shortly after the Big Bang to be of order $M^4 \exp(-4\pi M^2/M_p^2)$ where M_p is the Planck mass. Thus one would expect large numbers of black holes of a few Planck masses. If these all left naked singularities or Planck mass black holes, they would completely dominate the mass density of the universe at the present time and would produce an enormous negative or positive deceleration parameter respectively. I therefore believe that black holes evaporate completely. If this is the case and if most of the original baryons do not reappear, there must be a naked singularity when the black hole finally disappears. This naked singularity will mean that spacetime is not globally hyperbolic. One would have expected this anyway because of quantum fluctuations of the light cone.

If quantum effects caused the universe not to be globally hyperbolic,

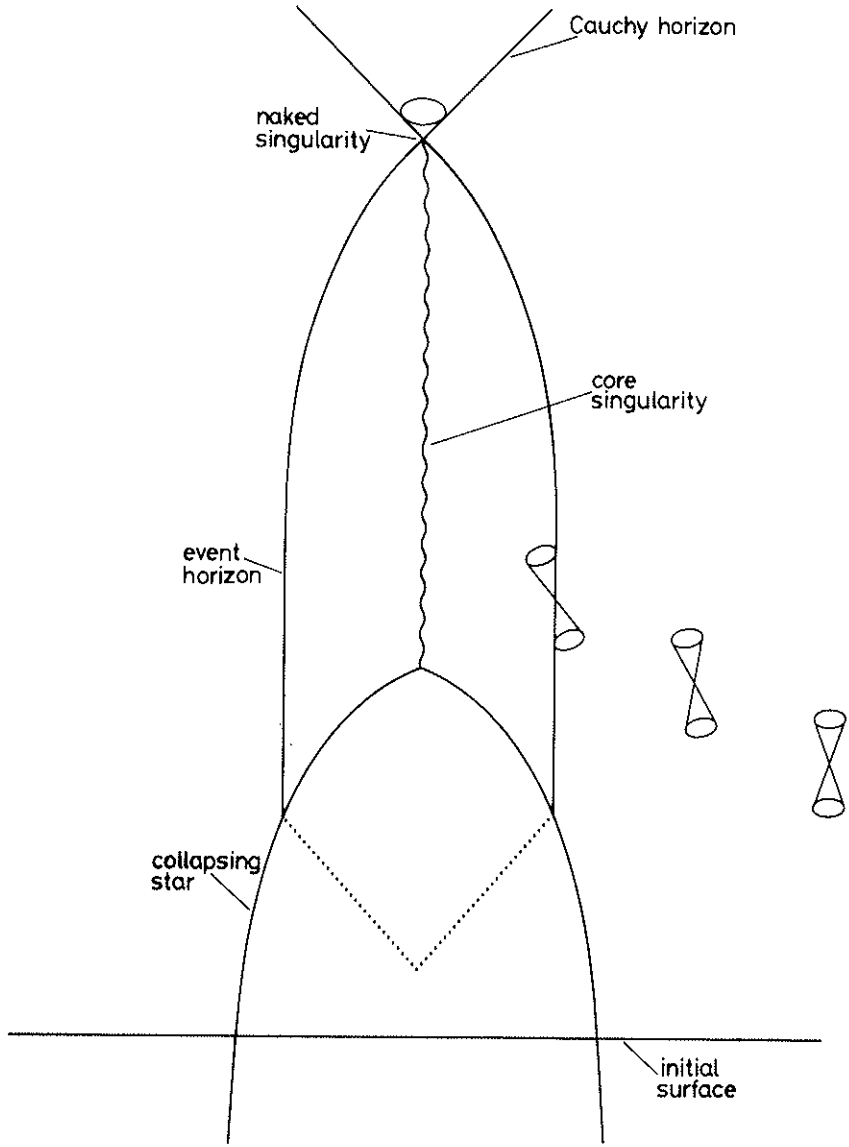


FIG. 3. A star of a few solar masses undergoes gravitational collapse to form a black hole which becomes a naked singularity.

it is obviously not sufficient to impose boundary conditions at the initial Big Bang singularity. It is also difficult to see how one could impose meaningful boundary conditions at a singularity which, by its very definition, is a place where the whole theory breaks down. This problem has been encountered by a number of authors who have tried to calculate particle production near the Big Bang singularity. They assumed some initial "no-particle" state (it could hardly be called a vacuum) and endeavoured to show that the rapidly varying metric would have created pairs of particles with an approximately thermal distribution. However it is not clear whether these "no-particle" initial conditions have any physical significance or, indeed, whether one should try to start off the universe with only gravity and nothing else. This would seem unfair discrimination in favour of gravitons over other particles.

I would like to suggest a different approach to the problem of the boundary conditions of the universe which is based on the so-called Euclidean approach to quantum gravity. In this one evaluates the path integral over all positive definite (Euclidean) metrics with certain boundary conditions. This improves the convergence of the path integral and enables one to handle topologically non-trivial manifolds which cannot really be done if one restricts the path integral to metrics of Lorentzian signature. It also improves the situation with regard to singularities because these will still exist in the complexified spacetime manifold but they can be avoided by the section (Euclidean section) of the manifold on which the metric is real and positive definite. An example of this is again provided by black holes. The partition function Z for the canonical ensemble at a temperature T is defined by a path integral over all metrics which are asymptotically flat in the spatial dimension and which are periodic with a period $\beta = T^{-1}$ in the Euclidean time direction. The Schwarzschild solution is normally given in the form.

$$d s^2 = - \left(1 - \frac{2 M}{r} \right) d t^2 + \left(1 - \frac{2 M}{r} \right)^{-1} d r^2 + r^2 d \Omega^2$$

Putting $\tau = - i t$ one obtains a positive definite metric for $r > 2 M$. The apparent singularity at $r = 2 M$ can be removed by defining a new radial coordinate

$$x = 4 M \left(1 - \frac{2 M}{r} \right)^{1/2} .$$

The metric then becomes

$$d s^2 = x^2 (d \tau/4 M)^2 + (r/2 M)^4 d x^2 + r^2 d \Omega^2$$

The apparent singularity at $x = 0$, $r = 2 M$ is like the singularity at the origin of the polar coordinates provided that $\tau/4M$ is regarded as an angular variable and is identified with a period 2π, i.e. τ is identified with a period $8 \pi M$. The resultant manifold, the Euclidean section of the Schwarzschild solution, is asymptotically flat and non-singular because it does not contain any points with $r < 2M$. The curvature singularity at $r = 0$ does not lie on the Euclidean section. Thus one does not need to specify the boundary values of physical fields at the singularity.

One might ask whether one could find such a singularity free Euclidean section for the Friedmann-Robertson-Walker metrics. The answer is that one can, at least for the spatially closed solutions. For example, the Tolman solution for $p = 1/3 \mu$ is

$$d s^2 = -d t^2 + (a^2 - t^2) d \Omega^2_3 \ ,$$

where $d \Omega^2_3$ is the metric on a three sphere. The quantity 'a' is the radius of the universe at the time of maximum expansion, $t = 0$. The Big Bang is at $t = - a$ and the Big Crunch is at $t = + a$. If one puts $\tau = i t$, one obtains a complete singularity free positive definite metric which, as shown in Figure 4, has two asymptotically Euclidean regions connected by a throat of radius 'a'.

This metric avoids the problem of having to specify boundary conditions at a singularity but one would have to explain why the positive definite metric should be asymptotically Euclidean in two regions and why the throat between them should be so large: the radius of 'a' of the throat is at least 10^{60} Planck lengths. One possible reason might be that this metric is in fact a constant time section of a five dimensional black hole. The entropy of this five dimensional black hole would be proportional to a^3. One might be able to interpret this entropy as the logarithm of the volume in space of all metrics occupied by metrics near the given metric. In this case the most probable value for 'a' would be infinite, i.e., one would have a $k = 0$ parabolic universe. Within the limits of observational accuracy, this is indeed what we have.

This positive definite metric does not, however, explain why the universe should have started off in thermal equilibrium at a high temperature, if indeed it did. To understand this I think that it is essential to consider the large fluctuations of the metric and of the topology of

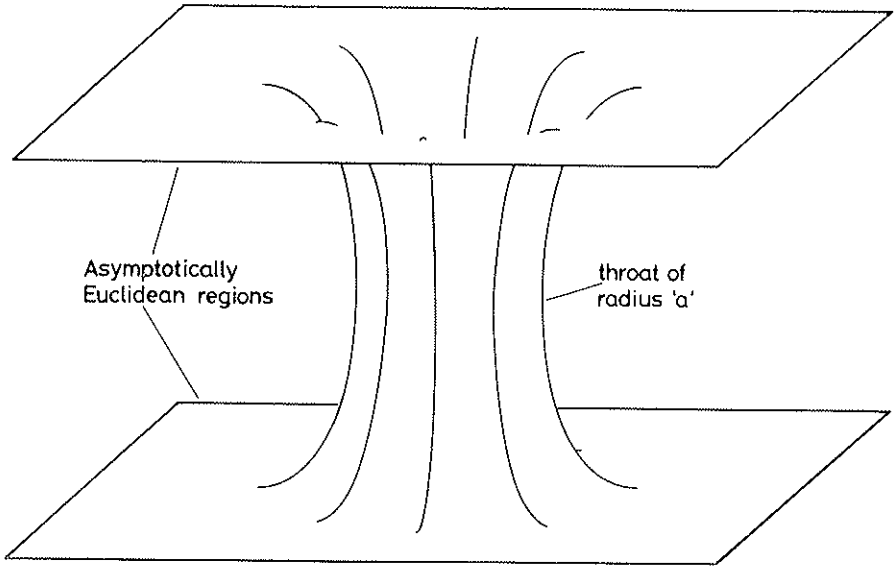


FIG. 4. A singularity-free positive definite metric which has two asymptotically Euclidean regions connected by a throat of radius, *a*.

spacetime that one would expect to occur on the order of the Planck length or less. One might imagine that spacetime would display a "foam-like" or even a fractal structure on these length scales but that it would appear smooth and nearly flat on larger length scales. In order to study this it is helpful to consider the volume canonical ensemble which is defined by a path integral over all compact positive definite metrics with an action that includes a cosmological term which is interpreted as a Lagrange multiplier for the four-volume, just as the inverse temperature is a Lagrange multiplier for the energy in the ordinary canonical ensemble. By evaluating the path integral over compact metrics, one eliminates one of the two parts of physics, the boundary conditions. There ought to be something very special about the boundary conditions of the universe and what can be more special than the condition that there is no boundary.

It seems reasonable to suppose that the path integral is dominated by metrics which are very complicated on Planck length scales but which appear smooth on a larger length scale and that the number of gravitational states with a four-volume V goes up steeply with increasing V. Any non-zero cosmological constant which arose from the vacuum fluctuations of the various fields or from non-zero Higgs potentials could be absorbed

into a shift in the Lagrange multiplier and a slight change in the small scale structure. Thus one might be able to explain both why the apparent cosmological constant is zero or nearly zero and why the universe is so near to a parabolic $k = 0$ model. The argument in the latter case would be that there were indeed states of the gravitational field which corresponded to universes which were only a few Planck lengths across, but that the number of gravitational states which they represented would be much smaller than for large universes. Normally, the value of the cosmological constant is regarded as belonging to the equation-of-motion part of the physics whereas the spatial flatness or otherwise of the universe is supposed to be a boundary condition. The volume canonical ensemble offers a possible way in which they might be related.

The complicated topology of the spacetime foam will mean that spacetime will not have the normal causal properties such as propagation only within the light cone and global hyperbolicity. These acausal effects will not be noticeable at low energies but they will be important near the Big Bang. One can think of the foam as being made up of virtual black holes which appear and disappear in about a Planck time, 10^{-43} seconds. Particles with normal energies will have wavelengths much longer than the Planck length and will have very little chance of falling into these virtual black holes. However, very high energy particles will be able to fall into the virtual black holes and will be re-emitted as particles of other species. This will mean that the universe will necessarily be in thermal equilibrium at the Planck time. The acausal properties of the spacetime foam might explain why the universe started off spatially homogeneous and isotropic even though different regions were not in causal contact according to tbe classical metric. This might also be a way of avoiding the production of too many monopoles, a problem tbat does not seem to have any satisfactory answer in conventional theories.

To summarise, because of quantum fluctuations of the causal structure of spacetime, it is not sufficient to impose boundary conditions at the Big Bang singularity, even if they could be well-defined there. To overcome this difficulty, I would advocate the Euclidean approach to quantum gravity in which the path integral is evaluated over positive definite metrics. If one uses compact metrics, one gets rid of the need to specify boundary conditions for the universe. Hopefully this approach would explain:

1. Why the apparent cosmological constant is zero;
2. Why the universe is spatially flat;
3. Why the universe was in thermal equilibrium.

DISCUSSION

PEEBLES

If the universe were in thermal equilibrium at the Planck time (strict Gibbs distribution ignoring gravitational energy) the thermal fluctuations could be disastrously large.

HAWKING

It seems that the universe was in *local* thermal equilibrium near the Big Bang, but I agree that if you assume that it was in thermal equilibrium over larger scales, you get too much large-scale inhomogeneity now. This is another of the problems that one gets into if one tries to impose a boundary condition at the initial singularity. On the other hand, if one imposes boundary conditions on the Euclidean section, one might expect the universe to be smooth on the Euclidean section. In the case of the $K = +1$ Tolman model, this would say that the universe would be smooth at the time of maximum expansion. The problem then would be to explain why it is slightly inhomogeneous now.

SILK

The most likely form of fluctuations at the quantum gravity epoch may be constant curvature fluctuations. These cannot have a large amplitude because of large-scale anisotropy constraints. Is it possible that, even on sub-horizon scales, black hole formation could be suppressed as a consequence of this?

HAWKING

In order to show that one would expect to form a large number of black holes of a few Planck masses, I assume only that one has thermal fluctuations within the particle horizon. As you and Peebles point out, the fluctuations on scales much larger than the horizon must be less than would be expected for thermal equilibrium.

SCIAMA

I understand that not all curved spaces of indefinite metric can be analytically continued into space with positive definite metric. Are you depending essentially on the fact that it can be done for a Robertson-Walker metric?

HAWKING

It is true that only rather special metrics, like the Schwarzschild and Robertson-Walker metrics, admit a Euclidean section on which the metric is real and positive definite. However, the idea is that a path integral over *all* positive definite metrics is equivalent to a path integral over all Lorentzian metrics. It is similar to what is done in ordinary field theory: one integrates over all fields that are real in Euclidean space and says that it is equivalent to integrating over fields that are real in Minkowski space.

SPONTANEOUS BIRTH OF THE CLOSED UNIVERSE AND THE ANTHROPIC PRINCIPLE

YA. B. ZELDOVICH
Space Research Institute, USSR Academy of Sciences
Moscow

1 - Introduction

The idea of the spontaneous birth of many, even an infinite number, of universes is not new; nor is it new that the actual universe may have been selected because it offered the possibility of the creation of life and consciousness. It appears that this conjecture, called the anthropic principle, was first considered by Idlis (1958). It was further dealt with by Carter (1973), and Wheeler (1974) pointed out that a definite set of physical constants is needed. The subject has been most recently reviewed by Rees (1981).

The general idea of quantum fluctuations of a metric, including the possibility of topology changes and the break-up of some domains, was emphasized by Wheeler (1962, 1968) and further developed by Hawking (1978). But the real possibility of a scientific approach to spontaneous birth of the universe has come about only recently in connection with the De Sitter cosmological solutions.

2 - Exponential expansion of the De Sitter solutions

In the De Sitter solutions the right hand side of the Einstein equation is given by the energy-stress tensor of vacuum polarization produced by the curvature of space-time. In this respect the solutions are self-consistent. Their cosmological significance was considered by Gurovich and Starobinsky (1978) and later in more detail, including the transformation into a plasma-matter phase, by Starobinsky (1980).

We may write down two non-singular De Sitter solutions, the flat one (1) and the closed one (2):

$$ds^2 = c^2\,dt^2 - (\mathrm{H}^{-1}\exp \mathrm{H}t)^2\,[dr^2 + r^2\,(d\,\Theta^2 + \sin^2\Theta\,d\,\varphi^2)] \tag{1}$$

$$ds^2 = c^2\,dt^2 - (\mathrm{H}^{-1}\,\mathrm{ch}\,\mathrm{H}t)^2\,[dr^2 + \sin^2 r\,(d\,\Theta^2 + \sin^2\Theta\,d\,\varphi^2)] \tag{2}$$

They both are unstable for small perturbations (Starobinsky 1979, 1980; Mukhanov and Chibisov 1981). Therefore, it is impossible to extrapolate these De Sitter solution to $t = -\infty$.

There have also been attempts to consider the De Sitter solutions depending on the energy of the Higgs fields in a symmetric state of energy maximum before spontaneous symmetry breaking leads to an energy minimum (see Turner 1981). This seems to me less probable compared with vacuum polarisation.

3 - Quantum spontaneous birth

We may imagine the quantum spontaneous birth of the universe as follows. Using the idea of path integration in superspace (Wheeler 1968, Hawking 1978) we may consider some particular scenario, for example the law of growth, $a\,(t)$, the scale of a three dimensionally closed world. One assumes that a changes from $a = o$ to $a = \mathrm{H}^{-1}$ during a finite time τ. At $a = o$ we have a singularity before which the universe did not exist. At the end the curve $a(t)$ is smoothly connected with the metrics (2), consistent with the Einstein equations.

The amplitude for this process is given by the action integral:

$$d\,\mathrm{A} = \exp(i\,\mathrm{S}) = \exp\left(i\int_0^{\tau}\mathcal{L}\,d\,\mathrm{V}\,dt\right) \tag{3}$$

The process violates tbe equations of general relativity, which means that the action S is not an extreme. The total amplitude:

$$\mathrm{A} = \int \exp\left(i\int^{\tau}\mathcal{L}\,d\,\mathrm{V}\,dt\right)d\omega \tag{4}$$

integrated over all scenarios $(d\omega)$ is an exponentially small quantity; but

the important point is that the universe is closed and, therefore, the integration over space, dV, is possible; otherwise we should obtain identically zero.

On the other hand it must be emphasized that the total energy of a closed universe is identically zero, as are its momentum and charge (Landau and Lifshitz 1973). In this respect the birth of a closed world does not violate the conservation equations. The critical or overcritical total density of all matter is now preferred on astronomical grounds (Peebles 1980).

In connection with the recent discovery of the rest mass of neutrinos it is quite probable that $\Omega > 1$ and that the universe is closed. Its age T in this case is limited by $T < 2/3\, H_0^{-1}$, where H_0 is the Huhble constant today, unless there is a small cosmological constant (Zeldovich and Sunyaev 1980) which could stretch T at given H_0 and Ω.

4 - The maximum expansion

It is important for spontaneous hirth that the minimum radius, $a_{min} = H^{-1}$, in (2) is not much larger than the Planckian length. The time span τ could be chosen to be of the order of the Planckian time. Therefore, the dimensionless action, S, could be of the order of several units but not very large.

The next step consists in an expansion of the universe in accordance with classical general relativity but with vacuum polarization on the right hand side of eq. (2). This expansion is asymptotically exponential, because $\mathrm{ch}Ht \cong \exp(Ht)$ at $Ht \gg 1$. This means that the role of curvature diminishes exponentially.

To switch from the De Sitter solution to the relativistic plasma Friedmann solution is considered as a jump of pressure from $p = -\varepsilon$ to $p = \varepsilon/3$. The radius $a(t)$ and the energy density ε are conserved during the jump.

The equations lead to the conservation of da/dt but also to a jump of d^2a/dt^2 which changes sign. The temperature of the plasma immediately after the jump is not much lower than the Planckian 10^{32} K.

It is easy to show, in accordance with the results of Starobinsky (1980) and Turner (1981), that if the jump occurs at $t = t_1$ when $a(t_1) = H^{-1} \exp(Ht_1)$, then the following expansion goes to the maximum radius (turnover point):

$$a_{max} = a(t_1) \exp(H t_1) = H^{-1} \exp(2 H t_1) \qquad (5)$$

The minimum temperature at this moment is:

$$T_{min} \cong T_{pl} \exp(-H t) \tag{6}$$

It is assumed that the cosmological constant is zero and moreover, as given above, that $p = \varepsilon/3$. In the late phase one has to deal with nonrelativistic matter, $p \ll \varepsilon$, but this does not change the results significantly.

The conclusion is that to explain the very large radius of the universe, $a(t) > c/H_0 = 10^{28}$ cm, and the very low temperature, $T_{min} < 3\ K = 10^{-28}\ T_{pl}$, we need a rather long phase of expansion in the De Sitter state.

5 - Anthropic principle

We assumed that the De Sitter formulae (1) and (2) are exact, but unstable, solutions. To make possible a prolonged expansion of the De Sitter state one needs a small initial perturbation of the quantum-born universe, because perturbations that are too strong lead to the birth of real particles (Starobinsky 1980).

The principle of quantum spontaneous birth dictates the closure of space, but does not exclude the possibility of large initial perturbations, i.e. large departures from the initial metric and its first derivatives from (2). The condition of small perturbations is perhaps explained by the anthropic principle. The strongly perturbed universe is too shortlived.

I feel also that there is a certain arbitrariness and fuzziness in the very concept of spontaneous birth. Does spontaneous birth emerge "out of nothing" or in a space of more dimensions or as a topological separation from an initially given empty Minkowskian space? Can one compare the probability of spontaneous birth of different universes? These unsolved questions and perhaps other questions which are not yet even understood should stimulate further work in this maturing area of research.

My thanks are due to L.P. Grischuk, L.E. Gurevich, G. Marx, V. Mukhanov, A. Poljakov, A. Starobinsky and G. Chibisov.

[Note by the editors: Details of the thoughts presented here by Professor Zeldovich are being prepared for publication in the Letters of the Russian Astronomical Journal and the Monthly Notices of the Royal Astronomical Society].

REFERENCES

Carter, B., 1974, in *Confrontation of Cosmological Theories with Observation* (ed. M.S. Longair), p. 291, Reidel, Dordrecht.

Gurovich, V. Ts. and Starobinsky, A.A., 1978, *Journ. Exp. Theor. Phys.*, **68**.

Hawking, S.W., 1978, *Nucl. Phys. B.*, **138**, 349.

Idlis, G.M., 1958, *Izvestia Astrophys.*, Institute of Kazach. SSR, **7**, 38.

Landau, L.D. and Lifshitz, E.M., 1973, *Field theory*, Moscow.

Mukhanov, V.F. and Chibisov, 1981, *Journ. Exp. Theor. Phys. Letters*, **33**, 549.

Peebles, P.J.E., 1980, *The Large Scale Structure of the Universe*, Princeton, Princeton University Press.

Rees, M., 1981, *Quarterly Journ. RAS*, **22**, 109.

Starobinsky, A.A., 1979, *Journ. Exp. Theor. Phys. Letters*, **22**.

— 1980, *Phys. Lett. B*, **91**, 99.

Turner, M., 1981, *Grand Unification and Fundamental Problems in Cosmology*, Preprint No 81-88, Enrico Fermi Inst., Chicago.

Wheeler, J.A., 1962, *Geometrodynamics*, Academic Press, N.Y.

— 1968, *Einstein Vision*, Springer, Berlin.

— 1974, in *Black Holes, Gravitational Waves and Cosmology* (eds. M.J. Rees, R. Ruffini and J.A. Wheeler), Gordon and Breach, New York and London, Chapter 19.

Zeldovich, Ya. B. and Sunyaev, R.A., 1980, *Astron. Journ. Letters*, **22**.

VII.

COSMOLOGY AND FUNDAMENTAL PHYSICS: CONCLUDING REMARKS

COSMOLOGY AND FUNDAMENTAL PHYSICS
CONCLUDING REMARKS

M.S. LONGAIR
Royal Observatory
Blackford Hill, Edinburgh

1 - Introduction

It is an impossible task to do justice to the level of scientific discussion which we have heard during the last five days. All the lectures have contained so much new and exciting material that the best I can do is to recommend to readers that they read each article with the closest attention and then they will have some idea of the wealth of physical phenomena which we now believe are important in understanding the large-scale structure of the universe. I will therefore restrict my attention to some personal remarks about what I found particularly impressive and how I believe we may be able to make real progress in elucidating many of the fundamental problems discussed.

Before discussing these aspects, let me make a few general observations about what has happened this week. First of all, it must be stated that not everything which could be said about cosmology and fundamental physics bas been said. With only twenty-two participants it is natural that some fields have only been lightly covered and others completely omitted. In the latter category we must include all non-conventional interpretations of cosmological data. Although regrettable, this omission was intentional on the part of the organisers. Our intention was to discuss how far conventional physics can account for the known facts of our observable universe. We must accept that there is a finite chance that this approach is not correct and that the non-conventional picture will prove to be right. In this case, this volume will go down in history as an interesting sidelight on intellectual history. However, if one were to ask the average

astronomer or physicist what approach was most likely to lead to a better understanding of the universe, I am confident that the overwhelming majority would adopt that contained in this volume.

A second striking aspect was the degree of unanimity of the participants on virtually all topics. This was not at all accidental, since it is obvious that there is a very strong Cambridge-Princeton-CalTech atmosphere about the meeting. According to my calculations, all the participants have either worked at these locations for substantial periods or have been regular visitors. Thus, the apparent unanimity is almost certainly deceptive and one must ask what the rest of the world would have said about the above lectures. In certain quarters, I know, there would be little sympathy for some of the ideas but, on the whole, I suspect there would be agreement about the basic ideas.

A third point which struck me was the way in which observational cosmology has become simplified. It is a credo among many cosmologists that the basic facts of cosmology should be simple. I am very reluctant to believe that the way in which the large scale structure of the universe came about depended upon a large number of different complicated physical phenomena. As an example of this simplifying trend, I would cite the picture of galaxies which is emerging. For virtually all purposes, galaxies were simply objects consisting of 10% visible matter and 90% invisible. These are the building blocks which define the large scale structure of the universe. Individual galaxies look very different and of bewildering complexity but there was an underlying trend in the discussions that much of this may well be of secondary importance. I helieve Jerry Ostriker's comment this morning was the culmination of this trend when he remarked that all you really need to know about a galaxy is its mass and all the rest is irrelevant detail. Thus, to a first approximation, all the cosmologist need explain is how objects of different masses formed with 10% visible matter and 90% dark. To anyone but a cosmologist it may be a depressing tbought that everything we know and love in the universe is simply referred to as "dissipation".

I will do three things. First, I will try to assess what facts have been established about the universe. I do this with some trepidation since I recall vividly a previous conference in which I tried to define the facts on another topic and was accused by Jerry Ostriker of being a Mr. Gradgrind. He was the character in Dickens' "Hard Times" whose children were only taught facts, facts, facts. They never saw horses dancing in firelight; they saw only the combustion of carbon and oxygen and so on.

Fortunately, my facts are very exciting and form the basis for the proper scientific study of cosmology. Second, I will try to answer Martin Rees' seven questions and, finally, I will look at future prospects.

2 - Modern Facts of Cosmology

It is not trivial to define precisely what one means by a fact, since they evolve with time. Being realistic, I will take a fact to be what everyone in this room would agree to be correct with, say, 95% confidence at 4.30 p.m. on 2 October 1981.

It is conventional to begin cosmology courses with the basic facts that: (a) the sky is dark at night; (b) the universe of galaxies is expanding more or less uniformly; (c) the universe exhibits a very high degree of isotropy on the largest scales as evidenced by the microwave background radiation. None of these facts are seriously questioned. To these, I suggest we add the following.

(i) *The hot Big-Bang.* There was no serious challenge to the basic hot Big-Bang picture of the history of our universe. The pieces of evidence which provide the strongest support for this picture are the coincidences between the predictions of light element synthesis, ^{3}He, ^{4}He, D, ^{7}Li, and the observed abundances which were discussed in detail by Jean Audouze. This coincidence is so striking that it seems inconceivable that this could occur by chance, although we know of other cases in astronomy in which such arguments can be misleading. It is intriguing that this type of evidence appears to contradict McCrea's law of evolutionary cosmologies that you always have sufficient freedom in evolving cosmologies to patch up anything which appears difficult to explain by reasonable or hypothetical astrophysical processes at various cosmological epochs. Therefore, you can never prove that evolutionary cosmologies are correct. To put it another way, if the abundances of the light elements had turned out to be totally irreconcilable with the predictions of the hot Big-Bang model, we would have tried to patch it up by appeal to suhsequent astrophysical processes. Undoubtedly, this would have been a severe blow to the hot Big-Bang model and would have weakened our confidence that the universe passed through a phase when the temperature exceeded 10^{10} K. I would argue that this coincidence is so striking that it substantiates convincingly the inference from the properties of the microwave background radiation

that the universe went through a hot phase. It would however be unwise to believe that the question of the details of the basic hot Big-Bang are settled. There remains the question of how much additional mass might be present in massive neutrinos.

Nonetheless, the evidence is sufficiently encouraging that cosmologists and physicists can now treat seriously the three basic problems of the hot Big-Bang, namely: (a) the origin of the baryon asymmetry in the universe; (b) the origin of the isotropy of the universe; (c) the origin of the fluctuations from which the fine-scale structure of the universe evolved. Until recently, cosmologists supposed that these properties had to be ascribed to the initial conditions from which the universe evolved — a totally unsatisfactory situation. We now begin to have the glimmering of an understanding of how these features might come about in evolutionary cosmologies at very early epochs. The papers by Steven Weinberg and Stephen Hawking indicated the directions in which progress may be expected in the not-too-distant future in understanding these very fundamental cosmological questions. It would not be appropriate to comment on these very exciting ideas at this stage. We must await the process of natural selection in cosmology to find out whether or not these ideas will survive further advances in elementary particle physics and quantum gravitation.

(ii) *The dark or hidden matter.* It was probably always foolish to hope that all the matter in the universe would make itself readily visible to astronomers. At least we have powerful dynamical methods of measuring how much mass is present in bound or almost bound systems and the general consensus seems to be that from the scales of Sc galaxies to groups and rich clusters of galaxies and probably beyond, most of the matter is unobservable at present other than by its gravitational influence. The discrepancy between the visible and invisible matter is agreed to be about a factor of 10 and possibly greater on the largest scale. Nobody is certain, however, about what form it is likely to take. Martin Rees illustrated this vividly by suggesting that at present the uncertainty in the form of the missing mass is about a factor of 10^{70} in mass, the possibilities ranging from massive neutrinos to massive black holes.

We must treat all these possibilities seriously. At the low mass end the possibility of massive neutrinos has many astrophysical attractions. Specifically, it may well go a long way to resolve the rather serious problems which have arisen in understanding how fluctuations in the mass distribution in the universe leave little trace of their presence at the epoch of

recombination and in reconciling dynamical estimates of Ω with those derived from primordial nucleosynthesis. My own view on this possibility is that it is not really up to the astronomers to provide the answer. It is up to the nuclear physicists to tell us whether the neutrino does indeed have mass of $\sim$ 10 to 100 eV and, if it does, that may settle the question.

If the neutrino does not have the appropriate rest mass, we have to use purely astronomical means to find the nature of the missing mass and this will not be easy. For example, it might consist of ultra-low mass stars or Jupiters. These may be detectable by their infrared emission. If it is in the form of black holes, certain masses may be detectable observationally by accretion phenomena or as gravitational lenses. That the latter might prove to be an effective approach is suggested by the argument that the excess of quasars around bright galaxies can be explained by gravitational lensing of background objects by discrete compact masses in and around galaxies. Incidentally, if this picture of gravitational enhancement of quasar images by stars in nearby galaxies turns out to be correct, it will be a classical example of a supposedly non-conventional phenomenon, advocated by Arp for years, becoming completely respectable and indeed conventional. It will be a remarkable vindication of Arp's approach to observational astronomy.

It is hard to be optimistic about our ability to determine the nature of the dark mass in view of the vast range of possibilities. For example, if the matter were in the form of mini-black holes, which form during phase transitions in the early universe as suggested by Stephen Hawking, they would be very difficult to detect.

In my optimistic moods I believe that the hidden mass is probably more or less distributed like the galaxies. I say this on the basis that the hidden mass ratio does not seem to vary greatly as one goes up in scale from Sc galaxies to the scale of superclusters. This probably makes its detection easier than if it were, say, concentrated in the space between galaxies. It is very improper for cosmologists to seek easy ways out of problems like this but it would be very convenient if the neutrino did turn out to have mass of $\sim$ 30 eV. The results of the forthcoming experiments to attempt to measure the neutrino mass are of the greatest interest cosmologically.

(iii) *The large scale structure of the universe — the great holes.* We now have a much greater appreciation of the way in which matter is distributed in the universe on the largest scale. Professor Oort and Marc

Davis showed convincing evidence that there exist large regions of enhanced galaxy density — superclusters — between which there are large holes which are almost completely void of galaxies. The reality of these holes and the filamentary supercluster structures are of the greatest interest for understanding the sequence of events which took place when galaxies first formed. Jim Peebles described his numerical experiments which try to mimic the filamentary structure seen in his magnificent map of the distribution of galaxies in the Shane-Wirtanen counts of galaxies. His model assumed a smooth covariance function without holes. You will recall his pleasant turn of phrase to the effect that, when he compared his simulations with the real universe, his attempts were like that of the rude apprentice compared with the work of the old master. The apprentice's picture was not as sharp as that of the old master. It was blurred and lacked the fine brush strokes of the original. He suggested that the general success of his simulations indicated that his smooth covariance function was a good model for the distribution of galaxies. My own personal view is that the old master knows his trade better than Jim. If Professor Oort will excuse my referring to him as the old master, I believe he is correct in identifying the filamentary structures, the cells and holes in the distribution of galaxies as real features which are telling us something important about how the largest scale structures formed.

It is, of course, rather difficult to prove the reality of these structures statistically. However, I take reassurance from the fact that it took Jim Peebles a long time and a great deal of hard work to establish the clustering of clusters of galaxies despite the evidence of the eye that they are clustered. I believe that if Jim works even harder on his statistics he will also be able to demonstrate the reality of filaments of galaxies and prove Professor Oort absolutely correct.

The reason for wanting to have an agreed answer to this question is that it is one of the few tests suggested to discriminate between those theories in which the largest scale structures form first and then form superclusters and filaments by collapsing into pancakes and those theories in which the galaxies form first and then develop into higher order clustering systems by hierarchical clustering. I am personally very impressed by the numerical simulations of Zeldovich, Doroshkevich and their colleagues which show convincingly how cell-like and filamentary structures can form in the first of these scenarios, commonly known as the pancake theory. It is not clear how these filamentary structures form in the hierarchical picture in which, almost by definition, fine scale structure tends to get washed out as the

clustering hierarchy grows to larger and larger scales. It is important to establish how critical a test this is of the way in which the large scale structure of the universe forms. If indeed it supports the pancake picture, the implications are profound for observational cosmology. The epoch of galaxy formation must have been relatively recent, $z \leqslant 5$ to 10, and then we may realistically hope to be able to observe directly phenomena associated with the formation of galaxies.

A further footnote to this "fact" is to note how rich Marc Davis is astronomically. His pictures of the three dimensional structure of galaxies in the universe are quite remarkable and suggest many marvellous tests which can be made on the clustering environment of different classes of objects. As yet the statistics are not very large, although definitely impressive, and one can look forward to obtaining answers to questions such as:

(a) Do spirals, ellipticals and S0 galaxies exhibit the same large scale structure or are their clustering properties different?

(b) What is the clustering environment about different classes of active galaxy — for example, Seyfert galaxies, pecular galaxies, radio galaxies, etc?

(c) What is the correlation between gas content and clustering environment?

This is only a short list of the many exciting things which can come out as by-products of Davis' massive survey. It is very important to pursue these studies with much larger bodies of data.

(iv) *Direct evidence for evolution of astrophysical objects over cosmological time-scales.* The evidence for evolution over cosmological time-scales has been convincing for some time for active systems such as quasars and strong radio sources. The presentations of Maarten Schmidt and Harry van der Laan indicated how these studies are now getting down to the details of how various classes of objects evolve with cosmic epoch. The only problem with these studies is that they concern the evolution of rather exotic objects whose astrophysics is not properly understood.

I found particularly intriguing the evidence which is now accumulating for the evolution of the ordinary stellar component of massive galaxies with cosmic epoch. Jim Gunn showed evidence for evolution of the properties of the brightest galaxies in clusters which may be ac-

countable for as a combination of cannibalism and stellar evolution. Harry van der Laan showed evidence that radio galaxies which are probably at redshifts between 0.5 and 1 are much bluer than nearby radio galaxies. In my own work with Simon Lilly, we have shown that the optical-infrared colours of radio galaxies show strong evolution of the stellar population out to redshifts exceeding 1. I regard all these studies as our first glimpse of what is now possible when the full power of modern instruments and techniques is applied to the study of galaxies at epochs significantly earlier than the present. There was *a priori* no reason why such evolutionary effects should be observable; it just happens that they have now been observed and they are bound to give us much deeper insight into the sequence of events which took place between the epochs when galaxies first formed and the present day.

One intriguing item concerns the question of the epoch at which galaxies formed and whether it is related to the apparent cut-off in the distribution of quasars described by Maarten Schmidt. It should be noted that we have known for a long time that the evolution of the comoving space density of quasars and radio galaxies must converge at redshifts $z \sim 3$ to 4 because otherwise the source counts would not decrease as dramatically as they do at low flux densities. However, it has not been possible to say whether the sources are absent beyond this redshift or whether the comoving density of sources simply increases much more slowly into the past.

It may be useful to look at the results described by Maarten Schmidt in a way which is independent of modelling the evolution of the quasar population. The question concerns the interpretation of Osmer's result on the absence of quasars in the redshift range $3.5 < z < 4.7$. Malcolm Smith has presented the result in the following way. There are 7 quasars in the Hoag and Smith survey in the redshift range $2.5 < z < 3.5$ which should be detectable in the Osmer deep survey with $3.5 < z < 4.7$ if such objects indeed exist in that redshift range. If it is assumed that the comoving space density of quasars remains constant between redshifts $z = 2.5$ and $z = 5$, i.e. no evolution over this redshift range, 7 to 9 quasars should be observed in the range $3.5 < z < 4.7$ whereas none are found. Thus, even on this very conservative basis, the decrease in comoving density of quasars at these redshifts must be taken seriously. My own impression is that the radio data are probably trying to tell us the same thing but the evidence is much less direct.

It has been remarked by a number of authors that such a cut-off

might be associated with the epoch when the bulk of the galaxies in the universe first formed and this would be entirely consistent with the pancake picture, in which it is the largest scale structures which collapse first at relatively small redshifts, $z \sim 5$.

3 - Tentative Answers to Some Basic Questions

In this opening lecture, Martin Rees listed seven questions which he hoped would at least be addressed, if not answered, during the last week. I will be so hold as to give tentative answers to these.

Question 1. *What is the value of Hubble's constant to a precision of 25%?*

I was rather surprised when I looked through my notes to find the degree of unanimity among the various estimates. I have summarised my understanding of what was said in Table 1.

Interpreted literally, these estimates tend to favour a value of the Hubble constant in the region of 50 to 60 km s^{-1} Mpc^{-1}. This is a case where I am sure the agreement is only apparent and due to the absence of certain views at this meeting. In particular, the infrared Fisher-Tully method was not properly discussed although Gustav Tammann did express some concerns about the technique. It is only fair to say that probably proponents of higher values of H_0 would have reservations about the way in which the low value was derived. I fear it is only prudent to reserve judgement on the answer to this question.

Table 1 - *Data relevant to the Hubble Constant*

Type I supernovae	Sandage and Tammann	$H_0 = 52 \pm 6$ kms^{-1} Mpc^{-1}
Globular cluster ages	Faber	$T = 16 \pm 3 \times 10^9$ years
	Sandage and Tammann	$T = 17 \pm 2 \times 10^9$ years
Age of the Local Group	Lynden-Bell	$T = 16 \times 10^9$ years
Rhenium/Osmium ages	Audouze	$T = 20 \pm 10 \times 10^9$ years

If $H_0 = 50$ kms^{-1} Mpc^{-1}, $H_0^{-1} = 20 \times 10^9$ years $\geqslant T_{\text{cosmological}}$.

While on the subject of cosmological parameters, it is as well to discuss the value of the deceleration parameter, q_0, and the density parameter, Ω. On the surface, it looks as if one has a strong result from the preferred value of Ω of about 0.03 to 0.1, which comes from the abundances of the light elements: D, ^{3}He, ^{4}He and ^{7}Li. None of the estimates of Ω was greater than 1. However, the trend of the dynamical estimates of Ω was to values in the range 0.1 to 0.7. These came from estimates of the ratios of dark to visible matter in the universe, from various versions of the cosmic virial theorem by Jim Peebles and from the Local Supercluster test described by Marc Davis. Although Gustav Tammann argued for a low value of Ω from the latter tests, there was some feeling that a literal interpretation of the data suggested values closer to $\Omega \sim 0.3$ than to 0.03. Among the ways of resolving this discrepancy is the possibility that the neutrinos have finite rest mass. If they have the canonical value of about 30 eV, they are relativistic when light element production takes place and do not affect the abundances of the elements, provided the baryon density remains low. However, they do influence the dynamics of the universe and it is an intriguing question whether or not it is possible to devise a consistent picture, so that the light elements can be synthesised in their correct proportions and there be sufficient mass in the neutrinos to give consistency with the dynamical estimates of Ω. My impression was that it is indeed possible and it makes the question of the laboratory value of the mass of the neutrino all the more urgent.

Question 2 (i). *Are there preferred scales in the distribution of galaxies?*

A tentative yes can be given to this question in the sense that the covariance function of galaxies appears to become zero or negative on scales $\sim$ 15 to 20 Mpc. In any case there must be a break in the covariance function from the value found on a scale of $1 \leqslant 8$ Mpc. Jim Peebles demonstrated how significant quadrupole anisotropies in the microwave background would become, if indeed there remains little or no correlation between structures on scales greater than about 20 Mpc. Nonetheless, we probably do not know the covariance function with any reliability on scales between 20 to 100 Mpc and this should be the subject of study on large scale deep plates of typical regions of the universe.

Question 2 (ii). *Is there evidence of dissipative non-gravitational effects?*

In my view, this is inevitable in all models of galaxy formation. I

was particularly impressed by Sandra Faber's diagram which showed that all known classes of galaxies now lie within the region of density and temperature space in which dissipation of energy hy radiative processes must have taken place. I do not believe this actually helps discriminate between models of the early evolution of large scale structures, because it must eventually occur in all of them so that galaxies with the properties we know can form. On the scale of clusters and superclusters, however, the question of dissipation was controversial and the arguments which suggest it is important are strongly model dependent. It would be very exciting if we could identify large scale systems in the process of dissipating energy during formation but to my knowledge none of these systems has yet been identified.

Question 3. *On what scales are galaxies tracers of the hidden mass?*

A provisional answer can be given to this question. The fact that the ratio of dark to visible matter is about 15 on all scales from Sc galaxies to giant clusters suggests that the dark matter is associated with visible matter. If the density parameter is low, $\Omega \approx 0.1$, then probably all the dark matter could be simply associated with the distribution of visible matter. If, however, the density parameter were closer to unity, the dark matter would have to be elsewhere, almost certainly in the spaces between clusters where the theoretician is free to endow it with a wide range of exotic properties, many of which would be essentially impossible to exclude by observation. This is just another aspect of the problem of resolving the discrepancy between the dynamical estimates of Ω and the amount of mass in known systems.

Question 4. *What is the relation between quasar and galaxy evolution?*

The basic problem here is that there is no theory of quasars and radio sources which is sufficiently secure that it can he used to predict what the relation between the galaxy and its active nucleus should be. We can all hazard guesses as to how it comes about. The central black hole is fed by gas and, possibly stars, from the surrounding galaxy and grows until it becomes a quasar-like object when its mass is $\sim 10^8\ M_\odot$ and there is a ready supply of fuel. But it is not at all clear how ideas like these can be put into a form which provides quantitative relations between galaxy and

quasar evolution. We need more empirical understanding of quasar, radio galaxy and galactic evolution to be able to put flesh on this question. We have seen bow these studies are now possible out to redshifts $z \simeq 1$ and I am optimistic that we will gain great insights from studies of these objects in this redshift range, rather than by studying those at the largest redshifts.

Question 5. *How can we probe observationally the redshift range* $5 < z < 1000$?

In principle, if there exist discrete objects in this redshift range, we should be able to investigate their properties and physical conditions along the line of sight to them. However, I believe it will be very difficult to find such objects. Clearly, they would have to be highly luminous objects to be observable and the best objects for this purpose, the quasars, show at least a convergence in their distribution, if not a cut-off, at redshifts about 4. In addition, there is the awkward fact that the comoving volume per unit interval of redshift decreases as $z^{-1.5}\,dz$ and so, even if the comoving space density of sources remains constant, the probability of finding large redshift objects decreases.

If we cannot use discrete objects, we have only the integrated properties of these objects to look for, i.e., diffuse emission from a background of sources or emission and absorption features in the background radiation. Distortions of the spectrum of the microwave background might provide some information but at present the evidence suggests that these distortions must be rather small.

I am not at all optimistic about studying this redshift interval on the basis of what we know now. I suspect we need a new discovery to make progress in this area.

Question 6. *How secure are the abundances of the light elements*: ^{4}He, ^{3}He, D, ^{7}Li?

Jean Audouze gave a complete answer to this question. I was particularly impressed by the new results on ^{7}Li and, by and large, the story looks rather convincing. However, there is a real question concerning abundance variations of deuterium. My inspection of his data suggested that many of them are consistent with a single deuterium abundance with one or two points which deviate from this mean value by a significant

amount. This is such an important question that the origin of these variations should receive urgent attention.

Question 7. *Are there comprehensible physical processes which are important for the time interval* $10^{-44} < t < 10^{-36}$ *seconds and which can explain simultaneously the overall isotropy of the universe and its finite degree of inhomogeneity on a fine scale?*

It is one of the most remarkable facts of modern cosmology that this question can now be asked with a straight face and that there are now the first tentative steps towards a real physical explanation of those profound cosmological problems concerning the isotropy of the universe, its baryon asymmetry and the origin of the fluctuations from which galaxies form. I leave it to each individual to decide how seriously he wishes to take these extrapolations of the known laws of physics.

4 - The way forward

I find it revealing to look at the observational basis for some of the new "facts" about the universe. The list shown in Table 2 is by no means complete but I bope it illustrates my point.

You will notice that all of these result from large and systematic studies. In other words, rarely does one make a single simple observation

TABLE 2 - *Some studies described during the Vatican Study Week.*

The covariance function for galaxies	Peebles
The three dimensional distribution of galaxies	Davis, Oort
The luminosity-velocity dispersion relation	Faber
Evolution of giant elliptical galaxies	Gunn
Distances of nearby galaxies	Tammann
Surveys of radio quiet quasars	Schmidt
Radio galaxy identifications	van der Laan, Longair
X-ray counts for quasars and active galaxies	Setti, Woltjer
D, ^{3}He, ^{4}He, ^{7}Li abundances	Audouze

which results in a quantum leap in our understanding. As soon as such a phenomenon is observed, the immediate question is "How common is it among similar objects of its class?" and true understanding only comes from further systematic studies. Perhaps the most striking exception to this rule is the discovery of the microwave background radiation but then I would argue that there is no other member of its class available for observation in cosmology.

Therefore, I firmly believe that the way forward is through the pursuit of yet more difficult systematic work in all the fields we have heard about this week. It may, of course, be that we are lucky and another single great discovery will be made which at a stroke answers some of the profound problems which we have heard about. Realistically speaking, however, such a discovery is most likely to come out of the systematic studies, as for example, happened in the cases of quasars, pulsars, X-ray hinaries, etc.

By and large, my list of top priorities would consist of more of the above. If I had to select those which I think may be particularly profitable, I would suggest: further detailed studies of nearhy galaxies, deep surveys of large areas of sky, the velocities and spectra of large samples of galaxies, the relation of active galaxies to the properties of the stellar and gaseous component of their parent galaxies and to their environments, measurement of fluctuations in the microwave background radiation with high precision. In particular, studies of how the properties of galaxies and the large scale distribution of galaxies and diffuse matter change with cosmic epoch are of particular significance.

Many of these programmes can be accomplished from ground based facilities. With the development of CCD cameras and spectrographs, infrared photometers and arrays of high sensitivity and complementary radio and X-ray studies, the redshift range $0 < z < 1$ is already available for detailed study, provided the bodies which award large telescope time recognise that, to make significant advances in these studies, a large amount of dedicated observing time is needed. This is not encouraged in the present climate where all large telescopes are grossly oversubscribed and it is necessary to make a quick killing to ensure that the next application will have a reasonable chance of success. I hope some of the influential people in this room will help ensure that some of these demanding large scale programmes are carried out in an effective manner.

The main problems in pursuing these programmes are technical. There is an urgent need for efficient CCD spectrographs so that the redshifts of

normal galaxies as faint as V = 22 can be achieved in a sensible observing time. The first effective infrared detector arrays will produce a new picture of the universe, very likely with most of the faint objects being galaxies at large redshifts. It may even be possible to detect those elusive objects, primaeval galaxies, in the near infrared waveband. To complement these infrared pictures, one needs infrared spectrographs. These are only now becoming available but without the sensitivities to undertake faint galaxy work. There is great scope for technological development in this area.

There are already many excellent large ground-based optical and infrared telescopes. The development of optical and infrared telescopes with apertures much greater than 5m is perhaps the most significant future development in ground-based astronomy. Their singular importance is their ability to increase, by a large factor, the number of photons received from faint galaxies and quasars, as compared with the present generation of detectors, and so probe the universe directly to the very faintest magnitudes.

In space astronomy the next major advance in cosmology will be made by the NASA Space Telescope due for launch in early 1985. This will open up the world of galaxies with very high angular resolution out to redshifts of one and greater. The types of astrophysics which we can now undertake at redshifts of 0.1 will be possible for objects at significantly earlier cosmological epochs than the present. The maximum gain will be for point objects and luminous quasars of the type we know could be observed at redshifts as large as 10 or 15, provided that: (a) they emit significant radiation at wavelengtbs $\lambda = 800/(1 + z)$ nm; (b) we know where to look for them; (c) they exist. I fully expect us to obtain definitive evidence about the detailed evolution of galaxies and clusters between redshifts of 1 and the present epoch. I also expect we will be able to see the effects of gravitational condensation in large scale structures between the same epochs. These are only a few of the many exciting cosmological questions which can be addressed and, I hope, answered with some confidence by the Space Telescope.

Beyond that, we can look forward to other exciting space observatories; for instance, COBE, which will perform detailed studies of the isotropy of the microwave background on scales $\Theta \geqslant 7°$, may be launched in 1987. In the 1990's we may look forward to the AXAF X-ray telescope which may be thought of as the X-ray equivalent of the Space Telescope.

5 - Concluding remarks

I repeat my feeling that the threads of the broad picture are beginning to be drawn together into a plausible physical picture for the origin and evolution of the universe. Although many of the issues are complex in their working out, the underlying physical ideas have an attractive simplicity. Not only are the physical ideas basically simple, but the present picture of the observed world of galaxies is becoming simpler.

I emphasised that facts are relative things and that in astronomy they tend to evolve with time. However, at least for a few hours, I have a vision of a comprehensible universe which we can account for by the best of contemporary physics. Our job is to find out whether this is illusory or whether we will build upon this hard-won understanding.

CLOSING STATEMENTS

J. H. Oort

Professor Brück, who is unfortunately unable to attend this meeting, has asked me to thank you on behalf of the Pontifical Academy for having come to Rome and for the way you have contributed by your communications and vivid discussion to the interest and success of this Study Week.

H. van der Laan

Before closing this final session of our Study Week, it is my privilege to express our gratitude and joy for this opportunity, on behalf of all participants.

The threatening divergences of art and science, of philosophy and analysis, of religious motives and ambitious careers are all too real in the lives of most of us. This week has been a great counterforce, an extraordinary experience of simultaneity in scientific endeavour and artistic appreciation, of analytic efforts in the quintessential environment of occidental historic christianity.

We thank Professor Chagas and the Pontifical Academy of Sciences over which he presides for the invitations and the unsurpassed hospitality. We ask you to convey our thanks to Pope John Paul II.

I wish to thank the Academy staff for all their services, large and small, to make this week a success and I single out particularly Father di Rovasenda and Signora Porcelli-Studer for their hard work and attention to detail before and during this week. We thank the technical staff for their help here and the kitchen staff for pleasant coffee breaks and superb lunches.

I thank Father Coyne and his associates, Father Boyle, Father Casanovas and Father McCarthy for their competent activities and their friendship. Their best reward is good manuscripts before 15 November! Please convey our appreciation to Dr. Roncalli for being so perceptive and enthusiastic a guide on our Vatican Museum tour!

Professor Brück, with Professors Longair and Rees, organized the scientific program. You will not be surprised that we all think you have perfect judgement in inviting each one of us. We praise you for your clever wisdom and your constructive labours. I include here the staff of the Royal Observatory, Edinburgh, particularly Miss Susan Hooper, who somehow moved the whole mountain of administrative affairs. Where would we be without dedicated secretaries.

Prof. Brück is unfortunately not here because of a slight indisposition. We send him our warm regards via, last but not least, the incomparable Mrs. Brück, whose charming leadership and constant attentive presence have done so much to make this week one of the most memorable of our lives! We thank you all. And herewith I close this Study Week on Cosmology and Fundamental Physics.