Astrophysics
Recent Progress and Future Possibilities

Invited reviews at a Symposium in honour of
Bengt Strömgren (1908-1987)
held in Copenhagen
May 25-26, 1988

Edited by B. GUSTAFSSON *and* P. E. NISSEN

Matematisk-fysiske Meddelelser **42**:4

Det Kongelige Danske Videnskabernes Selskab
The Royal Danish Academy of Sciences and Letters

Commissioner: Munksgaard · Copenhagen 1990

Printed in Denmark by Special-Trykkeriet Viborg a-s
ISSN 0023-3323 ISBN 87-7304-202-1

Astrophysics
Recent Progress and Future Possibilities

Bengt Strömgren

Preface

The present book contains the proceedings of the meeting "Astrophysics – Recent Progress and Future Possibilities" which was held at NORDITA in Copenhagen on May 25-26, 1988, in honour of Bengt Strömgren.

The meeting was organized by NORDITA, with a Scientific Organizing Committee consisting of B. Gustafsson (chairman), P. E. Nissen, C. J. Pethick and A. Reiz. Members of the Local Organizing Committee were P. Kjærgaard and H. Bergen.

The scientific programme consisted of nine lectures given by invited scholars, internationally leading in their respective fields. The content of the present proceedings is based on these lectures. In addition, a number of scientific results were reported at poster sessions. Former colleagues, students and friends of Bengt Strömgren, as well as a great number of younger Nordic astronomers, participated in the meeting.

We gratefully acknowledge substantial support from the Carlsberg Foundation, the Niels Bohr Institute and the Royal Danish Academy of Sciences and Letters.

The Organizing Committees

Contents

Introduction

Bengt Strömgren was born in 1908 and died in 1987. At an age of 13 he began regular observations with a transit instrument and took part in computations of the orbit of comet Baade. His first publication appeared when he was 17 years old and the last one shortly before his death. Throughout his long scientific life he made fundamental contributions to many different fields in astronomy and astrophysics, his research being characterized by the systematic development of methods and always directed towards important scientific problems, which he solved by combining observational data with physical theories through a careful numerical analysis. His work contributed significantly to making astrophysics an exact and recognized science in this century.

With references to the following bibliography of Bengt Strömgren a brief sketch of his main scientific contributions will be presented here. His first publications concerned celestial mechanics and meridian astronomy, including a pioneering work on photoelectric registrations of star transits [2]. Around 1930 he shifted to the field of astrophysics and on the basis of quantum mechanical calculations of the opacity of stellar matter he showed that hydrogen is the main constituent of stars [9]. In continuation of this work he interpreted the distribution of stars in the Hertzsprung-Russell diagram in terms of stellar evolution caused by the transformation of hydrogen into more complex elements [11] and determined the relative content of helium and hydrogen in stellar interiors [23]. In the late thirties he became interested in interstellar matter and in a classical paper [24] he showed that the transition zone between fully ionized and neutral hydrogen in a gas cloud around a hot star is rather thin, and thus explained the structure of HII-regions. Around 1940 he was the first to construct a realistic model of the solar atmosphere with the newly discovered H^- absorption taken into account [25]. The model was used to determine abundances of sodium, magnesium, potassium and calcium that agreed remarkably well with the composition of meteorites. Later on he also made a pioneering study of the physical state and chemical composition of the cold interstellar gas [31].

From about 1950 Bengt Strömgren's work was centered around a grand investigation of the structure, composition, dynamics and evolution of the Galaxy. The foundation was the *uvby*-β photometric system [59], and the very extensive and accurate photoelectric observations in this system, which he organized and actively took part in for more than 20 years, e.g. [54] and [66]. This work includes determination of ages of field stars [46, 57, 58], computation of orbits and birthplaces of stars [53, 62], three-dimensional mapping of interstellar reddening [64], determination of metal abundances of F-type stars [49, 63], discovery of possible helium abundance differences among young stars and clusters [71], and studies of the kinematics of different populations of stars [51, 74].

In his last paper [77] Bengt Strömgren reviewed the results of the very extensive investigation of F stars within 100 parsec which he carried out in collaboration with younger Danish astronomers. Interesting relations between stellar ages, metal abundances and velocity dispersions were presented. These relations are statistically much more significant than corresponding relations from other investigations and are of fundamental importance for the understanding of the evolution of the Galaxy. Yet it is characteristic of Bengt Strömgren that a large part of the paper is used for a discussion of potential errors and in pointing out a number of areas in which the analysis could be improved.

Bengt Strömgren's influence on 20th century astronomy and astrophysics goes far beyond what is documented in his own publications. He initiated and directed a number of large projects in the fields of astrometry, photoelectric photometry, stellar atmospheres, stellar interiors and galactic dynamics. He very seldom wanted to be a co-author on publications resulting from these investigations and as a result future generations will only be able to trace his influence through the acknowledgements in these publications.

Bengt Strömgren also undertook to serve on many expert commissions and carry out many administrative duties. He made important contributions to worldwide collaboration in astronomy as General Secretary (1948-52) and later President (1970-73) of the International Astronomical Union. During his years in the United States, first as director of the Yerkes and McDonald Observatories (1951-57) and later on as a faculty member at the Institute for Advanced Study at Princeton (1957-67), he played an important role in establishing Kitt Peak National Observatory and in defining NASA's astronomy program. After his return to Denmark in 1967 he became deeply involved in the European Southern Observatory and a very influential president of the ESO Council (1975-77). In his later years Bengt Strömgren substantially promoted collaboration in astrophysics and physics in the Nordic countries as professor and director at NORDITA.

When Bengt Strömgren was approached concerning a meeting on the occasion of his 80th birthday in 1988 he suggested that such a meeting should cover a broad field of contemporary astrophysics and emphasize recent progress and future possibilities rather than past developments. The main purpose of the meeting should be to stimulate the interest of young astronomers and physicists in astrophysical problems. In close consultation with Bengt Strömgren leading scientists were invited to speak on a number of fundamental subjects in astrophysics. In spite of the sad fact that Bengt Strömgren died before the meeting we think that it fulfilled its goals and we hope that these proceedings can serve as further inspiration for new, interesting work in the field of astrophysics.

Bengt Gustafsson *Poul E. Nissen*

Bibliography, Bengt Strömgren

1. Strömgren, B. 1925. Bestimmung der Rektaszensionen von 131 polnahen Sternen, *Kungl. Svenska Vetenskapsakademiens Handlinger* 3. ser. **2,** No. 2.
2. Strömgren, B. 1925. Photoelektrische Registrierung von Sterndurchgängen, *Astronomische Nachrichten* **226,** 81.
3. Strömgren, B. 1927. Tables giving tan(v/2) and $\tan^2$(v/2) in parabolic motion, *Mem. British astr. Assoc.* **27,** part 2, 41.
4. Strömgren, B. 1929. Formeln zur genäherten Störungsrechnung in Bahnelementen, *Kgl. Danske Videnskabernes Selskab, Mat.-Fys. Medd.* **10,** No. 2.
5. Strömgren, B. 1929. Formeln und Tafeln zur Bestimmung parabolischer Bahnen, *Kgl. Danske Videnskabernes Selskab, Mat.-Fys. Medd.* **10,** No. 3.
6. Strömgren, B. 1931. The possible solutions of the equations of fit on the standard model, *Mon. Not. R. astr. Soc.* **91,** 5.
7. Strömgren, B. 1931. The point-source model with coefficient of opacity $k=k_l \varrho T^{-3.5}$, *Z. f. Astrophys.* **2,** 345.
8. Strömgren, E. and Strömgren, B. 1931. *Lærebog i astronomi*, Gyldendal Norsk Forlag, Oslo.
9. Strömgren, B. 1932. The opacity of stellar matter and the hydrogen content of the stars, *Z. f. Astrophys.* **4,** 118.
10. Strömgren, E. and Strömgren, B. 1932. *Lehrbuch der Astronomie,* Springer, Berlin.
11. Strömgren, B. 1933. On the interpretation of the Hertzsprung-Russell diagram, *Z. f. Astrophys.* **7,** 222.
12. Strömgren, B. 1933. Photoelektrische Registrierung von Sterndurchgängen, *Vierteljahrsschrift der Astronom. Ges.* **68,** 365.
13. Strömgren, B. 1934. Tables and diagrams for dissecting a frequency curve into components by the half-invariant method, *Skandinavisk Aktuarietidsskrift 1934, p. 7.*
14. Strömgren, B. 1935. The influence of electron captures on the contours of Fraunhofer lines, *Z. f. Astrophys.* **10,** 237.
15. Strömgren, B. 1935. Das Schmidtsche Spiegelteleskop, *Vierteljahrsschrift der Astronom. Ges.* **70,** No. 1, 65.
16. Strömgren, B. 1936. Thermodynamik der Sterne und Pulsationstheorie, *Handbuch der Astrophysik* **7,** 21.
17. Strömgren, B. 1936. Die Ionisation in den Atmosphären der Himmelkörper, *Handbuch der Astrophysik* **7,** 203.
18. Strömgren, B. 1937. Aufgaben und Probleme der Astrophotometrie, *Handbuch der Experimentalphysik* **26,** 321.
19. Strömgren, B. 1937. Objektive photometrische Methoden, *Handbuch der Experimentalphysik* **26,** 797.
20. Strömgren, B. 1937. The boundary-value problem of the theory of stellar absorption lines, *Astrophys. J.* **86,** 1.
21. Kuiper, G.P., Struve, O., Strömgren, B. 1937. The interpretation of ε Aurigae, *Astrophys. J.* **86,** 570.
22. Strömgren, B. 1937. Die Theorie des Sterninnern und die Entwicklung der Sterne, *Ergebnisse d. exakten Naturw.* **16,** 465.
23. Strömgren, B. 1938. On the helium and hydrogen content of the interior of the stars, *Astrophys. J.* **87,** 520.
24. Strömgren, B. 1939. The physical state of interstellar hydrogen, *Astrophys. J.* **89,** 526.
25. Strömgren, B. 1940. On the chemical composition of the solar atmosphere, *Festschrift für Elis Strömgren,* Ejnar Munksgaard, Copenhagen, p. 218.
26. Strömgren, B. 1944. Tables of model stellar atmospheres, *Kgl. Danske Videnskabernes Selskab, Mat.-Fys. Medd.* **21,** 3.
27. Strömgren, B. 1945. Optical sine-tables, *Geod. Inst. Skrifter,* 3. Ser., **5.**

28. Strömgren, B. 1945. An aplanatic anastigmatic lens system suitable for astrograph objectives, *Kgl. Danske Videnskabernes Selskab, Mat.-Fys. Medd.* **23,** 9.
29. Strömgren, E. and Strömgren, B. 1945. *Lærebog i Astronomi, 2. rev. ed.,* Gyldendal Norsk Forlag, Oslo.
30. Strömgren, B. 1946. On certain mathematical problems connected with the determination of anthropometrical and diagnostical indices, *Acta Psychiatrica et Neurologica* **21,** 747.
31. Strömgren, B. 1949. On the density distribution and chemical composition of the interstellar gas, *Astrophys. J.* **108,** 242.
32. Strömgren, B. 1950. On the extent of the convections zones in the solar interior, *Matematisk Tidsskrift, B,* København, p. 96.
33. Strömgren, B. 1951. Problems of interstellar gas clouds. In *Problems of Cosmical Aerodynamics,* Central Air Documents Office, Dayton, Ohio, chapter 2.
34. Strömgren, B. 1951. On the development of astrophysics during the last half century. In *Astrophysics,* editor J. A. Hynek, McGraw-Hill, New York, chapter 1.
35. Strömgren, B. 1951. The growth of our knowledge of the physics of the stars. In *Astrophysics,* editor J. A. Hynek, McGraw-Hill, New York, chapter 5.
36. Strömgren, B. 1952. Evolution of stars, *Astron. J.* **57,** 65.
37. Strömgren, B. 1953. The sun as a star, *The Solar System* **1,** 36.
38. Morgan, W.W., Strömgren, B., Johnson, H.M. 1954. New features of some emission regions in The Milky Way, *Astron. J.* **59,** 188.
39. Strömgren, B. 1954. Dust and the origin and evolution of stars, *Mem. Soc. Roy. Sci. Liège* **15,** 615.
40. Strömgren, B. and Gyldenkerne, K. 1955. Spectral classification of G and K stars through photoelectric photometry with interference filters, *Astrophys. J.* **121,** 43.
41. Morgan, W.W., Strömgren, B., Johnson, H.M. 1955. A description of certain galactic nebulosities, *Astrophys. J.* **121,** 611.
42. Strömgren, B. 1956. The Hertzsprung-Russell diagram, *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* **3,** 49.
43. Strömgren, B. 1956. Two-dimensional spectral classification of F stars through photoelectric photometry with interference filters, *Vistas in Astronomy* **2,** 1336.
44. Strömgren, B. 1958. Composition differences between stellar populations, *Semaine d'Etude sur le Problème des Populations Stellaires,* publié par l'Académie Pontificale des Sciences, *Scripta varia,* No. 16, 245.
45. Strömgren, B. 1958. Spectrophotometric classification of the population groups, *Semaine d'Etude sur le Problème des Populations Stellaires,* publié par l'Académie Pontificale des Sciences, *Scripta varia,* No. 16, 385.
46. Strömgren, B. 1958. The composition of stars and their ages (The Halley Lecture for 1958), *The Observatory* **78,** 137.
47. Strömgren, B. 1962. Past distribution of the interstellar gas. In *The Distribution and Motion of Interstellar Matter in Galaxies,* editor L. Woltjer, W. A. Benjamin Inc., New York, p. 274.
48. Strömgren, B. 1963. Quantitative classification methods, *Stars and Stellar Systems* **3,** 123.
49. Strömgren, B. 1963. Problems of internal constitution and kinematics of main-sequence stars (George Darwin Lecture), *Q. Jl. R. astr. Soc.* **4,** 8.
50. Strömgren, B. 1964. Computations in astrophysics, *Proceedings of the IBM Scientific Computing Symposium on Large-Scale Problems in Physics,* p. 219.
51. Strömgren, B. 1964. On the chemical composition and kinematics of disc high-velocity stars of the main sequence, *Astrophysica Norvegica* **9,** 333.
52. Strömgren, B. 1964. Comparison of observed and theoretically calculated intensities in the continuous spectra of main-sequence B stars, *Reviews of Modern Physics* **36,** 532.
53. Contopoulos, G. and Strömgren, B. 1965. *Tables of plane galactic orbits,* Institute for Space Studies, NASA, New York.

54. Strömgren, B. and Perry, C. 1965. *Phoelectric uvby Photometry for 1217 Stars Brighter than V=6.5, mostly of Spectral Classes A, F and G,* Institute for Advanced Study, Princeton, New Jersey.
55. Strömgren, B. 1965. Stellar models for main-sequence stars and subdwarfs, *Stars and Stellar Systems* **8,** 297.
56. Crawford, D.L. and Strömgren, B. 1966. Comparison of the Hyades, Coma and Pleiades clusters based on photoelectric u, v, b, y and Hβ photometry, *Vistas in Astronomy* **8,** 149.
57. Strömgren, B. 1966. Age determination for main-sequence B, A, and F stars. In *Stellar Evolution,* Editors R.F. Stein and A.G.W. Cameron, Plenum Press, p. 391.
58. Kelsall, T. and Strömgren, B. 1966. Calibration of the Hertzsprung-Russell diagram in terms of age and mass for main-sequence B and A stars, *Vistas in Astronomy* **8,** 159.
59. Strömgren, B. 1966. Spectral classification through photoelectric narrow-band photometry, *Ann. Rev. Astron. and Astrophys.* **4,** 433.
60. Strömgren, B. 1967. The helium-hydrogen ratio for population I stars deduced from binary data, *Modern Astrophysics,* Paris, p. 186.
61. Strömgren, B. 1967. Results of *uvby* and Hβ photometry of B and A stars and their use in studies of peculiar A stars. In *»The Magnetic and Related Stars«,* editor R.C. Cameron, Mono Book Corp., Baltimore, p. 461.
62. Strömgren, B. 1967. Places of formation of young and moderately young stars, *Proceedings IAU Sympositum No. 31 »Radio Astronomy and the Galactic System«,* editor H. van Woerden, Academic Press, London, p. 323.
63. Strömgren, B. 1969. Quantitative Spektralklassifikation und ihre Anwendung auf Probleme der Entwicklung der Sterne und der Milchstrasse (Karl-Schwarzschild-Vorlesung), *Mitteilungen der Astronomischen Gesellschaft,* No. 27, p. 15.
64. Strömgren, B. 1972. Interstellar reddening within 200 pc of the sun, *Q. Jl. R. astr. Soc.* **13,** 153.
65. Strömgren, B. 1972. The rise of Astrophysics, *Annals of the New York Academy of Sciences* **198,** 245.
66. Grønbech, B., Olsen, E.H. and Strömgren, B. 1976. Standard stars for *uvby* photoelectric photometry south of declination +10°, *Astron. Astrophys. Suppl.Ser.* **26,** 155.
67. Strömgren, B. 1978. Possible use of electronographically recorded low-dispersion slitless spectra in optical quasar surveys, *Physica Scripta* **17,** 339.
68. Crawford, D.L., Mavridis, L.N. and Strömgren, B. 1979. Results of a search for population II stars with V between 14^m and 16^m in high galactic latitudes, *Abhandlungen aus der Hamburger Sternwarte,* Band X, Heft 2, 82.
69. Strömgren, B. 1980. Astrophysics, *Science and Future Choice, Proc. of the Twentieth Anniversary Commemoration Conference, 11-13 April 1978,* editor Philip W. Hemily, Nato Science Committee, Vol. I, p. 156.
70. Strömgren, B. 1980. Oort's Scientific Importance on a World-Wide Scale. In *Oort and the Universe,* editors H. van Woerden, W.N. Brouw, and H.C. van de Hulst, Reidel, Dordrecht.
71. Strömgren, B., Olsen, E.H. and Gustafsson, B. 1982. Evidence of helium abundance differences between the Hyades stars and field stars, and between Hyades stars and Coma cluster stars, *Publ. Astron. Soc. Pac.* **94,** 5.
72. Strömgren, B. 1983. Scientists I have known and some astronomical problems I have met, *Ann. Rev. Astron. Astrophys.* **21,** 1.
73. Strömgren, B. 1984. L'Evolution des Galaxies dans l'Univers en Expansion, *Proceedings of the Symposium on »Science and the Modern World«, Part III,* Plenary Session, Nov. 11-13, 1979, Pontificia Academia Scientiarum.
74. Strömgren, B. 1984. Investigations of the properties of stars of Population II in our Galaxy based on *uvby* photometry, *Proceedings of the Nordic Astronomy Meeting,* September 3-5, 1984, editor K.J. Donner, Observatory and Astrophysics Laboratory, University of Helsinki, Finland, p. 7.

75. Strömgren, B. 1985. Star Counts, Local Density and K_z Force, *Proceedings I.A.U. Symposium No. 106 »The Milky Way Galaxy«,* editors H. van Woerden et al., Reidel, Dordrecht, p. 153.

76. Strömgren, B. 1986. Niels Bohr and the Royal Danish Academy of Sciences and Letters, *The Lesson of Quantum Theory,* Proceedings of Niels Bohr centenary symposium, editors J. de Boer, E. Dal and O. Ulfbeck, Elsevier Science Publishers B.V., p. 3.

77. Strömgren, B. 1987. An investigation of the relations between age, chemical composition and parameters of velocity distribution based on *uvby*β photometry of F stars within 100 parsec, *Proceedings of the NATO Advanced Study Institute »The Galaxy«,* editors G. Gilmore and R. Carswell, Reidel, Dordrecht, p. 229.

The internal structure of late-type main-sequence stars

By D. O. Gough
Institute of Astronomy and Department of Applied Mathematics
and Theoretical Physics, University of Cambridge
Cambridge, CB3 OHA, U.K.

Abstract

Homology scaling laws for main-sequence stars are derived, and used to estimate how late-type stars evolve during their core hydrogen-burning phase. Though not exactly representative of realistic stellar models, the scaling laws do provide a useful method of making estimates of small perturbations either to the initial conditions or to the physics used in the so-called standard theory of stellar evolution. In particular, evolution with varying gravitational constant and varying mass are considered explicitly. The scaling laws are used to determine how gross observable parameters such as luminosity, neutrino flux or acoustic oscillation frequencies depend on the mass, composition and age of the star. By inverting the relations it is shown that with the precision of the best measurements of luminosity, effective temperature and the heavy-element to hydrogen abundance ratio, supplemented with a knowledge of the principle parameters characterizing the high-order acoustic oscillation spectrum, theoretical models could be calibrated to determine mass and age to within about 20 per cent. No useful information about the helium abundance can be obtained in this way.

Introduction

Much of Bengt Strömgren's work was concerned with or motivated by the desire to understand the nature of stars. In his early days he invested a substantial effort in modelling the internal structure, as also did Eddington and Milne at that time. The critical dependence of stellar structure on molecular weight and opacity implied that the position of a star on the Hertzsprung-Russell diagram depended not only on its mass but also on chemical composition, and therefore potentially afforded a means of inferring the proportion of hydrogen to heavy elements in stars. In 1937 the work by von Weizsäcker on nuclear transmutation made it evident that helium was the principal product of thermonuclear reactions in stars, and immediately Strömgren considered the implications of helium being a major constituent of stellar material. Now there was an additional important parameter to determine, and the problem of inferring the nature of a star from observation became richer, and correspondingly more difficult. The central question was then: What are the relative abundances X, Y and Z of hydrogen, helium and heavy elements in the interior of a star? And subsequently it was asked: Can we distinguish observationally between stars of different ages? These are amongst the questions I shall be adressing again in this lecture.

Most of the early work on stellar structure assumed that chemical composition was uniform throughout the interior of a star. This permitted the extensive use of homology scaling laws to compare the properties of one star with another, at least on the main sequence where issues of dense highly degenerate cores and extended giant envelopes do not arise. The sun occupied a central position in the investigations, providing an accurately determined standard with which to calibrate theory. Subsequently it became evident that stars are not chemically homogeneous, though it was not until the advent of electronic computers that it was possible to calculate, or perhaps I should say estimate, the material distribution in evolved stars. Now it is possible to carry out quite detailed numerical computations, and compare the outcome with a wealth of observational data.

So where does that leave the methods used by Strömgren and his contemporaries in those pioneering days? They are not outdated. Even though computers have taken over the role of constructing detailed models, simple scaling arguments, even under circumstances in which the conditions that justify them are not strictly satisfied, are extremely important aids to rationalizing the results of numerical computations, and so to increasing our understanding of the complicated balance of processes that determine stellar structure. By representing the results in rough analytical terms, people like me who are unskilled at interpreting vast arrays of precise numerical computer output can appreciate more readily what are likely to be the most important factors determining the observable properties of stars. In this lecture I shall illustrate this by discussing what is perhaps the most basic aspect of the subject, namely the hydrogen-burning main-sequence phase. I shall keep the discussion as simple as possible, ignoring unnecessary complications without justifying why they are unnecessary: unlike the pioneers in the days before electronic computing, I can always consult numerical solutions of more complicated and hopefully more realistic theoretical models to be reassured that my approximations do not distort the picture too severely.

One might well ask what the purpose of such an exercise is. Surely the main sequence is so well understood that there can be hardly any more to say at so elementary a level. To be sure we have the solar neutrino problem, but after so many thousands of hours of computer time have been dedicated to the unsuccessful search for but one theoretical model of the sun that is not in conflict with observation, rough analytical estimates can hardly be of any real value. That is certainly the view held by many workers in the field. However, I do not support it for the following reasons. First, by thinking in very simple terms one is forced to step back from the morass of detail that is present in the modern computer programmes, and perhaps then one can see more clearly what might be deficient in the theory. Second, with a simple picture in mind one can predict the results of new computations; this is important because it is extremely useful to know the answer to a problem in advance when trying to judge whether the inevitable errors that creep into new modifications to a computer programme have been eradicated. Finally, and most important of all, when one can really find no errors remaining and when the numerical results persist in disagreeing with expectation, one is forced to modify one's simple picture: that is what constitutes real learning.

I am not claiming that one can necessarily use simple arguments to make precise absolute comparisons with observation. One needs an accurately computed detailed model for that. But what one can do is to enquire how that model is modified when certain parameters are changed, or when certain physical phenomena are modified or even introduced into consideration, provided the modification is not too large. One is essentially carrying out approximate perturbation theory.

I must point out also that there is a new reason for rediscussing main-sequence evolution: it is provided by the body of new seismic data that have recently been gathered from solar observations, and some similar data that we anticipate will be obtained in the near future from other stars. These data provide additional and different constraints from those imposed by the bulk parameters that have been obtained by classical astronomical techniques, such as mass and position on the Hertzsprung-Russell diagram. In confronting theory with them it is again productive initially to think in very simple terms.

2. *Simple main-sequence evolution*

What I mean by 'simple' evolution is the theoretical study of spherically symmetrical stellar models whose temporal variation on the main sequence (possibly after an initial transient associated with the approach to the main sequence) is determined solely by the gradual nuclear transmutation of hydrogen into helium; in this description the star's mass is constant, and there is no transport of chemical species through the star except in convection zones.

The subject has been studied extensively. In particular, the theory has been applied to the mass-luminosity relation and the position of stars on the Hertzsprung-Russell diagram. The sensitivity of the results to chemical composition, and to uncertainties in the theoretical description of energy transport in the convection zones, has also been investigated.

As a result of the discrepancy between the observed and theoretical values of the solar neutrino flux, the theory has been quite highly refined. The microphysics especially has been reassessed, to provide a more secure basis for the procedures by which the nuclear reaction rates, the equation of state and the opacity are determined. The sensitivity of solar models computed in this simple way to the obvious uncertainties in the microphysics has been extensively investigated, and summarized recently in two important papers by Bahcall *et al.* (1982) and Bahcall and Ulrich (1988), which provide a useful basis for comparing other possibly more realistic models. Indeed, solar models computed in this simple way are now commonly called 'standard', even though the details are continually being modified. In the attempt to resolve the neutrino problem, uncertain parameters have understandably sometimes been set to extremes of plausibility.

Before proceeding into any detail it is useful to list some of the more obvious features of the so-called standard models:

(i) hydrostatic balance and thermal balance,

(ii) no rotation nor dynamically significant magnetic field, and hence spherical symmetry,

(iii) no accretion nor mass loss,

(iv) no macroscopic meridional motion, other than small-scale turbulence in the unstably stratified convection zone; therefore, in particular, no mixing of entropy from convection zones into radiative regions, and no mixing of the products of nuclear reactions in lower main-sequence stars that do not have convective cores, and therefore

(v) no wave transport of energy or momentum.

It is also assumed that the generally accepted laws of physics are valid. Thus, for example, with respect to appropriate units of mass, length and time in which Planck's constant h and the speed of light c are constant:

(vi) G is constant,

where G is the gravitational constant.

These features, which are written into the theory, are essentially assumptions, though they are not wholly unjustified. We note that most main-sequence stars do not appear to vary substantially on a dynamical timescale, and therefore that hydrostatic balance must be a very good first approximation. Moreover, since for all but perhaps the most massive stars the characteristic nuclear transmutation time substantially exceeds the thermal diffusion time, most main-sequence stars have had time to achieve thermal balance. Therefore they are presumably in thermal balance, unless some instability has recently upset it. Studies of thermal stability, particularly of the sun, generally provide little cause for doubt. Although the sun is observed to rotate (and spectrum line-width measurements of other stars suggest that the solar rotation is not grossly atypical of stars in its spectral class) the centrifugal force is extremely small compared with gravity. This is consistent with the figure of the sun having been observed to differ from being spherical by no more than about 1 part in 10^6. Early-type stars rotate more than 100 times faster, but except in extreme cases the neglect of centrifugal force in the hydrostatic equation is probably not a serious flaw in the models. For most main-sequence stars there is little evidence of substantial accretion or mass loss.

It is more difficult to justify the remaining assumptions. Although centrifugal force is unimportant to the hydrostatic balance in the radial direction, it is potentially important horizontally. Indeed, in a uniformly rotating star (or a star in which angular velocity Ω is instantaneously a function only of distance from the rotation axis) the rotational terms in the momentum equation can be represented as the gradient of a potential which can be added to the gravitational potential, thereby changing the meaning of horizontal; von Zeipel (1924) pointed out that except under very contrived circumstances the pressure cannot be made constant on surfaces of constant total potential, which Eddington (1925) and Vogt (1925) realised would lead to possibly significant circulatory motion. Basically, the reason is that thermal diffusion tends to make surfaces of constant temperature, and consequently surfaces of constant pressure, more nearly spherical than what is required for hydrostatic balance. If Ω is not a function of distance from the axis alone, the advection terms cannot even be derived from a potential; then they can never be balanced by a pressure gradient however contrived, and motion must necessarily ensue.

This conclusion might be invalidated if Lorentz forces associated with an internal magnetic field were taken into account. A general study of rotating magnetic equilibrium configurations of realistic stellar models has not been undertaken, but it seems likely that any that might exist would be unstable (Pitts and Tayler, 1985).

The first serious attempt to calculate the flow resulting from rotational imbalance in an isolated star was carried out by Sweet (1950). In his calculation, as in most that have followed, Ω was assumed known, and the influence upon it of advection of angular momentum by the circulation was not taken into account. A self-consistent steady solution of the equations governing a rotating nondegenerate nonmagnetic star has never been found; any that might exist is likely to be unstable (*e.g.* Gough, 1976; Tassoul, 1978; Zahn, 1989). Nevertheless, it appears that the characteristic circulation time is likely to be of order of the thermal diffusion time multiplied by the ratio of the gravitational to the centrifugal acceleration. It is called the Eddington-Sweet time, and for sun-like stars is about 100 times the characteristic nuclear evolution time, assuming the surface angular velocity to be characteristic of the interior rotation.

I appear to have digressed quite a long way from my simple picture of a star, and I have done so quite deliberately in order to draw attention to one of its possible deficiencies: although a rotationally driven circulation is too slow to be dynamically important, it could have a marked effect on the chemical evolution of the star by transporting the products of the nuclear reactions away from the site of their creation. An Eddington-Sweet time 100 times the nuclear time may at first seem too long to be significant for the structure of the sun, but a little thought makes one realise that that might not be so. In a subject with great uncertainty it is not difficult to erode confidence in a factor of 100, particularly when it is appreciated that the circulation rate is proportional to the square of the angular velocity. First, we know that the sun was rotating more rapidly in the past than it is now (the solar wind today is removing angular momentum on a timescale comparable with, though apparently somewhat greater than the age of the sun). This is consistent with the observation that the rotation rates of stars in young clusters, notably the Hyades and the Pleiades, are substantially greater than those of older but otherwise similar stars. So perhaps rotationally induced material mixing has significant consequences early in main-sequence evolution. Secondly, the scant seismological evidence that concerns the solar core indicates that even today the core might be rotating perhaps three times faster than the surface (Duvall *et al.*, 1984; Gough, 1985) though little confidence can yet be given to what at present is no more than a slight hint. Thirdly, we know that steep gradients of angular velocity can enhance the circulation rate, and a relatively rapid rotation of only the core means that substantial gradients might exist. And finally, the conclusion that a steady rotating star cannot be stable implies that the motion must actually vary with time. On what time scale we do not know, though one is tempted to ponder over the stellar cycle time, 22 years in the case of the sun, as a candidate. What are the consequences?

Transport of momentum and energy by waves is commonly thought to be negligible. Most nonadiabatic linear studies of g modes exicted by nuclear reactions in the core of the sun have found instability at some epoch on the main sequence. One might anticipate that these modes would have the capacity to redistribute not only momentum and energy, but also the helium produced by the nuclear reactions. However, it must be appreciated that the uncertain interaction between the modes and other forms of motion, such as convection, leaves considerable room for doubt. So perhaps in reality all the modes are stable. Moreover, Dziembowski (1983) has argued that even if the modes were excited, their nonlinear development would be so severely curtailed by resonant coupling to stable modes that their ability to induce substantial transport of material in the core would be negligible. However, it is not wholly out of the question that their influence on the distribution of angular momentum throughout the star is not insignificant.

Notwithstanding this list of concerns, it is very likely that the broad picture provided by the simple models is basically correct. Therefore without doubt it is extremely useful to study this picture, provided a healthy scepticism of the fine details is maintained.

3. *Simplified equations of stellar structure*

Assuming the star to be static and spherically symmetric, the equations of stellar structure may be written,

$$\frac{dp}{dr} = -\frac{Gm\rho}{r^2}, \tag{3.1}$$

$$\frac{dm}{dr} = 4\pi r^2 \rho, \tag{3.2}$$

$$\frac{dT}{dr} = -F, \tag{3.3}$$

$$\frac{dL}{dr} = 4\pi r^2 \rho \varepsilon, \tag{3.4}$$

where r is a radial coordinate, p, ϱ, and T are pressure, density, temperature, and the intermediate variable m is the mass enclosed within the sphere $r =$ constant; in addition $L(r)$ is the internal luminosity (the integral of the energy flux over the surface of the sphere $r =$ constant) and ε is the rate of generation of thermal energy per unit mass by nuclear reactions. If the star were slowly expanding or contracting, on a timescale much greater than the characteristic dynamical timescale of the star, the hydrostatic equation (3.1) would remain valid but a work term would need to be added to the energy conservation equation (3.4). The function F in the energy transport equation (3.3) depends on whether the radius at which it is defined is in a radiative or a convective region. In an optically dense radiative region

$$F = \frac{3\kappa\rho L}{16\pi r^2 acT^3}, \tag{3.5}$$

where $\varkappa$ is the Rosseland mean opacity, a is the radiation density constant and c is the speed of light. In convective cores and throughout most of the convective envelopes of dwarf stars the stratification is essentially adiabatic, and

$$F \simeq -T\left(\frac{\partial \ln T}{\partial \ln p}\right)_s \frac{d\ln p}{dr} = \frac{Gm\rho T}{r^2 p}\left(\frac{\partial \ln T}{\partial \ln p}\right)_s, \tag{3.6}$$

the partial thermodynamic derivative being taken at constant specific entropy s.

Equation (3.6) is invalid in the nonadiabatic boundary layers at the edges of convection zones. The boundary layer most significant to stellar structure is probably that at the top of convective envelopes, in and immediately beneath the photospheric layers. For my limited purposes I shall not need to know the detailed stratification of those boundary layers, so I shall not need to discuss the prescriptions that are employed to calculate it.

Notice that the Reynolds stress associated with the convective motion has been omitted from the hydrostatic equation (3.1). That stress is important only in the subphotospheric superadiabatic boundary layers of convective envelopes, whose details I have already decided to ignore.

Finally I must add an equation of state. The perfect-gas law is an adequate approximation for my purposes:

$$p = \frac{\Re \rho T}{\mu}, \tag{3.7}$$

where $\Re$ is the gas constant and μ is the so-called mean molecular weight of the stellar material; I ignore degeneracy and radiation pressure: These are small in sun-like stars, and I wish to keep the discussion simple. If all species were completely ionized

$$\mu^{-1} = 2X + \tfrac{3}{4}Y + \tfrac{1}{2}Z, \tag{3.8}$$

where X, Y and Z are the relative abundances by mass of hydrogen, helium and heavy elements respectively.

4. *Stellar scaling laws*

Because the right-hand sides of Equations (3.1)-(3.7) are all products or quotients of variables, one can seek scaling laws that preserve the functional form of the solution. That would require also that ε, $\varkappa$ and μ can also be similarly expressed; strictly speaking that is not the case, but one can make progress with power-law approximations provided the range of variation of conditions is not too great. In particular, I set

$$\varepsilon = \varepsilon_0 \rho T^{\eta} \tag{4.1}$$

$$\kappa = \kappa_0 \rho^{\lambda} T^{-\nu} \tag{4.2}$$

$$\mu = \mu_0 \tag{4.3}$$

where ε_σ, $\varkappa_0$, μ_0 are functions of X, Y and Z.

One now seeks homologous transformations of a solution to the structure equations. For any dependent variable q, say, one sets

$$q(r) = Q\tilde{q}(x), \tag{4.4}$$

where $x = r/R$, R being the radius of the star, and demands that the function $\tilde{q}(x)$ is independent of the scaling factor Q. Eq. (3.2) implies that ϱ scales as M/R^3. Then Equations (3.1) and (3.3)-(3.7) require the scaling factors to satisfy the relations

$$PR^{-1} \propto GMR^{-2}\left(M/R^3\right), \tag{4.5}$$

$$TR^{-1} \propto F, \tag{4.6}$$

$$LR^{-1} \propto \varepsilon_0\ R^2\left(M/R^3\right)^2 T^\eta, \tag{4.7}$$

where

$$F \propto \kappa_0 R^{-2}\left(M/R^3\right)^{1+\lambda} T^{-(3+\nu)} L \tag{4.8}$$

in a radiative star, or

$$F \propto \mu_0 GMR^{-2} \tag{4.9}$$

in a convective star, and

$$P \propto \mu_0^{-1}\left(M/R^3\right) T. \tag{4.10}$$

In a fully convective star the hydrostatic scaling apparently decouples from the energy scaling; relations (4.6) and (4.9) combine to reproduce relation (4.10) –

indeed, relation (4.9) was derived essentially by requiring that to be so – leaving the system (4.5)-(4.7), (4.9) and (4.10) incomplete. The system is closed by considering the radiative balance in the photospheric regions, thereby relating the energy flux F to the physical state of the atmosphere (e.g. Hayashi, Hōshi and Sugimoto, 1962). Note that the scaling factors M and L apply to the luminosity and mass variables for any fixed value of x, and in particular for $x = 1$. They are, therefore, measures of the total mass and luminosity of the star.

I should point out here that few stars are convective throughout, and no stars are everywhere in radiative equilibrium. Therefore these scaling laws cannot hold exactly, even if the simple approximations (4.1)-(4.3) were exact. Nevertheless, if a star were dominated by either a radiative or a convective zone, one might hope that the scaling laws were roughly valid, at least for that zone. It is with this hope, though in recognition of possible pitfalls, that my discussion proceeds.

Equations (4.5)-(4.10) can be solved to determine how R and L depend upon M and the coefficients μ_0, ε_0 and $\varkappa_0$. Except for very low-mass stars, most of the structure is determined by the radiative equilibrium condition (4.8). Then

$$R \propto (\mu_0 G)^{(\eta-\nu-4)k} (\varepsilon_0 \kappa_0)^k M^{(\eta-\nu+\lambda-1)k}, \tag{4.11}$$

where

$$k = (\eta - \nu + 3\lambda + 3)^{-1}, \tag{4.12}$$

and

$$L \propto (\mu_0 G)^a \varepsilon_0^{-b} \kappa_0^{-(1+b)} M^e, \tag{4.13}$$

where the exponents a, b and e (a no longer being the radiation constant) are given by

$$\begin{aligned} a &= \eta - (\eta + 3)(\eta - \nu - 4)k \\ b &= (\eta + 3)k - 1 \\ e &= \eta + 2 - (\eta + 3)(\eta - \nu + \lambda - 1)k. \end{aligned} \tag{4.14}$$

It is useful to rewrite the scaling laws in terms of chemical composition. To this end we note that $\varepsilon_0 \propto X^2$ with $\eta \simeq 4$ for the proton-proton chain, and recall Equations (4.3) and (3.8) determining μ_0. For stars somewhat more massive than the sun, the CNO cycle dominates the thermonuclear energy production and $\varepsilon_0 \propto XZ$ with $\eta \simeq 16$. If opacity is dominated by bound-free or free-free transitions, Kramers' opacity law holds: $\lambda = 1$ and $\nu = 3.5$. When bound-free transitions dominate (as for young Population I stars), $\varkappa_0 \propto (1+X)Z$, whereas when free-free transitions dominate (as for extreme Population II stars), then $\varkappa_0 \propto 1+X$. The sun is between these two extremes; for sun-like stars I shall adopt the approximation $\varkappa_0 \propto (1+X)Z^d$, with $d = 0.5$. [In very massive stars where electron scattering dominates, $\varkappa_0 \propto 1+X$, $\lambda = 0$ and $\nu = 0$.] Provided that only a limited range of X and Z are considered, these formulae can be approximated by power laws. Thus, Equation (4.13), for example, can be rewritten

$$L \propto G^a \, X^{-f} Z^{-g} M^e, \tag{4.15}$$

where

$$f \simeq \frac{5aX_0}{3+5X_0} + \frac{2b+(3b+1)X_0}{1+X_0}, \tag{4.16}$$

X_0 being a typical value of X, and $g = (1+b)d$, which varies between 0 and $1+b$ depending on the processes that dominate the opacity. In deriving Equation (4.16), which was carried out simply by equating the logarithmic derivatives of L given by Equations (4.13) and (4.15), Z was ignored in the expression (3.8) for μ_0.

As an example I evaluate the expressions for L for sun-like stars, taking $\lambda = 1$, $\nu = 3.5$, $\eta = 4$ and $X_0 \simeq 0.7$. Equations (4.13) and (4.15) then become

$$L \propto (\mu_0 G)^{7.8}\, \varepsilon_0^{-0.08} \kappa_0^{-1.1} M^{5.5} \propto G^{7.8} X^{-4.8} Z^{-0.55} M^{5.5}. \tag{4.17}$$

The corresponding result for somewhat more massive stars powered by the CNO cycle ($\eta \simeq 16$) but with the same composition and opacity law is

$$L \propto (\mu_0\, G)^{7.3}\, \varepsilon_0^{-0.03} \kappa_0^{-1.0} M^{5.2} \propto G^{7.3} X^{-5.8} Z^{-0.5} M^{5.2}. \tag{4.18}$$

Some care should be exercised in interpreting Equation (4.11). For upper main sequence stars the formula is roughly correct, with R being interpreted as the radius of the star. But lower main-sequence stars have extensive convective envelopes, which cause them to deviate from a homologous sequence. For a sun-like star the convective envelope, though extending over a substantial fraction of the radius (about 30 per cent in the case of the sun), has very little mass (and therefore very little weight), and does not have a serious influence on the radiative interior. Equation (4.11) might therefore be used as a rough guide for determining the characteristic scale of variation of the radiative interior, and hence Equation (4.13), which depends on that scaling, remains approximately valid. However, the convective envelope causes the radius of the star to be less than what it would have been had convection not been operative, and the increasing relative extent of the convective envelope as M decreases therefore implies that the actual radius R increases more rapidly with M than is suggested by Equation (4.11). The dependence on X and Z is modified too, because the extent of the convection zone is determined partly by opacity. This behaviour is strictly nonhomologous: to obtain analytical estimates of the stellar radius requires a considerably more sophisticated discussion than that which I am attempting here. In practice stellar models are commonly computed using a relation between heat flux and temperature gradient in the convection zone based on a mixing-length formalism, and the resulting stellar radius depends not only on M and the quantities $\mu_0 G$, ε_0 and $\varkappa_0$, but also on a parameter α, the ratio of mixing length to pressure scale height, which is usually regarded as a constant. Thus I write

$$R \propto X^h Z^j \alpha^u M^v. \tag{4.19}$$

For sun-like stars $h \simeq -1$, $j \simeq -0.5$, $u \simeq -0.2$ and $v \simeq 1.2$. The exponent v declines to about unity as M decreases below $M_\odot$. For high-mass stars $v \simeq 0.75$. The transition between the two extremes of mass occurs for stars of about a solar mass and somewhat higher, and is determined partly by the diminution of the extent of the convective envelope as M increases, and partly by the transition from the p-p chain to the CNO cycle. As I pointed out above, the luminosity L is quite insensitive to the structure of the convective envelopes of sun-like stars, so I add no dependence on α to the scaling (4.17).

Finally, I point out that the slope of the main sequence in the Hertzsprung-Russell diagram (actually the $\log L - \log T_e$ diagram) can be obtained from the scalings (4.15) and (4.19) and the black-body radiation law:

$$L = 4\pi R^2 \sigma T_e^4, \tag{4.20}$$

where σ is the Stefan-Boltzmann constant, and from here on L will be regarded as the (surface) luminosity $L(R)$ of the star.

5. *Main-sequence evolution*

The transmutation of hydrogen into helium causes X to decrease. Thus, μ_0 increases, and ε_0 and $\varkappa_0$ decrease, causing L, according to the scaling (4.13), to rise. The major influence is through the variation of μ_0, whose increase demands a higher core temperature to produce pressure enough to support the weight of the star above. The evolution of the star is not homologous, however, because the nuclear reactions take place only in the inner core, where the temperature is high. Lower main-sequence stars have radiative cores, and the products of the nuclear reactions remain *in situ*; in upper main-sequence stars the core is convective, which homogenizes its chemical composition, but the reaction products do not mix into the radiative envelope. Thus in either case the increase in μ_0 occurs only in the innermost regions of the star. Nevertheless, it is instructive first to assume the star to follow the homologous scaling law (4.13), and afterwards to consider the errors introduced by that assumption. Such an analysis was carried out by Strömgren (1952) for upper main-sequence stars, which at the time were believed to be fully mixed by rotationally induced (Eddington-Sweet) circulation currents.

The rate of change of the hydrogen abundance is given by

$$- ME \frac{dX}{dt} = L, \tag{5.1}$$

where $E \simeq 0.007c^2$ is the energy released per unit mass in converting hydrogen to helium. When coupled with the relation (4.13) and the appropriate expressions for μ_0, ε_0 and $\varkappa_0$, this determines how X and L vary with time t . The formula for $\varkappa_0$ may be evaluated at constant Z, since the effect of nuclear reactions on opacity has only a small influence on the structure of the star. The resulting equation is rather cumbersome to solve, but it can be simplified substantially if the scaling law (4.13) is replaced by the exponential approximation

$$L = L_0 \exp \left[-\beta (X - X_0)\right], \tag{5.2}$$

where X_0 and $L_0 = L(X_0)$ are constants whose values are characteristic of X and

L [and which in practice I shall take to be the initial values of X and L] , and β is a constant of order unity which is determined, like f , by equating the logarithmic derivatives of expressions (4.13) and (5.2) at the initial value X_0 of X. Thus, for example, taking $\varkappa_0 \propto 1+X$ (at constant Z) and assuming the p-p chain to dominate the nuclear reactions,

$$\beta \simeq X_0^{-1} f = \frac{5a}{3+5X_0} + \frac{2b}{X_0} + \frac{1+b}{1+X_0}. \tag{5.3}$$

Using the values of a and b calculated for formular (4.17), with $X_0 \simeq 0.7$, which are characteristic of sun-like stars, yields $\beta \simeq 6.8$.

The integration of Equations (5.1) and (5.2) is straightforward, yielding

$$X = X_0 + \beta^{-1} \ln\left(1 - t/\tau_n\right), \tag{5.4}$$

$$L = L_0 \left(1 - t/\tau_n\right)^{-1}, \tag{5.5}$$

where τ_n is a characteristic nuclear timescale:

$$\tau_n = \frac{EM}{\beta L_0}, \tag{5.6}$$

and I have chosen the constant of integration such that $X = X_0$, $L = L_0$ at $t = 0$. Note that according to Equation (5.4), the time t_n at which hydrogen is exhausted is given by

$$t_n = \tau_n \left(1 - e^{-\beta X_0}\right). \tag{5.7}$$

In practice the evolution is not homologous. Indeed, the inner regions of the star contract, as the scaling (4.11) suggests, whereas, according to the scaling (4.19), the outer regions of a sun-like star expand. Nuclear reactions modify the composition in only the central regions of the star, and consequently in those regions X varies

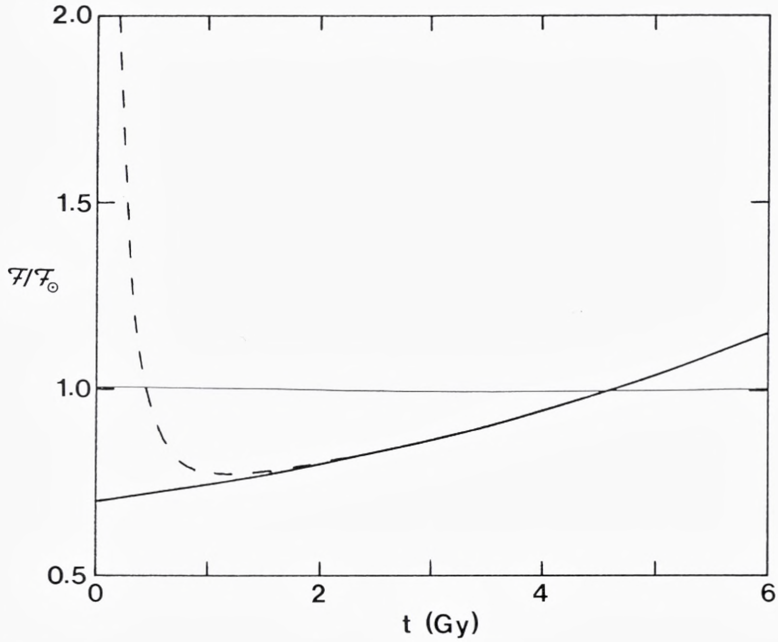

Fig. 1. The thick continuous curve illustrates the evolution of the solar constant $\mathcal{F}(t) \propto L$ according to Equation (5.8), normalized to the present value $\mathcal{F}_\odot$. The thin (almost horizontal) continuous curve shows the evolution of $\mathcal{F}/\mathcal{F}_\odot$ with $L/L_\odot$ given by Equation (8.4) when the gravitational constant declines according to Equation (8.1), with $q = 1$ and $\tau_G(T) = 1.2 \times 10^{11}$y. (A very similar result is obtained with $q = 0.1$ and $\tau_G = 1.3 \times 10^{11}$y.) The dashed curve is $\mathcal{F}/\mathcal{F}_\odot = L/L_\odot$, given by Equations (9.3) and (9.4), for a model losing mass according to Equation (9.1) with $t_0 = 0.05 t_\odot$ and $\Lambda = 0.45$. The initial luminosity is $L_0 = 5.8 L_\odot$. The two models with varying G or M both satisfy $\bar{L}(t_\odot) = L_\odot$.

substantially more rapidly than predicted by Equation (5.4). L varies more rapidly than predicted by Equation (5.5), but the discrepancy is not as great as for X. Indeed, in the case of the sun Equation (5.5) fits numerical computations very well if τ_n is replaced by $\tau_\odot = EM/\beta L_\odot$, yielding

$$L \simeq \frac{L_0}{1 - 0.3t/t_\odot} \simeq \frac{L_\odot}{1 + 0.4(1 - t/t_\odot)}, \tag{5.8}$$

where $t_\odot$ and $L_\odot$ are the current age and luminosity of the sun. The right-hand side of Equation (5.4) then estimates the variation of the mean hydrogen abundance of the star. Equation (5.7) can no longer be used to estimate accurately the main-sequence lifetime, however, because that is determined predominantly by conditions in the core, though it is likely that $\tau_\odot$ scales in roughly the same way as τ_n. According to Equation (5.8), $L(t)$ increases monotonically with t from its initial value $L_0 \simeq 0.7L_\odot$. Thus $\tau_n \simeq 1.3\tau_\odot$. The variation is illustrated in Figure 1.

6. *Calibration of solar models*

Equation (5.8) describing the temporal variation of the solar luminosity L has been calibrated such as to give the observed luminosity $L_\odot$ at the current age $t_\odot$. When computing theoretical models, that calibration is carried out by adjusting the initial composition. Usually Z/X_0 is specified and then X_0 is adjusted to give the correct luminosity. As mentioned above, in practice it is necessary also to adjust the mixing-length parameter α in order to obtain the correct radius $R_\odot$, but since changing α hardly influences the structure of the radiative interior I ignore it here.

To estimate how X_0, or equivalently the initial helium abundance Y_0, must be adjusted, one can use the homology scaling law (4.17) keeping L and M (and, of course, G) constant. Once again taking $X_0 = 0.7$, one thus obtains

$$\frac{\partial \ln Y_0}{\partial \ln (Z/X_0)} = -\frac{X_0}{1-X_0}\frac{\partial \ln X_0}{\partial \ln (Z/X_0)} \simeq 0.30. \tag{6.1}$$

This is the same as the value computed numerically by Bahcall *et al.* (1982) from standard solar models.

One can estimate how $L(t)$ is affected by changes in chemical composition using the arguments of the previous section. First, if one assumes evolution to be homologous, then the only modification arises from the dependence of β on X_0. In particular it is straightforward to demonstrate from the relations (4.17) and (5.3)-(5.6) that the initial luminosity L_0 satisfies

$$\frac{\partial \ln L_0}{\partial \ln \beta} \simeq -0.3, \qquad \frac{\partial \ln L_0}{\partial \ln (Z/X_0)} \simeq 0.02. \tag{6.2}$$

In reality, however, the evolution is not homologous, and the star evolves more rapidly than is predicted by Equations (5.3)-(5.6). This causes the sensitivity of L_0 to β and Z/X_0 to be some 70 per cent greater than is given by Equations (6.2).

7. *The solar neutrino flux*

In principle the solar neutrino luminosity L_ν provides an additional parameter for calibrating solar models. As is well known, that calibration yields an unacceptably low value for the initial helium abundance Y_0 . However, in any discussion of the solar interior it is important to bear in mind the value of L_ν, at least in order to relate it to predictions based on what I might call standard physics (in which, for example, neutrinos are taken to be massless).

According to standard theory, the dominant contribution to the neutrino flux measured by Davis's chlorine detector comes from the decay of ^8B in the p-p III chain. I shall therefore discuss only that contribution, $L_{\nu 8}$, to L_ν. If one assumes that evolved theoretical solar models scale homologously under variations of composition, it follows that $L_{\nu 8}$ is simply proportional to the abundance X_8 of ^8B. By balancing the nuclear reactions of the p-p chain it can easily be shown that the latter is given approximately by

$$X_8 \propto \frac{1-X}{1+X} X^2 \rho T^{24.5} \tag{7.1}$$

(e.g. Gough, 1988). The homology scaling laws of section 4 can now be used to obtain

$$L_{\nu 8} \propto G^{-10.7} M^{-11.8} L^{6.4} (Z/X_0)^{1.8} \propto G^{-13.6} M^{13.8} L^{6.3} Z^{2.0} . \tag{7.2}$$

Once again, I have taken $d = 0.5$ in the opacity formula and have evaluated the exponents for $X_0 = 0.7$. For comparison, partial derivatives quoted by Bahcall and Ulrich (1988) for their standard solar model are

$$\left(\frac{\partial \ln L_{\nu 8}}{\partial \ln L}\right)_{Z/X_0} = 6.8, \qquad \left(\frac{\partial \ln L_{\nu 8}}{\partial \ln (Z/X_0)}\right)_L = 1.3. \tag{7.3}$$

The signs of the exponents are easy to understand. First, notice from the proportionalities (4.17) that increasing either X or Z/X decreases L, requiring that any increase in Z/X requires a compensating decrease in X, given by Equation (5.1), to maintain the current value of L at the observed value $L_\odot$. An increase in Z/X leads to a greater opacity, despite the increased hydrogen abundance, and consequently a greater temperature gradient in the radiative interior. Therefore T is increased in the core. This is consistent with the requirement from the nuclear reactions that a lower concentration of fuel demands a higher temperature to keep the luminosity at $L_\odot$. It is evident from the relation (7.1) that $L_{\nu 8}$ is both a decreasing function of X and an increasing function of T; the dependence of ϱ on X is quite weak and therefore plays only a minor role. Consequently $L_{\nu 8}$ decreases with X and increases with Z/X at constant L.

The dependence of $L_{\nu 8}$ on t cannot be inferred from the simple scaling laws I have been using. The reason is that deviations from homology brought about by the variation of composition in the core increase the central temperature by more than the scaling laws imply, and since $L_{\nu 8}$ is so very sensitive to T (and therefore nearly all the ^8B neutrinos are produced in a small region around the centre of the sun), this dominates the variation of $L_{\nu 8}$. Thus, it is perhaps not surprising that if the sun were, say, older than is generally presumed, the increase in $L_{\nu 8}$ caused by the greater inhomogeneity would dominate the compensating decrease due to the increase in X_0 resulting from the solar calibration: $L = L_\odot$ at $t = t_\odot$. Here I simply quote the result from numerical computations reported by Bahcall and Ulrich (1988) for calibrated solar models:

$$\frac{d \ln L_{\nu 8}}{d \ln t_\odot} \simeq 1.3. \tag{7.4}$$

8. *Temporally and spatially varying gravitational constant*

Having discussed the principal aspects of standard main-sequence evolution, we are now in a position to estimate the consequences of relaxing some of the assumptions listed in section 2. I shall discuss two examples explicitly: in this section, that the gravitational constant G is not constant, and in the next section, that the mass M of the star is not constant.

In some cosmologies the gravitational constant decreases with time (using units of mass, length and time in which Planck's constant and the speed of light are constant). Since, according to the relations (4.17) and (4.18), L is a strongly increasing function of G, a declining gravitational constant would imply that main-sequence stars were considerably more luminous in the past than Equations (5.5) or (5.8) imply. The modification that is made can be estimated from an analysis similar to that in section 5.

The variation of G can often be represented by an equation of the form

$$\frac{G(t')}{G(T_u)} = \left(\frac{t'}{T_u}\right)^{-q} \tag{8.1}$$

where T_u is the age of the universe, t' is time measured from the Big Bang and q is a constant. Typically $0 < q \leqslant 1$. For the evolution of a homogeneous star, Equation (5.1) still holds, but Equation (5.2) must now be replaced by

$$\frac{L}{L_0} = \left(\frac{t'}{T_u}\right)^{-aq} e^{\beta(X-X_0)}, \tag{8.2}$$

where a is given by Equation (4.14). Hence

$$e^{\beta(X-X_0)}\frac{dX}{dt'} = -\frac{L_0}{ME}\left(\frac{t'}{T_u}\right)^{-aq} \tag{8.3}$$

How M varies with t' depends on the cosmology. In most discussions it is constant, and I shall assume that here; I discuss a variation of M separately in the next section. It is then straightforward to integrate Equation (8.3) to determine the variation of X, and thence to substitute the result into Equation (8.2) for the luminosity. If I specialize to the case of the sun, the result may be written

$$\frac{L}{L_\odot} = \left(1+\frac{t-t_\odot}{T_u}\right)^{-aq}\left\{1-\frac{\lambda T_u}{(aq-1)\,\tau_\odot}\left[1-\left(1+\frac{t-t_\odot}{T_u}\right)^{1-aq}\right]\right\}^{-1}, \tag{8.4}$$

where

$$\tau_\odot = \frac{EM}{\beta L_\odot} \simeq 1.5\times 10^{10} y \tag{8.5}$$

and, as was the case previously, $t = t' - (T_u - t_\odot)$ is time measured from the 'zero-age' main sequence. In Equation (8.4) I have introduced a factor λ to account for deviations from homology resulting from the inhomogeneity in the core that results from nuclear transmutations. (It is hoped that there will be no confusion with the exponent of ϱ introduced in Eq. (4.2) to represent the opacity). For equation (8.4) to be a solution of Equations (8.2) and (8.3) describing homogeneous evolution, λ must be set to unity; but for the more realistic inhomogeneous case, $\lambda > 1$. It was pointed out at the end of section 5 that standard solar evolution with G constant is well described by Equation (5.5) with τ_n replaced by $\tau_\odot$, which is equivalent to setting $\lambda \cong 1.4$. Thus we should expect Equation (8.4) with $\lambda \cong 1.4$ to provide quite an accurate description of solar evolution when G satisfies Equation (8.1), as indeed it does (c.f. Pochoda and Schwarzschild, 1964; Ezer and Cameron, 1965; Roeder and Demarque, 1966; Shaviv and Bahcall, 1969).

Because L is an increasing function of G, the effect of the variation (8.1) of G is to augment the past luminosity. For values of q and T_u of typical cosmologies, this predominates over the influence of the varying chemical composition represented by Equation (5.8), and L now decreases with time. A very high luminosity in the past would have severe implications for the Earth's climatic history, and can probably be ruled out. However, climatologists have had difficulty reconciling the relatively low past luminosity that is a consequence of standard physics, and it is therefore of interest to ask what values of q and T_u are required to maintain the solar irradiance on Earth at roughly a constant value. In carrying out that calculation one must take into account the influence of G on the radius $R_\oplus$ of the Earth's orbit. Since the timescale $\tau_G = -G/\dot{G}$ (where the dot denotes differentiation with respect to time) is very much greater than the Earth's orbital period, $R_\oplus$ is determined simply by the (Newtonian) orbit equations in which $\dot{G}$ is ignored and angular momentum (the adiabatic invariant) is held constant: thus $R_\oplus \propto G^{-1}$. The solar irradiance on Earth (the solar constant) is thus proportional to $F = [G(t')/G(T_u)]^2 L$. The values of q and T_u required to hold F approximately constant are then such that $\tau_G(T_u) = q^{-1}T_u \simeq 1.2 \times 10^{11} y$; the precise value is only very weakly dependent on the value of q. An example of $F(t)$, for $q = 1$, is illustrated in Figure 1.

Another possible modification to the theory of gravitation is that the potential V at a distance r from a point mass M is given by

$$V = GM/r \tag{8.6}$$

where

$$G = [1 + \tilde{\alpha} \exp(-r/R_G)]\, G_\infty, \tag{8.7}$$

where R_G is of order 10^2m. For $r \ll R_G$ and $r \gg R_G$ the inverse-square law of force applies, but for astronomical distances the gravitational constant is $(1 + \tilde{\alpha})^{-1}$ of the value measured in the laboratory. Since it is GM that is measured astronomically, it follows that the mass of the sun would be $1 + \tilde{\alpha}$ of the usually accepted value. If $\tilde{\alpha}$ and $G\infty$ are independent of time, the main-sequence evolution must simply follow Equation (5.8), though of course the adjustment of the chemical composition to obtain $L = L_\odot$ at $t = t_\odot$ would be different. The neutrino flux would also be somewhat different; according to the scalings (7.2),

$$L_{\nu 8} \propto (1+\tilde{\alpha})^{-1.1} (Z/X_0)^{1.8} \propto (1+\tilde{\alpha})^{-0.2} Z^{2.0} \tag{8.8}$$

at fixed GM and L. This result is to be compared with the numerical computations of Gilliland and Däppen (1987), who find $\partial \ln L_{\nu 8}/\partial \ln(1+\tilde{\alpha}) \cong -1.2$ at constant Z. For the small value of $\tilde{\alpha}$ suggested by geophysical data, about -7×10^{-3} (Fishbach et al., 1986), the effect on $L_{\nu 8}$ is negligible.

9. *Evolution with mass loss*

The second example I discuss is the consequence of losing mass, by a wind, during the early stages of main-sequence evolution. This possibility has been modelled recently by Guzik *et al.* (1987), who considered the implications for the sun. Here I consider a mass variation of the form

$$\frac{M}{M_\odot} = \frac{1 + \Lambda e^{-t/t_0}}{1 + \Lambda e^{-t_\odot/t_0}}, \tag{9.1}$$

where Λ and the characteristic mass-loss timescale t_0 are constants. This is similar to the variation considered by Guzik *et al.* for two of their three mass-losing models. Once again Equation (5.1) holds, but now Equation (5.2) is replaced by

$$\frac{L}{L_\odot} = \left(\frac{M}{M_\odot}\right)^e \exp\left[-\beta(X - X_\odot)\right], \tag{9.2}$$

where, according to the scaling (4.17), $e = 5.5$. Integrating these equations leads to

$$\frac{L}{L_\odot} = \left(\frac{M}{M_\odot}\right)^e \left\{1 - \frac{\lambda t_o}{\tau_\odot}\left(1 + \Lambda e^{-t_\odot/t_0}\right)^{e-1}\left[I_{e-1}\left(t/t_0\right) - I_{e-1}\left(t_\odot/t_0\right)\right]\right\}^{-2}, \tag{9.3}$$

where[1]

$$I_\sigma(x) \equiv \int \left(1 + \Lambda e^{-x}\right)^\sigma dx \tag{9.4}$$

and $\tau_\odot$ is given by Equation (8.5). I have again introduced the factor λ into the equation for the luminosity to account for the deviations from homology.

As with the case of a declining gravitational constant, mass loss causes the star to be more luminous in the past than is predicted by standard theory, because L is a rapidly increasing function of M. For the short timescales t_0 and the substantial initial mass $(2M_\odot)$ considered by Guzik *et al.*, there is first a rapid decline in L due to the decline in M, followed by the gentle rise, essentially at constant M, described by Equation (5.8). The initial phases of evolution are not very well described by Equation (9.3), because the star is powered predominantly by the CNO cycle in its high-mass phase. Thus L_0 is overestimated by the formula. However, the formula (9.3) should be a good description for cases in which M_0 is not very much greater than $M_\odot$.

Once again we can ask what parameters are required to maintain the luminosity of the sun at approximately a constant value throughout its main-sequence evolution. This can be carried out by demanding, for example, that $\bar{L}(t_\odot) = L_\odot$, where

$$\overline{L}(t) = t^{-1}\int_0^t L(t)dt \tag{9.5}$$

is the time-averaged luminosity. One thus determines a relationship between the initial mass $M_0 = M(0)$ and the mass-loss timescale t_0. Provided $t_0/t_\odot$ is not very small, M_0 varies slowly and is such that $L(t)$ is approximately constant. Indeed, as $t_0/t_\odot$ increases, M_0 tends to a limit for which $L(0) \simeq L_\odot$, which can be estimated from the scaling law (4.17) taking due account of the hydrogen consumed at con-

1. Integrals I_σ defined by Equation (9.4) can easily be generated for integral values of σ from the starting value $I_0 = x$ using the recurrence relation $I_\sigma = I_{\sigma-1} - \sigma^{-1}(1+\Lambda e^{-x})^\sigma$. I_σ is a continuous function of σ, and for the parameters considered here is quite accurately determined for nonintegral values of σ by linear interpolation of log I_σ between the integral values. Thus for all the results reported here, $I_{4.5}$ was approximated by the geometric mean of I_4 and I_5.

stant L in excess of that of the standard model. With the help of Equation (5.5) with τ_n replaced by $\tau_\odot$, this can be written

$$\begin{aligned}\frac{M_0}{M_\odot} &\simeq \left\{\left(1+\frac{\lambda t_\odot}{\tau_\odot}\right)\left[1+\frac{t_\odot}{\beta\tau_\odot X_{0st}}\left(1-\frac{\overline{L}_{st}(t_\odot)}{L_\odot}\right)\right]^{4.8}\right\}^{1/5.5}, \\ &\simeq 1.08,\end{aligned} \tag{9.6}$$

where $X_{0\mathrm{st}}$ is the initial hydrogen abundance of the calibrated standard model and

$$\overline{L}_{st}(t_\odot) = \frac{\tau_\odot L_\odot}{\lambda t_\odot}\ln\left(1+\frac{\lambda t_\odot}{\tau_\odot}\right) \simeq 0.83 L_\odot \tag{9.7}$$

is the mean main-sequence luminosity of the standard model. When $t_0/t_\odot << 1$, the mass of the model quickly approaches $M_\odot$, and the evolution then follows that given by Equation (5.8). Thus in this case it is not possible to maintain L at an approximately constant value. The constraint $\bar{L}(t_\odot) = L_\odot$ demands a very luminous inital phase, of relatively high mass, as is depicted in Figure 1 for the case $t_0/t_\odot = 0.05$.

10. *A further comment on the neutrino flux*

Although the precise profile $X(x,t_\odot)$ at the present time must depend on the detailed history of the star, the main factor influencing it is the total amount of hydrogen consumed. This is proportional to the evolutionary age

$$\tau = t/t_{ms} \tag{10.1}$$

where the main-sequence lifetime t_{ms} is a characteristic time taken to deplete hydrogen in the centre of the star[2]. It is evident, therefore, that the principal effect of a decreasing mass or gravitational constant is simply to make the sun look older by roughly a factor $\bar{L}/\bar{L}_{\mathrm{st}}$. (The adjustment due to changes in t_{ms} resulting from recalibrating X_0 is relatively small). According to the dependence (7.4), this increases

2. A convenient precise definition of t_{ms} is twice the time to deplete the central hydrogen abundance to half its initial value.

the ^8B neutrino flux by a factor $(\bar{L}/\bar{L}_{st})^{1.3}$, which represents an increase of 30 per cent for the models with $\bar{L} = L_\odot$ such as the two represented in Figure 1.

11. *Seismological analysis*

High-order p modes of low degree have been measured from whole-disk Doppler and intensity observations of the sun, and it is likely that similar observations will be successfully carried out on other stars in the near future. The cyclic frequencies ν of oscillations of degree l and order n satisfy for $n \gg l$:

$$\nu \sim \left(n + \tfrac{1}{2}\ell + \varepsilon\right)\nu_0 - \frac{A\,\ell(\ell+1) - B}{n + \frac{1}{2}\ell + \varepsilon}\nu_0 + \ldots, \tag{11.1}$$

where

$$\nu_0 = \left(2\int_0^R \frac{dr}{c}\right)^{-1} \tag{11.2}$$

and

$$A = \frac{1}{4\pi^2\nu_0}\left[\frac{c(R)}{R} - \int_{r_t}^R \frac{1}{r}\frac{dc}{dr}dr\right] \tag{11.3}$$

(e.g. Gough, 1986). Here $c(r)$ is the adiabatic sound speed in the star (rather than the speed of light), and ε and B are constants of the model that depend mainly on conditions in the very outer layers. (Note that ε introduced here is not to be confused with the nuclear energy generation rate.) The lower limit of integration r_t in Equation (11.3) is the radius of the inner turning point of the oscillation eigenfunction, and is given by

$$c(r_t)/r_t = 2\pi\nu/\ell. \tag{11.4}$$

The upper limit, R, is to be interpreted as the radius at which $\nu = \nu_c$, where

$$\nu_c = \frac{c}{2\pi H}\left(1 - 2\frac{dH}{dr}\right)^{\frac{1}{2}}, \tag{11.5}$$

H being the density scale height; for the high-order modes considered here, R is only very slightly less than the radius of the photosphere. Expressions (11.1)-(11.3) with $r_t = 0$ and R equal to the radius of the star can be deduced from the asymptotic analysis of Tassoul (1980), and as $n/l \to \infty$ this is formally equivalent to taking r_t and R as the turning-point radii.

Only modes with $l \leqslant 3$ are likely to be observed in stars other than the sun, at least initially, and from these can be constructed the quantities

$$\Delta_{n\ell} = \nu_{n,\ell} - \nu_{n-1,\ell} \simeq \left[1 + \frac{A\ell(\ell+1) - B}{\left(n + \frac{1}{2}\ell + \varepsilon\right)^2}\right]\nu_0 \simeq \nu_0, \tag{11.6}$$

$$d_{n\ell} = \frac{3}{2\ell+3}\left(\nu_{n,\ell} - \nu_{n-1,\ell+2}\right) \simeq \frac{6\,A\nu_0}{n + \frac{1}{2}\ell + \varepsilon}$$

$$\simeq -\frac{3}{2\pi^2\left(n + \frac{1}{2}\ell + \varepsilon\right)}\int_{r_t}^{R}\frac{1}{r}\frac{dc}{dr}dr \tag{11.7}$$

and

$$\Phi_{n\ell} = \frac{\nu}{\nu_0} - \left(n + \tfrac{1}{2}\ell\right)\nu_0 \simeq \varepsilon + \frac{B}{n + \frac{1}{2}\ell + \varepsilon} - \frac{\ell(\ell+1)d_{n\ell}}{6\nu_0}, \tag{11.8}$$

where, for simplicity, I have neglected in Equation (11.7) the small first term in square brackets in the expression (11.3) for A, and I have also ignored the l dependence of r_t. The quantity Δ_{nl} is an integral property of the entire star, d_{nl} depends predominantly on conditions in the core (since it is an integral of the sound-speed gradient weighted by r^{-1}), and $\varepsilon + B/(n+\frac{1}{2}l+\varepsilon)$ is a functional of the stratification principally in the surface layers.

I must emphasize that the asymptotic relations (11.6)-(11.8) are not accurate. In particular, the $(n+\frac{1}{2}l+\varepsilon)^{-1}$ dependence of the small frequency separation d_{nl} is not well satisfied by either numerically computed eigenfrequencies or the solar frequen-

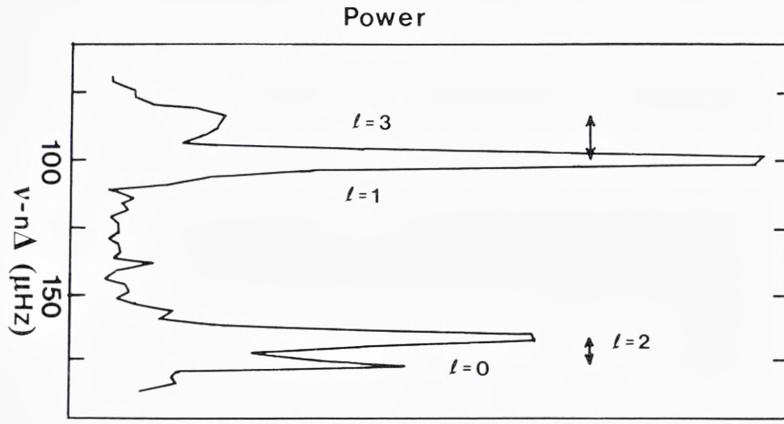

Fig. 2. Superposed power spectrum of solar whole-disk Doppler data obtained by Grec *et al.* (1980). It is a (somewhat optimistic) representation of what might soon be available from observations of other stars. The diagram is obtained by dividing the frequency axis into intervals (ν_{n-1}, ν_n), where $\nu_n = 136n$ μHz, and summing the power. The dissection interval (136μHz) was chosen to maximize some measure of overlap of the peaks, and determines $\triangle$. The double-headed arrows represent $d_0 \simeq 10$μHz and $(5/3)d_1 \simeq 17$μHz; the location of the $l = 0$ peak (177μHz) determines Φ ($\simeq 1.30$).

cies; this inaccuracy is due partly to the neglect of buoyancy and the perturbation of the gravitational potential in the asymptotic analysis, both of which are strongly l dependent. Nevertheless, Equations (11.6) and (11.7) do give some idea of the dominant term that determines the dependence of the measured quantities on the structure of the star, and I shall use them in this discussion as a guide to understanding more accurate quantitative comparisons.

The first stellar observations to reveal a profusion of high-order p modes are unlikely to resolve modes of different degree with like $n+\frac{1}{2}l$; the frequency difference d_{nl} is only of order 10μHz. However, since groups of modes with like $n+\frac{1}{2}$ are approximately uniformly spaced in frequency, a superposed frequency analysis of the kind first carried out by Grec *et al.* (1980) for the sun (illustrated in Figure 2) is likely to yield average values $\triangle$, d and Φ of $\triangle_{nl}$, d_{nl} and Φ_{nl}, from which ν_0, A and a measure of the surface conditions can be estimated. Since the physics of the outer layers of stars is very complicated and ill-understood, Φ_{nl} has not been satisfactorily reproduced theoretically even for the sun. I shall therefore confine my discussion to $\triangle$ and d.

12. *The solar calibration*

From the discussion in section 6 it is evident that standard solar models form a one-parameter sequence, which can be labelled by, say, the initial helium abundance Y_0. This assumes, of course, that the age $t_\odot$ of the sun since the zero-age main sequence is known. Knowledge of $\triangle$ or d therefore affords a basis for selecting the most representative model. The first attempt to calibrate Y_0, using both $\triangle$ and d, failed to produce consistent results (Christensen-Dalsgaard and Gough, 1980): a

Source	d_0(μHz)	d_1(μHz)
Observations	9.2±0.6	9.7
Standard solar models	10.3±0.3	10.0±0.6
Diffusively mixed models	15.5±2.0	13.6±2.3
Helium-deficient model	13	11
Model with varying G	9.0	9.2
Model with mass loss	9.0	9.2
Wimp-infested model	9.3±0.3	9.0±0.3

Table 1. Mean normalized solar p-mode frequency separations d_l, which are averages over n of the quantities d_{nl} defined by Equation (11.7). The observations are averages of d_1 from Claverie *et al.* (1981), Grec *et al.* (1983), Woodard and Hudson (1983) and Harvey and Duvall (1984) ± one standard deviation of the results quoted, taking no account of the (generally smaller) estimated observational errors. The separations for standard solar models are averages of the results of Christensen-Dalsgaard (1982) and Ulrich and Rhodes (1984) ± one standard deviation of those results taking no account of computational inaccuracy. The values of d_l for diffusively mixed models are averages of the results taken from Ulrich and Rhodes (1983), Berthomieu *et al.* (1984), Christensen-Dalsgaard (1986) and Cox and Kidman (1984); the standard deviations are higher here because the authors did not all make the same assumptions about how material was mixed. The entries for the model with wimps were taken from the estimates by Faulkner *et al.* (1986) and Däppen *et al.* (1986) of frequencies of a model by Gilliland *et al.* (1986). The helium-deficient model is that with $Y_0 = 0.19$ computed by Christensen-Dalsgaard, Gough and Morgan (1979). To provide a meaningful comparison the theoretical separations quoted here were computed by averaging the differences from the standard models obtained by each author and adding the result to the mean values quoted in the second row of the table. The models with decreasing G and decreasing M are those illustrated in Figure 1, for which $\bar{L}(t_\odot) = L_\odot$.

calibration based on $\triangle$ yielded a low value, about 0.19, for Y_0 whereas, as can easily be deduced from the entries in Table 1, a calibration based on d yields $Y_0 \simeq 0.27$. Since d depends primarily on conditions in the core where the chemical composition has been modified by nuclear reactions, whereas $\triangle$ depends on conditions throughout the star including the uncertain outer layers, one might at first sight put more trust in the value of d, despite the fact that the low value of Y_0 based on $\triangle$

yields a value for the neutrino flux which is not significantly at variance with observation. In the case of the sun, this opinion was upheld by a calibration using high-degree modes (Berthomieu *et al.*, 1980), but I shall not discuss that here since such modes are not going to be observed in other stars in the foreseeable future. Suffice it to say that a recent computation by Christensen-Dalsgaard *et al.* (1988) using an improved equation of state, which differs from previous equations of state mainly in the upper layers of the convective envelope, has diminished the inconsistency substantially, and thus suggests that the higher value of Y_0 is more-or-less correct.

It is important to realise that the standard view of main-sequence stellar evolution may not be correct, or that the generally accepted value of $t_\odot$ may be in error. So far as the influence on d is concerned, the major factor is the profile of mean molecular weight $\mu(r,t)$, which influences the sound speed c. Let us assume the perfect-gas law, so that

$$c^2 = \gamma \frac{\Re T}{\mu}, \tag{12.1}$$

where γ is the adiabatic exponent $(\partial \ln p/\partial \ln \varrho)_s$, the derivative being taken at constant specific entropy s. Then since the dependence of the nuclear energy generation rate on T is quite sensitive, the gross structure of the star tends to constrain the variation of T quite severely, leaving μ to have the major influence on c.

We are now in a position to make a rough estimate of the variation of d with time. By estimating the variation $X(r,t)$ from the hydrogen depletion by nuclear reactions, and considering the outcome to be but a small perturbation from an homologous evolution, the variation of d for late-type stars (with radiative interiors) is given approximately by

$$d(\tau) \simeq (1 - \Psi\tau)\,\tilde{d}, \tag{12.2}$$

where Ψ is a constant whose value (about 0.7 for sun-like stars) can easily be estimated from numerically computed stellar models, and $\tilde{d}$ is the value of d for a homogeneous stellar model of the same mass and radius. Thus one might expect $\tilde{d}$ to scale approximately according to the homology relation

$$\tilde{d} \propto M^{\frac{1}{2}} R^{-\frac{3}{2}}, \tag{12.3}$$

where, since $\tilde{d}$ is determined mainly by conditions in the radiative interior, R should be expected to characterize the interior scale rather than the entire star. The separation Δ, on the other hand, which depends more sensitively on the outer regions of the star where c is relatively low, is not substantially affected by the details of the μ profile in the core, and would be expected to scale as

$$\Delta \propto M^{\frac{1}{2}} R^{-\frac{3}{2}}, \tag{12.4}$$

where R is the stellar radius. In practice, the relation (12.4) is quite well satisfied, whereas the relation (12.3) is best approximated with a value of R that is intermediate between the stellar radius and that given by the homology law (4.11) which is presumed to represent the scale of the radiative interior. The appropriate R increases as the star evolves, and therefore d decreases with time. If one ignores the difference in the meanings of R in Equations (12.3) and (12.4), it is evident that the nonhomologous component of the evolution is characterized by

$$\delta = d/\Delta \simeq 1 - \Psi\tau. \tag{12.5}$$

As Ulrich (1986) has pointed out, this is a more direct measure of the evolutionary age.

As I mentioned earlier, the asymptotic formulae for the frequency separations Δ and d are not accurate enough for computing reliable theoretical values for comparison with observation. However, they are adequate, and indeed often very useful, for a first estimate of the frequency change produced by some variation to a stellar model. For example, as shown in Table 1, the value of d computed from the standard model 1 of Christensen-Dalsgaard (1982) is about 0.9μHz higher than the mean of the solar observations.[3] It is clear from Table 1 that at the present level of reliability of both the theory and the observations the discrepancy is not significant. Nevertheless, it is evident that that discrepancy could actually be removed if the sun were somewhat older, by an amount

$$\delta t \simeq -\frac{0.9 t_\odot}{d}\left(\frac{d\ln d}{d\ln t}\right)^{-1}, \tag{12.6}$$

3. Exactly how the averaging of d_{nl} is carried out to obtain d is not important for this discussion; all that is important is that the averaging be carried out in the same way for all data sets to be compared.

where d is measured in μHz. In making this estimate I ignored the contribution arising from the change in X required to maintain the observed luminosity; that is very small, as can easily be deduced from Equation (5.8) and the scaling (4.17). Thus from Equation (12.2), the scaling law (12.3) with R being the solar radius, and the temporal derivative of R quoted in Table 2, one obtains $\delta t \cong 6 \times 10^8$ y. Note, however, that gravitational settling of helium would also reduce d by increasing the μ gradient, but I make no attempt here to estimate the magnitude of the effect. Settling of heavier elements increases the opacity in the core; this also increases the central condensation and hence reduces d, though the effect is likely to be less than a comparable settling of helium.

It is evident that we can now immediately estimate the value of d for the models with varying G or M discussed in sections 8 and 9. As was recognized in section 10, it is the total amount of hydrogen consumed that is the principal determinant of the structure of an evolved main-sequence star, so that d can be estimated to be approximately the same as that of a standard model of age $(\bar{L}/\bar{L}_{st})t_\odot$. Hence, for the two models with $\bar{L} = L_\odot$ depicted in Figure 1, the age of the equivalent standard model would be $t_\odot/0.83$, which is about 9×10^8 y greater than $t_\odot$. Hence d is about 1.4μHz less than the standard value, and is currently within the limits set by the observations. According to this analysis, the two models satisfying Equation (9.1) that Guzick *et al.* (1987) considered, with $t_0 = 1.3 \times 10^8$ y and 3.3×10^8 y and each with $M_0 = 2M_\odot$, have present values $d \cong 8$μHz and 6μHz respectively; these values are somewhat lower than the values 9.5μHz and 8.5μHz computed recently by Turck-Chièze, Däppen and Casse (1988). The value of d for a model for which Equations (8.6) and (8.8) hold is influenced predominantly by the modification to the initial hydrogen abundance required for the solar calibration. Thus

$$\frac{\partial \ln d}{\partial \ln\,(1+\tilde{\alpha})} \simeq \frac{\partial \ln d}{\partial \ln X_0} \quad \frac{\partial \ln X_0}{\partial \ln\,(1+\tilde{\alpha})}, \tag{12.7}$$

the partial derivatives being taken at constant GM and Z/X_0. Using the values of the derivations in Table 2 one thus deduces a decrease in d below that of the standard model of $-1.1\tilde{\alpha}d \cong 0.08$μHz, if $\tilde{\alpha} = -7 \times 10^{-3}$. This is consistent with the very small values found numerically by Gilliland and Däppen (1987).

I must point out that these simple scaling laws do not always give the correct result. For example, one might try to estimate the value of d for a helium-deficient solar model. Since, as was pointed out above, the nonhomologous contribution to d due to composition changes is small, two calibrated models with the same value of τ have essentially the same value of d. Evidently $t_{ms} \propto X_0$ for models calibrated to have the correct luminosity, so $\delta\tau/\tau = -\delta X_0/X_0$ at fixed L. In other words, a model with a value of Y_0 that is, say, 0.08 smaller than the standard appears to be younger by $0.08X_0^{-1}t_\odot \cong 5 \times 10^8$ y. Thus we would expect d to be about 1.2μHz higher. As can be seen in Table 1, this is rather greater than the numerical calculations of d_1 (d_l is an average over n of d_{nl}), but substantially less than the computed values of d_0. The reason is presumably that neither the $l = 3$ not the $l = 1$ modes, which determine d_1, penetrate to the very centre of the star, whereas the $l = 0$ modes do. So perhaps there is a severe nonhomologous influence very close to $r = 0$.

Similar remarks apply to diffusively mixed models. The most extreme of a completely mixed model would be a zero-age $1M_\odot$ star calibrated to the solar radius and luminosity, for which we would expect $d \simeq 13$μHz. The scatter amongst the various numerical computations is too great to make a sound comparison, though the estimate given here appears to be somewhat too low.

Finally I include in Table 1 the results of an estimate made by Faulkner *et al.* (1986) from the asymptotic formula (11.7) of a solar model with a cloud of weakly interacting massive particles (wimps) in its interior. The effect of wimps is to introduce an additional energy-transporting agent into the inner regions, thereby making the core more nearly isothermal. Despite the resulting reduction in the μ gradient, the effect is to reduce the sound speed preferentially near the centre, and so decrease d. Numerical calculations by Däppen *et al.* (1986) have confirmed this result. The entries in Table 1 are averages of the two calculations.

13. *Asteroseismological calibration*

For stars of known chemical composition, the seismological diagnostic quantities Δ and d provide a very important supplement to the usual astronomical data. In particular, for zero-age main-sequence stars, the scaling (12.4), together with a mass-radius relation, should in principle permit one to infer the mass of a star. Note that if one ignores the difference between the meanings of R in the scalings (12.3) and (12.4), a knowledge of d would at first sight provide no new information. However, if that difference is taken into account, one would expect Δ and d to scale rather differently with chemical composition, and therefore some overall abundance information would also be available.

For older stars there is the added richness afforded by the temporal evolution. Since both d and Δ change with time, one might expect to be able to determine both the mass and the age of a star. This was first pointed out by Christensen-Dalsgaard (1986), but would evidently be the case only if all the other uncertain parameters determining the structure of the stellar model were known (and, of course, provided that the implementation of the theory of stellar evolution were correct).

Before addressing the issue of what might actually be learned from $\triangle$ and d to constrain the possible structure of a stellar model, it is useful first to consider how $\triangle$ and d depend on M and t. To this end it is necessary to determine how R varies with time. Since, according to the analysis of section 5, the evolution results solely from the depletion of X which causes an augmentation of L, I shall adopt the approximation $\partial\ln R/\partial\ln\tau = (\partial\ln R/\partial\ln L)\partial\ln L/\partial\ln\tau$, where the partial derivative $(\partial\ln R/\partial\ln L)$ is at constant M, τ, Z and α, and can thus be deduced from the scalings (4.17) and (4.19), and the time derivatives represent the evolution of a given model (at constant M, Z and α). Thus $\partial\ln L/\partial\ln\tau$ is obtained by differentiating Equation (5.8). The value of this derivative, together with the other partial derivatives of R and L obtained from the scalings (4.17) and (4.19) and Equation (5.8) are listed in the first two columns of Table 2. From the derivative of Equation (12.2) and the scalings (12.3) and (12.4), the partial derivatives of $\triangle$ and d can thus be determined. These too are listed in Table 2. (For this purpose the scaling

a_i/A_k	ln R	ln L	ln T_e	ln $\triangle$	ln d
ln X_0	-1	-4.8	-0.7	1.5	1.5
ln Z	-0.5	-0.55	0.11	0.75	0.75
ln M	1.2	5.5	0.78	-1.3	-1.3
ln α	-0.2	0	0.10	0.3	0.3
ln τ	0.09	0.4	0.06	-0.14	-0.68

Table 2. Partial derivatives $\partial A_k/\partial a_i$ of the properties A_k = (ln R, ln L, ln T_e, ln $\triangle$, ln d) of a standard solar model with respect to the control variables a_i = (ln X_0, ln Z, ln M, ln α, ln τ).

of $\tilde{d}$ has been assumed to be the same as that of $\triangle$.) It is evident that both $\triangle$ and d decrease with time, d decreasing roughly five times faster than $\triangle$. The results are plotted against each other in Figure 3, which, in view of the cavalier way in which I have simplified the analysis, can be regarded as only a very crude representation of what actually occurs. The careful numerical computations reported by Christensen-Dalsgaard (1988), however, are at least superficially similar.

Knowledge of $\triangle$ and d fixes a point in Figure 3, to which correspond values of M and τ (and hence t). But that is unambiguously the case only if the chemical abundance parameters, X_0 and Z, and the mixing-length parameter α are known, together with, say, the luminosity L. How well M and t can be determined depends on the errors in the other astronomical information about the star and on the sensitivity of the asteroseismological analysis to those errors. This issue has been partially addressed by Ulrich (1986, 1988), who determined the information required to obtain the entries in Table 2 from numerical computations. His results differ

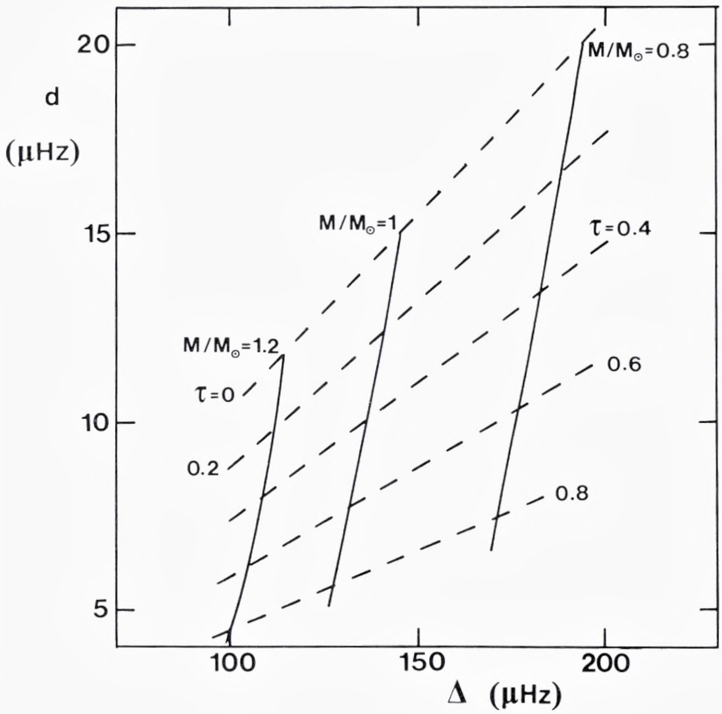

Fig. 3. Dependence of d upon Δ for late-type main-sequence stars, estimated from the rough scaling arguments described in the text. The solid lines represent the evolution of a given star (M = constant); the dashed lines are evolutionary isochrones (lines of constant $\tau = t/t_{ms}$). If δ rather than d were plotted against Δ, the evolutionary isochrones, in this approximation, would be uniformly spaced horizontal lines.

somewhat from the results I have obtained from the simple scaling laws, but that will not alter the qualitative nature of my principal conclusion.

Let $a_i = (\ln X_0, \ln Z, \ln\alpha, \ln M, \ln\tau)$ be the control parameters for the theoretical stellar model, and let $A_k = (\ln T_e, \ln L, \ln\Delta, \ln d)$ be the output. Then from the partial derivatives $\partial A_k/\partial a_i$ of Table 2 one knows that small changes δa_i in the control parameters produce changes δA_k in the output, according to the linearized equations

$$\frac{\partial A_k}{\partial a_i}\delta a_i - \delta A_k = 0. \tag{13.1}$$

One can now ask how any four parameters B_k, say, selected from the nine parameters a_i, A_k, depend on the remaining parameters b_i. More generally, the nine parameters b_i, B_k can be any set of independent functions of a_i, A_k. The appropriate relations between small perturbations δb_i, δB_k to b_i, B_k are, of course, simply Equations (13.1) rewritten in terms of the new variables thus:

$$\left(\frac{\partial A_k}{\partial a_i}\frac{\partial a_i}{\partial B_m}+\frac{\partial A_k}{\partial B_m}\right)\delta B_m+\left(\frac{\partial A_k}{\partial a_i}\frac{\partial a_i}{\partial b_j}+\frac{\partial A_k}{\partial b_j}\right)\delta b_j=0, \quad (13.2)$$

which can be solved formally for δB_m in terms of δb_j. If one then sets $b_j = \delta_{ji}$ (where δ_{ij} is the Kronecker delta), to obtain the solution $\delta_i B_k$, say, the appropriate sensitivity matrix is given by

$$\frac{\delta_i B_k}{\delta b_j}=\frac{\partial B_k}{\partial b_i}. \quad (13.3)$$

In Equation (13.2) the partial derivatives $\partial A_k/\partial a_i$ are defined regarding A_k to be functions of a_i as in Equation (13.1); thus they are the partial derivatives given in Table 2. The derivatives $\partial\alpha_i/\partial\beta_j$, where α_i is any of a_i or A_k and β_j is b_i or B_k, are the partial derivatives of the transformation $\alpha_i = \alpha_i(\beta_j)$ between the two sets of variables. The resulting partial derivatives in equation (13.3) are thus taken with the components B_k considered as functions b_i, analogously to those of Equation (13.1).

The results of such a transformation is given in Table 3, where I have taken $B_k = (\ln\tau, \ln M, \ln Y, \ln\alpha)$ in the hope that they can be determined in terms of $b_i = (200\ln T_e, 10\ln L, 10\ln(Z/X_0), \triangle, d)$, where $\triangle$ and d are measured in μHz. In so

b_i	δb_i	$\delta t/t$	$\delta M/M$	$\delta Y/Y$	$\delta\alpha/\alpha$
T_e	$0.005T_e$	0.003	−0.03	0.09	0.13
L	$0.1L$	−0.012	0.14	−0.38	−0.49
Z/X_0	$0.1Z/X_0$	−0.078	−0.17	1.8	−7.4
$\triangle$	1μHz	0.013	0.01	−0.05	−0.04
d	1μHz	−0.19	−0.01	0.08	0.08

Table 3. Sensitivity matrix $\partial B_k/\partial b_i$. The entries in columns 3-6 represent the contributions to the relative errors in calibrating t, M, Y or α from the errors δb_i in b_i that are listed in column 2.

doing, I have assumed that $t_{ms} \propto X_0/L$. The factors in the definition of b_i have been chosen such that a value of unity for any component δb_i represents a (perhaps optimistic) estimate of the smallest error one might make from the observation of an isolated sun-like star. Thus the entries in Table 3 represent the contributions from those errors in b_i to the resulting relative errors in t, M, Y and α. It is evident from the magnitude of the entries that from a knowledge of the position of a star on Figure 3 the mass and age of that star can each be determined to within about 20 %. However one cannot determine the initial helium abundance Y_0. A somewhat less optimistic conclusion is reached (Gough, 1987) if one replaces Table 2 by the numerical derivatives one deduces from the data provided by Ulrich (1986, 1988)[4].

The relatively high sensitivity of the components of B_k of Table 3 to Z/X_0 suggests that perhaps Z/X_0 should not be used as a control variable. This conclusion is strengthened by the realization that Z enters the equations principally via the opacity, and that therefore this analysis also reflects the sensitivity of the results to errors in the uncertain results of opacity calculations. Thus, if surface gravity, for example, were used instead, or if a determination of M could be made for a component of a binary system, then perhaps Z/X_0 and τ could be determined more reliably (Gough, 1987).

The ramifications of this seismological sensitivity analysis, even for sun-like stars, have not yet been fully explored. Nor has the analysis been carried out for stars whose structure differs substantially from that of the sun. Therefore it is not yet possible to make a quantitative assessment of the role the anticipated asteroseismic data will play. It is already evident from the difference between Table 3 and an analogous table (Gough, 1987) computed from presumably more accurate numerical calculations by Ulrich that variations in the values in Table 2 can produce quite substantial changes in the outcome of the analysis. Thus there might be regions in Figure 3 that are significantly less sensitive to the uncertain control parameters.

14. *Conclusion*

The purpose of my introductory discussion has been mainly to provide some insight into the structure and evolution of sun-like stars on the main sequence. I have avoided use of the results of complicated numerical calculations as much as possible, relying on simple scaling arguments wherever I could. This is in character with many

4. My earlier conclusion (Gough, 1987) was very much less optimistic than that arrived at here. The main reason is that errors in modern spectroscopic determinations of T_e and Z/X_0 for stars observed with the precision required for detecting a spectrum of p-mode oscillations (cf. Smith et al., 1986; Perrin et al., 1988; Spite et al., 1989) are substantially less than the older estimates which I had used.

of the arguments Strömgren himself advanced during his very productive career. Although the results of scaling arguments are not as accurate as detailed numerical calculations, and should not be used for quantitative comparisons between theory and observation if precise numerical results are available, they do provide a means by which some understanding of numerical computations can be achieved, and they are also often a ready tool for making preliminary estimates of the effects of some small change to the assumptions of the theory. Thus they are an extremely important complement to numerical modelling.

I have not attempted a serious review of the literature, but have instead concentrated on physical rationalizations. Much of what I have discussed is very well known, but I have also extended the technique to address some of the seismic properties of stars, which are of only quite recent interest and which promise to become an extremely important diagnostic of stellar structure. Thus I hope that this contribution will be of some practical scientific use.

One of the issues to which Strömgren devoted considerable attention is the helium content of main-sequence stars. With the addition of anticipated new asteroseismic data I have readdressed that subject, originally in the hope that substantial progress might soon be made. The initial results were perhaps disappointing, because the inferred helium abundance appears to be extremely sensitive to uncertainties in the astronomical data. However, sensitivity in one direction always implies that there is a direction, in some sense opposite, in which there is very little sensitivity; and that is perhaps the direction to go to make reliable physical deductions. Therefore we can look forward to the new data in the optimistic belief that important new constraints will be set on the possible structure of main-sequence stars, which no doubt will lead to an improvement in our understanding of stellar evolution.

Acknowledgements

I am very grateful to P. E. Nissen for pointing out that the work of Smith *et al.* (1986) and Perrin *et al.* (1988) shows that careful spectroscopic analysis can yield values of T_e and Z/X with only 0.5 % and 10 % errors respectively, and to P. S. Conti for drawing my attention to the work of Spite *et al.* (1989). I thank C. A. Morrow and E. Novotny for suggesting improvements to the original manuscript.

References

Bahcall, J.N. and Ulrich, R.K. 1988, *Rev. Mod. Phys.* **60,** 297.

Bahcall, J.N., Huebner, W.F., Lubow, S.H., Parker, P.D. and Ulrich, R.K. 1982, *Rev. Mod. Phys.* **54,** 767.

Berthomieu, G., Provost, J. and Schatzman, E. 1984, *Nature* **308,** 254.

Berthomieu, G., Cooper, A.J., Gough, D.O., Osaki, Y., Provost, J. and Rocca, A. 1980, *Nonradial and nonlinear stellar pulsation,* ed. H.A. Hill and W.A. Dziembowski (Springer, Heidelberg) p. 313.

Christensen-Dalsgaard, J. 1982, *Mon. Not. R. astr. Soc.* **199,** 735.

Christensen-Dalsgaard, J. 1986, *Seismology of the sun and distant stars,* ed. D.O. Gough (NATO ASI C169, Reidel, Dordrecht) p. 23.

Christensen-Dalsgaard, J., Däppen, W. and Lebreton, Y. 1988, *Nature* **336,** 634.

Christensen-Dalsgaard, J. 1988, *Advances in helio- and asteroseismology,* ed. J. Christensen-Dalsgaard and S. Frandsen, IAU Symp. **123** (Reidel, Dordrecht) p. 295.

Christensen-Dalsgaard, J. and Gough, D.O. 1980, *Nature* **288,** 544.

Christensen-Dalsgaard, J., Gough, D.O. and Morgan, J.G. 1979, *Astron. Astrophys.* **73,** 121; **79,** 260.

Claverie, A., Isaak, G.R., McLeod, C.P., van der Raay, H.B. and Roca Cortes, T. 1981, *Nature* **293,** 443.

Cox, A.N. and Kidman, R.B. 1984, *Theoretical problems in stellar stability and oscillations,* ed. A. Noels and M. Gabriel (Institute d'Astrophysique, Liège) p. 259.

Däppen, W., Gilliland, R.L. and Christensen-Dalsgaard, J. 1986, *Nature* **321,** 229.

Duvall Jr, T.L., Dziembowski, W.A., Goode, P.R., Gough, D.O., Harvey, J.W. and Leibacher, J.W. 1984, *Nature* **313,** 27.

Dziembowski, W.A. 1983, *Solar Phys.* **82,** 259.

Eddington, A.S. 1925, *Observatory* **48,** 73.

Ezer, D. and Cameron, A.G.W. 1965, *Canad. J. Phys.* **44,** 593.

Faulkner, J., Gough, D.O. and Vahia, M.N. 1986, *Nature* **321,** 226.

Fishbach, E., Sudarsky, D., Szafer, A. and Talmadge, C. 1986, *Phys. Rev. Lett.* **56,** 3.

Gilliland, R.L. and Däppen, W. 1987, *Astrophys. J.* **313,** 429.

Gilliland, R.L., Faulkner, J., Press, W.H. and Spergel, D.N. 1986, *Astrophys. J.* **306,** 703.

Gough, D.O. 1976, *Proc. Intern. Colloq. Drops and Bubbles,* ed. D.J. Collins, M.S. Plesset and M.M. Saffren (US Govn. Printing Office, Washington) p. 11.

Gough, D.O. 1985, *Future missions in solar heliospheric and space plasma physics,* ed. E.J. Rolfe and B. Battrick (ESA SP-235, Noordwijk) p. 183.

Gough, D.O. 1986, *Hydrodynamic and magnetohydrodynamic problems in the sun and stars,* ed. Y. Osaki (Univ. Tokyo Press, Tokyo) p. 117.

Gough, D.O. 1987, *Nature* **326,** 257.

Gough, D.O. 1988, *Solar-terrestrial relationships and the Earth environment in the last millenia,* ed. G. Cini-Castagnoli, XCV Corso Enrico Fermi (Soc. Ital. di Fisica, Bologna) p. 90.

Grec, G., Fossat, E. and Pomerantz, M. 1980, *Nature* **288,** 541.

Grec, G., Fossat, E. and Pomerantz, M. 1983, *Solar Phys.* **82,** 55.

Guzik, J.A., Willson, L.A. and Brunish, W.M. 1987, *Astrophys. J.* **319,** 957.

Harvey, J.W. and Duvall Jr, T.L. 1984, *Solar seismology from space,* ed. R.K. Ulrich, J.W. Harvey, E.J. Rhodes Jr and J. Toomre (JPL Publ. 84-84, Pasadena) p. 165.

Hayashi, C., Hōshi, R. and Sugimoto, D. 1962, *Prog. Theor, Phys. Suppl.* **22,** 1.

Perrin, M.N., Cayrel de Strobel, G. and Dennefeld, M. 1988, *Astron. Astrophys.* **191,** 237.

Pitts, E. and Tayler, R.J. 1985, *Mon. Not. R. astr. Soc.* **216,** 139.

Pochoda, P. and Schwarzschild, M. 1964, *Astrophys. J.* **139,** 587.

Roeder, R.C. and Demarque, P.R. 1966, *Astrophys. J.* **144,** 1016.

Schatzman, E., Maeder, A., Angrand, F. and Glowinski, R. 1981, *Astron. Astrophys.* **96,** 1.

Smith, G., Edvardsson, B. and Frisk, U. 1986, *Astron. Astrophys.* **165,** 126.
Shaviv, G. and Bahcall, J.N. 1969, *Astrophys. J.* **155,** 135.
Spite, F., Spite, M. and François, P. 1989, *Astron. Astrophys.* **219,** 25.
Strömgren, B. 1938, *Astrophys. J.* **87,** 520.
Strömgren, B. 1952, *Astron. J.* **57,** 65.
Sweet, P.A. 1950, *Mon. Not. R. astr. Soc.* **110,** 548.
Tassoul, J.-L. 1978, *Theory of rotating stars* (Princeton Univ. Press, Princeton).
Tassoul, M. 1980, *Astrophys. J. Suppl.* **43,** 469.
Turek-Chièze, S., Däppen, W. and Casse, M. 1988, *Seismology of the sun and sun-like stars*, ed. E.J. Rolfe (ESA SP-286, Noordwijk) p. 629.
Ulrich, R.K. 1986, *Astrophys. J.* **306,** 137.
Ulrich, R.K. 1988, *Advances in helio- and asteroseismology,* ed. J. Christensen-Dalsgaard and S. Frandsen, IAU Symp **123** (Reidel, Dordrecht) p. 299.
Ulrich, R.K. and Rhodes Jr., E.J. 1983, *Astrophys. J.* **265,** 551.
Ulrich, R.K. and Rhodes Jr., E.J. 1984, *Solar Seismology from Space,* ed. R.K. Ulrich, J.W. Harvey, E.J. Rhodes Jr. and J. Toomre (JPL Publ. 84-84, Pasadena) p. 371.
Woodard, M. and Hudson, H.S. 1983, *Nature* **305,** 589.
Vogt, H. 1925, *Astron. Nachr.* **223,** 229.
von Weizsäcker, C.F. 1937, *Physik. Zs.* **38,** 176.
Zahn, J.-P. 1989, *Astrophysical Fluid Dynamics,* Les Houches Session XLVII, ed. J.-P. Zahn and J. Zinn-Justin (Elsevier) in press.
von Zeipel, H. 1924, *Festschrift für H. von Seelinger,* p. 144.

Recent Progress and Future Prospects in the Study of Stellar Atmospheres

By DIMITRI MIHALAS
Department of Astronomy
University of Illinois
349 Astronomy Building, 1011 West Springfield Av., Urbana, IL 61801, USA

I first met Bengt Strömgren 25 years ago, in the autumn of 1963, having just finished my degree work at Caltech, and just arrived in Princeton as a Higgins Visiting Fellow. Naturally it was quite a thrill to meet personally a man who had made so many brilliant contributions to my field of study. As I got to know him I also got deeper and deeper in his debt for the many important suggestions he offered me, and for the considerable encouragement he gave me. So it is indeed a great privilege and pleasure to be here today to honor his memory.

What I hope to do this morning is to give a brief survey of where we started in this field, where we are now, and where we are going. Needless to say, I will emphasize those topics I am most interested in and know the most about; so it is likely that another worker in this field might sketch a somewhat different picture. I purposely intend to keep the talk simple and nontechnical, but I hope I don't offend anyone by being too elementary.

Stellar atmospheres was a favorite (but probably not *the* favorite) topic of Strömgren, to which he made many important contributions, and to which he returned repeatedly. Like many branches of astrophysics, the study of stellar atmospheres is relatively young, and is still changing and growing rapidly. Although radiative transport theory got started early in this century, stellar atmospheres theory per se couldn't move forward until the 1920's when Saha enunciated his ionization law, and quantum mechanics was worked out to the point that it could be used to compute reliable atomic data. Today this field is no longer being held up by the need for fundamental breakthroughs in basic physics, and progress in many aspects of the subject has become relatively dependable as computers become faster and bigger, and as atomic data become more complete and accurate.

I. *Why Study Stellar Atmospheres?*

When I first came to Princeton I was so wrapped up in the subject of stellar atmospheres that I was not sharply aware of many other things going on in astrophysics. So it was a bit of a surprise one day when Ed Salpeter, who was visiting for a few days, asked me, in effect, "Why does one even bother to study stellar atmospheres?". His point was that, after all, a stellar atmosphere contains only 10^{-11} or 10^{-10} of the whole mass of the star, and is therefore obviously inconsequential for structural or evolutionary considerations except insofar as it might influence slightly the outer boundary conditions one used to construct a stellar model. (Recall that at that time it was common to use "zero" boundary conditions – density, pressure, temperature all zero – at the stellar surface).

Given his background and interests, Salpeter's remark is perfectly reasonable. But for others of us, these layers have a great deal of intrinsic interest, far beyond the fact that they supply a boundary condition for the star as a whole. These, after all, are the layers we can *see*, and which our instruments can measure and probe (remotely of course). It is by some kind of "suitable analysis" of the light received from stellar atmospheres that we may hope to deduce: (a) The physical structure of the atmosphere, e.g. the run of temperature, density, pressure, and ionization degree as a function of depth. (b) The chemical composition of the atmosphere (and presumably, at least in most cases, the composition of the star). (c) The strength and topology of any magnetic fields in the star's atmosphere. (d) Velocity fields as a function of depth, and thus something about the dynamics of the atmosphere.

Clearly there is a rich reward for the effort expended! And, as Strömgren so beautifully demonstrated, information about effective temperatures, surface gravities, bolometric corrections, and compositions can be coupled directly into the theory of stellar evolution. And the resulting information can be coupled into a knowledge of stellar kinematics to build up a picture of the dynamical evolution of the galaxy, again a favorite area of study of Strömgren's.

II. *Development of the Basic Theory*

I would like to give here a brief summary of the development of stellar atmospheres theory based on the paradigm that the atmosphere is a radiatively dominated boundary layer that connects an equilibrium (or at least local equilibrium) interior to empty black space. In the first quarter of this century Schuster, Schwarzschild, Eddington, and Milne had formulated the transfer equation, identified the basic physical processes leading to the absorption, emission, and scattering of radiation, and had developed the basic theory of radiation transport. In the process they had been able to solve approximately the problem of radiative equilibrium (energy balance) in a grey medium, and had thus derived a rough estimate of the temperature distribution in the solar atmosphere.

But further progress was virtually halted until the development of quantum mechanics in the middle to late 1920's. Then it suddenly became possible to predict reasonably accurate values for spectrum line strengths and continuum cross-sections, hence opacities. Reliable detailed analyses of line profiles became possible for the first time, and these led to believable estimates of atmospheric conditions (e.g. temperature, density) and even some fairly trustworthy abundances. Discrepancies between predicted and observed profiles spurred people like Eddington, Milne and Strömgren to make deeper investigations into the nonequilibrium nature of the absorption/emission/scattering processes in spectral lines, and of the effects of partial coherence on line shapes. Thus many of the basic ideas necessary for a physically complete picture of line formation were already in hand by the 1930's.

By 1940 Strömgren broke new ground by developing the "model atmosphere method". In this approach one constructs, numerically, idealized stellar atmospheres based on a set of simplifying physical assumptions. The basic assumptions typically made in most of the early work are:

(1) The atmosphere is stratified in *plane homogeneous layers.*
(2) The atmosphere is in *hydrostatic equilibrium.*
(3) The atmosphere is in *radiative equilibrium.*
(4) The material is in *local thermodynamic equilibrium* (LTE).
(5) The material is *grey.*

Before proceeding farther, it is worthwhile to consider the significance and implications of these assumptions.

Curiously enough, the first assumption, which sounds relatively innocent, is probably the most far-reaching, and is certainly the most difficult to remove. By postulating homogeneous layers we reduce the problem to a strictly 1D geometry with complete translational invariance. Thus all the differential and/or integral equations to be solved are one-dimensional, and are therefore easily handled by even modest computing capabilities. If one says instead that we must treat 2D or even 3D structures, then we are immediately faced with a much more difficult problem. Not only do we require knowledge of our variables on many more grid-points, hence a much larger computational capacity, but also in order to specify the atmospheric structure we will have to solve a much more complex set of hydrodynamic or magnetohydrodynamic equations. It is only now, and then almost exclusively in solar work where we have some idea of what the structures look like, that multidimensional structures are being considered.

The second assumption merely says that the atmosphere is static (no motions), in which case the hydrodynamic equations degenerate to a hydrostatic stratification. Assumptions (1) and (2) are intimately connected because if we have dynamics the layers are unlikely to be homogeneous, and conversely if we have 2D or 3D structures they are unlikely to be static (except, perhaps, in a theoretician's model!).

Because we say the atmosphere is static (and we can, of course, neglect thermonuclear reactions) there can be no hydrodynamic work terms or time-dependent changes in the material energy density, so the atmosphere has no choice but to re-emit exactly as much radiation as it absorbs, hence to be in radiative equilibrium. It is improbable that any stellar atmosphere is strictly in radiative equilibrium; yet calculations based on that assumption agree well (sometimes astonishingly well) with observation, so the assumption is almost always invoked.

The idea of LTE has its roots in the stellar interior, where it is manifestly a very good approximation. Basically it says that we may calculate all physical variables from the standard equations of (equilibrium) statistical mechanics using the local value of the temperature and total density. The only exception to this rule is the radiation field, which is allowed to be nonlocal, being calculated from a transfer equation. In a stellar atmosphere whence photons may freely escape into interstellar space the assumption of LTE is manifestly inconsistent. Indeed, very simple order of magnitude estimates show that the state of the material is strongly dominated by radiative rates in the more rarefied outer layers where collisional rates become small, and therefore the material has no choice but to depart from equilibrium. A correct evaluation of the size of these effects had to wait another 25-30 years, and in the interim LTE was an essential (i.e. without which no modeling could be done) assumption which (amazingly!) usually gave what seemed to be good answers.

In the face of the mathematical complexity of the equations to be solved (typically nonlinear integrodifferential or partial integrodifferential equations), a numerical approach was (and still is) the only one possible. Strömgren realized that the calculation could not be mechanized at that time (computers didn't exist), but could be highly organized, which he proceeded to do. The resulting program was both powerful and flexible, but even with Strömgren's labor saving tricks it was very laborious (calculation on hand-driven or electrical desk calculators) and very slow. During the dark years of World War II Strömgren and his associates and students led the field and dominated it completely, becoming known as the "Copenhagen school" of stellar atmospheres. And after the war these same people spread out over Europe and North America, bringing the new techniques to a very wide audience.

Scientifically it was a remarkably fruitful time. In retrospect one marvels that so much could be done with such modest equipment and such crude algorithms. It is almost always the case in computational physics that the first few people who try to solve a new problem use old techniques, which usually turn out to be much less powerful and slower than the efficient methods that always seem to come once the problem is reasonably well understood. This remark certainly applies to Strömgren's pioneering efforts, which simply cried out for elaboration and generalization once automatic computation became possible.

In the 1950's the situation changed radically with the advent of electronic computers. It is almost impossible to communicate to the current generation of students, who have lived and played with computers all of their lives, what what a truly earth-shaking development even the most primitive "homemade" machines of the early 50's were! Possibly only one who has spent hundreds of hours doing a numerical integration by hand can have any idea what it meant to suddenly be able to do hours of work in seconds. Even with those early machines with their modest computational power one could begin to drop some of the restrictive assumptions that had been necessary earlier. For example it was possible to make nongrey models that allowed for the variation with wavelength of the continuous opacity (but omitted spectrum lines). And some ideas about how to enforce energy-balance (radiative equilibrium) in the atmosphere could be tested. For the first time we began to have at least a rough idea of the energy distribution of the hottest stars out into the ultraviolet. At about the same time exciting discoveries were being made using even the fairly crude models then available and primitive curve of growth techniques. In particular it was possible to establish the tremendous deficiency of heavy elements in the atmospheres of subdwarfs and thus map out some of the main features of cosmochemistry on a galactic scale. But in the absence of better models the analysis was effectively restricted to stars of near-solar temperature.

In the late 50's IBM released its famous IBM 7090, a transistorized follow-on to the 704 and 709 (which was not designed to satisfy the needs of scientists – not to mention astronomers! – but to provide a highly reliable machine for the US's Distant Early Warning [DEW] Line). A novel feature of this machine is that it came equipped with software: it had a compiler! And so the labor of coding diminished by a couple of orders of magnitude, and there was a virtual explosion of new work in the late 50's and throughout the 60's. Really good methods for solving transfer equations and enforcing the constraint of radiative equilibrium were rapidly developed. And with robust codes it became possible to make surveys of wide ranges of effective temperatures and gravities. Reasonably realistic nongrey continuum models were made from the O-stars down to the middle F's or early G's. Cooler than that line absorption becomes so heavy that models omitting lines are obviously seriously inadequate. From these efforts we were able to derive pretty good estimates of effective temperature as a function of spectral type. And it was possible to begin to make practical connections with observations by calculating, as a function of effective temperature and surface gravity, Strömgren's intermediate-band (*uvby*) colors or Johnson and Morgan's broadband (*UBV*) colors from the computed energy distributions of the models.

In the 1960's large grids of model atmospheres were constructed by several workers. And at about the same time Griem and his coworkers made a breakthrough in the broadening theory for lines of hydrogen and hydrogenic ions, and shortly afterward for He and hundreds of lines of most ions of astrophysical interest. For the first time astrophysicists were in a position to realize the promise of Pannekoek's early work and to use the hydrogen and helium lines as reliable temperature and density diagnostics (verified in the laboratory). In the early type stars (A-O) allowance could finally be made for the effects of blanketing by the lines of H, He, He+, and the strongest ultraviolet resonance lines of the astrophysically most abundant elements.

In this same era rockets were giving us our first glimpses of stellar ultraviolet spectra. A number of special-purpose codes were developed to make as accurate a fit to the H and He lines as possible.

In addition, some of the drudgery and inaccuracy of curve-of-growth analyses was relieved by codes designed to calculate profiles for many spectrum lines for a given model atmosphere, allowing fully for the depth-variation of an atom's excitation/ionization state, of line broadening constants and the Doppler velocity, and of the background continuum opacity. The codes were constructed so as to fit automatically the observed equivalent width of every line it was given and to derive an abundance estimate from each of the lines, which could then be averaged according to a prechosen weighting scheme. More sophisticated versions of this procedure included checks to see if the abundance derived from two (or more) ionization stages of the same element all gave the same result. If not, a new model could be chosen which was hotter and/or less dense, or the opposite. Likewise one could check the variation of deduced abundance with observed line-strength to see if it was necessary to introduce some ad hoc "nonthermal" velocity field in order to remove any correlation found. With these tools in hand quite a bit of new abundance work was done. Thanks to the increased precision achieved it was possible to discriminate relatively mild differences with fair confidence. For example, it was possible to show that Sirius had noticeably abnormal abundances relative to Vega, even though both exhibit what appears (at least to a casual inspection) to be a reasonably normal spectrum for its type. Moreover, attempts were made to analyze some stars which were much more exotic than hitherto attempted, for example the peculiar A-stars (Ap) and the metallic-line A-stars (Am). Not only did it emerge that each of these groups had characteristic abundance anomalies, but it was possible to demonstrate fairly convincingly that the atmospheric structure of the Ap stars is dominated by intense magnetic fields, and that elements are distributed nonuniformly into "patches" over the stellar surface. Likewise the first steps were being made towards unraveling the spectra of peculiar stars on the giant branch, and understanding the cosmochemical significance of their abundances. This is the subject of Professor Lambert's talk, so I will not pursue it further.

In the 1970's one can see at least three major lines of development of the theory. First, line-blanketed LTE models were pushed to a very high degree of realism as the quality and completeness of atomic data continued to improve (in some notable cases as a result of the efforts of the same people as were computing the models). Models were constructed with literally millions of spectral lines representing the first 3 or 4 ions of all the astrophysically abundant elements up through iron. In some cases extensive line lists from diatomic molecules were included. With the computing power available at that time (or even now, for that matter!) it was impossible to compute detailed profiles for all lines. Instead a couple of ingenious statistical schemes were devised to reduce the labor of the computation while preserving the final accuracy of the calculated energy distribution. In the *opacity distribution function* (ODF) *method*, one divides the whole spectrum into a fairly small number of bands (the goal is to use as large a band as possible while preserving accuracy). Then the total opacity (including the sum over all lines) is computed on a discrete frequency grid. Next, on the assumption that the band is narrow enough that the exact position of a line in the band doesn't matter, one rearranges the grid so that the opacity becomes a monotonic function of wavelength, resembling a "fat line". This smooth distribution is then sampled at a few representative points, and the calculation is done for each of these. The resulting reduction of the total number of wavelengths by two or three orders of magnitude makes the computation cheap enough to actually perform, while the effects of lines on both the emergent energy distribution and the temperature distribution in the atmosphere are still represented with good accuracy. Alternatively, in the *opacity sampling* (OS) *method*, much in the spirit of a Monte Carlo calculation, one simply chooses a wavelength at random, computes the total opacity at that point, and solves the transfer equation for the radiation field. As the number of sample points increases, each kind of point – pure continuum, weak line, strong line – is sampled with the correct frequency, in all parts of the spectrum, and the calculation converges to a stable result. Again, savings of several orders of magnitude are possible.

Comparison of the computed results with observed energy distributions has been extremely encouraging. For example, Kurucz has fitted the energy distribution of Vega very closely, and we can now say that we know the effective temperature and surface gravity of that star to quite acceptable precision. More recently he has been able to fill in enough missing atomic data in the ultraviolet to achieve quite a respectable fit of his calculations to the observed solar spectrum, a nontrivial achievement! Likewise it is now possible to calculate accurate photometric colors from the theoretical energy distributions, and thereby obtain reliable relations between colors on the one hand, and effective temperatures and gravities on the other. Further, predicted values of ultraviolet fluxes, which are essential in understanding the energy and excitation balance in the interstellar medium, are becoming trustworthy, a point of importance, e.g., below the Lyman limit where interstellar absorption forever prohibits our getting observed data.

In a second major thrust, efforts were made to remove the assumption of LTE, and to compute models in which the excitation/ionization state of the material is consistent with the radiation field it produces. With the advent of the CDC 7600 it was clear that we had enough computational power to handle at least the most basic features of the problem. As mentioned earlier, the primary reasons the material goes out of LTE are: (1) densities, therefore collision rates, are very low in the outer layers of a star, hence radiative rates dominate in determining the excitation/ionization equilibrium; and (2) because of the optically open surface of the star, and the presence of temperature gradients, the radiation field departs from the equilibrium (Planckian) distribution function. In reality the two effects are coupled and feed one another: the radiation field, which results from the material's absorptivity and emissivity, in turn determines the excitation and ionization of the material, hence its absorptivity and emissivity. To make the problem worse, it is easy to show that because of the dominant scattering component in NLTE source functions, photons are not absorbed and thermalized after each free flight, but rather only after some huge number of scatterings. This results in an essential delocalization of the radiation field, and implies that the interaction volume within which the radiation field at one point can affect the material state at some other point, is generally also huge. A corollary is that effects from the boundary, hence departures from LTE, can penetrate very deep (a full thermalization length) into an atmosphere. Moreover, the radiation field in any one line influences not only the two levels it couples, but, because a change in that line will perturb the rate matrix, it influences all other levels in the atom as well, and thus the radiation fields in all other lines in the entire transition array of the atom. Therefore all lines in the transition array are more or less strongly interlocked, and photons may be more or less freely shuffled back and forth among them. This fact was recognized 30 years ago by Jefferies, who pointed out that any particular photon does not have a unique "identity", but in reality belongs to a collective "photon pool". Finally, over and above these couplings, there is a global coupling across the entire spectrum via the constraint of radiative equilibrium.

It was clear that a powerful new algorithm would be required to overcome these multiple difficulties. After some experimentation Auer and I devised the »complete linearization« method, which like the Henyey method of stellar interiors, solves linearized versions of the original nonlinear equations iteratively. Mathematically the method essentially amounts to a multidimensional Newton-Raphson procedure, and converges quickly if the original solution falls within the domain of convergence. In more physical terms, the method accounts fully (to first order) for the effect of a change in any variable at any point in the medium on any other variable at any other point. Further, it allows free redistribution of photons within the transition array (thus realizing mathematically Jefferies's photon-pool idea) and over the entire spectrum (as implied by radiative equilibrium). The method thus appeared to have several promising features, and experience soon showed that the promise would be met in reality. Application of the method was not without difficulties; for example it was very hard to find adequate collision cross-sections for even hydrogen and helium, the two simplest atomic systems (and the situation has not improved much even today).

The algorithm we implemented was a direct (brute force) solution, and consequently was quite expensive. Therefore we were able to make only simplified models. We did a survey of O and B stars, including opacity from H and He, and allowing only 6 spectrum lines for hydrogen, and one each for He and He+. Despite the primitive nature of the model atoms we were able to show that departures from LTE in O-star spectra were major (essentially because of the intense radiation fields at those temperatures). In particular we showed that LTE models predicted a spurious weakening of the H lines, and led to absurdly high gravities if a fit were forced to observed equivalent widths. Likewise, LTE models predicted too-weak He lines, which would lead to spuriously high He abundances if we forced a fit to observed equivalent widths. The NLTE calculations removed both difficulties at a single blow. As modest as that achievement seems now, it was the first demonstration that a major input of new physics led to markedly better answers (a comfort to those of us who believe that good physics is the price for good answers).

The calculation of the NLTE spectrum of an impurity atom in a given atmosphere is a great deal simpler than construction of the atmosphere itself. Thus it was easy to adapt the new method to a study of abundance anomalies, and the formation of emission lines. The hardest part of setting out on one of these projects was to find adequate atomic data, and also to choose a model atom that is complete enough to treat all the important transitions explicitly while remaining small enough to fit into the constraints of machine speed and memory size. Calculations of the He I spectrum of the B-stars showed that lines located in the traditional blue-violet region of the spectrum were practically unaffected by departures from LTE, thus validating earlier LTE abundance analyses based on these lines. In contrast the lines and the yellow to red regions of the spectrum were predicted much too weak by LTE, and the new calculations removed much (but not all!) of the discrepancy with observation. The different behavior of lines in the two different spectral regions can be understood by a very simple physical argument. A similar analysis of the Ne I spectrum in B-stars showed qualitatively the same effects. But now the strong lines which were used for abundance analyses are in the red spectral region, and the too-weak lines predicted by LTE leads to spuriously high abundances. In fact it turned out that when the NLTE analysis was done the abundance dropped from 5×10^{-4}, a factor of 5 larger than the accepted solar/cosmic value, right down to 10^{-4}, the cosmic value. Another mystery solved. And similar results for Si III/IV, and a few other ions.

For the O-stars an analysis of the Mg II line at $\lambda4481$ showed major NLTE effects which had led to an overestimate of the Mg abundance in O-stars by a factor of 10. A more interesting result was that it was possible to reproduce the N III emission lines at $\lambda\lambda4634,40,41$ as seen in stars classified as O((f)) by Walborn. The two puzzles were why these lines, transitions from $2s^2\ 3d\ {}^2D$ to $2D^2\ 3p\ {}^2P$ are in emission, unlike any other line in the (visible) N III spectrum, while $\lambda4097$, the next line in the nominal cascade sequence $2s^2\ 3p\ {}^2P$ to $2P^2\ 3s\ {}^2S$ remained in absorption. The resolution of these puzzles was twofold: (1) It so happens that a double-excitation state $2s2p({}^1P)\ 3d\ {}^2(P,D,F)$ of N III exists just above the lowest state $2s^2\ {}^2S$ of N IV. Because the double-excitation state is only a fraction of a thermal energy width above the continuum threshold for N IV, electrons can efficiently recombine into that state by dielectronic recombination. Then the doubly-excited state stabilizes via a $2s2p\ {}^2P$

--> $2s^2\ {}^2S$ transition of the inner electron, leaving the system in the $2s^2\ 3d\ {}^2D$ state of N III, just where it is needed. The process is efficient enough that the upper state of the emission lines is sufficiently overpopulated to produce the observed emission. (2) The next step is that after 3d electrons decay to 3p, they do not preferentially decay to 3s, but instead to the $2s\ 2p^2\ {}^2(S, P, D)$ states. These *two-electron jumps* normally unimportant, provide the essential drain from $2s^2\ 3p$. The reason is that by another accident of nature a $2p^3\ {}^2P$ state, which *can* decay directly to $2p^2$, lies at almost the same energy as the $3p\ {}^2P$ state, so that the two states become very strongly mixed. A similar mechanism also works in C III.

While the work just described certainly yielded a substantial number of results of astrophysical interest, it had to be unwillingly brought a close for two reasons: First, to attack more complicated atomic systems, or even do the ones originally surveyed definitively, one would need to include many more atomic levels and transitions. With the existing algorithm that would have required a large increase in computer power. Alternatively we would have needed a more efficient algorithm. Both of these conditions have been met by now. Second, the underlying models were simply not good enough to trust for the analyses contemplated. In particular, the available NLTE models were unblanketed, which implied that they would give unreliable photoionization rates, particularly in the ultraviolet; these, unfortunately, entered the calculation in a sensitive way. On the other hand it would have been fatal to use LTE models because then a spurious thermal radiation field in the lines and in major photoionization transitions would artificially drive the atomic system under analysis towards LTE. The only solution seemed to be to face the NLTE line-blanketing problem for thousands of millions of lines. And that was a daunting prospect! Indeed, at the time it appeared impossible; nevertheless, as discussed below, we now stand on the threshold of this very achievement.

In a parallel effort, stimulated mainly by the desire to analyze solar chromospheric lines, powerful techniques were developed to treat partial redistribution (i.e. partial coherence) effects on resonance line profiles (a topic Strömgren had touched upon in the 30's). In brief it was found that photons emitted in the line core are essentially completely redistributed over the core, while photons emitted in the wings are emitted almost coherently. This coherence produces significantly different absolute intensities in the wings compared to profiles computed assuming complete redistribution over the entire line profile. A result of this change is that the temperature minimum in the solar chromosphere as inferred from strong resonance lines comes out a few hundred degrees cooler, and in essential agreement with estimates made from far-infrared data. This point is interesting because the semi-empirical temperature turns out to be *lower* than the pure radiative equilibrium value, which is counterintuitive because one expects the minimum to be a the region where shock heating of the chromosphere first becomes felt. Only when full dynamical models of shock trains in the solar atmosphere became available did we realize that radiative losses from the hot compressed material in shocks outweighs the radiative gains in the rarefaction phases, leading to a net energy loss, hence cooling of certain regions of the atmosphere!

A third line of investigation in the 70's, spurred by observations of high-velocity winds from early-type stars, was the development of techniques for treating lineformation in expanding envelopes. Sobolev had, of course, already contributed his brilliant escape-probability method, which works extremely well in relatively rapidly expanding flows (i.e. where a line shifts by about a line-width over a photon mean free path). But a full transfer treatment is indicated for trans-sonic winds where velocities range from quite small to very large. All the early work assumed planar geometry (clearly inappropriate for a wind), and formulated the problem in the laboratory frame (with the exception of a remarkable paper by McCrae and Mitra back in the 30's which solved the transfer problem in the comoving frame of the material). In the laboratory frame one must keep track of photons in a frequency band twice as wide as the frequency shift produced by the maximum flow velocity, so the method is practical only for low-velocity flows, and not too useful for winds.

In the comoving frame one needs to treat only the frequency band actually covered by line absorption, which is a decisive advantage. The differential operator becomes more complex (it contains a frequency derivative) in this frame, but the problem can be discretized and solved by techniques appropriate to partial differential equations. Solution of the equations yields the scattering term in a two-level-atom source function. To treat multi-level atoms one replaces the complete set of coupled transfer and statistical equilibrium equations by a sequence of equivalent-two-level-atom (ETLA) problems and iterates the entire set of them to self-consistency. The simple ETLA iteration scheme works well in an expanding atmosphere (or at least a lot better than for a static medium) because the medium is expanding, so that expansion-induced photon escapes outweigh the reshuffling of photons via interlocking of lines. One thus obtains a powerful and general tool for computing the spectrum from a multilevel, multi-ion medium in trans-sonic spherically-symmetric expansion. This methodology has been used for a number of investigations, including analysis of the impressive P-Cygni lines from several elements in the ultraviolet spectrum of O-stars, and He II $\lambda 4686$ in the visible.

In the 80's there has been a great resurgence of interest in creating efficient algorithms for solving various kinds of transfer problems, NLTE statistical equilibrium problems, and constructing model atmospheres. These new methods, coupled to the new generation of high-speed, large-memory machines (CRAY, ETA) make possible the rapid solution of hitherto unapproachable problems. Many of these developments are summarized in the two books "Methods in Radiative Transfer" and "Numerical Radiative Transfer" edited by Kalkofen (1984, 1987). The topics discussed include fast methods of solving the transfer equation, radiative transfer in spherical media, operator perturbation techniques, and transfer of polarized radiation. The "operator perturbation" methods ultimately derive from the elegant ideas of Cris Cannon, although they are now often cast in quite different mathematical form. Most current versions are based on an approach devised by Scharmer (1981). The basic idea is to use an approximate solution, which is then iterated to convergence. The method is easily generalized to NLTE line formation in moving media (Scharmer 1984), and to multilevel problems (Scharmer and Carlsson 1985).

Scharmer's method was modified by Werner and Husfeld (1985) to solve large statistical equilibrium problems, with up to 100 atomic levels in NLTE. And Hamann (1985, 1986, 1987) adapted it to solve statistical equilibrium problems for multilevel atoms in spherical expanding envelopes. By adding constraints of hydrostatic and statistical equilibrium Werner (1986, 1987) arrived at a very efficient scheme for constructing model atmospheres for material represented by model atoms having up to 100 levels. With this method Rauch and Werner have been able to evaluate the effects of various numerical approximations and assumptions about model atoms on the structure of, and line profiles from, NLTE stellar atmospheres. We now finally know how many atomic levels, angle-quadrature points, frequency-quadrature points, etc. are required in order to fit the high S/N data that we can obtain with modern spectrographs and receivers to its full accuracy. On the whole, it is amazing how well these methods work, and how large a speedup they yield. One senses that developments in this direction have not yet been exhausted, and that they hold much promise for the treatment of dynamical atmospheres.

A different kind of scheme has been developed by Anderson (1985, 1987) for computing lineblanketed NLTE atmospheres. In this ingenious method frequencies in all the lines and continua are cleverly regrouped into a small number of "blocks". As a result, the number of variables to be determined in the complete linearization scheme is much smaller, and therefore each iteration is relatively cheap. It is now possible to treat literally thousands of lines simultaneously, thus solving both the NLTE line blanketing problem and the statistical equilibrium problem for all interesting ions at the same time! And soon it may even be possible to do such calculations on a typical virtual memory workstation! This method is really a breakthrough, and may make it possible to produce large grids of NLTE line-blanketed models in the near future. Of course the construction of such models implies that we shall need huge amounts of atomic data: (1) continuum photoionization cross-sections, (2) line strengths, (3) line broadening parameters, and (4) collisional excitation and ionization rates. Work on items (1) and (2) continues at Los Alamos and Livermore, work on items (1)-(3) is being done by the UK/US opacity group, but work on item (4) still needs to be done. It is difficult to predict when all the data needed will become available. (Lack of adequate atomic data is also a problem Strömgren encountered in his work on stellar opacities and in constructing stellar interiors and atmospheres models!).

III. *Improvements in Instrumentation*

It goes almost without saying that the inferences we make from a theoretical structure, no matter how grand, can be no more trustworthy than the data we analyze are accurate and complete. The improvement in effectiveness of modern techniques for collecting spectroscopic and photometric data, compared to the methods of, say, 30 years ago, has had as large an impact on our knowledge about stellar atmospheres as the advent of computers has had on theory. One of the really important events has been the arrival of truly linear receivers of excellent stability and great sensitivity (e.g. CCDs and Reticons), which, in addition, produce digital data directly. To understand the importance of this event one needs to recall the arcane procedure used to reduce photographic spectra: to obtain the data one had to microphotometer the plate, and then (unless some kind of special electromechanical device had been constructed to do the procedure automatically as the plate was scanned) apply the nonlinear transformation between density and intensity, sometimes by hand. The whole procedure was slow, tedious, and error prone. Nowadays one can put the data tape from the telescope directly on a computer and do the whole reduction procedure in a few seconds. The nonlinearity of the photographic plate and the difficulty in obatining an accurate calibration often led to systematic errors of the order of 10 or 20 percent. Such errors can introduce serious errors in any subsequent analysis. Modern measurements yield accuracies of about 0.1 %

Modern receivers have incredibly high quantum efficiencies in their most sensitive spectral regions, and now cover a range that was largely unavailable to workers of 30 years ago. Actually, the sensitivities are already so good that it is probably not realistic to look for major improvements in the near future. In addition, new spectrograph designs, particularly with echelles, make available vast stretches of the spectrum in a single exposure, whereas before one would have needed to make multiple time-consuming exposures. And these new spectrographs are typically being fed by larger telescopes. This is an area where we may indeed see large improvements in the next few years as mirrors of huge dimensions are constructed. Moreover, we do not require images of excellent quality for this kind of work: the telescope need only operate as a huge light bucket. And even here we may learn new tricks, for example multiplexing by using optical fibers to pipe light from several sources into several spectrographs simultaneously. And finally, the availability of digital data and the speed of modern reduction and display facilities makes real-time display possible. That can be used as a tool to save the most precious resource of all, the astronomer's time, by letting us know when the exposure is adequate (or warning us to stop if, say, we have pointed at the wrong object!).

It is probably hard to overemphasize the importance of the new wavelength bands now accessible to a stellar astronomer. The different picture we get of what a stellar spectrum looks like at ultraviolet, infrared, and radio wavelengths, compared to the visible, truly challenges (which is a polite word for "destroys") the usual paradigm upon which we have based most of our theory. In the ultraviolet we see intense chromospheric/coronal emission lines in late-type stars, and massive, high-velocity winds in early-type stars. So we must admit the need for nonradiative heating/cooling in the outer atmospheric layers, and must find acceleration mechanisms for the observed flows. Similar challenges arrive from the infrared and radio bands. The point is that in these bands we are typically observing the outermost layers of the atmosphere, where conditions may be radically different from those relevant in the visible. Instead of essentially static, homogeneous layers, the structure of the medium may be quite inhomogeneous, perhaps structured by velocity and magnetic fields, and likely subject to various kinds of (magneto)hydrodynamic heating/cooling phenomena. These views are so very different that they raise the question of "just what is a stellar atmosphere anyway?", a point to which we return at the end of this talk.

IV. *Magnetic Fields*

I would now like to say a few words about magnetic fields in stellar atmospheres. It is a subject I don't know much about, so I can be (mercifully) brief. From observations of the Ap stars we know that at least some stars have very intense, highly-organized, global-scale fields that most probably completely dominate the atmospheric structure. However, no one, to my knowledge, has put together a self-consistent model of such a stellar atmosphere, including the radiation field. One of the reasons this has not been done is that even for a static atmosphere the incredible richness of possible solutions admitted by the underlying Maxwell equations and hydrodynamic equations (even without nonlocal radiative transport!) is so great that we literally don't know quite where to start. We need at least some idea of the shapes and scales of the features we are required to model. And unless we someday get some kind of spatial resolution of a stellar disk, we are likely to be stalled for a long time because the parameter space to be searched is simply too large, and there is no guarantee of a unique result.

The situation for the Sun is quite different. Here we already have enough spatial resolution to see magnetic fields on a variety of scales. Some of what we see is quite surprising. Since the turn of the century people have known about the intense bipolar fields in and around sunspots. With the advent of the coronagraph and birefringent filter it became possible to observe spicules (jet-like surges apparently confined within a magnetic flux tube), active prominences (where we see streams of fluid motion along magnetic loops and arches), and quiescent prominences (material suspended in, and shielded from, the intensely hot corona by magnetic fields). In the early 60's Leighton discovered the supergranulation network: a network of magnetic field that outlines large-scale (30,000 km) convective flows. And magnetograms measuring magnetic fields with a resolution of a couple of seconds of arc (1500 km) became routine. By ingenious work in the 70's it was finally proven that the "general" magnetic field of the Sun actually consists of very thin (< 100 km), unresolved flux tubes, having fields of the order of 1500 gauss (almost as strong as a sunspot). These flux tubes tend to occur in +/− pairs in close proximity, so that a measurement in a large region averages over many pairs, and the "general field" of the Sun on a global scale appears to be only a few gauss. With space observations we discovered gigantic, magnetically-controlled, surges of material (superspicules) and explosive ejections of material from the corona. We also see that the corona is dominated by closed magnetic arches containing very hot, dense material, whereas the wind originates in cooler, less dense open-field regions. Presumably all of these facts also apply to other stars, at least those with gross properties roughly like the Sun. And in fact observations of solar-type stars do reveal regions like sunspots and solar plage.

Several important conclusions can be drawn from these observations. First, the higher we look in the atmosphere, the more the material is structured by the magnetic field, the more inhomogeneous a radial shell of material becomes (as we sample across different nonspherical structures), the more dynamical its behavior, and the more concepts like hydrostatic or radiative equilibrium become irrelevant. We do *not* see an evolution from a smooth photosphere to chromosphere to corona. The chromosphere is very inhomogeneous (in the sense that a photon mean free path can span several different kinds of structures). The transition region between the chromosphere and the corona is positively ragged, having very steep gradients and a great deal of small-scale structure. The corona is also quite inhomogeneous, and structurally dominated by magnetic fields.

Thus the radiation field we observe is actually the result of some very complicated nonlinear averaging over regions with extremely different physical properties. This fact not only complicates the analysis of the solar spectrum but has ominous implications for the analysis of stellar chromospheric and coronal structures, which can certainly be expected to be every bit as inhomogeneous as in the Sun. The problem is that we haven't the faintest idea what those structures look like! Magnetohydrodynamic modeling of the solar chromosphere and other magnetically dominated structures in the solar atmosphere is currently underway. At present it is quite oversimplified, though more and more realistic treatments are surely coming, because in the Sun we can at least see the structures we are trying to model, and can tell when our models do in fact fit the data. But we have essentially no guidance for other stars, and I personally believe that we should consider our present stellar chromospheric models as being only a *very rough* caricature of reality. Even stellar photospheres may not be "safe". For example, although a 1500 gauss field in flux tubes is quite strong enough to have serious structural and energetic implications in a star's photosphere, we wouldn't even be aware of its existence in other stars where we have no spatial resolution and can infer, at best, only global averages.

In short, we must always keep in mind that our usual theoretical paradigm for a stellar atmosphere may fall far short of reality, and that in one star we may actually have several quite distinct "atmospheres". These are difficult problems, and we may have to content ourselves with partial answers for a long time!

V. *Dynamics*

In some stars the dominant phenomenon seems to be ordinary hydrodynamics (i.e. not magnetohydrodynamics). A very basic pattern observed in the Sun (and therefore probably present in solar-type stars) is granulation: individual time-dependent convective cells that rise into the photosphere, radiate, and die, typically in the form of an "exploding" granule. Positioning a slit on an image of the solar surface one observes "wiggly" spectrum lines, each line responding to the quasirandom up-and-down motions of the granules. For a solar-like stellar atmosphere, where we have no spatial resolution, we would observe the spatial average of these profiles, and would get a broadened line profile. This broadening has customarily been called *microturbulence*, a terminology that has often sparked intense arguments. The appellation "micro" is appropriate in the sense that the individual fluid elements are of the same order of size as a photon mean free path. However the quantity being measured is certainly in no sense what a hydrodynamicist means by "turbulence", but is simply the ensemble average of what may actually be completely laminar flow. The distinction becomes critical as the inferred "turbulent" velocities approach the speed of sound. One of the most impressive accomplishments is solar/stellar physics in the past few years has been the development of hydrodynamic codes that accurately model the convection in 3D, and allow for radiative transfer, and magnetic fields (Nordlund 1985, Stein and Nordlund 1989). These computations yield fairly realistic maps of both the intensity fluctuations and velocity fluctuations typically seen in granulation. This work has already greatly deepened our understanding of the solar atmosphere, and similar work is likely to add tremendously to our knowledge of the atmospheric properties of other stars.

On larger spatial scales one sees the solar atmosphere oscillating, with about a five minute period, in an extremely complicated and rapidly changing spatial pattern. We understand these motions today as the result of a superposition of thousands (millions?) of high-order nonradial pulsation modes of the solar envelope, although we do not have a good picture of what acutally excites and drives them. In the solar case we can identify the modes because we have spatial resolution and can carry out the required spatial transforms of the data; but this information is again unavailable for stars. It may prove possible ultimately to detect a few low-order modes, but in the meantime all we observe for stars is something called *macroturbulence* which presumably represents flows on scales much larger than a photon mean free path, changes the shape of a line profile without necessarily changing its strength. Sensitive fourier transform techniques applied to high signal-to-noise profiles now permit a dissection of the stellar velocity field into rotational, microturbulent, and macroturbulent components. But it must always be borne in mind that "micro" and "macro" "turbulence" give us at best only two characteristic points on a distribution function, and certainly do not provide enough information for unique hydrodynamic modeling.

Regular radial pulsations of stars driven by the He+ ionization zone occur in a fairly narrow strip in the H-R diagram, and a variety of semi-regular and irregular pulsations occur in late-type giants and supergiants. The underlying mechanism, an oscillation of the ionization front inward and outward, is by now well understood, and models of these phenomena can be used to place useful constraints on the theory of stellar evolution. However the atmospheric effects of the pulsations and the outward propagating shocks they drive have been relatively little studied theoretically. Part of the reason is that present atmospheric models of pulsating stars aren't very good. The codes are almost always Lagrangean and use the diffusion approximation in treating the radiation flow. As a result of coarse zoning and inadequacies of the diffusion approximation the computed emergent flux has large spurious "bumps" and "wiggles", and is not too reliable. These are not the result of numerical instabilities, but of a failure to resolve, on the computational mesh, critical structures such as the emerging shocks and the ionization front. Until the time that this difficulty can be overcome, we will not get even accurate light curves from the models, not to mention spectra! That's a pity because good spectra, showing fascinating shock phenomena such as line splitting and emission, have existed in the literature for over 25 years. But it is no surprise that the problem has not yet been solved because it would require a full NLTE treatment in a dynamical atmosphere (in which the important structures had been resolved!). That is a rather daunting prospect, yet I believe that it is within reach within the next decade provided that we do solve the computational problems with the dynamics.

One of the most interesting and successful developments of stellar atmospheres theory over the past 15 years has been the working out of a theory of radiatively driven winds, which are observed ubiquitously in high-luminosity stars. It was known from the outset that radiation forces on the continuum opacity alone could not drive a wind. Lucy and Solomon (1970) showed that direct radiative momentum deposition in strong resonance lines in the ultraviolet could, in fact, induce a transsonic flow, and Castor, Abbott, and Klein (1975) quickly showed that the observed mass-loss rates could be matched when one accounted for hundreds to thousands of lines in the spectrum. The first version of the theory has been elaborated and refined to include effects of the rotation and finite angular size of the star, a complete line spectrum, improved ionization balance, line overlap, and multiple scattering of photons back and forth across the volume contained within their successive resonance surfaces. This last point has turned out to be particularly important, because a photon has a very large energy relative to its momentum (compared to massive particles), and thus it can be "robbed" of its full momentum many times before its energy gets "used up". Thus photons can actually multiply their driving effect many-fold, and calculation shows that even the large mass-loss rates of the WR stars can be driven in this way. A related phenomenon is that because the expanding envelope can backscatter excaping photons back into the stellar photosphere where they thermalize, the star suffers a kind of "wind-blanketing" and backwarming which may have important consequences for the visible spectrum. This effect is now known to be so large as to have a major impact on the derivation of effective temperatures, surface gravities and abundances from O-star spectra.

I think that most people would agree today that CAK theory provides a pretty good basic description of the supersonic part of the flow, in particular of the acceleration mechanism that produces the high observed terminal velocities. Nevertheless there exist at least two outstanding problems that are being worked on actively at present: (1) The energetics in CAK type models is rudimentary; indeed the radiative scattering is assumed to be strictly conservative, and the gas is therefore adiabatic (or forced to be isothermal by fiat). This is probably not a serious problem for the supersonic flow region, but may be of much greater, even dominating, importance in the subsonic flow. And it should be remembered that it is the subsonic flow region that determines the mass loss rate, because once the flow passes sonic, no amount of energy or momentum input can alter the mass flux (only the terminal velocity). Accurate modeling of this flow regime, including an accurate treatment of the radiative transfer, full energetics of the gas, rotation, and perhaps magnetic fields, remains to be done, and may occupy us for some time. (2) Another important aspect of radiatively driven winds is that they are unstable. Small disturbances in the flow rapidly grow as they are swept downstream, and produce strong shocks. The interaction between radiation forces and the shocks seems to be reasonably well understood. The shocks are strong enough to produce very high temperature plasma that can emit soft X-rays. Early attempts to explain the X-rays in terms of a thin corona failed, and pointed to a source distributed throughout the wind, in agreement with the present picture.

Supernova envelopes are another example of coupling between radiation and hydrodynamic flow. The strong shock created in the bounce of the envelope from the collapsed core is supercritical, and strongly dispersed by radiation. At the time of shock breakout and unloading an intense burst of radiation emerges. Current codes describe these phases rather crudely, using, at best, some kind of flux-limited non-equilibrium diffusion, and usually not event that. A proper transport treatment of the radiation, taking into account that the opacity in the envelope is typically completely dominated by electron scattering, would be fruitful, and seems easily within reach. Modelling of supernova spectra is considerably more well advanced. At present the codes solve the full transport problem for an expanding spherical envelope, using a relativistic formulation that allows one to account for all velocity-dependent terms. In general one obtains quite convincing fits to the observed spectra, and it has been possible to infer both physical conditions and element abundances in the ejecta. An unusual complication, not normally met in ordinary stars, that enters is that element abundances in the envelope may be stratified because of the effects of thermonuclear reactions. At later stages a supernova envelope becomes very distended, and NLTE effects must become quite important owing to both dilution of the radiation field, and low densities. A few exploratory investigations have been done thus far, and this looks like a good area for rapid development in the future.

The situation with novae is much less satisfactory. Modeling a nova spectrum is extremely difficult because of the complicated geometry of the emitting medium: two stars, a disk, an expanding envelope, and perhaps other structures. There are no easy, but good, approximations: The material certainly does not have a 1D symmetry. A 2D calculation (difficult enough) fails because a disk in the system precludes axial symmetry around the axis joining the stars. So nothing less than a full 3D calculation is needed, and this appears to be beyond our present capabilities. Some novel Sobolev type treatments have been developed in the USSR; these may fill the gap until a full transport method becomes available. It will be interesting to see how much progress can be made on this problems in the next ten years.

VI. *Geometry*

We usually take for granted that we know the basic geometry of the objects we are trying to model, and we tend to forget how restricted our methods really are. As often as not our assumption is not really true, and it seems worthwhile to mention here a few points about global geometry. The solution of 1D transfer problems, whether in planar or the spherical geometry, is now well understood and needs little comment. One aspect of the spherical case worth mentioning is the question "What is a stellar radius?". In the planar limit the thickness of the atmosphere is so small compared to the stellar radius that the definition of the radius is unambiguous. But in a very distended envelope the radius of the surface of unit optical depth in the most transparent continuum may be a small fraction of the surface of unit optical depth in a strong line (or continuum edge). Consequently the size of the emitting surface of a star can be a strong function of wavelength, and it is not at all obvious what the usual stellar interiors convention of choosing the stellar radius to be the radius of the surface of Rosseland optical depth = 1 (or 2/3) means. It has always surprised me that those who make models of giants and supergiants, have not attempted to eliminate possible ambiguities and problems by using detailed model atmospheres for their outer boundary conditions.

For a rotating star we expect some kind of oblate spheroidal shape. We thus have

2D axial symmetry, and methods for treating radiation transfer in that symmetry exist. (At present they are rather clumsy and expensive, but it is reasonable to expect significant improvement over the next few years). These methods might also prove useful for, say, single (proto)stellar objects with bipolar jets, with or without an accretion disk. Axial symmetry would also be appropriate to weakly interacting binary stars if each star has a rotational symmetry around the line joining their centers (e.g. the stars are prolate spheroidal). Likewise for any diffuse matter in the system. But as the members of a binary get closer and interact strongly, the stars become distorted, and they become immersed in a common envelope of unknown shape. In addition, as sketched in bewildering detail in Otto Struve's old book "Stellar Evolution", there may be a considerable amount of radiating gas in disks, spirals, and streams. We are still a long way from being able to calculate the radiation field from such geometrically complex objects, except possibly by Monte Carlo methods. And even if that were not the case, we would still have the deeper problem of knowing what structure to choose to model!

VII. *Spatial Resolution*

I have alluded several times above to the problem that we have no spatial resolution for stars and therefore cannot choose a theoretical structure uniquely. Indeed, except for the Sun we cannot really answer even the question "What does a stellar atmosphere actually look like?" The things we would like to know are quite basic: We would like to know what the limb-darkening is for a given star, especially giants and supergiants, and most late-type stars. We want to know whether we can see convective patterns like granulation and supergranulation. We would like to know if there are sunspots, flux tubes, and magnetically-dominated active regions. We would like to know if the star has a chromosphere and corona, and, if so, how inhomogeneous they are. We want to know if the star has analogues of spicules and prominences. We want to know what the huge regions, which appear to be gigantic "starspots", on RS CVn stars really are. We would like to know what the magnetic field structures in an Ap atmosphere are, and what the regions of differing element abundances are and imply. We want to be able to detect nonradial pulsation modes. And we need even gross geometric information about the distribution of gas in close binary systems.

What are the prospects for ever getting this information? One immediately thinks of interferometry. Until now this technique has yielded only stellar diameters, which one might regard as the "zero-order" term of limb-darkening. At the present there are projects underway to improve greatly the amount of information we can recover from speckle interferometry, and to build a powerful new Michelson interferometer. Some of the information described above requires only a few resolution elements on the stellar disk, but the rest requires many. Even though I am hardly an expert on stellar interferometry I strongly suspect that we will not obtain the kind of information needed from ground-based observations because of the problem of overcoming seeing effects at long baselines. Beyond ground-based interferometry lies perhaps a huge telescope or a gigantic interferometer in space or on the moon. Such an instrument would be incredibly costly, and would take a very long time to build, even if it were ever approved. So at the risk of sounding unduly pessimistic, I would guess that at a meeting like this held, say, 20 years from now, we will have little, if any, *direct* new observational information about the structure of stellar atmospheres, and would still be talking about the same problems.

Of course there are some indirect techniques, but these have only a limited capacity to obtain the kinds of information we want, and always require considerable reduction and interpretation. In addition it may be possible to get some information, for example about granulation-like inhomogeneities, from numerical simulation, but then one is introducing theory into the process of deciding what kinds of structures must be considered by the theory, which may lead to circular reasoning. In short, it seems to me that we are not likely, with a high degree of confidence, to make major improvements in the structural assumptions underlying the models and that we will still be working with highly idealized and oversimplified models for some time to come. I therefore think that it will behoove all of us (especially the model builders!) to remember the significant limitations of our models, and to use the models with caution in areas they are unequipped to address.

VIII. *Prospects*

It seems appropriate at this point to adress briefly the questions "How good is our modeling, and where do we go from here?". I think that it may be fair to say that the quality of a model is, like beauty, in the eye of the beholder. Those who have developed the numerical techniques that have moved us so far forward in the past two decades can rightly be proud of the fact that we can now actually solve many problems, where before we could only make rough guesses. But those aware of all of the physics left out of the formulation, and those who have looked at solar observations long and hard in order to learn what a "typical" star might look like, will quickly point out the flaws. I would like to steer a middle course.

First of all, the place where we do best is near (or below) the main sequence, where stars have moderately compact atmospheres. I think it is safe to say that we essentially understand continuum and line formation in the photospheres of normal upper main sequence stars, at least to the extent that purely radiative models are appropriate at all. Thus Kudritzki and his coworkers, for example, have been able to demonstrate a very close correspondence between the computed spectrum of O-stars and high quality observations. It now appears that we can deduce effective temperatures, gravities, and abundances for these stars and can model their ultraviolet spectra with good reliability. On the other hand, it has long been known that most upper main-sequence stars show a considerable amount of "microturbulence", "macroturbulence", or both, in their atmospheres, along with rapid rotation. There is no question that hydrodynamic motions occur in the atmospheres, but we do not yet know what they are (waves? nonradial pulsations?), nor do we know whether they play an important role in either the energy or momentum balance in the atmosphere, or create inhomogeneities that have a significant effect on the radiative signature we receive. Unfortunately there doesn't seem to be much work being devoted to these important questions at present, and so it is impossible to guess when we might learn the answers.

We also seem to be able make qualitatively correct models of OB-star winds, and to estimate mass loss rates, though we certainly do not yet have self-consistent models including instabilities and shocks. And recently Hamann (1985) and Hillier (1987a, 1987b) have been successful in fitting the spectra of Wolf-Rayet stars, and in deriving a semiempirical model of a WN envelope. Nevertheless we must remember that the structure of the subsonic flow region has yet to be worked out, and the impact of nonradiative energy inputs (e.g. nonradial pulsation) from the interior of the star has yet to be understood. For all we know now, the ultimate driving force behind the wind might be photospheric and subphotospheric motions (as argued by Thomas). Things are much worse for the Be stars. Not only is there some kind of ill-defined flow (disk-like?) whose morphology we don't know and whose cause we don't really understand, but there is ample observational evidence by Doazan and her coworkers indicating that the Be phenomenon is episodic, recurring at irregular intervals. Furthermore they show that Be stars are not "special" objects in the H-R diagram (e.g. rapid rotators near "breakup" velocity), but that even "ordinary" B-stars can become Be-stars for a time, and vice versa. In my opinion, we have yet to take even the most basic steps towards a realistic model for Be stars.

As we go down the main sequence, line-blanketing becomes increasingly severe and at some point our present NLTE models fail because they do not yet account for line-blanketing adequately. We may be able to overcome this problem in the next few years thanks to Anderson's new method. But even if we can get through the G-stars, the K-stars and M-stars promise to be much harder because of molecular blanketing, and larger convective and magnetic effects on the structure of their atmospheres. In all cases we need to answer someday the question of to what extent convective/magnetic inhomogeneities in an atmosphere "average out" (nonlinearly!) so that we can describe the atmosphere as being "effectively homogeneous", and to what extent we must consider the atmosphere to be composed of physically distinct components.

If we turn to giants and supergiants, my reaction is to wring my hands. I know, of course, that some brave people have modeled such stars, but I personally feel that it is a problem of almost hopeless difficulty for our present tools and paradigm. From the great size and low densities of these objects it is obvious that departures from LTE will be extreme. Furthermore, from the small irregular fluctuations in their light, colors, and of their line profiles, it is obvious that there is a great deal of hydrodynamic motion going on, and indeed that the atmosphere may be "unstable" in the sense of local areas jittering around randomly with relatively large velocity amplitudes. It may well be that the key word is "random", and that we will need to develop some kind of *statistical models* of these stars, and a statistical transport theory (on which important work has already been done) to compute their emergent spectrum. I think that it is essentially impossible to predict the rate of progress here, but given the general reticence of astronomers for dealing with hydrodynamics, I am not too optimistic about it being large.

Likewise, prudence (or cowardice?) has delayed my mentioning stellar chromospheres and coronae. Here I have a rather pessimistic view: I think that until we have a detailed physical understanding of the solar chromosphere and corona; until we have some kind of empirical indication of the topology and strength of magnetic fields in stellar chromospheres and coronae, and of the nature of the material inhomogeneities they induce; and until we have the ability to calculate the magnetohydrodynamic behavior and NLTE radiant output of 3D structures, possibly embedded in an ambient wind, then I think that we are likely to make little if any real progress. I know of course that there have been many papers, often based on IUE data, written about stellar chromospheres. But, with all due respect to their authors, I personally think that we do not learn much of permanent value from them because they invariably assume a model that is manifestly extremely oversimplified.

Finally, we should return briefly to the very basic question "What is a stellar atmosphere (anyway)?". I have tried to make it clear that a real stellar atmosphere is, in general, radically different from the typical theoretician's conception of it. The primary reason that this is so is, I believe, that none of us has ever seen a stellar atmosphere (other than the Sun's), and so we can oversimplify with a clear conscience. Further, we don't know how far out in the atmosphere we have to go before interesting and important things stop happening. Is just the first few scale heights above optical depth unity enough? Or do we need to go out through the (inhomogeneous) chromosphere and corona? How about the wind? (Certainly needed for WR stars!) Is going out to the sonic point enough to avoid further effects of the wind on the star? (Not for an O-star in which the wind-blanketing heats the underlying atmosphere to a much higher temperature and changes the emergent energy distribution!) What about dust shells? Are they really separate objects surrounding the star, or should they be considered to be a part of the atmosphere, given that the optically thick shell can absorb a significant fraction of the stellar flux at some wavelengths and re-emit it at others? And what about binaries: when does one star stop and the other begin; what about a common envelope?

In no case do we know how to answer these questions precisely enough to formulate unique equations and algorithms. But that is not surprising because these questions are not easy, and the answers cannot be expected to be "tidy". And so for the next few years, at least, we will have to content ourselves with partial answers, which, nevertheless, if properly understood, can still guide our research. In any case I want to stress that the field is in a state of rapid change. The next 20 years or so will certainly be an era of rapid growth in our knowledge about stellar atmospheres; one that will surpass the past 20. It will be a time which, I am sure, that Strömgren would have enjoyed greatly.

Conclusion

To conclude, I would like to end this talk on a more personal note. Bengt is gone now. And there are many here who will miss him deeply, whether as friend, teacher, or colleague. But I would like to remind all of us of the old saying that "*No man is truly dead until the day his name is last spoken*". On that count Bengt will remain alive amongst us for many, many, years to come!

References

Anderson, L.S. 1985, *Astrophys. J.* **298,** 848.

Anderson, L.S. 1987, in *Numerical Radiative Transfer,* ed. W. Kalkofen (Cambridge University Press) p. 163.

Castor, J.I., Abbott, D., Klein, R.I. 1975, *Astrophys. J.* **195,** 157.

Hamann, W.-R. 1985, *Astron. Astrophys.* **145,** 443.

Hamann, W.-R. 1985, *Astron. Astrophys.* **148,** 364.

Hamann, W.-R. 1986, *Astron. Astrophys.* **160,** 347.

Hamann, W.-R. 1987, in *Numerical Radiative Transfer,* ed. W. Kalkofen (Cambridge University Press) p. 35.

Hillier, D.J. 1987a, *Astrophys. J. Suppl.* **63,** 947.

Hillier, D.J. 1987b, *Astrophys. J. Suppl.* **63,** 965.

Kalkofen, W. (ed.) 1984, *Methods in Radiative Transfer* (Cambridge University Press).

Kalkofen, W. (ed.) 1987, *Numerical Radiative Transfer* (Cambridge University Press).

Lucy, L., Solomon, P. 1970, *Astrophys. J.* **159,** 879.

Nordlund, Å. 1985, *Solar Physics* **100,** 209.

Scharmer, G.B. 1981, *Astrophys. J.* **249,** 720.

Scharmer, G.B. 1984, in *Methods in Radiative Transfer,* ed. W. Kalkofen, (Cambridge University Press) p. 173.

Scharmer, G.B., Carlsson, M. 1985, in *Progress in Stellar Spectral Line Formation Theory,* ed. J.E. Beckman and L. Crivellari (Reidel) p. 189.

Stein, R.F., Nordlund, Å. 1989, *Astrophys. J. Letters* **342,** L 95.

Werner, K. 1986, *Astron. Astrophys.* **161,** 177.

Werner, K. 1987, in *Numerical Radiative Transfer,* ed. W. Kalkofen (Cambridge University Press) p. 67.

Werner, K., Husfeld, D. 1985, *Astron. Astrophys.* **148,** 417.

The Chemical Composition of Stars

By David L. Lambert
McDonald Observatory and Department of Astronomy
University of Texas
Austin, Texas 78712-1083, USA

Abstract

This essay discusses how recent and continuing advances in astronomical spectrometers are providing novel opportunities for stellar spectroscopists to address a wide variety of problems of stellar nucleosynthesis and evolution

1. *Introduction*

Thirty years ago, Bengt Strömgren (1958) delivered The Halley Lecture in the University Museum in Oxford on the topic "The Composition of Stars and their Ages". His lecture sketched the theoretical principles, observational practices, and his initial goals for that system of photoelectric narrow-band photometry now universally known as Strömgren photometry. As Professor Strömgren was enlightening his learned audience on May 27, 1958, I may have been nearby as my first delightful year as an undergraduate reading Physics at Oxford was coming to an end. Although my curiosity about astronomy and, in particular, about stellar evolution and nucleosynthesis, had been piqued a year earlier by reading Fred Hoyle's *Frontiers of Astronomy*, I did not attend Bengt Strömgren's Halley Lecture.

Here, as in my talk in Copenhagen at the Niels Bohr Institute, I shall attempt to convey the excitement that now envelopes pursuit of the determination and interpretation of the chemical composition of stars. Although, in contrast to Strömgren's Halley lecture, I shall concentrate on spectroscopic rather than photometric methods of attack, a common thread connects these two lectures given 30 years apart. Strömgren referred in his lecture to exploratory studies of his photometric system using the 36 and 82-inch reflectors of the W.J. McDonald Observatory in West Texas. Observations obtained at this observatory with the 82-inch and the 107-inch reflectors will illustrate several of my topics. Emphasis in my essay on spectroscopic methods and neglect of photometric methods of abundance determination reflects a personal enthusiasm but certainly not astrophysical myopia. There is a critical need for a thorough review of the intimate connections between spectroscopic and photometric methods of abundance determination; the common view that the former is the preferred and secure basis for calibrating the latter deserves to be challenged. Perhaps, another occasion can be found to elaborate on this theme that, I suspect, would have provoked original ideas and suggestions from Professor Strömgren.

My initial ruminations on the instructions to address questions of "Recent Progress and Future Possibilities" led me to isolate three major themes:

(i) Interesting and important astrophysical problems whose solution is dependent on accurate data on stellar chemical compositions.
(ii) The stellar spectra that are basic ingredients with which an abundance analysis commences.
(iii) The methods and associated tools of the abundance analysis that are applied to the stellar spectra and other observations (e.g. photometry) to obtain the abundances. The tools include model stellar atmospheres and a library of atomic and molecular data.

Today's exciting expansion of our knowledge of stellar chemical compositions is occurring in large part because substantial, often dramatic, advances are occurring in all of the fields spanned by the above three themes. Larger telescopes with more sensitive and versatile spectrometers provide high quality stellar spectra over broader spans of the electromagnetic spectrum. Concurrently, advances are occurring in the key components of abundance analysis: model stellar atmospheres, spectral line formation including thorough treatments of the departures from Local Thermodynamic Equilibrium (LTE), and the library of atomic and molecular theoretical and experimental data. This happy conjunction of high quality stellar spectra and refined analytical tools is providing the imaginative observer with novel opportunities to tackle outstanding unresolved astrophysical problems.

To set the stage, I conclude this introduction with a few brief remarks on each of my three major themes.

Astrophysical Problems. The truly challenging problems may be defined succinctly: what was the origin, structure, and evolution of the Universe, the galaxies (including our Galaxy), the stars, and the solar system? A stellar spectroscopist who gathers the appropriate measures of stellar chemical compositions may probe almost any aspect of this question. Specific topics that follow in this essay are primarily devoted to questions of stellar evolution and nucleosynthesis. As a reminder that the potential scope of inquiry is very broad, I draw attention to just one recent focus of great interest: the discovery by Spite and Spite (1982) that Population II dwarfs have a measureable lithium abundance, ($\log \varepsilon(\mathrm{Li}) \simeq 2$ on the usual scale where $\log \varepsilon(\mathrm{H}) = 12$), has led to a flurry of claims and counterclaims about the yield of lithium from the primordial fireball, the nature of the fireball and alternative cosmologies, and theoretical ideas on the ways in which the surface Li abundance of Population I and II stars may be modified with age and reassessments of schemes for ^{7}Li synthesis by stars. The Spites' discovery has also led to a revival of observational studies of lithium in stars. Current studies seek to delineate not only how stars destroy lithium but also how a few stars may synthesize lithium in copious amounts.

Stellar Spectra. Recent and continuing advances in the tools of the trade include more large telescopes, more efficient spectrometers, and expanded spectral coverage. Further advances including the European Southern Observatory's Very Large Telescope are imminent, at least to observers with the long-term visions of a Bengt Strömgren. Such advances translate to the following enhanced opportunities for the observer to extend or to initiate programmes involving:

- fainter stars
- large samples of stars
- higher resolution spectra
- access to a preferred wavelength interval
- higher signal-to-noise (S/N) spectra
- higher temporal resolution

Often, outstanding advances towards the solution of fundamental problems will call for combinations of these opportunities. In the main body of my essay, I endeavour to illustrate, through a variety of examples drawn from the recent literature and our own work at the McDonald Observatory, the potential of these opportunities.

Methods of Abundance Analysis. The raw ingredients are astronomical observations (stellar spectra and photometry), model stellar atmospheres, and basic atomic and molecular data. Today, these ingredients are whipped together in a suite of computer codes that represent an increasingly sophisticated representation of the physics of stellar atmospheres.

Since Dimitri Mihalas reviews model atmospheres elsewhere in this volume, I limit my remarks to noting that, in the context of "Recent Progress and Future Possibilities", serious exploration of atmosphere construction is commencing in which the basic assumptions previously adopted are being modified or discarded. The 'new' atmospheres are beginning to impact our studies of stellar chemical compositions. The basic assumptions to which I refer involve questions of geometry, LTE, and hydrostatic equilibrium. A brief selective discussion of how the basic assumption pertaining to geometry is being modified must suffice to convey the scope of 'future possibilities' in the field of model atmospheres. Several studies have now been reported in which the basic assumption of homogeneous plane-parallel, semi-infinite layers is modified. Atmospheres for luminous cool stars with thick atmospheres have been computed from the same basic assumptions except that the sphericity of the thick atmosphere is recognized (e.g. Bessell *et al.* 1989). The assumption of homogeneity has been discarded in a few exploratory studies; e.g., Nordlund's (1982) modelling of solar granulation through numerical solutions of the hydrodynamic and radiative transfer equations. Other exploratory discussions of inhomogeneity have drawn attention to thermal instabilities that may arise when abundant molecules such as CO and SiO in O-rich stars exert a major influence on the radiative equilibrium of a star's upper atmosphere. The extreme sensitivity of such molecular opacity to temperature can lead to a 'bifurcation' of the solutions that satisfy the condition of radiative equilibrium (Muchmore 1986); i.e., a 'hot' column can coexist with a 'cool' column. Theoretical exploration and empirical testing is just beginning for these and other modifications of the basic assumptions that are the foundation of the various grids of model atmospheres in wide use for abundance analyses.

In the following sections, I aim to illustrate through specific examples how rapid and startling progress is being made towards the solution of a variety of astrophysical problems. My emphasis is unashamedly on the recent improvements in the quality and diversity of observed stellar spectra.

2. *Stellar Spectra in the 1980s*

In the Introduction, I listed six enhancements of the opportunities now open to the stellar spectroscopist relative to the more limited opportunities available less than a generation ago when telescopes with apertures larger than 100 inches were available only to a select few and serious spectroscopy was restricted to a narrow spectral interval recorded on photographic plates. Here, I amplify the discussion of the enhanced opportunities.

2.1 *Exploration of the Magellanic Clouds*

Perhaps the most dramatic exploitation of these opportunities is presently being made through moderate to high-resolution spectroscopy of faint stars and especially those in the Magellanic Clouds, the Galaxy's globular clusters, and the distant parts of our Galaxy's disk and central bulge.

Spectroscopy and photometry of Magellanic Cloud field and cluster stars have in recent years provided a stream of new results about stellar evolution. My focus here is on the red giants that belong to the asymptotic giant branch (AGB). The most evolved of AGB stars are predicted to experience He-shell flashes followed by the possibility of a mixing of products from the He-burning shell into the deep convective envelope and, hence, the spectroscopically accessible atmosphere. The principal products are ^{12}C from He via the 3α-process and heavy elements (e.g. Sr and Ba) synthesized by successive neutron captures (the s-process) from abundant lighter 'seeds' such as the iron-peak elements. The neutron source is considered to be $^{22}Ne(\alpha,n)^{25}Mg$ in the most massive AGB stars (say $M \simeq 3 - 8M_{\odot}$) and $^{13}C(\alpha,n)^{16}O$ in the less massive stars. This mixing (or third dredge-up) converts O-rich M giants into S stars and finally into C-rich N-type giants. Since bright Galactic S and N stars exist in great numbers and their chemical compositions may be analysed in detail from high-resolution spectra, it might appear that the much fainter AGB stars of the Magellanic Clouds play nothing more than a supporting rôle in the testing of the theory of the AGB stars and the third dredge-up. In fact, the rôles of principal and supporting player are reversed because reliable estimates of absolute luminosity are obtainable for Cloud members but not for the bright Galactic AGB stars. In short, one 'knows' from their location in the H-R diagram whether candidate AGB stars in the Clouds are or are not AGB stars, and, a closer comparison with theoretical expectations provides estimates of the stellar masses. The best that one can do for the Galactic stars is to estimate their luminosity from statistical parallaxes; a very few stars may belong to open clusters. As an extreme reflection on the uncertain estimates of luminosity, I note that just 10 years ago two authorities speculated on an alternative explanation for the Galactic stars: "Perhaps the simplest explanation of S stars and of many N-type carbon stars is that the surface composition characteristics originate during a helium [core] flash of an infrequently occurring nature" (Iben and Truran 1978). Although the possibility of He core flashes of "an infrequently occurring nature" may be retained as an explanation for peculiar rare stars, surveys of the Clouds have uncovered the general pattern of evolution on the AGB and the place of the S and N stars in that pattern.

The observed pattern provided one early surprise. Predictions that dredge-up of ^{12}C and s-process was the sole prerogative of intermediate-mass AGB stars were first shaken and then shattered by surveys of cool carbon stars in the Magellanic Clouds. The luminosity function of C stars in the Clouds peaks at $M_{bol} \sim -4.8$ or $\log L/L_\odot \sim 3.8$ (Cohen *et al.* 1981; Richer 1981). Carbon stars are not found with $\log L/L_\odot > 4.3$ (Cohen *et al.* 1981; Wood, Bessell, and Fox 1983 hereafter WBF; Wood 1987); this luminosity was the predicted *lower* limit for carbon star production by intermediate mass stars. The observed luminosity range is consistent with an identification of cool carbon stars as thermally pulsing *low* mass AGB stars. Between the observed upper limit ($M_{bol} \sim -6$) for cool carbon stars and the maximum luminosity for an AGB star ($M_{bol} \sim -7.1$), a limit set when the degenerate core reaches the Chandrasekhar limit, the AGB stars in the Clouds are oxygen-rich. This sample defined first by WBF contains S stars. WBF speculate that H-burning at the base of the convective envelope may through the cycling of C to N reconvert the carbon-star to an oxygen-star heavily enriched in nitrogen. The H-burning necessary to accomplish this conversion of C to N is predicted to occur at the base of the deep convective envelope of the most luminous AGB stars (Iben 1973; Scalo, Despain and Ulrich 1975; Renzini and Voli 1981). Other explanations for "The Carbon Star Mystery: Why Do the Low Mass Ones Become Such and Where have the High Mass Ones Gone?" (Iben 1981) are reviewed by Iben (1989).

These and other conclusions about the evolution of AGB stars in the Clouds are founded almost entirely on low resolution spectra and photometry. That this was possible is due to the controlling rôle of the CO molecule in the dissociation equilibrium of carbon and oxygen with the result that spectra of cool stars with an abundance ratio $C/O < 1$ (oxygen-rich) are readily distinguishable at even low resolution from spectra of cool stars with C/O > 1 (carbon-rich). In addition, enrichment of cool oxygen-rich stars with s-process elements is detectable through a strengthening of bands of metal-oxides such as ZrO and YO; i.e., S stars can be distinguished from M stars without access to high resolution spectra. But to retrieve finer details of the chemical composition, high resolution spectra must be obtained and analysed.

High resolution spectroscopy is now being undertaken of Cloud AGB and other stars in the upper part of the H-R diagram including OB stars, F supergiants and AGB stars. One goal of such analyses is to define differences in the stellar and interstellar (i.e., H II region) abundances and to relate these differences (e.g. N enrichment of a star) to mixing and mass loss by the star. These analyses have assumed an especial significance with the explosion of Supernova 1987A. If that explosion occurred in a star evolving to the blue after a period as a red supergiant, the envelope would have been enriched in N and indeed, there is evidence for such enrichment. Examination of C, N, and O abundances in a sample of OB stars should reveal the frequency of N enrichment and, hence, an upper limit on the fraction of stars that are evolving to the blue. Since N enrichment of the surface may also result from extensive mass loss near the main sequence or by mass exchange in a binary, it may not be a simple matter to isolate the post-red giant stars from such imitators. Studies of F supergiants reported by Spite *et al.* (1989) and Russell and Bessell (1989) have focused on determining the abundances for a wide variety of elements that sample the major sites of stellar nucleosynthesis. Comparison of the abundances in the Small and Large Magellanic Clouds with those in the Galactic stars of the same metallicity is expected to yield clues to the chemical evolution of these three galaxies.

High resolution spectra of the AGB stars in the Clouds promise to provide answers to some obvious questions as well as surprises that raise additional questions. It is with a surprise that I close this section. Recently, we (Smith and Lambert 1989) obtained spectra of five AGB stars in the Small Cloud. The stars were taken from WBF's list of O-rich AGB stars that are more luminous than the carbon-rich AGB stars and were classified as S stars because ZrO bands were detected on low resolution spectra. Our spectra from 5470-7900 Å at a resolution $\lambda/\Delta\lambda \simeq 20{,}000$ were obtained with CTIO's 4m telescope, the Cassegrain echelle spectrometer and a GEC CCD detector.

Two key questions may be answered with our spectra:

(i) Are these luminous massive AGB stars enriched in the s-process elements as WBF supposed from the appearance of the ZrO bands?

(ii) Is there evidence in the compositions for the presence of a hot-bottom convective envelope (HBCE) that converts the freshly synthesized C dredged from the He-shell into N?

The answer to the first of these questions is unambiguous. In Figure 1, I show a strip of spectrum near 7555 Å for two SMC stars and the Galactic stars δ Vir (a normal M giant) and HD 35155 (a S star). Inspection of the spectra shows that the

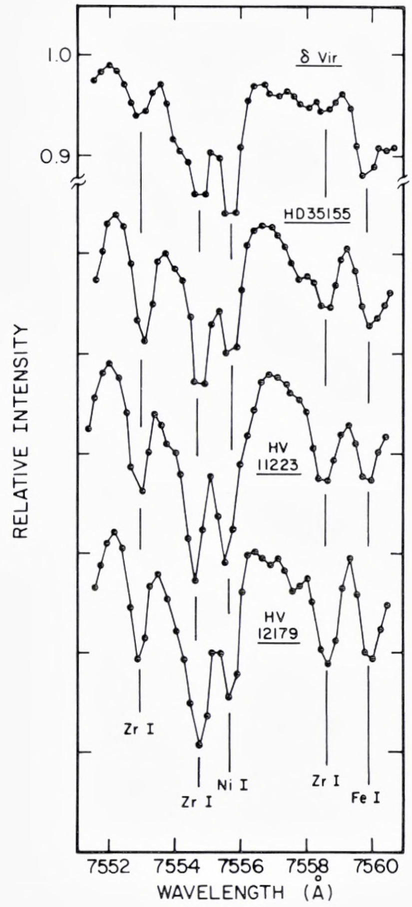

Fig. 1. Spectra showing the increased strengths of the Zr I lines in the SMC red giants relative to the normal-abundance red-giant δ Vir. The SMC giants are similar to the Galactic S-star HD 35155.

Zr I lines that are a monitor of the s-process products are enhanced (relative to δ Vir) in both SMC stars and the S star HD 35155. All 4 SMC stars show the s-process enhancements. Inspection of other wavelength intervals shows enhancements of other s-process elements. Since the enhanced equivalent widths of the s-process lines of the SMC and Galactic S stars are similar, we conclude that WBF's most massive AGB stars are enriched in the s-process, i.e., the predicted third dredge-up has occurred in these stars. Our abundance analysis confirms this conclusion.

Occurrence of the third dredge-up, as betrayed by the s-process enrichments, must add carbon to the envelopes of these stars. Indeed, these intermediate-mass stars were predicted by Iben (1975) in a series of pioneering calculations to evolve into carbon stars. However, as noted above, there are no luminous carbon stars in the clouds. Several obvious explanations have been advanced for their absence; e.g., evolution on the AGB is terminated by severe mass loss prior to transformation of the envelope from oxygen-rich to carbon-rich; severe mass loss shrouds a carbon-rich star in a thick graphite dust shell; a HBCE converts C to N to maintain an oxygen-rich envelope.

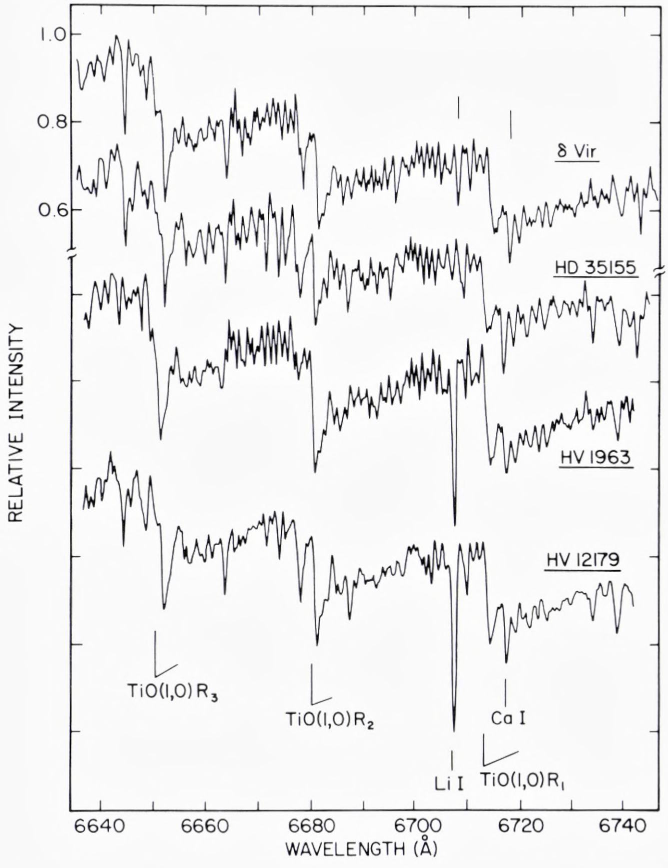

Fig. 2. Sample spectra illustrating the Li I resonance doublet. An entire echelle order is shown: note the very strong Li I feature in the SMC stars (HV 1963 and HV 12179) as compared to the two Galactic red giants. Note also that this region is blanketed by TiO absorption.

The presence of a HBCE may be signalled by the surprise provided by our spectra:

the great strength of the Li I resonance doublet at 6707 Å – see Figure 2 where it is shown that the Li I doublet is prominent in the SMC stars but very weak in δVir and HD 35155. Galactic counterparts of the Li-rich SMC stars are known but rare. The handful of Galactic examples have been gathered from surveys of two or more hundred cool giants. By contrast, each of the five SMC stars, a sample chosen randomly from the luminous AGB stars listed by WBF, is Li-rich with Li abundances within the range reported for the Galactic Li-rich stars, i.e., Li abundances of 10^2 to 10^4 higher than those found in normal red giants. In the latter, Li is diluted by a factor of about 50 below the abundance of the parental interstellar cloud. Reduction (destruction, diffusion, mass loss) of surface Li prior to evolution onto the first giant branch may have further reduced the Li abundance of some red giants. Lithium abundances are not yet available for SMC giants at a stage prior to the onset of thermally pulses on the AGB. Observations of the F and later type main sequence stars in which Li is expected to be seen are beyond the range of today's telescopes. A clue to the interstellar Li abundance in the LMC is provided by the upper limit, log ε(Li) $\lesssim$ 2, derived from the nondetection of the Li I 6707 Å interstellar line in the near maximum spectrum of SN 1987A (Baade and Magain 1988; Sahu, Sahu and Pottasch 1989) – see, however, Malaney and Alcock (1989) who argue that a more realistic upper limit is log ε(Li) $<$ 3.6. If log ε(Li) $\lesssim$ 2 is appropriate for the SMC's interstellar gas, the Li abundances of the 5 SMC AGB stars range from about the primordial value to up to almost 100 times larger. Since, however, the reference Li abundance for the AGB stars should be the lower (diluted) value expected for the pre-AGB red giants (i.e. log (Li) $\lesssim$ 0.3), the SMC AGB stars with log ε(Li $\simeq$ 2.2-3.8 must surely have synthesized the Li now observed in their atmospheres. Synthesis of Li has been previously proposed to account for the rare ($\lesssim$ 1 % of red giants, Scalo 1976) Li-rich Galactic giants. By contrast, Li appears to be enhanced in perhaps all of the most massive (4-8$M_\odot$) SMC AGB stars. We do not yet have spectra for less massive/luminous SMC AGB stars but we suspect that they will exhibit the low (diluted) Li expected for normal red giants. If this suspicion is confirmed, it will encourage the identification of the Galactic Li-rich giants as massive AGB stars. Their rarity is then a direct result of the bias of the Initial Mass Function toward low masses and the more rapid evolution of high mass stars along the AGB.

The possibility that ^{7}Li may be synthesized in red giants was recognized by Cameron and Fowler (1971) who proposed the "^{7}Be-transport mechanism" in which ^{7}Li is created via the sequence ^{3}He (α,γ) ^{7}Be (e^-, ν) ^{7}Li in a convection zone where the ^{3}He is a product of prior H-burning and the ^{7}Li (and ^{7}Be) is transported by the convection to cooler layers of the red giant and so avoids destruction by protons. Luminous AGB stars are predicted to develop convective envelopes with high temperatures ($T_b \simeq$ 20-60 $\times$ 10^6K) at their base where the ^{7}Be-transport mechanism and H-burning will occur. Theoretical studies suggest that, under certain conditions, ^{7}Li may be created and mixed to the surface of luminous AGB stars (Iben 1973;

Sackmann, Smith and Despain 1974; Scalo, Despain, and Ulrich 1975). Scalo *et al.* (1975) predict that a HBCE develops in AGB stars with $M_{bol} \lesssim -5.4$. Since our SMC stars have $M_{bol} \simeq -6$ to -7 and their Li abundances are within the (uncertain) range predicted by Scalo *et al.*, we associate the Li-rich SMC stars with the occurrence of a HBCE.

A part of the fascination for these massive AGB stars is their betrayal of the secrets of internal nucleosynthesis including the subtle mechanisms needed to synthesize and transport Li to the surface. Another reason for interest in these stars is their potential rôle as a leading producer of ^{7}Li within a galaxy. A firm calibration of the yield of ^{7}Li from Li-rich AGB stars may terminate the vigorous debate on the primordial ('big-bang') Li abundance in favour of the value observed in Galactic Pop. II dwarfs ($\log \varepsilon(\mathrm{Li}) \simeq 2$) rather than the value ($\log \varepsilon(\mathrm{Li}) \simeq 3$) seen in the youngest Pop. I stars.

Our observations suggest that the Li rich stars are massive AGB stars. Although our sample is small, the fact that all are enriched in Li suggests that all AGB stars in this mass range synthesize and retain substantial amounts of Li. Observations of less luminous AGB stars are needed to determine the lower mass limit for Li enrichment. With some simple assumptions, we may estimate whether mass loss from the Li-rich AGB stars results in a progressive enrichment of the interstellar ^{7}Li abundance. Let the Li-rich stars of mass 4 to $8M_\odot$ have an abundance $\varepsilon(\mathrm{Li})_s$ and their ejected mass be equal to the inital stellar mass less about 1 $M_\odot$, the mass of the white dwarf remnant. The ejecta of lower mass stars are depleted in Li by a factor of about 50. If the initial mass function follows the form, $\phi(m) \propto m^{-2.35}$, interstellar Li will increase as long as

$$\varepsilon(\mathrm{Li})_s \gtrsim 5\varepsilon(\mathrm{Li})_0$$

where $\varepsilon(\mathrm{Li})_0$ is the prevailing interstellar abundance. Our results suggest that $\varepsilon(\mathrm{Li})_s$ satisfies this condition. Our identification of the most massive AGB stars as an important source of galactic ^{7}Li confirms Scalo's (1976) conclusion based on crude estimates of mass loss from Li-rich AGB stars.

The hypothesis that Li production is a consequence of a HBCE would seem to demand that the atmospheres of the Li-rich AGB stars contain the products of H-burning that must also occur at the base of the convective envelope; i.e., the atmospheres should be N-rich. Unlike Li that may be eventually destroyed throughout the convective envelope, the N enrichment is seemingly a permanent signature of a HBCE. Brett (1989) used low-resolution near-infrared and infrared spectra plus spectrum synthesis of molecular bands to estimate the C, N and O abundances of a sample of luminous SMC AGB stars that included two of our stars. He concluded that the stars are not as extremely N-rich as would be expected for a HBCE and that these AGB stars have experienced the third dredge-up but have not developed a HBCE. Analysis of CN lines in our spectra does suggest, however, that the N abundances in these SMC stars probably exceed both the C and O abundances, thus these red giants might actually be quite N rich. A full analysis of the CN lines requires knowledge of the C abundance, thus, a proper CNO analysis must await improved spectra of the infrared vibration-rotation CO bands. Since Scalo *et al.*'s (1975) calculations suggest that the timescale for ^{7}Li-production is significantly shorter than for ^{14}N production, the Li-rich stars may be identified as having just developed a HBCE and, if the HBCE is maintained for long enough, these stars may later develop a N-rich and Li-poor atmosphere.

This discussion of the composition of massive AGB stars in the SMC is but a beginning of detailed spectroscopic exploration of the AGB stars in the Clouds. Much can be done with existing equipment. A list of key questions to be answered at the telescope would surely include: Is Li production restricted to the most massive AGB stars? Are these stars maintained as O-rich by a HBCE that converts C into N? Is there spectroscopic evidence for the predicted neutron source – ^{22}Ne(α, n)^{25}Mg – in these stars? Did this source operate in the He-shell at the predicted neutron densities? Are there systematic differences between the s-process enrichments and abundance patterns in high and low mass AGB stars as the neutron source is expected to change from ^{22}Ne(α,n) ^{25}Mg to ^{13}C(α,n) ^{16}O? How do the abundance patterns change for S and N type stars (i.e., with metallicity) between the Galaxy, the LMC and the SMC? Do the cool supergiants with masses above the limit for AGB stars have the expected low (diluted primordial) Li abundance?

2.2 *Statistical Spectroscopy of Red Giants*

With today's telescopes, spectrometers and computing facilities, it is possible to execute observing programs that are designed to search large numbers of stars for extremely rare examples of an abundance anomaly. In this section I discuss two programs of 'statistical spectroscopy' completed recently at our McDonald Observatory.

2.2.1 *Lithium in G and K Giants*

Lithium is expected to be destroyed in all but the outermost layers (1-2 % by mass) of a main sequence star. On ascent of the red giant branch, a deepening convective envelope dilutes the remaining lithium and so reduces the surface (i.e. observable) lithium content by a large factor. G and K giants are expected to have a low lithium abundance. This expectation can be expressed quantitatively. Currently, the lithium abundances in the local interstellar gas and in those young stars believed to retain the primordial abundance is log ε(Li) = 3.1 ± 0.2. Although the interpretation is not without its critics, the current Li abundance appears to have grown from the level log ε(Li) = 2.0 ± 0.2 seen at the surfaces of unevolved Population II stars with [Fe/H] ≤ -1. A compilation of Li abundances in unevolved stars (Rebolo, Molaro, and Beckman 1988) suggests how the abundance may have grown with metallicity. This suggested trend describes the *maximum* Li abundance at the surface of a main sequence star. In general, the cooler stars exhibit significantly lower Li abundances.

Iben's (1967) models predict the dilution at the tip of the red giant branch to amount to a factor of between 60 at 3 $M_\odot$ to 28 at 1 $M_\odot$. Therefore red giants of near solar metallicity are expected to have Li abundances of $\log \varepsilon(\mathrm{Li}) \approx 1.3$ at 3 $M_\odot$ to ≈ 1.7 at 1 $M_\odot$, assuming that their main sequence progenitors retained the primordial lithium over the shallow layer predicted by standard Iben models. However, since many main sequence stars show significant lithium depletion, these predictions surely refer to the *maximum* expected in a red giant. In short lithium in red giants should not exceed an abundance $\log \varepsilon(\mathrm{Li}) = +1.5 \pm 0.2$ and a majority of giants are expected to show a Li abundance below this maximum value.

Wallerstein and Sneden's (1982) discovery of an apparently ordinary K-giant with a nearly cosmic Li abundance challenged the simple picture of Li destruction, depletion, and dilution: the giant HD 112127 (spectral type K2 III: CN+3) has a Li abundance $\log \varepsilon(\mathrm{Li}) = 3.0 \pm 0.2$. A detailed analysis of the spectrum of HD 112127 demonstrated that the spectral type assignment is accurate; the atmosphere parameters are those of a giant. Finally, the star evidently is a giant with a deep convective envelope because the ^{13}C content is enhanced (^{12}C/^{13}C ≈ 22).

Stars as Li-rich as HD 112127 are extremely rare and all previously known examples have other striking abundance anomalies. Lithium abundances for the weak G-band G and K giants range up to the cosmic abundance of $\log \varepsilon(\mathrm{Li}) \simeq 3$ but their atmospheres are severely contaminated with the products of H-burning CN-cycling: i.e., ^{12}C/^{13}C ~ 3, and a low ^{12}C with a high ^{14}N abundance such that ^{12}C plus ^{14}N is conserved. By contrast, HD 112127 has the ^{12}C, ^{13}C and ^{14}N abundances expected of normal giants on the first ascent of the giant branch. The Li-rich S and C stars discussed in the preceding section are much more highly evolved than giants like HD 112127.

The significance of HD 112127 to our picture of stellar evolution depends on the answer to the question: Is HD 112127 a very peculiar, even unique star or is it a representative of a rare subclass of G and K giants? To answer this question, we conducted a survey of the Li I 6707 Å doublet in 644 bright giants (Brown *et al.* 1989).

Our spectra were obtained with the coudé spectrograph of the 82-inch telescope at the McDonald Observatory. The Reticon detector recorded 95 Å around the Li I doublet at a resolution of 0.3 Å and a high S/N ratio (> 150). The Li abundance was extracted by spectrum synthesis.

The results of our survey are conveniently displayed as a histogram (Figure 3). All but a handful of the 644 giants have a Li abundance within the predicted range for giants having a deep convective envelope, i.e., $\log \varepsilon(\mathrm{Li}) \lesssim 1.5$. The lower histogram in Figure 3 is compiled from observations of main sequence stars. As expected, the two histograms are similar but that for the giants is displaced by a factor of about

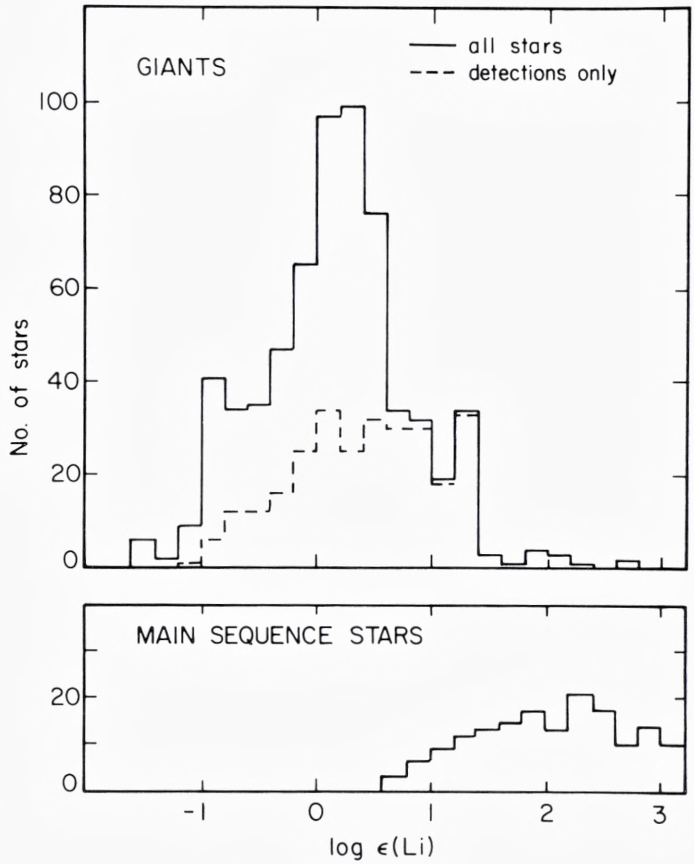

Fig. 3. Histograms of Li abundances for red giants (top panel – Brown *et al.* 1989) and main sequence stars (bottom panel – Boesgaard and Tripicco 1986 and Duncan 1981).

50 ($\triangle \log \varepsilon(\mathrm{Li}) \simeq 1.7$) to lower Li abundances. This factor represents the predicted dilution introduced by a giant's convective envelope. (Note: both histograms are partly shaped by selection effects.)

Ten of the 644 giants have a Li abundance greater than is predicted for a red giant. The ten stars are listed in Table 1 and plotted in an HR diagram in Figure 4. One star, HD 9746, resembles the prototype HD 112127 in having a near-cosmic Li abundance. An additional Li-rich ($\log \varepsilon(\mathrm{Li}) = 2.8$) giant HD 39583 is described by Gratton and D'Antona (1989). In searching for explanations for these peculiar stars, we dismiss temporarily the supergiant HD 205349 (No. 10 in Table 1) and the subgiant HD 126868 (No. 7). These latter stars may not yet have developed the

Table 1
Properties of the Strong Li Stars

Star	HD	log ε(Li)	[Fe/H]	log T_{eff}	M_{bol}
1	787	1.8	+0.07	3.63	−1.3
2	9746	2.7	−0.13	3.65	0.0
3	30834	1.8	−0.17	3.62	−1.4
4	108471	2.0	−0.02	3.70	−0.2
5	112127	2.7	+0.31	3.64	−0.2
6	120602	1.9	−0.07	3.70	−0.1
7	126868	2.3	−0.25	3.74	+1.8
8	148293	2.0	+0.23	3.67	+0.6
9	183492	2.0	+0.08	3.67	−0.2
10	205349	1.9	...	3.65	−5.2

convective envelope that dilutes the surface lithium. The supergiant is reminiscent of HR 8626 (Baird *et al.* 1975) and a few other stars.

The $^{12}C/^{13}C$ ratio is taken to be a monitor of a convective envelope. Our observations of CN red system lines show that both HD 9746 (No. 2) and HD 108471 (No. 4) have a $^{12}C/^{13}C$ ratio that is representative of normal giants possessing a convective envelope. This conclusion was reached earlier for HD 112127 (No. 5). Other stars on the list of ten have yet to be analysed for their $^{12}C/^{13}C$ ratio. It is conceivable that some of the remainder will prove to have retained their main-sequence ^{13}C abundance and, hence, be identifiable as giants at the base of the first giant branch with an incompletely developed convective envelope. The approximate locations for such giants with Li depleted by factors of 1.5 and 12 are shown on Figure 4. HD 120602 (No. 6), HD 148293 (No. 8) and HD 183492 (No. 9) are possibly normal giants with, as yet, a very shallow convective envelope. The more evolved stars HD 787 (No. 1) and HD 30834 (No. 3) appear to have an anomalously high Li abundance. Determinations of the $^{12}C/^{13}C$ ratios and C, N, and O abundances for the full set of 10 Li-rich giants will clarify their status. For the present, we reject only HD 126868 and 205349 and retain 8 of the 10 stars in Table 1 as peculiar Li-rich giants.

Convincing explanations for the preservation or production of Li do not exist. Several speculations may be worth noting for the observational tests they may stimulate:

(i) *Production at the He-core flash?* Six of the 8 peculiar stars are located near the clump containing He-core burning giants at the base of the AGB. The remaining 2 stars are plausibly identifiable as either evolved clump (i.e., early AGB) stars or younger giants on the first giant branch. Can Li be synthesized at the time of the He-

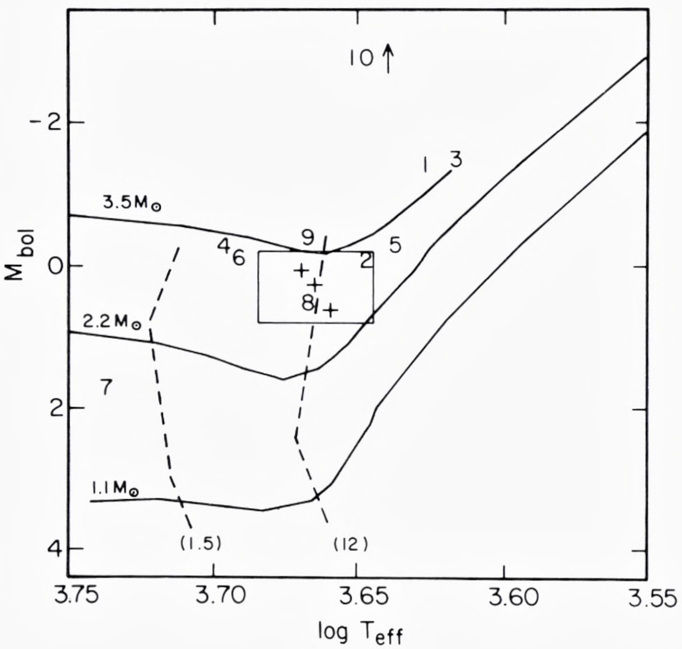

Fig. 4. A H-R diagram for the Li-rich red giants. The numbers identify the stars listed in Table 1. The box indicates the position of the He-core burning giants ('clump' giants). Crosses show the positions of the clump for 3 Galactic clusters, and the dashed lines are loci of constant Li depletions with the values of the depletion given in parenthesis.

core flash that initiates the He-core burning phase? The ^{7}Be-transport mechanism invoked for the AGB stars may be here working under different conditions.

(ii) *Preservation of Li through diffusion?* If much of a star's initial Li were to diffuse outwards, it may escape destruction by warm protons. Then, the giant's convective envelope would redistribute the Li to provide a Li-rich (relative to normal giants) envelope with an abundance in the range $1.5 \lesssim \log \varepsilon(\mathrm{Li}) \lesssim 3$. This is the process sketched by Lambert and Sawyer (1984) who noted that this diffusive process provided a natural explanation for the fact that the maximum Li abundance reported for weak G-band giants is a star's presumed initial or cosmic abundance. The same maximum applies apparently to the Li-rich G and K giants.

(iii) *Replenishment of Li by ingestion of major planets or brown dwarfs?* Alexander (1967) suggested that a giant could replenish its Li by ingesting surrounding planetary material. It is readily shown (Brown *et al.* 1989) that ingestion of terrestrial-like planets (i.e. no H or He) formed from a mass (initially including H and He) about equal to that of a red giant will restore the Li abundance to about the initial/cosmic value. Ingestion of a brown dwarf that has retained H and He has a smaller effect on the giant's Li abundance; Livio and Soker (1983, 1984) suggest that, if the mass of the brown dwarf (or terrestrial planet) exceeds about 1 percent of the giant's mass, the brown dwarf accretes a large fraction of the giant's envelope and becomes a low mass stellar companion to the giant's core.

(iv) *Li-rich giants have an active chromosphere?* There is a tantalising hint that some of these giants may possess an active chromosphere. The causal connection between Li and an active chromosphere is unclear. It could be simply that these stars possess large cool plages whose presence is ignored by standard model atmospheres. A careful examination of the abundances of elements having neutral atoms with low ionization potentials (Li, Na, K, Al, Rb) should uncover the existence of plages and result in a revision downwards of the Li abundance.

One goal of our survey was accomplished. HD 112127 is a rare but not a unique example of a G-K giant with a near cosmic Li abundance as its sole abundance anomaly. About 1 in a 100 G-K giants has an anomalously high Li abundance. A second goal proved elusive: the evolutionary origin of these Li-rich giants was left to speculation but some of the speculations are open to observational tests. For an observer, this is not an entirely unsatisfactory end – unemployment is postponed!

2.2.2 *The s-process in G and K Giants*

Barium stars were identified as a class of peculiar G-K giants by Bidelman and Keenan (1951) who noted the stars' enhanced atomic lines of heavy elements (e.g. Sr, Ba) and molecular lines of the C-containing molecules CH, CN and C_2. Overabundances of the heavy elements are now attributed to their synthesis by neutrons in the s-process in a He-burning layer that is also the site of the C enrichment. Thanks to the discovery by McClure and colleagues (McClure, Fletcher, and Nemec 1980; McClure 1985) that all Barium stars are spectroscopic binaries, the search for the origin of the stars shifted to scenarios involving binary rather than single stars. In particular, Barium stars appear to be created when an AGB S or C star transfers mass via the stellar wind or a Roche-lobe overflow to a companion that is converted to a Barium star with the core of the former AGB star remaining as a white dwarf. This hypothesis predicts that the compositions of AGB S and C stars and the Barium stars should be similar. This prediction is confirmed observationally (Lambert 1985, 1988).

Classical Barium stars for which the s-process elements are overabundant by a factor of 3 to 10 are readily identified on classification spectra. Mild Barium stars are not so easily identified and secure identification may require inspection of higher dispersion spectra. What is the frequency distribution for Barium stars of differing degrees of s-process enrichment? Can a single nucleosynthetic history account for all species of Barium stars? Before the mass-transfer hypothesis was identified, we suggested that the mild Barium stars might be quite normal giants whose parental clouds happened to be slightly more polluted with s-process elements than more typical clouds (Sneden, Lambert, and Pilachowski 1981). (This suggestion arose in part because the C abundances of mild Barium stars were very nearly normal but this is predicted on the mass-transfer hypothesis.) Is mass transfer across a binary the sole method for production of a Barium star? Perhaps, the He-core flash can trigger the running of the s-process. This flash was a leading suspect before McClure's discovery that Barium stars were spectroscopic binaries.

To investigate these and other questions, McWilliam (1988) undertook a survey of s-process elements in 570 G-K giants. Three 100 Å intervals were observed to provide a selection of s-process and iron-group lines that yielded abundance ratios s/Fe that are minimally dependent on the adopted effective temperature and surface gravity. In fact, the dominant error was, in general, provided by the uncertainties of the measured equivalent widths. Effects of systematic errors and particularly their variation across the HR diagram were minimized by examining stellar samples drawn from restricted regions of the diagram. In Figure 5, I show a series of labelled areas (A through G) whose combined area contains 70 % of the observed stars. Of the

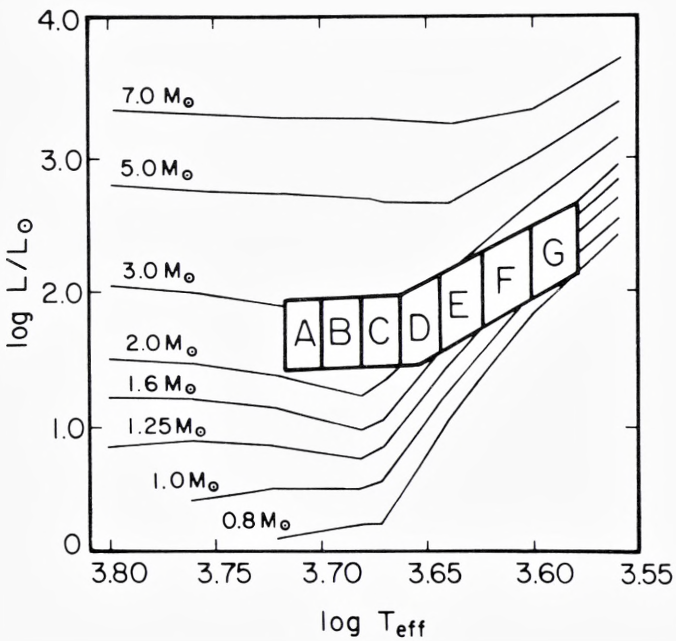

Fig. 5. A H-R diagram showing the subgroups A through G defined by McWilliam (1988).

other 30 %, approximately 60 % lie in the HR diagram at luminosities above that of the strip. My brief discussion is restricted to the Y/Ti abundance ratio of stars in the areas A-G.

Table 2
Abundance and other data for the groups A-G

Group	No. Stars	T_{eff}(K) Range	[Fe/H]	$\varepsilon(Y)/\varepsilon(Ti)$	Ba[b] Limit
A	35	5000-5200	-0.14 ± 0.14[a]	-2.73 ± 0.12[a]	-2.50
B	67	4800-5000	-0.16 ± 0.15	-2.73 ± 0.17	-2.50
C	74	4600-4800	-0.15 ± 0.14	-2.83 ± 0.13	-2.61
D	63	4400-4600	-0.09 ± 0.14	-2.86 ± 0.13	-2.65
E	55	4200-4400	-0.14 ± 0.16	-2.81 ± 0.19	-2.60
F	49	4000-4200	-0.14 ± 0.15	-2.85 ± 0.09	-2.65
G	52	3800-4000	-0.14 ± 0.12	-2.77 ± 0.11	-2.56

[a] 3σ estimate from the frequency distribution of abundances.
[b] See text.

Derived abundance ratios and other information for the 7 groups are summarized in Table 2. The frequency distributions for groups A+B, F, and G are shown in Figure 6. It is clear that the Y/Ti abundance ratios are similar across the groups. There is a hint that the groups A+B containing the hottest giants have a slightly higher mean Y/Ti ratio and a smaller dispersion. These differences are about equivalent to a 3 σ event. One may speculate that these differences are linked to the identification of the A+B stars as clump or He-core burning giants; i.e., the He-core flash that initiates He-core burning mixed s-process products into the giant's convective envelope. In offering this speculation as a hypothesis to be tested by additional observations, McWilliam (1988) notes several confirmatory hints.

First, the frequency distribution of the A+B group is slightly narrower than that of the cooler groups. Since the former is probably dominated by He-core giants and the latter groups are a mix of pre- and post-He core flash giants, the larger dispersion for the cooler stars could be the result of a superposition of two displaced distributions with the He-core flash inducing a s-process enrichment in post-He core flash giants.

Second, the Y/Ti abundance ratio of the subgiants is similar to that for the cooler groups that may contain a high proportion of pre-He core flash giants. For 41 stars with 4800 K $< T_{eff} <$ 5200 K and $\log L/L_\odot < 1.30$ (this is about 0.3 dex below the minimum luminosity of a He-core burning giant), the mean $\log(\varepsilon(Y)/\varepsilon(Ti)) = -2.86$ which differs at about the 3σ level from the value (-2.73) for the A+B group.

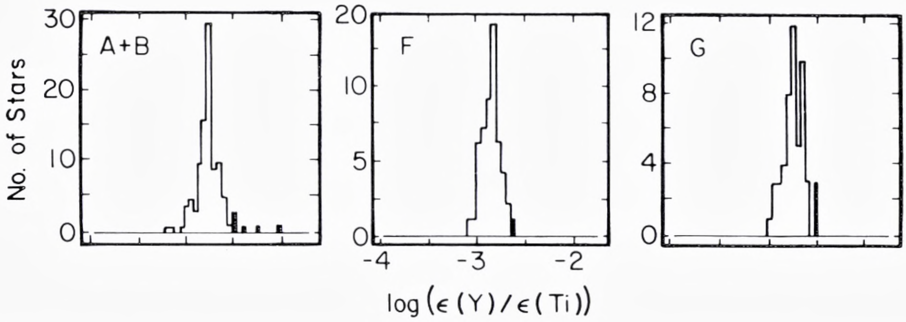

Fig. 6. Histograms of the Y/Ti abundance ratio for subgroups A+B, F, and G. The shaded portion of each histogram denotes the Barium stars that satisfy McWilliam's definition (see text).

Third, the frequency of Barium stars is higher in the A+B groups (see Figure 6). McWilliam defines a Barium star by applying a form of Chauvenet's criterion to the frequency distributions. A star is deemed a Barium star if the probability of the measurement of the Y/Ti ratio is greater than half the observed frequency of that measurement occurring by chance. The latter is calculated on the assumption that the frequency distribution is Gaussian with a dispersion derived from the distribution's width (FWHM); the lower limit to the abundance ratio for a star to be termed a Barium star is given in the far right-hand column of Table 2. These Barium stars are identified as the shaded portions in Figure 6. These Barium stars are predominantly mild Barium stars but the few classical Barium stars in the sample were extracted successfully. It is clear from the H-R diagram (Figure 7) that the Barium stars are concentrated to the area populated by the He-core burning or clump giants. Among the more luminous stars is the classical Barium supergiant ζ Cap. The frequency of Barium stars in McWilliam's sample of 568 giants and with his definition of a Barium star is 4.2 %.

As noted earlier, the mass-transfer hypothesis accounts satisfactorily for the classical and most probably a majority of the mild Barium stars. It is not yet clear whether all of the mildest of Barium stars uncovered by McWilliam's survey can yet be ascribed to mass transfer. McWilliam noted that a majority of his stars are reported to have a variable radial velocity by the *Bright Star Catalogue* and several are known as spectroscopic binaries. If, on close scrutiny, the vast majority are revealed to be spectroscopic binaries, they may be identified as products of mass-transfer across a binary system. If, however, a significant minority show no radial velocity variations to within a narrow limit, alternative hypotheses will need to be invoked; e.g. (i) contamination of the giant's envelope by s-process products at the time of the He-core flash; (ii) a 'cosmic' dispersion of the Y/Ti (equivalently, s/Fe) ratios in interstellar clouds as a result of their pollution by ejecta from evolved stars. A survey of elemental

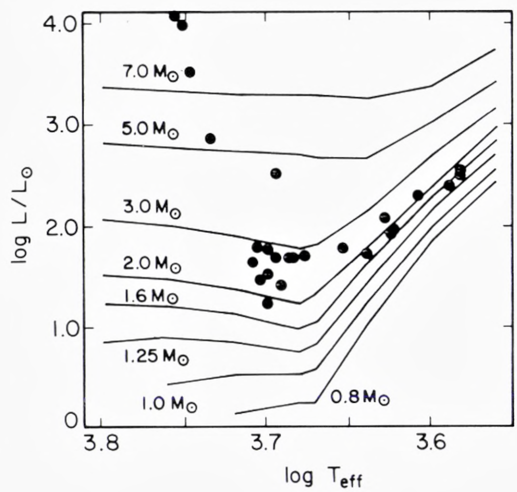

Fig. 7. A H-R diagram showing the locations of the Barium stars in McWilliam (1988).

abundances in a large sample of *single* main sequence stars would provide an estimate of the cosmic dispersion. Subgiant and even main sequence Barium stars are known (Bond 1974; Tomkin *et al.* 1989) and the majority are spectroscopic binaries (McClure 1985). An examination of a larger sample of subgiants (log $L/L_\odot \lesssim 1.3$) would also be instructive. If the He-core flash is the culprit creating mild Barium stars, the frequency of Barium stars among subgiants should be lower than is observed for group A+B.

2.3 *High-resolution Stellar Spectra*

Perhaps, the principal motivation for acquiring and analysing high-resolution stellar spectra is to detect elusive elements and isotopes whose lines, often weak, fall in crowded spectral regions or are blended with lines with more abundant isotopes. Other motivations exist and are certainly important: e.g., the measurement of intrinsic line profiles in order to characterize the atmospheric velocity field (microturbulence, macroturbulence, granular velocities, radial and non-radial pulsations).

2.3.1. *Thorium and the Age of the Disk*

Thorium is certainly an elusive element for the stellar spectroscopist. The Th II resonance line at 4019.1 Å is a weak line in the solar spectrum and blended with a Co I line. The solar Th abundance derived from the 4019 Å line is within 0.04 dex of the meteoritic value (Anders and Grevesse 1989). This agreement between solar and meteoritic Th abundances is expected and shows that the Th II contribution to the

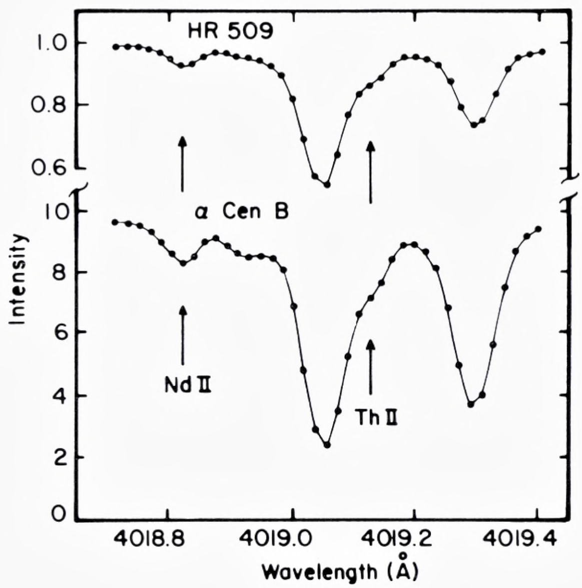

Fig. 8. High-resolution ($\lambda/\Delta\lambda \simeq$ 100,000) spectra of two dwarfs in the region of the Th II 4019 Å resonance line (after Butcher 1987, 1988). The neighbouring Nd II line is identified.

solar line has been correctly assessed. (If the Co I line's contribution to the solar line is overlooked, the solar Th abundance is 0.10 dex larger than the meteoritic value).

Thorium is of a special interest because it represents a 'nuclear cosmochronometer'. The only isotope expected to be present in any but the most peculiar of stars is ^{232}Th with a half-life of 14 Gyr. Since Th is the only known nuclear cosmochronometer for material outside the solar system, accurate determinations of the Th abundance in stars are of great interest because they may lead to an estimate of the Galactic age that, unlike the fitting of colour-magnitude diagrams, is independent of stellar evolution theory.

Recently, Butcher (1987, 1988) obtained high-resolution ($\lambda/\Delta\lambda \simeq$ 100,000) spectra near 4019 Å with ESO's fine Coudé Auxiliary Telescope and associated Coudé Echelle Spectrometer. Lower main-sequence disk stars of differing (stellar evolutionary) ages were observed. The blended Th II line is shown in Figure 8 for two dwarfs. It is clear that, without high resolution ($\lambda/\Delta\lambda \gtrsim$ 100,000), the Th II (+ Co I) line would be irretrievably blended with neighbouring lines. Butcher's simple but probably adequate method of abundance analysis gave the Th/Nd abundance ratio. He chose Nd as the reference element because a low excitation Nd II line of comparable strength to the Th II line occurs only 0.3 Å to the blue (see Figure 8). The Th II and Nd II lines fall between the same "two excellent continuum windows". Conversion of the equivalent widths of these lines to a Th/Nd ratio should be insensitive to the continuum placement and to the adopted effective temperature and surface gravity of the stars. The striking result of this pioneering survey is that the Th/Nd ratios are independent of the stellar evolutionary age except perhaps for a slight increase beyond 15 Gyr.

If the Th cosmochronometer is to be exploited, a model of the rates of nucleosynthesis within the Galaxy must be adopted and its predictions fitted to the observed Th/Nd ratios. Thorium is synthesized by neutron capture in the r-process. In solar system material, the r-process is responsible for about half the Nd with the s-process providing the other half. On the assumption that Th and Nd have been synthesized at a constant rate, Butcher (1987) gave an upper limit to the Galactic (nucleosynthetic) age of about 10 Gyr. Clearly, this age is in conflict with the stellar evolutionary ages, but is consistent with the age for the disk, $T_0 = 9.3 \pm 2.0$ yr, derived from the observed luminosity function for white dwarfs (Winget *et al.* 1987). Clayton (1987, 1988) attempted to recognize that the yields of the r- and s-processes might vary differently with age (i.e., metallicity). For example, he considered a model in which the r-process abundances are primary and the s-process abundances are secondary with their relative yields adjusted to provide solar s/r ratio for Nd. This model fits the observed Th/Nd ratios using the stellar evolutionary ages. In a related approach, Mathews and Schramm (1988) conclude that "a model dependent (3σ) best upper limit of $T_0 \simeq 18$ Gyr to the age of the galaxy is derived". Butcher's (1988) riposte to these alternative models of the chemical evolution of heavy elements is direct: such models do not reproduce the observed constant ratio of Eu/Ba over the range in [Fe/H] spanned by disk stars. Eu is an almost pure r-process and Ba an almost pure s-process element.

Malaney and Fowler (1989) concoct a recipe for primary s-process production by AGB stars such that the r/s ratio of disk stars is constant. Their model of chemical evolution combines early spikes in the production of the r- and s-process elements with a subsequent period of nucleosynthesis in which the r/s ratio is maintained fixed. This model leads to "a relatively young age of $\lesssim$ 12 Gyr for the galaxy".

Two points about Malaney and Fowler's approach deserve comment: (i) this empirical approach to the observed r/s ratio yields a fair fit to the Eu/Ba observations. The galactic age is obtained by fitting predicted Th/Nd vs age relations to the observations after a simple scaling of the stellar evolutionary ages: $t = t_{ev}\, T_o/20$ where T_0 (in Gyr) is the galactic age and t_{ev} is the evolutionary age quoted by Butcher. However, if the evolutionary ages are indeed in error, it is not clear that this simple scaling will be an adequate correction. A slight stretching of the correction with age or metallicity could lead to a quite different T_0 in light of the fact that the predicted Eu/Ba and Th/Nd ratios vary little in the range 5 to 20 Gyr. (ii) Although both Eu and Th are assigned to the r-process, there is no unique prescription or even well established series of recipes for the r-process, and, hence, the relative yields of Eu and Th could vary from site to site and with galactic age. Without tighter observational constraints on the r-process yields, an even larger uncertainty must be attached to applications of the Th cosmochronometer.

(Butcher notes that his observation of a uniform Th/Nd ratio independent of a stellar age is consistent with the synthesis of all heavy elements in a 'spike' before the onset of star formation. This conclusion overlooks the fact that the abundances of the r- and s-process along with those of the lighter elements have increased by an order of magnitude since the formation of the disk. Continuing synthesis is required to account for this increase. If early synthesis of Th and Nd (and related heavy elements) is ascribed entirely to an initial "spike", the abundance ratio Th/Nd would be held constant, but their abundances relative to lighter elements (e.g., Fe) would decline in contrast to the observations).

Observers should be encouraged to pursue this unique cosmochronometer by extending the measurement of the Th II line to more metal-poor stars. Butcher's sample contained no stars with $[Fe/H] < -0.8$. Examination of Th in samples of old disk and halo stars would be informative because studies show that the r-process is the dominant source of heavy elements for $[Fe/H] \lesssim -2$ (Gilroy *et al.* 1988). An exhaustive search might reveal lines attributable to other very heavy elements whose abundance could constrain the r-process yield of Th. Finally, the prescriptions of chemical evolution must be refined by observations in order that Th may be a more accurate cosmochronometer. Thorium deserves a place on the list of the contributions from high-resolution spectroscopy to "Recent Progress on Future Possibilities" in stellar and galactic nucleosynthesis.

2.4 *Selection of Spectral Interval*

Of the several expanded opportunities now available to the observer, the ability to select the wavelength interval appropriate to a particular problem is of especial importance. In 1958, the stellar spectroscopist was constrained by the available photographic emulsions to the blue and red. Today, exploration beyond the red is made routinely with solid-state detectors: the near-infrared is accessible with CCDs and Reticons and a variety of detectors are in use for infrared (1-10μm) spectroscopy. Ultraviolet spectroscopy is now possible with the IUE satellite. Launch of the Hubble Space Telescope will greatly enhance the opportunities of ultraviolet stellar spectroscopy. In the area of 'future possibilities' for ground-based stellar spectroscopy, I would note the appearance of cryogenic echelle spectrometers with infrared array detectors as perhaps the most significant of the spectrometers now in development. One could devote an entire review to a discussion of how access to the entire (almost!) electromagnetic spectrum permits observational solutions to challenging problems in stellar physics. Here, I must confine my remarks to two topics: a search for meridional mixing in rapidly-rotating early-type stars, and detection of H- and He-burning products in the atmospheres of red giants. This pair of topics has one common feature: neither can be addressed through photographic spectra of the blue and red.

2.4.1 *Meridional mixing in rapidly-rotating early-type stars*

Do meridional currents within a rapidly rotating star mix the envelope and hot interior and, hence, change the chemical composition of the stellar atmosphere? Paczyński (1973) predicted that the surfaces of 3-10 $M_\odot$ stars rotating at typical velocities should be significantly deficient in ^{12}C and enriched in ^{14}N as a result of CN-cycling at the deepest points reached by the mixing currents. Oxygen was predicted not to decrease because the meridional currents did not extend to the deeper interior when ON-cycling operates.

In subsequent theoretical explorations, penetration of the currents was suggested to be so restricted by the gradient in mean molecular weight (μ) introduced by nuclear burning of hydrogen, that surface changes in ^{12}C and ^{14}N could be much less than suggested by Paczyński, who did not explicitly account for the inhibiting effects of the μ-gradient. Tassoul and Tassoul (1984) show that the partial penetration of the μ-gradient is possible and ought to lead to observable changes in the surface C and N abundances. Noting that predictions of the abundance changes are sensitive to "too many unknown parameters" (e.g. how does the interior rotate? Is the core rotating rapidly?), Tassoul and Tassoul concluded with a prophecy "one may thus look forward to the time when the surface abundances of C, N, and O will be used to probe the star's inner rotation and concomitant turbulent eddies."

This prophecy stands as a challenge to the observer. Until very recently, the only observational search for the predicted C deficiency appears to have been by Preston and Paczyński (1974) who measured the equivalent width of the C II 4267 Å doublet in B3 to B5 stars and found it and, hence, the carbon abundance to be independent of the projected rotational velocity (v sin i) over the range 20-300 km s^{-1}. Two recent searches for meridional mixing have exploited the enhanced opportunities listed in the Introduction. 'Selection of wavelength interval' is the key to these searches. In the rapidly rotating late-B and early-A stars, a search for a C-deficiency must exploit strong carbon lines falling either in the near-infrared (C I lines at 9000-11000 Å) or the ultraviolet (C II lines at 1335 Å, C I lines at 1657 Å).

Since the near-infrared C I lines become very shallow features in the most rapidly rotating stars, access to the near-infrared must be combined with high S/N in order to determine the C abundance in these stars. In our search (Lambert, McKinley, and Roby 1986), we obtained Reticon spectra of C I 9100 Å lines in 22 early-A stars main sequence stars with projected rotational velocities of up to 180 km s^{-1}. Although several C-deficient stars were found, the C abundance did not show the decline with increasing v sin i that might be expected from meridional mixing. Note that we selected early A stars because their longer main-sequence lifetime may enable the slow meridional currents to produce compositional changes more severe than those expected in the shorter-lived B stars.

Hardorp *et al.* (1986) and Cugier and Hardorp (1988) used IUE high-resolution spectra to derive C abundances for a sample of B3 to A0 main sequence stars with projected rotational velocities of up to about 300 km s^{-1}. The primary lines were the C II 1335 Å multiplet and the more recent results are based on a non-LTE analysis of carbon line formation. A majority of the stars have a near-solar C abundance with an interesting minority showing as severe a C deficiency as other examples reported by Lambert, McKinley and Roby (1986); e.g. ψ^2 Aqr (B5V, v sin i = 280 km s^{-1}) has log $\varepsilon(C) = 7.1 \pm 0.2$ or $[C/H] = -1.5$ relative to the Sun (Cugier and Hardorp 1988). It must be noted that the minority of the C-deficient stars are found over the entire range of projected rotational velocities. Most rapidly rotating (i.e. high v sin i) stars have a normal C abundance.

Two possible origins for the C-poor stars may be suggested as the basis for additional spectroscopic tests:

(i) *Meridional Mixing.* A naive view is that the efficiency of the meridional mixing currents in mixing CN-cycled material to the surface is fully determined by the star's mass, age, and rotational velocity. Then, the C deficiency is expected to increase with increasing projected rotational velocity: the upper envelope to the $\varepsilon(C)$ –vsin i distribution should be well defined with lowest $\varepsilon(C)$ at highest vsin i. Rapidly rotating stars seen pole-on will show a low vsin i and a low C abundance. This simple picture is not supported by the observations. It could be that the "too many unknown parameters" sketched by Tassoul and Tassoul (1984) are effective in controlling the meridional mixing and, hence, the lower surface C abundances may not show a simple correlation with the observed vsin i. One helpful test would be to measure the N abundance in the C-poor (and other) stars. If deep mixing has led to the C-deficiency, the surface must now be enriched in N. Of course, the deep mixing need not necessarily be the exclusive consequence of meridional mixing. Hardorp *et al.* (1986) intimate that at least some of the C-poor stars are not N-rich relative to normal stars of the same type; the ultraviolet N I lines are not strengthened in these stars.

(ii) *Chemically Peculiar Rapidly-rotating Stars.* Chemically peculiar (CP) stars are known to include stars having substantial C-deficiencies (Roby 1987; Roby and Lambert 1990). Since the distinctive lines by which CP stars are classified are very shallow in spectra of stars having a large vsin i, one may suspect that some CP broad-lined stars have gone undetected; e.g. the Hg II and Mn II lines that define the Ap (HgMn) class may escape detection at classification dispersions when vsin i $\gtrsim$ 100 km s^{-1} (Preston 1974). If the C-poor stars identified in the recent surveys are CP stars, high S/N spectra should now be capable of detecting the various trademarks of Bp, Ap and Am stars. Indeed, a survey of rapidly rotating A and B stars for evidence of chemical peculiarities could provide a valuable database for testing theories on the origins of CP stars.

2.4.2 *Red Giants and Dredge-Up*

To the spectroscopist fascinated by the coolest red giants, the infrared is of critical interest because

(i) these giants emit most of their flux in the infrared, and

(ii) the infrared contains the molecular transitions (Table 3) that provide the C, N, and O elemental and isotopic abundances that are key monitors of the dredge-up from H and He burning shells.

Table 3
Indicators of CNO Abundances in Red Giants

Spectral Type	Primary Lines[a]	Secondary Lines[a]	Refs.[b]
G and K	*C_2 Swan $\Delta v=0,-1$	*CH A-X $\Delta v=0,-1$	LR,K,G
	*Cn Red $\Delta v=4$ to 2	CN Red $\Delta v=-1,-2$	
	*[OI] 6300 and 6363 Å	*CN Violet $\Delta v=0,-1$	
		CO V-R $\Delta v=3$	
		OH V-R $\Delta v=2$	
M,MS,S	CO V-R $\Delta v=3,2$	H_2O V-R	SL
	OH V-R $\Delta v=2,1$		
	NH V-R $\Delta v=1$		
	CN Red $\Delta v=-2$		
SC,C	CO V-R $\Delta v=3,2$	C_2 Ballik-Ramsay	DWS
	*C_2 Phillips $\Delta v=3,2,1$	*CN Red $\Delta v\geqslant 0$	
	C_2 Phillips $\Delta v=-2$	CH V-R $\Delta v=1$	LGEH
	CN Red $\Delta v=-2$	NH V-R $\Delta v=1$	
		HCN V-R $\Delta v=1$	

[a] Unless marked by an asterisk the indicated lines are in the infrared ($\lambda \gtrsim 1.3\ \mu m$).

[b] DWS = Dominy, Wallerstein, and Suntzeff (1986)
LGEH = Lambert *et al.* (1986)
SL = Smith and Lambert (1985, 1986, 1990)
G = Gratton (1985)
K = Kjærgaard *et al.* (1982)
LR = Lambert and Ries (1981)

In the determination of C, N, O abundances in the warmer (G and K) red giants, a few atomic transitions provide useful data and the visible and near-infrared provides an adequate set of molecular transitions – see the references given with Table 3. For the cooler red giants, the atomic lines are unuseable because molecular formation depletes the partial pressure of the atoms and the weakened atomic lines fall in regions of intense molecular absorption. Since the useful lines of the various molecules containing C, N, and O are in the infrared, spectra in that region are an essential prerequisite for a C, N, and O analysis. Recent papers on the oxygen and carbon rich cool red giants (see Table 3) show how high resolution infrared spectra, model atmospheres, and a suite of basic molecular data are combined to yield the C, N, and O elemental and isotopic abundances.

Here, I comment briefly on how analysis of infrared spectra led to the resolution of a puzzle of long-standing: What is the $^{12}C/^{13}C$ ratio in cool carbon stars? These carbon stars and their C_2 Swan system bands provided the first detection of ^{13}C in extraterrestrial objects, but the same characteristic that underlay this early discovery – namely, the great strength of C_2 Swan bands – bedevils attempts to obtain an accurate estimate of the $^{12}C/^{13}C$ ratio. The same comments apply to the CN Red system lines in the visible and near-infrared which have proven a popular source of a $^{12}C/^{13}C$ ratio. A compilation of published ratios would bewilder the reader unfamiliar with the spectra (see the summary in Lambert 1980): estimates for a single star may range from $^{12}C/^{13}C$ = 3 to 100! The root cause of the general lack of agreement is not that the ^{13}C ^{12}C or ^{13}CN lines are extremely weak but that they and their $^{12}C_2$ and ^{12}CN counterparts are strong.

It is important to obtain accurate estimates of the $^{12}C/^{13}C$ ratio because it is a valuable monitor of the material dredge-up into the giant's convective envelope and, hence, the atmosphere. Two limiting cases may be identified;

(i) A ratio $^{12}C/^{13}C \sim 50$ or so would suggest that pure ^{12}C, a product of He-burning, has been added to the atmosphere to convert the O-rich star to a C-rich star without appreciable conversion of the freshly synthesized ^{12}C to ^{14}N by the CN cycle.

(ii) A ratio $^{12}C/^{13}C \sim 3$ would suggest exposure to the CN cycle. Quite a weak exposure may suffice to give $^{12}C/^{13}C \sim 3$ but if the CN cycle operates for an extended period at a sufficently high temperature, this cycle can also reduce the O abundance to create a C-rich envelope. Extended operation at lower temperatures will convert the C-rich envelope back to an O-rich envelope.

Visible and near-infrared spectra of carbon stars are so crowded with strong lines that the location of the continuum level is uncertain and weak lines are rare. These spectroscopic facts of life surely account for the discrepancies between observers who may have analysed spectra of similar quality. In selected intervals of the infrared, the line density is lower and several favorable opportunities exist for a measurement of the $^{12}C/^{13}C$ ratio. We exploited the CN Red system $\Delta v = -2$ lines near 2μm and the CO V-R $\Delta v = 2$ and 3 lines (Lambert *et al.* 1986).

We described two methods of extracting the $^{12}C/^{13}C$ ratio from the CO V-R lines. The $\Delta v = 3$ bands near 1.6 μm provide weak ^{13}CO lines. A comparison of weak ^{12}CO and ^{13}CO lines within this sequence necessarily pairs high excitation ^{12}CO and lower excitation ^{13}CO lines and, hence, the $^{12}C/^{13}C$ ratio is dependent on the temperatures in the line-forming region; i.e., the effective temperature and the chemical composition which, through the line blanketing, influences the temperature profile. An alternative scheme combines the ^{13}CO lines from the stronger (i.e., larger f-value) $\Delta v = 2$ bands near 2.5 μm with the ^{12}CO lines at 1.6 μm, and then ^{12}CO and ^{13}CO lines of similar excitation potential are compared and the sensitivity of the $^{12}C/^{13}C$ ratio to the atmospheric structure is slight. These two methods yield similar results.

In the CN $\Delta v = -2$ sequence near 2 μm, weak ^{12}CN lines are identifiable; most are satellite lines and a few are high rotational members of the main (P, Q, R) branches. In the typical carbon star, weak ^{13}CN lines from the main branches are present in large numbers. The difference in f-values between the main and satellite lines is such that for a $^{12}C/^{13}C$ ratio of about 20 to 40, the satellite ^{12}C line and typical ^{13}CN line have comparable (and small) equivalent widths. Since the lines also have similar excitation potential, the derived $^{12}C/^{13}C$ ratio is insensitive to the adopted excitation or effective temperature. The advantage gained by combining satellite ^{12}CN and main ^{13}CN lines was noted first by Fujita and his colleagues in analyses of near-infrared CN ($\Delta v = +2$) lines – see, for example, Fujita and Tsuji (1977). Indeed, our results confirm the suggestions by Fujita and colleagues that carbon stars have, in general, a low ^{13}C content.

The CO and CN infrared lines yield consistent results. Our new results show that the typical cool carbon star has a higher $^{12}C/^{13}C$ ratio than the M giants from which the star evolved. If J-type stars rich in ^{13}C are excluded, the mean is $^{12}C/^{13}C$ = 60. Inspection of the $^{12}C/^{13}C$ and $^{12}C/^{16}O$ ratios shows that they are consistent with the hypothesis that the carbon stars were produced from M giants by the third dredge-up on the AGB of nearly pure ^{12}C. For the J(^{13}C-rich) carbon stars, our infrared spectra confirm many earlier claims the ^{13}C-rich carbon stars exist; for example, for RY Dra, T Lyr, Y CVn, the $^{12}C/^{13}C$ ratios (3.6, 3.2, and 3.5, respectively) are not significantly different from the predicted ratio (3.4) for the CNO cycle in equilibrium. A fourth star commonly put with the above trio is WZ Cas, but its ratio, $^{12}C/^{13}C$ = 4.5, is distinctly above this predicted value. These low $^{12}C/^{13}C$ ratios surely denote severe contamination of the envelopes with CN-cycled material. When and where this contamination occurs in the life of the J-type cool carbon star remains obscure even when additional clues offered by the C, N, and O elemental and $^{16}O/^{17}O/^{18}O$ ratios are provided. Prospective candidates for the progenitors of the cool J stars may be found among the early R stars which are carbon-rich giants with the effective temperature and luminosity of K giants. Evolution of such R stars into the cooler and more luminous cool J carbon stars is seemingly inevitable. The origin of the early R stars is unclear. Dominy (1984) suggested that a violent He-core flash in low mass stars led in a few rare cases to release of freshly synthesized carbon into the envelope of the He-core burning clump giant. Earlier I stressed the vital contributions of studies of the Magellanic Clouds to our understanding of stellar evolution. Future spectroscopic studies of carbon-rich stars at pre- and post-AGB phases of evolution are sure to shed light on the origins and history of the ^{13}C-rich and other types of carbon stars.

2.5 *High S/N Spectra*

Several of the preceding discussions involved a pairing of the enhanced opportunity to acquire high S/N spectra with the opportunity that was the focus of that section. The opportunity to obtain high S/N spectra is, in the pursuit of stellar chemical compositions, generally exploited in the detection of weak lines from trace species; e.g. lithium and the 6707 Å Li I resonance doublet in metal-poor dwarfs, thorium and the 4019 Å Th II resonance line. Other applications of high S/N spectra with relevance to the determination of chemical compositions include the accurate measurement of weak lines of common species in order to determine a star's effective temperature and surface gravity, and the accurate definition of line profiles in order to characterize the atmosphere's velocity field.

Of the many examples that might be discussed, I close this discussion of the opportunities that will define "future possibilities" with a commentary of the oxygen abundance in young and old stars. Data on the abundance of oxygen, the third most abundant element after H and He, are of especial interest to studies of the chemical evolution of the Galaxy and the ages of stars. Models of chemical evolution must account for the overabundance of O (relative to Fe) in metal-poor stars; the standard explanation is that massive stars, which are leading producers of oxygen, were more common in earlier generations of stars. With its high abundance, O is a leading contributor to the opacity of stellar interiors and, hence, an influence on main sequence (and other) lifetimes. Ages of stellar clusters derived from fitting theoretical isochrones to color-magnitude diagrams of the main sequence and subgiant branch are dependent on the assumed O abundance. Since a cluster's metallicity is readily derivable but its O abundance may be unknown, the assumption about the O abundance reduces to one about the O/Fe ratio. Current choices for this ratio are usually based on recent abundance analyses of O in dwarfs and giants. A cluster's derived age may be reduced by 2-4 Gyr as [O/Fe] is raised to the upper limit set by the available observations. With [O/Fe] at the upper band, the ages of the most metal-poor globular clusters are near 14 Gyr according to standard models of stellar interiors and evolution (VandenBerg 1988).

The presence of oxygen in the atmosphere of a metal-poor dwarf or giant is betrayed by few O I lines; e.g. the [O I] 6300 and 6363 Å lines for giants and the O I 7770 Å triplet for dwarfs. The need for high S/N spectra is well demonstrated by Barbuy's (1988) montage of high-resolution spectra of the [O I] 6300 Å line in halo giants. The [O I] line, the stronger of the two forbidden lines, has a central depth of about 20 % ($W_\lambda \simeq 50$ mÅ) in the most metal-rich halo stars ([Fe/H] ~ -1.2). In the most metal-poor giants observed by Barbuy, this line has a central of only about 3 % ($W_\lambda \simeq 8$ mÅ) at [Fe/H] ~ -2.5. Since one would like to extend the determinations of the O abundance to the most metal-poor giants known ([Fe/H] ~ -4), it is clear that high S/N is essential. In addition, high S/N spectra would permit the detection of the [O I] lines in subgiants and dwarfs where, of course, the predicted equivalent is much smaller than for a giant of comparable metallicity and temperature; the 6300 Å line is reduced to $W_\lambda \sim 2.5$ mÅ for main sequence ($\log g \sim 4$) stars as metal-rich as [Fe/H] ~ -0.6 (Barbuy and Erdelyi-Mendes 1989).

Oxygen in halo dwarfs and subgiants is detectable through the high excitation triplet of O I at 7770 Å and other similar lines. Observations of the 7770 Å lines provided the first real evidence of an O over abundance (relative to Fe) in halo stars (Sneden, Lambert, and Whitaker 1979). The strength of the lines is sensitive to effective temperature. In the warmer stars, these O I lines should be measureable off high S/N spectra in the most extreme halo stars; Sneden *et al.* found $W_\lambda = 19$ mÅ for the combined triplet in HD 140283 with [Fe/H] $\simeq -2.3$.

An association of the [O I] lines with giants and the O I lines with dwarfs should not be considered immutable. When possible, O I and [O I] lines should be observed and analysed in the same objects. These lines offer different advantages and disadvantages. In particular, the high excitation O I lines may be susceptible to non-LTE effects but the [O I] lines are expected to be formed close to LTE. Preliminary theoretical studies of the non-LTE effects on the 7770 Å lines in halo dwarfs were reported by Sneden *et al.* (1979). Empirical evidence for non-LTE enhancement of the 7770 Å lines in F and G dwarfs was given by Clegg, Lambert and Tomkin (1981) who noted that these lines when strong gave a systematically higher (LTE) abundance than other weaker permitted lines. Although an improved theoretical estimate of the non-LTE effects on the O I lines is now achievable, I would suggest that a thorough application of high S/N spectroscopy to the [O I] lines in metal-poor dwarfs and giants is likely to lead to the most reliable estimates of the O abundance. Finally, complete dependence on the O I spectrum may be eliminated through observations of the OH $A^2\,\Sigma^+ - X^2\Pi$ ultraviolet system (Bessell and Norris 1987).

The reader interested in the run of the O/Fe ratio with metallicity is referred to reviews by Lambert (1989) and Wheeler, Sneden, and Truran (1989) and to recent work by Barbuy (1988) and Barbuy and Erdelyi-Mendes (1989).

3. *Basic Atomic and Molecular Data*

3.1 *Introduction*

In the preceding sketches of astrophysical problems that may be addressed through stellar spectroscopy, I placed the emphasis on the expanded opportunities available to the observer. The methods of extracting the chemical composition from the spectra were not discussed. To conclude this essay, I offer a few illustrations of the spectroscopists' need for the accurate data on atoms and molecules of astrophysical interest which are essential components in the analytical techniques linking spectra and compositions. The data are used both in the construction of model stellar atmospheres and in the applications of the models to the computation of the synthetic spectra to be fitted to the observed spectrum. A contributing factor to my enthusiasm for quantitative stellar spectroscopy is the fact that accurate basic atomic and molecular data are being made available in increasing quantities. Rapid growth of our understanding of atoms and molecules is stimulating refinements of the analytical techniques applied to stellar spectra. I illustrate these refinements with commentaries on three recent studies of non-LTE line formation.

As long as the assumption of LTE is retained, the basic data needed by a spectroscopist includes the observed (and predicted) spectrum, the associated term diagram (i.e., excitation and ionisation/dissociation energies) and the transition probabilities for emission and absorption of photons in transitions between the terms, including continuum processes as well as lines. The extent of the required data varies from species to species and with the particular spectroscopic problem under consideration.

When the assumption of LTE is discarded, lists of basic data must be enlarged to include cross-sections for interactions between the atom or molecule of interest and the abundant particles in the atmosphere. These interactions include those leading to internal excitation of a species and others resulting in a change of species (e.g. ionization and dissociation). Each interaction has a direct inverse interaction. Often, the dominant interaction is collisional excitation (or ionisation) of an atom by free electrons. In the cooler stars, the free electrons are greatly outnumbered by H and He atoms and, in the coolest stars, by H_2 molecules. Collisional excitation by these atoms and molecules has been widely supposed to be negliglible with respect to excitation by the electrons. Recently, Holweger and colleagues have challenged this supposition (see below). Their challenge means that non-LTE studies of atoms and molecules in cool stars will now require accurate rate constants for excitation by H, He, and H_2 in addition to electrons. Theoretical and experimental data for excitation by electrons is available for many atoms and generally successful approximations for rate constants may be used when detailed studies have not been reported. However, rate constants for electronic excitation by H and He atoms or H_2 molecules have not been determined theoretically or experimentally at the low energies of interest except for a few specific cases. This is virgin territory for a quantum or experimental chemist. (Excitation within the vibration-rotation ladder of a molecule is primarily by H and He atoms or H_2 molecules – see Hinkle and Lambert [1975]).

3.2 *Line Lists*

Stellar atmospheres constitute a family of spectroscopic sources that cannot be simulated in detail in the laboratory. As a result, stellar spectra contain many absorption and emission lines neither recorded on laboratory spectra nor predicted from the available sets of energy levels. These unidentified lines present a variety of problems to the stellar spectroscopist.

Suppose that a trace element is being sought whose only imprint on the spectrum is a single resonance line (e.g. Th II at 4019 Å or Li I at 6707 Å). An unidentified line that is blended with the resonance line will compromise the abundance determination of the trace element. The presence of an extra line will often be revealed on high resolution spectra. In rare cases the wavelengths of the lines will be coincident and the stellar line profile may not reveal the contaminant. As a recent example, I note that an investigation of the Li abundance in cool Ap stars was compromised by blending with an unidentified line (Mathys *et al.* 1989). Although empirical methods may be found to correct for the unidentified blend, the proper and accurate separation of the Li line's contribution to the blend can only come when the blending line is identified. This identification will require further intensive laboratory spectroscopy of the ions abundant in atmospheres of cool Ap stars.

One may identify a second class of problems in which a statistical representation of lines suffices and the precise wavelength of individual lines is unimportant. Detailed laboratory spectroscopy may not be needed in these cases; *ab initio* quantum predictions of the transitions' wavelengths and strengths may suffice and be more readily provided. These problems include the representation of the atomic and molecular line blanketing required in the computation of a model stellar atmosphere, and the prediction of a stellar spectrum in which the continuum is depressed by quasi – continuous opacity contributed by overlapping molecular lines.

For an example of the former problem, I cite the interpretation of the H_2 quadrupole vibration-rotation lines in the spectra of cool carbon stars. Goorvitch, Goebel and Augason (1980) noted that the H_2 lines in the spectra of cool carbon stars were much weaker than predicted. The authors suggested that the stars were H deficient. However, this conclusion is sensitive to the molecular line blanketing. If the blanketing is increased above the levels introduced by Goorvitch *et al.*, the atmosphere is further backwarmed and association of H into H_2 is hindered so that the predicted strengths of the H_2 lines are reproduced *without* the introduction of H deficiency. In our work on the carbon stars (Lambert *et al.* 1986), we suggested that the additional opacity overlooked in the early models came from HCN and C_2H_2, and probably C_3 too. With preliminary estimates of the HCN and C_2H_2 opacity, we were able to reconcile the predicted and observed H_2 lines and retain a normal He/H ratio. We do not yet have a fully consistent interpretation of the spectra of these cool carbon stars – in part, the situation is compromised by a lack of a detailed prescription of the molecules' vibration-rotation band strengths. This prescription is being supplied – particularly for HCN – by a combination of laboratory measurements and *ab initio* quantum chemistry calculation, notably by Uffe Jørgensen here at Nordita, and his colleagues (Jørgensen *et al.* 1985; Smith, Jørgensen and Lehmann 1987).

3.3 *Collision Cross-sections and non-LTE*

Where a statistical representation of a spectrum suffices, quantum chemistry may be tapped to provide the necessary wavelengths and the transition probabilities. However, when unambiguous identification and precise wavelengths are required, high-resolution laboratory spectra must be obtained. Many challenging problems may be posed to the experimental and laboratory physical chemists by the stellar spectroscopists. The challenge is amplified when studies of non-LTE line formation are considered. Examples drawn from the recent literature must serve to illustrate the prevalence of non-LTE effects and the concomitant demands for accurate atomic data.

A series of non-LTE studies of common lines in the spectra of B stars is being undertaken in Munich: see Becker and Butler (1988 a and b) on O II, Becker and Butler (1988 c and 1989) on N II and Eber and Butler (1988) on C II. A summary of the C II study is given here. About 100 levels of C^+ ion were included in the model atom: all energy levels up to n = 6 for terms converging to the $2s^2\ {}^1S$ ground state of C^{2+} and up to n = 4 for terms converging to the $2s^22p\ {}^3P^0$ level of C^{2+}. Additional C^+ levels (n = 7 and 8, n′ = 5 and 6) were included but with populations constrained to LTE values relative to the C^{2+} ground state. A total of 73 transitions were included in the linearization scheme providing the non-LTE populations. The calculation was primarily directed at the excitation equilibrium of C^+ but ionisation equilibrium was considered by including the above two levels of C^{2+} and the ground state of C^{3+}. Neutral C was ignored but it has a negligible abundance in the investigated atmospheres ($T_{eff} \gtrsim 15000$ K).

The equations of statistical equilibrium included radiative and collisional transitions as well as contributions from dielectronic recombination. Eber and Butler note that the radiative rates generally dominate the rates between levels and, hence, "errors in the collision constants are thus unimportant". Indeed Eber and Butler are content to adopt van Regemorter's (1962) formula or Allen's (1973) semi-empirical recipe for the rates for excitation by electron collisions and a comparable prescription (Seaton 1962) for ionization by electron collisions. Modern quantal calculations would yield more accurate results but the extensive computational effort would be hard to justify for this problem. (For C^{2+}, collisional excitation rates obtained by the R-matrix method (Dufton *et al.* 1978) were available and were adopted.) The radiative rates were computed, when possible, using the modern predictions of the oscillator strengths (e.g., Yu Yan, Taylor and Seaton 1987 for C II lines) and the photoionization cross-sections. In the absence of such predictions, the Coulomb approximation (Bates and Damgaard 1949) was used.

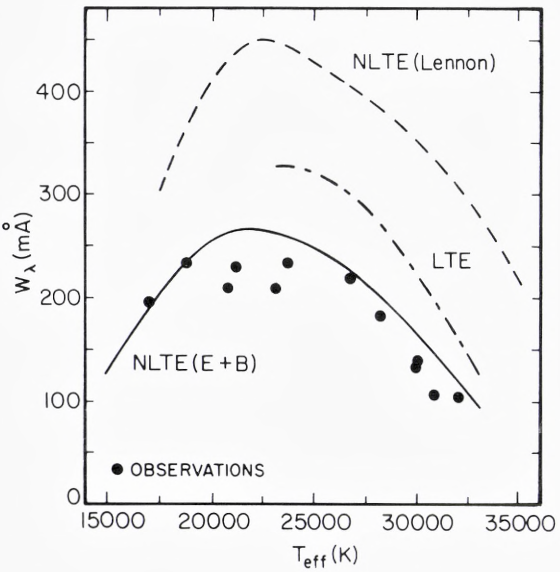

Fig. 9. Equivalent width predictions and observations for the C II 4267 Å 3d ^{2}D-4f ^{2}F multiplet. LTE as well as non-LTE predictions by Lennon (1983) and Eber and Butler (1988) are shown. The observations are from several sources – see Lennon (1983).

The non-LTE populations were used to predict the equivalent widths of a sample of 5 C II lines including the 4267 Å (3d ^{2}D-4f^{2}F) feature that is a prominent line in B-type spectra. Lennon (1983) in an earlier non-LTE study had pointed out that the observed equivalent width of the 4267 Å line was much weaker than the LTE and his non-LTE predictions. Eber and Butler point out that their non-LTE predictions for their more extensive model C^+ atom are close to the observed values (see Figure 9). (Lennon's model C^+ atom consisted of 14 doublet levels and the quartet levels were ignored. Sixteen transitions were linearized). The illustrated predictions correspond to a surface gravity $g = 10^4$ cm s^{-2}, and no microturbulence. The discrepancy between the observations and these non-LTE predictions line may be resolved through a combination of the following factors: further improvements to the non-LTE calculations including the use of improved model atmospheres, adoption of lower surface gravities and higher effective temperatures for the observed stars, and a stellar C abundance that is about 0.2 dex below the solar value. The stellar C abundance should, of course be based on a wide selection of the available lines of C II and C III. Indeed, Eber and Butler point out that the other C II lines considered by them are relatively insensitive to non-LTE effects and, hence, might be preferred for abundance studies.

These calculations by Eber and Butler resolve the discrepancy between the predicted and observed equivalent widths of 4267 Å line that was highlighted by Lennon who believed that the explanation lay in improving the model C^+ atom. Eber and Butler do not identify specific processes that control the non-LTE populations of the 3d ^2D or 4f ^2F levels but simply remark "there is no simple explanation for the different results of the current calculations compared with those of Lennon ... It would seem that the increased complexity of our model atom, compared to that of Lennon, is responsible for the improved results".

Since statistical equilibrium in cooler atmospheres than those considered by Eber and Butler is maintained with contributions from inelastic collisions, accurate rate constants must be known for all the important collisional processes. All investigations of the statistical equilibrium of atoms and ions include terms representing the inelastic collisions with free electrons. In cool stars, hydrogen atoms outnumber electrons by a considerable number (e.g., $n(H)/n(e) \sim 10^4$ in the upper solar photosphere) but inelastic collisions with H atoms have rarely been included in the equations of statistical equilibrium. In general, the justification for this omission is not given. There appear to be two factors whose product encourages the neglect of the H collisions. The collision rate (at unit density) is a product ($<\sigma v>$) (really an integral) of a cross-section (σ) and the relative velocity (v) of the target atom and the projectile (H atom or electron). For equal cross-sections, the rate of inelastic collisions with electrons will outnumber those with H atoms because the electrons' thermal velocities are higher by a factor of about 40. Then, there is a general argument that shows that the cross-section for excitation of optical transitions by electrons will be much larger than the cross-section for excitation by H atoms.

This argument about the cross-sections is discussed by Massey (1949). A collision occurs with a characteristic time scale $t \sim r/v$ where r is the effective range and v is the relative velocity. The transition in the perturbed molecule (or atom) corresponds to a frequency $\nu \sim \triangle E/h$. Classically and, also, quantum mechanically, the expectation is that the cross-sections for excitation and de-excitation will be small unless $1/t \sim \nu$ or $t\nu \sim 1$. It is instructive to examine this limit:

$$t\nu \approx \frac{r}{v}\frac{\Delta E}{h} = \frac{4r\Delta E}{\sqrt{T}} \quad \text{(electrons)}$$

$$= \frac{170r\,\Delta E}{\sqrt{T}} \quad \text{(hydrogen atoms)}$$

In these formulae, the effective range is given in Å units and the energy ΔE in eV. In the limit $tv \rightarrow o$, the cross-section will decrease from a maximum near $tv \lesssim 1$ but it will remain significantly large for small tv values. On the other hand the cross-section is expected to be very small for $tv >> 1$.

A typical optical transition, has $\Delta E \sim 3eV$ so that at $T \sim 5000$ K, $tv \sim 1/6$ for an electron collision and ~ 7 for a hydrogen atom collision. Then, the cross-section for an inelastic H atom collision is expected to be much smaller than for an inelastic electron collision. In an atom, fine-structure and some term-to-term transitions correspond to small ΔE. Also, in a molecule, the rotational and vibrational transitions corresponds to small excitation energies: $\Delta E_{rot} \sim 0$ to 0.05 eV and $\Delta E_{vib} \sim 0.2$ to 0.4 eV for typical molecules. Then, the product $tv < 1$ and excitation by hydrogen (also, helium) atoms must be included.

The total rate appearing in the equations of statistical equilibrium is the product of $<\sigma v>$ and the density of projectiles: can the small $<\sigma v>$'s expected for collisions with H atoms be offset by the large ratio $n(H)/n(e)$? Recent calculations beginning with Steenbock and Holweger (1984) suggest the answer to this question is often 'yes'. They introduced collisions with H atoms in their study for non-LTE line formation of the Li I lines in cool stars. Estimates of $<\sigma v>$ were drawn from a generalization of a 'modified classical Thomson formula' (Drawin 1968, 1969). This simple recipe is expected to yield an 'order-of-magnitude estimate of collisional excitation and ionization cross-sections'.

Here, I comment on later work on Fe I and Fe II lines. A thorough empirical LTE analysis of the iron lines in the spectrum of the K0 III giant β Gem was reported by Ruland *et al.* (1980) who noted that the high and low excitation Fe I lines gave different iron abundances. These differences are shown in the top panel of Figure 10 for lines with equivalent widths of 200 mÅ or less. The only reasonable interpretation of these differences is that they are due to non-LTE effects in the excitation of neutral iron atoms. This assertion is wonderfully supported by calculations done by Steenbock (1985 – see also Holweger 1988) whose Fe I/II/III model atoms comprise 79/20/1 levels with 52/23/0 transitions. Predicted corrections to the LTE abundances for a model atmosphere representative of β Gem are shown in the lower panel of Figure 10. Inspection shows that the NLTE calculations predict fairly well the sense of the difference between the high and low excitation lines over the entire range of equivalent widths. The illustrated predictions from Holweger (1988) were obtained with the cross-sections for the inelastic H collisions scaled by a factor $S_H = 0.2$ from the values expected on Drawin's recipe (Watanabe and Steenbock 1986). The predicted abundance spread is slightly smaller than observed. If the H collisions are neglected ($S_H = 0$), the spread in $\log\varepsilon_{LTE}/\varepsilon_{NLTE}$ is increased from 0.3 dex to 0.6 dex, a spread somewhat larger than indicated by the empirical results. As S_H is increased, the non-LTE effects are reduced. At $S_H = 1$, Holweger remarks that 'the calculated NLTE effects become too small«. Similar calculations and their fit to empirical

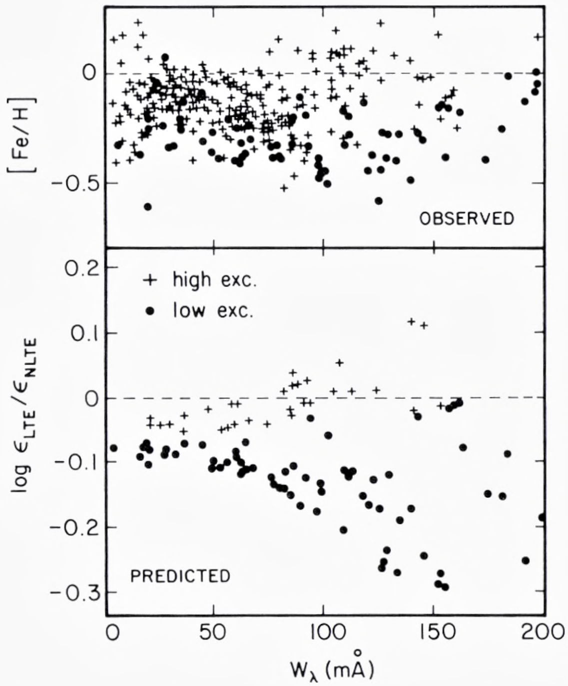

Fig. 10. Observed LTE iron abundances [Fe/H] (top panel) and the predicted corrections to LTE abundances (bottom panel) for Fe I lines in the spectrum of β Gem. The observed abundances are taken from Ruland *et al.* (1980). The predicted log $\varepsilon_{LTE}/\varepsilon_{NLTE}$ are taken from Holweger (1988).

results have been undertaken for the metal-poor K2 giant α Boo, the Sun, and main sequence stars Procyon (F5 IV-V) and Vega (A0V) – see Gigas (1986) for a discussion of Vega. (The predicted NLTE corrections for the Fe II lines are very small [Steenbock 1985]. The Fe I – Fe II abundance difference for giants cannot be readily established empirically because this and other atom-ion pairs are used to determine the surface gravity by imposing the condition that Fe I and Fe II lines yield the same Fe abundance).

This fascinating series of non-LTE calculations shows that (i) the observed departures from LTE in complex spectra such as Fe I may be understood semi-quantitatively, and (ii) firm quantitative predictions and, hence, abundances unbiased by non-LTE effects cannot be made until accurate information is available on cross-sections for excitation and ionisation by electrons *and* H atoms. While the theoretical and experimental literature on electron collisions is extensive, Drawin's crude recipe would appear to be sole extant tool for the astrophysicist needing estimates of the cross-sections for inelastic collisions with H atoms at the low thermal energies encountered in cool stellar atmospheres. Perhaps, a brave theoretician here at NORDITA will assume the challenge.

To close this discussion of non-LTE studies, I comment on one remarkable abundance anomaly that was uncovered through LTE analyses but is confirmed by a fairly exhaustive non-LTE analysis. I refer to the Na overabundance seen in F-K supergiants – see summary by Boyarchuk and Lyubimkov (1983) and also Sasselov (1986). After noting that the Na abundances were to be considered preliminary for lack of a non-LTE study of Na I line formation, Boyarchuk and Lyubimkov suggested that the Na overabundance may indicate the occurrence of deep convective mixing leading to the dredge-up of material exposed to the NeNa cycle (Marion and Fowler 1957). This suggestion was examined by Denisenkov and Ivanov (1987) who showed that the ^{22}Ne(p,γ)^{23}Na reaction can enrich the H-burning core of massive (M $\gtrsim$ 1.5 $M_\odot$) main sequence stars by factors of 5 to 6. Substantial amounts of this core material must be mixed with the envelope to account for the observed (LTE) overabundances. Such mixing is not predicted by standard models of the massive stars that evolve into the F-K supergiants. However, Denisenkov and Ivanov point to a correlation between the Na overabundance and the measured ^{12}C/^{13}C ratio as evidence for deep mixing.

The overabundance of Na has been confirmed by non-LTE calculations made by Boyarchuk *et al.* (1988a,b). The model atom consisted of 19 levels plus the ground state of the Na^+ ion. Nine transitions were linearized and many more were fixed during the iterations. Collisional exicitation and ionization by electrons but not H atoms was considered; the latter omission is most probably unimportant. The primary uncertainty in the calculations would appear to come from the representation of the ultraviolet radiation field that controls the photoionization rates. The pleasing (surprising?) result is that the Na abundances are changed only slightly by the introduction of non-LTE: $\triangle \log\varepsilon = \log\varepsilon_{\mathrm{NLTE}} - \log\varepsilon_{\mathrm{LTE}}$ ranges from $+0.06$ to -0.17 for four supergiants examined by Boyarchuk *et al.* (1988b).

The sceptic is left to ponder two questions: How does so much Na-rich material from the core get into the atmosphere? Are there further extensions to be made to the model atom and the atomic processes that would dramatically increase the predicted Na I equivalent widths for the non-LTE case and so lead to lower, even normal, abundances? The observer will hasten to the telescope to search for further spectroscopic evidence of extensive contamination by material from the H-burning core.

4. *Epilogue*

A majority of Bengt Strömgren's audience in 1958 would have entered the University Museum through the main door on their way to the lecture room in the far lefthand corner. A stuffed dodo bird from Mauritius would have greeted those members who took one of the possible routes to the lecture room. An hour later, attentive listeners would have recognized that Strömgren photometry would not rapidly follow the dodo into extinction. The informed listeners may even have predicted that Strömgren's scheme would become the premier photometric system.

A well-known historian in contemporary Oxford has been reported as despondent because, in his view, historians have run out of questions to answer. Professor Strömgren demonstrated in his 1958 lecture and by his energetic pursuit of problems to the end of his life that astronomy has not yet run out of questions and that the imaginative observer can forge the tools with which to extract the answers. In this essay, I have endeavoured to show how quantitative stellar spectroscopy may serve to answer a wide variety of the remaining outstanding questions. History may be an academic dodo but astronomy is far from extinction!

I thank colleagues in Austin and in Scandinavia for help and inspiration with the research that is described here. My research is supported in part by the U.S. National Science Foundation (currently through grant AST 86-14423) and the Robert A. Welch Foundation of Houston, Texas.

References

Alexander, J.B. 1967, *The Observatory,* **87,** 238.
Allen, C.W. 1973, *Astrophysical Quantities,* 3rd ed. (London: Athlone Press).
Anders, E. and Grevesse, N. 1989, *Geochim. Cosmochim. Acta,* **53,** 197.
Baade, D. and Magain, P. 1988, *Astr. Ap.,* **194,** 237.
Baird, S.R., Roberts, W.J., Snow, T.P., and Wallerstein, G. 1975, *Pub. A.S.P.,* **87,** 385.
Barbuy, B. 1988, *Astr. Ap.,* **191,** 121.
Barbuy, B. and Erdelyi-Mendes, M. 1989, *Astr. Ap.,* **214,** 239.
Bates, D.R. and Damgaard, A. 1949, *Phil. Trans. Roy. Soc. A.,* **242,** 101.
Becker S.R. and Butler, K. 1988a, *Astr. Ap.,* **201,** 232.
Becker S.R. and Butler, K. 1988b, *Astr. Ap. Suppl.,* **74,** 211.
Becker S.R. and Butler, K. 1988c, *Astr. Ap. Suppl.,* **76,** 331.
Becker S.R. and Butler, K. 1989, *Astr. Ap.,* **209,** 244.
Bessell, M.S., Brett, J.M., Scholz, M., and Wood, P.R. 1989, *Astr. Ap.,* **213,** 209.
Bessell, M.S. and Norris, J. 1987, *J. Astr. Ap.,* **8,** 99.
Bidelman, W.P. and Keenan, P.C. 1951, *Ap. J.,* **114,** 473.
Boesgaard, A.M. and Tripicco 1986, *Ap. J.,* **303,** 724.
Bond, H.E. 1974, *Ap. J.,* **194,** 95.
Boyarchuk, A.A. and Lyubimkov, L.S. 1983, *Izv. Krym. Astrofiz. Obs.,* **66,** 130.
Boyarchuk, A.A., Gubeny, I., Kubat, I., and Lyubimkov, L.S. 1988a *Astrofizika,* **28,** 335.
Boyarchuk, A.A., Gubeny, I., Kubat, I., and Lyubimkov, L.S. 1988b *Astrofizika,* **28,** 343.
Brett, J.M. 1989, preprint.
Brown, J.A., Sneden, C., Lambert, D.L., and Dutchover, E. Jr. 1989, *Ap. J. Suppl.,* **71**, 293.

Butcher, H.R. 1987, *Nature,* **328,** 127.

Butcher, H.R. 1988, *The Messenger,* No. 51, p. 12.

Cameron, A.G.W. and Fowler, W.A. 1971, *Ap. J.,* **164,** 111.

Clayton, D.D. 1987, *Nature,* **329,** 397.

Clayton, D.D. 1988, *M.N.R.A.S.,* **234,** 1.

Clegg, R.E.S., Lambert, D.L., and Tomkin, J. 1981, *Ap. J.,* **250,** 262.

Cohen, J.G., Frogel, J.A., Persson, S.E., and Elias, J.H. 1981, *Ap. J.,* **249,** 481.

Cugier, H. and Hardorp, J. 1988, *Astr. Ap.,* **197,** 163.

Denisenkov, P.A. and Savanov, V.V. 1987, *Soviet Astr. Letters.,* **13,** 214.

Dominy, J.F. 1984 *Ap. J. Suppl.,* **55,** 27.

Dominy, J.F., Wallerstein, G. and Suntzeff, N.B. 1986, *Ap. J.,* **300,** 325.

Drawin, H.W. 1968, *Z. Phys.,* **211,** 404.

Drawin, H.W. 1969, *Z. Phys.,* **225,** 483.

Dufton, P.L. Berrington, K.A., Burke, P.G., and Kingston, A.E. 1978, *Astr. Ap.,* **62,** 111.

Duncan, D.K. 1981, *Ap. J.,* **248,** 651.

Eber, F. and Butler, K. 1988, *Astr. Ap.,* **202,** 153.

Fujita, Y. and Tsuji, T. 1977, *Pub. A.S. Japan,* **29,** 711.

Gigas, D. 1986, *Astr. Ap.,* **165,** 170.

Gilroy, K.K. Sneden, C., Pilachowski, C.A., and Cowan, J.J. 1988, *Ap. J.,* **327,** 298.

Goorvitch, D., Goebel, J.H., and Augason, G.C. 1980, *Ap. J.,* **240,** 588.

Gratton, R.G. 1985, *Astr. Ap.,* **148,** 105.

Gratton, R. G. and D'Antona, F. 1989, *Astr. Ap.,* **215,** 66.

Hardorp, J., Cugier, H., Koratkar, A., and Scott, J. 1986, in *New Insights in Astrophysics – 8 Years of Astronomy with IUE* (ESA: Paris).

Hinkle, K.H. and Lambert, D.L. 1975, *M.N.R.A.S.,* **170,** 447.

Holweger, H. 1988, in *The Impact of Very High S/N Spectroscopy on Stellar Physics,* ed. G. Cayrel de Strobel and M. Spite (Dordrecht: Kluwer).

Iben, I. Jr. 1967, *Ap. J.,* **147,** 624 and 651.

Iben, I. Jr. 1973, *Ap. J.,* **185,** 209.

Iben, I. Jr. 1975, *Ap. J.,* **196,** 525.

Iben, I. Jr. 1981, *Ap. J.,* **246,** 278.

Iben, I. Jr. 1989, in *Evolution of Peculiar Red Giants,* ed. H.R. Johnson and B. Zuckerman (Cambridge: CUP) in press.

Iben, I. Jr., and Truran, J.W. 1978, *Ap. J.,* **220,** 980.

Jørgensen, U.G., Almlöf, J., Gustafsson, B., Larsson, M., and Siegbahn, P. 1985, *J. Chem. Phys.,* **83,** 3034.

Kjærgaard, P., Gustafsson, B., Walker, G.A.H., and Hultqvist, L. 1982, *Astr. Ap.,* **115,** 145.

Lambert, D.L. 1980, in »*Les Spectres des Molecules Simple an Laboratoire et en Astrophysique*«, (Liège: Inst. d'Ap.) p. 173.

Lambert, D.L. 1985, in *Cool Stars with Excesses of Heavy Elements,* ed. M. Jaschek and P.C. Keenan (Dordrecht: Reidel), p. 191.

Lambert, D.L. 1988, in *The Impact of Very High S/N Spectroscopy on Stellar Physics,* ed. G. Cayrel de Strobel and M. Spite, (Dordrecht: Kluwer), p. 563.

Lambert, D.L. 1989, in *Cosmic Abundances of Matter,* ed. C.J. Waddington, (New York: AIP), p. 168.

Lambert, D.L. Gustafsson, B., Eriksson, K., and Hinkle, K.H. 1986, *Ap. J. Suppl.,* **62,** 373.

Lambert, D.L. McKinley, L.K., and Roby, S.W. 1986, *P.A.S.P.,* **98,** 927.

Lambert, D.L. and Ries, L.M. 1981, *Ap. J.,* **248,** 228.

Lambert, D.L. and Sawyer, S. 1984, *Ap. J.,* **283,** 192.

Lennon, D.J. 1983, *M.N.R.A.S.,* **205,** 829.

Livio, M. and Soker, N. 1983, *Astr. Ap.,* **125,** L12.

Livio, M. and Soker, N. 1984, *M.N.R.A.S.*, **208,** 763.
Malaney, R.A. and Alcock, C.R. 1989, preprint.
Malaney, R.A. and Fowler, W.A. 1989, *M.N.R.A.S.*, **237,** 67.
Marion, J.B. and Fowler, W.A. 1957, *Ap. J.*, **125,** 221.
Massey, H.S.W. 1949, *Rept. Prog. Phys.*, **12,** 248.
Mathews, G. and Schramm, D.N. 1988, *Ap. J. Letters,* **324,** L67.
Mathys, G., Maitzen, H.M., North, P., Hensberge, H., Weiss, W.W., Ansari, S. Catalano, F.A., Didelo, P., Faraggiana, R., Fuhrmann, K., Gerbaldi, M., Renson, P., and Schneider, H. 1989, *The Messenger,* No. 55, p. 41.
McClure, R.D. in *Cool Stars with Excesses of Heavy Elements,* ed. M. Jaschek and P.C. Keenan (Dordrecht: Reidel), p. 315.
McClure, R.D., Fletcher, J.M., and Nemec, J.M. 1980, *Ap. J. (Letters),* **238,** L35.
McWilliam, A. 1988, Ph. D. Thesis, University of Texas at Austin.
Muchmore, D. 1986, *Astr. Ap.*, **155,** 172.
Nordlund, Å. 1982, *Astr. Ap.*, **107,** 1.
Paczyński, B. 1973, *Acta Astr.*, **23,** 191.
Preston, G. 1974, *Ann. Rev. Astr. Ap.*, **12,** 257.
Preston, G. and Paczyński, B. 1974, *Carnegie Institution of Washington Yearbook,* **73,** 133.
Rebolo, R., Molaro, P., and Beckman, J.E. 1988, *Astr. Ap.*, **192,** 192.
Renzini, A., and Voli, M. 1981, *Astr. Ap.*, **94,** 175.
Richer, H.B. 1981, *Ap. J.*, **243,** 744.
Roby, S.W. 1987, Ph. D. thesis, University of Texas at Austin.
Roby, S.W. and Lambert, D.L. 1990, *Ap. J. Suppl.*, in press.
Ruland, F., Holweger, H., Griffin, R., and Biehl, D. 1980, *Astr. Ap.*, **92,** 70.
Russell, S.C. and Bessell, M.S. 1989, *Ap. J. Suppl.*, **70**, 865.
Sackmann, I.J., Smith, R.L., and Despain, K.H. 1974, *Ap. J.*, **187,** 555.
Sahu, K.C., Sahu, M., and Pottasch, S.R. 1989, *Astr. Ap.*, in press.
Sasselov, D.D. 1986, Pub. A.S.P., **98,** 561.
Scalo, J.M. 1976, *Ap. J.*, **206,** 795.
Scalo, J.M., Despain, K.H., and Ulrich, R.K. 1975, *Ap. J.*, **196,** 809.
Seaton, M.J. 1962, in *Atomic and Molecular Processes,* (New York: Academic Press).
Smith, A.M., Jørgensen, U.G., and Lehmann, K. 1987, *J. Chem. Phys.*, **87,** 5649.
Smith V.V. and Lambert, D.L. 1985, *Ap. J.*, **294,** 326.
Smith V.V. and Lambert, D.L. 1986, *Ap. J.*, **311,** 843.
Smith, V.V. and Lambert, D.L. 1989, *Ap. J. Letters,* **345**, L75.
Smith V.V. and Lambert, D.L. 1990, *Ap. J. Suppl.*, in press.
Sneden, C., Lambert, D.L., and Pilachowski, C.A. 1981, *Ap. J.*, **247,** 1052.
Sneden, C., Lambert, D.L., and Whitaker, R.W. 1979, *Ap. J.*, **234,** 964.
Spite, M., Barbuy, B., and Spite, F. 1989, *Astr. Ap.*, **222**, 35.
Spite, F. and Spite, M. 1982, *Astr. Ap.*, **115,** 357.
Steenbock, W. 1985 in *Cool Stars with Excesses of Heavy Elements,* ed. M. Jaschek and P.C. Keenan (Drodrecht: Reidel) p. 231.
Steenbock, W. and Holweger, H. 1984, *Astr. Ap.*, **130,** 319.
Strömgren, B. 1958, *The Observatory,* **78,** 137.
Tassoul, J.L. and Tassoul, M. 1984, *Ap. J.*, **279,** 384.
Tomkin, J., Lambert, D.L., Edvardsson, B., Gustafsson, B., and Nissen, P.E. 1989, *Astr. Ap.*, **219,** L15.
Van Regermorter, H. 1962, *Ap. J.*, **136,** 906.
VandenBerg, D.A. 1988, in *The Extragalactic Distance Scale,* ed. S. van den Bergh and C.J. Pritchet (San Francisco: Astr. Soc. Pac.) p. 187.

Wallerstein, G. and Sneden, C. 1982, *Ap. J.,* **255,** 577.

Watanabe, T. and Steenbock, W. 1986, *Astr. Ap.,* **165,** 163.

Wheeler, J.C., Sneden, and Truran, J.W. 1989, *Ann. Rev. Astr. Ap.,* **27**, 279.

Winget, D.E., Hansen, C.J., Liebert, J., Van Horn, H.M., Fontaine, G., Nather, R.E. Kepler, S.O. and Lamb, D.Q. 1987, *Ap. J. Letters,* **315,** L77.

Wood, P.R. 1987, in *Late Stages of Stellar Evolution,* ed. S. Kwok and S.R. Pottasch (Dordrecht: Reidel) p. 197.

Wood, P.R., Bessell, M.S., and Fox, M.W. 1983, *Ap. J.,* **272,** 99.

Yu Yan, Taylor, K.T., and Seaton, M.J. 1987, *J. Phys. b.,* **20,** 6399.

The Story of AGB Star Evolution – An Intimate Connection between Theory and Observation

By ICKO IBEN JR.
University of Illinois
349 Astronomy Building, 1011 West Springfield Av., Urbana, IL61801, USA

Abstract

A summary is given of results of theoretical and observational studies of asymptotic giant branch (AGB) evolution. High and intermediate mass AGB model stars activate the ^{22}Ne neutron source during thermal pulses, produce s-process isotopes in a non-solar distribution, and dredge these isotopes along with fresh ^{12}C up to the surface. Observations suggest that real counterparts do not live long enough to become carbon stars, but the actual distribution of s-process isotopes is not yet known. Low mass AGB models of low metallicity activate the ^{13}C neutron source during thermal pulses, produce s-process isotopes in the solar-system distribution, and dredge freshly produced isotopes and ^{12}C to the surface when convective overshoot is assumed. Low mass AGB models of solar metallicity have not yet been persuaded to activate the ^{13}C neutron source, although they do dredge up fresh ^{12}C. Observations show that, independent of metallicity, real low mass AGB stars dredge up both carbon and s-process isotopes, the latter in the solar-system distribution.

I. *Preamble*

Over the past year, in the wake of SN 1987a's first appearance, we have been treated to a marvelous example of how theory and observation interact in astrophysics, with both theory and observation playing absolutely essential roles in guiding us to an understanding of an extraterrestial phenomenon. For the first time, we have direct evidence that stars of initial mass in the range 20 ± 5 $M_\odot$ actually develop a neutron star remnant with theoretically anticipated properties, including the release of gravitational potential energy of the expected order of magnitude and the expulsion of envelope matter containing freshly produced iron-peak elements. For years, there have been conflicting theoretical (numerical, model based) inferences as to how much, if any, material from the imploding iron-nickel core of a massive star would be expelled. Now, thanks to SN 1987a, we have a quantitative understanding of how much of this core is expelled, something that really cannot be estimated unambiguously from first principles. And yet, without the prior theoretical exploration and numerical modeling, we would not have been able to interpret aspects of the observed light curve in terms of the release of nuclear energy by radioactive nickel and cobalt.

A no less important, but certainly less immediately spectacular example is our growth in understanding of asymptotic giant branch (AGB) stars. Over the past two decades, a combination of theoretical and observational discovery has given us insight into the last stages of the evolution of low and intermediate mass stars prior to their becoming white dwarfs. Both theory and observation have made essential contributions. Without observation, theory alone would have led us astray; and without a theoretical framework, the observations would be of little use in adding to our understanding.

We are now persuaded that the AGB phase is the last nuclear-burning phase which all stars of mass less than about 8 $M_\odot$ experience. Hydrogen and helium burn alternately in thin shells above an inert electron-degenerate carbon-oxygen (CO) core. That this phase is indeed the last burning stage for all stars of low and intermediate mass was not an initial prediction of the theory, although a few tentative theoretical speculations were advanced that this might be the case. It is really observational evidence that has taught us most convincingly that the AGB phase is terminal. For example, if all stars of initial mass in the range (1.4 - 8) $M_\odot$ were to remain AGB stars long enough for their CO core to grow to the Chandrasekhar mass of 1.4 $M_\odot$, the supernova rate in galaxies similar to our own would be over 20 times the observed rate (Iben 1981), and one could infer from this that most stars initially less massive than 8 $M_\odot$ must somehow lose essentially all of their hydrogen-rich envelope before their CO core grows to 1.4 $M_\odot$. The properties of planetary nebulae, including their occurrence frequency, offer further observational evidence that mass loss terminates the AGB phase. This same story is told even more directly and emphatically by the paucity of luminous AGB stars in the Magellanic Clouds and by the observed rates of mass loss from AGB stars in our own Galaxy.

On the other hand, theory has been able to show how AGB stars make carbon and s-process isotopes (such as radioactive ^{99}Tc) in their interiors and bring these freshly produced elements to the surface. Theory is also now beginning to converge on how mass is ejected from AGB stars. Ingredients include the formation of grains, which are pushed outward by radiation pressure to inflate the stellar envelope, and shock heating of an expanding atmosphere induced by acoustical pulsations which are driven by thermodynamic conditions below the photosphere.

In an earlier review this year (Iben 1988), I emphasized the role of observations in guiding our understanding of AGB star evolution. In this essay, I will summarize what we have learned about the activation of the neutron source in AGB stars and about the dredge up of freshly processed carbon and neutron rich isotopes to the surface, emphasizing theoretical insights.

II. *AGB Stars of Intermediate Mass*

A) *Basic Structure and Thermal Pulses*

By intermediate mass I mean stars which are sufficiently massive that they do not

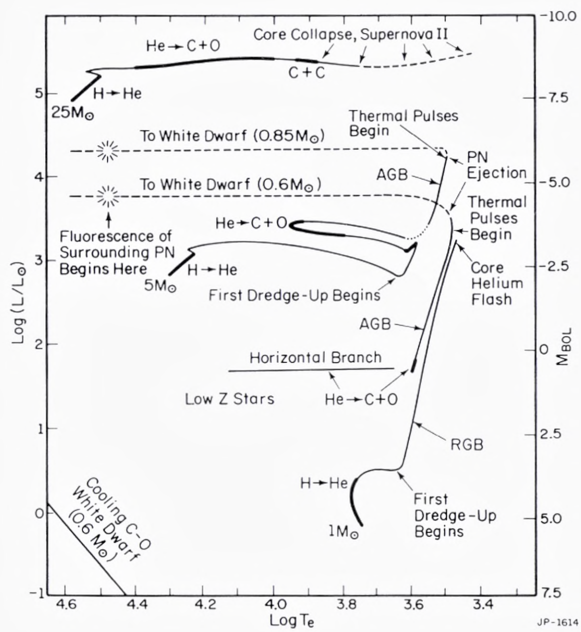

Figure 1. The evolutionary paths in the HR diagram of model stars of population I composition and of initial mass 1 M$_\odot$, 5 M$_\odot$, and 25M$_\odot$. The 25 M$_\odot$ model burns hydrogen in its core as a hot main-sequence star, helium in its core as a blue star, carbon in its core as a blue star, and experiences a core collapse and type II supernova explosion shortly after exhausting central carbon. The 5 M$_\odot$ model becomes an AGB star after exhausting helium in its core (the solid curve at the highest luminosities along the 5 M$_\odot$ evolutionary track). Observations suggest that, shortly after it enters the thermally pulsing phase, the real analogue loses most of its hydrogen-rich envelope and evolves rapidly to the blue, eventually to become a white dwarf. Along the way, it excites the nebular material about it into flourescence. A low mass star becomes a horizontal branch or red giant "clump" star while it burns helium in its core and hydrogen in a shell. After it exhausts central helium, it becomes an AGB star before losing most of its hydrogen-rich envelope and evolving into a white dwarf configuration.

form an electron-degenerate core until after they have exhausted helium at their centers, but light enough that they develop such a core before igniting carbon. In practice, this means stars with an initial main-sequence mass in the range about 2-8 M$_\odot$. Both the lower and upper limits to this mass range depend on the choice of composition, and both are highly uncertain due to the uncertainty in the treatment of convective overshoot during the main-sequence and core helium-burning phases.

After the exhaustion of central helium, helium burning takes place in a shell. The material at the helium-hydrogen interface and beyond is pushed outward to such low temperature and densities that hydrogen burning effectively ceases until the helium-burning shell almost reaches this interface. Then, hydrogen is reignited and helium burning dies down temporarily. Thereafter, hydrogen and helium burning alternate in supplying surface luminosity. When this alternation begins, the mass of the CO core is about 0.3 $M_\odot$ for stars of initial mass near the lower limit of ~ 2 $M_\odot$ and increases to about 1.1 $M_\odot$ for stars of initial mass near the upper limit of ~ 8 $M_\odot$. The evolutionary track of a 5 $M_\odot$ model star in the Hertzsprung-Russell diagram all the way to the beginning of the alternate-burning (or thermally pulsing) phase is shown in Figure 1.

Each time the hydrogen-burning shell has laid down a thick enough layer of helium, temperatures and densities in this layer become large enough to ignite helium

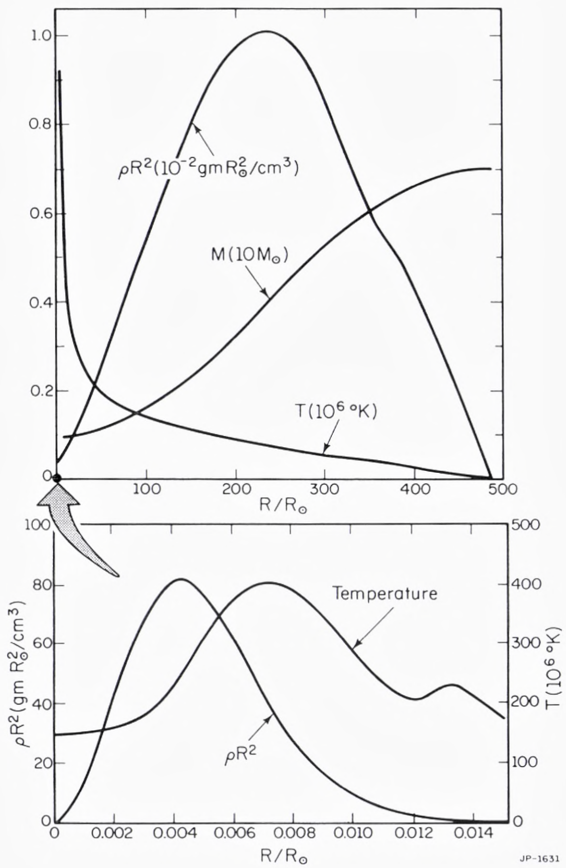

Figure 2. The density, temperature, and mass distributions within an AGB model of core mass $M_{CO} \sim 0.95$ $M_\odot$ and total mass $M_{\bullet} = 7$ $M_\odot$. The core characteristics are shown in the lower panel and the envelope characteristics are shown in the upper panel.

explosively. Matter at and above the helium-hydrogen interface is again pushed out to such an extent that hydrogen burning ceases. In a matter of few years to decades, depending on the mass of the CO core, the thermonuclear runaway is quenched and the star embarks on a phase of quiescent helium burning which continues until the amount of mass which has been converted into carbon and oxygen equals the amount of mass which has passed through the hydrogen-burning shell during the preceding interpulse phase. Then hydrogen burning takes over and continues until another "helium shell flash" or "thermal pulse" is excited.

The structure of a thermally pulsing AGB model of mass $M_* = 7\ M_\odot$ is illustrated in the two panels of Figure 2. At the center of the model star is a very hot white dwarf of mass $M_{co} \sim 0.95\ M_\odot$ and radius $R_{co} \sim 0.01\ R_\odot$ (lower panel). The maximum in the temperature occurs where the rate of cooling by neutrino losses is just balanced by the rate of heating due to compression. Most of the matter in the model resides in a very low density, low temperature, giant envelope (upper panel). Between the giant envelope and the central white dwarf is the "nuclear-active" region. A thermal pulse has just begun and this is reflected in the "bump" in temperature that occurs at a radius $\sim 0.013\ R_\odot$. The progress in time of this bump is illustrated in Figure 3. The outward movement and cooling of the boundary between hydrogen-rich matter and hydrogen-exhausted matter (the "XY discontinuity") is evident in this figure.

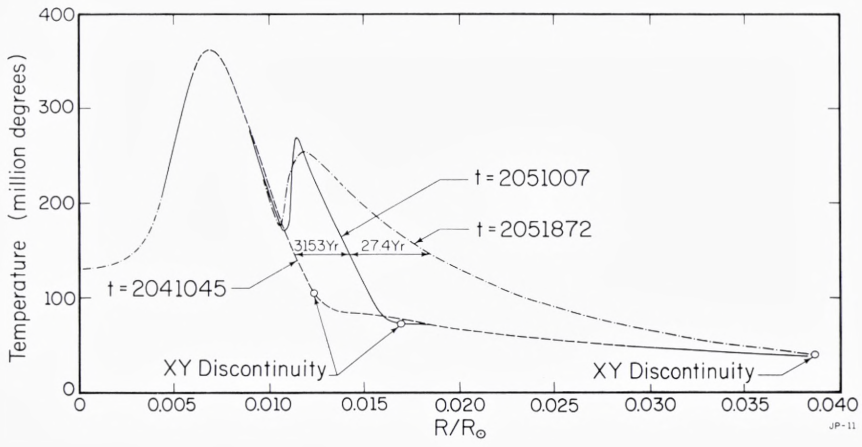

Figure 3. Temperature profiles before and during a thermal pulse in a model star of core mass $0.95\ M_\odot$ and total mass $7M_\odot$. The three times are in units of 10^{12}s.

B) *Nucleosynthesis and Dredge Up*

The dominant products of helium burning are, of course, carbon and oxygen. However, by far the most interesting aspect of nucleosynthesis during thermal pulses is the production of neutron-rich isotopes. In AGB models of large core mass (say $M_{co} \gtrsim 0.95\ M_{\odot}$), the major source of neutrons is the $^{22}Ne(\alpha,n)^{25}Mg$ reaction (Iben 1974a, 1976, 1977). During hydrogen burning, ^{12}C and ^{16}O are converted into ^{14}N and, during the early portion of a helium shell flash, ^{14}N is converted into ^{22}Ne within the convective shell which is formed due to the high fluxes generated by the $3\alpha \rightarrow {}^{12}C$ reactions. When the temperatures near the base of the convective shell approach and exceed 300×10^6 K, neutrons are released by the endoergic $^{22}Ne(\alpha,n)^{25}Mg$ reaction. Most of the neutrons are captured by light element filters (such as ^{22}Ne and ^{25}Mg) but enough are captured by the "seed" nucleus ^{56}Fe and by its neutron-capture progeny to build up a substantial overabundance of the so-called s-process isotopes. The basic reactions and where they occur during a thermal pulse are described in Figure 4, where the outer "speckled" region depicts the base of the convective envelope and the inner speckled region depicts the convective shell which lies between the CO core and the radiative hydrogen-rich zone.

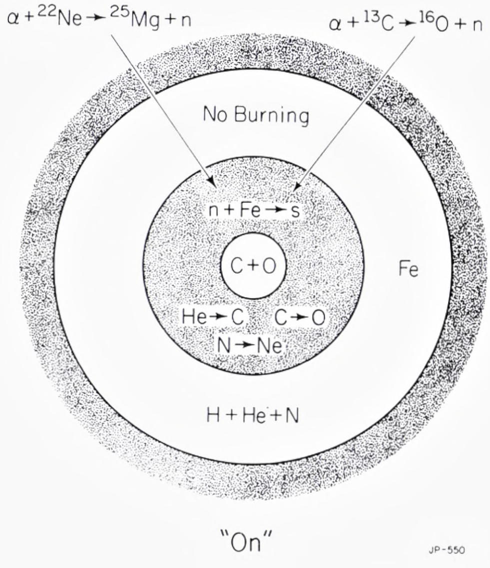

Figure 4. Nuclear burning activity and convective zones during a helium shell flash. The speckled regions are convective zones. Only the ^{22}Ne neutron source operates in intermediate mass AGB models. In low mass AGB models, the ^{13}C neutron source is active and the ^{22}Ne neutron source operates as well, but only weakly.

In general, the s-process isotopes are not made in the solar-system distribution because the neutron densities which are formed as a balance between the rate of neutron production and the rate of neutron consumption are too large by many orders of magnitude (Despain 1980, Cosner, Iben, and Truran 1980). Nevertheless, there are strong similarities between the distributions formed and the solar-system distribution. These similarities are a consequence of (a) the unique characteristic of the ^{22}Ne neutron source that the number of light element filters made during a helium shell flash is comparable with the number of neutrons released and (b) the fact that some fraction of the material that appears in the convective shell during any given flash has also appeared in an earlier flash (Iben 1975b, Truran and Iben 1977). This second feature, which is illustrated in Figure 5, leads to an exponential distribution of exposures for the matter in any given convective shell (Ulrich 1973) and it is well known that such an exponential distribution is essential for producing s-process isotopes in the solar-system distribution (Clayton et al 1961, Seeger, Fowler, and Clayton 1965).

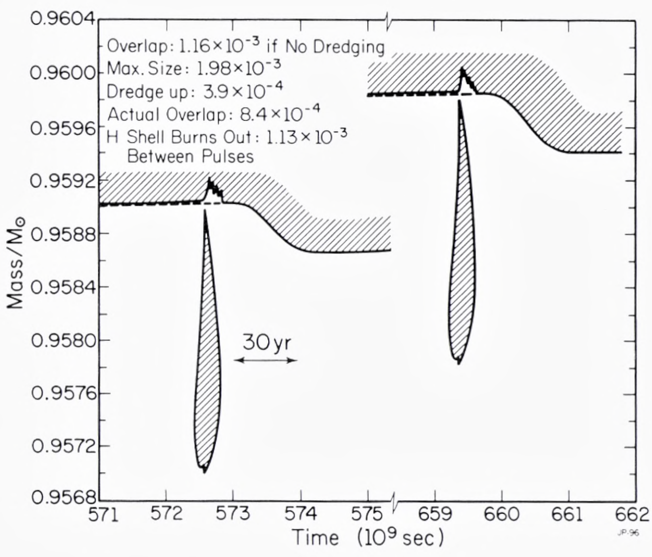

Figure 5. Convective zones (hatched regions) as a function of time in a model of core mass $M_{CO} = 0.95\ M_{\odot}$, total mass $M_* = M_{\odot}$. Both the properties of overlap (some of the matter appearing in a convective shell during a given pulse has also appeared in the convective shell formed during the previous pulse) and of dredge up (following a pulse, the base of the convective envelope extends into the region containing freshly made ^{12}C and s-process isotopes and these nuclei are carried to the surface by convection) are evident.

Figure 5 also demonstrates the property of "dredge up" (Iben 1975a, 1976) which occurs in intermediate mass AGB models after the helium shell flash subsides. The base of the convective envelope moves inward (in mass) and into the outer portions of the region once occupied by the convective shell during the height of a flash. Fresh ^{12}C and s-process isotopes are then dredged to the surface.

Enough studies have now been conducted of the nucleosynthesis expected in the environment provided by intermediate mass AGB stars which activate the ^{22}Ne neutron source that it appears to be inescapable that such stars are not responsible for the production of the bulk of the s-process isotopes in the solar-system (e.g., Mathews and Ward 1985, Howard et al 1986, Malaney and Boothroyd 1987, Busso et al 1988). From this one might infer that: (a) real intermediate mass stars do not reach the thermally pulsing phase; (b) such stars do reach this phase but do not dredge up to their surfaces material which has been processed through convective shells powered by helium burning during a thermal pulse; or (c) such stars reach the thermally pulsing AGB phase, dredge up material "nuclearly" processed in convective shells, but do not live long enough as thermally pulsing AGB stars to contribute substantially to the galactic abundances of neutron-rich isotopes.

The model studies indicate that option (b) is not likely. That is, if the CO core mass is large enough to permit activation of the ^{22}Ne neutron source, then dredge up of processed material will occur in a natural fashion, without the necessity of invoking

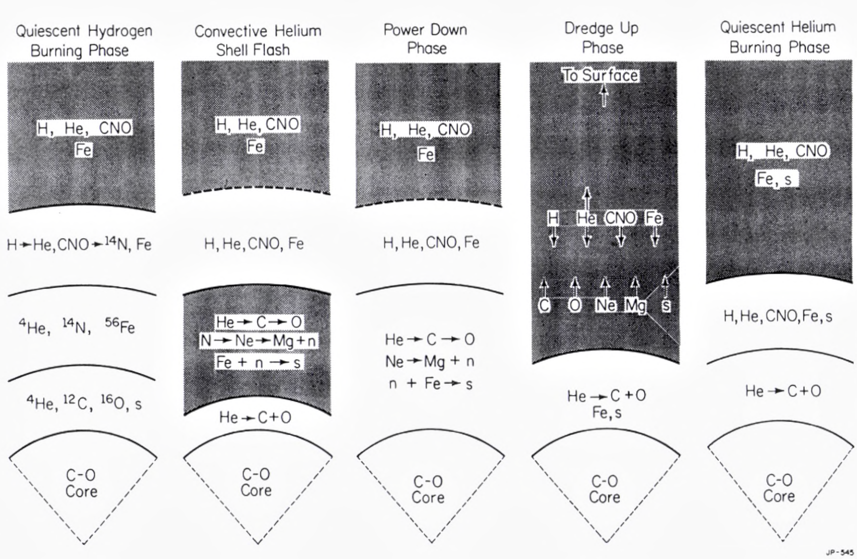

Figure 6. Schematic showing convective zones and nucleosynthesis activity during five stages of a thermal pulse cycle. Zone sizes are not to scale.

a physical process (such as overshoot) for which there exists as yet no easily quantifiable theory. As discussed in the next subsection, observations rule out option (a) and support at least the first portion of option (c).

The nucleosynthesis activity and the convective mixing activity which take place over a complete thermal pulse-interpulse cycle of an intermediate-mass AGB model with a large CO core is summarized in Figure 6.

C) *Lessons from the Observations*

In the preamble it is noted that, if all intermediate mass stars were to evolve along the AGB until the mass of their CO core reached the Chandrasekhar limit, the supernova rate in a galaxy such as ours would far exceed the observed rate. The logical inference is that real AGB analogues must lose most of their hydrogen-rich envelopes before this occurs and the existence and properties of planetary nebulae and of their central stars provides some direct confirmation of this inference. The kinematical and mass loss characteristics of OH/IR sources suggest that these sources are the consequence of mass loss from an underlying AGB star and are in fact in the process of becoming planetary nebulae (de Jong 1983, Habing 1986, Kwok 1987).

Limits on how long a real AGB star of large core mass can spend in the thermally pulsing phase is provided by observations of bright stars in the Magellanic clouds, coupled with theoretical estimates of how long an AGB star must spend in the AGB phase to (a) achieve carbon star characteristics and (b) contribute significantly to the galactic nucleosynthesis of neutron-rich isotopes.

That thermally pulsing AGB stars of large core mass exist and that the theoretical predictions of dredge up and neutron-capture nucleosynthesis are basically correct is (in my mind) demonstrated unequivocally by the long period variables in the Magellanic Clouds. The strengths of ZrO lines in the LPV's with bolometric magnitudes brighter than M_{bol} = - 6 mag (corresponding to CO core masses larger than $\sim$ 0.85 $M_\odot$) imply overabundances of Zirconium (an s-process element) and therefore suggest both the present activity of a neutron source and the reality of dredge up (Wood, Bessel, and Fox 1983). The paucity of bright LPV's ($\sim$ 100-300) relative to the number of Cepheids ($\sim$ 2000-4000), which are presumably the core helium-burning progenitors of LPV's, suggests that the lifetime in the thermally pulsing AGB phase is of the order of 10 % of the Cepheid lifetime. This latter lifetime is estimated in a semi-empirical fashion to be of the order of 10^6 yr (Becker, Iben, and Tuggle 1977). Since the CO core mass of an AGB model grows by $\sim$ 0.1 $M_\odot$ per 10^6 yr and the mean brightness of the model increases by $\sim$ 1 mag in this time, the inference is that shortly after a real star of intermediate mass reaches the thermally pulsing AGB phase with a CO core mass $\gtrsim$ 0.85 $M_\odot$ it "evaporates". That is, it loses its hydrogen-rich envelope in 10^5 yr or less after having increased its initial CO core mass by only a few percent, but not before having produced and dredged to the surface the results of some fresh neutron-capture nucleosynthesis.

The inferred short lifetime of thermally pulsing AGB stars of large core mass explains why there are essentially no carbon stars brighter than $M_{bol} \sim -6$ mag – it requires $\sim 10^6$ yr for C-star characteristics to be achieved (Iben and Truran 1978, Renzini and Voli 1981, Iben and Renzini 1983) – and reconciles (a) the fact that model AGB stars of large core mass produce s-process isotopes in an apparently non-solar-system distribution with (b) the fact that, in real objects, s-process isotopes tend to be in approximately the solar-system distribution. Iben and Truran (1978) show that, if AGB stars of large core mass were to live long enough to increase their initial core mass by $\sim 0.1\ M_\odot$, they should be able to account for ~ 2 times the estimated galactic abundance of s-process isotopes. Observation and theory suggest that AGB stars evaporate before having increased their core mass by over ~ 0.01-$0.02\ M_\odot$ and, thus, all is well: the stars which activate the ^{22}Ne source contribute 10-20 % at most to the galactic nucleosynthesis of s-process isotopes.

What is urgently required is that high dispersion spectroscopy and s-process isotope abundance analysis be undertaken for the LPV's in the Magellanic Clouds to see whether or not the abundance distributions are consistent with the ^{22}Ne neutron source, which nuclear reaction data and stellar model theory together suggest is operating in these stars.

III. *AGB Stars of Low Mass*

A) *Development of a Common CO Core and Thermal Pulse Characteristics*

By definition, a star of low mass is one which develops an electron-degenerate helium core after exhausting central hydrogen. As such a star evolves upward along the giant branch (see Figure 1), its helium core grows until its mass reaches ~ 0.45-$0.5\ M_\odot$, at which point helium is ignited. After a series of shell flashes (e.g., Mengel and Sweigart 1981), the degeneracy of the core is lifted and the star continues to burn helium in the core, but now quiescently, and to burn hydrogen in a shell. If it is of population I composition, a star spends this phase confined to a small "clump" region along the giant branch in the H-R diagram; if it is of population II composition, it resides on the "horizontal branch" (see Figure 1). The clump or horizontal branch phase lasts for approximately 10^8 yr. During this time, the hydrogen-burning shell processes approximately $0.05\ M_\odot$ of matter, so that the mass of the hydrogen-exhausted core of a low mass star becomes ~ 0.5-$0.55\ M_\odot$, nearly independent of the total mass of the star.

After exhausting central helium, a low mass star evolves over a period of $\sim 10^7$ yr along the "early" asymptotic giant branch, processing helium into carbon and into oxygen in a shell above a growing electron-degenerate CO core. Hydrogen does not burn. When the CO core mass reaches $\sim$ 0.5 $M_\odot$, hydrogen-burning is reactivated and the star enters the thermally pulsing AGB phase.

The properties of the thermal pulse cycle of a low core mass AGB star are in most respects qualitatively the same as those of a high core mass AGB star. That is, there is the same long period of quiescent hydrogen burning interrupted periodically by a helium-burning thermonuclear runaway which relaxes into a quiescent helium-hurning phase lasting about 10 percent of the duration of the quiescent hydrogen-burning phase. The duration of each phase, however, is much longer for stars of small core mass than for those with large core mass. The time between thermal pulses varies inversely as the tenth power of the core mass, being about 2000 yr when M_{CO} $\sim$ 0.95 $M_\odot$ and about 200,000 yr when M_{CO} $\sim$ 0.6 $M_\odot$. The light curve of a low mass model is shown in Figure 7.

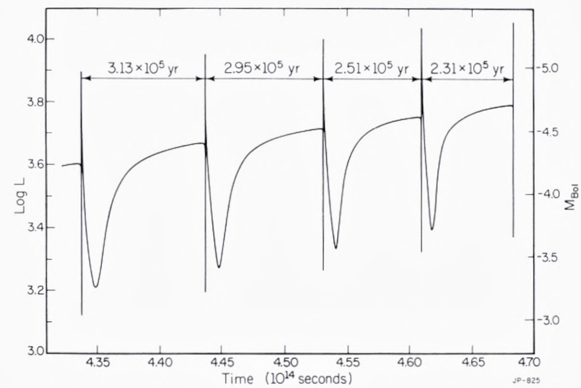

Figure 7. The light curve of an AGB model of low metallicity (Z = 0.001), low mass ($M_* = 0.7\ M_\odot$), and small core mass ($M_{CO} \sim 0.57 - 0.61\ M_\odot$).

A more important difference is the fact that the dredge up of freshly produced carbon does not occur in an unforced way. It is necessary to assume that some form of convective overshoot at the base of the convective envelope occurs. In the work of Iben and Renzini (1982a,b), for example, dredge up is achieved by forcibly mixing material into regions which are, initially, formally stable against convection and which lie successively further below the formal base of the convective envelope. If the matter in the freshly mixed-in region becomes formally unstable against convection, this procedure is continued until, on tentatively adding one final zone to the fully mixed region, it transpires that the matter in this zone is still formally stable against convection. Thus, a self consistent inward motion (in mass) of the base of the convective envelope is achieved. Hollowell (1988) adopts a diffusive mixing algorithm with various choices for the mean distance which a convective element can overshoot the formal base of the convective envelope. Wood and Zarro (1981) and Boothroyd and Sackmann (1988a,b) show that by increasing sufficiently the mixing length to scale height ratio in a mixing length treatment of convection, dredge up can also be achieved, given large enough envelope mass.

Another very important difference between AGB models of small and large core mass is that, in models of small core mass, the ^{22}Ne neutron source is only mildly activated, with at most only about 1 % of the neon which enters the convective shell during a pulse reacting with α particles (Becker 1981, Iben 1982, Hollowell 1988). Since, as argued in section II, s-process isotopes are produced in a non-solar distribution when ^{22}Ne is the dominant source of neutrons, this weakness may be a virtue. On the other hand, the activation of the ^{13}C neutron source has been demonstrated to occur only in low core mass models of low metallicity (Iben and Renzini 1982a,b, Iben 1983, Hollowell 1988, Hollowell and Iben 1988, 1989); as a word of caution, it must be mentioned that other independent investigations (Lattanzio 1986, 1987, 1988, Boothroyd and Sackmann 1988a,b) have not succeeded in confirming this.

B) *Carbon Recombination and Semiconvection*

Sackmann (1980) pointed out that, following the disappearance of the convective shell in a low mass model, the carbon-rich matter which was once at the outer edge of the convective shell at its maximum extent is propelled outward to such low temperatures and densities that the contribution of carbon to the opacity may become significant and play an important role in the dredge up process. An explicit suggestion as to how this might come about and as to how it might also lead to the activation of the ^{13}C source was made by Iben (1982). The essential features of this suggestion are illustrated in Figure 8. The occurrence of convective motions and mixing is denoted by shading. The idea was that, after the disappearance of the main convective shell (the snail-like shapes in Figure 8) and upon cooling of matter at the edge of the carbon-rich zone left behind by the primary convective shell, the opacity of the carbon-rich material would lead to a secondary phase of convective shell mixing. Overshoot at the edge of this secondary zone would carry fresh carbon outward, causing an increase in opacity, thereby forcing the outer edge of the convective zone to move into a region which earlier contained only hydrogen and helium.

Once the secondary convective shell vanished, the base of the fully convective envelope would extend inward, dredging up the fresh products of helium burning and neutron-capture nucleosynthesis contained in the region once occupied by the secondary convective shell, even though the base of the convective envelope did not, as extant models suggested, extend as far inward as the point defined by the outermost extent of the primary convective shell.

The ^{13}C source would be activated in the following way. In the lower portion of the region once within the secondary convective shell (the portion not affected by dredge

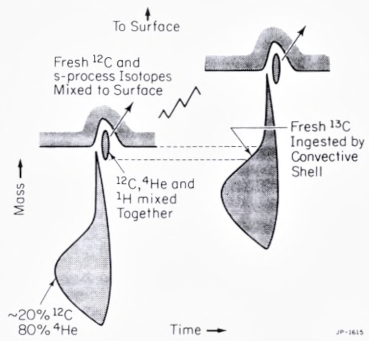

Figure 8. Schematic showing how convective zones might behave in an AGB star of small core mass. The primary convective shell (lowermost of the three shaded zones) and the base of the convective envelope (lower boundary of the uppermost convective zone) are consequences of model evolution when the opacity due to partially recombined carbon is neglected. The small, intermediate convective zone is a fabrication, based on the hope that recombination opacity may force the development of a fully convective zone which brings fresh ^{12}C and ^{1}H together at comparable number abundances.

up) both ^{12}C and ^{1}H are to be found. When this region heats up sufficiently, ^{12}C is converted into ^{13}C following a proton capture and a β decay. If the initial number abundances of ^{1}H and ^{12}C in the region are comparable, then most of the freshly formed ^{13}C is not destroyed by further proton capture. When the next thermal pulse occurs, the ^{13}C is ingested by the growing primary convective shell; it will be convected to the base of the convective shell where temperatures become large enough ($\geqslant 150 \times 10^6$K) to activate the $^{13}C(\alpha,n)^{16}O$ reaction.

This hypothetical scenario is close to, but not quite like what actually happens in current model calculations. Instead of being fully convective, the secondary mixing zone is actually semi-convective (Iben and Renzini 1982a,b, Hollowell 1988, Hollowell and Iben 1988, 1989). The reason for this is that, in the region where carbon is partially recombined with electrons, the opacity is nearly proportional to the carbon abundance. Small, shifting convective regions appear near the outer edge of the zone formerly contained in the primary convective shell and convective overshoot carries some carbon outward into a region containing hydrogen, raising the opacity there. The opacity in the region from which this carbon comes is reduced, thus lowering the degree of instability against convection. Ultimately, the abundance of ^{12}C throughout a large region readjusts in such a way that the radiative gradient equals the adiabatic gradient. This is just the classical requirement for semiconvection.

The time-dependent behavior of convection in the semiconvective zone of a model of low core mass, low mass, and low metallicity is shown in Figure 9 (Hollowell 1988, Hollowell and Iben 1988, 1989). Analytic approximations to Cox-Kidman (1986) opacities are used and overshoot has not been explicitly taken into account.

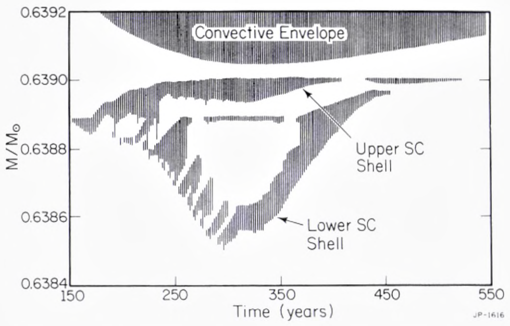

Figure 9. Structure of semiconvective (SC) regions in a model AGB star of low metallicity when the opacity due to partially recombined carbon is included. Hydrogen is carried downward by the lower of two main branches and carbon is carried outward by the upper branch. Dredge up does not occur in this instance, as overshoot has not been assumed.

The vertical bars depict the extent of convection in selected models. Once fresh carbon is mixed with some hydrogen, the progress of convective flow proceeds in two directions. An upper "semiconvective shell" carries carbon further outwards into hydrogen-rich material and a lower semiconvective shell carries hydrogen deeper into carbon-rich regions. The total mass of hydrogen mixed in with fresh carbon is of the order of 10^{-6} $M_\odot$.

Long after the semiconvective episode is completed and toward the end of the ensuing quiescent helium-burning phase, the matter once within the semiconvective zone is heated to the extent that protons react with ^{12}C. Within the inner portion of this zone, the final product is mostly ^{13}C (at a total mass of $\sim 5 \times 10^{-6}$ $M_\odot$). In the outer portion of the zone, the final product is mostly ^{14}N. Thus a thin layer of matter containing ^{13}C is topped by another thin layer containing ^{14}N.

When the next thermal pulse occurs the two layers lie approximately half-way between the base of the primary convective zone which is formed and the location of the hydrogen-helium discontinuity. As the primary convective shell grows, its outer edge encounters the ^{13}C containing layers and, over the ensuing $\sim$ 10 yrs, ^{13}C flows into the convective shell.

C) *Nucleosynthesis of Neutron-Rich Isotopes*

When the ^{13}C is ingested by the primary convective shell, it participates in the convective flows which carry matter back and forth within the shell at the rate of one transit per $\sim 3\times 10^4$ s. The lifetime of ^{13}C against α capture with neutron release varies from ~ 100 yr at the top of the shell, where the temperature is ~ 10^8K, to ~ 10 days at the base of the shell, where the temperature is ~ 1.5 × 10^8K. Hence, ^{13}C burns effectively only near the base of the shell and each ^{13}C nucleus which is introduced experiences roughly 30 transits of the convective shell before capturing an α particle and releasing a neutron near the base. It requires approximately $\tau_{in} \sim 10$ years for the ^{13}C-rich layer to be ingested by the outwardly growing convective shell, so that the effective rate at which neutrons are released is governed, not by the rate at which ^{13}C transits within the shell nor by the rate at which ^{13}C burns at the base of the shell, but by the rate at which ^{13}C is engulfed by the shell (Iben 1983, Hollowell 1988, Hollowell and Iben 1988, 1989).

The neutron density which results at the base of the shell where neutron-capture nucleosynthesis occurs is particularly interesting. In the model constructed by Hollowell (1988), the total mass of ^{13}C ingested by the convective shell is $M_{13} \sim 5 \times 10^{-6}\ M_\odot$. One may assume, in first approximation that neutrons are released in a "burning" zone of mass M_{burn} near the base of the convective shell and that the rate at which neutrons are released in this zone is

$$\dot{N}_n^+ = (M_{13}/13M_H)\ \frac{1}{\tau_{in}}\ M_{burn}\ /M_{CS}, \tag{1}$$

where $M_H \simeq$ mass of a neutron and M_{CS} ($\sim 0.01\ M_\odot$) is the mass of the convective shell during the ingestion phase.

The rate of neutron captures in the burning zone is

$$\dot{N}_n^- = n_n \sum_i n_i <\sigma_i\ v_i> (M_{burn}/\ \varrho), \tag{2}$$

where n_n = neutron number density (cm^{-3}), ϱ = density (gm cm^{-3}), n_i = number density of the ith neutron absorber, σ_i = neutron capture cross section of the absorber, v_i = relative velocity of neutron and absorber, and brackets denote an average over a Maxwell-Boltzmann distribution.

Equating expressions (1) and (2), one has

$$\frac{M_{13}}{M_{CS}}\ \frac{1}{\tau_{in}} = n_n \sum_i < \sigma_i v_i > X_i\ (13/A_i), \tag{3}$$

where X_i is the abundance by mass of the ith absorber and A_i is its atomic mass (in units of M_H). From Hollowell (1988) and Hollowell and Iben (1988, 1989), one has that

$$\sum_i < \sigma_i v_i > X_i\ (13/A_i) \sim 4 \times 10^{-19}\ Z\ \mathrm{cm}^3, \tag{4}$$

where Z is the abundance by mass of elements heavier than helium and it has been assumed that these elements are in the solar-system distribution. This assumption is a dangerous one to make since both ^{22}Ne and ^{12}C are present at abundances far in excess of solar (relative to each other).

Equating expressions (3) and (4), one obtains

$$n_n \sim 4.2 \times 10^6 \ \mathrm{cm}^{-3} \ /Z, \tag{5}$$

which, for the $Z = 10^{-3}$ model constructed by Hollowell gives $n_n \sim 4.2 \times 10^9 \ \mathrm{cm}^{-3}$, a value too large by perhaps one order of magnitude or so to give a solar system distribution of s-process isotopes.

The situation is improved if one takes into account that both ^{12}C (Gallino et al 1988) and ^{22}Ne (Hollowell and Iben 1988, 1989) are overabundant in the convective shell relative to solar, since both are products of α burning. As a next approximation, one may write

$$\begin{aligned} \Sigma < \sigma_i v_i > X_i \ (13/A_i) &\sim [4 \times 10^{-19} Z \\ &+ 0.2 \times 7.1 \times 10^{-20} \ X_{22} \\ &+ (0.003 - 0.2) \times 1.3 \times 10^{-19} \ X_{12}] \ \mathrm{cm}^3 \ \mathrm{s}^{-1}, \end{aligned} \tag{4'}$$

where the cross section for neutron capture on ^{22}Ne at $T = 150 \times 10^6$ K is assumed to be $\sim$ 0.2 mb and that of ^{12}C is assumed to be between 0.003 mb (Fowler 1967) and 0.2 mb (Bao and Käppeler 1987). The abundance by mass of ^{12}C is typically 0.2 and that of ^{22}Ne is $\sim$ (0.5-1.5)Z (Hollowell and Iben 1988, 1989). Hence,

$$\Sigma \ (\sigma_i v_i \ X_i \ 13/A_i) \sim [5.4 \times 10^{-19} Z + 2.6 \times 10^{-20} (0.003\text{-}0.2)] \mathrm{cm}^3 \ \mathrm{s}^{-1} \tag{4''}$$

and, in the $Z = 0.001$ model, $n_n \sim 2.7 \times 10^9 \ \mathrm{cm}^{-3}$ if $\sigma_{12} \sim 0.003$ mb and $n_n \sim 2.9 \times 10^8 \ \mathrm{cm}^{-3}$, if $\sigma_{12} \sim 0.2$ mb.

The situation is improved further by taking into account the fact that neutron capture on ^{12}C produces ^{13}C which can then, on α capture, recycle neutrons (Gallino et al. 1988). The effect is to spread the neutron-capture episode over a longer time, thereby reducing the average neutron density. The final improvement comes when it is recognized that the final s-process distribution "freezes out" at a neutron density considerably less than the average one (Cosner, Iben, and Truran 1980). Gallino et al. (1980) show that freezeout occurs at $n_n \sim 2 \times 10^8 \ \mathrm{cm}^{-3}$, almost precisely the density required for producing the solar-system distribution! Not only that, but the abundance distributions currently being found at the surfaces of carbon stars in our own galaxy (and these must be the product of thermal pulse evolution) are showing that real AGB stars are producing s-process isotopes in nearly the solar-system distribution (Lambert 1988). The only untidy note in this otherwise beautifully developing scenario is that it has not thus far been demonstrated that models of population I composition can activate the ^{13}C neutron source, although dredge up does occur (Iben 1983).

Observations and analysis of peculiar red giants in the Galactic disk (e.g., Scalo and Miller 1979, Lambert 1988, and Jura 1988) show that population I AGB stars of small initial mass ($\sim$ 1.5 $M_\odot$) and small core mass certainly activate the ^{13}C neutron source and dredge up freshly processed material. It is obvious that those of us in the theoretical modeling business have not yet completed our task.

IV. *Epilogue*

The tantalizing promise of ultimate concordance between theory and observation occasions me to close with a personal observation.

Over 20 years ago I gave a lecture at a Stonybrook conference organized by H. Y. Chiu. Bengt Strömgren was in the audience. In my lecture I very proudly demonstrated how a very simple back-of-the-envelope calculation using basic physics provided as good an estimate of the central temperature of the Sun as that provided by the most sophisticated stellar model calculation with the most sophisticated input physics. After the lecture, Bengt came up to me and said: "you were very lucky."

Some fourteen years ago I found by accident that intermediate-mass AGB model stars activate the ^{22}Ne neutron source and dredge up freshly made s-process isotopes and freshly made carbon to their surfaces. In response to a referee's comment about my first cursory speculations concerning the production of s-process isotopes, I spent several months becoming acquainted with the nuclear astrophysics lore in this field and wrote a companion paper extolling the virtues of the ^{22}Ne source. In subsequent papers with Jim Truran and Ken Cosner, and in countless reviews I continued to extol the virtues of this source, absolutely convinced that intermediate-mass AGB stars produce s-process isotopes in the solar-system distribution and not fully appreciative of the knowledgeable nucleosynthesists' arguments that the neutron densities during the neutron-capture episode are too large.

Over the next ten years, the observations of Magellanic Cloud AGB stars showed that intermediate-mass AGB stars do not live long enough in the thermally pulsing phase to be major contributors to the galactic nucleosynthesis of s-process isotopes, and theoretical nucleosynthesis studies showed this to be just as well, as otherwise a major discrepancy between theory and observation would have persisted (see, e.g., Iben 1988 for a summary). Had Bengt Strömgren spoken to me after I had given a talk describing the debacle of a lovely, but oversimplified theory, he might have said: "you were very unlucky."

On this occasion, I am reluctant to claim that I know where and how solar-system s-process isotopes are made, even in stars of low metallicity. Past experience cautions that, once again, those of us in the model making business may have been (temporarily) very lucky.

Acknowledgements

I would like to thank David Hollowell and Roberto Gallino for enlightening conversations on neutron-capture nucleosynthesis in the context of low mass, low metallicity AGB stars and to thank Robert MacFarlane for his expert draftsmanship. Preparation of this paper was supported in part by the U. S. National Science Foundation grant AST 84-13371.

References

Anders, E., and Ebihara, M. 1982, *Geochim. Cosmochim. Acta.*, **46**.
Bao, Z. Y., and Käppeler, F. 1987, *Atomic Data Nucl. Tables*, **36**, 411.
Becker, S. A., Iben, I. Jr., and Tuggle, R. S. 1977, *Ap. J.*, **218**, 633.
Boothroyd, A. I., and Sackmann, I.-J. 1988a, *Ap. J.*, **328**, 653.
Boothroyd, A. I., and Sackmann, I.-J. 1988b, *Ap. J.*, **328**, 671.
Busso, M., Picchio, G., Callino, R., and Chieffi, A. 1988, *Ap. J.*, **326**, 196.
Clayton, D. D., Fowler, W. A., Hull, T. C., and Zimmerman, B. 1961, *Ann. Phys.*, **12**, 121.
Cosner, K., Iben, I. Jr., and Truran, J. W. 1980, *Ap. J. Lett.*, **238**, L91.
Cox, A. N., and Kidman, R. 1986. Private communication.
de Jong, T. 1983, *Ap. J.*, **274**, 252.
Despain, K. H. 1980, *Ap. J. Lett.*, **236**, L165.
Fowler, W. A., Coughlin, G. R., and Zimmerman, B. A. 1975, *Ann. Rev. Astron. Astrophys.* **13**, 69.
Gallino, R., Busso, M., Picchio, G., Raiteri, C. M., and Renzini, A. 1988, *Ap. J. Lett.*, **334**, L45.
Habing, H. J. 1986. In *The Galaxy*, ed. G. Gilmore and B. Carswell (Dordrecht: Reidel), p. 173.
Hollowell, D. 1988. Ph.D. Thesis, University of Illinois.
Hollowell, D., and Iben, I. Jr. 1988, *Ap. J. Lett.*, **333**, L25.
Hollowell, D., and Iben, I. Jr. 1989, *Ap. J.*, **340**, 966.
Howard, W. M., Mathews, G. J., Takahashi, K., and Ward, R. A. 1986, *Ap. J.*, **309**, 633.
Iben, I. Jr. 1975a, *Ap. J.*, **196**, 525.
Iben, I. Jr. 1975b, *Ap. J.*, **196**, 549.
Iben, I. Jr. 1976, *Ap. J.*, **208**, 165.
Iben, I. Jr. 1977, *Ap. J.*, **217**, 788.
Iben, I. Jr. 1981. In *Effects of Mass Loss on Stellar Evolution*, ed. C. Chiosi and R. Stalio (Dorcrecht: Reidel), p. 373.
Iben, I. Jr. 1982, *Ap. J.*, **260**, 821.
Iben, I. Jr. 1983, *Ap. J. Lett.*, **275**, L65.
Iben, I. Jr. 1985, *Q. Jl. R. Astr. Soc.* **26**, 1.
Iben, I. Jr. 1988. In *Astronomy in the Southern Hemisphere*, ed. V. M. Blanco and M. M. Phillips (Provo: Brigham Young), p. 220.

Iben, I. Jr., and Renzini, A. 1982a, *Ap. J. Lett.*, **259**, L79.
Iben, I. Jr., and Renzini, A. 1982b, *Ap. J. Lett.*, **263**, L23.
Iben, I. Jr., and Renzini, A., 1983, *Ann. Rev. Astron. Astrophys.* **21**, 271.
Jura, M. 1988, *Ap. J. Suppl.*, **66**, 33.
Kwok, S. 1987, *Phys. Reports*, **157**, 3, 111.
Lambert, D. L. 1988. This conference.
Lattanzio, J. 1986, *Ap. J.*, **311**, 708.
Lattanzio, J. 1987, *Ap. J. Lett.*, **313**, L15.
Lattanzio, J. 1988. In *The Origin and Distribution of the Elements*, ed. G.J. Mathews (Singapore: World Scientific), p. 398.
Malaney, R. A., and Boothroyd, A. I. 1987, *Ap. J.*, **320**, 866.
Mathews, G.J., and Ward, R. A. 1985, *Reports on Prog. in Phys.*, **48**, 1371.
Mengel, J. G., and Sweigart, A. V. 1981. In *Astrophysical Parameters for Globular Clusters*, ed. A. G. D. Philip (Dordrecht: Reidel), p. 277.
Renzini, A., and Voli, M. 1981, *Astron. and Ap.*, **94**, 175.
Sackmann, I.-J. 1980, *Ap. J. Lett.*, **241**, L37.
Scalo, J. M., and Miller, G. E. 1979, *Ap. J.*, **233**, 596.
Seeger, P. A., Fowler, W. A., and Clayton, D. D. 1965, *Ap. J. Suppl.*, **11**, 121.
Truran, J., and Iben, I. Jr. 1977, *Ap. J.*, **216**, 797.
Ulrich, R. K. 1973. In *Explosive Nucleosynthesis*, ed. D. N. Schramm and D. W. Arnett (Austin: U. Texas), p. 139.
Wood, P. R., Bessell, M. S., and Fox, M. W. 1983, *Ap. J.*, **272**, 99.
Wood, P. A., and Zarro, D. M. 1981, *Ap. J.*, **247**, 247.

Nucleosynthesis in Stars

By DAVID ARNETT
Departments of Physics and Astronomy
University of Arizona
Tucson, Az 85721, USA

Abstract

The supernova 1987A in the Large Magellanic Cloud is the brightest since the invention of the telescope. It provides an exceptional test of well-developed theoretical ideas concerning the evolution and death of massive stars. Such objects are thought to be the primary contributors to nucleosynthesis, and therefore central to understanding the evolution of galaxies. The observation of a antineutrino burst from the event provides a confirmation of the theory of gravitational collapse of stellar cores, and of modern theories of the weak interaction.

I. *Introduction*

The evolution of stars, and the use of their composition as a tracer og galactic evolution, were central themes in the work of Bengt Strömgren. Supernova 1987A has provided striking insight into these problems, and in particular into the evolution, death and heavy element yield of a massive star – just the sort of object thought to be a major contributor to nucleosynthesis.

In addition to the work presented here, that of two other groups (Nomoto *et al.*, and Woosley *et al.*) was conducted at about the same time and with similar results. Detailed references to this and other work may be found in the review by Arnett, Bahcall, Kirshner, and Woosley (1989). This discussion will focus on the broad features of what has been learned from SN1987A.

II. *Nucleosynthesis and Structure*

The first step in investigating nucleosynthesis in a star is to determine which nuclei and which reactions are to be considered. This defines a nuclear reaction network. Because of the enormous variation in reaction time it is possible (and for economy, advisable) to ignore classes of reactions and nuclei.

In Figure 1 is shown the reaction network necessary for a correct treatment of energy generation and electron capture through oxygen burning. Because of their low thresholds, ^{31}P, ^{33}S, and ^{35}Cl actively capture electrons at the densities at which hydrostatic oxygen burning occurs, and they are produced by secondary reactions in oxygen burning. This affects the star by reducing the number of electrons available

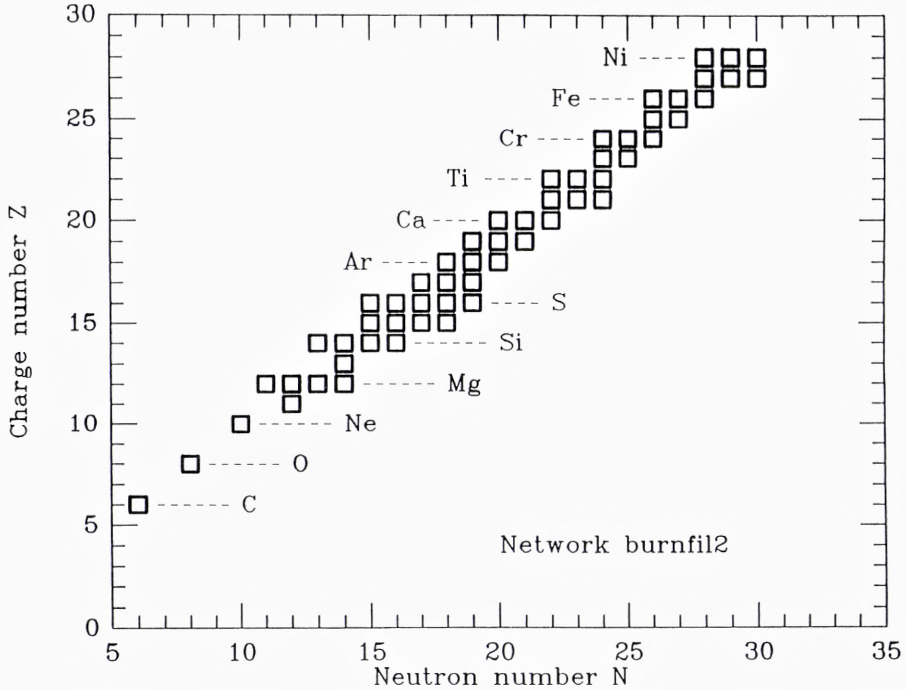

Figure 1. Reaction Network Used.

for pressure, and by reducing the energy by that carried away by $\nu\bar{\nu}$ in an URCA cycle of e-capture and decay. It affects the nucleosynthesis (and hence the energy release) by changing the ratio of neutrons to protons in the reacting matter. Consider ^{33}S: $Y_e \approx Z/A = 16/33 = 0.4848$ is in one sense optimum for the production of this nucleus. For larger Y_e there are excess protons, and for smaller values, excess neutrons. Oxygen burning proceeds, making a trace of ^{33}S. It captures an electron, reducing Y_e from an initial value of about 0.4985 toward 0.4848, which increases the amount of ^{33}S, and so on until $Y_e \approx 0.4848$. Note that the crucial quantity is neutron excess $\eta = (N - Z)/A$, which is 0.003 and 0.030 for Y_e equal 0.4985 and 0.4848, respectively; η changes by a factor of ten! Further reduction in Y_e then reduces the abundance of this nucleus, and therefore its contribution to electron capture. The process "saturates". The other two nuclei have similar Z/A, and the three dominate electron capture under typical oxygen burning conditions. The result is that Y_e approaches 0.48 and tends to stay at this value for oxygen burning. Because electron degeneracy pressure is important in supporting the stellar core, smaller core masses result. However, since the electrons are only partially degenerate, the heating/cooling implied by such processes also can modify the core mass. This entropy increase/

decrease depends upon the coupled hydrodynamics and reaction dynamics of the convective, burning flow. This problem has not been analyzed; only complete mixing approximations (spherical symmetry) have been used to date.

This matter evidently does not escape the star in large quantities. The most abundant nuclei with similar Z/A are ^{54}Fe and ^{58}Ni; they comprise only 0.1 of the solar system abundance of ^{56}Fe. This suggests that matter which undergoes *hydrostatic* oxygen burning ends as part of the neutron star; *explosive* oxygen burning occurs too fast for electron captures, and – for η near the expected value of 0.003 – tends to make an isotopic distribution like that of the solar system.

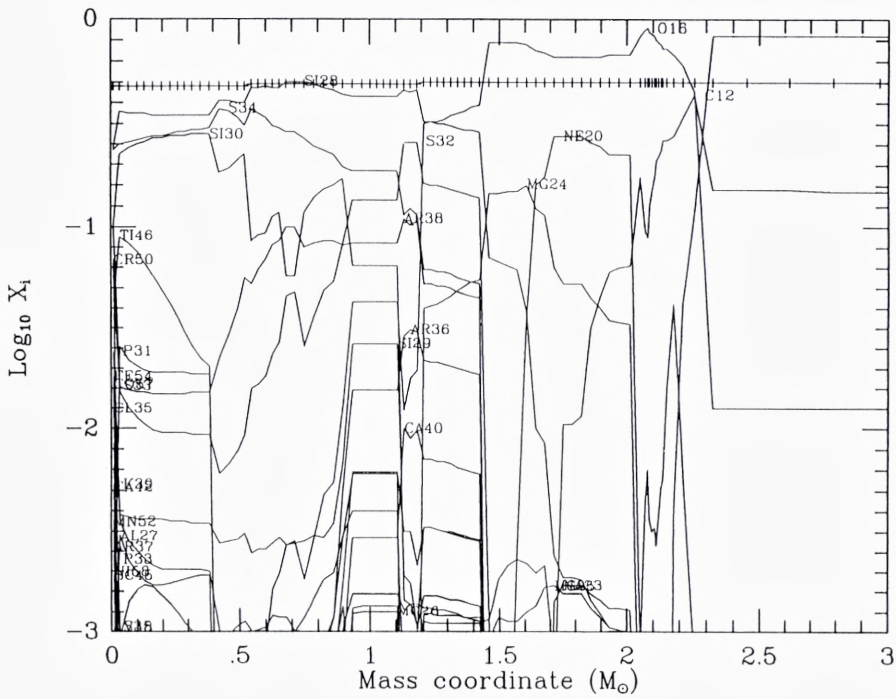

Figure 2. Detailed Abundance in the Inner 3 Solar Masses.

Figure 2 shows the complexity of the behavior of abundance versus mass coordinate in the core of an evolved massive star (Arnett 1988a). While this complication is important for understanding the approach to core collapse, the abundance patterns beyond 1.5 solar masses are considerably simpler. These are the regions that are ejected; the innermost of which have Y_e close enough to 0.5 to make the ^{56}Ni which powered the light curve of SN1987A and the daughter nucleus ^{56}Co.

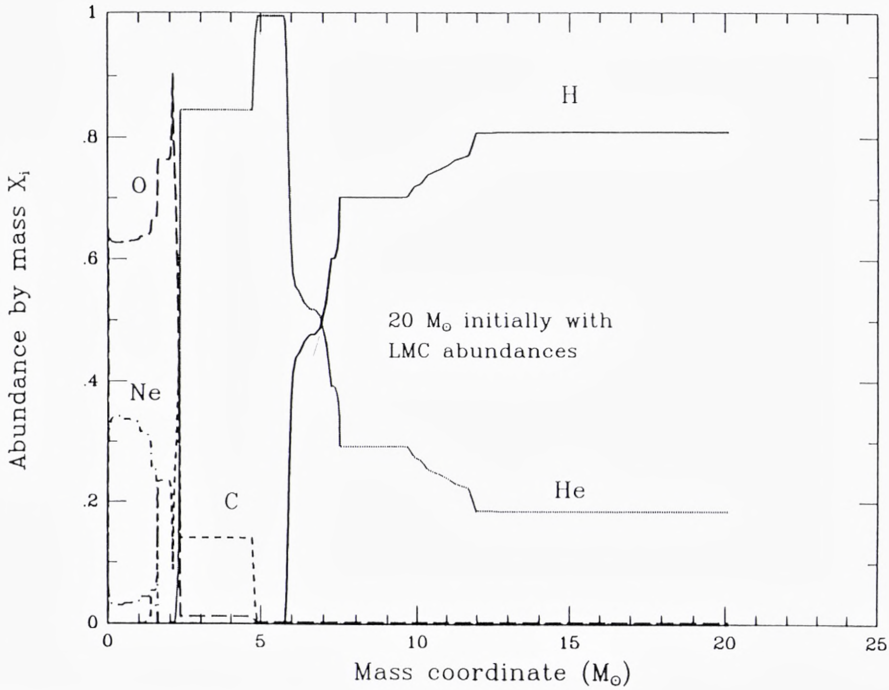

Figure 3. Abundance Structure of a 20 Solar Mass Star.

Figure 3 gives the abundance distribution of the major species for the whole star (of which Figure 2 showed the innermost region). Outside the core there is a He mantle and a H-rich envelope. There is steep density gradient between the core and the He mantle, and a lesser one between the He mantle and the H-rich envelope. The core-mantle gradient is a consequence of $\nu_e\bar{\nu}_e$ emission during carbon, neon, oxygen and silicon burning. This behavior does not occur without a direct e - ν coupling of about the strength predicted by Conserved Vector Current and Weinberg-Sahlam neutral current theory of the weak interactions, and subsequently detected experimentally. It is the pronounced core-mantle gradient which gives the small yield of ^{56}Co that was seen in SN1987A. By steep density gradient we imply a small mass which has a large range in density. At high density, Y_e is too small (*i.e.*, too neutron rich) to allow ^{56}Ni production. At low density, the shock does not heat the ejected matter enough to burn to ^{56}Ni.

A typical density structure is shown in Figure 4. The steep density gradient near 1.5 solar masses is the core-mantle interface, and is similar in stars of, say, 10 to 30 solar masses. This is the result of "core convergence" due to $\nu_e\bar{\nu}_e$ emission mentioned above. At 4 solar masses there is a smaller bump, which is the interface between the

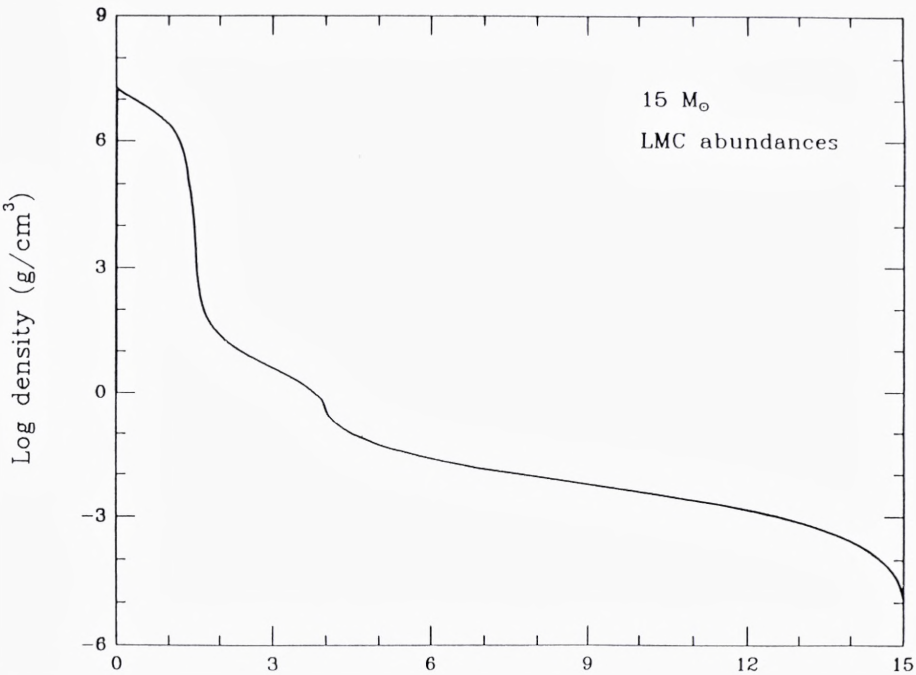

Figure 4. Typical Density Structure for a Presupernova.

He mantle and the envelope. The position in mass of this bump depends upon the mass of the star during hydrogen burning. For SN1987A we can estimate this from the luminosity of the progenitor; there is a consensus on 6 $\pm$ 1$M_\odot$. This in turn suggests a initial stellar mass of 20 $M_\odot$ or so. The bump at the mantle-envelope interface may cause the shock to become nonspherically symmetric by generating an entropy bubble which will be Rayleigh-Taylor unstable. Asphericity was noted in the hydrogen lines at velocities appropriate to matter near this interface.

III. *The Hertzsprung Russell Diagram*

One of the major new facts that SN1987A provided was the nature of the presupernova star: it was a B3 supergiant with $M_{bol} \approx -7.8$ (the star Sanduleak -69 202; see Humphreys and McElroy 1984). Because massive stars evolve quickly and are rare, their evolution is more difficult to unravel from Hertzsprung-Russell diagrams of clusters. Young clusters are too rare and too sparse in massive stars to give good statistical accuracy. The theoretical problem is exacerbated by the observed importance of mass loss, binary companions, and more rapid rotation. Further, the increasing importance of radiation pressure means that mixing, an irreversible process, is relatively easier in massive stars. Finally, a doubly diffusive instability involving radiation and composition called "semiconvection" is thought to be important and has certainly been controversial.

Most calculations of the evolution of massive stars through hydrogen and helium burning agreed that such objects would be red supergiants, not blue. Observed supernovae of Type II had light curves, temperatures and velocities which were consistent with the large radii of red supergiants, although it had been suggested that some massive stars might become supernovae as more compact (blue) objects (Arnett 1977).

Initial abundance seems to play a major role in producing a blue presupernova. Mass loss may be important also, but most of the hydrogen envelope must remain on the star till explosion to fit the observations (Arnett 1988a; Woosley, Pinto and Ensman 1988). Variations in the profile of the composition gradient in H has long been known to cause a massive star to make "blue loops" in the HR diagram (*e.g.*, Chiosi and Summa 1970). This could be brought about by semiconvection or other sorts of mixing of this region.

Some mass was lost prior to explosion because slow moving matter, which appears to be nitrogen rich, has been observed. Since only a few percent of H consumption is required for massive stars to covert CNO nuclei to ^{14}N, this does not necessarily imply extensive mass loss; it might imply extensive but slow mixing. Observations also seem to suggest an enhancement of He in this matter. This is consistent with such mixing, and would tend to drive the presupernova blueward (Nomoto *et al.* 1988).

Much work is necessary to sort out the complexity of this aspect of stellar evolution. Fortunately this ambiguity relates to the path to explosion more than to the nature of the object at explosion. The core and mantle structure is oblivious to the radius of the envelope (except in the extreme case that the surface convective zone reaches down into the mantle). Models which arrive by different paths to the correct region of the HR diagram, have relatively similar envelope structure as well. The problem is not whether blue presupernova models can be constructed, but *which* way nature makes them.

Figure 5 shows one way to make a blue presupernova (Arnett 1987b). The key was to use abundances one quarter of solar and Ledoux semiconvection (more or less). These models need to be modified to provide some mass loss of N-rich matter; this is easy to say, but providing the uniquely correct physical mechanism requires more effort. The presupernova had a luminosity within a factor of two of 10^5 suns; its temperature is shown by the two vertical dashed lines. For this evolution, a star of slightly below 20 $M_\odot$ would fit this error box nicely. From shock calculations of the

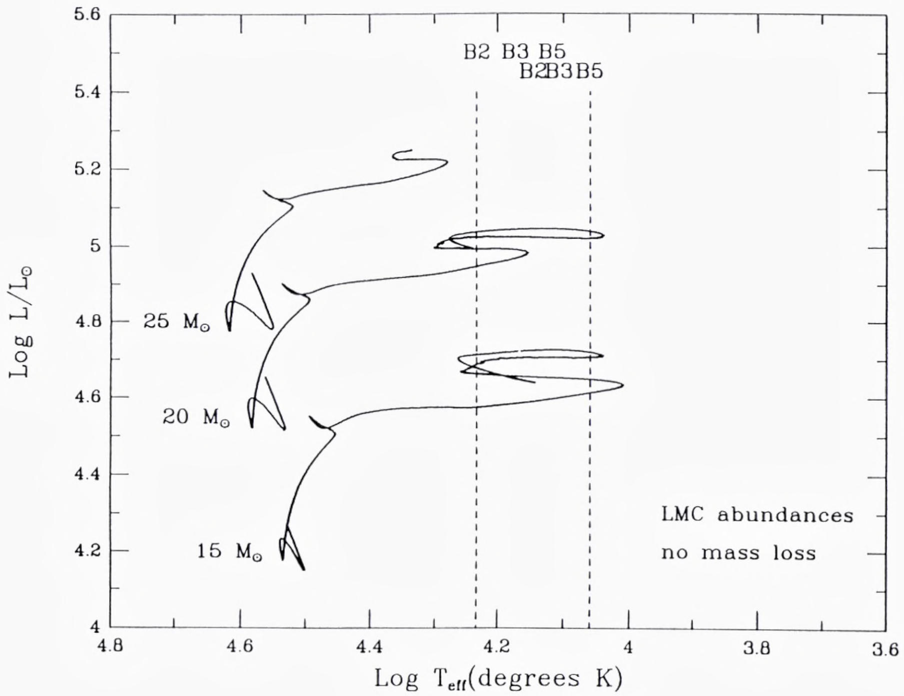

Figure 5. HR Diagram. The evolution is shown for stars of 15, 20 and 25 solar masses. The initial abundance of heavy elements ($Z > 2$) was 0.25 of solar. This is one of several ways to make a blue presupernova; it probably fails in that it does not give a trip to the red before explosion.

early light curve, to be presented below, we can infer a presupernova radius of (3 to 6) $\times$ 10^{12} cm. This passes through the error box nicely.

Theories which do not take the presupernova to have been the Sanduleak −69 202 star (*e.g.*, most binary models) clash with the first two weeks of observation of SN1987A (see below), and therefore seem unattractive.

IV. *Core Dynamics*

As the nuclear burning in the core exhausts the fuel, heat loss due to neutrino emission drives further contraction. As the compressional heating procedes (the star has an effectively negative specific heat), nuclear photodissociation occurs, giving rise to a hydrodynamic instability toward collapse. As density rises, electron fermi energies rise as well, but the threshold for electron capture on nuclei inhibits this process.

Eventually some electron capture occurs, but at sufficiently high density that the neutrinos produced do not escape freely; their diffusion time is longer than the collapse time. Thus "neutrino trapping" occurs. This inhibition of electron capture and lepton loss keeps the entropy low. Thus the inner core (a mass of about 0.7 $M_\odot$) collapses as a unit until it reaches nuclear density, at which point the nucleon-nucleon interaction becomes repulsive enough to stiffen the equation of state. The collapse is halted for the inner core, it rebounds, and the outer core material now rains down supersonically. A shock wave is formed which begins to propogate outward, reversing the infall and photodissociating the nuclei in the infalling matter. So far there is general agreement among theorists (see Brown 1988, and other papers in that volume).

At present the mechanism of explosion is unclear. In the "prompt shock" picture, the shock continues outward, ejecting the mantle and envelope. Unfortunately the best calculations to date agree that for realistic physics, the energy losses due to photodissociation of nuclei and to neutrino emission as the shock moves to lower densities where it becomes transparent, conspire to kill the shock. The competition is the "delayed mechanism". In this picture the shock dies (at about 20 milliseconds after bounce), but after some time (hundreds of milliseconds) heat transfer by the slowly diffusing neutrinos heats the infalling matter. As the pressure rises, the infall slows and reverses itself. This then drives off the mantle and envelope. There are no numerically reliable computation of this process as yet, just some interesting pioneering studies.

The precise nature of the explosion mechanism is important in that it determines the energy of the explosion, the mass of the condensed remnant left, and the yield of the heavy elements in the innermost region of the ejecta.

Figure 6 indicates the nature of the ejection process for a "toy" theory, in which the neutrino processes were artificially frozen to represent a limiting case in which neutrino cooling did not kill the shock (Arnett 1987c). It is to be taken as roughly representative of the sucessful realistic calculation which we still seek. The "mantle-envelope" shock (which may or may not be the "core" shock) propogates into the mantle and then the envelope, leaving behind a hot neutron star. Notice the extraordinary density gradient, comprising 18 powers of ten! The interface between the ejected mass and the edge of the new neutron star is a quasihydrostatic region, at a density near that of the post-shock matter. As the shock moves to lower densities, it lays down a hydrostatic trace which becomes the outer part of the neutron star. This trace involves increasingly less matter. This is the hydrodynamic process which determines the "mass cut". For the toy calculation shown, the explosion energy was a bit too large, the remnant mass too small, and too much deep neutron rich matter was ejected. All these problems could be solved in principle by the action of some mechanism to damp the explosion. The question is *what* mechanism? Either of the two theories mentioned above could accomplish this in principle; what did nature do?

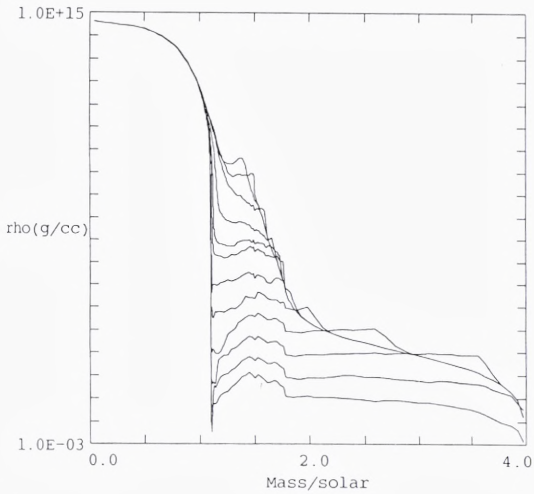

Figure 6. A Toy Model of Supernova Collapse and Explosion. By supressing neutrino transport, a hydrodynamic explosion resulted. The "mass cut" (the interface between remnant and ejecta) was carefully resolved in this calculation.

V. *Neutrinos*

The most dramatic observation of SN1987A was the detection of 11 neutrinos by the Kamiokande group (Hirata *et al.* 1987) and 8 by the IMB (Irvine, Michigan and Brookhaven) group (Bionta *et al.* 1987). There are possibly supporting observations from Baksan (Alekseev 1988). The number of facts that have been claimed to have been implied by these observations, like the number of papers concerning them, far exceed the number of events. We must not overinterpret. Statistics of small numbers are relevant here.

How many parameters may be inferred from 19 events? Every added parameter implies a division of the information content and a corresponding increase in statistical error. An astrophysicist would not be embarrassed to infer three parameters from this data, but even so the error is not negligible. For example, the three might be total energy release, neutrino temperature, and release time; these might map into the experimental data on number of neutrinos detected, energy of neutrinos detected, and time interval of detection, for example.

What was to be expected?

The energy release should reflect the binding energy of a neutron star. Typical theoretical models gave a gravitational binding energy of about 0.1 of the rest mass. Core convergence mentioned above gave core masses of roughly a Chandrasekhar mass, or about 1.5 $M_{\odot}$. This implies a binding energy of $B \approx 3 \times 10^{53}$ ergs.

The neutron star equation of state becomes stiff at densities at which repulsive interactions between nucleons become important. The atomic nucleus balances the attractive and repulsive nucleon forces for stability. Because these short range forces change rapidly with density, and also because the average density in a neutron star is slight less than its central value, the neutron star will settle to an average density close to nuclear density, or about 4×10^{14} grams cm^{-3}. This implies a radius of $R \approx 12$ km.

The luminosity in neutrinos of all types is the energy release divided by the diffusion time, $L \approx B/\tau$, and also the surface area times the emissivity per unit area, $L = 4\pi R^2 2 f\sigma T_e^4$, where f is the number of neutrino flavors and σ is 7/8 the usual Stefan-Boltzmann constant for photons (the 7/8 is due to Fermi-Dirac rather than Bose-Einstein statistics). This gives $T_e \approx 8.6[3/2f\tau]^{1/4}$ MeV for this radius R. For diffusion from a sphere, $\tau = 3R^2/\pi^2\lambda c$, where the mean-free-path λ is $1/N\sigma$. The number of interacting centers per unit volume is $N \approx \varrho N_A \approx 2.4 \times 10^{38}$, where N_A is Avogadro's number. The cross section is $\sigma \approx 2{\times}10^{-44}\varepsilon_\nu^2$ cm^2. Because of neutrino trapping, the lepton number is not much less than it was at the onset of collapse, $Y_e = 0.42$. The fermi momentum may then be scaled from that of the nucleon in the nucleus, which is at comparable number density. Doing this carefully gives $\mu_\nu = 100$ MeV, and $\varepsilon_\nu^2 = \mu_\nu^2/2$. Thus $\lambda = 3$ cm. In turn, this gives $\tau \approx (10^6)^2/(3(3)3{\times}10^{10}) = 3$ seconds.

Using this and assuming e, μ, and τ type neutrinos, so $f = 3$, we have $T_e \approx 5.5$ MeV. This will decrease slightly if gravitational redshift is corrected for.

Numbers of this sort were put together by many people before SN 1987A exploded, but unfortunately not concisely collected in a single publication.

From the observations (for example, see Burrows and Lattimer 1987, Bahcall, Piran, Press, and Spergel 1987, and Lamb, Melia and Laredo 1988), $B = (2 \text{ to } 4) \times 10^{53}$ ergs, $T_e \approx 3.5$ to 5 MeV, and $\tau \approx 4$ seconds. The agreement is dramatic. Unfortunately the limited number of neutrino events precludes a discrimination between the two pictures of the explosion mechanism discussed above.

VI. *Early Light Curves*

For the first two weeks the behavior of SN1987A was dominated by the effects of the shock and its heating. These were quickly calculated by dumping some amount of energy inside a presupernova model and calculating the hydrodynamics and radiative diffusion with a one dimensional hydrocode. Figure 7 shows shapshots of the structure of a representative model (Arnett 1988b). At 54 minutes the shock has reached the surface of the presupernova. Within another 50 minutes the structure has assumed a constant shape which then expands homologously. Subsequent behavior depends upon further heating and cooling.

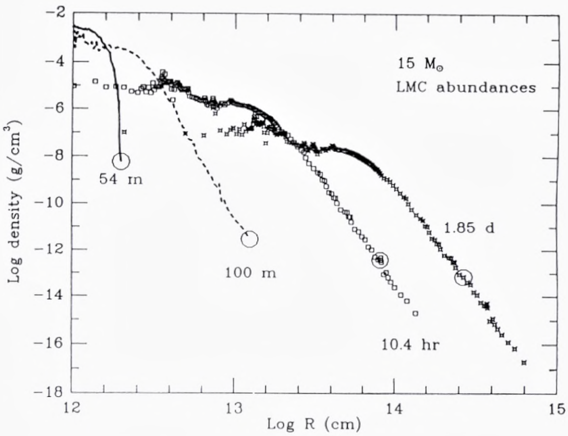

Figure 7. Snapshots of Density Structure in Supernova Ejecta.

The luminosity behaves as shown in Figure 8. The solid curve represents the visual magnitude V, and the dashed curve the bolometric magnitude. The time scale is normalized so the explosion occurred at the Kamiokande-IMB detection time for neutrinos. The theoretical curve gives a good representation of the fast rise implied by the earliest observations and limits. Note that there was a brief but intense flash, mostly in the ultraviolet, about an hour after the neutrino detection. This occurs when the shock hits the stellar surface. The agreement is good between observation and theory; the adjusted parameter is the shock energy (here 2×10^{51} ergs).

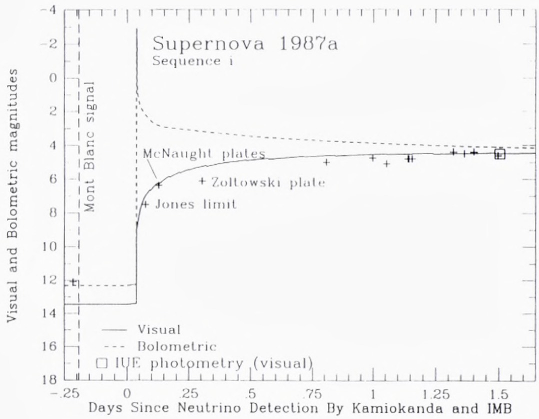

Figure 8. Comparison of Theory and Observation of the Early Light Curve of SN1987A.

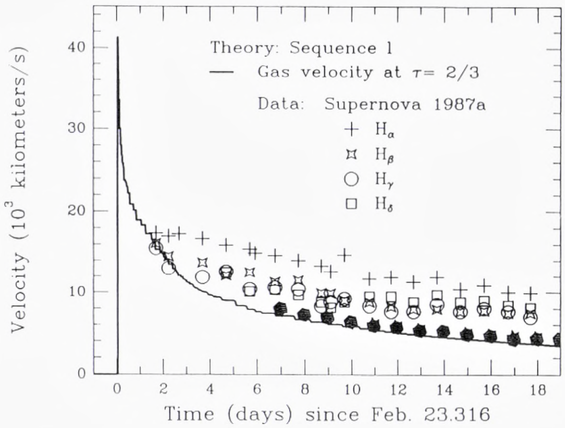

Figure 9. Velocities in SN1987A.

Figure 9 shows the change in photospheric velocity over the first few weeks. The solid curve is the theoretical model and the points are various upper limits to the photospheric velocity inferred from different spectral lines. No new parameter adjustments were made.

Figure 10 shows the change in effective temperature (now of photons, not neutrinos!) over the first few weeks. The solid curve is the theoretical model and the points are for values inferred from UBV observations (pluses) and IUE data (boxes); detalied references and discussion is to be found in Arnett (1988b). Again, no new parameters were adjusted.

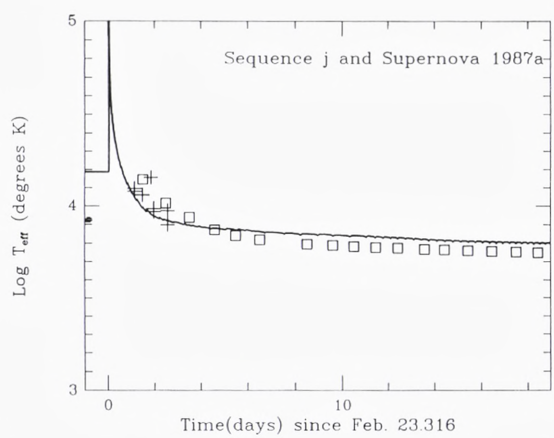

Figure 10. Effective Temperature in SN1987A.

The shock model provides a superb representation of the observational data over the first few weeks. After that, the decay of ^{56}Co becomes evident.

VII. *The Nickel Bubble*

The most tightly bound nucleus having equal numbers of protons and neutrons ($Z = N$) is ^{56}Ni. Therefore, any fuel with $Z = N$ will tend to burn to ^{56}Ni.

Hydrogen burning converts H to ^{4}He, which has $Z = N$, and rearranges the CNO isotopes to ^{14}N, which also has $Z = N$. Helium burning converts ^{4}He to ^{12}C and ^{16}O (again $Z = N$), while $^{14}\mathrm{N}(\alpha,\gamma)^{18}\mathrm{F}(\beta^+\nu_e)^{18}\mathrm{O}$ increases the neutron excess to $\eta = (N - Z)/A \approx 0.002$ in matter initially having solar abundance. The burning of carbon and neon change this little (to $\eta \approx 0.003$), and are essentially a rearrangement of $Z = N$ nuclei.

For $\eta \leqslant 0.003$, these fuels will burn to ^{56}Ni if heated to explosive temperatures.

With hydrostatic oxygen burning the picture changes, as discussed above, approaching $\eta \approx 0.030$. For such high neutron excess, the matter no longer burns to ^{56}Ni.

Production of ^{56}Ni indicates that matter *around* the stellar core has been *explosively* heated.

This has two implications: (1) because it is an endpoint in burning, ^{56}Ni can be made in relatively large abundance, and (2) because it is radioactive, it and its daughter ^{56}Co will store energy until they decay.

Type I supernovae have fast radiative diffusion times, so that the ^{56}Ni decay is evident, and produces the peak in the light curve. For more massive (or more slowly expanding) supernovae, a longer diffusion time will smooth out the ^{56}Ni peak, allowing the ^{56}Co one to dominate. In SN1987A this begins to occur in the third week, and is observed to continue for at least the next year and a half.

As we saw above, the ^{56}Ni is made in matter which just escapes the star, *i.e.*, just outside the "mass cut". The shock leaves the ejected matter moving almost homologously ($v \propto r$). The ^{56}Ni is therefore some of the most slowly moving matter; in models it moves at about 1,000 km/s. As ^{56}Ni decay occurs, the gamma-rays are trapped, heating the matter. This increases the pressure relative to unheated matter, setting up a pressure gradient and driving additional mass motion. A hot "nickel bubble" is formed, underlying slowly moving matter. It becomes Rayleigh-Taylor unstable, and overtakes overlying material. This intrusion gives macroscopic mixing, but the mean free paths are too small for complete microscopic mixing on this time scale.

Figure 11 shows the beginning of this process (Arnett 1988a). The horizontal scale is expansion velocity, which is proportional to radius. The top panel indicates the composition; many elements are omitted for clarity. The bottom panel shows the temperature of the matter. Two jumps are indicated. The one at higher velocity is

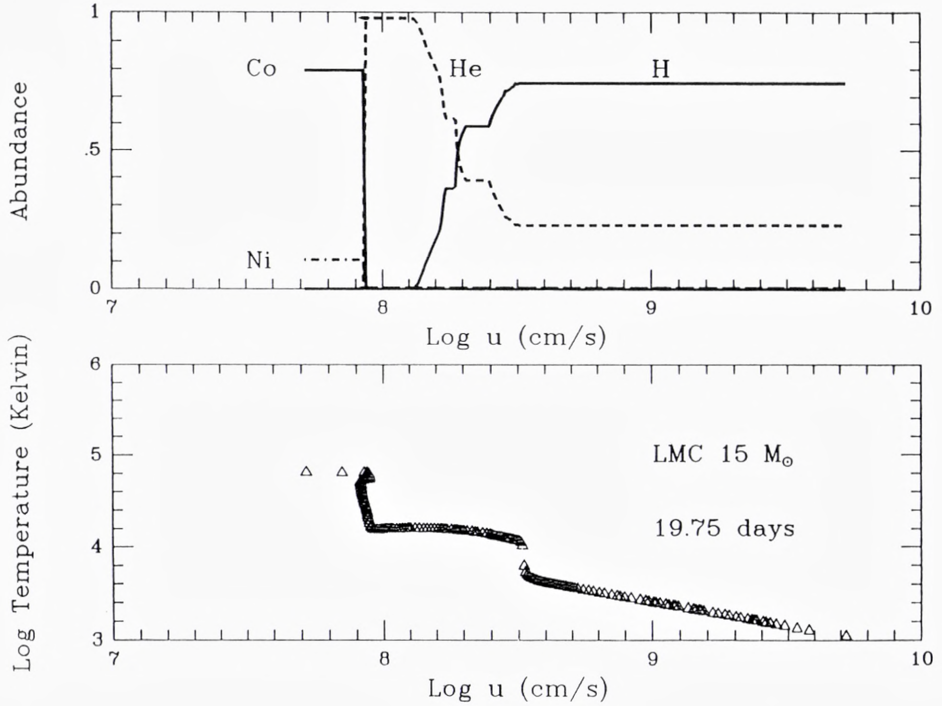

Figure 11. Temperature and Abundance of Selected Elements as a Function of Velocity at 19.75 Days after the Explosion.

due to the recession of the photosphere into the mass of the ejecta; it is a recombination wave. The one at low velocity is due to heating by ^{56}Ni decay; here the matter is Rayleigh-Taylor unstable. A multidimensional calculation is needed to follow the hydrodynamic behavior further. A crude approximation is to define a mixing velocity from the acceleration implied by the Rayleigh-Taylor instability, and microscopically mix the spherically symmetric zones. This underestimates the penetration of ^{56}Ni into overlying regions; this gives mixing up to velocities above 2,000 km/s.

Such penetration by the ^{56}Ni and ^{56}Co has dramatic implications for γ- and x-ray luminosities.

VIII. *The Light Curve at Later Times*

The expansion is homologous except for slow aspherical mixing motions at the interface between the He mantle and the H-rich envelope, and the nickel bubble. This

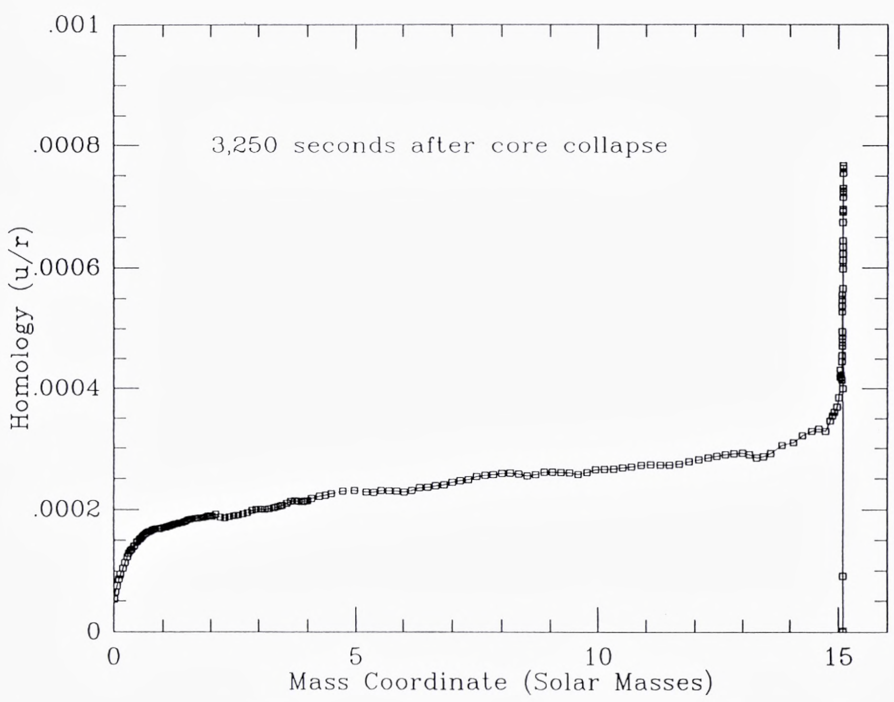

Figure 12. Homology (u/r) versus Mass Coordinate in a Supernova Model after the Shock reaches the Surface.

homology allows an accurate analytic treatment of the supernova light curve (Arnett and Fu 1989). Figure 12 explicity shows the degree of homology found in a numerical computation.

Figure 13 compares the bolometric luminosity from (1) the SAAO data (shown as crosses; Menzies *et al.* 1987, Catchpole *et al.* 1988), (2) a numerical computation (solid line), and (3) an analytic solution (open circles). The numerical solution differs from the analytic in two important ways. First, it correctly deals with the first few weeks of shock related behavior, which are not included in the analytic model (the behavior near time zero in the figure). After this poor start, the analytic solution reproduces the observations well. The numerical calculations do less well, having a jagged behavior. This is mostly due to the unrealistic assumption of strict spherical symmetry. As the recombination wave sweeps in through zones of varying composition, it is not reasonable to ignore nonspherical motions (see above) which would destroy the precise phase coherence which gives rise to the jagged effect. Numerical calculations which attempt to introduce some "mixing" do smooth out this effect.

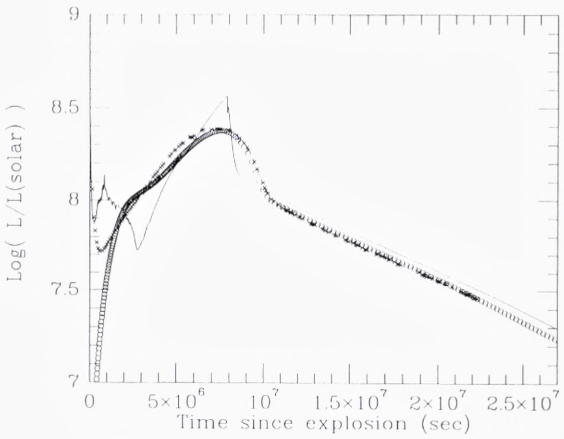

Figure 13. Comparison of Observed, Numerical and Analytic Light Curves.

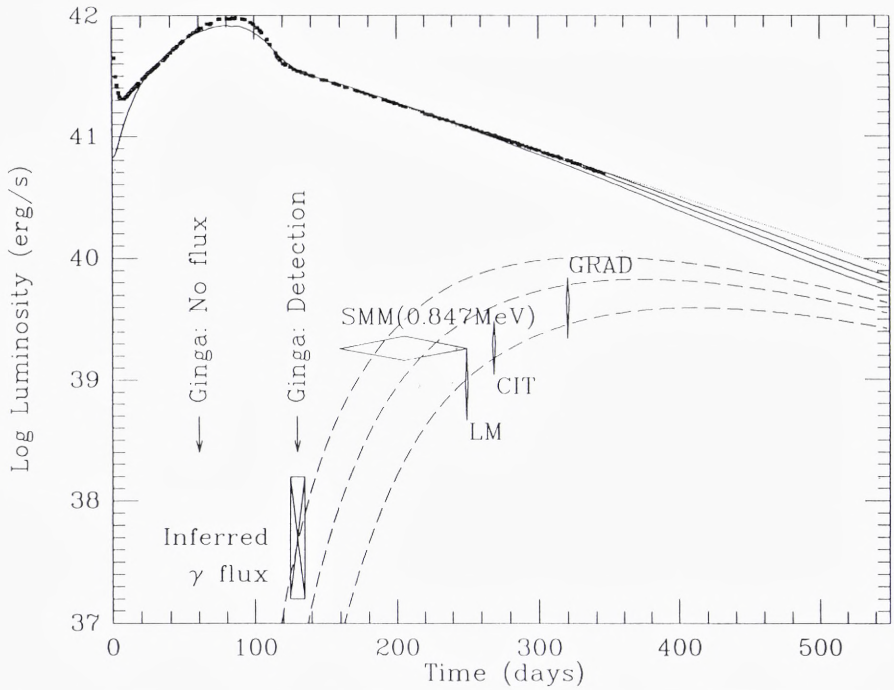

Figure 14. Thermal and Gamma Ray Light Curves for the First 550 Days of SN1987A. SAAO data are represented by solid dots, solid lines are the theoretical curves for infrared, visual and ultraviolet wavelengths combined, and dashed curves are for gamma luminosity. Several gamma line detections are shown.

Such smoothing is a natural feature of the analytic models, which use an average opacity inside the photosphere.

Figure 14 shows the light curve in "thermal" (infrared, visual, ultraviolet) radiation (solid lines) and gamma rays (dashed lines). If the first Ginga detection of *x*-rays is interpreted as due to comptonized gamma rays, the corresponding gamma luminosity can be estimated; this is shown as a crossed box. The luminosities corresponding to several detections of gamma lines are shown as diamonds, of size corresponding to quoted errors. Given the great experimental difficulty, the agreement is startling. As time passes, the easier escape of gammas will cause the "thermal" curve to sag. This has been observed. If there is another source of energy it will cause the curve to decay less steeply, or rise. The accurate exponential decay follows the meanlife of ^{56}Co quite well for the SAAO curve, but less well for the bolometric curve inferred by the CTIO group. This seems to be due to different pass bands for filters, resulting in CTIO missing some lines and therefore giving a lower limit for the luminosity (Menzies 1989). For this reason the SAAO curves are plotted; an agreement on this point by the observers would be welcome.

Figure 15 shows the light curves for still later times. Three possibilities are given,

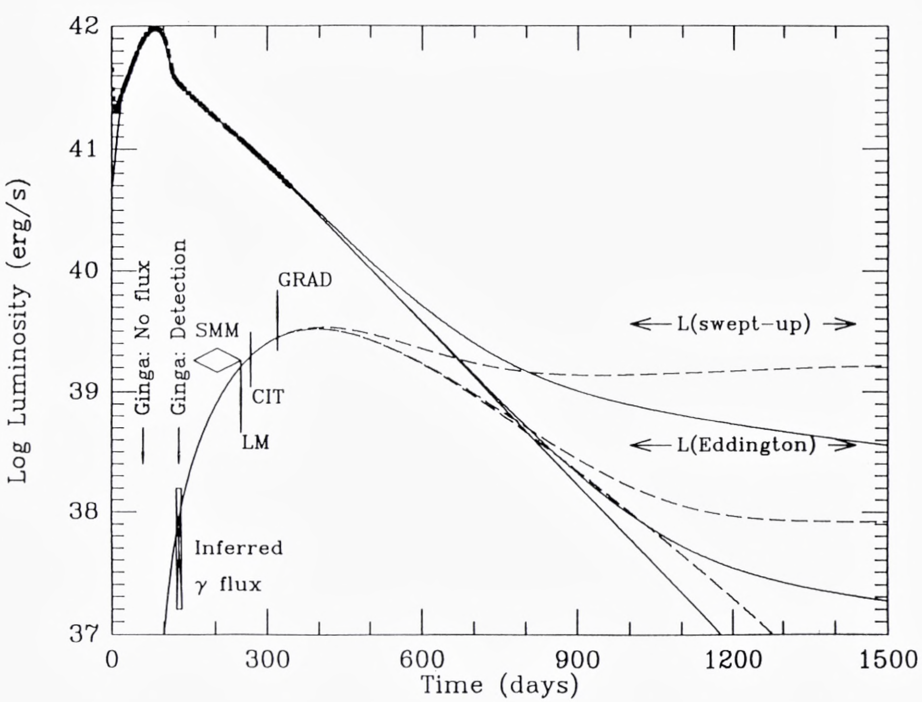

Figure 15. Similar to Figure 14, but for 1500 Days, and showing the Effects of a Possible Pulsar.

corresponding to no pulsar, a pulsar of luminosity of 10^{39} erg/s and one of 2×10^{38} erg/s. This pulsar luminosity is assumed to be emitted in e^+e^- annihilation radiation, which maximizes its chance of escape. At the time of writing, the light curve follows the ^{56}Co decay so well that no pulsar brighter than that in the Crab Nebula could be in SN1987A. It might be that the pulsar is subluminous because it rotates more slowly, or it may not yet have turned on. The neutrino detection has assurred us at least that a neutron star was formed.

Future dramatic events to be expected are the detection of the neutron star in electromagnetic radiation, and of ^{57}Co decay.

IX. *The Uniqueness of SN1987A*

Observationally, SN1987A was unprecedented. To what extent is this event anything more than a freak?

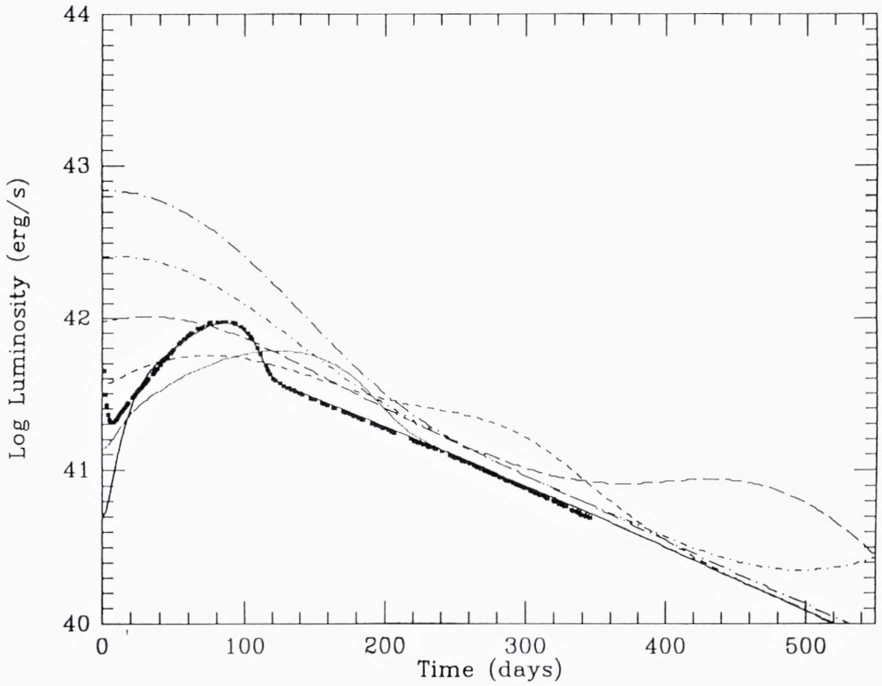

Figure 16. SN1987A and SNII. The SAAO data are shown as solid squares, and the fiducial model (Arnett and Fu 1989) which fits the data as a solid line. The sequence is of models which are identical except for radius, which increases by factors of 2.667.

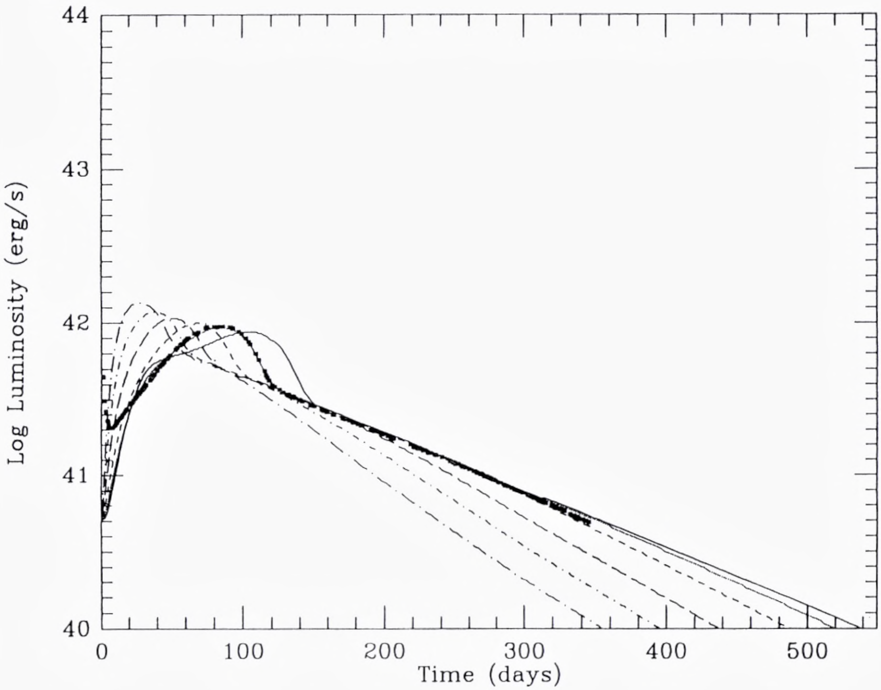

Figure 17. SN1987A and SNI. The SAAO data are shown as solid squares, and the fiducial model (Arnett and Fu 1989) which fits the data as a solid line. The total explosion energy is fixed, so that there is an increase in energy per unit mass, and therefore velocity scale, as the mass decreases.

The only thing atypical about SN1987A is probably the fact that is was exceptionally well observed. Dim supernovae are under-represented in surveys, and would not be well observed. SN1987A is likely to differ from typical Type II supernovae (that is, the ones observed till now) only in that it was a blue supergiant when it exploded, not a red one. This is a phenomenon of the envelope structure, and not of the core of mantle.

Figure 16 illustrates this point; shown are light curves for models identical to that for SN1987A except that they have increasingly larger radii. They form a sequence which extends to behavior appropriate for canonical Type II supernovae.

This is not to say there are not unanswered questions, such as the mystery spot, the disagreement with speckle and theoretical radii, the absence of a pulsar signal, the nature of the observed asphericities, and the yield of nuclei other than $A = 56$, for example. These will undoubtedly teach us more; we have already leaned much.

Further, SN1987A has a simple connection to Type I supernovae. Figure 17 shows light curves for models identical to that for SN1987A except that they have increasingly smaller mass. This gives larger expansion velocities for the fixed explosion energy. The shorter radiative diffusion times begin to show the ^{56}Ni peak, and look like dim SNI. By increasing the mass of ^{56}Ni in these models, they would fit the light curves of Type I supernovae.

SN1987A is a natural member of a theoretical sequence, not a freak. Therefore it provides an empirical determination of the general process by which massive stars die, and of their nucleosynthesis yield. We are extraordinarily lucky to have such a splendid test of a fundamental aspect of astrophysical theory.

References

Alekseev, E. N., Alexeyeva, L. N., Krivosheina, I. V., and Volchenko, V. I. 1988, *Phys. Lett.* **B 205**, 209.

Arnett, W. D. 1977, *Ann. N. Y. Acad. Sci.*, **302**, 90.

Arnett, W. D. 1987a, *ESO Workshop Proceed.*, **26**, 373.

Arnett, W. D. 1987b, *Ap. J.* **319**, 136.

Arnett, W. D. 1987c, *Origin and Evolution of Neutron Stars*, ed. D. J. Helfand and J.-H. Huang, (Dordrecht: D. Reidel), p. 273.

Arnett, W. D. 1988a, *Supernova 1987A in the Large Magellanic Cloud*, (Cambridge: Cambridge University Press), p. 301.

Arnett, W. D. 1988b, *Ap. J.* **331**, 377.

Arnett, W. D., and Fu, A. 1989 *Ap. J.*, **340**, 396.

Arnett, W. D., Bahcall, J. N., Kirshner, R. and Woosley, S. E. 1989, *Ann. Rev. Astron. Ap.*, **27**, 629.

Bahcall, J. N., Piran, T., Press, W. H., and Spergel, D. N. *Nature* **327**, 682.

Bionta, R. M. *et al.* 1987, *Phys, Rev. Letters* **58**, 1494.

Brown, G. E. 1988, *Phys. Reports* **163**, 1.

Burrows, A. and Lattimer, J. M. 1987, *Ap. J.* **318**, L63.

Catchpole, R. M., Whitelock, P. A., Feast, M. W., Menzies, J. W., Glass, I. S., Marang, F., Liang, J. D., Spencer Jones, J. H., Roberts, G., Balona, L. A., Carter, B. S., Laney, C. D., Lloyd Evans, T., Sekiguchi, K., Hutchinson, M. G., Maddison, R., Albinson, J., Evans, A., Allen, D. A., Winkler, H., Fairall, A., Corbally, C., Davies, J. K. and Parker, Q. A. 1988, *Mon. Not. R. astr. Soc.* **229**, 15P.

Chiosi, C. and Summa, C. 1970, *Ap. Space Sci.* **8**, 478.

Hirata, K. *et al.* 1987, *Phys. Rev. Letters* **58**, 1490.

Humphreys, R. M. and McElroy, D. B. 1984, *Ap. J.* **284**, 565.

Lamb, D. Q., Melia, F., and Loredo, T. J. 1988, *Supernova 1987A in the Large Magellanic Cloud*, (Cambridge: Cambridge University Press), p. 204.

Menzies, J. W., Catchpole, R. M., van Vuuren, G., Winkler, H., Laney, C. D., Whitelock, P. A., Cousins, A. W. J., Carter, B. S., Marang, F., Lloyd Evans, T. H. H., Roberts, G., Kilkenny, D., Spencer Jones, J., Sekiguchi, K., Fairall, A. P., and Wolstencroft, R. D. 1987, *Mon. Not. R. astr. Soc.* **227**, 39P.

Menzies, J. W. 1989, preprint.

Nomoto, K., Shigeyama, T., and Hashimoto, M. 1987, *Proc. ESO Workshop on the SN1987A*, ed. I. J. Danziger, (European Southern Observatory: Munich), p. 325.

Nomoto, K., Shigeyama, T., and Hashimoto, M. 1988, *Astron. Ap.*, in press.

Woosley, S. E., Pinto, P. A., and Ensman, L. 1987, *Ap. J.*, **324,** 466.

Hot Gas in Interstellar Space

By Lyman Spitzer Jr
Princeton University Observatory
Peyton Hall
Princeton, NJ 08544, USA

Abstract

Supernova explosions produce shock waves which heat the interstellar gas to temperatures exceeding 10^6K. This hot gas expands, interacts with clouds present between the stars, rises to appreciable distances from the galactic plane and generally affects the structure, dynamics and evolution of the interstellar medium. Recent theories of such processes are briefly reviewed.

I. *Introduction*

It is highly appropriate to include interstellar matter in a series of scientific papers in honour of Bengt Strömgren, whom I remember warmly as mentor, colleague and close friend. His two major theoretical contributions to this subject, made some 40 to 50 years ago, are still of fundamental importance. The differences which he pointed out between HII regions, in which hydrogen atoms are nearly all ionized, and HI zones, where they are nearly all neutral, pervade all our discussions of the interstellar gas. His interpretation of interstellar absorption lines has served as a model for subsequent investigations; his results on the chemical composition of the gas, especially on the overall ratio of calcium to sodium, are still valid qualitatively.

More recent work on interstellar problems has profited from new observational tools. In particular, the existence of a hot gas between the stars has been demonstrated in the last two decades by observations from instruments outside the Earth's atmosphere. To be sure, ground based detection of apparently normal absorbing clouds far from the galactic plane, in the galactic halo, had suggested much earlier the presence of a surrounding hot gas, whose pressure could keep these clouds from expanding. Definite identification of such a hot gas has been obtained not in the halo but in the galactic disc; this result was achieved with two types of observations from space: ultraviolet absorption lines of OVI and other highly ionized species, and soft X rays emitted by hot plasma in extended neighbouring regions of the Galaxy.

The extensive observational material concerning this hot interstellar gas, with a kinetic temperature in the range from 10^5 to 10^7K, has been clearly summarized in recent broad reviews by Cox and Reynolds (1987), Jenkins (1987) and Savage (1987). The present paper makes no attempt to duplicate these summaries, but treats instead some of the theoretical work that has been done, especially during the last few years, on the origin and development of this hot gas.

Understanding the processes which occur as the hot interstellar gas evolves is an ambitious goal which we are far from achieving. The dynamics of a compressible gas, subject to the photons and cosmic rays in interstellar space, is a complex topic. Some progress has been made through the development of idealized models, which are so simplified that one can hope to understand them and to compute their properties. In terms of such models one can distinguish three scenes in the unfolding drama of the hot interstellar gas. First the explosion of a supernova ejects a rapidly expanding envelope, whose interaction with the surrounding medium creates the hot gas in which we are interested. Next, as this heated gas expands it encounters regions whose internal density is well above the average. These regions, which we call clouds, are compressed by the hot gas, are heated by conduction and sometimes evaporate or are disrupted. In the final scene the remnant of heated gas surrounding one or more supernovae can rise to appreciable distances from the galactic plane, and may produce a hot galactic corona before it falls back down or escapes the Galaxy entirely.

In actuality these three scenes overlap so much that their mutual interactions are important. In most theoretical models these scenes have been kept somewhat separate to simplify the theory and to clarify what happens in at least a few highly simplified situations.

During the last few years theorists have constructed a number of such simplified models. To describe the details of all these models would require a substantial monograph. The present paper comments briefly on a few models, indicating the simplifications made and the general character of the results, together with some of the chief problems remaining. After discussions of the three evolutionary scenes listed above, a final section treats the vertical structure of the interstellar medium, again through discussion of simple models. The active dynamical processes in which the hot gas participates, often as the primary driving force, must play a major role in an overall account of the interstellar medium, particularly its structure and evolution.

II. *Expansion of Supernova Remnants*

The phenomena which follow a supernova explosion can in principle be followed in rather full theoretical detail if spherical symmetry is assumed. Such a spherical model is applicable if the stellar explosion itself produces this symmetry, if the initial properties of the surrounding interstellar gas are functions only of r, the distance from the supernova, and if the magnetic field **B** is ignored. In addition one must assume that no non-spherical instabilities will arise. Under these conditions, all quantities are functions of radius r and time t, and the relevant differential equations can be integrated. While this spherical symmetry is not likely to be realized in detail, it may provide an adequate first approximation, especially in those regions where the interstellar gas is resonably homogeneous; an initial particle density independent of r is usually assumed.

Most theoretical models also make the restrictive "hydrodynamic" assumption that the mean free path of all particles is much less than r. This assumption has the great advantage that it yields the familiar equations of fluid dynamics, for which the techniques of numerical solution have been much studied. Physically the hydrodynamic assumption leads to a thin shock wave and to negligible conductive heat flow. In fact the mean free path of protons and electrons for 90° deflections in two-body encounters is many parsecs for a newly born supernova remnant, decreasing to about 1 pc when $T = 10^7$K, if the proton density is $0.1/cm^3$. A magnetic field restricts the travel of charged particles transverse to **B**, but does not yield the hydrodynamic approximation for motions parallel to **B**. Other processes have been suggested which can reduce the effective mean free path. Energetic particles moving through an ionized gas are sometimes slowed down by the plasma instabilities which they excite. This complex effect can perhaps provide full justification for the hydrodynamic assumption, which is also consistent (McKee and Hollenbach 1980) with the relatively sharp boundaries observed for the X-ray emission from some young supernova remnants, notably around the entire circumference of Cas A. However, definite confirmation is lacking.

We ignore initially here both the magnetic field, which is almost certainly present, and the uncertainties associated with the hydrodynamic assumption. The effects of thermal conductivity and of a **B** field are discussed briefly at the end of this section. Effects produced by the relativistic particles constituting cosmic rays have not been much considered in models of supernova remnants and are ignored here; in some circumstances such effects may be highly important. The processes which result when the initial ambient distribution is cloudy or has a vertical density gradient are treated in subsequent sections.

The spherical models based on these assumptions have yielded a substantial body of knowledge on supernova remnants. Most such models assume that there is no important energy source in the supernova core after the explosion; e.g., any radiation from a rapidly rotating neutron star, produced by the collapsing stellar core, is ignored. Three familiar evolutionary stages are then distinguished. First there is free expansion of the ejected material, a stage which lasts as long as the ejected mass is large compared to the mass of the swept-up interstellar gas. Subsequently, the interstellar mass swept up by the outwards moving shock, of radius r_s, much exceeds the ejected mass. As long as the kinetic temperature is high enough throughout the remnant that radiative cooling is slight, this is the well known "Sedov-Taylor stage" (see the monograph by Ostriker and McKee 1988); the density increases rather steeply outwards, with half the mass in the outer six percent of the radius, and the shock velocity V_s varies as $r_s^{-3/2}$, giving $r_s \propto t^{2/5}$.

With increasing r_s, the postshock temperature T_s decreases as V_s^2. When T_s falls below roughly 10^6K the increased rate of radiation then cools the postshock gas to a temperature below 10^4K, and the density increases by a large factor behind the shock, forming a cold shell much thinner than before; this is the third or "snowplow stage." As long as the internal pressure deep within the remnant remains high, the outwards momentum of the cold shell gradually increases, and V_s now varies as $r_s^{-5/2}$ (Cox 1972), giving $r_s \propto t^{2/7}$.

For an actual remnant these stages are approximations. An exact numerical solution of the fluid dynamical equations (Cioffi et al. 1988), including radiative emission in the appropriate energy equation, shows spherical disturbances moving inward and outward, resulting in large part from transitions between successive stages. As a result of these disturbances and the limited duration of each stage, there is only rough agreement with the predictions for the various stages in isolation. The combination of numerical calculations with approximate analytic results gives a reasonably complete understanding of the idealized spherical model, based on the hydrodynamic assumption and an initially uniform interstellar gas density.

In subsequent discussions we shall take as typical parameters for supernova remnants the results obtained in this numerical model, with an ambient particle density of 0.1 atoms/cm^3 and an initial kinetic energy of 0.93×10^{51} ergs in an ejected envelope of mass $3M_\odot$. In this model the swept-up and ejected masses are equal at roughly 10^3 years. The cold shell forms during the interval from 1.2 to 1.7×10^5 years as the shock velocity drops below about 120 km/s; the shock radius r_s is about 55 pc during this interval.

We turn now to the effects which thermal conduction and magnetic fields can produce in supernova remnants. A model which takes into account thermal conduction by electrons as well as energy exchange between ions and electrons shows (Cowie 1977) that after the initial free expansion stage the electron temperature T_e becomes nearly constant with radius inside the remnant, a marked change from the Sedov-Taylor solution. When the electrons are nearly isothermal the positive ion temperature, T_i, assumed to equal T_e immediately behind the shock, is found to increase inward. For the particular case treated the rate of expansion is not much altered by such effects, with r_s at each time increased by not more than 8 percent. These detailed results depend on the assumption that no heat conduction occurs through the moving shock front. More detailed computations (Cox and Edgar 1983 and 1984) show the effects of thermal conduction and of electron-ion energy exchange on various properties of an evolving supernova remnant.

If a magnetic field **B** is present, as seems very likely, thermal conduction transverse to **B** will be almost completely suppressed. Thus T_e will tend to be constant along **B**, but to show a variation of Sedov-Taylor type in planes transverse to **B**. The positive-ion temperature T_i will exceed T_e deep within the remnant, since approach to equipartition through electron-ion encounters is relatively slow (see refs. in preceding paragraph); conduction of heat by positive ions contributes significantly to holding down dT_i/dr along **B**.

The effects which a magnetic field produces on a supernova remnant are more conspicuous if the energy density $B^2/8\pi$ is at least comparable with the material pressure nkT, where n is the total number of particles per cm^3. An equivalent condition is that the Alfvén speed $V_A = B/(4\pi\varrho)^{1/2}$ be at least comparable with the isothermal sound speed $C_s = (kT/m)^{1/2}$, where $m = \varrho/n$ is the mean mass per particle. This situation can arise even if the magnetic field in the gas surrounding the supernova is initially very weak.

One such case for which detailed calculations have been made (Kulsrud et al. 1965) is the hydromagnetic flow around a conducting spherical shell, or "piston," assumed to be expanding at a constant rate into a conducting gas permeated by an initially weak and uniform magnetic field. The lines of force which have been pushed outward by the piston tend to accumulate in a thin boundary layer, where the magnetic field grows steadily until its energy density becomes comparable with that of the streaming gas. In an actual remnant the ionized ejected gases from the supernova may take the place of the expanding piston. The high magnetic pressure in the surrounding boundary layer would then tend to decelerate the inner ejected gases and the Rayleigh-Taylor instability should occur. While the analysis is evidently idealized and other effects will certainly be present, such a process may play a part in producing the filaments of high magnetic field observed in the inner region of the Crab supernova. The shock which moves outward from the piston into the surrounding gas is not much affected by the magnetic field, whose energy density just behind the shock is relatively small.

A uniform interstellar **B** field can produce important effects during the snowplow stage of supernova expansion. With representative parameters for the warm interstellar medium ($n_H = 0.15$ cm^{-3}, $T = 6300K$, $B = 3\mu G$) the Alfvén speed V_A is 14 km/s, substantially greater than the isothermal sound speed C_s of 6.1 km/s. As pointed out above, when the cold radiative shell starts to form, the shock velocity V_s is about 120 km/s; we adopt 40 km/s as a representative value of V_s during the snowplow phase. For these parameters the relative increase of density across an isothermal shock (Spitzer, 1978) is a factor $(V_s/C_s)^2 = 43$ if **B** vanishes or is parallel to $\mathbf{V}_s$, but is only 3.4 (approximately $2^{1/2}$ V_s/V_A) for propagation transverse to the assumed magnetic field. This latter compression is not only weak but nearly reversible, since the energy stored in compressing the magnetic field can drive a reexpansion when the pressure falls. If the high compression "parallel shocks" are assumed to be relatively infrequent and the low-compression transverse ones are regarded as dominant, one may conclude (Cox 1986 and 1988) that the late expansion stages of a supernova remnant produce only a very modest compression of the interstellar medium.

To evaluate the compression expected in an actual interstellar situation one must consider oblique shocks, with a wave normal at some arbitrary angle to the magnetic field. The physical principles governing such shocks are summarized in the accompanying Appendix. It turns out that for the conditions specified above, a single shock, even if parallel to the magnetic field, produces a relatively low compression (a factor 8.2). However, two successive parallel shocks (with an inclined magnetic field between them) can produce the same high compression found above for a single shock with $B = 0$, and constituting effectively the "parallel shock" discussed above. Similar results are presumably possible for two successive shocks within some range of directions relative to **B.** Until such possibilities have been explored, the average compression in the late stage of a supernova shock is essentially unknown, except that it presumably lies between $(V_s/C_s)^2$ and about $2^{1/2}V_s/V_A$.

The amount of this compression has important effects on the structure and dynamics of the interstellar gas. A familiar picture of the interstellar medium, often used as a standard of reference, is based on the sweeping synthesis by McKee and Ostriker (1977), which brings into one theoretical framework many different aspects of this medium. If the compression in an isothermal shock during the snowplow phase were much reduced by magnetic forces, with the compression ratio decreasing from 4 to 1 as V_s decreases from 40 down to about 14 km/s, some aspects of this picture would require modification (Cox 1986). In particular, the warm neutral medium (WNM) would not be swept up into dense shells but would occupy an appreciable part of the remnant's volume, reducing the fraction of this volume occupied by the hot gas. The overall filling factor f_h of the hot gas (the fraction of the volume of the galactic disc occupied by this gas) would be correspondingly reduced. In addition, the evolution of the remnant after expansion ceased would be greatly altered in detail.

In reality strong compression is likely to occur along some lines of the magnetic field. Coupling between motions parallel and transverse to **B** may conceivably convert the enhanced magnetic energy of the compressed WNM into radiation from dense clumps of cooling gas. In any case, magnetic tensions and pressures in the complex interstellar medium are likely to produce unexpected consequences. Computers are reaching the power necessary to follow such hydromagnetic processes approximately. Until more knowledge is available either from analysis or from simulations, theory cannot indicate the hot gas filling factor. The topology of this gas – isolated hot bubbles at one extreme and isolated cooler clouds (either warm or cold) at the other – is equally unclear. Observational information on these questions is also not definite.

III. *Interaction of Remnants with Clouds*

When a star explodes, the expanding remnant may sweep through a gas quite different from the uniform interstellar medium discussed above. It is well known (Spitzer 1985) that the gas between the stars has inhomogeneities with a wide variety of sizes, ranging from filaments less than a parsec across to giant molecular clouds and cloud complexes a hundred parsecs in size. If the exploding star was particularly luminous or its companions had exploded a short time before, the local gas may have been greatly modified and perhaps somewhat homogenized by ultraviolet photons and by expanding hot gases. We do not treat here the details of these complex environments, but consider some of the physical processes which can occur when a supernova remnant engulfs a cloud. Since a full review of such processes has recently appeared (McKee 1988), the present discussion is brief and highly selective.

Strong compression of clouds is believed to be a major effect produced by supernova remnants. A combination of analytic theory and numerical simulation gives an approximate indication of how this "cloud crushing" might proceed. A passing supernova shock generates a slower shock in the denser cloud material. Behind this cloud shock the velocity field leads both to compression and shear, and can produce instabilities which may disrupt the cloud at least in part. An analysis of cloud compression when a magnetic field is present indicates (Oettl et al. 1985), as expected, that if the magnetic pressure is dominant motions transverse to **B** are suppressed. The effect of instabilities may also be less when the compression is one-dimensional, though further study would be needed to establish this conclusion.

Another major effect produced by remnants is cloud evaporation as a result of thermal conduction from the hot gas. The shock itself is not directly involved in this process, which is usually modelled with the hot gas in pressure equilibrium with the cloud and with no systematic velocity of the cloud with respect to the gas. The nature of this process depends on a global saturation parameter σ_0 (Cowie and McKee 1977), essentially the ratio of the electron mean free path to the cloud radius a, and equal to $0.4\ T_{f7}^2/n_{ef}a_{pc}$; T_{f7} is the asymptotic temperature of the hot gas, far from the cloud, in units of 10^7K, n_{ef} is the asymptotic electron density and a_{pc} is the cloud radius in pc. When σ_0 is small compared to unity, the thermal conductivity is given by its classical value, the temperature distribution (in this three-dimensional situation) has a quasi-steady state, and the velocity of the gas in the expanding envelope is subsonic everywhere, However, if σ_0 is too small, less than about 0.03, (corresponding to a_{pc} exceeding about 10 pc in a typical situation) the heat flow is inadequate to offset radiative cooling, and condensation of the hot gas replaces evaporation of the cloud (McKee and Cowie 1977). On the other hand, when σ_0 exceeds unity the physical situation becomes more complicated; the heat flow in this "saturated" condition can be estimated from observations of the solar wind (Cowie and McKee 1977) and the resultant mass loss computed.

A recent investigation (Draine and Giuliani 1984) indicates that for large σ_0 the effect of viscosity, produced by atomic ions, must also be considered. Calculations based on the simplifying assumption that $T_i = T_e$ show that for $\sigma_0 \geqslant 100$ the pressure in the cold cloud can substantially exceed that in the hot gas, and that the viscosity can increase the mass loss rate by about an order of magnitude above the inviscid case; the viscosity requires a pressure increase within the cloud to drive the flow, and at the resultant higher density the heat flow and the resultant mass-loss rate are increased. If the positive ions are heated only by two-body encounters with electrons, T_i will be much less than T_e and these viscous effects will be much reduced.

A further modification in the theory is needed if a magnetic field is present; as pointed out above, there is virtually no conductive flow of heat across an interstellar magnetic field. Limitation of the conductive flux to the direction of **B** has been taken into account (Balbus 1986) for the evaporative flow from an infinite plane surface, initially separating a cold gas on one side from a hot gas extending infinitely far on the other. In this model all variables depend only on t and on z, the distance from the plane; both $B^2/8\pi$ and ϱv^2 are assumed small compared to p. No steady state is possible in this one-dimensional situation (unless a surface at some fixed temperature is located a finite distance away). Instead a self-similar solution is obtained, with a similarity variable proportional to $z/t^{1/2}$; thus as t increases the mass loss rate decreases and the conductive front thickens.

An interesting result of this analysis is that the initial rate of evaporation varies roughly as $\cos^2\Theta_\infty$, where Θ_∞ is the angle between **B** and the z axis at large z. This may be understood physically, since the temperature gradient parallel to **B** equals $\cos\Theta dT/dz$, and the component of the heat flux parallel to z varies as $\cos^2\Theta dT/dz$.

Since the magnetic fields within diffuse clouds (determined from the Zeeman effect of 21-cm lines) are apparently about the same as those in the warm ionized medium (Heiles 1987), a model of straight, parallel lines of force extending into the hot gas may, perhaps, provide a reasonable approximation. On this basis the thickening of the conduction front with time, resulting from the one-dimensional character of the heat flow, should be an important effect, probably more so than the dependence of the flow on Θ_∞.

One attribute of these various conductive evaporating envelopes, with or without magnetic fields, is that they contain highly ionized atoms such as O^{+5}, whose OVI absorption features have been observed along numerous lines of sight through the interstellar gas. At a kinetic electron temperature above 10^5K oxygen atoms will be highly ionized by electron collisions. If collisional ionization equilibrium is assumed in conductive envelopes, the fraction of oxygen in O^{+5} ions is greatest at $T = 3\times10^5$K.

In these envelopes the temperature of each fluid element in the moving gas is changing with time, and hence collisional ionization equilibrium will not be fully reached. The relative numbers of atoms in different stages of ionization must be computed from the relevant differential equations, including the rates of ionization and recombination. Such computations (Ballet et al. 1986) for an outwardly expanding conductive envelope give a substantial increase in the total number of O^{+5} ions in the envelope. We denote by F the ratio of this number to its value in collisional equilibrium; this factor F depends on $n_{ef}a_{pc}$, where again n_{ef} is the asymptotic electron density, far from the cloud, and a_{pc} is the cloud radius in parsecs. Detailed calculations show that for an asymptotic temperature T_f equal to 10^6K, F increases from 2.5 at $n_{ef}a_{pc} = 0.10$ pc/cm^3 to 40 at $n_{ef}a_{pc} = 0.01$ pc/cm^3.

These results, combined with an assumed cosmic composition, indicate that $\langle n(O^{+5})\rangle$, the mean particle density of O^{+5} ions in the galactic disc, is about 2×10^{-7}cm^{-3} for a = 5 pc, $n_{ef} = 0.01$ cm^{-3} and $T_f = 10^6$K; the filling factors f_c and f_h for the cold clouds and the hot gas in the galactic disc are set equal to 0.02 and 1, respectively. If n_{ef} is decreased to 0.001 cm^{-3}, the computed O^{+5} density rises to 5×10^{-7}cm^{-3}. These densities exceed by an order of magnitude or somewhat more the observed mean value of about 2×10^{-8}cm^{-3} (Jenkins 1987).

Later computations for the ionization distribution in expanding conductive envelopes (Böhringer and Hartquist 1987) give similar results. These were used to compute $N(O^{+5})$, the column density of O^{+5} ions, integrated over dr from r = a to r = 10a. The resultant values were found to be clustered around 10^{13}cm^{-2}, with a relatively slow dependence on $n_{ef}a$; as this parameter increases, the increased mass loss rate is offset by a decrease in F. This theoretical column density of 10^{13}cm^{-2} agrees with the observed mean value (Jenkins 1978b) for two-thirds of the O^{+5} gas. (The remaining components have larger column densities and were not included in the value of $\langle n(O^{+5})\rangle$ cited above.) However, the observed values should exceed these theoretical ones by a geometrical factor. For a line of sight through the cloud center the column density is twice the theoretical one. For lines passing tangentially through the shell of highest $n(O^{+5})$ the increase will be somewhat greater; an average increase by a factor two should provide a rough approximation.

While one would not expect such idealized models to correspond closely with reality, it is of interest to note that plausible changes in the assumed parameters can bring theory and observation into rough agreement. It is known (York et al. 1983) that oxygen is depleted in interstellar clouds, with a depletion factor δ_O between 0.4 and 0.7. If we set $\delta_O = 1/2$, this factor compensates for the geometrical factor discussed above, leaving the observed and computed column densities in agreement. If also f_h, the filling factor of the hot gas, is set equal to 0.2, the theoretical value for $\langle n(O^{+5})\rangle$ is then reduced by an order of magnitude and agrees with the observations to within the many uncertainties involved. About this same value of f_h is needed to give the observed average number (Jenkins 1978b) of about six O^{+5} conductive envelopes (each with a radius typically of about 2a) in the line of sight per kiloparsec.

The O^{+5} velocity distribution computed for the conductive model appears to be not inconsistent with the observations. For the thermal velocity spread of these ions within a single envelope the model yields a value between 14 and 18 km/s "in most cases" (Böhringer and Hartquist 1987), corresponding to temperatures between 4×10^5 and 6×10^5K. Because of overlapping components the observed line profiles can give only the minimum values of v_m, the rms velocity dispersion. The envelope expansion velocity can increase v_m appreciably for some lines of sight, but will have little effect on the minimum values, corresponding to lines of sight passing tangentially through the outer layers. Thus one would expect from the theory that for envelopes in general the minimum v_m should be about 14 km/s.

While a number of the observed OVI profiles (Jenkins 1978a) have values of v_m less than 14 km/s, ranging down to 10 km/s, Jenkins (1978b) points out that because of observational errors one can infer only that values as low as 14 km/s are "indeed rather common." More precise observations would be required to determine whether the actual distribution of v_m values is consistent with the conductive envelope theory.

The observed dispersion of radial velocities for the system of OVI absorbing regions is about 26 km/s (Jenkins 1978b), higher than the 6 km/s observed for most interstellar clouds but perhaps consistent with the velocities anticipated for clouds which have recently been compressed and accelerated by a supernova shock wave. Along individual lines of sight the velocity structure of the OVI lines appears correlated with that for lines of the less highly ionized species, SiIII and NII (Cowie et al 1979), supporting a common origin in the same set of conductive envelopes for at least some components of these different lines.

Two additional problems must be considered before the theory of expanding conductive envelopes could be accepted as an approximate quantitative explanation of the OVI observations. The first is the part played by the warm gas (WNM). According to the theory, $\langle n(O^{+5})\rangle$ is proportional to the filling factor of the gas which is evaporating, which for the WNM is at least an order of magnitude greater than for the cold diffuse clouds. Perhaps evaporation of warm gas does not contribute many O^{+5} ions because this low-density gas evaporates so quickly when it is in contact with hot gas. The second problem is the effect of a magnetic field, which as we have seen above is likely to produce a marked thickening of the conductive envelope with time, as a result of the resultant one-dimensional flow. The theoretical calculation of O^{+5} densities should be repeated, taking this effect into account.

Some absorption by O^{+5} ions should also be produced in the inner regions of an idealized spherical supernova remnant, when some of the remnant gas cools through temperatures of some 3×10^5K. Since the initial temperature of the remnant decreases outwards, the outer layers cool first and a radiative cooling front gradually eats its way into the inner hot bubble, surrounded by the cold shell of the snowplow stage. As time goes on, the outwards velocity of the gas which is cooling will become less, and by the time one remnant overlaps with others this velocity may be sufficiently low to be consistent with the OVI observations. If a compressed transverse magnetic field is present, the restoring force on the gas may cause a more abrupt drop in the expansion velocities. The inner layers of hot gas which are cooling down to 3×10^5K and below have been proposed (Cox 1986) as sites for the observed OVI absorption. If effects of initial inhomogeneities and of thermal conduction were included, these cooling layers might have a somewhat similar appearance, though a different course of development, as the conductive envelopes between cold clouds and hot gas discussed above.

The development of a supernova remnant must be strongly affected by the various interactions with clouds, especially by cloud compression and evaporation. A detailed numerical calculation of this dependence was carried out some years ago (Cowie et al 1981), based on the clouds present in the galactic disc generally. Recently a different hydrodynamic technique has been applied (Wolff and Durisen 1987) to this same situation with general overall agreement. In particular, vigorous evaporation from clouds keeps the density nearly independent of r. As a result, the cold shell which is formed by radiative cooling, at a remnant age of typically some 10^5 years, first appears well inside the remnant rather than immediately behind the shock layer.

IV. *Effects far from the Galactic Plane*

The possibility that a hot gas might form a galactic corona at kiloparsec distances from the galactic plane (Spitzer 1956) has given particular interest to the possible role of supernova remnants in this connection, since the hot gases in such remnants provide an obvious heat source. While the remnant from a single stellar explosion can provide heating far from the galactic plane if the supernova is located in the halo or if the ambient particle density is 10^{-2}cm^{-3} or less, more energetic events provide a better source. It has long been clear (Chevalier and Gardner 1974) that adjacent explosions of several supernovae, as might be expected in young stellar groups, would more easily break out of the gaseous galactic disc and pervade the halo. Recent studies have emphasized such sequential explosions of supernovae and the "superbubbles" which they produce in this and other galaxies.

A primary reason for this emphasis has been the accumulating observational evidence (McCray and Snow 1979; Heiles 1979 and 1987a; Tomisaka et al. 1981; McCray and Kafatos 1987) for the existence of superbubbles not only in our own Galaxy but in other Local Group spiral and irregular systems. Extensive arcs and filaments seen in 21-cm emission reveal "supershells" of neutral H. X rays are detected from the hot gas within some of these supershells, and surveys of Balmer emission lines confirm such extended structures. The observed radii range from a hundred to a thousand parsecs; the total energies range up to 10^{53} erg or more, corresponding to about a hundred supernovae.

A second reason for the recent focus on superbubbles is the strong theoretical expectation that such concentrations of supernovae in time and space should in fact exist. Some supernovae (most of the Type I class) originate in old stars, which in the galactic disc show virtually no concentration in groups. However, the Type II supernovae and a few of Type I (Wheeler and Levreault 1985) result from core collapse in young, massive stars, which are apparently formed to a large extent in groups, including clusters and associations. Some of these escape as runaways. According to a recent survey (Gies 1987) about 70 percent of the O stars are now in groups and most of the supernovae produced when these massive stars die will be found within a radius of a few times 10 pc and a time interval of several times 10^7 years (McCray and Kafatos 1987). If several stellar groups are formed within the same cloud complex, the various superbubbles produced may overlap, producing an even more spectacular explosion.

Theoretical models of these superbubbles may be constructed, based on most of the assumptions made for a one-supernova remnant. For a superbubble, however, the disturbance can spread so far from the galactic plane that all physical quantities must be functions of two spatial dimensions, cylindrical radius r and vertical height z. Offsetting somewhat this complication is the initial homogeneity which may characterize the ambient interstellar gas, thanks to the processing of this medium by energetic stellar winds and intense ultraviolet stellar radiation emitted from the massive bright stars before these die and explode.

The nature of blast waves and outgoing winds under these conditions has received much study (see Schiano 1985 for a review). An important aspect of such disturbances in an exponential atmosphere, for example, is that they can lead to "blowout," in which the region of the outgoing shock at the greatest height starts to accelerate, and attains a much increased velocity in the high layers of very low density.

Numerical computations have shown the development of galactic superbubbles produced by consecutive supernovae (Tomisaka and Ikeuchi 1986; Mac Low and McCray 1988). While some details are still controversial, the general outline seems clear. With realistic assumptions for the ambient density as a function of z, blowout does not occur unless the center of the superbubble is about 100 pc or more from the galactic plane. In any case, an energetic superbubble expands rapidly in z and rises far above the galactic plane. In one example (Mac Low and McCray 1988), 80 supernovae occur uniformly during 10^7 years at $z = 0$; the ambient particle density at $z = 0$ equals 1.0 cm^{-3} (0.19 cm^{-3} of warm gas and 0.81 cm^{-3} of cold; I am indebted to R. H. McCray for providing me with the values actually used in the computations) and for large z varies as $0.19 \times \exp(-z/H)$ cm^{-3}, where $H = 500$ pc. After 10^7 years the hot expanding gas has reached $z \approx 500$ pc, as compared with $r \approx 300$ pc reached at $z = 0$. The temperature of the rising gas is of order 10^6K.

In a second example (Tomisaka and Ikeuchi 1986), with 50 supernovae in 10^7 years, again centered at $z = 0$, the total mass of the ambient gas in a column 1 cm^2 in cross-section is less by an order of magnitude; the interstellar particle density for large z equals $0.035 \times \exp(-z/H)$ cm^{-3}, where $H = 250$ pc. In this case the superbubble expands more rapidly, reaching $z \approx 1000$ pc in 10^7 years, as compared with $r \approx 400$ pc reached at $z = 0$.

The actual ambient gas density at high z is probably between the values assumed in these two examples. Hence one may conclude that some hot gas from superbubbles reaches z values of at least 500 pc and probably substantially more. This conclusion is strengthened by calculations of superbubbles centered initially at z = 100 pc. Such computations were made for each of the two examples cited above, and even for the case with the higher ambient density show hot gas rising to some 1500 pc, where blowout begins. In view of the many uncertainties affecting the dymanical calculations and also the overall energy budget of superbubbles (Heiles 1987a), definitive models are not yet to be expected.

If the superbubble gas rising to kiloparsec heights is mostly confined to the Galaxy, the material will recirculate, falling towards the galactic plane as cooler gas, presumably concentrated in clouds formed through thermal instabilities. Such models have been called galactic fountains; the superbubbles, spewing out hot gas at great altitudes, have also been likened to smoking chimneys (Ikeuchi 1987).

The distribution of such falling clouds in velocity and in space when they reach the base of the corona has been computed with an idealized model (Bregman 1980). First the formation of clouds was analyzed, using two-dimensional hydrodynamic calculations for the uprushing hot gas; condensation into cold clouds was assumed whenever the gas temperature fell below 10^4K as a result of adiabatic expansion and radiative cooling. As boundary conditions, in a thin layer at z = 0 and at all times after t = 0, the gas temperature was taken to be about 10^6K, with a particle density of about 10^{-3}cm^{-3}, varying slowly with distance from the galactic center. Since the initial density at high z was negligibly small, the boundary layer generates outward winds, expanding into a vacuum; such models can be regarded as exploding gaseous discs, which approach a quasi-steady state as a result of cloud condensation.

After the clouds formed, these were assumed to move ballistically, independently of the ambient pressure and density, falling back to z = 0. The final distribution of these clouds in velocity and in galactocentric distance is in general agreement with the observed high-velocity clouds, seen mostly in 21-cm radiation. While this model is too idealized to permit definitive conclusions, the general agreement with available 21-cm data on cloud velocities seems impressive. In a related dynamical investigation (Corbelli and Salpeter 1988) the downward cooling flow from superbubbles is analyzed and its effects in extending or compressing the galactic HI disc are discussed.

One may infer from these discussions that hot supernova remnants ejected into the halo, especially from successive explosions within a young stellar group, provide a substantial source of thermal energy to the halo and can possibly maintain the coronal gas at a high temperature. However, the observational evidence for the existence of such a corona at high z is not yet conclusive (see below). While such a corona provides a natural explanation for a large scale height of the halo gas, implied by the observations (Savage 1987; Jenkins 1987), alternative methods of supporting the gas, including magnetic fields and cosmic rays, must also be considered; these topics are treated in the following section.

In an exploration of alternatives to a hot coronal gas, the high state of ionization found for C and Si atoms at high z may be attributed at least in part to photoionization rather than to collisional ionization (Jenkins 1987; Savage 1987). A detailed computation (Bregman and Harrington 1986) has led to the conclusion that these observed C^{+3} and Si^{+3} ions could be produced by photons from hot O-type stars and from the central stars of planetary nebulae. This analysis assumed that 20 percent of the energetic photons from such stars could escape from the Galaxy, either because the radiating stars are at high z, above the absorbing layer of the galactic disc, or because there are gaps in the distribution of neutral H. This escape fraction is highly uncertain. Looking outward from the Sun one finds (for $\delta > -40°$) a minimum HI column density of about 5×10^{19} cm^{-2} (Lockman et al 1986b). For this column density the optical thickness of neutral H for a 48-eV photon, just capable of ionizing C^{+2}, is about 9. If ionizing stellar radiation cannot escape from the galactic disc, photons from active galaxies and quasars can, perhaps, account for the observed C^{+3}, if n_e is less than about 2×10^{-3} cm^{-3} (Fransson and Chevalier 1985).

In any case the N^{+4} ions observed along a few lines of sight at high z are difficult to account for by photoionization. It is primarily for this reason that the presence of hot gas in the halo seems likely.

V. *Structure of the Halo Gas*

The injection of a superbubble into the halo will certainly have dramatic effects on the properties of the local gas, with consequences that are difficult to predict. After such an eruption is over, the local halo will presumably relax to some physical state that may even remain somewhat constant, at least in a statistical sense, until the next great explosion nearby. Thus one may reasonably ask what sort of steady state may be physically possible, subject to what little we know about conditions in the halo. The simplest assumption is that in such a steady state all quantities are functions only of z, the distance from the galactic plane. We discuss here two recent such one-dimensional models of halo gas.

A basic element in any such model is the assumed topography of **B,** the magnetic field. The magnetic pressure p_B, equal to $B^2/8\pi$, makes a major contribution to the pressure in the galactic disc provided that $B_z = 0$. On the other hand, cosmic rays can effectively stream only parallel to the magnetic lines of force, and hence the escape of these energetic particles from the Galaxy is most simply explained if **B** has an appreciable z component. The first model is a hydrostatic equilibrium configuration which is based on the magnetic pressure, with **B** assumed everywhere parallel to the galactic plane; this approach is consistent with the observed direction of **B** in the galactic disc, as determined both from pulsar rotation measures and from the optical polarization of starlight, resulting from alignment of interstellar grains. The second model is a cosmic-ray supported configuration which is based on the escape of cosmic rays, with B_z the only non-vanishing component of **B**; this approach is consistent with the escape of relativistic particles, which seems required by the observed composition of cosmic rays.

The most detailed discussion of models in hydrostatic equilibrium with **B** parallel to the galactic plane includes (Bloemen 1987) all the components of the interstellar gas, cold, warm and hot. A central feature of this discussion is a condition for hydromagnetic stability, particularly against instabilities of Parker type (Parker 1966), in which each line of magnetic force is raised in some regions and depressed in others, with the gas flowing along **B** into the depressed regions. This stability condition (Lachièze-Rey et al. 1980) requires that p_G, the total gas pressure (including the turbulent pressure which is produced by cloud motions), exceed $(dp_{TOT}/dz)/(\gamma d\ln\varrho/dz)$, where p_{TOT} is the sum of all the pressures acting on the gas, – p_G, p_B and the cosmic-ray pressure p_R; as usual, γ is the ratio $\delta \ln p_G/\delta \ln \varrho$ during perturbations of a gaseous element. This criterion, which is derived only for the idealized hydrostatic equilibrium model, may not be a sufficient condition for stability of the model.

One of the significant features included is the large scale height of the warm HI gas, which is set equal to 400 pc in accordance with recent observations (Lockman et al. 1986a). Since the gravitational acceleration increases with z out to at least 1 kpc, the total weight of this component is relatively large, requiring (Cox 1986; Bloemen 1987) a greater value than previously assumed for $p_{TOT}(0)$, the value of p_{TOT} at $z = 0$.

Since $p_G(0)$ and $p_R(0)$ are reasonably well known, this increase in $p_{TOT}(0)$ requires increasing $p_B(0)$, giving an rms B field of about 6μG. While so large a field agrees with a variety of estimates (Bloemen 1987; Heiles 1987b), its consistency with the precise measures of B from pulsar data is a primary requirement. These data give a mean field B_m in the solar neighbourhood which ranges from about 1.6 to 3.5μG depending on the choice of region averaged and on other features of the data analysis (Heiles 1987b). We set B_m equal to 2.5μG. While the rms dispersion of measured fields is less than B_m, the rms random field B_r may considerably exceed B_m if the scale size is substantially less than the pulsar distances.

Such a random field will have dynamical consequences, producing oscillations of the lines of force, together with the attached interstellar clouds. The mean magnetic energy of such oscillations, equal to $B_r^2/8\pi$, should be roughly equal to the mean kinetic energy density $\varrho v^2/2$; if for ϱ we take the smoothed density in cold diffuse clouds (corresponding to $n_H = 0.7 cm^{-3}$) and for v the two-dimensional rms cloud velocity of 8.4 km/s, then $B_r = 4$μG. This value is consistent with the observed dispersion of pulsar measures if the scale size for field variations is in the range from 100 to 200 pc (Thomson and Nelson 1980), depending on the relative importance of fluctuations in n_e and B. The quadratic sum of B_m and B_r then gives about 5μG, not far below the value obtained from $p_{TOT}(0)$. One inference from this discussion is that above about 100 pc, where the density of the cold gas is much reduced and the rms gas velocity is not much greater, B_r^2 may be significantly below its value in the galactic disc.

While the numerical calculations ignore a number of effects (for example, the gradient of $p_B + p_{CR}$ usually assumed to support the cold gas in the galactic disc), the detailed discussion indicates that models which include an increased $p_{TOT}(0)$ together with the hydromagnetic stability criterion promise to fit together various aspects of the interstellar medium at high z. To provide sufficient p_G to satisfy the stability criterion a gas of high pressure but low density apparently suffices. For an assumed halo exponential scale height H_h of 6 kpc the halo temperature needed at high z is about 10^6K, with lower values possible at about a kiloparsec. The value of $n_h(0)$, defined as the atomic particle density of the hot gas extrapolated to $z = 0$, is determined from the added constraint which the observed synchrotron radio emission places on p_{CR} and p_B; it turns out that $n_h(0)$ varies about as H_h^{-2}. If H_h is assumed to equal 6 kpc, a value of 5×10^{-3} provides a "best estimate" for $n_h(0)$.

How such a physical state might be established and maintained raises several physical problems. In particular, how can the temperature of the coronal gas be held in a steady state, with heating balanced by radiative losses? A gas in such radiative equilibrium at $T = 2 \times 10^5$K, for example, is thermally unstable and tends to heat up or cool down. A transient situation, with the gas either heating or cooling through these high temperatures, can provide a more likely explanation for the highly ionized atoms observed in the halo. For example, with plausible assumptions for the flow of initially hot gas through the galactic corona, a calculation of effective recombination rates gives N^{+4} column densities vertically through the corona (Edgar and Chevalier 1986) in good agreement with observed values.

Another problem is how do cosmic rays escape from the Galaxy if **B** is parallel to the galactic plane. Streaming along spiral arms to large distances and then spraying out is conceptually possible but may require diffusion over too great a distance. Diffusion of cosmic-ray particles transverse to **B** can be produced by small-scale magnetic turbulence (Cesarsky 1980). However, this turbulence is probably weak at high z, where $\varrho v^2/2$ is small compared to $B^2/8\pi$; as a result, any turbulent diffusion is likely to be confined to the galactic disc.

We turn to the second theoretical model, in which the escape of cosmic rays is a basic element of the picture, with **B** taken parallel to z. This model (Hartquist and Morfill 1986) takes into account the detailed physical processes involved when energetic ions stream outwards along lines of magnetic force, exciting Alfvén waves. These waves scatter the ions, which drift along **B** by a combination of diffusion and convection. The model calculations assume hydrostatic equilibrium, with dp_R/dz balancing $-g\varrho$, and two energy equations, one for p_R (which includes convection, diffusion and conversion of cosmic-ray energy into wave energy) and one for the wave energy density. In this latter equation the wave damping normally provided by collisions of ions with neutral atoms is ineffective because of low density and high ionization. Instead damping is attributed to interactions between upward travelling and downward travelling Alfvén waves. It is not clear what physical process will produce downward waves of the necessary strength. The stability of such an equilibrium model is also unclear.

Apart from these problems, the model presents a full and self-consistent picture for the particular magnetic topography assumed. Since $p_R(0)$ is only a fraction of $p_{TOT}(0)$, this model applies only to the halo gas, at values of z where the particle density does not exceed about $10^{-3}cm^{-3}$. The energy input from the diffusing cosmic rays into Alfvén waves and into thermal energy via wave-wave damping provides a heat source for the gas. Depending on the parameters assumed, such models would be consistent either with equilibrium kinetic temperatures somewhat below 10^5K and photon ionization as a source of C^{+3}, or with higher temperatures and collisional ionization. Since thermal instability is probable, a transient solution, with the temperature of a fluid element rising or falling, may again be needed to explain the observed CIV and SiIV line strengths.

A time-dependent solution would also be helpful in reconciling the magnetic field perpendicular to the galactic plane, which is suggested by the cosmic ray data, with the parallel field, which is observed in the solar neighbourhood and which seems needed to support the cold and warm components of the interstellar gas. Such a situation is provided, of course, by the occasional eruption of superbubbles. If some fraction of the rising hot gas escapes from the Galaxy, the magnetic lines of force will be stretched far out, with some reconnection perhaps recurring. As pointed out by several researchers (Cesarsky 1980), such a time-dependent sequence might permit the intermittent escape of cosmic-ray particles.

Evidently progress in understanding the hot gas takes place through a succession of idealized models, a familiar state of affairs in much of modern astrophysics. Many of these models have been proposed, some have been developed in considerable detail, and a few have been described here. For the most part they are not very realistic, but they may provide a tentative picture of the hot gas, how it arises and what range of effects it produces. In any case these models are helpful in suggesting new theoretical questions and new observational programs.

Acknowledgements

The preparation of this paper has been greatly aided by comments and suggestions from several astronomers, including J. B. G. M. Bloemen, J. N. Bregman, B. T. Draine, L. L. Cowie, T. W. Hartquist, E. B. Jenkins, R. McCray, C. F. McKee, J. P. Ostriker, B. D. Savage and M. Wardle. Several very detailed discussions with D. P. Cox have been particularly helpful.

Appendix: Oblique Isothermal Shocks

We summarize some relevant information (Landau et al. 1984; Kantrowitz and Petschek 1966) on a plane shock travelling at some oblique angle across a magnetic field. The shock front is stationary and in the preferred reference frame **v,** the local fluid velocity, is parallel to **B**. We denote by θ the angle between **B** and **n**, the unit vector normal to the shock front. Subscripts 1 and 2 denote preshock and postshock quantities; a subscript n denotes a component normal to the shock front.

It is straightforward to write all the shock jump conditions and to solve these for the various postshock quantities. We assume an isothermal shock, with $T_2 = T_1$. The resultant cubic equation for ϱ_2/ϱ_1 gives values which in Figure 1 are plotted against θ_1, the preshock angle between **n** and $\mathbf{B}_1$. The values of V_A (for $B = B_1$) and of C_s taken in the computations are those given above for the warm neutral medium, with $V_s = |v_{1n}|$ set equal to 40 km/s. The values of θ_2 corresponding to various points are indicated in the Figure. The analysis is not strictly applicable for $\theta_1 = 90°$, since for this transverse shock there is no reference frame in which $\mathbf{v}_1$ is parallel to $\mathbf{B}_1$. However, the limiting results as θ_1 approaches 90° agree with those found separately for **n** perpendicular to **B.**

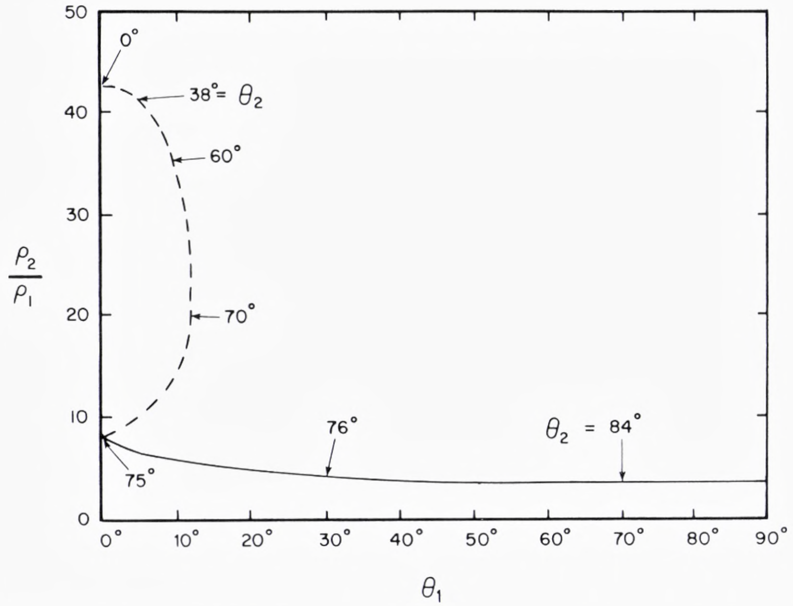

Figure 1. Compression factor across an oblique hydromagnetic shock. The density ratio ϱ_2/ϱ_1, determined from the shock jump conditions, is plotted against θ_1, the preshock angle between the magnetic field and the shock-front normal. Values of the postshock θ_2 are indicated at various points. The dashed curve represents extraneous solutions, not realizable in a single shock.

Figure 1 shows that large compressions are possible only for a narrow range of θ_1 ($\theta_1 \leqslant 12.8°$). The two solutions which at $\theta_1 = 0$ have low compression ($\varrho_2/\varrho_1 = 8.2$) are "switch-on shocks," with $\mathbf{B}_1$ parallel to $\mathbf{n}$, while $\mathbf{B}_2$ is steeply inclined, at an angle θ_2 which here equals 75°. For these shocks the azimuthal angle of $\mathbf{B}_2$ cannot be determined from the shock jump equations but depends on the boundary conditions.

Two additional features must be considered in connection with these results. In the first place, only the lower curve in Figure 1 is physically realizable. The shocks represented by points on the upper dashed line, including all those with $\varrho_2/\varrho_1 > V_s^2/V_A^2 = 8.2$, would, if formed, immediately disintegrate; this disruption is much more rapid than an exponential growth of small perturbations and cannot be regarded as an instability. Such shocks are termed "non-evolutionary" (Landau et al. 1984) or "extraneous" (Kantrowitz and Petschek 1966). The fact that θ_2 is in the opposite direction from θ_1 for these higher-compression solutions of the jump conditions is characteristic of many such extraneous shocks.

In the second place, the end result which would be produced by an extraneous

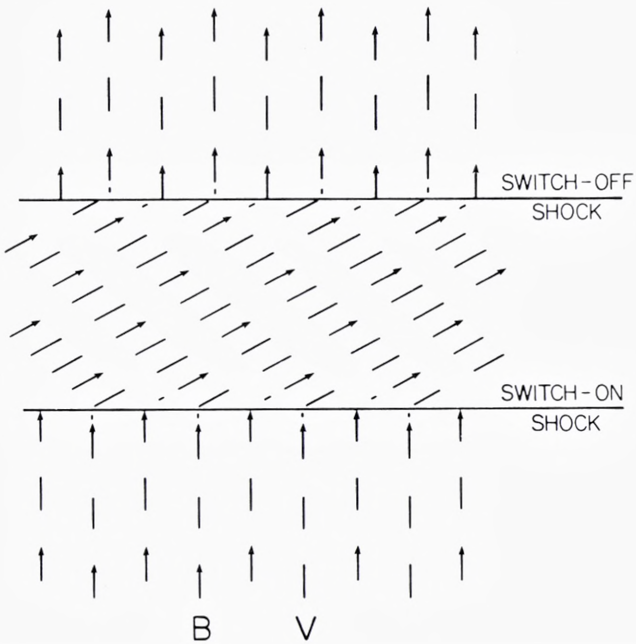

Figure 2. Direction of magnetic field and velocity in a double shock. The leading switch-on shock produces a transverse magnetic field component, which is eliminated in the subsequent switch-off shock. The distance between the two shocks is arbitrary but constant. The overall compression ratio $\varrho_3/\varrho_1 = v_1/v_3$ is that produced by a single shock with the same v_1 but with $B = 0$.

shock, if this could exist, can under some conditions be produced by two successive shocks. This possibility has been pointed out (Kantrowitz and Petschek 1966) in particular for the high-compression shock travelling parallel to $\mathbf{B}_1$; i.e., with $\theta_1 = 0$. If V_A exceeds C_s and if, for an isothermal shock, V_s has any value greater than V_A, this shock is extraneous and cannot exist. However, the switch-on shock which appears instead can be followed by a switch-off shock, which travels at a fixed separation from the first shock and which bends the magnetic field back to its original direction, normal to the two shock fronts. The resultant geometry is indicated in Figure 2. The second shock produces an additional compression by a factor 5.2 for the parameters assumed here, giving an overall compression factor of 43, the value found above for $B = 0$. What is possible for $\theta_1 = 0$ should also be possible for some other values of θ_1, especially if the two shocks are allowed to separate gradually and if the final direction of **B** is allowed to differ somewhat from the original direction. Further analysis is needed to indicate under what conditions a supernova remnant in the snowplow stage can strongly compress the warm interstellar medium.

References

Balbus, S. A 1986, *Astrophys. J.*, **304**, 787.

Ballet, J., Arnaud, M. and Rothenflug, R. 1986, *Astron. Astrophys.*, **161**, 12.

Bloemen, J. B. G. M. 1987, *Astrophys. J.*, **322,** 694.

Böhringer, H. and Hartquist, T. W. 1987, *Mon. Not. R. astr. Soc.* **228**, 915.

Bregman, J. N. 1980, *Astrophys. J.*, **236**, 577.

Bregman, J. N. and Harrington, J. P. 1986, *Astrophys. J.*, **309**, 833.

Cesarsky, C. J. 1980, *Ann. Rev. Astron. Astrophys.*, **18**, 289.

Chevalier, R. A. and Gardner, J. 1974, *Astrophys. J.*, **192**, 457.

Cioffi, D. F., McKee, C. F. and Bertschinger, E. 1988, *Astrophys. J.* **334**, 252.

Corbelli, E. and Salpeter, E. E. 1988, *Astrophys. J.*, **326**, 551.

Cowie, L. L. 1977, *Astrophys. J.*, **215**, 226.

Cowie, L. L., Jenkins, E. B., Songaila A. and York, D. G. 1979, *Astrophys. J.*, **232**, 467.

Cowie, L. L. and McKee, C. F. 1977, *Astrophys. J.*, **211,** 135.

Cowie, L. L., McKee, C. F. and Ostriker, J. P. 1981, *Astrophys. J.*, **247**, 908.

Cox, D. P. 1972, *Astrophys. J.*, **178**, 159.

Cox, D. P. 1986, in *Workshop on Model Nebulae*, Observ. de Meudon; ed: D. Pequinot (Paris: Observ.) p. 11.

Cox, D. P. 1988, in IAU Colloq. 101, *Supernova Remnants and the Interstellar Medium*, eds: T. L. Landecker and R. S. Roger (Cambridge: Cambridge Univ. Press) p. 73.

Cox, D. P. and Edgar, R. J. 1983, *Astrophys. J.*, **265**, 443.

Cox, D. P. and Edgar, R. J. 1984, *Astrophys. J.*, **283**, 833.

Cox, D. P. and Reynolds, R. J. 1987, *Ann. Rev. Astron. Astrophys.*, **25**, 303.

Draine, B. T. and Giuliani, J. L. 1984, *Astrophys. J.*, **281**, 690.

Edgar, R. J. and Chevalier, R. A. 1986, *Astrophys. J. Letters*, **310**, L27.

Fransson, C. and Chevalier, R. A. 1985, *Astrophys. J.*, **296**, 35.

Gies, D. R. 1987, *Astrophys. J. Suppl.*, **64**, 545.

Hartquist, T. W. and Morfill, G. E. 1986, *Astrophys. J.*, **311**, 518.

Heiles, C. 1979, *Astrophys. J.*, **229**, 533.

Heiles, C. 1987a, *Astrophys. J.*, **315**, 555.

Heiles, C. 1987b, in *Interstellar Processes,* eds.: D.J. Hollenbach and H.A. Thronson, Jr. (Dordrecht: Reidel) p. 171.

Ikeuchi, S. 1987, in *Starbursts and Galaxy Evolution,* Moriond Astrophys. meeting, eds.: T.X. Thuan, T. Montmerle and J.T. Thanh Van (Paris: Editions Frontières) p. 27.

Jenkins, E.B. 1978a, *Astrophys. J.*, **219**, 845.

Jenkins, E.B. 1978b, *Astrophys. J.*, **220**, 107.

Jenkins, E.B. 1987, in *Exploring the Universe with the IUE Satellite,* ed: Y. Kondo (Dordrecht: Reidel) p. 531.

Kantrowitz, A. and Petschek, H.E. 1966, in *Plasma Physics in Theory and Application,* ed.: W.B. Kunkel (New York: McGraw-Hill) p. 148.

Kulsrud, R.M., Bernstein, I.B., Kruskal, M., Fanucci, J. and Ness, N. 1965, *Astrophys. J.*, **142**, 491.

Lachièze-Rey, M., Asséo, E., Cesarsky, C.J. and Pellat, R. 1980, *Astrophys. J.*, **238,** 175.

Landau, L.D., Lifschitz, E.M. and Pitaevskii, L.P. 1984, *Electrodynamics of Continuous Media,* 2nd edit. (New York: Pergamon) § 72 and 73.

Lockman, F.J., Hobbs, L.M. and Shull, J.M. 1986a, *Astrophys. J.*, **301**, 380.

Lockman, F.J., Jahoda, K. and McCammon, D. 1986b, *Astrophys. J.*, **302**, 432.

Mac Low, M.-M. and McCray, R. 1988, *Astrophys. J.*, **324**, 776.

McCray, R. and Kafatos, M. 1987, *Astrophys. J.*, **317**, 190.

McCray, R. and Snow, T.P. 1979, *Ann. Rev. Astron. Astrophys.*, **17**, 234.

McKee, C.F. 1988, IAU Colloq. 101, *Supernova Remnants and the Interstellar Medium,* eds.: T.L. Landecker and R.S. Roger (Cambridge: Cambridge Univ. Press), p. 205.

McKee, C.F. and Cowie, L.L. 1977, *Astrophys. J.*, **215**, 213.

McKee, C.F. and Hollenbach, D.J. 1980, *Ann. Rev. Astron. Astrophys.*, **18**, 219.

McKee, C.F. and Ostriker, J.P. 1977, *Astrophys. J.*, **218**, 148.

Oettl, R., Hillebrandt, W. and Miller, E. 1985, *Astron. Astrophys.*, **151**, 33.

Ostriker, J.P. and McKee, C.F. 1988, *Rev. Mod. Phys.*, **60**, 1.

Parker, E.N. 1966, *Astrophys. J.*, **145**, 811.

Savage, B.D. 1987, in *Interstellar Processes,* eds.: D.J. Hollenbach and H.A. Thronson, Jr., (Dordrecht: Reidel) p. 123.

Schiano, A.V.R. 1985, *Astrophys. J.*, **299**, 24.

Spitzer, L. 1956, *Astrophys. J.*, **124**, 20.

Spitzer, L. 1978, *Physical Processes in the Interstellar Medium* (New York: Wiley) Chapt. 10.

Spitzer, L. 1985, *Physica Scripta,* **T11,** 5.

Thomson, R.C. and Nelson, A.H. 1980, *Mon. Not, R. astr. Soc.*, **191**, 863.

Tomisaka, K., Habe, A. and Ikeuchi, S. 1981, *Astrophys. Space Sci.*, **78**, 273.

Tomisaka, K. and Ikeuchi, S. 1986, *Publ. Astron. Soc. Japan,* **38**, 697.

Wheeler, J.C. and Levreault, R. 1985, *Astrophys. J. Letters,* **294**, L17.

Wolff, M.T. and Durisen, R.H. 1987, *Mon. Not. R. astr. Soc.*, **224**, 701.

York, D.G., Spitzer, L., Bohlin, R.C., Hill, J., Jenkins, E.B., Savage, B.D. and Snow, T.P. 1983, *Astrophys. J. Letters,* **266**, L55.

The Dense Interstellar Medium and the Birthplaces of OB Stars

By Frank H. Shu
Astronomy Department, University of California, Berkeley, CA 94720, USA

Abstract

We review Bengt Strömgren's research interests in the relationships among ionized, atomic, and molecular hydrogen in the Galaxy and the spatial distribution and kinematics of OB stars. Our current perception of these problems owes much to Strömgren's deep influence and early insight.

I. *Introduction*

In 1939, Bengt Strömgren wrote perhaps his most celebrated paper, introducing the seminal notion that ionizing radiation from luminous OB stars sharply partitions the regions of interstellar hydrogen into two forms, H I and H II. As a measure of the care Strömgren invested in this landmark work, I should mention that he briefly considered the possibility (suggested by Eddington) that the hydrogen might also exist in molecular form (Strömgren 1939, p. 541). Strömgren rejected the idea on the grounds that if molecular hydrogen were created and destroyed by processes similar to those he had analyzed for atomic hydrogen, then negligible amounts of H_2 would be found in interstellar space. Today, we consider this last conclusion wrong – molecular clouds *do* indeed constitute the third great resevoir for interstellar gas – but only because H_2 is formed and destroyed by processes entirely different than those known to astronomers fifty years ago.

Although the parallel idea of the formation of OB stars from the neutral gas which they subsequently ionize must have occurred to Strömgren, he refrained from direcly speculating on the exact connection. It was only after the development of his narrow-band photometric system (see the review of Strömgren 1966), when he had in hand a quantitative tool to measure the ages of the B stars, that he turned again the full force of his attention to this problem. In 1961, he organized a meeting at the Institute for Advanced Study in Princeton on the distribution and motion of interstellar matter (Woltjer 1962). C. C. Lin, a hydrodynamicist on sabbatical leave from M.I.T. at the Institute for the year, attended the conference at Strömgren's invitation. As part of the program, Per Olof Lindblad gave a description of some computer simulations which supported the ideas of Bertil Lindblad concerning spiral structure in disk galaxies. This work, plus the succinct description given by Jan Oort of the problem, caught Lin's imagination, and from that interest grew the modern theory of spiral density waves.

Almost from the start, Strömgren (1967) saw to the heart of the central thesis of density-wave theory, the implication that the pattern of the formation of OB stars in a spiral galaxy should rotate at a constant angular speed Ω_p, which differs generally from the mean material speed $\Omega(r)$ of the interstellar gas. If one combines measured space motions of B stars with their ages deduced from narrow-band photometery, one could integrate the equations of motion backward in time (with some adopted Galactic potential) to find the birthplaces of these B stars. On the other hand, one could also rotate the present-day spiral pattern backwards in time at a constant angular rate Ω_p to deduce the theoretical location of spiral arms at the time of the births of the corresponding stars. If the birthplaces of B stars deduced by the star-migration calculation agreed with the location of spiral arms determined by the modal picture, the entire undertaking would gain credence.

The initial results proved ambiguous if one only took into account the kinematical aspects of the problem. Yuan (see Lin, Yuan, and Shu 1969) found better fits by including the focusing of stellar orbits by the perturbational gravitational field associated with the spiral arms. Within the uncertainties of the age and space-motion determinations, one could then force all of the then-known sample stars to have birthplaces inside spiral arms with reasonable choices for the parameters of the Galactic density-wave pattern. Later study by Yuan and Grosbøl (1981) extended these results to assess the expected color variations and surface brightnesses across spiral arms, but definitive conclusions belong to the future and will probably come only with completion of Preben Grosbøl's ongoing program to obtain the space motions and ages of a very large number of B stars.

II. *Triggered Star Formation*

In the interim, theorists began to examine in detail how spiral structure could help to trigger OB star formation. Motivated by the calculations of Fujimoto (1968), Roberts (1969) championed the idea of a central role for *galactic shocks*, regions of high compression (identified with dust lanes) that follow the supersonic flow of interstellar gas into the spiral potential minima defined by the disk stars. In a cloudy model for the interstellar medium, however, the increase in average gas density arises merely by the bringing of the centers of individual clouds closer together. Why should such a process by itself lead to enhanced star formation? Shu et al. (1972) adopted the two-phase models of Field, Goldsmith, and Habing (1969) and Spitzer and Scott (1969) to compute that a true hydrodynamic shock in a warm intercloud medium could cause the implosion of normal H I clouds, leading possibly to gravitational collapse and star formation (see also Woodward 1976). Beneath the train of thought during this period lay the assumption that star formation proceeds from atomic hydrogen clouds, which, because they lack appreciable self-gravitation to begin with, had to be imploded by an increase of the external pressure load in order to give star formation.

Two developments in the 1970s served to undermine this point of view. First, ultraviolet and x-ray observations led to the picture (Cox and Smith 1974, McKee and Ostriker 1977) that much of the volume of interstellar space might be filled with gas too hot to contain continuum galactic shocks of the type envisaged by density-wave theorists. Discrete clouds could still respond in a sufficiently nonlinear manner as to yield a "shocklike" distribution, but the pressure of the hot intercloud medium would not suffer a sharp upward jump inside a dust lane. Second, molecular-line studies (see, e.g., the review of Burton 1976), particularly in the J = 1-0 transition of CO, demonstrated that OB star formation occurred almost exclusively in giant molecular clouds (GMCs). Unlike H I clouds prevented from free expansion by the surface pressure of an external medium, GMCs represent gravitationally bound objects. Thus, GMCs do not need external triggering to undergo star formation *spontaneously*; indeed, the theoretical difficulty lies, as we shall se in the next section, entirely in the opposite direction. Nevertheless, the concept that star formation needs to be *induced* became a fixed idea in the thinking of many astrophysicists, and it led to a torrent of ingenious new proposals long after the original motivation (the birth of stars from non-self-gravitating H I clouds) had vanished.

III. *The Mechanical Support of Molecular Clouds*

Zuckerman and Evans (1974) first gave voice to the crucial issue. The actual mass of a typical GMC much exceeds its Jeans mass; thus, unless agents of support other than thermal pressure exist, a GMC should collapse gravitationally on the order of a free-fall time to convert its entire mass into stars. Such a scenario would yield star formation rates in the Galaxy two to three orders of magnitude larger than the observed value. The central dynamical problem with GMCs consists, therefore, not of how to induce them to form stars, but of how to prevent them from doing it as fast as natural processes would seemingly dictate.

When viewed in this light, the problem of star formation in molecular clouds acquires a completely different perspective. Whatever constitutes the agent of molecular cloud support, it must (1) be difficult to dissipate (or else, star formation would proceed much more efficiently than it actually does), and (2) be present at a dynamically significant level over the wide range of scales in which molecular cloud structures are observed to be self-gravitating. Given these two criteria, of the three mechanisms conventionally invoked to supplement thermal pressure in cloud support – rotation, turbulence, and magnetic fields – we can now single out magnetic fields for special consideration.

If a cloud has internal temperature T and is subject to an external pressure P_{ext}, we may use the results of Mouschovias and Spitzer (1976) to show that the maximum mass M_{cr} that a cloud can support against its own gravity satisfies the equation,

$$M_{\rm cr}\left[1-(M_\Phi/M_{\rm cr})^2\right]^{3/2}=M_{\rm BE}, \tag{1}$$

where M_{BE} is the Bonnor-Ebert mass

$$M_{\rm BE}\equiv 1.4\,\frac{(kT/m)^2}{G^{3/2}P_{\rm ext}^{1/2}}. \tag{2}$$

We have followed Mouschovias and Spitzer in modifying the numerical coefficient in eq. (2) from that given by Ebert (1955) and Bonnor (1956) to provide better overall agreement with the nonspherical models of Mouschovias (1976). In equation (1), we have defined the magnetic critical mass M_Φ as

$$M_\Phi = 0.13\,\frac{\Phi}{G^{1/2}}, \tag{3}$$

with Φ equalling the total magnetic flux that threads the electrically conducting cloud. In the field freezing approximation, Φ is a conserved quality in the mechanical evolution of an isolated cloud.

Gravitational collapse occurs for $M > M_{cr}$; whereas the cloud is stable if its mass $M < M_{cr}$. If we can ignore magnetic effects, i.e., if we can set $M_\Phi = 0$, we would get the critical mass from equation (1) as M_{BE}. The typical parameter regime that applies to actual molecular clouds corresponds to $M_\Phi \approx 10^5\ M_\odot$ for $B = 30\ \mu G$ and $R = 20$ pc, whereas $M_{BE} \approx 6\ M_\odot$ for $T = 10$ K and $P_{ext}/k = 10^4\ cm^{-3}$ K. As a consequence, equation (1) has the approximate solution,

$$M_{\rm cr}\approx M_\Phi\left[1+\frac{1}{2}\left(\frac{M_{\rm BE}}{M_\Phi}\right)^{2/3}\right]\approx M_\Phi, \tag{4}$$

when $M_\Phi >> M_{BE}$. Except for small dense cores, thermal support plays a relatively minor role in comparison with magnetic (and, possibly, turbulent) support in molecular clouds. Moreover, provided cloud masses do not rise with increasing size faster than $M \propto R^2$ (e.g., as indicated by the observations of Solomon et al. 1987), support by magnetic fields of a given strength B can keep pace with gravity at all scales since $M_\Phi \propto \Phi \propto BR^2$ also increases as R^2 for fixed B. (The characteristic mass associated with the "turbulence" observed in molecular clouds also has this property and may have an explanation in MHD fluctuations being self-regulated at some fixed fraction of the Alfvén speed; see, e.g., Lizano and Shu 1989). These conclusions have, as we shall see below, some dramatic consequences for the process of stimulated star formation.

We suppose that at the present epoch of galactic evolution almost all clouds have subcritical masses, $M < M_{cr}$, since initially supercritical clouds must have collapsed a long time ago. An increase in the external pressure P_{ext} – due, e.g., to the passage of a galactic or supernova shock – can lower M_{BE} but not M_Φ (if field freezing holds); thus, external inducement of gravitational collapse and star formation cannot occur unless M happens to satisfy: $M_\Phi < M < M_{cr}$, which constitutes an extremely narrow range (cf. eq. [4]) if $M_{BE} \ll M_\Phi$. In this circumstance, the vast majority of molecular clouds must be magnitically subcritical, $M < M_\Phi$, and no amount of external compression can induce a cloud (under the constraint of field freezing) to collapse.

In a sense, magnetic fields play the same role for our theory for molecular clouds that electron degeneracy pressure does for the theory of white dwarfs. In this analogy, M_Φ replaces the Chandrasekhar limiting mass M_{Ch}. No increase of external pressure can cause a magnetically supported cloud, with $M < M_\Phi$, to collapse, any more than it can cause a white dwarf, with $M < M_{Ch}$, to do so, because the restoring force due to the compressed magnetic fields (or degenerate electrons) rises in direct proportion to the increase in self-gravitation (for a "gas" whose internal "pressure" is proportional to the 4/3 power of the density in three-dimensional compression). Clearly, magnetic fields provide such a formidable obstacle to gravitational collapse that we should no longer be surprised, to zeroth order, that star formation in magnetized molecular clouds proves to be generally a very inefficient process.

IV. *The Relationships among H_2, H I, and H II*

If we accept the arguments of the previous section, we see that there exists logically only two ways for a magnetized molecular cloud to become unstable to gravitational collapse (see also the reviews of Mestel 1985; Shu, Adams, and Lizano 1987). To increase the ratio, M/M_Φ, either we can lower Φ (and therefore M_Φ) at fixed M, or we can increase M relative to M_Φ (i.e., increase the mass-to-flux ratio M/Φ). Because molecular clouds are only lightly ionized (Elmegreen 1979), the first process will occur inexorably via ambipolar diffusion in small dense cores. Current belief holds this to occur for the mode of low-mass star formation. Because GMCs appear to be aggregates of smaller cloud clumps (Blitz 1987), the second process can occur through the agglomeration of clumps, which can enhance the mass to flux ratio M/Φ (if the agglomeration occurs parallel to field lines) and lead to overall gravitational collapse of a relatively large piece of a GMC. Some evidence exists to indicate that this may constitute the mode of high-mass star formation.

The interesting question for the line of research originated by Strömgren then becomes: Does the formation of GMCs and subsequent clump agglomeration due to orbit focusing in spiral arms occur through the collection of small H I clouds or small H_2 clouds?

On this question astronomical opinion remains divided. Part of the difficulty lies in the demonstration by Kaufmann et al. (1987) that the details of the cross-arm distribution of H II regions and other spiral tracers in galaxies like M81 cannot be reproduced by single-component fluid models of the interstellar medium. In any case, the traditional answer, an outgrowth of the ideas outlined in § II, holds in favor of the following pathway: Atomic clouds give rise to molecular clouds; molecular clouds yield luminous OB stars, which then produce the large photoionized regions that show up so dramatically in optical photographs of spiral galaxies. The traditional picture then predicts the following sequence as we follow the gas flow into and out of a galactic shock: well-separated H I clouds → gathered H I superclouds (dust lanes) → H_2 clouds → OB stars → H II regions. More recently, detailed studies of gas-rich galaxies like M83 and M51 (Allen, Atherton, and Tilanus 1986; Lo et al. 1987; Tilanus et al. 1988; Vogel et al. 1988) indicate that the dust lanes are composed of GMCs, and that the atomic gas lies downstream from the dust lanes, well mixed (on a large scale) with the ionized gas. These observations suggest that the H I and H II arise by photodissociation and photoionization of the giant molecular clouds giving birth to young OB stars. In this scheme, we have the sequence; small molecular clouds → GMCs (dust lanes) → OB stars → H I and H II regions.

To complicate the situation, investigations of gas-poor galaxies like M31 (e.g., Lada et al. 1988) suggest that the original scheme involving the formation of H_2 clouds from H I clouds may be more appropriate after all! In the final analysis, it may turn out that both pictures have validity; GMCs may be assembled primarily from pre-existing dwarf molecular clouds in gas-rich galaxies, and from pre-existing H I clouds in gas-poor galaxies. It may even turn out that the inner and outer regions of the *same* galaxy may rely on different formation mechanisms for giant molecular clouds.

I think Bengt Strömgren would have been pleased that the questions he first asked concerning the spatial relationships among ionized, atomic, and molecular hydrogen in the Galaxy and their associated OB stars have had such a bountiful set of implications. It forms a true tribute to his insight and inspiration that these problems continue to excite and fascinate the current generation of observational and theoretical astronomers.

This work has been supported in part by the National Science Foundation through grant No. AST86-14743 and in part under the auspices of a special NASA astrophysics theory program which funds a joint Center for Star Formation Studies at NASA Ames Research Center, UC Berkeley, and UC Santa Cruz.

References

Allen, R. J., Atherton, P. D., and Tilanus, R. P. 1986, *Nature,* **319,** 296.

Blitz, L. 1987, in *Physical Processes in Interstellar Clouds,* ed. G. Morfil and M. Scholer (Dordrecht: Reidel), p. 35.

Bonnor, W. B. 1956, *M.N.R.A.S.,* **116,** 351.

Burton, W. B. 1976, *Ann. Rev. Astr. Ap.,* **14,** 275.

Cox, D. P., and Smith, B. W. 1974, *Ap. J. (Letters),* **189,** L105.

Ebert, R. 1955, *Zs. Ap.,* **37,** 217.

Elmegreen, B. G. 1979, *Ap. J.,* **232,** 729.

Field, G. B., Goldsmith, D. W., and Habing, H. J. 1969, *Ap. J. (Letters),* **155,** L149.

Fujimoto, M. 1988, in *Proc. IAU Symp. No. 38,* p. 368.

Kaufman, M., Bash, F. N., Kennicutt, R. C., and Hodge, P. W. 1987, *Ap. J.,* **319,** 61.

Lada, C. J., Margulis, M., Sofue, Y., Nakai, N., and Honda, T. 1988, *Ap. J.,* **328,** 143.

Lin, C. C., Yuan, C., and Shu F. H. 1969, Ap. J., **155,** 721.

Lizano, S., and Shu, F. H. 1989, *Ap. J.,* **342,** 834.

Lo, K. Y., Ball, R., Masson, C. R., Phillips, T. G., Scott, S., and Woody, D. P. 1987, *Ap. J. (Letters),* **317,** L63.

McKee, C. F., and Ostriker, J. P. 1977, *Ap. J.,* **218,** 148.

Mestel, L. 1985, in *Protostars and Planets II.,* ed. D. C. Black and M. S. Matthews (Tucson: University of Arizona Press), p. 320.

Mouschovias, T. Ch. 1976, *Ap. J.,* **207,** 141.

Mouschovias, T. Ch., and Spitzer, L. 1976, *Ap. J.,* **210,** 326.

Roberts, W. W. 1969, *Ap. J.,* **158,** 123.

Shu, F. H., Adams, F. C., and Lizano, S. 1987, *Ann. Rev. Astr. Ap.,* **25,** 23.

Shu, F. H., Milione, V., Gebel, W., Yuan, C., Coldsmith, D. W., and Roberts, W. W. 1972, *Ap. J.,* **173,** 557.

Solomon, P. M., Rivolo, A. R., Barret, J. W., and Yahil, A. 1987, *Ap. J.,* **319,** 730.

Spitzer, L., and Scott, E. H. 1969, *Ap. J.,* **158,** 161.

Strömgren, B. 1939, *Ap. J.,* **89,** 526.

Strömgren, B. 1966, *Ann. Rev. Astr. Ap.,* **4,** 433.

Strömgren, B. 1967, in *Radio Astronomy and the Galactic System, Proc. IAU Symp. No. 31,* ed. H. van Woerden (London: Academic Press), p. 323.

Tilanus, R. P., Allen, R. J., van der Hulst, J. M., Crane, P. C., and Kennicutt, R. C. 1988, *Ap. J.,* **330,** 667.

Vogel, S. N., Kulkarni, S. R., and Scoville, N. Z. 1988, *Nature,* **334,** 402.

Woltjer, L. (ed.) 1962, *The Distribution and Motion of Interstellar Matter in Galaxies* (New York: W. A. Benjamin).

Woodward, P. R. 1976, *Ap. J.,* **207,** 484.

Yuan, C., and Grosbøl, P. 1981, *Ap. J.,* **243,** 432.

Zuckerman, B., and Evans, N. J. 1974, *Ap. J. (Letters),* **192,** L149.

The Structure and Evolution of the Galaxy

By K. C. Freeman
Mount Stromlo and Siding Spring Observatories
The Australian National University
Woden Post Office, ACT 2606, Canberra, Australia

Abstract

The structure and evolution of the Galaxy is discussed, with particular emphasis on the properties of the old disk, the thick disk and the metal weak halo. The dynamics of satellite accretion may be fundamental to understanding the kinematics and structure of the thick disk and halo.

I very much regret that I did not know Strömgren well. I greatly admire his farsighted work on galactic evolution, and the work of those that follow him, and it is a pleasure to give a talk on the structure and evolution of the Galaxy at this meeting in his honor.

Our Galaxy has several major structural components, which I will discuss in turn. They include the thin disk (total mass of about 6.10^{10} $M_\odot$), the metal weak halo (2.10^9 $M_\odot$), and the thick disk (4.10^9 $M_\odot$). There is also the bulge (5.10^9 $M_\odot$) and the dark corona (10^{12} $M_\odot$), which I will not discuss here. For recent reviews, see Frogel (1988), Wyse and Gilmore (1988), Strömgren (1987) and Freeman (1987).

1. *The Thin Disk*

Near the sun, the star formation rate in the disk has been approximately constant over the last 10 Gyr (Twarog 1980; Miller and Scalo 1979). The older stars, in the mean, have a higher velocity dispersion and a lower chemical abundance [Fe/H]. There are several versions of the age-velocity dispersion relation (AVR) and the age-metallicity relation (AMR): see Strömgren (1987) for recent estimates from evolved F stars. The work of Nissen *et al.* (1985) shows that the AMR has substantial intrinsic scatter; we will return to this subject later. The AVR (eg Wielen 1974) shows a fairly continuous increase of velocity dispersion σ with age, which indicates that the disk is evolving dynamically. The dynamical heating process is not yet fully understood. It has been known for many years that an isotropic diffusion process reproduces the observed shape of the AVR and the observed ratio of the vertical component of the velocity dispersion to the planar component (see Fuchs and Wielen 1986). Two-body interactions between stars and giant molecular clouds (GMC's) have long been considered as the primary heating source (Spitzer and Schwarzschild 1951). Recent detailed studies (eg Lacey 1984) show that this mechanism alone is not fully successful in reproducing the shape and the amplitude of the velocity ellipsoid. However, the interaction of a disk of stars and GMC's will be more complex: spiral waves will increase the plane component of σ, and scattering by GMC's will deflect the stellar orbits out of the plane, increasing the vertical component of σ. Carlberg (1987) has shown that this process appears to work well in reproducing the shape of the velocity ellipsoid.

An interesting complication in the AVR arises from recent work by Knude *et al.* (1987). For disk stars, the equilibrium ratio of the azimuthal to the radial components of the velocity dispersion, σ_V/σ_U, is $[-B/\ (B-A)]^{1/2} \approx 0.7$, from epicyclic theory, where A and B are the Oort constants. We would expect this ratio to be established on a timescale of a few epicyclic periods, ie a few $\times\ 10^8$ yr. Knude *et al.* studied a sample of evolved F stars at the NGP, with known distances, ages and [Fe/H] between 0.2 and -0.2. Their data suggest that σ_V/σ_U does not reach its equilibrium value for stars younger than about 4 Gyr. Why does it take so long? We can speculate that this effect is associated with the dissolution of aggregates along Lindblad dispersion orbits, for which phase mixing can take much longer than the expected few $\times\ 10^8$ yr.

We now turn to the vertical structure of the thin disk. Optical observations of other galaxies suggested that the vertical density distribution of the disk follows the $\mathrm{sech}^2(z/z_o)$ law, where z is the height above the plane, and z_o is a scalelength which is independent of radius (see van der Kruit and Searle 1981a). This form is dynamically simple; it represents an isothermal sheet. Dust in the galactic plane prevents a definitive test of this law at optical wavelengths for other spiral galaxies. For our Galaxy, star counts (Pritchet 1983) suggest that the vertical density distribution is closer to exponential. Recently Wainscoat *et al.* (1989) have made infrared observations of the edge-on spiral IC 2531, which is structurally fairly similar to our Galaxy. In the K-band, the dust absorption is low, and they find that the vertical structure of this galaxy is very well represented by a simple exponential distribution. This light distribution has now been found by Wainscoat (unpublished) for several other edge-on spirals. This is interesting, because the vertical density profile of the disk represents the sum of components of different age and vertical velocity dispersion, which add together to give the observed exponential light distribution. It means that the star formation rate and the disk heating rate are coupled in a similar way in all these galaxies. This becomes particularly interesting if the mean age of the disk changes with radius, because the coupling is then able to maintain this vertical structure.

As mentioned earlier, the equilibrium value of the ratio σ_V/σ_U for disk stars is determined from epicyclic theory by the local Oort constants. There is no such constraint on the ratio σ_W/σ_U, where σ_W is the vertical velocity dispersion: σ_W depends on the local disk heating. However, there is evidence now that σ_W/σ_U is constant between about 2 kpc and 18 kpc from the galactic center. The argument depends partly on two well observed properties of other disk galaxies: (i) the disks have an exponential radial surface brightness profile $\mu(R) \propto \exp(-R/h_R)$, where h_R is the radial scalelength, and (ii) the vertical scaleheight of the disk is independent of radius (see van der Kruit and Searle 1982). We assume that the mass to light ratio of the luminous matter in the disks is independent of radius, and that most of the mass of the disk is in old stars with velocity dispersion $(\sigma_U, \sigma_V, \sigma_W)$. From the vertical equilibrium of the disk, it then follows that

$$\sigma_W \propto \exp(-R/2h_R) \tag{1}$$

and this has been observed directly in some face-on galaxies (eg van der Kruit and Freeman 1986). Recently Lewis and Freeman (1989) measured the radial component $\sigma_U(R)$ for the old disk of our Galaxy, for $2 < R < 18$ kpc. They found that σ_U also follows the law

$$\sigma_U \propto \exp(-R/2h) \tag{2}$$

where $h \approx$ the photometrically determined radial scalelength h_R of the Galaxy. It follows that the ratio σ_W/σ_U is approximately constant (at about 0.5) over most of the galactic disk. This provides another strong constraint on the theory of disk heating.

Finally, we should consider another problem which will hopefully soon be resolved: the local matter density ϱ_o and the surface density Σ of the galactic disk, as determined from the vertical distribution and kinematics of tracer populations (eg F stars, K dwarfs, K giants: see Oort 1965). At this time, the best estimates (Bahcall 1984) are $\varrho_o \approx 0.20\ M_\odot\ pc^{-3}$ (which is about twice the density of known matter near the sun) and $\Sigma \approx 70\ M_\odot\ pc^{-2}$. These estimates are far from definitive, because of deficiences in the tracer samples. Several groups, including a Danish group, are working on this problem with samples that are much better understood, and we should expect results soon. This work is important not only for the problem of local dark matter in the disk but also for understanding the radial equilibrium of the galactic disk. For example, if it turns out that $\Sigma \approx 80\ M_\odot\ pc^{-2}$, then the bulge and the disk together provide most of the local radial gravitational force. On the other hand, if $\Sigma \approx 50\ M_\odot\ pc^{-2}$, then the dark corona dominates the radial forcefield near the sun. Evidence from other galaxies suggests that the disk and bulge probably dominate at locations corresponding to the position of the sun (about $2h_R$ from the galactic center) but this is still contentious (see van Albada and Sansici 1986).

2. *The Metal Weak Halo*

The metal weak halo is a slowly rotating component. The kinematics of nearby stars show a fairly abrupt transition at [Fe/H] ≈ -1, from the rapidly rotating disk (and thick disk: see belov) to the slowly rotating ($V_{rot} \approx 0$) metal weak halo (eg Norris 1986). The halo shows little direct evidence for an abundance gradient or for any dependence of mean rotational velocity on [Fe/H].

At the time of the Eggen, Lynden-Bell and Sandage (1962) paper on the relationship between the abundance and kinematics of halo stars, the metal weak stars ([Fe/H] < -1) were kinematically selected and had orbital eccentricities $e > 0.4$; ie they were kinematically members of the spheroidal halo. More recent work on non-kinematically selected samples of metal weak stars shows that about 20 % of stars with [Fe/H] < -1.2 have $e < 0.4$ (Norris *et al.* 1985); their mean rotation is high ($V_{rot} \approx 180$ kms^{-1}) and their vertical velocity dispersion $\sigma_W \approx 45$ km s^{-1}. These kinematical parameters for the metal weak stars with $e < 0.4$ are very similar to those of the thick disk, and it seems that they form a metal weak tail of the thick disk.

The globular clusters of the Galaxy fall into two components (Zinn, 1985): clusters more metal rich than [Fe/H] ≈ -0.8 belong to a rapidly rotating disklike system with kinematical properties that are again very similar to those of the thick disk (Armandroff 1989), while the more metal poor clusters belong to the slowly rotating halo. The halo clusters show no abundance gradient. However, the second parameter effect (the anomalous distribution of stars along the horizontal branch) increases with increasing galactocentric distance, and suggests that the outer halo clusters formed later and over a longer period (Zinn 1980). It is not clear yet whether the disk clusters and the halo clusters have similar ages.

The formation of globular clusters is not well understood. Fall and Rees (1985) proposed that clusters form by fragmentation of the collapsing protogalaxy. Thermal instability of the low abundance gas leads to a two-phase medium, the two phases having temperatures of about 10^6 K (the virial temperature) and 10^4 K (H-recombination) with a density contrast of about 400. The critical mass for gravitational instability of the low temperature phase is about 10^6 $M_\odot$. Searle and Zinn (1978) suggested that the clusters form in small disklike satellite galaxies which are then accreted by the Galaxy. (We note that globular clusters are forming now in some disklike galaxies, like the LMC, SMC and M33, although not at this time in our Galaxy.) Mass loss from these small satellites helps to explain the observed abundance distribution of the galactic globular clusters.

In the Searle-Zinn picture, the globular clusters form in metal weak satellites such as dwarf irregulars or nucleated dwarf ellipticals, which are then accreted. The globular clusters are dense and can survive the accretion event. However, most of the accreted satellite is disrupted and becomes part of the field halo. We might then expect to see some moving stellar groups in the halo, and there is some limited evidence that such groups do exist (eg Eggen 1979). In this picture for the formation of the metal weak halo, the halo comes from these pre-formed weak satellites, so the kinematics of the halo then depends on the dynamics of sinking and merging satellites. This picture does offer a natural explanation for the striking kinematical discontinuity between the rapidly rotating disk(s) and the non-rotating halo.

Recently Quinn and Goodman (1986) have studied the dynamics of satellites sinking into a parent disk galaxy. Here are their main conclusions:

1) Satellites with masses of a few percent of the parent mass and with orbits that come within about 8 of the parent's scale lengths are captured by dynamical friction in a few dynamical times (somewhat longer for satellites in retrograde orbits). 2) If the orbits do not come within about 8 scalelengths, then the satellite survives. 3) If the orbital inclination of the satellite is less than about 60° to the plane of the parent, then the satellite orbit is dragged down into the plane and then decays radially. 4) The orbit energy goes into heating the parent disk, radially and in z. It follows that, if satellite accretion produced the metal halo, then this accretion must have occurred while the disk was mainly gaseous, to allow the disk to settle down again and form the presently observed thin disk. The heating of the stellar component of the early partly stellar disk could produce the thick disk. In this picture, the thick disk should be very old, ie older than most disk stars.

There is a problem in understanding the shape of the metal weak halo. For metal weak stars, the velocity dispersion in the direction of the SGP is constant, at about 75 km s^{-1}, out to distances of about 25 kpc. At lower latitudes, the observed velocity dispersion is in the range 120 to 140 km s^{-1} (Ratnatunga and Freeman 1989). Dynamical models (eg White 1989) then indicate that the halo must be significantly flattened ($c/a \approx 0.5$). On the other hand, most direct indicators (star counts, distributions of RR Lyrae stars and BHB stars and globular clusters) suggest that the halo is nearly spherical (see Freeman 1987 for references). Recent developments, which show that the metal weak halo is not just a simple nonrotating system, may give some insight into this problem.

Hartwick (1986) investigated the galactic distribution of RR Lyrae stars with $[Fe/H] < -1$. To represent their distribution, he found that two components were needed (i) an inner ($R \lesssim 8$ kpc) flattened component, with $c/a \approx 0.6$, and (ii) a spherical outer component. By analogy with the halo globular clusters, he argued that both components are slowly rotating, so the inner component is flattened by its anisotropic velocity distribution. Sommer-Larsen (1986) compared the kinematics of nearby non-kinematically selected metal weak stars in the abundance ranges -1.2 to -1.5 and $\lesssim -1.5$. Both subsamples have similar values of σ_U and σ_V and are slowly rotating. However the values of σ_W are very different: 59 km s^{-1} for the metal richer subsample and 102 km s^{-1} for the metal poorer stars. Again, this indicates that the metal richer halo stars form a flatter subsystem. These observations suggest that the local kinematics are perhaps dominated by the flatter component, while the star count data on the shape of the halo come mainly from the more spherical component.

However, the situation is not yet satisfactory: the anisotropy ($\sigma_U/\sigma_W > 1$) appears to persist to large values of R,z ($\approx$ 25 kpc), which would not be expected if the spherical component dominates at large distances.

How do these observations fit into the Searle-Zinn/Quinn-Goodman accretion picture? Here are a few comments: 1) Larger satellites are (now) more metal rich and are denser (see Dekel and Silk 1986). They will survive orbit decay (against tidal disruption) to smaller galactocentric radii, and their orbits at destruction will have lower inclinations and their debris will have lower σ_W. The debris from these larger satellites would contribute to an inner, flatter, metal richer population within the metal weak halo. 2) The disk globular clusters are metal rich ([Fe/H] $\gtrsim -0.8$). The absence of metal weak *disk* clusters suggests, in the accretion picture, that the metal weak clusters formed in fragile satellites (which could not survive orbit decay), rather than in larger more robust satellites. Nucleated dwarf ellipticals would seem to be attractive candidates for the formation sites of metal weak clusters: their dense nuclei would survive the accretion event, as present day globular clusters, while the fragile envelope would be disrupted to become part of the metal weak field halo. 3) The second parameter effect (the outer clusters formed later and over a longer period of time than the inner clusters) is also readily understood in the accretion picture. The theory shows that the outer clusters survive longer and have a larger spread in survival time.

3. *The Thick Disk*

The vertical density distribution of pure disk galaxies has the sech^2 or exponential form described above. The disks of galaxies with small bulges show extra light in a thick disk distribution for $z \gtrsim 1$ kpc, in excess of that expected from the thin disk alone. This thick disk was first observed photometrically in other galaxies by Burstein (1979), and then in more detail by van der Kruit and Searle (1981b), and was identified in our Galaxy from star counts by Gilmore and Reid (1983). We will see that this thick disk is kinematically quite distinct from the metal weak halo (which would make a negligible contribution to the surface brightness of our Galaxy if it were seen edge-on from outside).

Stars of the galactic thick disk have now been identified *in situ* in several fields (see Freeman 1987 for references). The mean abundance of the thick disk stars is about -0.7, with a dispersion of about 0.3. The vertical velocity dispersion σ_W is about 40 kms^{-1}, and its scaleheight is about 1 kpc. In [Fe/H] and σ_W, the thick disk is intermediate between the thin disk and the metal weak halo, and in this sense it is an intermediate population. However, its rotation is almost as rapid as that of the thin disk: the asymmetric drift is in the range 30 to 50 kms^{-1}, so the thick disk is close to centrifugal equilibrium, with somewhat more internal energy than the thin disk. This intermediate population was of particular interest to Strömgren; it holds many clues to the formation of the Galaxy, as he saw long ago.

The vertical structure, kinematics and chemical properties of the thick disk are fairly similar to those of the Zinn/Armandroff population of disk globular clusters. It may be that the stellar thick disk, as defined by stars with [Fe/H] ≈ -0.7, is a little younger: see Schuster and Nissen (1987) and Norris and Green (1989). However, Nissen *et al.* (1985) have shown that the age-metallicity relation for nearby stars has a large intrinsic spread, and in particular that there are very old stars (ages $\gtrsim$ 12 Gyr) at all abundances in the range $0 >$ [Fe/H] > -1. If accretion is important in forming the thick disk, we would expect that the stars of the thick disk are as old or older than the oldest stars of the thin disk, and it would be very interesting to know whether this is true.

There is some evidence that the galactic thick disk may not extend much beyond about 10 kpc from the galactic center. Ratnatunga and Freeman (1989) found thick disk stars *in situ* out to galactocentric distances r of about 10 kpc in their survey at $l = 270°$, $b = 39°$. Friel's (1988) survey at $l = 194°$, $b = -49°$ showed no stars with thick disk kinematics, although many thick disk stars were discovered in her survey at $l = 36°$, $b = 55°$. Armandoff's (1989) compilation of disk globular clusters has no clusters with r $\gtrsim$ 8 kpc. This apparent limited extent of the thick disk has a direct and interesting implication for the epoch of satellite accretion, if the thick disk was indeed formed by heating of the stellar component of the early thin disk by accretion of satellites. It suggests that, by the end of the epoch of accretion, star formation in the early thin disk had not progressed much beyond about 10 kpc.

There is some controversy now about the nature of the thick disk: is it a discrete component, or is it just the high energy, low abundance tail of the thin disk (see Freeman 1987 for references). My view is that it is probably discrete. We know that not all disk galaxies have thick disks; there is an apparent association of bulges and thick disks, which suggests that the thick disk does not come from the normal secular evolution of the thin disk. Also, the velocity dispersion – abundance relation as given by Strömgren (1987) shows an abrupt rise in σ_W at [Fe/H] ≈ -0.7, from values of about 20 to 25 km s^{-1} (which we would associate with the old thin disk) to about 40 kms^{-1} (which is the characteristic thick disk value). This is a serious question, because the thick disk surely provides some of the potentially most useful clues to understanding the early formation events of our Galaxy, and it is important that we establish observationally whether or not it is a discrete component. This will probably be settled by careful derivation of the local AVR and AMR, to which Strömgren's large F-star program is very well suited.

References

Armandroff, T. 1989. *Astron. J.,* **97**, 375.
Bahcall, J. 1984. *Astrophys. J.,* **287**, 926.
Burstein, D. 1979. *Astrophys. J.,* **234**, 829.
Carlberg, R. 1987. *Astrophys. J.,* **322**, 59.
Dekel, A., Silk, J. 1986. *Astrophys. J.,* **303**, 39.
Eggen, O. 1979. *Astrophys. J.,* **229**, 158.
Eggen, O., Lynden-Bell, D., Sandage, A. 1962. *Astrophys. J.,* **136**, 748.
Fall, S. M., Rees, M. 1985. *Astrophys. J.,* **298**, 18.
Freeman, K. 1987. *Ann. Rev. Astron. Astrophys.,* **25**, 603.
Friel, E. 1988. *Astron. J.,* **95**, 1727.
Frogel, J. 1988. *Ann. Rev. Astron. Astrophys.,* **26**, 51.
Fuchs, B., Wielen, R. 1986. In *The Galaxy,* ed. Gilmore, G., Carswell, R. (Dordrecht: Reidel), p. 375.
Gilmore, G., Reid, N. 1983. *M.N.R.A.S.,* **202**, 1025.
Hartwick, F. D. A., 1986. In *The Galaxy,* ed. Gilmore, G., Carswell, R. (Dordrecht: Reidel), p. 281.
Knude, J., Schnedler Nielsen, H., Winther, M. 1987. *Astron. Astrophys.,* **179**, 115.
Lacey, C. 1984. *M.N.R.A.S.,* **208**, 687.
Lewis, J., Freeman, K. 1989. *Astron. J.,* **97**, 139.
Miller, G., Scalo, J. 1979. *Astrophys. J. Suppl.,* **41**, 513.
Nissen, P.E., Edvardsson, B., Gustafsson, B. 1985. In *ESO Workshop on Production and Distribution of C,N,O Elements,* ed. Danziger, I., Matteucci, K., Kjär, K (Garching: ESO), p. 131.
Norris, J. 1986. *Astrophys. J. Suppl.,* **61**, 667.
Norris, J., Bessell, M., Pickles, A. 1985. *Astrophys. J. Suppl.* **58**, 463.
Norris, J., Green, E. 1989. *Astrophys. J.,* **337**, 272.
Oort, J. 1965. In *Galactic Structure (Stars and Stellar Systems* Vol 5), ed. Blaauw, A., Schmidt, M. (Chicago: Univ. Chicago Press), p. 455.
Pritchet, C. 1983. *Astron. J.,* **88**, 1476.
Quinn, P., Goodman, J. 1986. *Astrophys. J.,* **309**, 472.
Ratnatunga, K., Freeman, K. 1989. *Astrophys. J.,* **339**, 126.
Schuster, W., Nissen, P.E. 1987. In *Stellar Evolution and Dynamics in the Outer Halo of the Galaxy,* ed. Azzopardi, M., Matteucci, F. (Garching: ESO), p. 141.
Searle, L., Zinn, R. 1978. *Astrophys. J.,* **225**, 357.
Sommer-Larsen, J. 1986. Licentiate thesis, Copenhagen University.
Spitzer, L., Schwarzschild, M. 1951. *Astrophys. J.,* **114**, 385.
Strömgren, B. 1987. In *The Galaxy,* ed. Gilmore, G., Carswell, R. (Dordrecht: Reidel), p. 229.
Twarog, B. 1980. Astrophys. J. Suppl., **44**, 1.
van Albada, T., Sansici, R. 1986. *Phil. Trans. Roy. Soc. Lond.,* **A 320**, 447.
van der Kruit, P., Freeman, K. 1986. *Astrophys. J.,* **303**, 556.
van der Kruit, P., Searle, L. 1981a. *Astron. Astrophys.,* **95**, 105.
van der Kruit, P., Searle, L. 1981b. *Astron. Astrophys.,* **95**, 116.
van der Kruit, P., Searle, L. 1982. *Astron. Astrophys.,* **110**, 61.
Wainscoat, R., Freeman, K., Hyland, A. R. 1989. *Astrophys. J.,* **337**, 163.
White, S. 1989. *M.N.R.A.S.,* **237**, 51P.
Wielen, R. 1974. in *Highlights of Astronomy,* ed. Contopoulos, G. (Dordrecht: Reidel), **3**, 395.
Wyse, R., Gilmore, G. 1988. *Astron. J.,* **95**, 1404.
Zinn, R. 1980. *Astrophys. J.,* **241**, 602.
Zinn, R. 1985. *Astrophys. J.,* **293**, 424.

Galaxies, Galactic Nuclei and Dark Matter

By Martin J. Rees
Institute of Astronomy
Madingley Road, Cambridge, CB3 0HA, U.K.

I. *Introduction*

The properties of our own Milky Way Galaxy have been expertly reviewed by Professor Ken Freeman. My aim will be to describe some features of galaxies in general. Here we are still groping for answers to the most basic questions:

1. We do not know why such things as galaxies should exist at all – why these assemblages of stars and gas with fairly standardised properties are the most conspicuous large-scale features of the cosmos.
2. About 90 % of the mass associated with galaxies is hidden. The luminous stars and gas contribute only about a tenth of the gravitating material inferred from dynamical arguments. What the rest consists of is still a mystery.
3. It is unclear why the nuclei of some galaxies flare up, and release the colossal amount of non-stellar radiation emitted from quasars and radio galaxies.

We are perplexed about these issues, just as 50 years ago our predecessors were perplexed about the nature of stars. But some of us are hopeful that the physical processes underlying galaxies are coming into focus, and can at least be seriously addressed. I must apologise in advance to specialists on this topic for the 'broad brush' and inevitably distorted exposition I shall be giving.

In their already-classic book on galactic dynamics, Binney and Tremaine (1987) make the point that galaxies are to astronomy what exosystems are to biology. They are not only dynamical units, but chemical units as well. The atoms we are made of come from all over our Milky Way Galaxy, but few come from other galaxies. The ecological analogy reflects other features of galaxies: their complexity, ongoing evolution, and relative isolation.

Single stars, the individual organisms in the galactic ecosystem, can be traced from their birth in gas clouds through their lifecycle. And we have come to understand why *stars* exist with the general properties we see. The question why *galaxies* exist is less straightforward than the equivalent question for stars. Galaxies formed at an earlier and remote cosmic epoch. We don't know how much *can* be explained in terms of ordinary processes accessible to study now, and how much has its causes in the earliest universe.

Figure 1. 'Cartoon' showing three stages in the traditional picture of protogalactic collapse.

There is an elaborate taxonomy for galaxies, but the most obvious categories are disks and spheroids or ellipticals. There is a well-known cartoon model, dating back about 30 years, to account for this basic morphological distinction. Suppose that a galaxy started life as an irregularly-shaped gas cloud contracting under gravity, and that the collapse of such a gas cloud were highly dissipative, in the sense that any two globules of gas that collided would radiate their relative kinetic energy and merge (Figure 1). The end result of the collapse of such a cloud would be a rotating disk. This is the lowest energy state that the cloud can reach if it does not lose or redistribute its angular momentum. On the other hand, *stars* do not collide with each other, and are unable to dissipate energy in the same fashion as gas clouds. So the *rate of conversion of gas into stars* could be the crucial feature determining the type of galaxy that results. Elliptical galaxies would be those in which the conversion is fast, so that most stars have already formed before the gas has had time to settle down in a disk. Disk galaxies result when the star formation is delayed until the gas has already settled into a disk. According to this traditional picture, disk galaxies are those with slower metabolism, which have not yet got so close to the final state in which essentially all the gas is tied up in low mass stars of dead remnants. The main challenge is to elucidate this process, and to determine how and when the protogalactic clouds emerged during the overall expansion of the Universe.

II. *What is Special about Galactic Dimensions?*

This story depicted in Figure 1 has many inadequacies, and I'll return to some of them later. In particular, there is no scale in the picture. Is there any physics that singles out clouds of galactic dimensions, just as, since Eddington and Chandrasekhar, we have known the natural scale of stars? All we have for galaxies is a simple but suggestive physical argument. Two timescales are important in determining how a self-gravitating gas cloud evolves. The first of these is the dynamical or freefall time, which is of order $(G\varrho)^{-1/2}$, its precise value depending on the geometry of the collapse. The second is the radiative cooling timescale. This depends on the gas temperature T_g, and can be written $T_g/\varrho\Lambda(T_g)$ where Λ can be calculated from atomic physics.

If $t_{\rm cool}$ exceeds $t_{\rm dynamical}$, a cloud of mass M and radius r can be in quasi-static equilibrium, with the gas at the virial temperature. But if $t_{\rm cool} < t_{\rm dynamical}$ such equilibrium is impossible (Figure 2). The cloud cools below the virial temperature

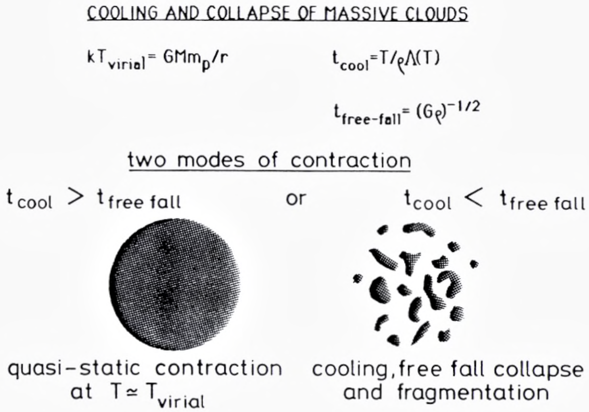

Figure 2. Cooling and contraction of self-gravitating gas clouds.

and undergoes freefall collapse or fragmentation. We would expect clouds to collapse and fragment in the fashion depicted in Figure 1 only if they enter the part of M - r plane where cooling is faster than freefall. A simple calculation shows that this criterion involves a characteristic mass-independent radius of order 75 kpc and a characteristic mass $M_{\rm crit}$ of order 10^{12} $M_\odot$. Clouds less massive than $M_{\rm crit}$ will readily fragment, but above $M_{\rm crit}$ fragmentation is impossible unless the cloud contracts until its radius is below $r_{\rm crit}$. This characteristic mass and radius, consequences of straight-

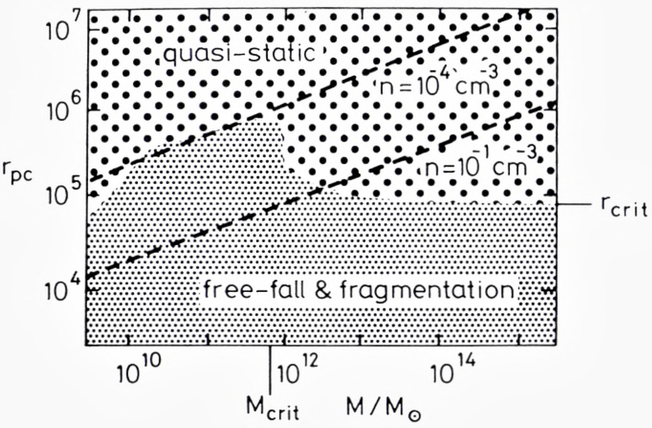

Figure 3. The quasi-static and free-fall regions (cf. Fig. 2) are here presented in a mass-radius plot (from Rees and Ostriker 1977). M_{crit} and r_{crit} should set characteristic upper limits to the dimensions of galaxies. Clouds with $M \gg M_{crit}$ would be quasi-static unless at very high densities (cf. gas in clusters of galaxies).

forward physics (Figure 3), feature in many cosmogonic schemes as at least setting an upper limit to the scale of galaxies.

Eddington claimed that a physicist on a cloud-bound planet could have predicted the properties of the gravitationally-bound fusion reactors that we call stars. But these simple considerations don't suffice to predict galaxies, even with hindsight. This is because any true explanation of galaxies must involve setting them in a cosmological context.

III. *The Cosmological Context*

In a memorable invited discourse at the Patras IAU General Assembly, Zel'dovich (1982) discussed the hot big bang model, which he opined was as sure as that the Earth goes round the Sun. We may not all quite share his exuberant certitude. But most of us regard the hot big bang as the 'best buy' cosmology, more than 50 % likely to be essentially correct. According to this picture everything emerged from a universal thermal soup which was initially smooth, and almost featureless, but not quite. There were, we don't really know why, small fluctuations from place to place in the expansion rate. Structures emerged via gravitational instability as over-dense regions lagged more and more behind the universal expansion, and eventually condensed out as embryo galaxies and clusters.

Theorists trace back the history of the hot big bang over 60 decades of logarithmic

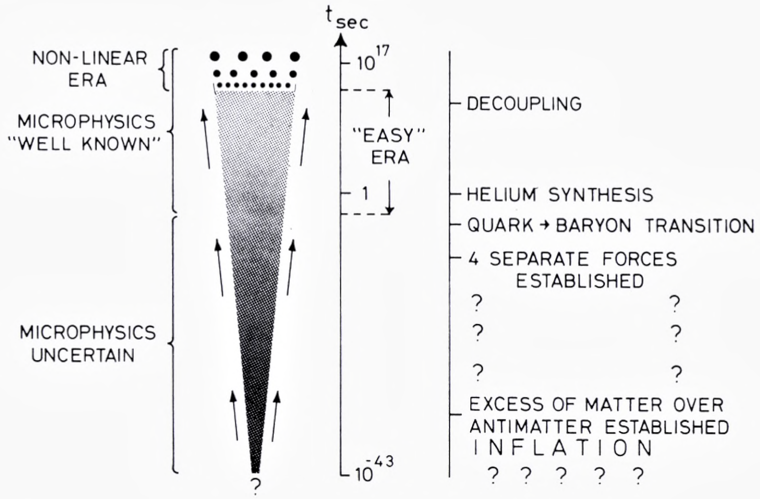

Figure 4. Stages in the evolution of the 'standard' big bang model universe.

time. The events and stages in the cosmic expansion are summarised in Figure 4, which goes back to the earliest era, the intellectual habitat of the 'gee whiz' fringe of particle physicists. For our present purposes the uncertain details are irrelevant. It may, though, be conceptually useful to divide cosmic history into 3 parts. For the first 40 decades the *micro*physics is uncertain. When the universe cools below 10 MeV and the density falls below nuclear density, the microphysics become straightforward. Initial irregularities, owing their origin to the first era, amplify via gravitational instability, and things become *less* straightforward when the first of these condense out. Then we confront a set of new difficulties. The physics is just Newtonian gravity and gas dynamics, but the complications are those of non-linearity. The 'recent' universe is hard to understand for the same reason that weather prediction is difficult.

A key question is how much can be explained by processes occurring at the range of epochs accessible to observational astronomers, and how much has to be attributed to the uncertain physics at ultra-early eras.

The main types of relevant data are morphological classifications (dating back to Hubble); correlations between luminosity, velocity dispersion and size; and the statistics of galaxy clustering. Any quantitatively satisfactory theory must explain these things. We have no generally agreed theory yet. Indeed, as Saslaw has put it, 'if galaxies didn't exist, we'd have no problems explaining the fact'. Moreover, the seekers for any such theory must first face a most embarrassing circumstance: this is the *dark matter problem:* evidence that 90 % of the mass of galaxies is unaccounted for, and takes some unknown form.

IV. *Dark Matter*

The evidence for dark matter dates back more than 50 years, but has firmed up since the classic papers of Einasto, Kaasik and Saar, and Ostriker, Peebles and Yahil, both published in 1974. The masses inferred from relative motions of galaxies in apparently bound groups and clusters exceed by a factor 10 those inferred from the internal dynamics of the luminous parts of galaxies. This apparent discrepancy could be resolved if galaxies were embedded in extensive dark haloes. The halo hypothesis can be checked in some edgeon disk galaxies, where emission from gas can be observed out at radii far exceeding the extent of the conspicuous stellar disk. The mass of this gas is itself negligible, but rotation velocities derived from its spectral lines do not fall off as $R^{-1/2}$, as would be expected if the gas were orbiting a mass distribution concentrated at much smaller radii. Instead the velocity remains almost constant, implying that the mass within radius R is proportional to R out to 80 kpc in some cases. Direct lower limits on the mass-to-light ratio in the outlying parts of some galaxies exceed 300 solar units.

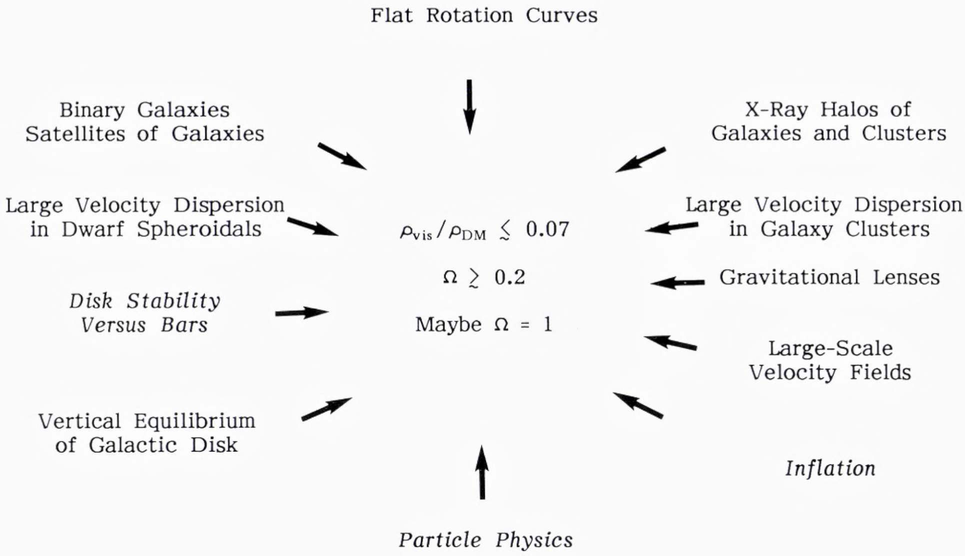

Figure 5. The various lines of evidence for dark matter. Items in italics refer to theoretical arguments (diagram due to J. Kormendy).

In some elliptical galaxies also, the mass seems to increase proportional to R out to large R. In M87, such evidence comes from globular cluster orbits, and, still further out, is inferred from the X-ray temperature and profile of the diffuse gas. On a larger scale, we have evidence from *clusters of galaxies*, along the lines first discussed by

Zwicky and Sinclair Smith in the 1930s. Many independent lines of evidence point towards the existence of dark matter (these are summarised in Figure 5). This has as good a claim to be termed a paradigm shift as any development one can think of in modern astronomy.

The dynamically-inferred dark matter, though ten times the luminous matter, still amounts to only 10 or 20 per cent of what is required for a closed universe: the corresponding value of the density parameter Ω, the ratio of the actual density to the cosmological critical density, is 0.1 or 0.2.

V. *The Nature of the Dark Matter: Baryonic or not?*

What is the halo dark matter? The first possibility that comes to mind is faint stars or stellar remnants. Figure 6, due to Carr, Bond and Arnett (1984), quantifies the maximum hidden contribution to Ω that could be made by stars or their remnants in the mass range from 10^{-2} to 10^{8} $M_\odot$. There are two tenable dark matter candidates: very low mass stars, below 0.1 $M_\odot$, or the remnants of very massive stars. Ordinary stars above 0.1 $M_\odot$ would contribute too much background light unless they had all

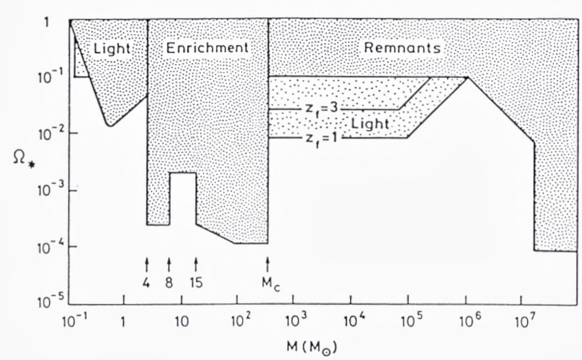

Figure 6. Constraints on the fraction of the critical density that would be present in stars or stellar remnants of various masses. The stars are presumed to have formed at some redshift z_f. Possible candidates for the dark matter are low mass stars (brown dwarfs or 'Jupiters') or very massive objects (VMOs).

evolved and died, leaving dark remnants. But the remnants of ordinary massive stars of 10-100 $M_\odot$ would produce too much material in the form of heavy elements. Stars with core masses above 200 $M_\odot$ end their lives, via the pair production instability, by collapsing rather than exploding. These very massive objects, (VMOs for short), do not eject heavy elements, and leave black hole remnants. Such objects, if they constitute our own galactic halo, can't however exceed 10^6 $M_\odot$ each, because otherwise dynamical friction, whereby a hole transfers energy to lighter stars close to its path, would have led to excessive thickening of the galactic disk.

Is it likely or unlikely that a forming galaxy should convert most of its mass into either ultra-low mass stars or objects heavier than a few hundred suns? We don't understand enough about star formation, even close at hand in for instance the Orion nebula, to be confident in saying how the initial mass function might be affected by intense background radiation, absence of heavy elements, lack of magnetic fields, and the rest. Theory therefore cannot arbitrate reliably between low mass and VMO options (Figure 7).

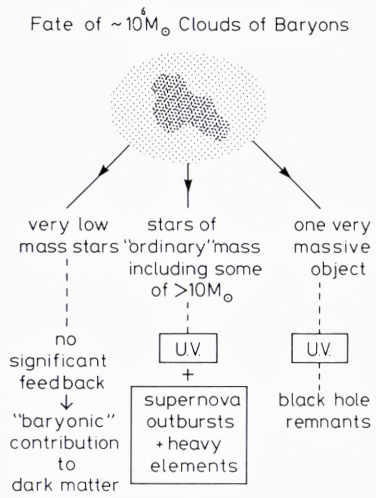

Figure 7. In many cosmogonic schemes, star formation would be initiated in baryonic clouds of around 10^6 $M_\odot$, but we have no firm theoretical basis for deciding the characteristic mass, or the IMF, of these first stars.

Can we learn from observations about what the dark matter is? Low mass objects would be perhaps detectable in the infrared: the nearest would be less than a parsec away, with high proper motions. There are two handles on VMO remnants. They might reveal their presence by accretion on passage through interstellar clouds. Also, they imply that galaxies would be bright when young – there are constraints from the sky brightness, and from the faint galaxy counts, but the quantitative interpretation of these limits depends on the uncertain redshift of galaxy formation.

One way of detecting compact dark objects, and discriminating between the Jupiter and VMO options, is by searching for evidence of gravitational lensing. The probability of seeing lensing due to an object in our own halo is only about 10^{-6}. But

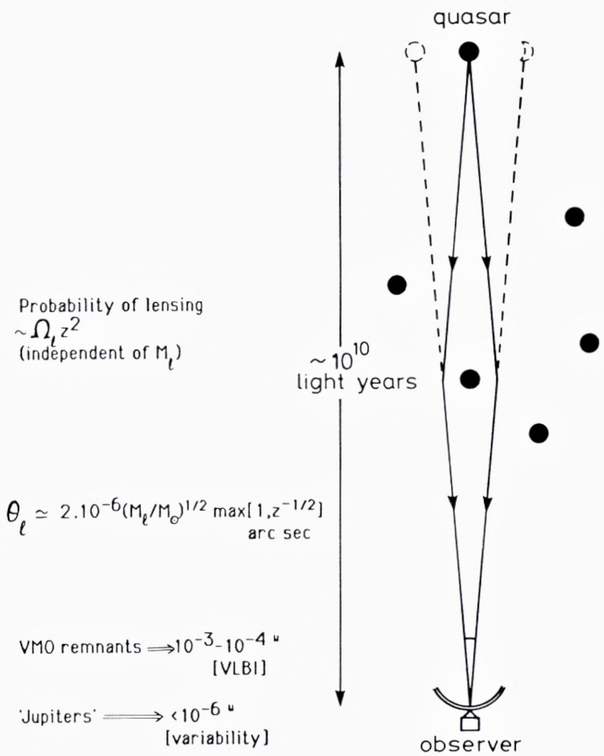

Figure 8. Properties of gravitational microlensing by a population of compact objects along the line of sight to a cosmologically-distant quasar.

the cross-section for effective lensing is proportional to distance, so there is, perhaps surprisingly, much more chance of detecting objects in the haloes of galaxies half way out to the Hubble radius (Figure 8). The probability that a compact source at redshift $z > 1$ is significantly microlensed by objects along the line of sight is of order Ω, independent of the individual lens mass involved (Refsdal 1970, Press and Gunn 1973). The *angular separation* of the images, proportional to (lens mass)$^{1/2}$ is however a diagnostic of the masses. For masses above 10^5 $M_\odot$, very long baseline radio interferometers provide adequate resolution. We could probably already exclude $\Omega = 1$ in such objects.

For brown dwarfs of below 0.1 $M_\odot$, the angular scale is less than a micro arcsecond. This cannot be directly resolved by any technique, until optical interferometers are deployed in space. There is nevertheless a genuine prospect of detecting lensing of this kind because of the variability that would ensue if the lens were to move transversely (*e.g.* Gott 1981). An object at the Hubble distance moving at 100 km s^{-1} takes only a few years to tranverse a micro arcsecond. The image structure and time variation are more complicated if the line of sight passes through, for example, a galactic halo, thereby encountering an above-average column density of dark matter. Several objects may then contribute to the imaging, yielding a frosted glass effect, whose pattern, though too small to be seen directly, would vary on a timescale of months or years.

Those I've just discussed are the 'dull man's' options for dark matter. The big bang may have left not just baryons and radiation, but other species as well, which may contribute to Ω. In the standard big bang model, neutrinos are almost as abundant as microwave background photons, outnumbering baryons by around 10^9. Their mass would only need to be a few eV to make them dynamically important. More than 15 years ago Cowsik and McClelland (1973) and Marx and Szalay (1972) conjectured that neutrinos could provide the dark mass in galactic haloes and clusters. At that time the suggestion was not followed up very extensively. But by the 1980s physicists had become more openminded about non-zero neutrino masses. A change in theoretical attitude, coupled with experimental claims that the electron neutrino had a mass around 36 eV (Lyubimov *et al.* 1980), stimulated astrophysicists to explore scenarios for galaxy formation in which neutrino clustering and diffusion played a key rôle. More recently, other kinds of non-baryonic matter have also been considered.

Provided that we know the mass and annihilation cross-section for any species of elementary particle, we can in principle calculate how many survive from the big bang, and the resultant contribution each species makes to Ω. Progress in experimental particle physics may therefore reveal a particle which must contribute significantly to Ω, unless we abandon the hot big bang theory entirely.

Neutrinos have the virtue of being known to exist, but particle physicists are inventive, and have come up with a long shopping list of relics that *might* exist. The most theoretically-favoured option is some kind of electrically neutral weakly interacting massive particles, WIMPs for short. These have attractive cosmogonic consequences which I'll come back to in a moment. What is perhaps more remarkable is that such particles may be looked for in the lab.

If our Galactic Halo were composed of WIMPs with individual masses of a few GeV, they would be swarming through this room, with a density of $10^5 m^{-3}$ and speeds of around 300 kms^{-1}. Collision cross-sections are small, but whenever a WIMP collided with an atomic nucleus, the nucleus would recoil with a similar velocity, and an energy around a keV. The collision rates depend on the physical details and the target nucleus, but are in the range 1-1000 events per day per kilogram of detector.

These collision events may be detectable by a variety of cryogenic teckniques in a low background environment, at quite modest cost. Such experiments are being planned in various countries (see Primack *et al.* 1988, for a review). Ingenious schemes for detecting a halo background of exotic particles are surely among the most worthwhile and exciting high risk experiments in physics or astronomy today – potentially as important as those that led to the discovery of the microwave background in the 1960s. A null result, with just upper limits, would surprise nobody. On the other hand, such experiments could reveal new particles, as well as determining what 90 % of our universe consists of. Because the detection is sensitive to velocity, they would even reveal the halo's velocity dispersion and rotation. The mean velocity of halo particles relative to the detector would change during the year, owing to the Earth's motion round the Sun. The resultant annual modulation, with an amplitude of a few per cent and a peak in June, would be an unambiguous signature discriminating against spurious background.

VI. *Dark Matter and Galaxy Formation*

A less direct line of attack on pinning down the dark matter entails exploring the consequences of each option for galaxy formation. If it is dynamically dominant, then non-baryonic matter plays a key rôle in the process whereby small primordial pertubations evolve into protogalaxies and clusters.

The key parameter is the spectrum of density fluctuations, the rms amplitude as a function of mass scale, at the recombination epoch, $z = 1000$. Density contrasts on all relevant scales amplify at the same rate thereafter, so the first bound systems to arise via gravitational instability will have mass scales for which this amplitude peaks. The spectrum depends on what is imprinted initially, possibly modified by preferential damping of smaller scales before recombination.

The left-hand panel in Figure 9 shows the spectrum expected if the universe is dominated by neutrinos with masses 10 or 20 eV. These are moving sufficiently fast that everything is homogenised on scales at least up to 10^{14} $M_{\odot}$. The first bound systems would then be superclusters, and galaxies would result from some kind of secondary fragmentation process.

On the right is shown a 'white noise' spectrum, with amplitude larger for smaller scales. Here we have a hierarchical 'bottom-up' cosmogony, with the emergence first of subgalactic scales, then galaxies, and then clusters. (There may then be an interesting complication: radiative or explosive output from the first small bound objects could create secondary large-scale inhomogeneities that swamp those already present.)

At recombination, when the universe was 10^6 years old, the microwave background shifted redward of the visible band, and the universe entered a literal dark age. The universe remained a simple place until the first bound systems condensed. We don't know when 'first light' occurred. The dark age may have been brief, as in the right-hand panel of Figure 9, or it could have lasted a billion years if the left-hand panel is

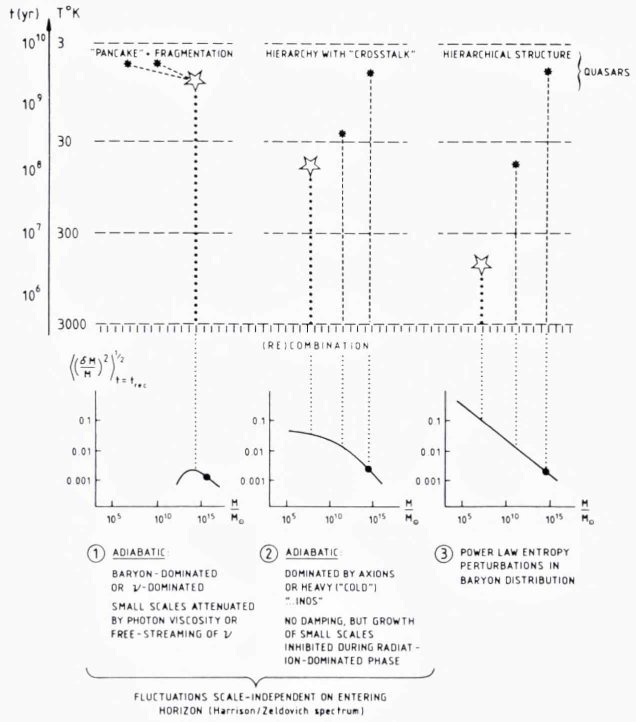

Figure 9. Cosmogonic scenarios corresponding to three different spectra for the post-recombination density perturbations. See text for further explanation.

closer to the truth. We remain more confused and ignorant about this phase of cosmic history than many seem to be about the first 10^{-35} seconds.

Let us focus now on the middle panel in Figure 9. The fluctuation spectrum here has the shape unambiguously calculable for WIMPs, or for any non-baryonic dark matter that is 'cold', in the sense that the individual particles move too slowly for damping due to free-streaming to occur, as it does for neutrinos. This 'cold dark matter' spectrum is nearly flat for small masses, so the typical fluctuation of 10^6 $M_\odot$ would collapse no earlier than the epoch corresponding to $z \approx 10$. The build-up of structure is hierarchical, in the sense that smaller scales tend to form earlier. However, because of the flat spectrum, there would be complicated 'cross talk' between many different scales. The 3σ peaks in the density distribution on galactic scales, 10^{11} $M_\odot$, would have the same amplitude as more typical peaks of mass 10^6 $M_\odot$, and would therefore collapse at the same time. It is consequently hard to analyse, either analytically or numerically, even the purely dynamical and non-dissipative aspects of the clustering. However, those studies that have been done are encouraging, in that when the amplitude of the fluctuations is normalised so as to match the data on galaxy clustering, the finer scale disposition of the dark matter closely reproduces the sizes and profiles of individual galactic haloes. An example of how such clustering develops, based on simulations of Frenk *et al.* (1985), is shown in Figure 10. Figure 11 shows the final spatial disposition of the dark matter for a slightly different model.

Even though the dark matter may be dynamically dominant, it manifests itself only gravitationally. To predict what the universe would actually look like in this model – the luminosity function of galaxies and how they are clustered – we need to develop more understanding of several physical processes. Baryons are presumed to condense in virialised haloes of dark matter in the mass range 10^8-10^{12} $M_\odot$. For larger masses, dissipative cooling may be inefficient for the reason mentioned earlier (*cf.* Figure 3). Below 10^8 $M_\odot$ the potential wells may be too shallow to capture primordial gas. The mass distribution of isolated virialised systems can in principle be learned from N-body simulations. But even if the dissipationless clustering of the dark matter is

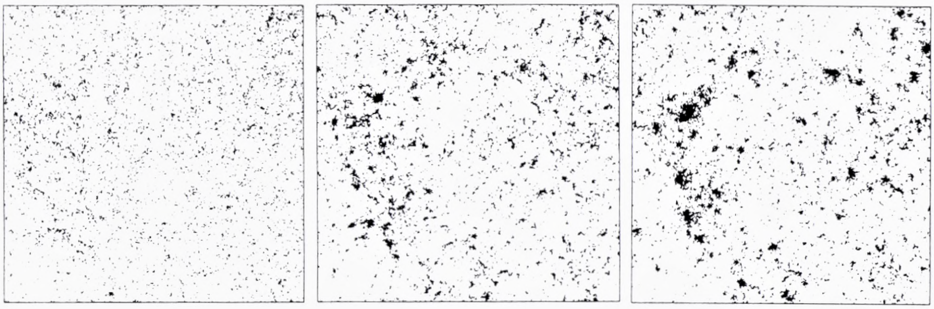

Figure 10. Three stages in the evolution of non-dissipative gravitational clustering within a comoving cubical volume, for an initial spectrum with $[(\delta M/M))^2]^{1/2}$ proportional to $M^{-1/3}$. If the right-hand panel is taken to represent the present epoch, then the middle panel is z = 0.9 and the left-hand panel z = 3.5.

accurately known, the fate of the baryonic component – how much gas falls into each potential well and how much is retained – involves complex gas dynamics. We need also to understand how the baryonic component behaves during mergers. If we trace back the history of the large haloes in Figure 10, half have experienced a merger since z = 2.

Theoretical fashions are often transient. But the cold dark matter model (Peebles 1982, Bond and Szalay 1983, Blumenthal *et al.* 1984 and references cited therein) has survived for more than 5 years. Insofar as it can account for galaxies and their haloes it offers circumstantial support for the idea that the dark matter is in WIMPs or axions. But this evidence is only circumstantial. The nature of the dark matter is still an open question. I am personally agnostic and would bet 25 % on Jupiters, 25 % on black holes, 25 % on WIMPs or other cold dark matter, leaving the remaining 25 % for things not yet thought of.

All things considered, the existence of dark matter is unsurprising. There are all too many forms it could take, and the aim of theorists and observers alike must be to narrow down the range of options. What is encouraging is that various lines of observations, experiments, and theoretical modelling should over the next few years do just this.

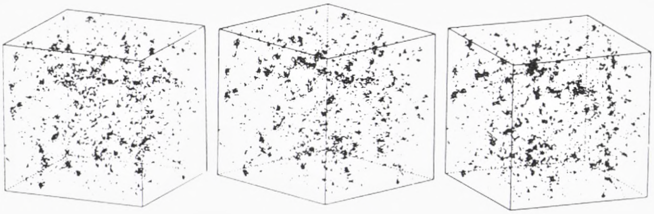

Figure 11. Three views to illustrate the spatial structure within a simulated cubical volume of the expanding universe. In this model the initial fluctuations were Poissonian, with amplitude proportional to $M^{-1/2}$.

It would be specially interesting if we could, by astronomical methods, discover some fundamental particle which has been predicted by theorists. If such particles turned out to account for the dark matter, we would however have to view the galaxies, the stars, and ourselves, in a downgraded perspective. Copernicus dethroned the Earth from any central position. Early this century, Shapley and Hubble demoted us from any privileged location in space. But now even baryon chauvinism might have to be abandoned: the protons, neutrons, and electrons of which we and the entire astronomical world are made could be a kind of after-thought in a cosmos where photinos or neutrinos control the overall dynamics. Great galaxies could be just a puddle of sediment in a cloud of invisible matter ten times more massive and extensive.

VII. *Active Galactic Nuclei*

It has now been realised for 30 years that not all galaxies are 'mere' assemblages of stars and gas, but that in some there is a central power source – manifested as a quasar, radio galaxy, or Seyfert galaxy – which involves release of gravitational energy in and around a supermassive compact object (see *e.g.* Rees 1984 for a review). Close to such an object, a Newtonian '$1/r$' potential no longer applies. Indeed it is here, in the deepest part of the potential well, that the main power output is generated. We must therefore take the inherently relativistic features of the gravitational field into account.

The physics of dense star clusters and of supermassive objects are complex and poorly understood. In contrast, the final state of such a system, when gravitational collapse occurs, is comparatively simple, at least if we accept general relativity. According to the so-called 'no-hair' theorems, the endpoint of a gravitational collapse, however messy and asymmetric it may have been, is a standardised black hole characterised by just two parameters, mass and spin, and described exactly by the Kerr metric. If the collapse occurred in a violent or sudden way, it would take a few dynamical timescales for the black hole to settle down, and during that period gravitational waves would be emitted. But the final stationary state would be described by the Kerr solution, provided only that the material left behind outside the hole did not provide a strong gravitational pertubation.

Such models can, in broad terms, provide acceptable models for quasars. However, one cannot reliably predict the spectrum, nor whether the radiation is thermal or non-thermal; the hardest thing to estimate is what fraction of the power dissipated by viscous friction would go into relativistic particles (via shocks, magnetic reconnection, etc.) rather than being shared among all the particles. Nor do we know how steady or stable the inflow pattern might be. This is a topic where detailed numerical simulations would be worthwhile, particularly if these allowed us to treat unsteady accretion, non-axisymmetric instabilities, and realistic radiative emission and transfer processes.

Despite the lack of quantitative understanding of AGNs in general, the strong radio galaxies (such as Cygnus A) have a distinctive property that offers a clue to their central mechanism. The remarkable feature of these particular AGNs is that the 'kinetic' power required to energise the extended radio lobes (transmitted by the jets in the form of relativistic particles or Poynting flux) exceeds the radiative luminosity of the nucleus itself. Is there a mechanism that could generate an intense plasma outflow, even if the accretion rate and nuclear luminosity were low?

There is indeed another possible source of power over and above the gravitational energy released by infalling matter. The part of a spinning black hole's rest mass that is associated with its spin can in principle be extracted, as was first emphasised by Penrose (1969). By exploiting the analogy between a black hole's horizon and an electrical conductor, Blandford & Znajek (1977) suggested a realistic astrophysical context whereby electromagnetic torques can extract this energy, rather as the unipolar inductor mechanism brakes an ordinary spinning conductor. Three conditions are necessary, all of which can be fulfilled if the hole is surrounded by a small amount of plasma (as could result from low-level accretion).

(i) Magnetic fields threading the hole must be maintained by an external current system. The requisite flux could have been advected in by slow accretion; even if the field within the inflowing matter were tangled, around the hole it would nevertheless be well ordered. The surrounding plasma would be a good enough conductor to maintain surface currents that could confine such a field within the hole's magnetosphere. The only obvious upper limit to the field is set by the requirement that its total energy should not exceed the gravitational binding energy of the infalling gas.

(ii) There must be a current flowing into the hole. Although the relativistic plasma expected around the hole when $\dot{m}$ is low radiates very little, it emits some bremsstrahlung γ-rays. Some of these will interact in the funnel to produce a cascade of electron-positron pairs, yielding more than enough charge density to 'complete the circuit' and carry the necessary current; enough, indeed, to make the magnetosphere essentially charge-neutral, in the sense that $|(n_+ - n_-)| \ll (n_+ + n_-)$, so that relativistic m.h.d. can be applied.

(iii) The proper 'impedance match' must be achieved between the hole and the external resistance. Phinney (1983) has explored the physics of the relativistic wind whose source is the pair plasma created by $\gamma + \gamma \rightarrow e^+ + e^-$ in the hole's magnetosphere, which flows both outward along the funnel and into the hole. He finds consistent wind solutions in which $\sim$ ½ of the hole's spin energy is transformed into Poynting flux and a relativistic electron-position outflow. [The ubiquity of pair-dominated plasmas in strong compact objects is, incidentally, something which has only been properly appreciated in recent years; this subject owes much to the contributions of NORDITA scientists].

The general scheme is depicted in Figure 12. Even a low-level and inefficient

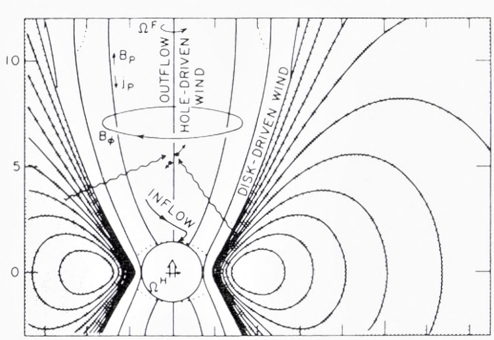

Figure 12. Schematic illustration of the hypothetical 'central engine' in radio sources (Rees *et al.* 1982; Phinney 1983). Interaction between a Kerr hole and the magnetic field generates a hydromagnetic wind. External plasma (stippled) confines a poloidal magnetic field B, of strength 10^3-$10^4 G$, to the hole. The precise geometry is unimportant; that shown is appropriate for a pressure-supported torus with constant specific angular momentum. γ-rays (wavy lines) radiated by the external plasma create pairs in the otherwise empty volume near the symmetry axis from which accreting material is excluded. On field lines which cross the event horizon, these pairs carry a current which extracts rotational energy from the hole in the form of a direct-current Poynting flux.

accretion flow can 'anchor' a magnetic field that threads the hole, and thereby tap the hole's spin energy; in these conditions the extracted power naturally goes predominantly into a relativistic bifurcated outflow. This mechanism seems specially appropriate for strong radio galaxies such as Cygnus A (Rees *et al.* 1982) – objects where the energy flowing along the jets dominates the radiative output of the AGN itself. Electron-positron pairs moving with Lorentz factors $\sim$ 100 would transport some kinetic energy, but most of the power outflow would initially be in the form of Poynting flux associated with the magnetic field coiled round the jet axis, and 'frozen in' to the pair plasma. This Poynting flux may be converted into fast particles where the jet encounters ambient material (perhaps on the scale of the VLBI radio components).

According to this general idea, radio galaxies harbour massive black holes formed long ago via catastrophic collapse (maybe during a quasar phase of activity). The holes lurked quiescent, the galaxy being swept clean of gas, for billions of years. Then some event (perhaps interaction with a companion) triggered renewed infall; maybe at a low rate, but sufficient to reactivate the nucleus by applying a magnetic field. This 'engages the clutch', tapping the hole's latent spin energy, and converting it into non-thermal directed outflow (Poynting flux and $e^+ - e^-$ plasma), which ploughs its way out to scales $\sim 10^{10}$ times larger. If this is indeed what happens in Cygnus A and M87, then these very large-scale manifestations of AGN activity could offer the most direct evidence for inherently relativistic effects.

How much energy is available? Up to 29 % of a hole's mass can be extracted in principle; models for relativistic winds with realistic efficiency factors allow conversion of rest mass with at least a few per cent efficiency – more than enough to power the largest radio galaxy, provided that the hole mass is greater than 10^8 $M_\odot$.

So there are two quite distinct ways in which massive black holes can generate a high luminosity: straightforwardly by accretion; or via the electromagnetic process just described, where the energy comes from the hole itself. The latter process tends to give purely non-thermal phenomena, whereas accretion yields an uncertain mixture of thermal and non-thermal power. The properties of an AGN must depend, among other things, on the relative contributions of these two mechanisms, which are functions of $\dot{M}$ and the spin of the hole. Other parameters may also be relevant, such as the nuclear mass M, the orientation and properties of the host galaxy, etc.

VIII. *What can we realistically expect from Theories and Models?*

How will we ever really know whether there are black holes in AGNs? The evidence can never be more than circumstantial. But we should not be too downcast by that. After all, the evidence that the Sun is powered by nuclear fusion, a cherished dogma never seriously contested, is also really just circumstantial. However, the confrontation of theory with observations, indirect even for stars, is more ambiguous still for quasars: in stars, energy percolates to the observable surface in a relatively steady, symmetric, and well-understood fashion; but in galactic nuclei it is reprocessed into all parts of the electromagnetic spectrum on scales spanning many powers of 10, in a way that depends on poorly-known environmental and geometrical effects in the host galaxy.

The generic black hole model is not infinitely flexible, and not invulnerable. It could be refuted in at least three ways:

1. by finding very regular periodicities, particularly on timescales below 1 hour;
2. by showing that the central masses were $\ll 10^6\ M_\odot$ in Seyferts or $\ll 10^8\ M_\odot$ in radio galaxies, or;
3. by developing a theory of gravity more convincing than general relativity which prohibits black holes.

But a clean-cut refutation, leading to abandonment of some theory, happens only rarely in astrophysics; many models have persisted unrefuted for a long time. A cynic might argue that they have survived only because they don't go beyond generalities, or else because their proponents have been adept at replacing or modifying faulty parts to keep shaky old models 'roadworthy'. Such a cynical attitude is not necessarily justified, and to explain why I must digress briefly into methodology. The way we are told science is done is like this: the data suggest a model, which suggests further tests, whereby the original model is either refuted or refined. Such a simple scheme is realistic in, for instance, particle physics, where the fundamental entities may be exactly reducable to a few basic constants and equations. But other sciences deal with inherently complex phenomena and no theoretical scheme can be expected to account for every detail. In geophysics, for instance, the concepts of continental drift and plate tectonics have undoubtedly led to key advances; but they cannot be expected to explain the shape – the precise topography – of the continents. What we should aim to do, in our attempts to understand quasars, is focus on those features of the data which genuinely test crucial ideas, and not to be diverted into measuring or modelling something which is accidental or secondary.

IX. *The AGN Population*

Ideally, one would like a unified model that explains the multifarious types of AGN in the same way that our theories for the Hertzsprung-Russell diagram do this for stars. Still more ambitiously, one would like to place quasars in the general context of galactic evolution. Do quasars die and get resurrected as radio galaxies? How and by what route does a condensed mass first accumulate in the centres of galaxies? What are the masses involved? How do the pyrotechnics in the nucleus react back on the rest of the galaxy? How common is it for massive holes, remnants of past activity, to lurk quiescent in the centres of normal galaxies? Already we have some clues from quasar demography and from the study of nearby galaxies.

Even if AGNs are precursors on the route towards black hole formation rather than involving black holes that have already formed, massive black holes should exist in profusion as remnants of past activity; they would be inconspicuous unless infall onto them recommenced, and generated a renewed phase of accretion-powered output or catalysed the extraction of latent spin energy.

Let us consider further the collective properties of quasars. The total energy output from all quasars in known to within a factor of order 2: it is $\sim$ 3000 $M_\odot$ c^2 per Mpc^3, about half coming from quasars with apparent magnitudes in the range 19-21. Galaxies contribute $\sim 10^5$ in the same units, and the microwave background $\sim 7.5 \times 10^6$. So, even though quasars may influence their host galaxies, they are collectively rather modest contributors to the cosmic energy budget, because of their low space density. (They may nevertheless have a crucial cumulative effect on the entire intergalactic medium, because their energy emerges largely in forms such as an ionizing continuum and high velocity jets and winds.)

The prime era of quasar activity is at $z = 2$ or 3. It is from these redshifts that most of the quasar background light originates. The population thereafter decays on a timescale of order $t_{\mathrm{Evo}} = 2 \times 10^9$ years. This is, however, merely an *upper limit* to the lifetime t_0 of each object: many generations of individual quasars could be born, and could die, in the period over which the population declines. Several important numbers depend on what t_0 actually is: the mass and the number of quasar remnants,

Table 1

(i)	(ii)
$t_0 \simeq t_{Evo}$	$t_0 \simeq 4 \times 10^7$ yr $\simeq 0.02\, t_{\mathrm{Evo}}$
$M = 2.5 \times 10^9\ L_{46}\ \varepsilon_{0.1}^{-1}\ M_\odot$	$M = 5 \times 10^7\ L_{46}\ \varepsilon_{0.1}^{-1}\, M_\odot$
$L << L_{\mathrm{Ed}}$	$L \simeq L_{\mathrm{Ed}}\ \varepsilon_{0.1}^{-1}$
Broad–line regions gravitationally bound	Broad–line region *not* gravitationally bound
Very massive remnants in $\sim$2% of galaxies	$\sim 10^8$ $M_\odot$ remnants in most bright galaxies

the ratio of their luminosity to the Eddington limit, and the issue of whether the broadline region can be gravitationally bound.

Table 1 brings out these points. It contrasts two hypotheses: (i) that there was, in effect, only one generation of quasars, which were long-lived and massive; *versus* (ii) that there were $\sim$ 50 generations of quasars, so that their individual masses (for a given efficiency) need not have built up to such high values, and quasar remnants would be more numerous. In reality t_0 would not be a single number, but there would be a spread of ages, perhaps correlated with luminosity L.

X. *Evidence of Physical Conditions at Z $\gtrsim$ 2: the first Quasars*

There are four lines of evidence which imply that conditions at redshifts of 2-4 were very different from those at the present epoch. These are:

(a) *Quasars.* These are now observed at redshifts exceeding 4. They offer evidence on the time-dependence of galactic activity. It has long been known that at z $\simeq$ 2 the universe was more violently active, in the sense that the comoving density of powerful AGNs was then much higher. Enough high-z objects have now been found by well-defined search procedures to allow provisional attempts to infer how the luminosity function has evolved. These results, which should still be interpreted cautiously, suggest that there may be a decrease beyond z $\simeq$ 2.5. There is, however, no incontrovertible evidence for a really sharp cut-off; and even if there was, it would be hard to decide whether this indicated a cut-off in the sources themselves, or if they became undetectable beyond a certain redshift because of stifling effects within the host galaxy (Rees 1986) or absorption in the intervening intergalactic medium by neutral gas (Rees 1969) or by dust (Ostriker and Heisler 1984). The first bound systems may have formed even earlier than the first quasars: different cosmogonic models (see Figure 9) place the first non-linear behaviour anywhere from z $\simeq$ 4 to z $\simeq$ 1000.

(b) *Quasars as probes of intervening gas.* The true nature of the gas that gives rise to multiple absorption redshifts in quasar spectra remains controversial. There can be no doubt, however, that the inferred gas clouds have great relevance to galaxy formation. Progress in interpreting these absorption features is, fortunately, not stymied by our poor understanding of the intrinsic properties of quasars: the quasar itself merely serves as a probe of material along the entire line of sight. Some of the absorption lines are caused by HI with a column density as high as 10^{21}cm^{-2} (Wolfe *et al.* 1986), and could involve a protogalaxy or a protogalactic disc. The much larger number of 'clouds' responsible for the Lyman α forest involve much lower HI columns; the relationship of these to galaxies is less clear. Although the details are still controversial, quasar spectra certainly reveal that the amount of HI increases with z over the range 1.8-3.8 where Lyman α can be studied from the ground using known quasars. It is unclear to what extent this implies that the amount of gas is increasing with z, rather than that the ionization level is decreasing with z.

(c) *Direct imaging of (proto)galaxies at high z*. Several extended emission-line objects have been recently observed, which are probably gas-rich galaxies at an early stage in their evolution (see Djorgovski 1988 and Cowie 1988 for reviews).

(d) *Radio structures*. Recent evidence from the radio – the band that can offer the most detailed structural information on high redshift objects – suggests that AGNs were not merely more numerous at high redshifts, but that their radio morphologies were qualitatively different. Barthel and Miley (1988) find that the high redshift radio sources, whether categorised as quasars or as radio galaxies, are more distorted than local sources of similarly high power. (At small redshifts, the powerful sources tend to be symmetric doubles; only those of lower power display strong bending or asymmetry). This suggests that at high z the medium around the sources is more disturbed; it might also suggest that these sources involve interacting galaxies (see, however, Gopal-Krishna and Wiita 1987). Studies of objects at lower redshift suggest that the rarer renewed activity in nuclei at recent epochs is triggered by close encounters or mergers.

XI. *When was Galaxy Formation completed?*

Redshifts $z \simeq 3-4$ correspond to a cosmic time $t \simeq 10^9$ yrs. The first bound systems *may* (though they need not) have formed *much* earlier than this. However, a general argument allows us to infer that galaxy formation would not have been completed much before this time: it should therefore not surprise us to find that conditions at $z \simeq 3$ were indeed very different from those prevailing now.

The luminous parts of many galaxies are now known to be embedded in dark halos extending out to $R \sim 100$ kpc. The free-fall time for protogalactic material is of order $(2GM/R^3_{turn})^{-1/2}$, where R_{turn} is its radius at turnaround. The collapse factor before a non-dissipative system virialises is ~ 2 (though it can, of course, be larger when radiative cooling permits dissipation). If galaxies were only 10 kpc in size, they could therefore have formed on a timescale of 10^8 years from material that 'turned around' at redshift $z \gtrsim 20$. But the collapse phase of galaxies whose diffuse halos extend out beyond 100 kpc must have taken much longer.

The mean density of the Universe at the turnaround time of the halo material must be lower than the density of the proto-halo itself (by a factor 5.5 in the simplest model for a spherical protogalaxy), so a system whose mean density at turnaround was very low cannot have collapsed until a correspondingly late epoch. The low densities and long dynamical timescales in the outer parts of halos therefore have the crucial implication – irrespective of what these halos actually consist of – that galaxy formation (and, more specifically, the formation of the outer parts of disc galaxies) was not completed until the Universe was $\gtrsim 2.10^9$ years old: *i.e.* not until redshifts $z \simeq 2$, even if it started much earlier. If the angular momentum of discs were acquired via tidal torques (Figure 13) between neighbouring protogalaxies, the 'spin up' must have occurred at a radius $\gtrsim$ 10 times exceeding the present radius. This is a separate argument in favour of 'recent' galaxy formation, at least for disc galaxies (Fall and Efstathiou, 1980; Gunn, 1982).

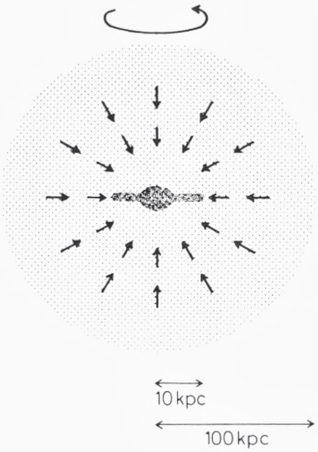

Figure 13. Tidal torques arise between neighbouring protogalaxies, because these typically have non-zero quadrupole moments, but these typically impart less than 10 per cent of the rotational velocity needed for centrifugal support. The material that ends up in a disc of radius 10 kpc must therefore have fallen in from $\gtrsim$ 100 kpc. This infall timescale is $\sim 10^9$ yrs, so the formation of discs cannot have been completed before the epoch corresponding to a redshift of 2 or 3.

Quasar activity peaks at $z \simeq 2$; thereafter the comoving density of luminous quasars declines rapidly as the universe expands, on a timescale much shorter than the expansion timescale. The observed number density at $z \lesssim 2$ is so much smaller than the number density of bright galaxies that the formation of an ultra-luminous quasar within a bright galaxy must be a transient event and therefore sensitive to some (as yet not well understood) feature of the system.

XII. *Black Hole Remnants in Quiescent Galaxies*

Since quasars may involve an atypical subset of galaxies, we must be cautious about inferring anything about typical galaxies from the quasar redshift distribution. The same is true of radio galaxies, because they are exceptional too. Until recently, hardly anything was known about *ordinary* galaxies sufficiently far back in time for evolutionary changes to really show up. But large telescopes and more sensitive detectors are changing this. Images can now go so faint that there are 100,000 galaxies per square degree, and counts can be compiled down to fainter than 26 magnitude. These counts cannot be uniquely modelled in the absence of knowledge of the redshift. But models suggest that the dominant faint galaxy population may be being seen at the stage when they are acquiring disks. We must await the sharper images the ST will give to test this hypothesis. Next-generation telescopes should give us snapshots of galaxies at different redshifts, (different epochs), thereby allowing us to check how galactic evolution actually occurred.

These faint galaxies vastly outnumber quasars and radio sources, which could mean one of two things (see Table 1). *Either* a very small fraction of galaxies have long-lived active nuclei; *or* more do, but the activity represents a relatively brief phase in each galaxy's life history. Most authors favour something closer to this second option, because if individual quasars were too long-lived, they would build up to unacceptably large masses. If there were many generations of quasars, we would expect that dead quasars, massive black holes now starved of fuel, should lurk in the nuclei of many nearby galaxies.

Evidence on the masses, and frequency of occurrence, of black holes in present-day galaxies therefore offers important clues to quasar lifetimes, and to the relationships between different classes of AGNs. The dynamics of stars in the inner regions of nearby galaxies such as M31 and M32 indicate the presence of central concentrated dark masses (Tonry 1984, Kormendy 1988, Dressler and Richstone 1988), of $\sim 10^7$ $M_\odot$, and one would like to know whether these might indeed be massive black holes. Since the number of surrounding stars, and their motions, are roughly known, the capture rate by the putative central hole can be estimated; the question then arises of whether the apparent quiescence of the nuclei of these galaxies is compatible with a massive hole's presence. The answer depends on what happens to the debris from each disrupted star. How much is accreted or expelled and with what associated luminosity or radiative efficiency? And, more specifically, *how long* does it take to digest or expel the debris from one star, in comparison with the interval between successive captures? If the stellar density in the central few parsecs is known, it is in principle straightforward to estimate how often a star gets close enough to the central hole to be tidally disrupted (or at least captured). The debris from one solar-mass star per ten thousand years, swallowed steadily with 10 per cent radiative efficiency, would yield a luminosity of 6.10^{41} erg s^{-1} – higher than is observed from M31 or M32.

Several features of the stellar capture process have as yet not been fully analysed:

(i) What fraction of the debris goes down the hole, rather than being expelled?

(ii) What is the radiative efficiency for the accretion process? In other words, how many ergs are radiated for each gram that is swallowed?

(iii) How long do viscous processes, etc., take to 'process' the debris?

The fate of the debris is discussed more fully elsewhere (Rees 1988). The bulk of the bound debris would be swallowed or expelled *rapidly* compared with the interval between successive stellar captures – the only conspicuous luminosity being a flare (predominantly of thermal UV X-ray emission) with $L = L_E$, the hole's Eddington luminosity, fading within a few years. The integrated output from this flare could in principle amount to a few per cent of the star's rest mass, but would probably be far less, because most debris would be 'fed' to the hole far more rapidly than it could be accepted if the radiative efficiency were high; much of the bound debris would then escape in a radiatively-driven directed outflow or be swallowed with low radiative efficiency.

Black holes formed at the era of peak quasar activity ($z \sim 2$) could be reactivated, perhaps as radio galaxies or Seyferts, if the host galaxy were disturbed by a merger. Otherwise they would be quiescent, but not quite. Now and again a star would wander so close that tidal forces ripped it apart. We would then see a flare persisting for as long as it took the debris to be swallowed or expelled, maybe a year or so. Searches for such a phenomenon would be a crucial test of the reality of these quiescent black holes.

XIII. *Conclusions*

There is darkness at the centre of even the most familiar galaxies. Moreover, 90 % of the gravitating stuff that binds them may be a dark relic of the hot early phases of the big bang, whose elucidation transcends the physics we understand, and perhaps points to new links between the cosmos and the microworld.

I argued earlier that the mundane physics of gas cooling and Newtonian collapse singles out a galactic mass and lengthscale, so that a favoured mass need not be imprinted *ab initio*. But there must have been some initial fluctuations. Otherwise the universe would still be amorphously uniform, with no galaxies, no stars, and no astronomers. There is still no agreed understanding of why the universe combines the small-scale roughness needed to initiate galaxies with the large-scale uniformity that has allowed it to expand smoothly for 10 billion years.

The various problems of large-scale cosmogony are so intermeshed that we will not really solve any until the whole picture comes into sharper focus. For instance, we cannot test theories of galaxy formation and evolution until we understand the gas dynamics of star formation, and the possible rôle of active nuclei, as well as the exotic physics of the initial fluctuations.

The empirical data – observations in all wavebands, and laboratory experiments as well – are burgeoning and all advancing the subject. And theorists are injecting a range of not necessarily compatible ideas whose vector sum at least pushes the subject forward. Hubble's great book, 'The Realm of the Nebulae', concludes with these words. 'With increasing distance our knowledge fades and fades rapidly. Eventually we reach the dim boundary, the utmost limits of our telescope. There we measure shadows, and we search among ghostly errors of measurement for landmarks that are scarcely more substantial. The search will continue. Not until the empirical resources are exhausted need we pass on to the dreamy realm of speculation.'

This search *has* continued as more powerful telescopes and detectors have been deployed. Observers have colonised the speculators' former territory, and theory itself now has a speculative range undreamt of by Hubble's contemporaries. The origin of the nebulae, and the emergence of cosmic structure, are still mysterious but the key questions are at least in clearer focus.

I am grateful to many colleagues for discussions and collaboration on topics mentioned in this talk, and to Judith Moss for her careful preparation of the typescript. Some of this written material is adapted from an invited discourse presented at the IAU General Assembly in August 1988.

References

Barthel, P. D. and Miley, G. K. 1988, *Nature* **333**, 319.
Binney, J. and Tremaine, S. 1987, *Galactic Dynamics*, (Princeton U.P.).
Blandford, R. D. and Znajek, R. L. 1977, *Mon. Not. R. astr. Soc.* **169**, 395.
Blumenthal, G. R., Faber, S. M., Primack, J. R. and Rees, M. J. 1984, *Nature* **311**, 517.
Bond, J. R. and Szalay, A. S. 1983, *Astrophys. J.* **274**, 443.
Carr, B. J., Bond, J. R. and Arnett, W. D. 1984, *Astrophys. J.* **277**, 445.
Cowie, L. L. 1988. In *'The Post-Recombination Universe'*, ed. N. Kaiser and A. Lasenby (Reidel).
Cowsik, R. and McClelland, J. 1973, Astrophys. J. **180**, 7.
Djorgovski, G. 1988, in *'Towards Understanding Galaxies at Large Redshift'*, eds. A. Renzini and R. Kron, (Reidel).
Dressler, A. and Richstone, D. O. 1988, *Astrophys. J.* **324**, 701.
Einasto, J., Kaasik, A. and Saar, E. 1974, *Nature* **250**, 309.
Fall, S. M. and Efstathiou, G. P. 1980, *Mon. Not. R. astr. Soc.* **193**, 189.
Frenk, C. S., White, S. D. M., Davis, M. and Efstathiou, G. P. 1985, *Nature* **317**, 595.
Gopal-Krishna and Wiita, P. J. 1987, *Mon. Not. R. astr. Soc.* **226**, 531.
Gott, J. R. 1981, *Astrophys. J.* **243**, 140.
Gunn, J. E. 1982, in *'Astrophysical Cosmology'*, ed. H. A. Bruck *et al.* (Vatican Publications).
Kormendy, J. 1988, *Astrophys. J.* **325**, 128.
Lyubimov, V. A., Novikov, E. G., Nozik, V. Z., Tretyakov, E. F., and Kozek, V. S. 1980, *Phys. Lett.* **394**, 266.
Marx, G. and Szalay, A. S. 1972, *Proc. Neutrino 72, Technoinform*, Budapest, p. 191.

Ostriker, J. P. and Heisler, J. 1984, *Astrophys. J.* **278**, 1.
Ostriker, J. P., Peebles, P. J. E. and Yahil, A. 1974, *Astrophys. J. Letters* **193**, L.1.
Peebles, P. J. E. 1982, *Astrophys. J. Letters* **263**, L.1.
Penrose, R. 1969, *Revista Nuovo Cim* **1**, 252.
Phinney, E. S. 1983, Cambridge Ph.D. thesis.
Press, W. H. and Gunn, J. E. 1973, *Astrophys. J.* **185**, 397.
Primack, J. R., Seckel, D. and Sandoulet, B. 1988, *Ann. Rev. Nucl. and Particle Sci.* **38**, 751.
Rees, M. J. 1969, *Astrophys. Lett.* **4**, 61.
Rees, M. J., Begelman, M. C., Blandford, R. D. and Phinney, E. S. 1982, *Nature* **295**, 17.
Rees, M. J. 1984, *Ann. Rev. Astr. Astrophys.* **22**, 471.
Rees, M. J. 1986. in *'Structure and Evolution of AGNs'*, eds. G. Giuricin *et al.* (Reidel) p. 447.
Rees, M. J. and Ostriker, J. P. 1977, *Mon. Not. R. astr. Soc.* **179**, 541.
Rees, M. J. 1988, *Nature* **333**, 523.
Refsdal, S. 1970, *Astrophys. J.* **159**, 357.
Tonry, J. 1984, *Astrophys. J. Letters* **283**, L27.
Wolfe, A. M., Turnshek, D. A., Smith, H. E., and Cohen, R. D. 1986. *Astrophys. J. Suppl.* **249**, 304.
Zel'dovich, Y. B. 1982, *Highlights of Astronomy* **6**, 29.

Submitted to the Academy August 1989
Published May 1990