NOBEL LECTURES

PHYSICS

1981 – 1990

World Scientific

NOBEL LECTURES IN PHYSICS

1981-1990

NOBEL LECTURES

Including Presentation Speeches And Laureates' Biographies

PHYSICS

CHEMISTRY

PHYSIOLOGY OR MEDICINE

LITERATURE

PEACE

ECONOMIC SCIENCES

NOBEL LECTURES

INCLUDING PRESENTATION SPEECHES
AND LAUREATES' BIOGRAPHIES

PHYSICS

1981-1990

EDITOR-IN-CHARGE

TORE FRÄNGSMYR

EDITOR

GÖSTA EKSPONG

World Scientific
Singapore • New Jersey • London • Hong Kong

Published for the Nobel Foundation in 1993 by

World Scientific Publishing Co. Pte. Ltd.
P O Box 128, Farrer Road, Singapore 9128
USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661
UK office: 73 Lynton Mead, Totteridge, London N20 8DH

NOBEL LECTURES IN PHYSICS (1981-1990)

ISBN 981-02-0728-X
ISBN 981-02-0729-E (pbk)

Printed in Singapore by Continental Press Pte Ltd

Foreword

Since 1901 the Nobel Foundation has published annually "Les Prix Nobel" with reports from the Nobel Award Ceremonies in Stockholm and Oslo as well as the biographies and Nobel lectures of the laureates. In order to make the lectures available to people with special interests in the different prize fields the Foundation gave Elsevier Publishing Company the right to publish in English the lectures for 1901-1970, which were published in 1964-1972 through the following volumes:

Physics 1901-1970	4 volumes
Chemistry 1901-1970	4 volumes
Physiology or Medicine 1901-1970	4 volumes
Literature 1901-1967	1 volume
Peace 1901-1970	3 volumes

Elsevier decided later not to continue the Nobel project. It is therefore with great satisfaction that the Nobel Foundation has given World Scientific Publishing Company the right to bring the series up to date.

The Nobel Foundation is very pleased that the intellectual and spiritual message to the world laid down in the laureates' lectures will, thanks to the efforts of World Scientific, reach new readers all over the world.

Lars Gyllensten
Chairman of the Board

Stig Ramel
Executive Director

Stockholm, June 1991

PREFACE

The present volume contains the lectures by the Nobel laureates in physics during the years 1981-1990. Included for each year is also a translation of the presentation speech by a member of the Nobel Committee for Physics within the Royal Academy of Sciences delivered during the ceremony on December 10 - the day of Alfred Nobel's death. The autobiographies of the laureates are reprinted with updated information according to the wishes of the respective author.

The number of laureates is 23, since during the decade, three prizewinners were honoured during each of the five years 1981, 1986, 1988, 1989, and 1990, two prizewinners in 1983, 1984, and 1987, whereas a single laureate was awarded the prize in 1981 and again in 1985.

Although there is no rule of rotation between the fields of physics, it is interesting in retrospect to note the following. Atomic physics was in focus in 1981 with the prize going to Nicolaas Bloembergen, Arthur Schawlow and Kai Siegbahn and again in 1989 to Norman Ramsey, Hans Dehmelt and Wolfgang Paul. In 1983 the prize to Subrahmanyan Chandrasekhar and William Fowler was awarded for researches in astrophysics, while achievements within the field of elementary particle physics was the motivation on three occasions, in 1984 to Carlo Rubbia and Simon van der Meer, in 1988 to Leon Lederman, Melvin Schwartz and Jack Steinberger, and in 1990 to Jerome Friedman, Henry Kendall and Richard Taylor. Discoveries or inventions within solid state physics appeared four times in the citations, namely in 1982 to Kenneth Wilson, in 1985 to Klaus von Klitzing, in 1986 to Ernst Ruska, Gerd Binnig and Heinrich Rohrer and in 1987 to Georg Bednorz and Alex Müller. The last two mentioned prizes went in consecutive years not only to the same field of physics but also to people working in the same research institute, something unique in the history of the Nobel prizes in physics.

The division of physics into subfields as done above is debatable, since the borderlines are not sharp. Take the case of lasers, which have importance and applications within large areas of science and technology, or take that of the electron microscope, which is of similar wide ranging importance. A new, important method or invention may be first used in conjunction with making a new discovery, but it may later turn out that the method or invention has much broader, significant applications within or outside the initial area of research.

Gösta Ekspong

CONTENTS

Physics 1981

NICOLAAS BLOEMBERGEN, ARTHUR L SCHAWLOW

for their contribution to the development of laser spectroscopy

and

KAI M SIEGBAHN

for his contribution to the development of high-resolution electron spectroscopy

THE NOBEL PRIZE FOR PHYSICS

Speech by Professor INGVAR LINDGREN of the Royal Academy of Sciences. Translation from the Swedish text

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen,

This year's Nobel prize in physics is shared between three scientists-Nicolaas Bloembergen and Arthur Schawlow, both from the United States, and Kai Siegbahn from Sweden-for their contributions to the development of two important spectroscopic methods-laser spectroscopy and electron spectroscopy.

Both of these methods are based upon early discoveries of Albert Einstein. One of the major problems for physicists of the last century was to explain, with the "classical" concepts, the so-called photoelectric effect, i.e. the emission of electrons from a metal surface irradiated with light of short wavelength. In 1905 Einstein explained this phenomenon in a simple and elegant way, using the quantum hypothesis introduced by Max Planck five years earlier. According to this model, light is a wave motion, but it is quantized, i.e. it is emitted in small pieces - light quanta or photons -which in some respects behave like particles. This discovery was the first foundation stone to be laid in the building of the "new" physics - the quantum physics - which was to develop rapidly during the first decades of this century.

The photoelectric effect is the basis of the spectroscopy which Kai Siegbahn has developed together with his collaborators in Uppsala. When a photon of high energy-e.g. from an X-ray tube-hits an atom, it can penetrate deeply into the atom and expel an electron. By analyzing the electrons expelled in this way, it is possible to extract valuable information about the interior of the atom. Early experiments of this kind where performed in the second decade of this century, but the method was not sufficiently developed to probe the atomic structure until the 1950's. At that time Kai Siegbahn had for a number of years developed more and more sophisticated instruments for analyzing electrons emitted at the decay of certain radioactive nuclei-so called beta decays. When he and his collaborators applied this technique to analyze the electrons emitted in the photoelectric process, the new era of electron spectroscopy was born.

With this spectroscopy it became possible to determine the binding energy of atomic electrons with higher accuracy than was previously possible. This was of great importance for testing new atomic models and computation schemes, which were being developed at the same time, partly due to the simultaneous rapid development of computers. It was furthermore found that the electronic binding energy was to some extent dependent upon the chemical environment of the atom, and this led to a new method for chemical analysis-ESCA-which stands for "Electron Spectroscopy for Chemical Analysis". Nowadays, this method is being applied at hundreds of laboratories all over the world, particu-

larly in the investigation of surface reactions, such as corrosion and catalytic reactions, i.e. reactions where a substance may initiate or stimulate a chemical reaction, seemingly without taking part in it. Such reactions are of vital importance for the process industry, and the spectroscopy developed by Siegbahn and his collaborators can be of great help in our efforts to understand processes of this kind.

The second form of spectroscopy which is awarded with this years's Nobel prize-the laser spectroscopy-is based upon another early discovery of Einstein. It had been known for a long time that atoms and molecules could absorb light as well as spontaneously emit light of certain wave lengths. In 1917 Einstein found that light could also *stimulate* atoms or molecules to emit light of the same kind. This is the basic process in the laser. Photons emitted in such a stimulated process have not only the same wave length but they also oscillate in phase with each other. We call light of this kind *coherent.*

Coherent light could be compared with a marching military troop-where the soldiers correspond to the photons - while non-coherent light in this model could be compared with people on a busy shopping street on a Saturday morning. Those who have done their military service know that marching soldiers should keep in step. However, there is one occasion, namely when the troop crosses a small bridge, when it is necessary to break step, otherwise the strong, coherent vibrations of the troop could break the bridge. The situation is similar for the coherent light. Due to the fact that the photons oscillate in phase, such light will have a much stronger effect than incoherent light on the irradiated material, and this gives laser light its very special character.

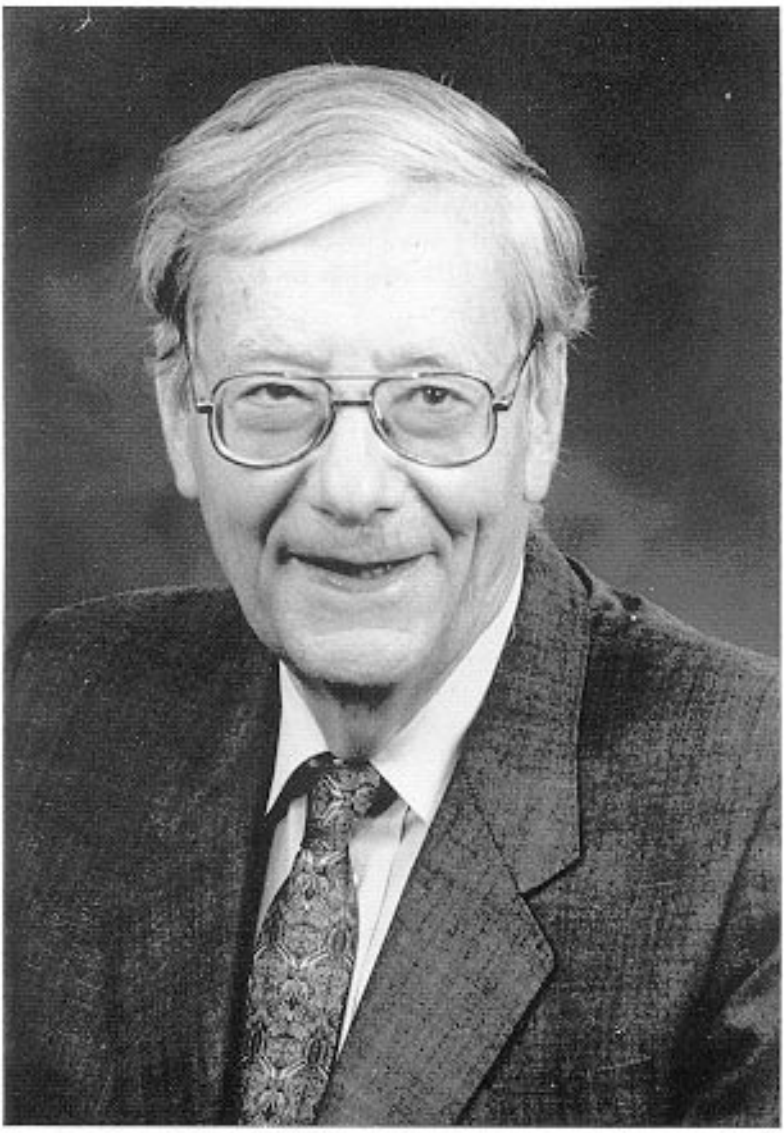
Coherent radiation was first produced in the microwave region, using an instrument we call maser (MASER=Microwave Amplification by means of Stimulated Emission of Radiation). The idea of the maser was conceived in the middle 1950's by the American Charles Townes and by Basov and Prokhorov from the Soviet Union, who shared the Nobel prize in Physics in 1964. Townes and Arthur Schawlow extended the idea of the maser to the optical region-i.e. for visible light-and this led to the construction of the laser two years later (LASER= Light Amplification by means of Stimulated Emission of Radiation). At Stanford University Schawlow has led a research group, which has developed a number of advanced methods, where the laser is used to study the properties of atoms and molecules with extreme accuracy. This has stimulated the development of new theoretical models and improved appreciably our knowledge of these building blocks of matter.

Nicolaas Bloembergen has contributed to the development of laser spectroscopy in a different way. Laser light is sometimes so intense that, when it is shone on to matter, the response of the system could not be described by existing theories. Bloembergen and his collaborators have formulated a more general theory to describe these effects and founded a new field of science we now call *non-linear optics.* Several laser spectroscopy methods are based upon this phenomenon, particularly such methods where two or more beams of laser light are mixed in order to produce laser light of a different wave length. Such methods can be applied in many fields, for instance, for studying combustion

processes. Furthermore, it has been possible in this way to generate laser light of shorter as well as longer wave lengths, which has extended the field of application for laser spectroscopy quite appreciably.

Professor Bloembergen, Professor Schawlow, Professor Siegbahn: you have all contributed significantly to the development of two spectroscopic methods, namely the laser spectroscopy and the electron spectroscopy. These methods have made it possible to investigate the interior of atoms, molecules and solids in greater detail than was previously possible. Therefore, your work has had a profound effect on our present knowledge of the constitution of matter.

On behalf of the Royal Swedish Academy of Sciences I wish to extend to you the heartiest congratulations and I now invite you to receive this year's Nobel prize in Physics from the hands of His Majesty the King!



Nicolaas Bloembergen

NICOLAAS BLOEMBERGEN

My parents, Auke Bloembergen and Sophia Maria Quint, had four sons and two daughters. I am the second child, born on March 11, 1920, in Dordrecht, the Netherlands. My father, a chemical engineer, was an executive in a chemical fertilizer company. My mother, who had an advanced degree to teach French, devoted all her energies to rearing a large family.

Before I entered grade school, the family moved to Bilthoven, a residential suburb of Utrecht. We were brought up in the protestant work ethic, characteristic of the Dutch provinces. Intellectual pursuits were definitely encouraged. The way of life, however, was much more frugal than the family income would have dictated.

At the age of twelve I entered the municipal gymnasium in Utrecht, founded as a Latin school in 1474. Nearly all teachers held Ph.D. degrees. The rigid curriculum emphasized the humanities: Latin, Greek, French, German, English, Dutch, history and mathematics. My preference for science became evident only in the last years of secondary school, where the basics of physics and chemistry were well taught. The choice of physics was probably based on the fact that I found it the most difficult and challenging subject, and I still do to this day. My maternal grandfather was a high school principal with a Ph.D. in mathematical physics. So there may be some hereditary factor as well. I am ever more intrigued by the correspondence between mathematics and physical facts. The adaptability of mathematics to the description of physical phenomena is uncanny.

My parents made a rule that my siblings should tear me away from books at certain hours. The periods of relaxation were devoted to sports: canoing, sailing, swimming, rowing and skating on the Dutch waterways, as well as the competitive team sport of field hockey. I now attempt to keep the body lit by playing tennis, by hiking and by skiing.

Professor L. S. Ornstein taught the undergraduate physics course when I entered the University of Utrecht in 1938. He permitted me and my partner in the undergraduate lab, J. C. Kluyver (now professor of physics in Amsterdam) to skip some lab routines and instead assist a graduate student, G. A. W. Rutgers, in a Ph.D. research project. We were thrilled to see our first publication, "On the straggling of Po-α-particles in solid matter", in print (Physica 7, 669, 1940).

After the German occupation of Holland in May 1940, the Hitler regime removed Ornstein from the university in 1941. I made the best possible use of the continental academic system, which relied heavily on independent studies. I took a beautiful course on statistical mechanics by L. Rosenfeld, did experi-

mental work on noise in photoelectric detectors, and prepared the notes for a seminar on Brownian motion given by J. M. W. Milatz. Just before the Nazis closed the university completely in 1943, I managed to obtain the degree of Phil. Drs., equivalent to a M.Sc. degree. The remaining two dark years of the war I spent hiding indoors from the Nazis, eating tulip bulbs to fill the stomach and reading Kramers' book "Quantum Theorie des Elektrons und der Strahlung" by the light of a storm lamp. The lamp needed cleaning every twenty minutes, because the only fuel available was some left-over number two heating oil. My parents did an amazing job of securing the safety and survival of the family.

I had always harbored plans to do some research for a Ph.D. thesis outside the Netherlands, to broaden my perspective. After the devastation of Europe, the only suitable place in 1945 appeared to be the United States. Three applications netted an acceptance in the graduate school at Harvard University. My father financed the trip and the Dutch government obliged by issuing a valuta permit for the purchase of US$l,850. As my good fortune would have it, my arrival at Harvard occurred six weeks after Purcell, Torrey and Pound had detected nuclear magnetic resonance (NMR) in condensed matter. Since they were busy writing volumes for the M.I.T. Radiation Laboratory series on microwave techniques, I was accepted as a graduate assistant to develop the early NMR apparatus. My thorough Dutch educational background enabled me to quickly profit from lectures by J. Schwingcr, J.H. Van Vleck, E.C. Kemble and others. The hitherto unexplored field of nuclear magnetic resonance in solids, liquids and gases yielded a rich harvest. The results arc laid down in one of the most-cited physics papers, commonly referred to as BPP (N. Bloembergen, E. M. Purcell and R. V. Pound, Phys. Rev. 73, 679, 1948). Essentially the same material appears in my Ph.D. thesis, "Nuclear Magnetic Relaxation", Leiden, 1948, republished by W. A. Benjamin, Inc., New York, in 1961. My thesis was submitted in Leiden because I had passed all required examinations in the Netherlands and because C. J. Gorter, who was a visiting professor at Havard during the summer of 1947, invited me to take a postdoctoral position at the Kamerlingh Onnes Laboratorium. My work in Leiden in 1947 and 1948 resulted in establishing the nuclear spin relaxation mechanism by conduction electrons in metals and by paramagnetic impurities in ionic crystals, the phenomenon of spin diffusion, and the large shifts induced by internal magnetic fields in paramagnetic crystals.

During a vacation trip of the Physics Club "Christiaan Huyghens" I met Deli (Huberta Deliana Brink) in the summer of 1948. She had spent the war years in a Japanese concentration camp in Indonesia, where she was born. She was about to start her pre-med studies. When I returned to Harvard in 1949 to join the Society of Fellows, she managed to get on a student hospitality exchange program and traveled after me to the United States on an immigrant ship. I proposed to her the day she arrived and we got married in Amsterdam in 1950. Ever since, she has been a source of light in my life. Her enduring encouragement has contributed immensely to the successes in my further career. After the difficult years as an immigrant wife, raising three children on

the modest income of a struggling, albeit tenured, young faculty member, she has found the time and energy to develop her considerable talents as a pianist and artist. We became U.S. citizens in 1958.

Our children are now independent. The older daughter, Antonia, holds M.A. degrees in political science and demography, and works in the Boston area. Our son, Brink, has an M.B.A. degree and is an industrial planner in Oregon. Our younger daughter, Juliana, envisages a career in the financial world. She has interrupted her banking job to obtain an M.B.A. in Philadelphia.

In this family setting my career in teaching and research at Harvard unfolded: Junior Fellow, Society of Fellows 1949-1951; Associate Professor 1951-1957; Gordon McKay Professor of Applied Physics 1957-1980; Rumford Professor of Physics 1974-1980; Gerhard Gade University Professor 1980-present. While a Junior Fellow, I broadened my experimental background to include microwave spectroscopy and some nuclear physics at the Harvard cyclotron. I preferred the smaller scale experiments of spectroscopy, where an individual, or a few researchers at most, can master all aspects of the problem. When I returned to NMR in 1951, there were still many nuggets to be unearthed. My group studied nuclear quadrupole interactions in alloys and imperfect ionic crystals, discovered the anisotropy of the Knight shift in noncubic metals, the scalar and tensor indirect nuclear spin-spin coupling in metals and insulators, the existence ofdifferent temperatures of the Zeeman, exchange and dipolar energies in ferromagnetic relaxation, and a variety of cross relaxation phenomena. All this activity culminated in the proposal for a three-level solid state maser in 1956.

Although I was well aware of the applicability of the multilevel pumping scheme to other frequency ranges, I held the opinion - even after Schawlow and Townes published their proposal for an optical maser in 1958 - that it would be impossible for a small academic laboratory, without previous expertise in optics, to compete successfully in the realization of lasers. This may have been a self-fulfilling prophesy, but it is a matter of record that nearly all types of lasers were first reduced to practice in industrial laboratories, predominantly in the U.S.A.

I recognized in 1961 that my laboratory could exploit some of the new research opportunities made accessible by laser instrumentation. Our group started a program in a field that became known as “Nonlinear Optics”. The early results are incorporated in a monograph of this title, published by W. A. Benjamin, New York, in 1965, and the program is still flourishing today. The principal support for all this work, over a period of more than thirty years, has been provided by the Joint Services Electronics Program of the U. S. Department of Defense, with a minimum amount of administrative red tape and with complete freedom to choose research topics and to publish.

My academic career at Harvard has resulted in stimulating interactions with many distinguished colleagues, and also with many talented graduate students. My coworkers have included about sixty Ph.D. candidates and a similar number of postdoctoral research fellows. The contact with the younger genera-

tions keeps the mind from aging too rapidly. The opportunities to participate in international summer schools and conferences have also enhanced my professional and social life. My contacts outside the academic towers, as a consultant to various industrial and governmental organizations, have given me an appreciation for the problems of socio-economic and political origin in the "real" world, in addition to those presented by the stubborn realities of matter and instruments in the laboratory.

Sabbatical leaves from Harvard have made it possible for us to travel farther and to live for longer periods of time in different geographical and cultural environments. Fortunately, my wife shares this taste for travel adventure. In 1957 I was a Guggenheim fellow and visiting lecturer at the École Normale Supérieure in Paris, in 1964-1965 visiting professor at the University of California in Berkeley, in 1973 Lorentz guest professor in Leiden and visiting scientist at the Philips Research Laboratories in the Netherlands. The fall of 1979 I spent as Raman Visiting Professor in Bangalore, India, and the first semester of 1980 as Von Humboldt Senior Scientist in the Institut für Quantum Optik, in Garching near Munich, as well as visiting professor at the College de France in Paris. I highly value my international professional and social contacts, including two exchange visits to the Soviet Union and one visit to the People's Republic of China, each of one-month duration. My wife and I look forward to continuing our diverse activities and to enjoying our home in Five Fields, Lexington, Massachusetts, where we have lived for 26 years.

Honors

Correspondent, Koninklijke Akademie van Wetenschappen, Amsterdam, 1956
Fellow, American Academy of Arts and Sciences, 1956
Member, National Academy of Sciences, Washington, D. C., 1959
Foreign Honorary Member, Indian Academy of Sciences, Bangalore, 1978
Associé Étranger, Académic des Sciences, Paris, 1980
Guggenheim Fellow, 1957
Oliver Buckley Prize, American Physical Society, 1958
Morris E. Liebman Award, Institute of Radio Engineers, 1959
Stuart Ballantine Medal, Franklin Institute, Philadelphia, 1961
National Medal of Science, President of the United States of America, 1974
Lorentz Medal, Koninklijke Akademie van Wetenschappen, Amsterdam, 1979
Frederic Ives Medal, Optical Society of America, 1979
Von Humboldt Senior Scientist, 1980

(added in 1991) : In June 1990 I retired from the faculty of Harvard University and became Gerhard Gade University Professor Emeritus. During the past decade I was also a visiting professor or lecturer for extended periods at the California Institute of Technology, at Fermi Scuola Nationale Superiore in Pisa, Italy, and at the University of Munich, Germany.

In 1991 I serve as President of the American Physical Society. I became an honorary professor of Fudan University, Shanghai, People's Republic of China,

and received honorary doctorates from Lava1 University, Quebec, the University of Connecticut and the University of Hartford. In 1983 I received the Medal of Honor from the Institute of Electrical and Electronic Engineers.

My research in nonlinear optics continued with special emphasis on interactions of picosecond and femtosecond laser pulses with condensed matter and of collision-induced optical coherences. My personal life and professional activities during the past decade have been a natural continuation of what I described in my autobiographical notes in 1981.

NONLINEAR OPTICS AND SPECTROSCOPY

Nobel lecture, 8 December, 1981

by

NICOLAAS BLOEMBERGEN

Harvard University, Division of Applied Sciences, Cambridge, Massachusetts 02138, USA

The development of masers and lasers has been reviewed in the 1964 Nobel lectures by Townes (1) and by Basov (2) and Prokhorov (3). They have sketched the evolution of the laser from their predecessors, the microwave beam and solid state masers. Lasers are sources of coherent light, characterized by a high degree of monochromaticity, high directionality and high intensity or brightness. To illustrate this last property, consider a small ruby laser with an active volume of one 1 cc. In the Q-switched mode it can emit about 10^{18} photons at 694 nm wavelength in about 10^{-8} sec. Because the beam is diffraction limited, it can readily be focused onto an area of 10^{-6} cm^2, about ten optical wavelengths in diameter. The resulting peak flux density is 10^{13} watts/cm*. Whereas 0.1 Joule is a small amount of energy, equal to that consumed by a 100 watt light bulb, or to the heat produced by a human body, each one-thousandth of a second, the power flux density of 10 terawatts/cm^2 is awesome. It can be grasped by noting that the total power produced by all electric generating stations on earth is about one terawatt. (The affix "tera" is derived from the Greek τερασ = monstrosity, not from the Latin “terra”!) Indeed, from Poynting’s vector it follows that the light amplitude at the focal spot would reach 10^{8} volts/cm, comparable to the electric field internal to the atoms and molecules responsible for the binding of valence electrons. These are literally pulled out of their orbits in multiphoton tunneling processes, and any material will be converted to a highly ionized dense plasma at these flux densities. It is clear that the familiar notion of a linear optical response with a constant index of refraction, i.e., an induced polarization proportional to the amplitude of the light field, should be dropped already at much less extreme intensities. There is a nonlinearity in the constitutive relationship which may be expanded in terms of a power series in the electric field components.

$$P_i = \chi_{ij}^{(1)} E_j + \chi_{ijk}^{(2)} E_j E_k + \chi_{ijkl}^{(3)} E_j E_k E_l + \ldots \qquad (1)$$

Such nonlinearities have been familiar at lower frequencies for over a century. For example, power and audio engineers knew about the nonlinear relationship between magnetic field and induction, $B = \mu(H)H$, in transformers and solenoids containing iron. Waveform distortion results (4). Such nonlinear phenomena at optical frequencies are quite striking and can readily be calculated

by combining the nonlinear constitutive relation (1) with Maxwell's equations. In the first decade of this century Lorentz (5) calculated $x^{(1)}$ with the electron modeled as a harmonic oscillator. If he had admitted some anharmonicity, he could have developed the field of nonlinear optics seventy years ago. It was, however, not experimentally accessible at that time, and Lorentz lacked the stimulation from stimulated emission of radiation.

Nonlinear effects are essential for the operation of lasers. With dye lasers it is possible to cover the range of wavelengths from 350-950 nm continuously, including the entire visible spectrum. A variety of nonlinear processes, including harmonic generation, parametric down conversion and the stimulated Raman effects extend the range for coherent sources throughout the infrared and into the vacuum ultraviolet. Thus the field of nonlinear laser spectroscopy could be developed rapidly during the past two decades, aided considerably by previous investigations of related phenomena at radiofrequencies. It is, therefore, appropriate to start this review by recalling some nonlinear phenomena first discovered in the field of magnetic resonance.

NONLINEAR PRECURSORS IN MAGNETIC RESONANCE

As a graduate student of Professor E. M. Purcell at Harvard University, I studied relaxation phenomena of nuclear magnetic resonance in solids, liquids and gases. A radiofrequency field at resonance tends to equalize the population of two spin levels, while a relaxation mechanism tries to maintain a population difference, corresponding to the Boltzmann distribution at the temperature of the other degrees of freedom in the sample. The reduction in population difference is called saturation. It is a nonlinear phenomenon, as the magnitude of the susceptibility tends to decrease with increasing field amplitude. In 1946 we found that "a hole could be eaten", or a saturation dip could be produced, in an inhomogeneously broadened line profile (6). Figure la shows the proton spin resonance in water, broadened by field inhomogeneities of the available magnet. Figures 1b and lc show saturation of a particular packet in the distribution, which is subsequently probed by sweeping through the resonance with a weaker signal after various time intervals. The disappearance rate of the hole is determined by the spin lattice relaxation time. This was also the first indication of the extremely sharp features of NMR lines in liquids, due to motional narrowing, on which the widespread use of NMR spectroscopy is founded.

If two pairs of levels have one level in common, saturation of one resonance may influence the susceptibility at another resonance. This was also observed early in NMR in spin systems with quadrupole splitting and quadrupolar relaxation (7). The detection of Hertzian resonances by optical methods described by Kastler (8) is another manifestation of this phenomenon. A change in the population of sublevels with different values of the spatial quantum number m_j induced by a radiofrequency field produces a change in the polarization of the emitted light. The Overhauser effect (9) describes the change in population of nuclear spin levels in metals (10) due to an application of a

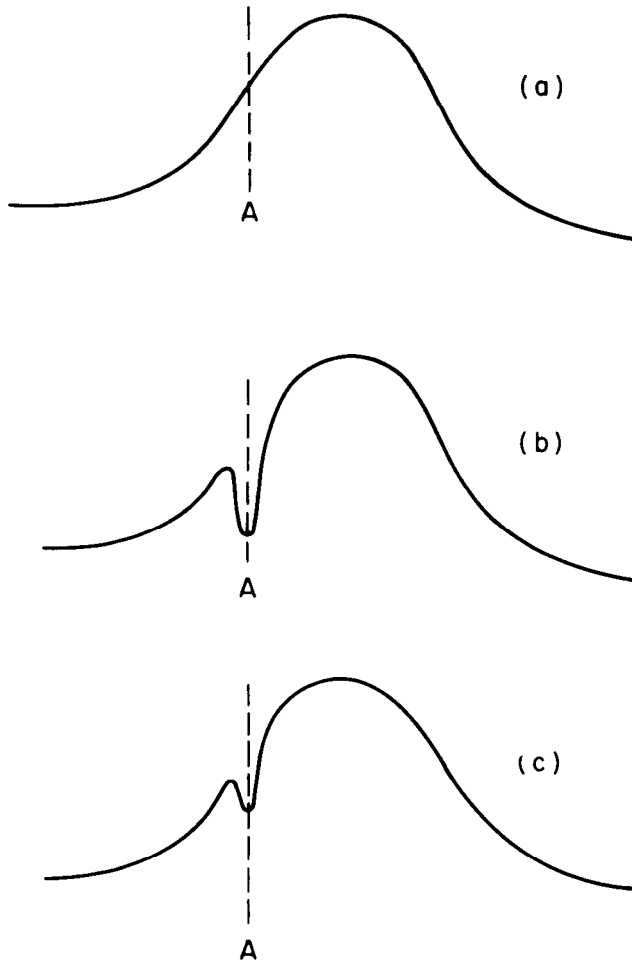


Fig. 1. (after reference 6)
a) Inhomogeneous broadened profile of NMR in water.
b) Saturation dip in inhomogeneous profile, observed in 1946.
c) As in b), but with longer delay between pump signal and probing scan

microwave field at the electron spin resonance. Both optical and microwave pumping methods have been used to obtain nuclear spin polarized targets (11).

It is possible to maintain a steady state inverted population, in which a level with higher energy is more populated than another level with lower energy (12). This pair of levels may be said to have a negative temperature. The principle of the method, displayed in Fig. 2, is based on frequency selective pumping between a pair of nonadjacent energy levels, with the simultaneous action of a suitable relaxation mechanism. The pump tends to establish a high temperature for a pair of levels separated by a higher frequency, while at the same time relaxation maintains a low temperature between a pair with a smaller frequency separation. Stimulated emission will dominate over absorp-

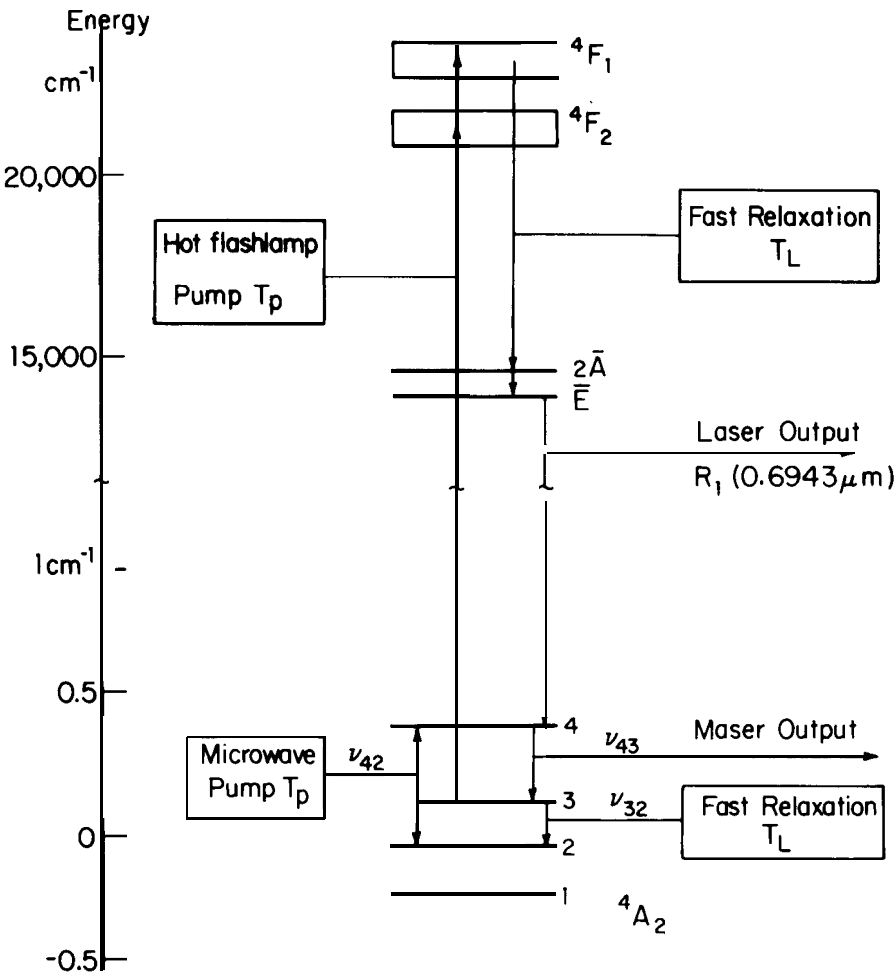


Fig. 2. Energy level diagram of Cr^{3+} in ruby. Note change in vertical scale between microwave maser and optical laser action.

tion at the third pair of a three-level system. Basov and Prokhorov (13) had proposed a frequency selective excitation mechanism for molecular beam masers without explicit discussion of relaxation.

The spin levels of paramagnetic ions in crystals arc useful to obtain maser action at microwave frequencies. The stimulated emission may be considered as the output of a thermodynamic heat engine (14), operating between a hot pump temperature and a low relaxation bath temperature. These two temperatures occur in the same volume element in space, while in a conventional heat engine there is, of course, a spatial separation between the hot and cold parts. The question of thermal insulation between the paramagnetic spin transitions is based on frequency differences and differentials in relaxation rates. This question was addressed in a study of cross-relaxation phenomena (15), which

determine the heat transfer between different parts of the spin Hamiltonian. It turns out that concentrated paramagnetic salts cannot be used in masers, because no large thermal differentials can be maintained in the magnetic energy level system. As a historical curiosity I may add that the biggest hurdle for me in working out the pumping sheme was the question of how to obtain a nonvanishing matrix element between nonadjacent spin levels. This, of course, is resolved by using states which are a superposition of several magnetic quantum numbers m,. This can be obtained by applying the external magnetic field at an arbitrary angle with respect to the axis of the crystal field potential. The multilevel paramagnetic solid state maser is useful as an extremely low noise microwave amplifier. Such a maser, based on the energy levels of the Cr^{3+} ion in ruby, was used, for example, by Penzias and Wilson in their detection of the cosmic background radiation (16).

The same principle has subsequently been used to obtain a medium with gain in most lasers. It was incorporated in the basic proposal for an optical maser by Schawlow and Townes (17). It is noteworthy that the first operating laser by Maiman (18) also used the $Cr.^{3+}$ ions in ruby as the active substance. Of course, a different set of energy levels is involved in the two cases, and the change in frequency scale in the top and bottom part of Fig. 2 should be noted. The amplitude of the laser output is limited by a nonlinear characteristic, as for any feed-back oscillator system. It is the onset of saturation by the laser radiation itself which tends to equalize the populations in the upper and lower lasing levels.

NONLINEAR OPTICS

With the development of various types of lasers, the stage was set for a rapid evolution of the study of nonlinear optical phenomena. The demonstration by Franken and coworkers of second harmonic generation of light by a ruby laser pulse in a quartz crystal marks the origin of nonlinear optics as a new separate subfield of scientific endeavor (19). The straightforward experimental arrangement of this demonstration is shown in Fig. 3.

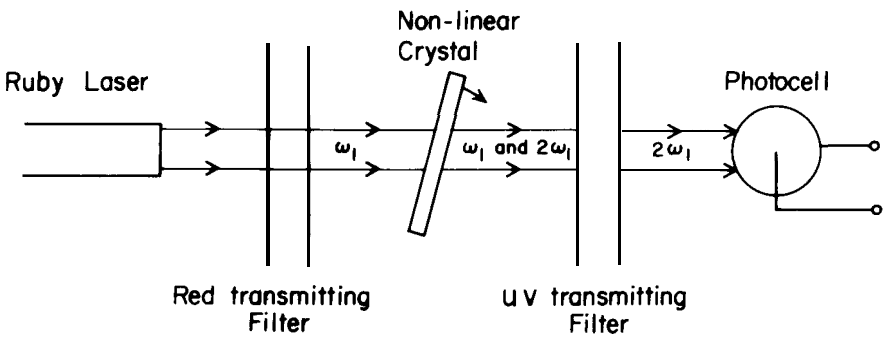


Fig. 3. Second harmonic generation of light

The lowest order nonlinear susceptibility $x^{(2)}$ in equation (1) has only nonvanishing tensor elements in media which lack inversion symmetry. The polarization quadratic in the field amplitude leads to the optical phenomena of second harmonic generation, sum and difference frequency mixing, as well as to rectification of light. These properties of a device with a quadratic response were, of course, well known in radio engineering. The photoelectric emission current is a quadratic function of the light field amplitudes, and it is modulated at a difference frequency when two light beams with a small frequency difference are incident on it (20).

In general, the terms in $x^{(2)}$ provide a coupling between sets of three electromagnetic waves. Each wave has its own frequency ω_i, wave vector $\underline{k}_i$, state of polarization $\hat{e}_i$, as well as a complex amplitude $E_i = A_i \exp(i\phi_i)$. In the same manner the term in $x^{(3)}$ causes a coupling between four electromagnetic waves. A general formulation of three- and four-wave light mixing was developed by our group at Harvard (21). The quantum mechanical calculation of the complex nonlinear susceptibilities, based on the evolution of the density matrix, was also quickly applied to optical problems (22). Generalizations of the Kramers-Heisenberg dispersion formula result. The nonlinear susceptibilities are functions of several frequencies and have more than one resonant denominator. They are tensors of higher order, and each clement has a real and an imaginary part in the presence of damping. They describe a large variety of nonlinear optical effects. At the same time Akhmanov and Khokhlov (23) also extended the formulation of parametric nonlinearities from the radiofrequency to the optical domain.

Returning to the generation of optical second harmonics in transparent piezo-electric crystals, the problem of momentum matching in the conversion

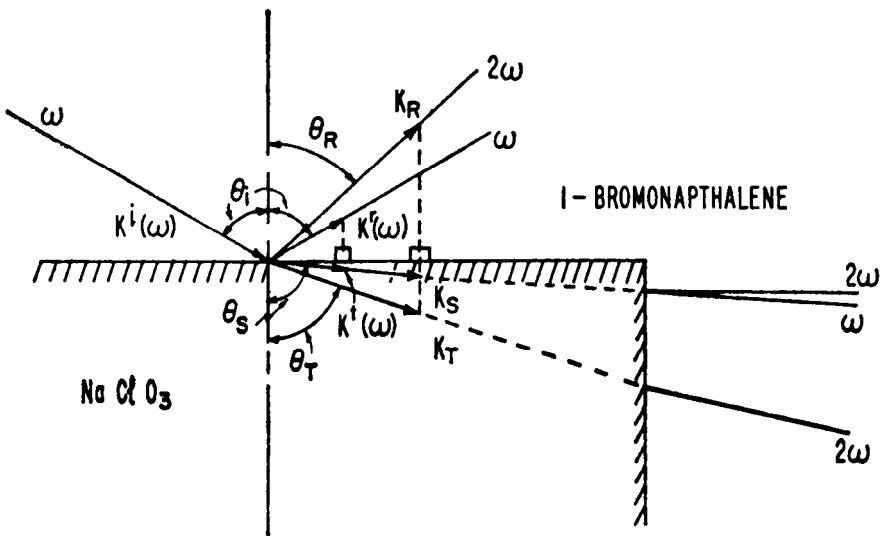


Fig. 4. Wave vectors of fundamental and second harmonic light waves at the boundary of a cubic piezoelectric crystal immersed in an optically denser fluid.

of two fundamental quanta to one quantum at the second harmonic frequency presents itself. Due to color dispersion, one usually has $\underline{k}_2 - 2\underline{k}_1 = \Delta\underline{k} \neq 0$. The mismatch in phase velocities between the second harmonic polarization and the freely propagating wave at 2ω leads to the existence of two waves at 2ω n the nonlinear crystal, a forced one with wave vector $\underline{k}_s = 2\underline{k}_1$ and another with wave vector $\underline{k}_T = \underline{k}_2$, for a freely propagating wave at 2ω In addition, there is a reflected second harmonic wave with wave vector $\underline{k}_R$. Figure 4 depicts the geometry for the case that the nonlinear crystal is embedded in a liquid with a higher linear index of refraction. Conservation of the components of momentum parallel to the surface determines the geometry (24). The amplitudes of the free waves, which are solutions of the homogeneous wave equations, are determined by the condition that the tangential components of the second harmonic electric and magnetic field at the boundary are continuous. Thus a very simple procedure, based on conservation of the component of momentum parallel to the boundary, yields the generalizations of the familiar optical laws of reflection and refraction to the nonlinear case (24). Table 1 illustrates the enormous compression in the time scale of the development of linear and nonlinear geometrical optics. This compression is made possible, of course, by the establishment of a general formulation of electromagnetic phenomena by Maxwell in the second half of the nineteenth century. Lorentz showed in his Ph.D. thesis (25) how the laws of linear reflection, recorded by Hero of Alexandria (first century A.D.), Snell's laws (1621) and Fresnel's laws (1823) for the intensities and polarizations all followed from Maxwell's equations.

It is also suggested by the geometry of Fig. 4 that, on increasing the angle of incidence θ_i, nonlinear analogues for total reflection and evanescent surface waves should occur. Indeed, all such predictions have been verified (26), and in

Table 1. Historical dates of linear and nonlinear optical laws

	Linear	*Nonlinear*
Law of Reflection	1st century (Hero of Alexandria)	I962 (Bloembergen and Pershan)
Law of Refraction	1621 (Snell)	1962 (Bloembergen and Pershan)
Intensity of Reflected and Refracted Light	1823 (Fresnel)	1962 (Bloembergen and Pershan)
Conical Refraction		
Theory	1833 (Hamilton)	1969 (Bloembergen) and Shih)
Experiment	1833 (Lloyd)	1977 (Schell and Bloembergen)

particular the nonlinear coupling between surface excitations is of active current interest (27). In 1833 Hamilton, who was to formulate Hamiltonian mechanics three years later, predicted the phenomenon of conical refraction based on Fresnel's equations of light propagation in biaxial optical crystals. The experimental confirmation in the same year by Lloyd was considered a triumph of the Fresnel equations for the elastic nature of optical propagation! The time lag between the prediction of nonlinear conical refraction and its experimental confirmation was much longer (28), as shown in Table I. In the twentieth century the description of electromagnetic propagation is not in doubt, and most researchers were too busy with more important applications of laser beams than the rather academic problem of nonlinear conical refraction.

The parametric coupling of light waves in a nonabsorbing medium may be considered as the scattering of photons between eigenmodes or waves of the electric field by the material nonlinearity. Heisenberg (29) and others had discussed an intrinsic nonlinearity of the vacuum.

The virtual intermediate states in that process are the electron-hole pair creation states which lie about a million times higher in energy than the excited states of electrons bound in a material medium. Since the energy mismatch of the intermediate states enters as the cube in the expression of $x^{(3)}$, the vacuum nonlinearity has not been detected. It would be difficult to exclude the nonlinear action of one atom or molecule in the focal volume of extremely intense laser beams used in attempts to detect the nonlinearity of vacuum.

In parametric, nondissipative processes the energy and momentum between incident and emerging photons must be conserved, $\Sigma_i \hbar\omega_i = 0$ and $\Sigma_i \hbar\underset{\sim}{k}_i = 0$, where the frequencies and wave vectors of the incident photons are taken to be negative. As noted above, color dispersion generally gives rise to a momentum mismatch $\Delta\underset{\sim}{k} = \underset{\sim}{k}_2 - 2\underset{\sim}{k}_1$. This limits the active volume of emission to a layer of thickness $|\Delta\underset{\sim}{k}|^{-1}$. It is possible, however, to compensate the color dispersion by optical birefringence in anisotropic crystals. This was demonstrated independently by Giordmaine (30) and by Terhune (31). For $\Delta\underset{\sim}{k} = 0$, the polarization in all unit cells in the crystal contributes in phase to the second harmonic field, and if the crystal is long enough and the light intensity high enough, the fundamental power may be quantitatively converted to second harmonic power (21). Phase coherence is essential. For random phases the final state would be one of equipartition with equal power in the fundamental and the second harmonic mode. More than eighty percent of the fundamental power at 1.06 pm wavelength in a large pulsed Nd-glass laser system has recently been converted to third harmonic power (32) at 0.35 nm. In the first step two-thirds of the fundamental power is converted to second harmonic power. Then equal numbers of fundamental and second harmonic photons are combined to third harmonic photons in another crystal. This conversion may be important for the inertial confinement of fusion targets, as the laser-plasma coupling is improved at higher frequencies. The Manley-Rowe relations, which describe the balance in the photon fluxes of the beams participating in a parametric process, are here put to practical use. A few simple conservation laws thus determine many fundamental features of nonlinear optics.

NONLINEAR SPECTROSCOPY

Processes which are described by the imaginary part of the nonlinear susceptibility, $x^{(3)}$", include saturation and cross saturation, twophoton absorption and stimulated Raman effect. The corresponding real part $x^{(3)}$' describes the intensity dependent index of refraction. It plays a role in self-focusing and defocusing of light, and in creating dynamic optical Stark shifts.

Saturation dip spectroscopy is used extensively to eliminate the effects of Doppler broadening in high resolution spectroscopy and in frequency stabilization of lasers. Consider the case of two traveling waves incident on a gas sample with the same frequency ω, but with opposite wave vectors, $\underline{k} = -\underline{k}'$. The wave with $\underline{k}$ produces a saturation dip in the Doppler profile for the velocity packet of molecules satisfying the relation $\omega = \omega_{ba} - \underline{k} \cdot \gamma$, where ω_{ba} is the atomic resonance frequency. The beam in the opposite direction probes the packet satisfying $\omega = \omega_{ba} - \underline{k}' \cdot \underline{v}' = \omega_{ba} + \underline{k} \cdot \underline{v}'$. The two packets coincide only for $\omega = \omega_{ba}$. If ω is scanned across the Doppler profile, the probe beam will register a saturation dip exactly at the center. The correspondence with the NMR situation described earlier is clear. At optical frequencies the effect was first demonstrated as a dip in the output of a helium neon laser (33, 34), and is known as the Lamb dip (35). It is experimentally advantageous to observe the effect in an external absorption cell with a strong pump beam in one direction and a weak probe beam in the opposite direction. While the Doppler width of

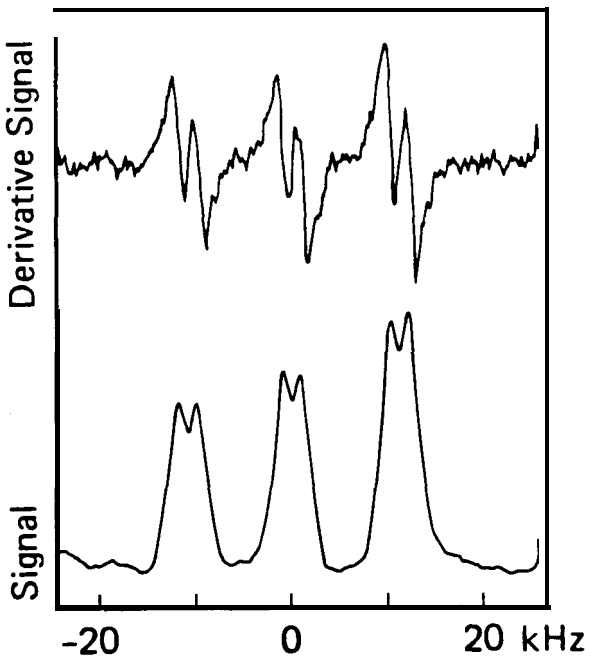


Fig. 5. High resolution (< 1 kHz) saturation spectroscopy of a $^{12}C\,H_4$ spectral line near 3.39 μm wavelength. Each of the three hyperfine components is split into a doublet from the optical recoil effect. The upper curve is the experimental derivative trace (after reference 36).

the vibrational rotational transition of methane near 3.39 μm wavelength is about 300 MHz, spectral features of about 1 kHz have been resolved by Hall and Borde (36). Figure 5 shows the features of the saturation dip, as the frequency of the probe beam was modulated. Saturation spectroscopy reveals not only a hyperfine structure of the molecular transition due to spin-rotational interaction, but also the infrared photon recoil effect which doubles each individual component. With a resolution approaching one part in 10^{11} many

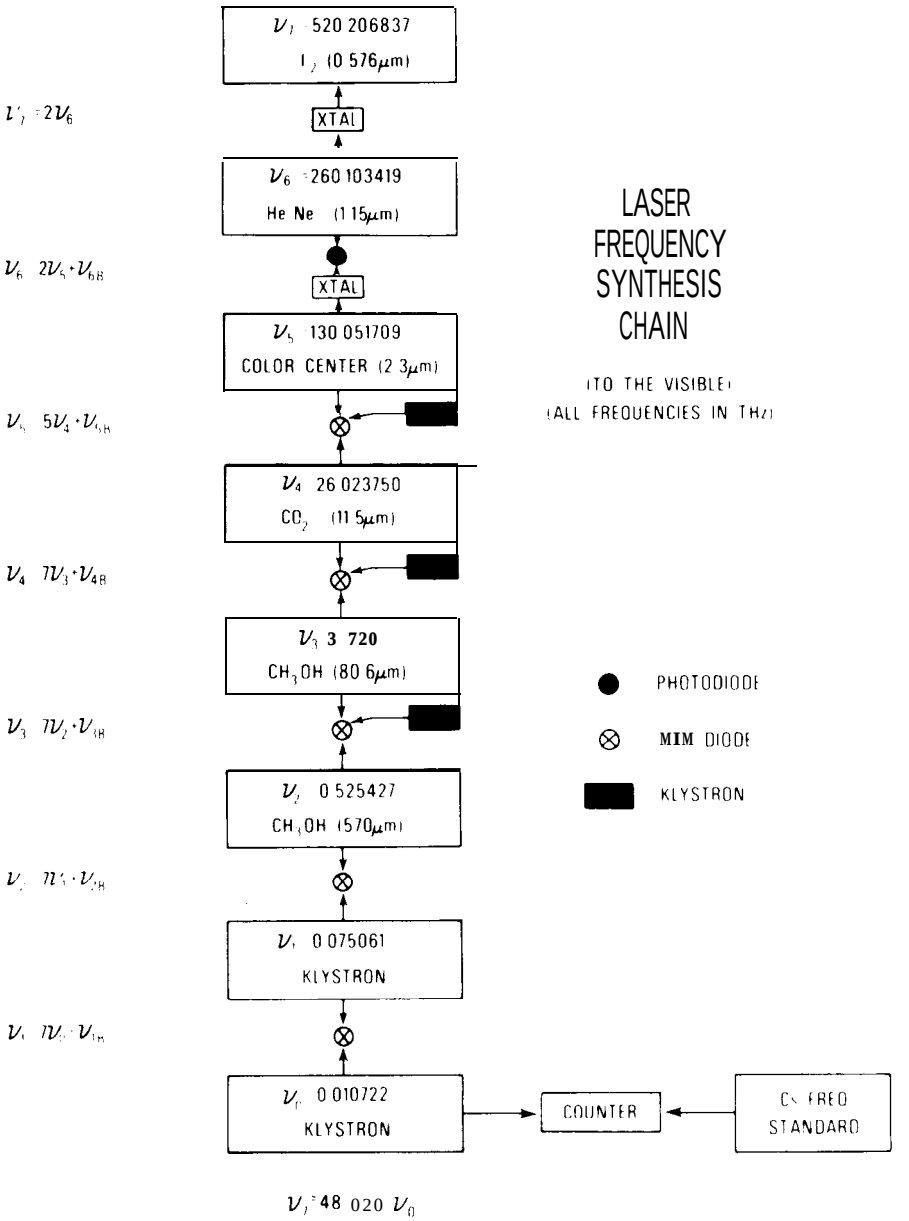


Fig.6. Laser frequency **synthesis chain (after** reference **38)**

other effects, such as curvature of the optical phase fronts and dwell time of the molecules in the beam, must be considered.

The frequencies of lasers throughout the infrared, each stabilized on the saturation dip of an appropriate molecular resonance line, have been compared with each other by utilizing the nonlinear characteristics of tungsten whisker-nickel oxide-nickel point-contact rectifiers (37, 38). The difference frequency between one laser and a harmonic of another laser is compared with a microwave frequency, which in turn is calibrated against the international frequency standard. Thus it has been possible to extend absolute frequency calibrations to the visible part of the spectrum (38), as shown by the chain in Fig. 6. Since the wavelength of the laser is independently compared with the krypton source length standard, it is possible to determine the velocity of light (39) with a precision set by the length standard definition, c = 299 792 458.98 ± 0.2 m/s. It is proposed to define the velocity of light by international agreement, with length measurements then being tied directly to the frequency standard.

The application of saturation spectroscopy to a determination of the Rydberg constant and many other spectroscopic advances are discussed by Schawlow (40). Further details may be found in several comprehensive books on the subject (41-43). Optical saturation spectroscopy has also been carried out in solids, for example for Nd^{3+} ions in a crystal of LaF_3. Here the analogy with NMR techniques is more striking (44).

Two-photon absorption spectroscopy at optical frequencies, predicted by Goeppert-Mayer (45) was first demonstrated by Kaiser and Garrett (46) for E u^{2+} ions in CaF_2. When the two photons have different wave vectors, an excitation with energy $2\hbar\omega$ and wave vector $\underline{k}(\omega) + \underline{k}'(\omega)$ may be probed. Fröhlich (47) applied wave vector-dependent spectroscopy by varying the angle between $\underline{k}$ and $\underline{k}$' to the longitudinal and transverse excition branches in CuCl.

It was suggested by Chebotayev (48) that Doppler-free two-photon absorption features may be obtained in a gas. Consider again two counter-propagating beams. Tune the frequency ω so that $2\omega = \omega_{ba}$ corresponds to the separation of two levels with the same parity. For processes in which one photon is taken out of the beam with wave vector $\underline{k}$ and the other photon is taken out of the beam with wave vector $\underline{k}' = -\underline{k}$, all atoms regardless of their velocity are resonant. The apparent frequencies $\omega + \underline{k}\cdot\underline{v}$ and $\omega - \underline{k}\cdot\underline{v}$ of the photons in the two beams in the rest frame of an atom always add up to ω_{ba}. The two-photon absorption signal thus exhibits a very sharp Doppler-free feature, which was demonstrated experimentally in three independent laboratories (49-51). Thus high energy levels, including Rydberg states, of the same parity as the ground state may be studied in high resolution (52). The reader is again referred to the literature of further details (41, 43).

There is, of course, a close correspondence between two-photon absorption and Raman processes. A medium with a normal population difference between two levels la > and lb >, which permit a Raman active transition, will exhibit a gain at the Stokes frequency, $\omega_s = \omega_L - \omega_{ba}$, in the presence of a strong pump

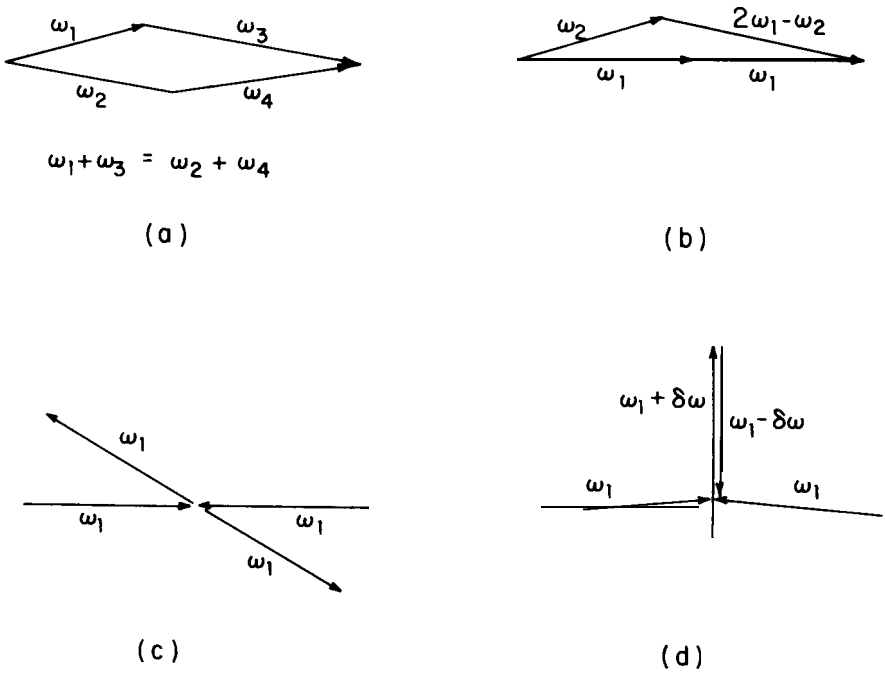


Fig. 7. Some typical wave vector geometries of four-wave light mixing

beam at ω_L. Owyoung (53), for example, has resolved the line structure in the Q-branch of a vibrational-rotational band of the methane molecule by the technique of stimulated Raman scattering. It is also possible to compare directly the Raman gain and a two-photon absorption loss with these nonlinear techniques.

FOUR-WAVE MIXING SPECTROSCOPY

The nonlinearity $X^{(3)}$ describes a coupling between four light waves, and some typical wave vector geometries which satisfy both energy and momentum conservation of the electromagnetic fields are shown in Fig. 7. The generation of a new beam at the frequency $2\omega_1 - \omega_2$, due to one incident beam at ω_1 and another at ω_2, corresponding to the geometry in Fig. 7b, was first demonstrated by Maker and Terhune (54, 55). They detected coherent antistokes raman scattering in organic liquids, where the nonlinear coupling constant $x^{(3)}$ exhibits a Raman-type resonance at the intermediate frequency $\omega_1 - \omega_2$, as shown schematically in Fig. 8b. Enchancement can also occur by a resonance at the intermediate frequency $2\omega_1$. It is thus possible, using light beams at visible wavelengths in a transparent crystal, to obtain information about resonance and dispersive properties of material excitations in the infrared (56, 57) and the utltraviolet (58). An example of this type of nonlinear spectroscopy is shown in Fig. 9. The two-dimensional dispersion of $x^{(3)}(-2\omega_1 + \omega_2, \omega_1, \omega_1, -\omega_2,)$ in

CuCl is measured as $2\omega_1$, is varied in the vicinity of the sharp Z_3 exciton resonance, while at the same time $\omega_1 - \omega_2$ is varied in the vicinity of the infrared polariton resonance. The interference of two complex resonances with each other, as well as the interference of their real parts with the nonresonant background contribution to $x^{(3)}$, leads to a direct comparison of these nonlinearities.

Wave vector-dependent four-wave mixing spectroscopy by variation of the angle between the incident beams was first performed by De Martini (59). The case of enhancement of the CARS process by one-photon absorptive resonances was investigated by several groups (60-62). Figure 8d shows an example of this situation. The CARS technique is used to monitor the composition and temperature profile in flames. In this and other situations with a large incandescent or fluorescent background, the coherent technique provides additional discrimination (63).

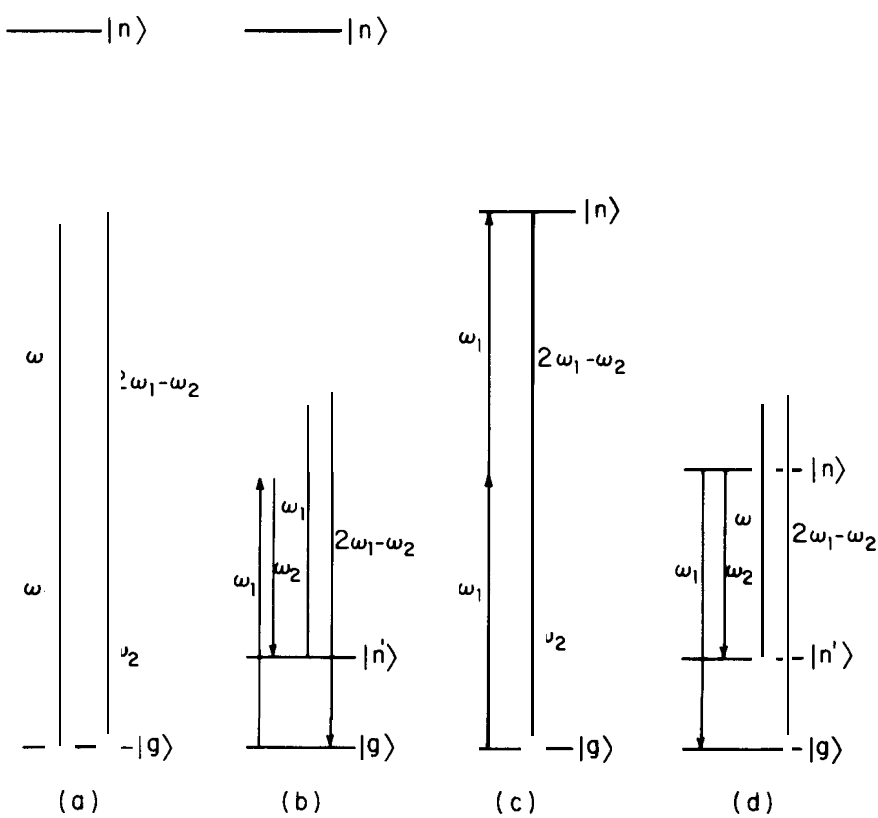


Fig. 8. The creation of a new beam at $2\omega_1 - \omega_2$ by two incident beams at ω_1 and ω_2, respectively. according to the geometry of Fig. 7b.

a) Nonresonant mixing

b) Intermediate Raman resonance (coherent antistokes Raman scattering, or CARS)

c) Intermediate two-photon absorption resonance

d) One-photon resonantly enhanced CARS

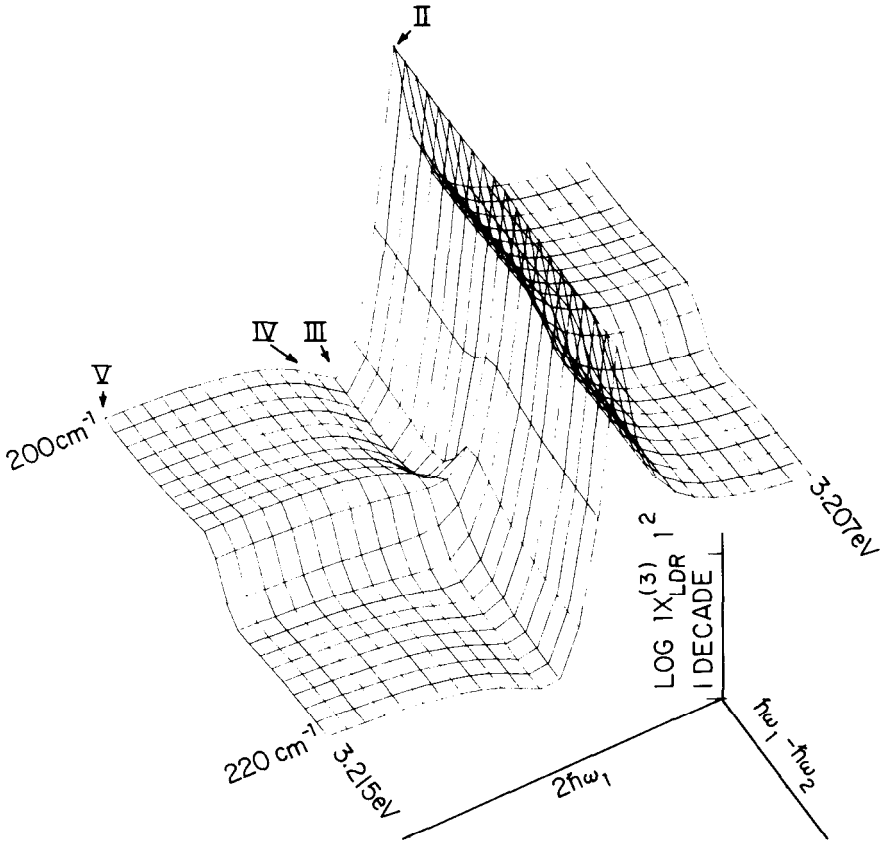


Fig. 9. Two-dimensional frequency dispersion of the nonlinear susceptibility x^{3} in cuprous chloride (after reference 58).

An important recent application of four-wave mixing is phase conjugation (64). Time-reversed phase fronts are obtained by the frequency-degenerate scattering geometry depicted in Fig. 7c. A strong standing wave pump field provides two beams at ω with equal and opposite wave vectors, $\underset{\sim}{k}_1 = -\underset{\sim}{k}_3$. The nonlinear medium may be liquid CS_2, Na vapor, InSb, an absorbing fluid, a molecular gas, or any other medium (65). A signal beam at the same frequency ω has a wave vector $\underset{\sim}{k}_2$, which makes an arbitrary small angle with $\underset{\sim}{k}_1$. In the four-wave scattering process a new beam with wave vector $\underset{\sim}{k}_4 = -\underset{\sim}{k}_2$ is created by the nonlinear polarization

$$P_4(\omega) = x^{(3)}(-w,\ \omega,\ \omega,\ -\omega)E_1E_2^*E_3e^{-i\underset{\sim}{k}_2\cdot\underset{\sim}{r}}.$$

Note that not only the wave vector but also the phase is reversed, because $E_2^* = |E_2|\ e\ x\ p\ (-i\phi_2)$. This implies that the backward wave is the time reverse of the signal wave. If the phase front of the latter has undergone distortions in propagation through a medium, these will all be compensated as the backward wave returns through the same medium. The amplitude of the backward wave may show gain, because the parametric process involved **takes** one photon each

out of the two pump beams and adds one each to the signal and its phase-conjugate beam. The process may also be viewed as real-time instant holography (66). The signal wave forms an intensity interference pattern with each of the pump beams. The physical cause for the grating may be a variation in temperature, in carrier density, in bound space charges, in molecular orientation, depending on the material medium. The other pump reads out this hologram and is scattered as the phase-conjugate wave.

Another variation of nearly degenerate frequency four-wave light mixing has resulted in the recent demonstration of collision-induced coherence (67). Two beams at frequency ω_1 were incident in a vertical plane on a cell containing Na vapor and helium buffer gas. A third beam at a variable frequency ω_2 is incident in the horizontal plane. The generation of a beam in a new direction in the horizontal plane is observed at frequency $2\omega_1-\omega_2$. The intensity of this new beam displays resonances for $\omega_1 = \omega_2$ and $\omega_1 - \omega_2 = 17$ cm-', corresponding to the fine structure splitting of the 3P doublet of the Na atom. These resonances, however, occur only in the presence of collisions. Their intensity varies linearly or quadratically with the partial pressure of helium (68). The paradox that a phase-destroying collisional process can give rise to the generation of a coherent light beam is resolved as follows. In four-wave mixing many different scattering diagrams contribute to the final result (60, 61). These different coherent pathways happen to interfere destructively in the wave mixing case under consideration. Collisions of the Na atoms destroy this destructive interference.

HIGHER ORDER NONLINEARITIES

Higher order terms in the perturbation expansion of equation (1) are responsible for the generation of higher harmonics and multiphoton excitation processes. Akhmanov (69) has studied the generation of fourth harmonics in a crystal of lithium formate and the fifth harmonic in calcite. Reintjes et al. (70) have generated coherent radiation in the vacuum ultraviolet at 53.2 nm and 38.02 nm, as the fifth and seventh harmonic of a laser pulse at 266 nm which was focused into helium gas. The intensity at 266.1 nm was itself derived by two consecutive frequency doublings from a Nd^{3+} glass laser at 1.06 μm. Radiation at this infrared wavelength can induce photoelectric emission from tungsten. The energy of four photons is necessary to overcome the work function. This photoelectric current is proportional to the fourth power of the laser intensity (71) .

Studies of multiphoton ionization of atoms and molecules has been pioneered by Prokhorov and coworkers (72). There is clear evidence for ionization of xenon by eleven photons at 1.06 μm. The ion current increases as the eleventh power of the intensity (73). The required laser intensities are so high that extreme care must be taken to avoid avalanche ionization started by electrons created from more readily ionizable impurities.

Atoms and molecules may, of course, also be ionized stepwise. A real excited bound state may be reached, whence further excitation beyond the ionization

limit proceeds. The spectroscopy of auto-ionizing states has also been furthered by multiphoton laser excitation (74).

The intermediate resonances in the stepwise ionization process are species selective. The ionization of single atoms may be detected with a Geiger-Müller counter. Resonance ionization spectroscopy (75) uses this device in combination with one or more tunable dye lasers. The presence of a single atom amidst 10^{20} atoms of other species may be detected. Thus rare stable or unstable daughter atoms may be identified in coincidence with the decay of the parent atom. Ultralow level counting may also aid in measuring inverse B-decay products induced by the solar neutrino flux (75).

Many polyatomic molecules with absorption features near the infrared emission lines from pulsed CO_2 lasers can be dissociated without collisions in a true unimolecular reaction (76,77). In many cases more than thirty infrared quanta at λ = 9.6 or 10.6 μm wavelength are needed to reach the dissociation limit. Nevertheless the rate determining the step appears to be a succession of one-photon absorption (and emission) processes (78). The dissociation yield depends on the total energy fluence in the pulse and is largely independent of pulse duration (or peak intensity). This may be understood in terms of the large density of states in polyatomic molecules with a high degree of vibrational excitation. The energy absorbed by one mode is rapidly shared (equipartitioned) with the other degrees of freedom. Intramolecular relaxation times in highly excited polyatomic molecules are often quite short, on the order of one picosecond (10^{-12} sec). Infrared photochemistry of molecules in highly excited states has been stimulated by the availability of high power lasers. Both multiphoton dissociation and ionization processes can be applied to laser isotope separation (77).

OPTICAL TRANSIENTS

The perturbation expansion in equation (1) converges only if the Rabi frequency, $\hbar^{-1}|ex|_{ba}|E|$, proportional the magnitude of the electric dipole matrix element and the field amplitude, is small compared to the detuning from resonance ω-ω_{ba}, or small compared to the homogeneous width or damping constant Γ_{ba} of the resonance. When this condition is not satisfied, very interesting nonlinear optical phenomena occur. They again have their precursors in magnetic resonance and include, among others, free induction decay (79), optical nutation (79), optical echoes (80, 81) and split field resonances (82). The one-to-one correspondence of the evolution of any two-level system with the motion of a spin 1/2 system in magnetic resonance offers a convenient basis for description and also has heuristic value (83, 84).

Self-induced transparency (85) describes the propagation of a solitary optical wave or “soliton” which develops when an intense light pulse enters a material medium at a sharp absorbing resonance. The front part of the pulse excites the resonant transition; then the excited resonant state feeds back the energy to the trailing part of the pulse. The net result is that each two-level member of the ensemble executes a complete revolution around the effective field in the

rotating frame of reference (83). In this 2π pulse no electromagnetic energy is dissipated in the medium, but the propagation velocity of the energy is slowed down. The fraction of energy stored in the medium does not contribute to the propagation.

The spontaneous emission process in the presence of a large coherent driving field (86, 87), the cooperative radiation phenomena associated with the superradiant state (88), and the statistical properties (89, 90) of electromagnetic fields with phase correlations have increased our understanding of the concept of the photon.

Short optical pulses have been used extensively for time-resolved studies of transient phenomena and the measurement of short relaxation times. Very powerful pulses of about 10 picosecond (10^{-11} sec) duration are readily obtained by the technique of mode locking. Generally, the medium is excited by the first short pulse and probed by a second pulse with a variable time delay. The first pulse, for example, may excite a molecular vibration by stimulated Raman scattering. This coherent vibration will interact with the second pulse to give an antistokes component. A picosecond pulse traversing a cell of water generates a nearly continuous white spectrum due to phase modulation. This white picosecond pulse may be used to probe variations in absorption due to the first pulse. These techniques have been developed in depth by Kaiser (91) and others (92). More recently, the creation of light pulses as short as 4×10^{-14} sec has been achieved.

It is also possible, with a picosecond pulse, to melt a thin surface layer of a metal, alloy or semiconductor. After the light pulse is gone, this layer (10-20 nm thick) resolidifies rapidly by thermal conduction to the cool interior. Cooling rates of 10^{13}"C/sec are attainable. Thus it is possible to freeze in amorphous phases or other normally unstable configurations (93). New regimes of solid state kinetics are thus opened up for investigation.

CONCLUSION

Nonlinear optics has developed into a significant subfield of physics. It was opened up by the advent of lasers with high peak powers. The availability of tunable dye lasers has made detailed nonlinear spectroscopic studies possible throughout the visible region of the spectrum, from 0.35 to 0.9 nm. Conversely, nonlinear techniques have extended the range of tunable coherent radiation. Harmonic generation, parametric down conversion, and stimulated Raman scattering in different orders have all extended the range from the vacuum ultraviolet (94) to the far infrared (95). The soft X-ray region still presents a challenge.

Nonlinear optical processes are essential in many applications. Modulators and demodulators are used in optical communications systems. Saturable absorption and gain play an essential role in obtaining ultrashort pulses. The domain of time-resolved measurement may be extended to the femtosecond domain. This opens up new possibilities in materials science and chemical kinetics. A detailed understanding of nonlinear processes is essential in pushing

the frontiers of time and length metrology, with applications to geological and cosmological questions.

The field of nonlinear spectroscopy has matured rapidly but still has much potential for further exploration and exploitation. The applications in chemistry, biology, medicine, materials technology, and especially in the field of communications and information processing arc numerous. Alfred Nobel would have enjoyed this interaction of physics and technology.

I wish to express my indebtedness to my coworkers and graduate students, past and present, as well as to many colleagues, scattered in institutions around the globe, whose work in nonlinear optics and spectroscopy, cited or uncited, is also honored by this award.

REFERENCES

1. Townes, C. H., in ***Nobel I.ectures in Physics*** (Elsevier, Amsterdam, 1972). Vol. 4. p. 58.
2. Basov, N. G., ***ibid, p. 89.***
3. Prokhorov. A. M,, ***ibid.,*** p. 110.
4. Salinger, H., Archiv. f. Elektrotechnik 12, 268 (1923).
5. Lorentz, H. A., ***Theory of Electrons (Teubner.*** Leipzig. 1909).
6. Bloembergen, N., Purcell. E. M. and Pound, R. V.. Phys. Rev. 73. 679 (1948).
7. Pound. R. V., Phys. Rev. 79, 685 (1950).
8. Kastler, A., in ***Nobel*** Lectures in Physics (Elsevier, Amsterdam. 1972). Vol. 4, p. 186.
9. Overhauser, A. W., Phys. Rev. 91, 476 (1953).
10. Carver, T R. and Slichter, C. P., Phys. Rev 102. 975 (1956).
11. See. for example. Abragam. A., Principles of Nuclear Magnetism (Oxford University Press. London. 1960). Goldman, M., .Spin *Temperature and* Nuclear .*Magnetic* Resonance in Solids (Oxford University Press. London, 1970).
12. Bloembergcn, N., Phys. Rev. 104, 324 (1956).
13. Basov, N. G. and Prokhorov. A. M., Zh. Eksp. Teor. Fiz. 28. 249 (1955).
14. Scovil. H. E. D. and Schulz-du Bois, E. O., Phys Rev. Lett. 2. (1955).
15. Bloembcrgen, N., Shapiro, S.. Pcrshan, P. S. and Artman. J. O., Phys. Rev. 114, 445 (1959).
16. Wilson, R. W., "Nobel Lecture 1978," in Rev. Mod. Phys. 51, 767 (1979).
17. Schawlow, A. I,., and Townes. C H.. Phys. Rev. 112. 1940 (1958)
18. Maiman. T. H., Nature 187, 493 (1960).
19. Franken, P., Hill. A. E., Peters, C W. and Weinreich, G., Phys. Rev. Lett. 7. 118 (1961).
20. Forrester. A. T., Gudmundsen. R. A. and Johnson. P. 0.. Phys. Rev. 99. 1961 (1955).
21. Armstrong. J. A. Bloembergen, N., Ducuing, J. and Pershnn, P. S., Phys. Rev. 128. 606 (1962).
22. Bloembergen, N. and Shen. Y. R.. Phys. Rev. 133, A 37 (1963).
23. Akhmanov, C. A. and Khokhlov, R. V., *Problems in Nonlinear Optics* (Academy of Sciences. USSR. Moscow, 1964).
24. Blocmbergen, N. and Pershan. P. S., Phys. Rev. 128. 606 (1962).
25. Lorentz, H. A., *Collected Papers* (Martinus Nyhoff, The Hague. 1935). Vol. I.
26. Bloembergen, N., Simon, H.J. and Lee, C. H.. Phys. Rev. 181. 1261 (1969).
27. DeMartini. F.. Colocci. M., Kohn. S. E. and Shen, Y. R., Phys Rev. Lett. 38. 1223 (1977).
28. Schell. A. J, and Bloembergen. N., Phys. Rev. A18 2592 (1978).
29. Hcisenbcrg, W and Euler. H., Z., Phys. 98. 714 (1936).
30. Giordmainc, J.A, Phys. Rev. Lett. 8. 19 (1962).
31. Maker, P. D., Terhune. R. W., Nisenhoff, M and Savage. C. M., Phys. Rev. Lctt. 8, 21 (1962).

32. Seka, W., Jacobs, S. D., Rizzo, J. E., Boni, R. and Craxton, R. S., Opt. Commun. 34, 469 (1980).
33. Szoke, A. and Javan, A., Phys. Rev. Lett. 10, 521 (1963).
34. McFarlane, R. A., Bennett, W. R. Jr. and Lamb, W. E. Jr., Appl. Phys. Lett. 2, 189 (1963).
35. Lamb, W. E., Phys. Rev. 134, 1429 (1964).
36. Hall, J. L., Bordé, C. J. and Uehara, K., Phys. Rev. Lett. 37, 1339 (1976).
37. Hocker, L. O., Javan, A. and Ramachandra Rao, D., Appl. Phys. Lett. 10, 147 (1967).
38. Jenning, D. A., Petersen, F. R. and Evenson, K. M., in *Laser Spectroscopy* ***IV,*** edited by H. Walther and K. W. Rothe (Springer, Heidelberg, 1979), p. 39.
39. Baird, K. M., Smith, D. S. and Whitford, B. C., Opt. Commun. 31, 367 (1979); Rowley, W. R. C., Shotton, K. C. and Woods, P. T., ibid. 34, 429 (1980).
40. Schlawlow, A. L., ***Les Prix Nobel 1981*** (Almqvist and Wiksell International, Stockholm, 1982); also Rev. Mod. Phys., to be published (1982).
41. ***High-Resolution Laser Spectroscopy,*** edited by K. Shimoda, Topics in Applied Physics 13 (Springer-Verlag, Berlin, Heidelberg, 1976).
42. Letokhov, V. S. and Chebotayev, V. P., *Nonlinear Laser Spectroscopy,* Springer Series in Optical Sciences 4 (Springer-Verlag, Berlin, 1977).
43. Levenson, M. D., ***Introduction*** to ***Nonlinear Laser*** **Spectroscopy** (Academic Press, New York, 1982). The author is indebted to Dr. M. D. Levenson for having received a preprint of his manuscript
44. McFarlane, R. M. and Shelby, R. M., Opt. Lett. 6, 96 (1981).
45. Goeppert-Meyer, M., Ann. Phys. 9, 273 (1931).
46. Kaiser, W and Garrett, G. C. B., Phys. Rev. Lett. 7, 229 (1961).
47. Fröhlich, D., Staginnus, B. and Schönherr, E., Phys. Rev. Lett. 19, 1032 (1967).
48. Vasilinko, L. S., Chebotayev, V. P. and Shishaev, A. V. JETP Lett. 12, 113 (1970).
49. Cagnac, B., Grynberg, G. and Biraben, F., Phys. Rev. Lett. 32, 643 (1974).
50. Levenson, M. D. and Bloembergen, N., Phys. Rev. Lett. 32, 645 (1974).
51. Hänsch, T. W., Harvey, K. C., Meisel, G. and Schawlow, A. L., Opt. Commun. 11, 50 (1974).
52. Stoicheff, B. P. and Weinberger, E., in *Laser Spectroscopy* ***IV,*** edited by H. Walther and K. W. Rothe (Springer, Heidelberg, 1979), p. 264.
53. Owyoung, A., Patterson, C. W. and McDowell, R. S., Chem. Phys. Lett, 59, 156 (1978).
For a historical review of the early work in the stimulated Raman effect, see:
Bloembergen, N., Am. J. of Phys. 35, 989 (1967).
54. Terhune, R. W., Solid State Design 4, 38 (1963).
55. Maker, P. D. and Terhune, R. W., Phys. Rev. 137, A801 (1965).
56. Levenson, M. D. and Bloembergen, N., Phys. Rev. B 1O,4447 (1974).
57. Lotem, H., Lynch, R. T. Jr., and Bloembergen. N., Phys. Rev. A ***14, 1748*** (1976).
58. Kramer, S. D. and Bloembergen, N., Phys. Rev. B 14, 4654 (1976).
59. Coffinet, J. P. and DeMartini, F., Phys. Rev. Lett. 22, 60 (1969).
60. Bloembergen, N., Lotem, H. and Lynch, R. T. Jr., Ind. J. of Pure & Appl. Phys. 16, 151 (1978).
61. Attal, B., Schnepp, O. O., Taran, J.-P. E., Opt. Commun. 24, 77 (1978).
Druet, S. A. J., Taran, J.-P. E. and Bordé, Ch. J., J. Phys. (Paris) 40, 819 1979).
62. Carreira, L. A., Goss, J. P. and Malloy, T. B., Jr., J. Chem. Phys. 69, 855 (1978).
63. Eckbreth, A. C., Appl. Phys. Lett. 32, 421 (1978).
64. Hellwarth, R. W., J. Opt. Soc. Am. 67, 1 (1977).
65. See, for example, a number of papers by various authors published in Opt. Lett. 5, 51, 102, 169, 182 and 252 (1980), and other references quoted therein.
66. Gabor, D., Les *Prix Nobel* 1971 (P. A. Norstedt & Söner, Stockholm, 1972), p. 169.
67. Prior, Y., Bogdan, A. R., Dagenais, M. and Bloembergen, N., Phys. Rev. Lett. 46, 111 (1981).
68. Bloembergen, N., Bogdan, A. R. and Downer, M., in Laser Spectroscopy ***V,*** edited by A. R. W. McKellan, T. Oka and B. Stoicheff (Springer-Verlag, Heidelberg, 1981), p. 157.
69. Akhmanov, S. A., in *Nonliner Spectroscopy,* edited by N. Bloembergen, Fermi, Course 64 (North-Holland Publishing Co., Amsterdam, 1977), p. 239.
70. Reintjes, J., Eckardt, R. C., She, C. Y., Karangelen, N. E., Elton, R. C. and Andrews, R. A., Phys. Rev. Lett. 37, 1540 (1976); Appl. Phys. Lett. 30, 480 (1977).

71. Bechtel, J. H., Smith, W. L. and Bloembergen, N., Opt. Commun. 13, 56 (1975).
72. Bunkin, F. V. and Prokhorov, A. M., Zh. Eksp. Teor. Fiz. 46, 1090 (1964); Bunkin, F. V., Karapetyan, R. V. and Prokhorov, A. M., ***ibid. 47,*** 216 (1964).
73. LeCompte, C., Mainfray, G., Manus, C. and Sanchez, F., Phys. Rev. Lett. 32, 265 (1972).
74. Armstrong, J. A. and Wynne, J. J., in *Nonlinear Spectroscopy,* edited by N. Bloembergen, E. Fermi, Course 64 (North-Holland Publishing Co., Amsterdam, 1977), p. 152.
75. Hurst, G. S., Payne, M. G., Kramer, S. D. and Young, J. P., Rev. Mod. Phys. 51, 767 (1979).
76. Isenor, N. R. and Richardson, M. C., Appl. Phys. Lett. 18, 225 (1971).
77. Ambartsumian, R. V., Letokhov, V. S., Ryabov, E. A. and Chekalin, N. V., JETP Lett. 20, 273 (1974).
78. Bloembergen, N., Opt. Commun. 15, 416 (1975).
79. Brewer, R. G., in ***Nonlinear Spectroscopy,*** edited by N. Bloembergen, E. Fermi, Course 64 (North-Holland Publishing Co., Amsterdam, 1977), p. 87.
80. Kurnit, N. A., Abella, I. D. and Hartman, S. R., Phys. Rev. Lett. 13, 567 (1964); Phys. Rev. ***141,*** 391 (1966).
81. Mossberg, T. W., Whittaker, F., Kachru, R. and Hartmann, S. R., Phys. Rev. A 22, 1962 (1980).
82. Salour, M. M. and Cohen-Tannoudji, C., Phys. Rev. Lett. 38, 757 (1977).
83. Rabi, I. I., Ramsey, N. F. and Schwinger, J. S., Rev. Mod. Phys. 26, 167 (1954).
84. Feynman, R. P., Vernon, F. L. and Hellwarth, R. W., J. Appl. Phys. 28,49 (1957).
85. McCall, S. L. and Hahn, E. L., Phys. Rev. Lett. 18, 908 (1967); Phys. Rev. ***183, 457,*** (1969).
86. Mollow, B. R., Phys. Rev. 188, 1969 (1969).
87. Cohen-Tannoudji, C. and Reynaud, S., J. Phys. B 10, 345 (1977).
88. Gibbs, H. M., Vrehen, Q. H. F. and Hikspoors, H. M. J., Phys. Rev. Lett. 39, 547 (1977).
89. Glauber, R., Phys. Rev. 131, 2766 (1963).
90. See, for example, *Coherence and Quantum Optics IV,* edited by L. Mandel, and E. Wolf, (Plenum Press, New York, 1978).
91. van der Linde, D., Laubereau, A. and Kaiser, W. Phys. Rev. Lett. 26, 954 (1971).
92. See, for example, *Ultrashort Light Pulses,* edited by Shapiro, S. L., Topics in Applied Physics 18 (Springer-Verlag, Heidelberg, 1977).
93. Liu, J. M., Yen, R., Kurz, H. and Bloembergen, N., Appl. Phys. Lett. 39, 755 (1981).
94. Harris, S. E., Appl. Phys. Lett. 31, 398 (1977); Harris, S. E. ***et al.,*** in *Laser Spectroscopy V,* edited by A. R. W. McKellan, T. Oka, and B. Stoicheff, (Springer-Verlag, Heidelberg, 1981), p, 437.
95. Byer, R. L. and Herbst, R. L., in ***Nonlinear Infrared Generation.*** edited by Y. R. Shen, Topics in Applied Physics 16 (Springer-Verlag, Heidelberg, 1977), p. 81; V. T. Nguen, and T. J. Bridges, *ibid.,* p, 140; J. J. Wynne, and P. P. Sorokin, *ibid.,* p. 160.

Arthur L. Schawlow

ARTHUR L. SCHAWLOW

I was born in Mount Vernon, New York, U.S.A. on May 5, 1921. My father had come from Europe a decade earlier. He left his home in Riga to study electrical engineering at Darmstadt, but arrived too late for the beginning of the term. Therefore, he went on to visit his brother in New York, and never returned either to Europe or to electrical engineering. My mother was a Canadian and, at her urging, the family moved to Toronto in 1924. I attended public schools there, Winchester elementary school, the Normal Model School attached to the teacher's college, and Vaughan Road Collegiate Institute (high school).

As a boy, I was always interested in scientific things, electrical, mechanical or astronomical, and read nearly everything that the library could provide on these subjects. I intended to try to go to the University of Toronto to study radio engineering, and my parents encouraged me. Unfortunately my high school years, 1932 to 1937, were in the deepest part of the great economic depression. My father's salary as one of the many agents for a large insurance company could not cover the cost of a college education for my sister, Rosemary, and me. Indeed, at that time few high school graduates continued their education. Only three or four out of our high school class of sixty or so students were able to go to a university.

There were, at that time, no scholarships in engineering, but we were both fortunate enough to win scholarships in the faculty of Arts of the University of Toronto. My sister's was for English literature, and mine was for mathematics and physics. Physics seemed pretty close to radio engineering, and so that was what I pursued. It now seems to me to have been a most fortunate chance, for I do not have the patience with design details that an engineer must have. Physics has given me a chance to concentrate on concepts and methods, and I have enjoyed it greatly.

With jobs as scarce as they were in those years, we had to have some occupation in mind to justify college studies. A scientific career was something that few of us even dreamed possible, and nearly all of the entering class expected to teach high school mathematics or physics. However, before we graduated in 1941 Canada was at war, and all of us were involved in some way. I taught classes to armed service personnel at the University of Toronto until 1944, and then worked on microwave antenna development at a radar factory.

In 1945, graduate studies could resume, and I returned to the University. It was by then badly depleted in staff and equipment by the effects of the depression and the war, but it did have a long tradition in optical spectroscopy. There were two highly creative physics professors working on spectroscopy,

Malcolm F. Crawford and Harry L. Welsh. I took courses from both of them, and did my thesis research with Crawford. It was a very rewarding experience, for he gave the students good problems and the freedom to learn by making our own mistakes. Moreover, he was always willing to discuss physics, and even to speculate about where future advances might be found.

A Carbide and Carbon Chemicals postdoctoral fellowship took me to Columbia University to work with Charles H. Townes. What a marvelous place Columbia was then, under I. I. Rabi's leadership! There were no less than eight future Nobel laureates in the physics department during my two years there. Working with Charles Townes was particularly stimulating. Not only was he the leader in research on microwave spectroscopy, but he was extraordinarily effective in getting the best from his students and colleagues. He would listen carefully to the confused beginnings of an idea, and join in developing whatever was worthwhile in it, without ever dominating the discussions. Best of all, he introduced me to his youngest sister, Aurelia, who became my wife in 1951.

From 1951 to 1961, I was a physicist at Bell Telephone Laboratories. There my research was mostly on superconductivity, with some studies of nuclear quadrupole resonance. On weekends I worked with Charles Townes on our book *Microwaue Spectroscopy,* which had been started while I was at Columbia and was published in 1955. In 1957 and 1958, while mainly still continuing experiments on superconductivity, I worked with Charles Townes to see what would be needed to extend the principles of the maser to much shorter wavelengths, to make an optical maser or, as it is now known, a laser. Thereupon, I began work on optical properties and spectra of solids which might be relevant to laser materials, and then on lasers.

Since 1961, I have been a professor of physics at Stanford University and was chairman of the department of physics from 1966 to 1970. In 1978 I was appointed J. G. Jackson and C. J. Wood Professor of Physics. At Stanford, it has been a pleasure to do physics with an outstanding group of graduate students, occasional postdoctoral research associates and visitors. Most especially the interaction with Professor Theodor W. Hänsch has been continually delightful and stimulating. Our technicians, Frans Alkemade and Kenneth Sherwin have been invaluable in constructing apparatus and keeping it in operation. My secretary for the past nineteen years, Mrs. Fred-a Jurian, provides whatever order that can be found amidst the chaos of my office. Much of the time, my thoughts are stimulated there by the sounds of traditional jazz from my large record collection.

My wife is a musician, a mezzo soprano and choral conductor. We have a son, Arthur Keith, and two daughters, Helen Aurelia and Edith Ellen. Helen has studied French literature at Stanford, the Sorbonne, and at the University of California in Berkeley, and is now on the staff of Stanford University. Edith graduated from Stanford this year with a major in psychology.

Awards

Stuart Ballantine Medal (1962); Thomas Young Medal and Prize (1963); Morris N. Liebmann Memorial Prize (1964); California Scientist of the Year (1973); Frederick Ives Medal (1976); Marconi International Fellowship (1977).

Honorary doctorates from University of Ghent, Belgium (1968), University of Toronto, Canada (1970), University of Bradford, England (1970). Honorary professor, East China Normal University, Shanghai (1979).

Member, U.S. National Academy of Sciences.
Fellow, American Academy of Arts and Sciences.
President, Optical Society of America (1975)
President, American Physical Society (1981)

(added in 1991) : I retired from teaching and became Professor Emeritus in 1991. My wife died in an automobile accident in May, 1991. My daughter Helen is now Assistant Professor of French at the University of Wisconsin. From Helen and her sister Edith, I now have four grandchildren.

Awards:

Arthur Schawlow Medal, Laser Institute of America (1982)
U.S. National Medal of Science (1991)
Honorary doctorates from University of Alabama, U.S.A. (1984) Trinity College, Dublin, Ireland (1986) University of Lund, Sweden (1988)

SPECTROSCOPY IN A NEW LIGHT

Nobel lecture, 8 December, 1981

by
ARTHUR L. SCHAWLOW
Department of Physics, Stanford University, Stanford, California 94305, USA

INTRODUCTION

Scientific spectroscopy really began in Uppsala, Sweden, where Anders Angström in 1853 showed that some of the lines in the spectrum of an electric spark come from the metal electrodes and others from the gas between them.' Even earlier, Joseph Fraunhofer had charted the dark lines in the spectrum of the sun, and had measured their wavelengths.' But it was Angström who first identified some of these lines as corresponding to bright lines emitted by particular substances in the spark. Most importantly, he showed the red line of hydrogen, now known as H\α. In subsequent years, Angström found several more visible lines from hydrogen, and measured their wavelengths accurately. When W. Huggins[3] and H. W. Vogel[4] succeeded in photographing the spectra of stars in 1880, they found that these visible lines were part of a longer series extending into the ultraviolet. J. J. Balmer[5] in 1885 was able to reproduce the wavelengths of these lines by a formula, which we might write as

$$\frac{1}{\lambda} = \nu = \frac{R}{n^2} - \frac{R}{2^2}.$$

Balmer obtained the values of the constants from Angström's measurements.

Five years later Johannes Rydberg[6], without knowing of Balmer's work, developed a more general formula for the atomic spectra of alkali metals such as sodium.

$$\frac{1}{\lambda} = \nu = \frac{R}{(n-d)^2} - \text{Constant},$$

Rydberg's formula includes Balmer's as a special case where d = 0. The constant R is now universally known as the Rydberg constant. We know now that it measures the strength of the binding between electrons and nuclei in atoms.

It is well known that the Balmer equation for the hydrogen spectrum helped Niels Bohr to introduce a quantum theory of atoms. In the 1920s atomic and molecular spectroscopy was the principal experimental tool leading to the discoveries of the laws of quantum mechanics, from which comes most of our understanding of modern physics and chemistry.

In the 1940s when I was a graduate student at the University of Toronto,

nuclear physics seemed to be the most active branch of the subject but we had no accelerator. Therefore, I worked with two other students, Frederick M. Kelly and William M. Gray, under the direction of Professor M. F. Crawford to use high-resolution optical spectroscopy to measure nuclear properties from their effects on the spectra of atoms. The shifts and splittings of spectral lines from the interactions between electrons and nuclei were so small that they are known as hyperfine structures. To resolve them, we needed to build high resolution spectroscopic equipment. We also had to reduce the widths of the spectral lines from our light source, because broad lines cause overlapping that could completely hide much of the detail that we were seeking. When the gas density is so low that collisions could be neglected, the principal source of the line widths is the Doppler-broadening from the thermal motions of the atoms. Atoms moving toward the observer emit light that is shifted upward in frequency, while atoms moving away emit light of lower frequency than atoms at rest. Since there is a distribution of velocities, the line is broadened, with a Doppler width given by

$$\Delta\nu = \frac{2\nu}{c}\sqrt{2k\,N_o\,\ell n\,2}\,\sqrt{\frac{T}{M}},$$

where ν is the line frequency, k is Boltzmann's constant, N_o is Avogadro's number, T is the absolute temperature, and M is the molecular weight.

This Doppler width, as a fraction of the line frequency is of the order of $\bar{v}/c$, where $\bar{v}$ is the atomic velocity and c is the velocity of light, or typically about 10^{-5}. We were able to reduce it by a factor of ten or so, by using a roughly collimated beam of atoms, excited by an electron beam, as had been done earlier by Meissner and Luft[7] and by Minkowski and Bruck[8], and by observing the emitted light from a direction perpendicular to the atomic beam. The hyperfine structures we sought could be resolved, but four hours exposure time on our photographic plates was required. It seemed that there really ought to be an easier method that would give still sharper spectral lines, and indeed a large part of our work in laser spectroscopy has been devoted to finding such methods.

LASER SPECTROSCOPY

By the time that Charles Townes and I were working to see if it was possible to make a laser', in 1957 and 1958, both of us had experience in microwave spectroscopy. Thus, it was a familiar idea to us that spectra could be observed without a spectrograph, by tuning a narrowband source across the spectrum. At some wavelengths the light would be absorbed, at others it would be transmitted. This was one of the few applications we could then foresee for lasers. But each of the early lasers gave its own characteristic wavelength, depending on the material used. They could not be tuned very far, just across the width of the laser line. More tuning could be obtained sometimes by applying an external magnetic field, or by changing the temperature of a solid laser material. For instance, a ruby laser emits at 6943 Angstrom units at room

temperatures and 6934 Angstrom units at liquid nitrogen temperature of 77K. William Tiffany, Warren Moos and I used a temperature-tuned ruby laser to map out a small portion of the absorption spectrum of gaseous bromine and see how changing the laser wavelength affected the chemical reactivity of the bromine.[10] We studied bromine because it had a rich spectrum with many absorption lines within the range of the available laser.

Others studied the spectra of the atoms used in a gas laser, particularly neon, and interesting phenomena were discovered. To understand them, we must note that the output wavelength of a laser is determined only roughly (within the Doppler line width) by the amplifying medium, and more precisely by the tuning of the laser resonator. By changing the spacing between the end mirrors, the laser can be tuned over the frequency range for which amplification is enough to overcome the losses. One might expect that in the center of the atomic line, where the gain is largest, the laser output will be greatest. But Willis E. Lamb, Jr. predicted, from a detailed theoretical analysis of laser principles, that there would be a dip in power output at the center of the line." It would occur because light beams travel in both directions inside the laser resonator. At the center of the line both beams stimulate the same excited atoms, those with zero velocity, as indicated in Figure 1. At any other tuning, the light waves interact with those atoms whose velocity provides just the

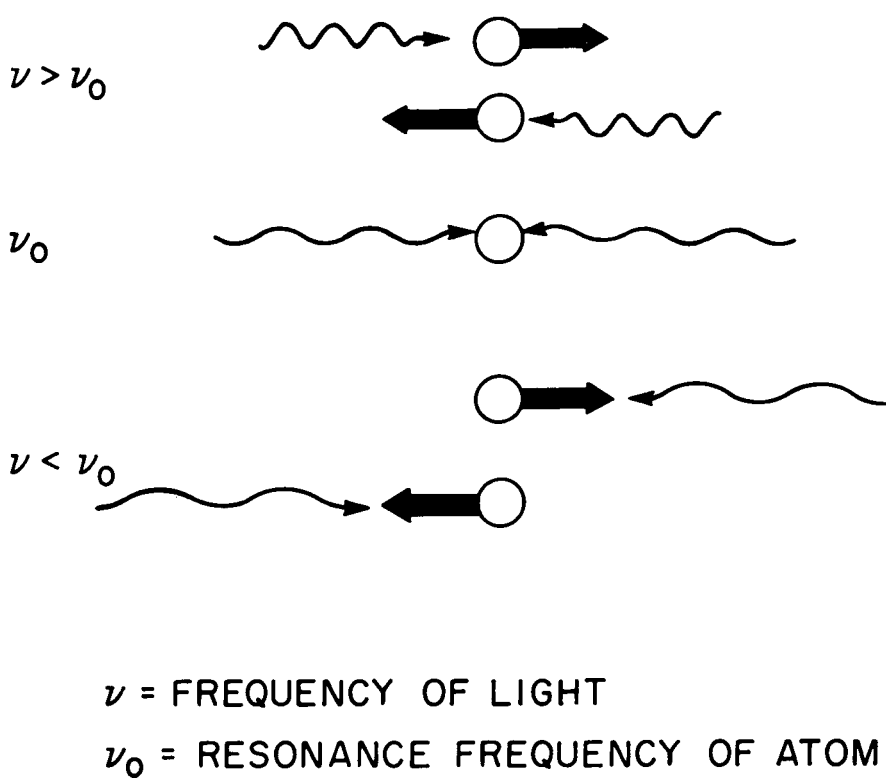


Fig. 1. Moving molecules interact with an approaching lower frequency wave Doppler-shifted upward in frequency, or a following higher-frequency wave shifted downward.

Doppler shift needed to bring the light into resonance with them. Thus there are two groups of atoms, with the same speed but opposite directions, which can be stimulated to provide the laser output. This "Lamb dip" at the center of the laser line was very soon observed by Bennett, Macfarlane and Lamb.[12] It was used for spectroscopy by Szoke and Javant[13], who also showed that the narrow resonance at the dip, free from Doppler-broadening, is sensitive to broadening by collision unless the gas pressure is quite low.

Paul Lee and M. L. Skolnick[14] also showed that if an absorbing gas is present inside the laser resonator, an "inverse Lamb dip" can occur, in which the laser output shows a peak at the center of the absorption line where the absorption of the molecules is saturated by the beams from both directions. The narrow, Doppler-free, optical resonances revealed by the Lamb dip and its inverse have been used for stabilizing the wavelength of lasers.

Thus by the middle of the 1960s it could be seen that, for spectroscopy, laser light possesses several advantages in addition to monochromaticity. Its intensity makes it possible to at least partially saturate absorption or stimulated emission transitions, and so to burn a narrow absorption or emission hole in a Doppler-broadened line. The directionality permits us to observe the combined effects of oppositely directed beams. Thus we could recognize the absorption from just those atoms or molecules which have zero velocity component along a chosen direction, and observe spectral details without Doppler-broadening. But at that time, we could only do so inside the resonator of some laser, and we could only work at those wavelengths where there happened to be laser lines. Later in the decade, Theodor W. Hänsch and Peter Toschek prepared the way for the subsequent advances by using the beam from a second laser on a cascade transition to probe the distribution of molecules as it was affected by saturation inside a laser. [15]

SATURATION SPECTROSCOPY

Laser spectroscopy became much more widely useful when, in 1970, Theodor W. Hänsch,[16,17] and Christian Bordé[18] independently introduced a method which uses these properties of laser light to give Doppler-free spectra of gases external to the laser. As shown in Figure 2, the light from the laser is divided by a partial mirror into two beams which pass through the sample in nearly opposite directions. The stronger "pump" beam is chopped at an audio frequency. When it is on, it is strong enough to partially saturate the absorption of the molecules in the region through which it passes. The probe beam then is less attenuated by its passage through the gas, and a stronger signal reaches the detector. When the chopper obstructs the pump beam, the gas absorption returns, and less of the probe's light reaches the detector. Thus the probe beam is modulated as the pump is alternately turned on and off by the chopper. However, this modulation occurs only when the two beams interact with the same molecules, and that happens only when the laser is tuned to interact with molecules at rest, or at least with zero velocity component along the direction of the beams. Any molecule moving along the beams sees one wave as shifted up

SATURATION SPECTROMETER

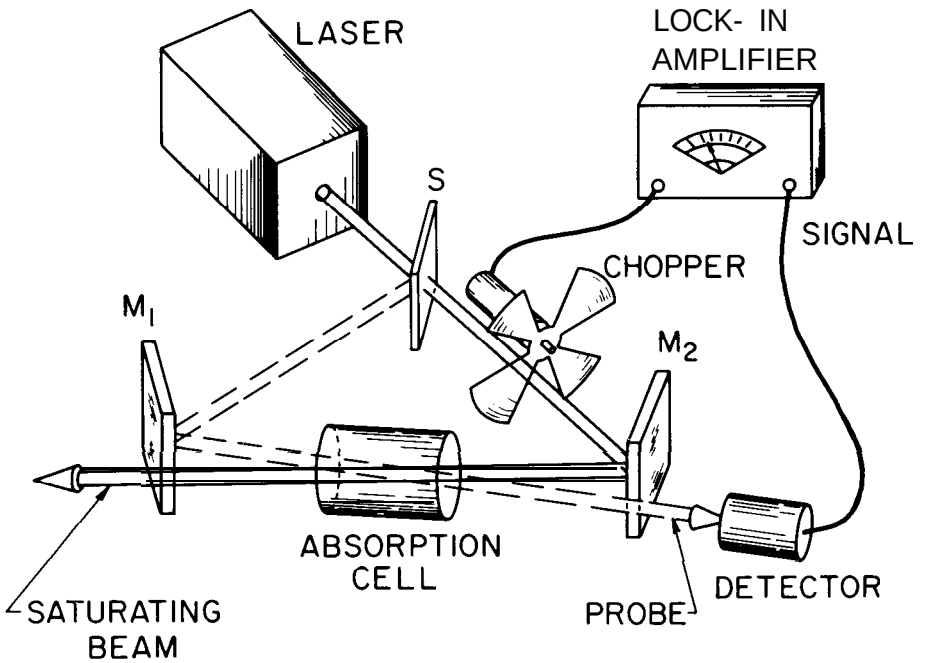


Fig. 2. Apparatus for Doppler-free spectroscopy by saturated absorption in an external cell.

in frequency and the other shifted down, and so a moving molecule cannot be simultaneously in resonance with both beams.

Hänsch and Marc Levenson applied this method first using a single-mode krypton ion laser which could be tuned over a range of about a twentieth of a reciprocal centimeter, that is, 1500 Megahertz, around the wavelengths of a few visible lines. This tuning range, although still quite limited, was enough to explore the details of several of the many lines in the dense visible absorption spectrum of the iodine molecule, I_2. For example, Figure 3 shows the hyperfine structure of a single line in that spectrum, produced by the interaction between the molecular axial field gradient and the quadrupole moments of the two iodine nuclei. Although other workers subsequently attained considerably better resolution by more carefully stabilizing the laser wavelength, the power of the method was already a spectacular advance over what could be done before. Thus, if Figure 3 is projected on a screen two meters wide, on the same scale the visible portion of the spectrum would have a width of more than 500 kilometers! Moreover, the individual lines in the pattern, although still limited by pressure broadening and laser frequency jitter, had a width of about 6 MHz, or about one part in 10^8. The hyperfine structures revealed, which had up to then

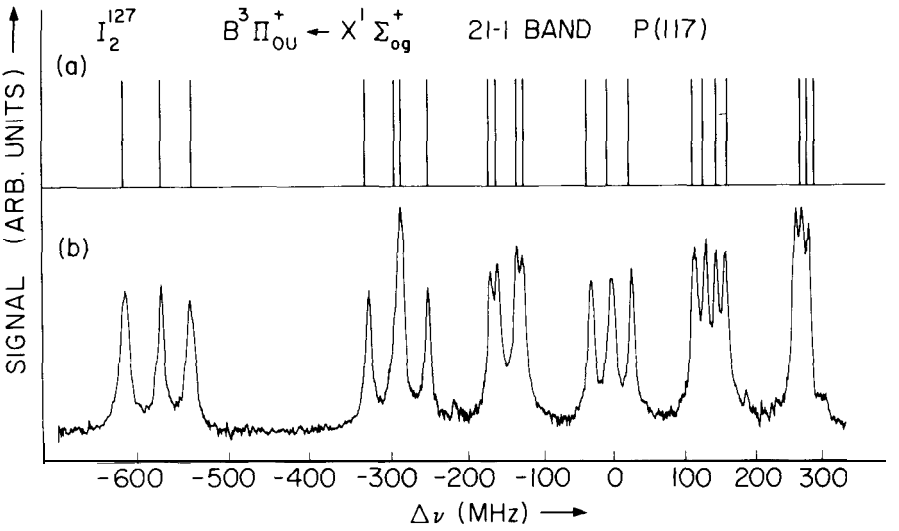


Fig. 3. Hyperfine structure of the P (117) 21-1 B ← X transition of molecular iodine at 568.2 nm (a) theoretical (b) experimental.

always been obscured by the Doppler-broadening of 600 MHz even with the best spectrographs, can be interpreted as in microwave spectroscopy to provide information about the distribution of electrons in the molecule.'"

Narrow as these lines were, they were still broadened by intermolecular collisions at the operating pressure of about one torr. It was easy to reduce the vapor density by cooling the iodine cell, but then the absorption of the probe beam would be negligible whether the pump beam was on or not. However, C. Freed and Ali Javan[20] had shown, in some infrared spectroscopic studies, that when absorption is saturated, any fluorescence that follows from the absorption also shows a saturation. That is, the fluorescence intensity is not linearly proportional to the laser power, but levels off when the laser intensity is enough to deplete the number of molecules contributing to it. In our case the Javan-Freed method was not immediately applicable because, if the two oppositely directed beams were to work together to saturate one of the hyperfine-structure lines, it would cause only a very small change in the unresolved fluorescence in all of the components. Sorem and I, therefore, introduced a method of intermodulated fluorescence.[21] We chopped both of the counterpropagating beams at different audio frequencies, by using two rings with different numbers of holes on the chopping wheel. Our fluorescence detector was tuned to respond to modulation at the sum of the two chopping frequencies, which arose when the stationary molecules were simultaneously excited by the two laser beams. Thus we obtained a good signal, free from Doppler-broadening, even at pressures as low as one millitorr or a thousand times less than we had been able to use with the saturated absorption method.

This is still far from the sensitivity that could ultimately be attained. When continuous wave, broadly tunable lasers became available, William M. Fairbank, Jr. and Theodor W. Hänsch tuned the laser to the orange-yellow wavelength of the sodium resonance lines. With a sodium cell designed to avoid stray light from the walls, we were able to measure the intensity of the light scattered from the sodium atoms, down to as few as a hundred atoms per cubic centimeter.[22] At that density, attained when the cell was cooled to -30" C, there was on the average only one or two atoms at a time in the beam. With this method, we were able to measure the vapor density of sodium metal with a million times greater sensitivity than could previously be obtained. It was evident that laser methods could be much more sensitive than other techniques, such as radioactive methods, for detecting small amounts of suitable substances. A single atom can scatter very many light quanta without being destroyed, and so it should be possible to observe and study a single atom or molecule of a substance. Indeed, in favorable cases this sensitivity can already be achieved, by a method that uses resonant laser excitation followed by ionization.[23] The principal difficulty in making such methods broadly applicable is the lack of suitable lasers at some of the wavelengths needed, especially in the ultraviolet regions.

BROADLY TUNABLE LASERS

During the 1960s there was a rapid growth in discoveries of new laser materials and ways to excite them. Solids, liquids, and gases were made to produce laser action under optical excitation, as well as electrical discharges in gases and semiconductors. But each of them operated at its own characteristic wavelength, determined by the properties of the material, and there was no way to obtain an arbitrary wavelength even in the same spectral region. We did not at first expect to be able to produce laser operation over a continuous band of wavelengths, because we knew that the available optical amplification was inversely proportional to the width of the lasing line or band. Nevertheless, Peter Sorokin and J. R. Lankard[24] and, independently, Fritz Schäfer[25] were able to use intense flashlamps to excite laser action in organic dyes, whose emission bands could be as wide as a hundred Angstrom units or more.

A further advance came when it was realized that the high light intensity to pump broadband laser materials could be best obtained from another laser. I must admit that at first I wondered why anyone would want to compound the inefficiencies by using one laser to pump another. But when you need concentrated light for pumping, a suitable laser is a good way to get it. Thus J. A. Myer, C. L. Johnson, E. Kierstead, R. D. Sharma, and I. Itzkan[26] used a pulsed ultraviolet nitrogen laser to pump a tunable dye laser. As shown in Figure 4, the dye laser can consist of just a cell containing a dye in solution, an output mirror and a diffraction grating. The grating replaces the second mirror of the ordinary laser structure, and acts as a good mirror for one wavelength that changes as the grating is rotated. By now, dyes are available to give laser action at all wavelengths in the visible, extending into the near ultraviolet and

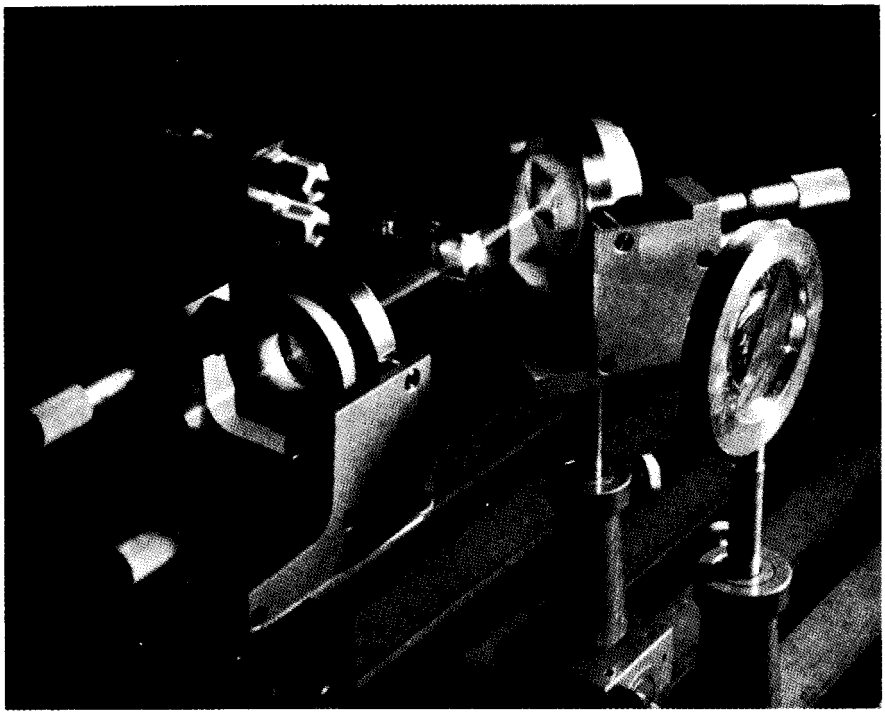


Fig. 4. Photograph of a simple pulsed dye laser, pumped by an ultraviolet beam from a nitrogen laser, and tuned by a diffraction grating.

infrared. When pumped by a nitrogen laser, the dye laser typically gives pulses of a few nanoseconds duration. The amplification is very high, so that the end mirror and the diffraction grating or other tuning element do not need to have very high reflectivity.

But such a simple dye laser of this kind typically gives an output too widely spread in wavelength to be useful for high resolution spectroscopy. Hänsch was able to obtain narrow line output by adding a telescope between the dye cell and grating. Then the light at the grating was spread over more of the rulings, and was better collimated, so that the sharpness of its tuning was improved. To get output with sub-Doppler narrowness limited only by the length of the light pulse, or about 300 Megahertz, he placed a tilted etalon in front of the grating (Figure 5).[27] Even more monochromatic output, with a corresponding increase in pulse length could be obtained by filtering the output through a passive resonator.

Continuous-wave dye laser operation was obtained in 1970 by Peterson, Tuccio, and Snavely,[28] who used an argon ion laser to excite it. The pumping laser and dye laser beams were collinear, as had been used for C.W. ruby lasers by Milton Birnbaum,[29] rather than in the transverse arrangement used in most earlier lasers. Refining the output of continuous-wave lasers presented difficulties because the available amplification was small, so that any tuning

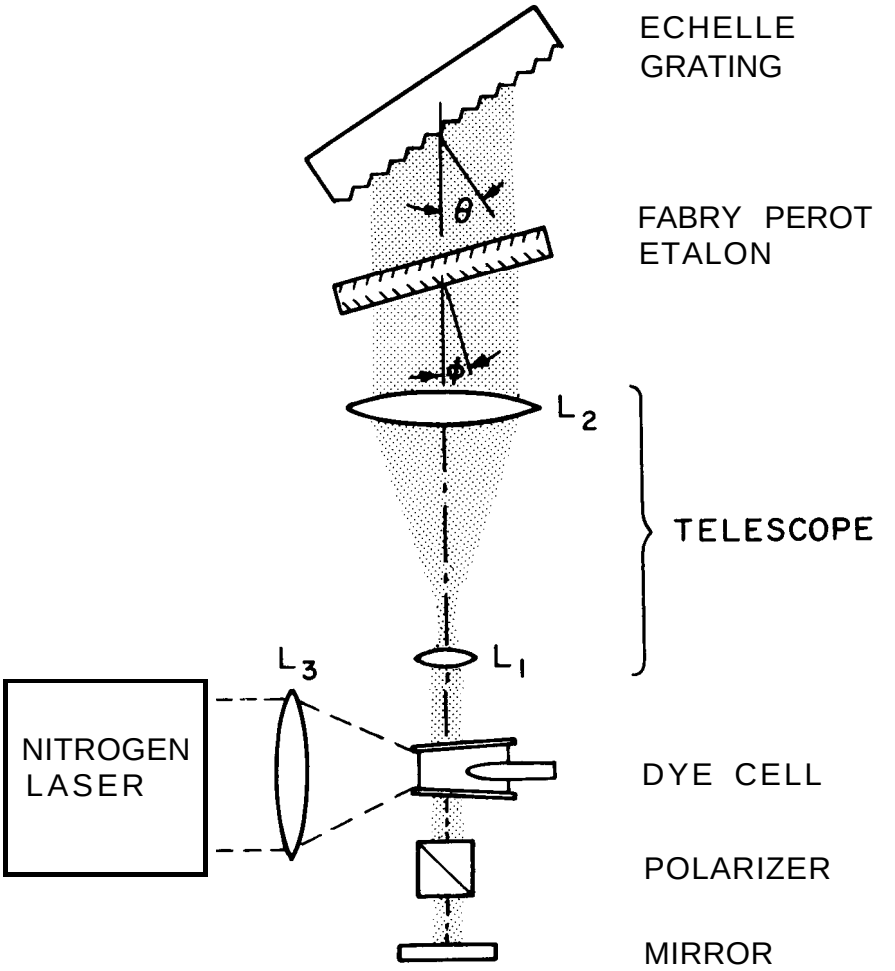


Fig. 5. Diagram of an improved dye laser, with a telescope and an etalon between the dye cell and the diffraction grating.

elements had to present low losses. But by now, extremely stable, narrowband lasers have been made, with line widths much less than one Megahertz.

Once a powerful, narrowband but broadly tunable, laser was available, it became possible to adapt the laser to the problem rather than the reverse. The methods of saturation spectroscopy could be applied to examine in detail the spectral lines of atoms simple enough to be of theoretical interest. With the pulsed dye laser, Hänsch and Issa S. Shahin first obtained Doppler-free spectra of the sodium atom's D lines at 5890 and 5896 Angstrom units, with the ground state hyperfine structure clearly resolved.[30] Then they applied it to study the fine structure of the red line H_α of atomic hydrogen."' For this purpose, they constructed an electric glow discharge tube with end windows through which the two beams could pass. One beam was, as before, the saturating beam, while the other, weaker beam was the probe to detect the absorption from atoms with no velocity component along the beam direction.

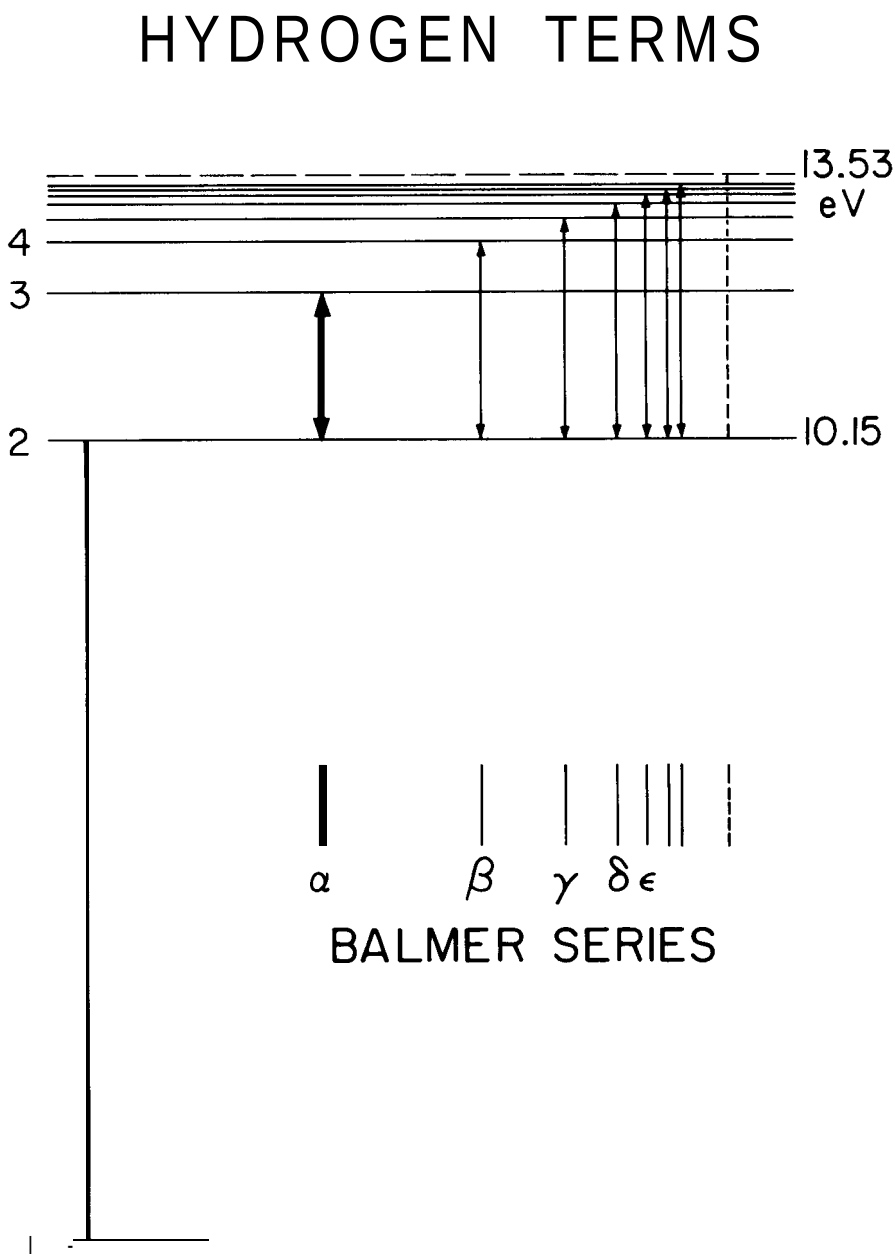


Fig. 6. Energy levels and transitions of atomic hydrogen

Figure 6 recalls the energy levels and spectral lines of the hydrogen atom according to the quantum theory of Niels Bohr. The transitions from higher levels to the level with principal quantum number, n, equal to 2 give rise to the Balmer series spectrum, drawn at the top of Figure 7. Below it is shown the fine structure of the red line, on a scale expanded by a factor of 40000, as it would be

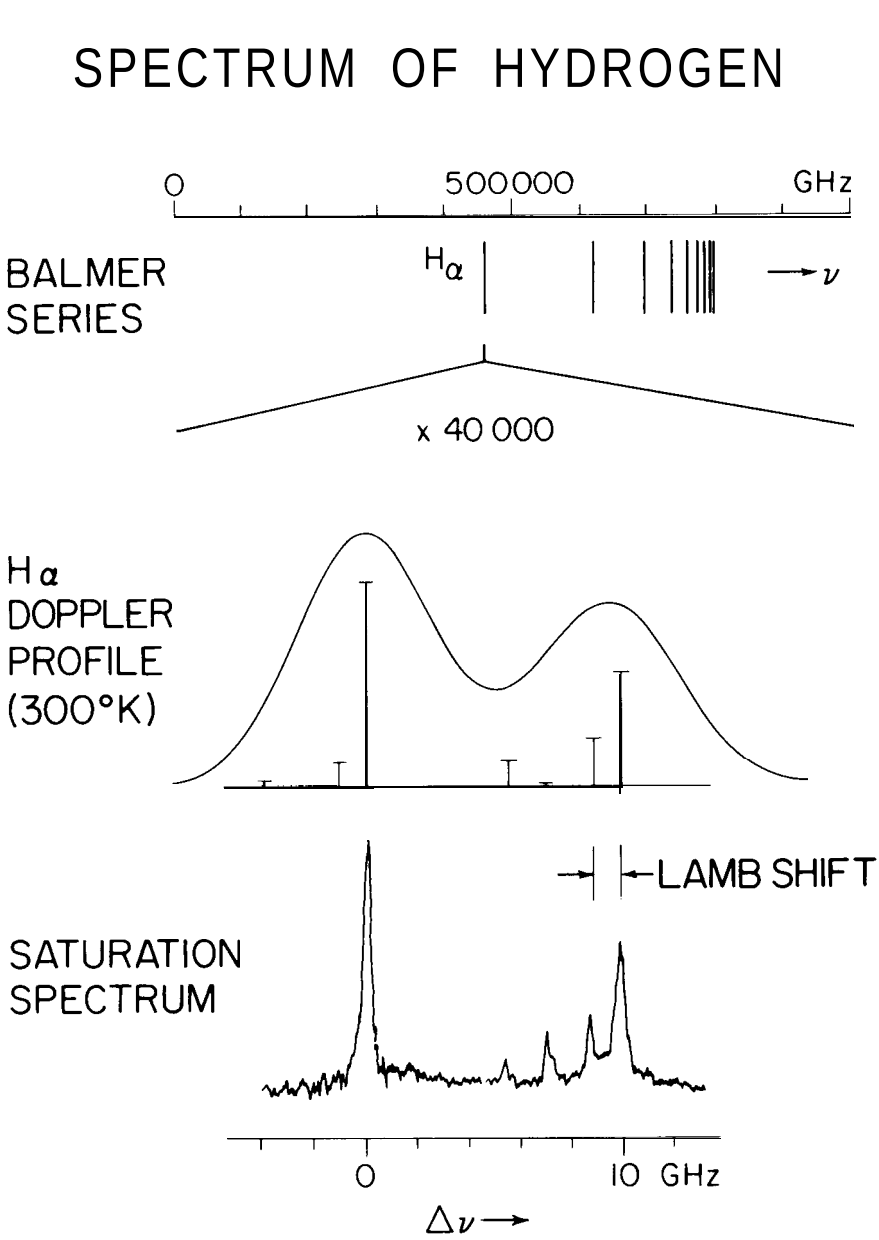


Fig. 7. Hydrogen Balmer series and fine structure of the red line Hα, resolved by saturation spectroscopy.

revealed by a perfect conventional spectrograph. The line is known, from theory and from radiofrequency measurements, to have the several line structure components indicated, but they would be nearly obscured by the large Doppler width. At the bottom of the Figure is shown the tine structure of this line revealed by laser saturation spectroscopy. The improvement is dramatic, and most of the details of the fine structure can be clearly seen. In particular,

the Lamb shift between the $2s_{1/2}$ and the $2p_{1/2}$ levels is clearly revolved, which had not been possible previously in hydrogen, although Gerhard Herzberg had resolved the Lamb shift in the corresponding line of ionized helium, where the shift is four times greater.[32]

Microwave measurements had already given an accurate account of all of these details, and the optical resolution had little hope of improving on them. What could be done much better than before was to make an accurate determination of the wavelength of one of the components, and thereby obtain an improved value of the Rydberg constant. Of course, if the positions and relative intensities of the components are known, it is possible to compute the line shape and compare it precisely with the shape and position determined by optical spectroscopy. But the relative intensities are determined by the detailed processes of excitation and deexcitation in the gas discharge, and their uncertainty was the principal source of error in earlier measurements of the line wavelength.

Hänsch and his associates Issa S. Shahin, Munir Nayfeh, Siu Au Lee, Stephen M. Curry, Carl Wieman, John E. M. Goldsmith, and Erhard W. Weber, have refined the measurement of the wavelength of the line and thereby of the Rydberg constant, through a series of careful and innovative researches.[33] They have improved the precision by a factor of about eighty over previous work, so that the Rydberg is now one of the most accurately known of the fundamental constants. The value obtained, $R_\infty = 109737.3148 \pm .0010$ cm^{-1} is in good agreement with that obtained in a recent experiment using laser excitation of an atomic beam, by S. R. Amin, C. D. Caldwell, and W. Lichten.[34]

In the course of these investigations, Wieman and Hänsch[35] found a new method to increase the sensitivity of the saturation method for avoiding Doppler-broadening. As shown in Figure 8, they used a polarized pump beam. It preferentially excites molecules with some particular orientation, leaving the remainder with a complementary orientation. The probe beam is sent through two crossed polarizers, one before and one after the sample region, so that no light reaches the detector except at the wavelengths where the light is depolarized by the oriented molecules. The saturation signal then appears as an

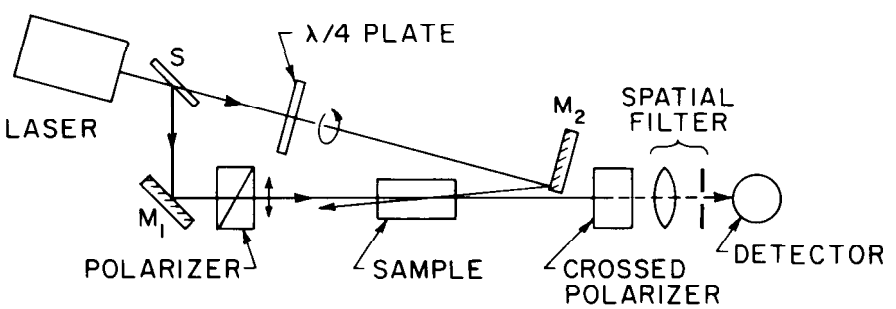


Fig. 8. Apparatus for Doppler-free polarization spectroscopy

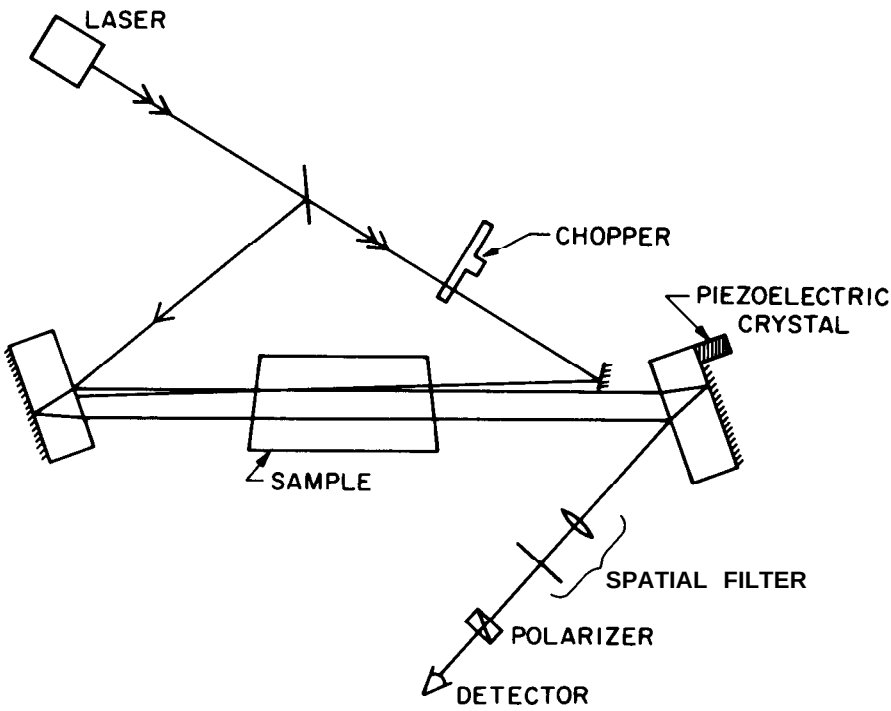


Fig. 9. Apparatus for Doppler-free saturated interference spectroscopy.

increased transmission with nearly no interfering background. Thus the noise caused by fluctuations in intensity of the probe laser is nearly eliminated, and the spectra can be observed at lower density or with lower light intensity. This method is now known as polarization spectroscopy.

Another way of balancing out the background was introduced by Frank V. Kowalski and W. T. Hill[36] and, independently, by R. Schieder.[37] They used a configuration like a Jamin interferometer, in which the probe beam is split into two parts which travel parallel paths through the sample cell, as shown in Figure 9. The beams are recombined in such a way that they cancel each other. Then when a saturating beam reduces the absorption along one of the paths, the interferometer becomes unbalanced and a Doppler-free signal is seen. In a way, polarization spectroscopy can be thought of as a special case of saturated-interference spectroscopy. The plane polarized probe wave is equivalent to two waves circularly polarized in opposite senses. They combine to produce a plane wave of the original polarization, which is stopped by the second polarizer, unless one of the two circularly polarized components experiences a different absorption or a different effective path length than the other.

SIMPLIFYING SPECTRA BY LASER LABELING

Spectra of molecules are very much more complicated than those of atoms. Even a diatomic molecule such as Na_2 has dozens of vibrational and hundreds of rotational levels for every electronic level. We have, therefore, sought systematic ways to use lasers to simplify molecular spectra so as to identify their various states. Even before lasers, something like this could be done by using a monochromatic light source, such as a filtered mercury lamp, to excite just one level, and observing the fluorescence from it to lower levels. With monochromatic, tunable lasers this can be used, for instance, to explore the vibrational and rotational structure of the ground electronic state of molecules. The upper state in this case may be said to be "labeled," since it is identified by having molecules excited to it, while neighboring states have none.

But if anything at all is known about a molecule, it is likely to be the constants of the ground electronic state, which can also be studied by microwave, infrared and Raman spectroscopy. Mark E. Kaminsky, R. Thomas Hawkins, and Frank V. Kowalski therefore inverted the process, by using a laser to pump molecules out of a chosen lower level.[38] All of the absorption lines originating on this chosen level were, then, weakened. If the pumping laser was chopped, the absorption lines from the labeled level were modulated at the chopping frequency. Thus when a high-resolution optical spectrometer was scanned across the spectrum, the lines from the labeled level could be recognized by their modulation, even if perturbations displaced them far from their expected position.

Almost as soon as Hänsch and Wieman introduced the method of polarization spectroscopy, it was apparent to us that it could be adapted for searching for and identifying the levels of molecules or complex atoms.[39] Apparatus for the polarization labeling method is shown in Figure 10. A polarized beam from a repetitively pulsed dye laser is used to pump molecules of a particular orientation from a chosen lower level, and leave the lower level with the complementary orientation. A broadband probe from a second laser is directed through two crossed polarizers, before and after the sample, and then into a

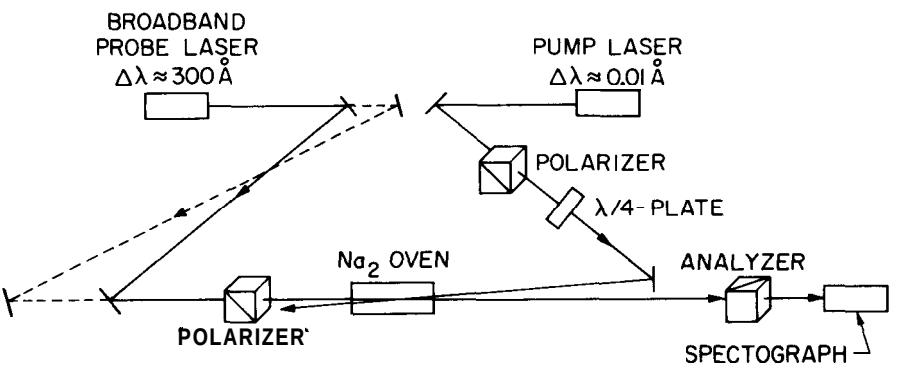


Fig. 10. Apparatus for simplifying spectra by polarization labeling.

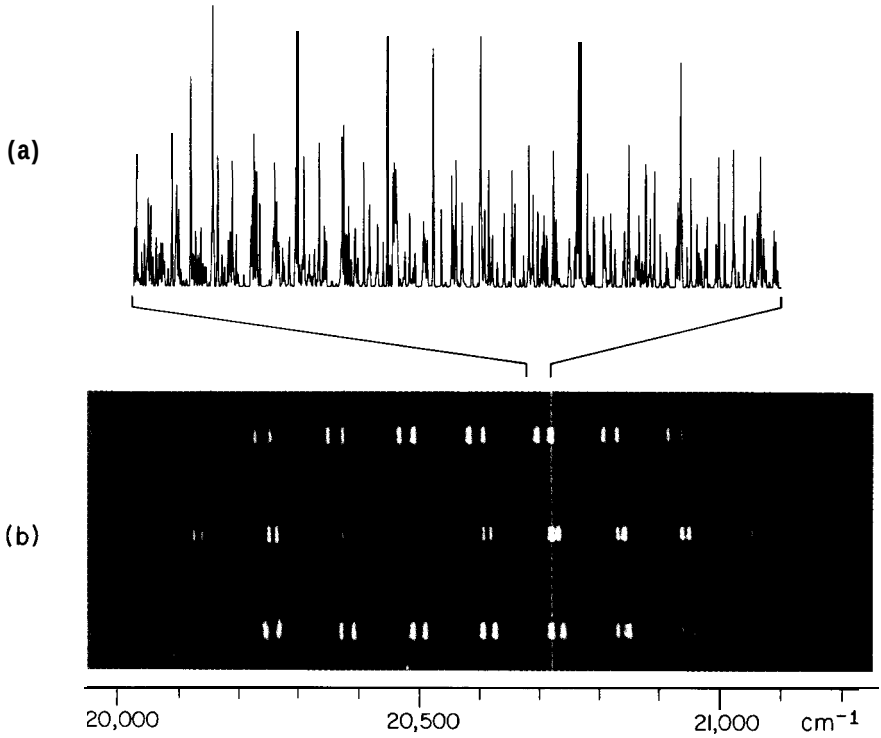


Fig. Il. A small section of the Na_2 spectrum revealed by conventional spectroscopy and by polarization labeling.

photographic spectrograph. Figure 11 shows the spectra observed by Richard Teets and Richard Feinberg,[39] as several neighbouring lines of N_2 are pumped. It is seen that from each labeled level, there are just two rotational lines for each vibrational level (J' = J" + 1 and J' = J" - 1). A small portion of the spectrum as obtained by simple absorption spectroscopy is shown for comparison. As the tuning of the pump laser is changed slightly, different groups of lines appear. For each of them, the upper state vibrational quantum number can be inferred simply by recognizing that the lines of lowest frequency end on the v' = 0 level of the upper electronic state.

As the molecules raised to the excited electronic level by the polarized pump laser are also oriented, the probe can record transitions from that to still higher levels. Nils W. Carlson, Antoinette J. Taylor, and Kevin M. Jones[40, 41, 42] have identified 24 excited singlet electronic states in Na_2 by this method, whereas all previous work on this molecule had only produced information about 6 excited states. The new levels include 1, II, and △ states from the electron configurations 3sns and 3snd, as indicated in Figure 12. For larger values of n, these arc molecular Rydberg states, with one electron far outside the core of two Na^+ ions bound mostly by the single 3s electron. In the △ states the outer electron contributes something to the bonding, so that the depth of

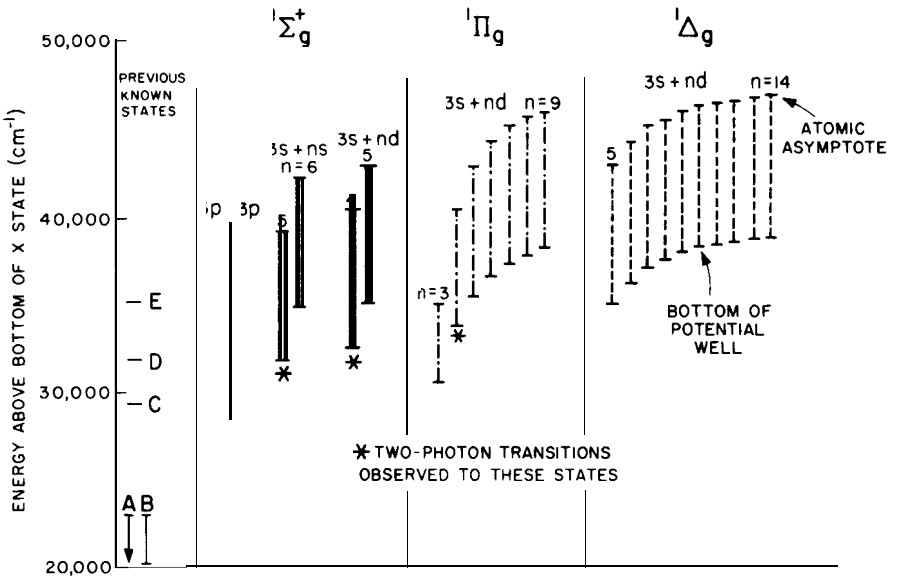


Fig. 12. Excited electronic states of Na_2, as revealed by two-step polarization labeling

the potential well increases as n is decreased which brings the outer electron closer to the core, as is seen in Figure 13. In the II states, the outer electron is antibonding, and so it decreases the molecular bonding when it is close to the core. Corresponding behaviors are observed for the other molecular constants, vibrational energy and bond length. Thus they can all be extrapolated to obtain good values for the constants of the ground state of the Na_2^+ ion.[36] This method and the several related techniques of optical-optical double resonance arc making increasing contributions to the analysis of complex atomic and molecular spectra.

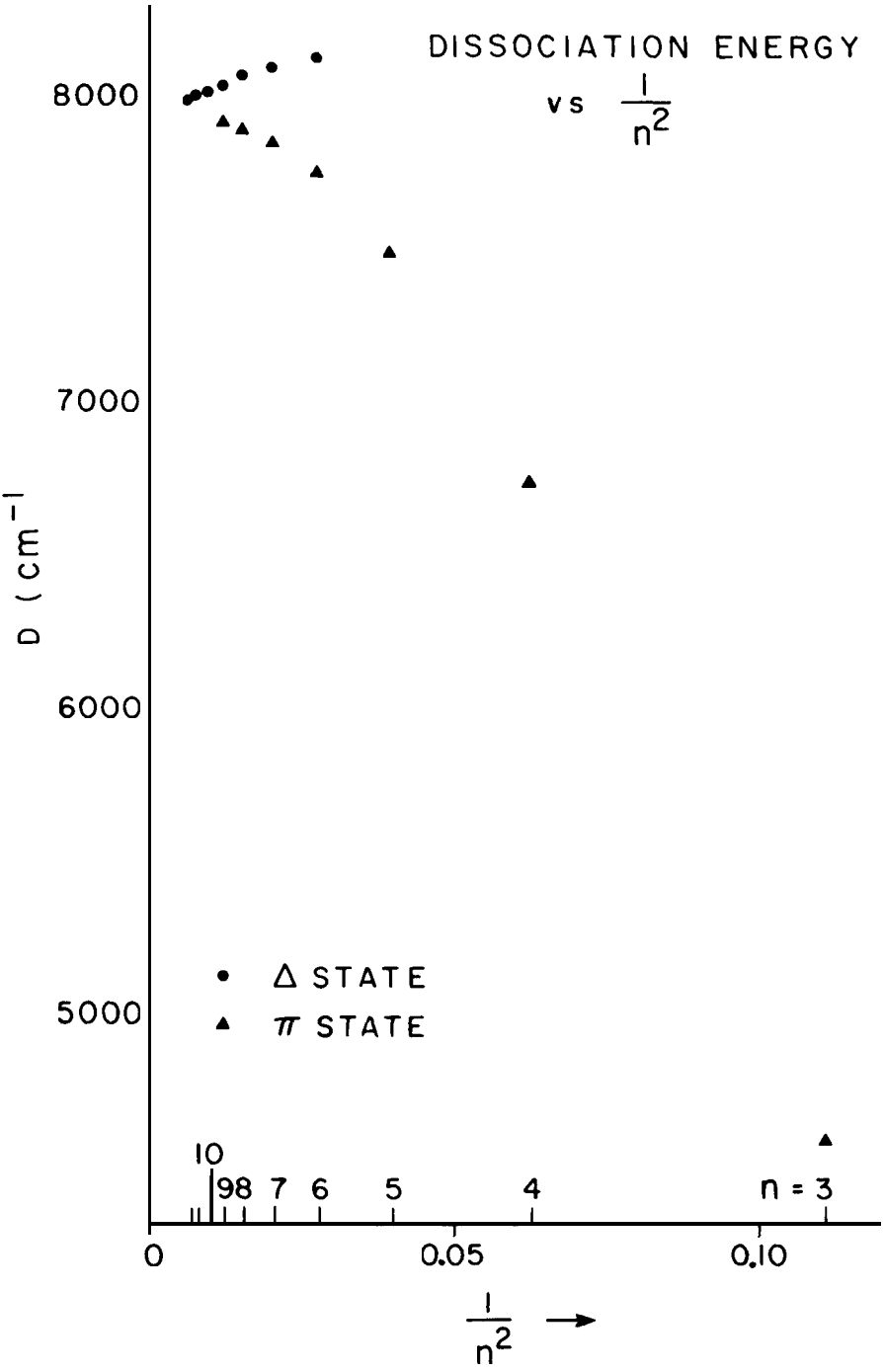


Fig. 13. Dissociation energy for Na_2 molecular Rydberg states as a function of l/n^2.

TWO-PHOTON DOPPLER-FREE SPECTROSCOPY

In 1970, L. S. Vasilenko, V. P. Chebotayev, and A. V. Shishaev proposed a method for obtaining two-photon spectral lines without Doppler-broadening.[43] As shown schematically in Figure 14, a molecule moving along the

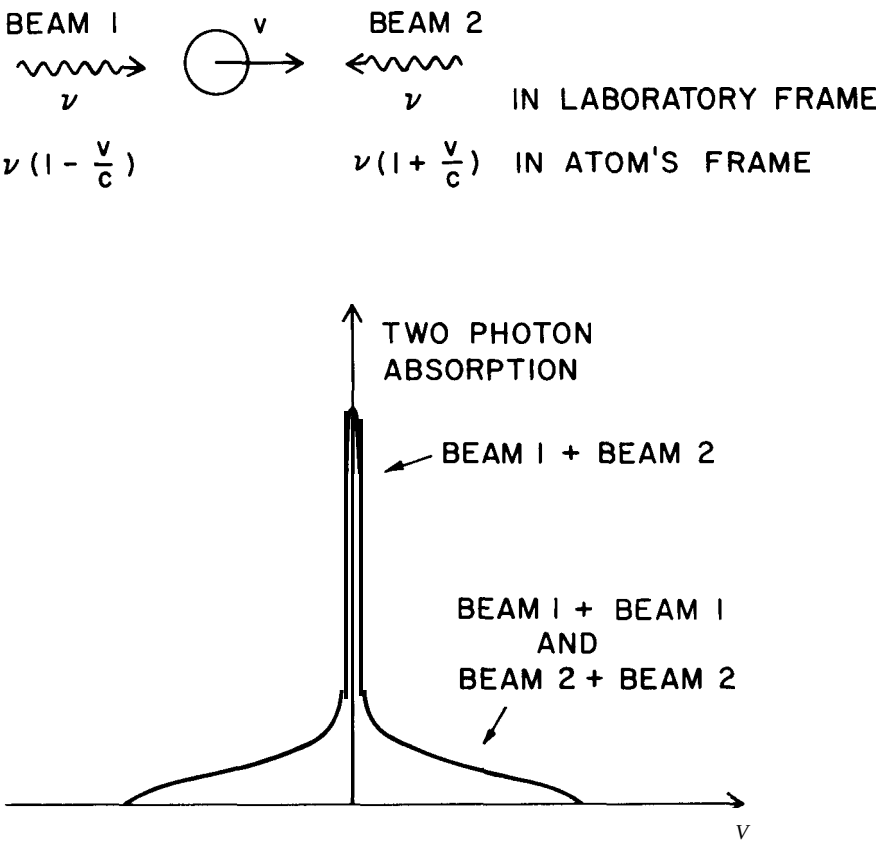


Fig. 14. Principles of Doppler-free two-photon spectroscopy.

direction of one of two oppositely-directed light beams, from the same laser, sees one of them shifted up in frequency and the other shifted down by just the same amount. Thus the sum of the photon frequencies is, to first order, unaffected by the Doppler shifts. All molecules contribute equally to the Doppler-free two-photon line. The predicted effect was observed in atomic sodium vapor by B. Cagnac, G. Grynberg, and F. Biraben,[44] by M. D. Levenson and N. Bloembergen,[45] and by Hänsch, K. C. Harvey, and G. Meisel.[46] They observed transitions from the 3s ground state to 4d or 5s levels (Figure 15). The lines were not only sharp, but remarkably prominent and easy to detect because of the presence of very strong allowed transitions, the well known sodium D lines, less than 100 Angstroms from the wavelength needed

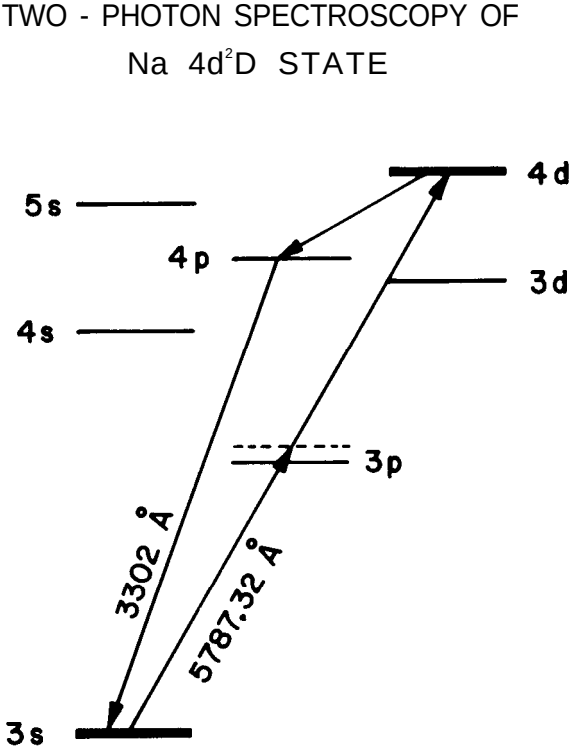


Fig. 15. Energy levels and some two-photon transitions in sodium atoms.

for the two-photon transitions. The allowed transition makes the atom more polarizable at the light frequency. Thus it enhances the two-photon absorption coefficient by a factor proportional to the inverse square of the offset, or frequency difference between the light frequency and the frequency of the allowed transition. Subsequently, R. T. Hawkins, W. T. Hill, F. V. Kowalski, and Sune Svanberg[47] were able to use two lasers of different frequencies in the beams, and so to take advantage of different enhancing lines to reach a number of other levels in the sodium atom, and to measure the Stark shifts caused by an applied electric field. They used a roughly collimated atomic beam illuminated transversely by the lasers, to provide further reduction of the Doppler-broadening.

It was rather surprising, in Kenneth Harvey's early work, that some other two-photon lines were seen in the neighborhood of the expected atomic lines.[48] Since they did not exhibit the well known hyperfine structure of the sodium ground state, they could only come from molecular sodium, Na_2 (Figure 16). But that was remarkable, because the number of molecules at that temperature was very small in comparison with the number of atoms, and there would be still fewer in any individual level. Yet the molecular lines were as strong as the atomic lines. We realized that the explanation must be a more or less accidental close coincidence with some allowed, and therefore enhancing, molecular line.

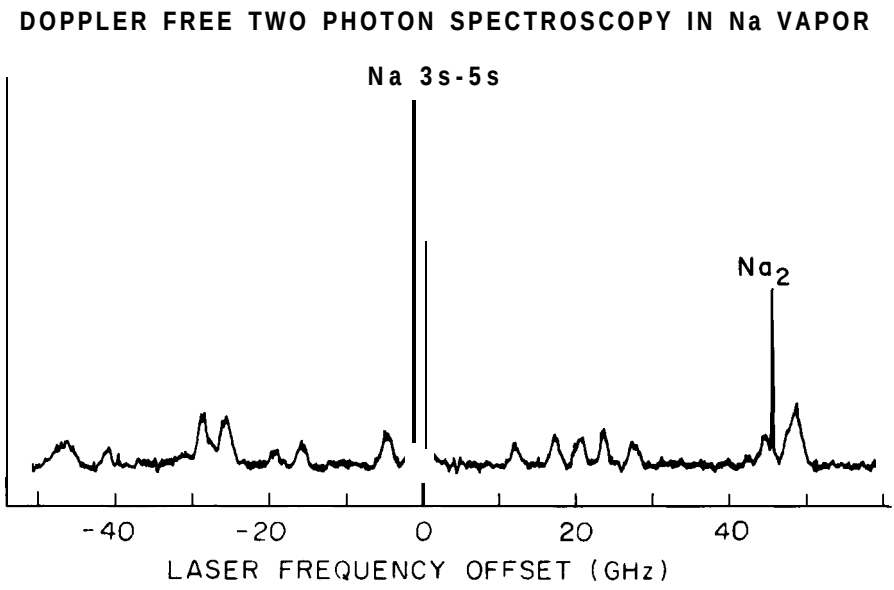


Fig. 16. Atomic and molecular two-photon lines in Na_2.

J. P. Woerdman[49] also observed some of these lines, and was able to identify the rotational quantum number through the nearby enhancing line of the $A \leftarrow X$ band of Na_2. Recently Gerard P. Morgan, Hui-Rong Xia, and Guang-Yao Yan[50,51] have found and identified a large number of these strong two-photon lines in Na_2. The offsets from neighboring enhancing lines have been measured by simultaneous one-photon and two-photon Doppler-free spectroscopy. They are indeed small, ranging from 0.1 cm^{-1} to as little as 38 Megahertz or about .001 cm^{-1}. Thus we see how it is possible to have the probability of two-photon absorption, and thus two steps of excitation, nearly as strong as that of a single step.

On the other hand, two-photon Doppler-free lines may also be observed if there is no enhancing state anywhere near, if there is enough laser intensity and sufficiently sensitive detection. Thus Hänsch, Siu Au Lee, Richard Wallenstein, and Carl Wieman[52,53] have observed the 1s to 2s two-photon transition in atomic hydrogen, excited by the second harmonic (2430 Angstroms) of a visible dye laser which simultaneously scans the blue H_β Balmer line. They have, thus, made an accurate comparison between the 1s to 2s interval and four times the 2s to 4s interval in hydrogen. According to the Bohr theory, the ratio of these level spacings should be exactly 4 to 1. The deviation observed is a measure of the Lamb shift in the ground 1s state, which is otherwise not measurable.

The 1s to 2s transition is particularly intriguing, because the lower state is stable, and the upper state is metastable so that it has a lifetime of 1/7 second. Thus the lifetime width need to be no more than one Hertz, or a part in 10^{15}. Since we can usually locate the center of a line to one percent of the linewidth, it

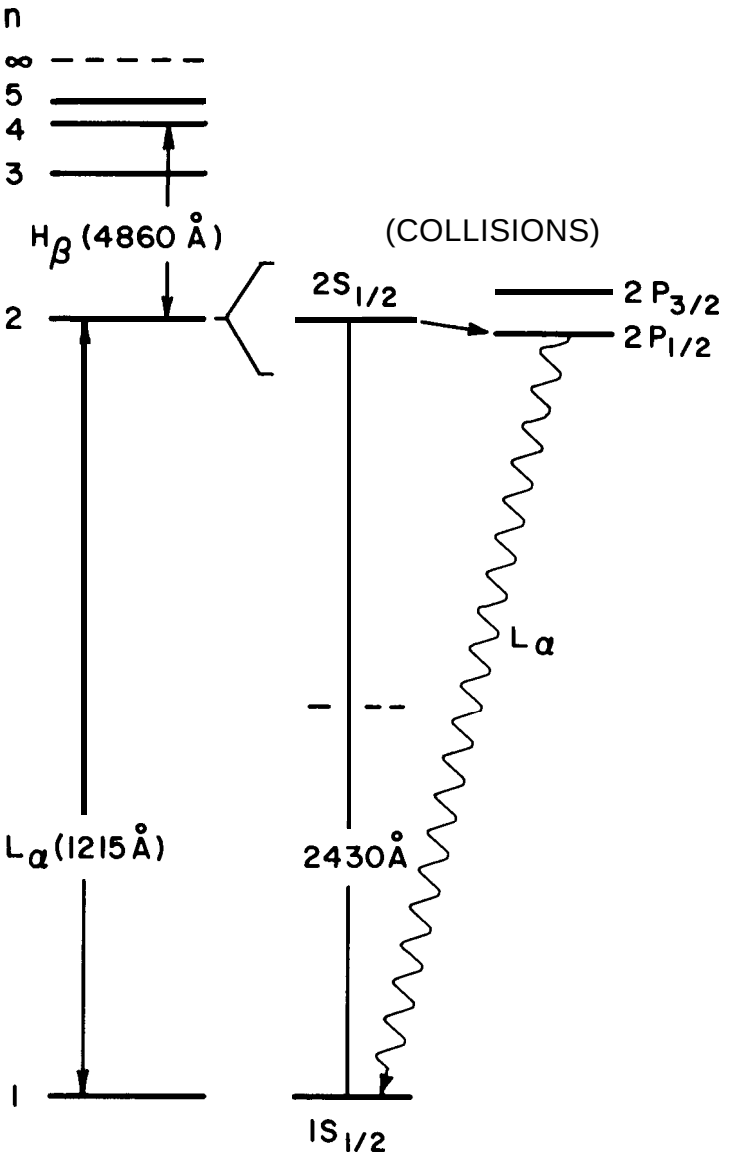


Fig. 17. Energy levels and transitions for measuring the Lamb shift of the 1s level of atomic hydrogen.

should be possible eventually to measure this line to one part in 10^{17} or so. But nobody measures anything to a part in 10^{17}! Before we can hope to achieve that, such things as second order Doppler effect, transit time broadening, radiation recoil and power broadening will have to be eliminated. The challenge is great, and should occupy experimental physicists for some years.

OTHER METHODS

There is not enough room to discuss all of the laser spectroscopic methods that have interested our colleagues. Serge Haroche and Jeffrey A. Paisner have used short, broadband laser pulses to produce quantum beats in fluorescence, by exciting a coherent superposition of several hyperfine levels.“” James E. Lawler extended the methods of optogalvanic detection of laser absorption”” to detect Doppler-free intermodulation and two-photon lines.“” In turn, this method has been extended by Donald R. Lyons and Guang-Yao Yan,[57] to use electrodeless radiofrequency detection of Doppler-free resonances.

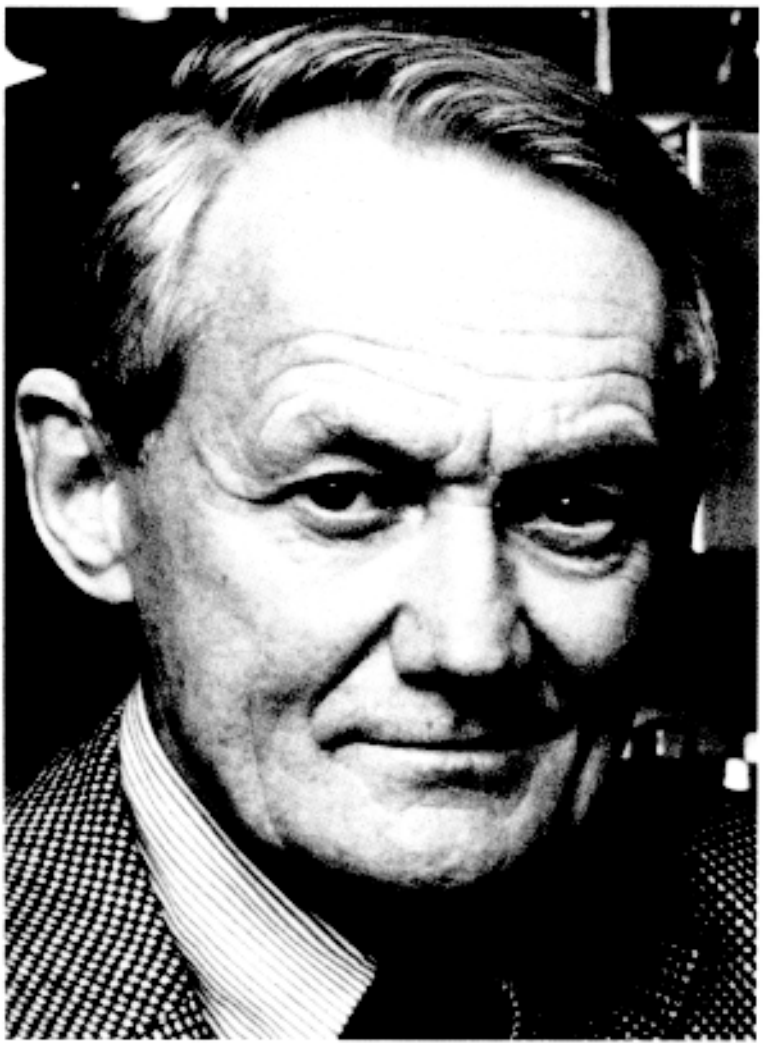
Even less is it possible to begin to describe the many exciting discoveries and developments from other laboratories. Some indication of them can be obtained from the proceedings of the five biennial conferences on Laser Spectroscopy.“s The field has had an almost explosive growth, and laser spectroscopy in some form or other extends from the submillimeter wavelengths in the far infrared to the vacuum ultraviolet and soft x-ray regions.

Thus in the powerful, directional, coherent and highly monochromatic new light of lasers, we are learning to do entirely new kinds of spectroscopy. We can resolve fine details hitherto obscured by thermal broadening, can observe and study very small numbers of atoms, and can simplify complex spectra. We can take the measure of simple atoms with a precision that is providing a real challenge to the best theoretical calculations. Our experimental capabilities have been extended so rapidly in the past few years, that there has not been time to bring them fully to bear on the interesting, fundamental problems for which they seem so well suited. But the spectroscopy with the new light is illuminating many things we could not even hope to explore previously, and we are bound to encounter further intriguing surprises.

REFERENCES

1. Angström, A. J., Kungl. Svenska Vetensk. Akad. Handl., p. 327 (1852); Poggendorff Annalen 94, 141 (1855); Phil. Mag. 9, 327 (1855)
2. Fraunhofer, J., Gilbert’s Annalen 56, 264 (1817).
3. Huggins, W., Phil. Trans. Roy. Soc. London 171, 669 (1880).
4. Vogel, H. W., Monatsbericht der Königl. Academie der Wissenschaften zu Berlin, July 10, 1879, reported by Huggins, op. cit.
5. Balmer, J. J., Poggendorff's Annalen, N. F. 25, 80 (1885).
6. Rydberg, J. R., Kungl. Svenska Vetensk. Akad. Handl. 32, 80 (1890); Phil. Mag. 29, 331 (1890).
7. Meissner K. W., and Luft, K. F.. Ann. der Physik, 28, 667 (1937).
8. Minkowski, R., and Bruck, G., Zeits. f. Phys. 95, 284 (1935).
9. Schawlow, A. L., and Townes, C. H., Phys. Rev. 112, 1940 (1958).
10. Tiffany, W. B., Moos, H. W., and Schawlow A. L., Science, 157, 40 (1967).
11. Lamb, W. E. talk at the *Third International Conference on Quantum Electronics*, Paris. February 11 15, 1963; Phys. Rev. *134A*, 1429 (1964).

12. McFarlane, R. A., Bennett, W. R. Jr., and Lamb, W. E., Jr, Appl. Phys. Letters, 2, 189 (1963).
13. Szoke. A., and Javan, A., Phys. Rev. Letters IO, 521 (1963).
14. Lee, P. H., and Skolnick, M. L., Appl. Phys. Letters 10, 303 (1967).
15. Hänsch, T. W. and Toschek, P., I.E.E.E. J. Quant. Electr. 4, 467 (1968).
16. Smith, P. W., and Hänsch, T. W., Phys. Rev. Letters 26, 740 (1971).
17. Hänsch, T. W., Levenson, M. D., and Schawlow, A. L., Phys. Rev. Letters 27, 707 (1971).
18. Bordé, C., C. R. Acad. Sci. Paris 271, 371 (1970).
19. Levenson, M. D., and Schawlow, A. L., Phys. Rev. A6, 10 (1972).
20. Freed, C., and Javan, A., Appl. Phys. Letters 17, 53 (1970).
21. Sorem, M. S., and Schawlow, A. L.. Opt. Comm. 5, 148 (1972).
22. Fairbank, W. M., Hänsch, T. W., Jr., Schawlow, A. L., J. Opt. Soc. Am. 65, 199 (1975).
23. Hurst, G. S., Payne, M. G., Kramer, S. D., and Young, J. P., Rev Mod. Phys. *51, 767* (1979).
24. Sorokin, P. P., and Lankard, J. R., IBM J. Res. Develop. *IO, 306 (1966).*
25. Schafer, F. P., Schmidt, W., and Volze, J., Appl. Phys. Letters 9, 306 (1966).
26. Myer, J. A., Johnson, C. L., Kierstead, E., Sharma, R. D., and Itzkan, I., Appl. Phys. Letters 16, 3 (1970).
27. Hänsch, T. W., Appl. Opt. *11, 895 (1972).*
28. Peterson, 0. G., Tuccio, S. A., and Snavely, B. B., Appl. Phys. Letters 17, 245 (1970).
29. Birnbaum, M., Wendizowski, P. H., and Fincher, C. L., Appl. Phy. Letters *16, 436* (1970).
30. Hänsch, T. W., Shahin, I. S., and Schawlow, A. L., Phys. Rev. Letters 27, 707 (1971).
3 I. Hänsch, T. W., Shahin, I. S., and Schawlow, A. L., Nature 235, 63 (1972).
32. Herzberg, G., Physik, 146, 269 (1956).
33. Goldsmith, J. E. M., Weber, E. W. and Hänsch, T. W., Phys. Rev. Letters 41, 1525 (1978).
34. Amin, S. R., Caldwell, C. D., and Lichten, W., Phys. Rev. Letters 47, 1234 (1981).
35. Wiemann C., and Hänsch, T. W., Phys. Rev. Letters 36, 117O (1976).
36. Kowalski, F. V., Hill, W. T., and Schawlow, A. L., Optics Letters, 2, 112 (1978).
37. Schieder, R., Opt. Comm. 26, 113 (1978).
38. Kaminsky, M. E., Hawkins, R. T., Kowalski, F. V., and Schawlow, A. L., Phys. Rev. Letters 36, 671 (1976).
39. Teets, R. E., Feinberg, R., Hänsch, T. W., and Schawlow, A. L., Phys. Rev. Letters 37, 683 (1976).
40. Carlson, N. W., Kowalski, F. V., Teets, R. E., and Schawlow, A. L., Opt. Comm. 18, 1983 (1979).
41. Carlson, N. W., Taylor, A. J., Jones, K. M., and Schawlow A. L., Phys. Rev. *A24,822* (1981).
42. Taylor, A. J., Jones, K. M., and Schawlow, A. L., Opt. Comm. 39, 47 (1981).
43. Vasilenko, L. S., Chebotayev, V. P., and Shishaev, A. V., JETP Letters 12, 113 (1970).
44. Cagnac, B., Grynberg, G., and Biraben, F., J. Phys. (Paris) 34, 845 (1973).
45. Levenson, M. D., and Bloembergen, N., Phys. Rev. Letters 32,645 (1974).
46. Hänsch, T. W., Harvey, K. C., Miesel, G. and Schawlow, A. L., Opt. Comm. 11, 50 (1974).
47. Hawkins, R. T., Hill, W. T., Kowalski, F. V., Schawlow, A. L., and Svanberg, S., Phys. Rev. *A15, 967* (1977).
48. Harvey, K. C., thesis, Stanford University, M. L. Report No. 2442 (1975).
49. Woerdman, J. P., Chem. Phys. Letters 43, 279 (1976).
50. Xia, H.-R., Yan, G.-Y., and Schawlow, A. L., Opt. Comm. 39, 153 (1981).
51. Morgan, G. P., Xia, H.-R.; and Schawlow A. L., J. Opt. Soc. Am., 72, 315 (1982).
52. Lee, S. A., Wallenstein, R., and Hänsch, T. W., Phys. Rev. Letters 35, 1262 (1975).
53. Wieman, C., and Hänsch, T. W., Phys. Rev. *A22,* 1 (1980).
54. Haroche, S., Paisner, J. A., and Schawlow, A. L., Phys. Rev. Letters 30, 948 (1973).
55. Green, R. B., Keller, R. A., Luther, G. G., Schenck, P. K., and Travis, J. C., Appl. Phys. Letters 29, 727 (1956).
56. Lawler, J. E., Ferguson, A. I., Goldsmith, J. E. M., Jackson, D. J., and Schawlow, A. L., Phys. Rev. Letters 42, 1046 (1979).
57. Lyons, D. R., Schawlow, A. L., and Yan, G.-Y., Opt. Comm. 38, (1981).
58. The most recent of these is *Laser* Spectroscopy *V,* McKellar, A. R. W., Oka, T., and Stoicheff, B. P., editors, Springer-Verlag, Berlin, Heidelberg and New York (1981).



Kai Siegbahn

KAI SIEGBAHN

Born April 20, 1918, in Lund, Sweden. Parents: Manne Siegbahn and Karin Högbom. Married May 23, 1944, to Anna Brita Rhedin. Three children: Per (1945), Hans (1947) and Nils (1953). Attended the Uppsala Gymnasium; Studied physics, mathematics and chemistry at the University of Uppsala from 1936 until 1942. Graduated in Stockholm 1944. Docent in physics that year. Research associate at the Nobel Institute for Physics 1942 - 1951. Professor of physics at the Royal Institute of Technology in Stockholm from 1951 to 1954. Professor and head of the Physics Department at the University of Uppsala since 1954. Member of the Royal Swedish Academy of Sciences, Royal Swedish Academy of Engineering Sciences, Royal Society of Science, Royal Academy of Arts and Science of Uppsala, Royal Physiographical Society of Lund, Societas Scientiarum Fennica, Norwegian Academy of Science, Royal Norwegian Society of Sciences and Letters, Honorary Member of the American Academy of Arts and Sciences, Membre du Comité International des Poids et Mesures, Paris, President of the International Union of Pure and Applied Physics (IUPAP).

Awards

The Lindblom Prize 1945, Björkén Prize 1955, Celsius Medal 1962, Sixten Heyman Award, University of Gothenburg 1971, Harrison Howe Award, Rochester 1973, Maurice F. Hasler Award, Cleveland 1975, Charles Frederick Chandler Medal, Columbia University, New York 1976, Björkén Prize 1977, Torbern Bergman Medal 1979, Pittsburgh Award of Spectroscopy 1982. Doctor of Science, honoris causa: University of Durham 1972, University of Basel 1980, University of Liege 1980, Upsala College, New Jersey, 1982.

Research in physics covering atomic and molecular physics, nuclear physics, plasma physics and electron optics. Main research activity in the field of electron spectroscopy, ESCA. Books: Beta- and Gamma-Ray Spectroscopy, 1955; Alpha-, Beta- and Gamma-Ray Spectroscopy, 1965; ESCA - Atomic, Molecular and Solid State Structure Studied by Means of Electron Spectroscopy, 1967; ESCA Applied to Free Molecules, 1969.

Surveys on ESCA:

Electron Spectroscopy for Chemical Analysis, Phil. Trans. Roy. Soc. London A, 33-57, 1970.

Electron Spectroscopy, Encyclopedia of Science and Technology, McGraw-Hill, 1971.

Perspectives and Problems in Electron Spectroscopy, Proc. Asilomar Conference 1971, Ed. D. A. Shirley, North Holland, 1972.

Electron Spectroscopy-A New Way of Looking into Matter, Endeavor 32, 1973.

Electron Spectroscopy for Chemical Analysis, Proc. of Conf. on Atomic Physics 3, Boulder, 1972, Ed. S. J. Smith and G. K. Walters, Plenum, 1973.

Electron Spectroscopy for Chemical Analysis (together with C. J. Allan), MTP Int. Rev. of Science, Vol. 12, Analytical Chemistry, Part 1, Butterworths, 1973.

Electron Spectroscopy - An Outlook, Proc. Namur Conference 1974, Elsevier 1974.

Electron Spectroscopy and Molecular Structure, Pure and Appl. Chem. 48, Pergamon, 1976.

Electron Spectroscopy for Solids, Surfaces, Liquids and Free Molecules, in Molecular Spectroscopy, Ch. 15, Heyden 1977.

ELECTRON SPECTROSCOPY FOR ATOMS, MOLECULES AND CONDENSED MATTER

Nobel lecture, 8 December, 1981

by

KAI SIEGBAHN

Institute of Physics, University of Uppsala, Sweden

In my thesis [1], which was presented in 1944, I described some work which I had done to study β decay and internal conversion in radioactive decay by means of two different principles. One of these was based on the semi-circular focusing of electrons in a homogeneous magnetic field, while the other used a big magnetic lens. The first principle could give good resolution but low intensity, and the other just the reverse. I was then looking for a possibility of combining the two good properties into one instrument. The idea was to shape the previously homogeneous magnetic field in such a way that focusing should occur in two directions, instead of only one as in the semi-circular case. It was known that in betatrons the electrons performed oscillatory motions both in the radial and in the axial directions. By putting the angles of period equal for the two oscillations Nils Svartholm and I [2, 3] found a simple condition for the magnetic field form required to give a real electron optical image i.e. we established the two-directional or double focusing principle. It turned out that the field should decrease radially as $\frac{1}{\sqrt{R}}$ and that double focusing should occur after $\pi \cdot \sqrt{2} \sim 255°$. A simple mushroom magnet was designed, the circular pole tips of which were machined and measured to fit the focusing condition. ThB was deposited on a wire net and put into position in the pole gap. A photographic plate was located at the appropriate angle and the magnet current set to focus the strong F line of ThB on the plate. Already the first experiment gave a most satisfactory result. Both the horizontal and the vertical meshes of the wire net were sharply imaged on the plate. A more detailed theory for the new focusing principle was worked out and a large instrument with R = 50 cm was planned and constructed [4]. Due to the radially decreasing field form an additional factor of two was gained in the dispersion, compared to the homogeneous field form. Since all electrons, for reasonably small solid angles, were returning to the symmetry plane of the field at the point of focus no loss in intensity was experienced by increasing the radius of curvature of the instrument. Very large dispersion instruments with good intensity and much improved resolving power could therefore be designed to record β spectra and internal conversion spectra from radioactive sources. The magnetic double focusing was convenient for the fairly high energy electrons (50 keV-2 MeV)

normally occurring in radioactive decay and the field form could easily be achieved by means of shaping the poles of an iron magnet. In my laboratory and in many other nuclear physics laboratories double focusing spectrometers frequently became used for high resolution work [5]. This type of focusing was also used subsequently by R. Hofstadter [6] in his well-known work on high energy electron scattering from nuclei and nucleons.

During the late forties, the fifties and the early sixties I was much involved in nuclear spectroscopy. This was a particularly interesting and rewarding time in nuclear physics since the nuclear shell model, complemented with the collective properties, was then developed, which to a large extent was founded on experimental material brought together from nuclear decay studies. Nuclear disintegration schemes were thoroughly investigated, and the spins and parities of the various levels were determined, as well as the intensities and multipole characters of the transitions. During this period the discovery of the non-conservation of parity added to the general interest of the field. Also the form of the interaction in β decay, appearing originally in Fermi's theory, was extensively investigated. A large part of my own and my students' research was therefore concerned with nuclear spectroscopy of radioactive decay [7-23]. In 1955 I edited a volume [24] on "Beta and Gamma Ray Spectroscopy". In 1965 I concluded my own career as a nuclear spectroscopist by publishing [25] "Alpha, Beta and Gamma Ray Spectroscopy". In this extensive survey of nuclear spectroscopy I had been able to collect the prodigious number of 77 coauthors, all being prominent authorities and in many cases the pioneers in the various fields. Although my own scientific activity at that time had almost entirely become directed towards the new field which is the subject of the present article I have still kept my old interest in nuclear physics much alive as the editor for the journal Nuclear Instruments and Methods in Physics Research (NIM) ever since its start in 1957.

Let me now return to the situation around 1950. At that time my coworkers and I had already for some time been exploring the high resolution field by means of our large dispersion double focusing instrument and other methods, such as the high transmission magnetic lens spectrometer and coincidence techniques. Often I found, however, that my experimental work had to stop and wait for radioactive samples, the reason being a capricious cyclotron. It then came to my mind that I should try to simulate the radioactive radiation by a substitute which I could master better than the cyclotron. I had found that a very convenient way to accurately investigate gamma radiation from radioactive sources was to cover them with a γ-ray electron converter, i.e. a thin lead foil which produced photoelectrons to be recorded in the spectrometer. I now got the idea that I should instead use an X-ray tube to expel photoelectrons from ordinary materials, in order to measure their binding energies to the highest possible accuracy. In my nuclear physics work such binding energies had to be added to the energy values of the internal conversion lines from the radioactive sources in order to get the energies of the nuclear transitions. I studied what had been done along these directions before [26, 27], and I finally got a vague feeling that I could possibly make an interesting and perhaps big

step forward in this field if I applied the experience I had from nuclear spectroscopy using the above mentioned external photoeffect and my high resolution instruments. The previous investigations had confirmed that the atomic electrons were grouped in shells, and by measuring on the photographic plates the high energy sides of the extended veils from the various electron distributions approximate values of the binding energies could be deduced. On the other hand, since the observed electron distributions had no line structure and consequently did not correspond to atomic properties, the attained precision and the actual information was far inferior to what could be obtained by X-ray emission and absorption spectroscopy. I realized that electron spectroscopy for atoms and solids could never become competitive with X-ray emission or absorption spectroscopy unless I was able to achieve such a high resolution that really well-defined electron lines were obtained with linewidths equal to or close to the inherent atomic levels themselves.

I thought of these problems considerably and started to make plans for a new equipment which should fulfill he highest demands on resolution at the low electron energies I had to be concerned with, ten to a hundred times smaller than in radioactive work. I recall I sat down for some days early in 1950 to try to make a thorough calculation about the expected intensities. I designed [28, 29, 30] an ironfree double focusing spectrometer with R = 30 cm, in which I should be able to measure the current with a precision of better than 1 part in 10^4. The spectrometer was surrounded by a big, three-component Helmholtz coil system to eliminate, to better than 1 part in 10^3, the earth's magnetic field over the entire region of the spectrometer. If I had an X-ray tube with a Ka radiation in the region of 5 keV, this would enable me to measure expelled photoelectrons with a precision of a fraction of an electron volt. This I thought was about sufficient in atomic physics. I also hoped to observe phenomena of chemical interest provided I could realize the resolution I aimed at, but at that time my ideas in this latter respect were of course very vague, centering around atomic level shifts in alloys, etc. When I calculated the expected intensities of the photoelectron lines, I started from the very beginning, i.e. with a certain number of mA·s in the X-ray tube, I then calculated by means of existing knowledge the number of Ka X-ray photons, next I put in all solid angles both in the X-ray tube and the electron spectrometer and made some assumptions about the effective photoelectron cross sections to expel electrons from the outermost layers in a solid surface. Those electrons could not be expected to suffer much energy loss and were the interesting ones upon which I should base my spectroscopy. In retrospect, this last stage in my considerations was of course of some interest, in view of the later development of electron spectroscopy into a *surface* spectroscopy. I guessed that what is now called the "escape depth" of the electrons should be less than a light wavelength and more than a few atomic layers and so I used 100 Å in my calculations. This was not too bad a guess in consideration of later studies, indicating a lower figure for metals and 100 Å for organic multilayers. I finally arrived at an estimated counting rate on a photoline in my apparatus of several thousands of electrons per minute as recorded in the Geiger-Müller (G-M) counter placed at the focal plane in the

double focusing spectrometer. Afterwards I was satisfied to find that this calculation turned out to correspond fairly well to reality. This step, however, in fact took several years to make.

The equipment which I had to build and test was at that time very complicated. The resolution ultimately achieved turned out to be high enough to enable recording even of the inherent widths of internal conversion lines [24]. This was done in 1954 and in 1956 I published [31] together with my collaborator, Kay Edvarson, an account of this phase of the work under the title "β-ray Spectroscopy in the Precision Range of $1:10^5$". In the next phase I had, however, to overcome many difficulties in handling the low energy electrons excited by X-rays and to record them by the G-M counter. This had an extremely thin window through which gas diffused continuously and so was compensated for by an automatic gas inlet arrangement. I did not realize, to start with, the precautions I had to take when dealing with surfaces of solids in order to record resolved line structures.

After some further testing of the equipment, concerning the influence of the finite nuclear size on the conversion lines in some nuclei [32, 33], my two new coworkers, Carl Nordling and Evelyn Sokolowski, and I finally made the transition to atomic physics and were able to record our first photoelectron spectrum [34, 35] with extremely sharp lines and with the expected intensities. These electron lines definitely had all the qualities which I had set as my first goal. They were symmetric, well defined and had linewidths which could be deduced from the linewidth of the X-ray line used and the width of the atomic level of the element under study, plus of course a small additional broadening due to the resolution of the instrument. Fig. 1 shows an early recording of MgO. The exact position of the peak of the electron lines could be measured

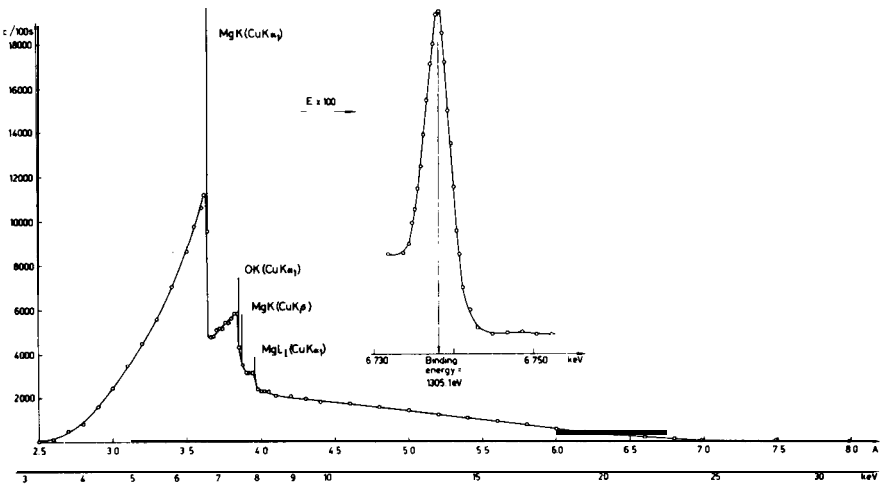


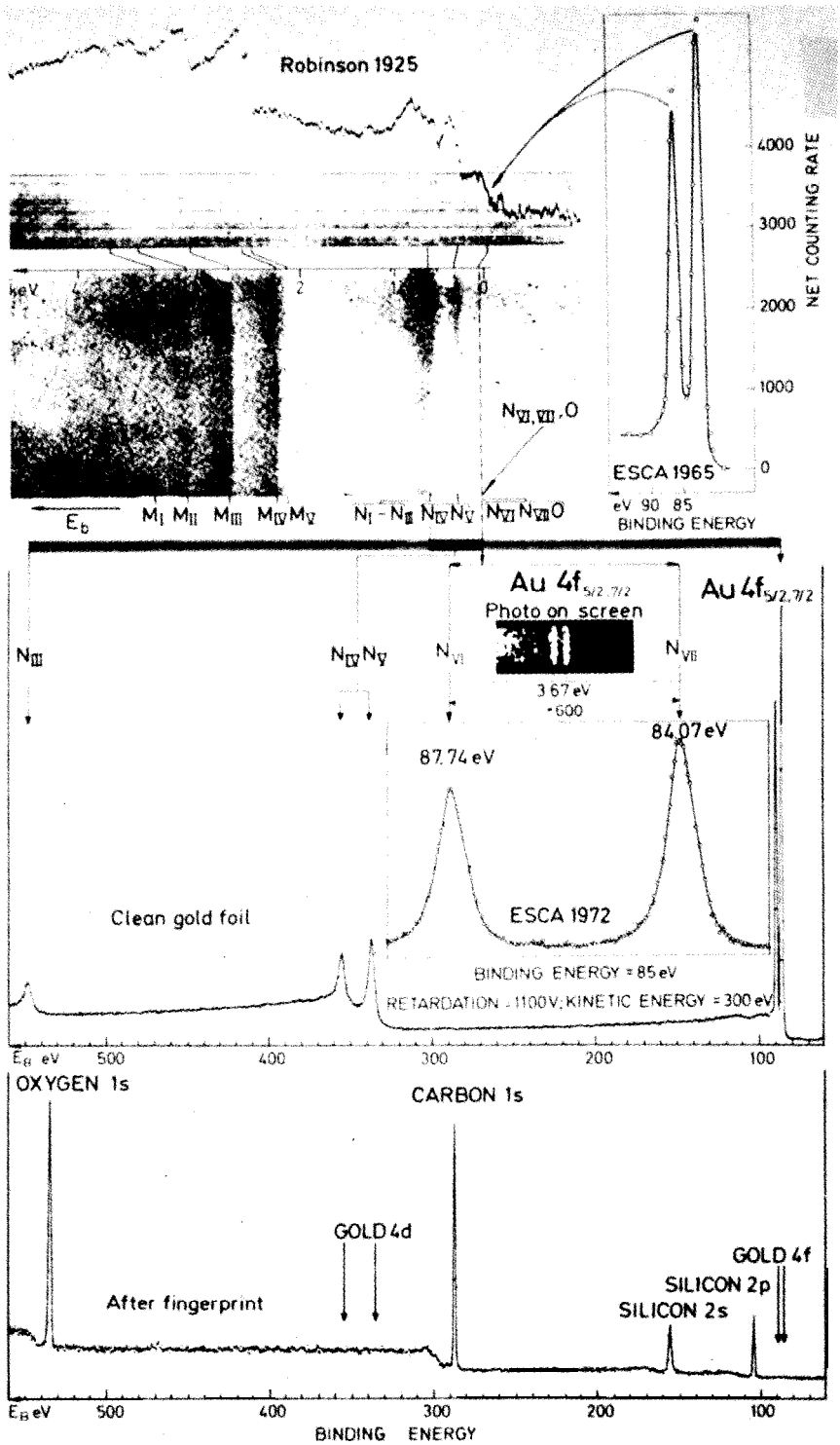
Fig.1. Electron spectrum obtained from magnesium oxide with copper X-radiation. Edges are found at energies corresponding to atomic levels of magnesium and oxygen. A very sharp electron line can be resolved from each edge. Such an electron line is shown in the insert figure with the energy scale expanded by a factor of one hundred to bring out the finite width of the line.

with considerable accuracy. Electron spectroscopy for atoms could be developed further with confidence.

Fig. 2 illustrates the steps which we took from the earlier recording of the photoelectrons expelled from a gold foil by Robinson [36] in 1925 to the introduction of the electron line spectroscopy in 1957. The dotted line inserted in Robinson's spectrum should correspond to the place where our spin doublet $N_{vI}N_{vII}$ to the right in the figure is situated. The distance between the two well resolved lines in our spectrum would correspond to about 0.1 mm in the scale of Robinson's spectrum. Below this spectrum the black portion has been enlarged (the gray scale) to show the corresponding part in our spectrum. This spectrum was taken at a later stage of our development. Within this enlarged spectrum a further enlargement of the $N_{v\ I}N_{vII}$ doublet is inserted. The spin-orbit doublet has now a distance between the lines which corresponds to a magnification of 600 times the scale in Robinson's photographic recording.

A comparison between the middle and the lower spectrum in the figure further demonstrates the extreme surface sensitivity of electron spectroscopy. The difference between the two spectra is caused by a slight touch of a finger. At the beginning this sensitivity caused us much trouble, but later on, when electron spectroscopy was applied as a surface spectroscopy, it turned out to be one of its most important assets.

In 1957 we published some papers [34, 35, 37] describing our first results, which really did indicate great potential for the future. We also obtained our first evidences of chemical shifts [37-42] for a metal and its oxides and for Auger electron lines. I thought, however, that we should first improve our techniques and explore purely atomic problems until we had achieved a greater knowledge to enable us to progress to molecular problems. We therefore systematically measured atomic binding energies for a great number of elements with much improved accuracy compared to previous methods, in particular X-ray absorption spectroscopy [40, 41, 43-61]. We were surprised to find how inaccurate previously accepted electron binding energies for various shells and elements could be. We made so-called "modified" Moseley diagrams. We were bothered by uncertainties due to the chemical state and therefore tried to use only metals or at least similar compounds of the elements in our systematic studies. We also devoted much effort to the investigation of the Auger electron lines [62-69] which appeared in our spectra with the same improved resolution as our photoelectron lines. As one of the results of such studies we were able to observe for a group of elements around Z = 40 all the nine lines expected in the intermediate coupling theory as compared to the observed six lines in pure j-j coupling [63]. In general, in the spectroscopy we developed photoelectron and Auger electron lines were found side by side. Later on, therefore, we avoided any notation for this spectroscopy which could give the false impression that only one of the two types of electron lines were present. Auger electron lines can *in addition* to the X-ray mode of excitation also be produced by electrons. Much of the above basic work on atomic energy levels is described in theses by E. Sokolowski [70], C. Nordling [71], P. Bergvall [72], 0. Hörnfeldt [73], S. Hagström [74] and A. Fahlman [75].


Robinson 1925
4000
3000
2000
1000
0
NET COUNTING RATE
eV
2
1
0
$N_{VI,VII}$,O
ESCA 1965
eV 90 85
BINDING ENERGY
E_b
M_I M_{II} M_{III} $M_{IV}M_V$ N_I - N_{II} $N_{IV}N_V$ $N_{VI}N_{VII}O$
Au $4f_{5/2,7/2}$
Photo on screen
Au $4f_{5/2,7/2}$
N_{III}
$N_{IV}N_V$
N_{VI}
N_{VII}
367 eV
·600
84.07 eV
87.74 eV
Clean gold foil
ESCA 1972
BINDING ENERGY = 85 eV
RETARDATION -1100V; KINETIC ENERGY = 300 eV
E_B eV 500 400 300 200 100
OXYGEN 1s
CARBON 1s
GOLD 4d
GOLD 4f
SILICON 2p
SILICON 2s
After fingerprint
E_B eV 500 400 300 200 100
BINDING ENERGY

After some years' work in electron spectroscopy on problems in atomic physics the next step came to the fore, namely to make systematic studies of the chemical binding. This step was taken together with my coworkers Stig Hag-Strom and Carl Nordling when $Na_2S_2O_3$ was found to give two well resolved K photoelectron lines from the sulphur [76]. This showed that two differently bonded sulphur atoms could be separated in the molecule, which according to classical chemistry were in the -2 and the +6 valence states, respectively. This was a more clear-cut case than the copper-copperoxide case we had studied before, since the reference level for the two sulphur atoms could be traced to the same molecule. The systematic investigation of chemical binding by means of electron spectroscopy is described in theses by S.-E. Karlsson [77], R. Nord-berg [78], K. Hamrin [79], J. Hedman [80], G. Johansson [81], U. Gelius [82] and B. Lindberg [83].

Fig. 3 shows the chemically shifted Cls spectrum of ethyl trifluoroacetate [84, 85]. Fig. 4 [86] shows how the chemical shift effect can be used to identify groups linked together in branched chains in polymers [87-90]. The intensi-ties of the lines are correlated to the different branchings in the two viton polymers.

In the interpretation of the electron spectra the first step is to consider the electron structure as 'frozen' under the photoelectron emission process. In this approximation the measured electron binding energies can be identified with the Hartree-Fock energy eigenvalues of the orbitals. One then disregards the fact that the remaining electronic structure, after electron emission, is relaxing to a new hole state. This relaxation energy is by no means negligible and an accurate calculation of the relevant binding energies has to include both the ground state and the hole state energies as the difference between them. The inclusion of relativistic effects in this treatment is essential for inner core ionization and heavier elements (I. Lindgren [91]). More recently, methods have been devised to describe the photoelectron emission by means of a transition operator which properly accounts for the relaxation process [92-94]. Various conceptual models complement the computational procedures on an ab initio level.

For chemical shifts in free molecules, it is usually sufficiently accurate to consider only the ground state properties [95-106]. This is due to the circum-

Fig.2. Electron spectra of gold.
Upper left: Spectrum recorded by Robinson in 1925 /36/. (Reproduced by due permission from Taylor & Francis Ltd, London.)
Upper right: ESCA spectrum recorded in Uppsala before 1965 by non-monochromatized $MgK\alpha$ excitation, The N_{vi}, N_{vii} levels are seen as two completely resolved lines in this spectrum whereas the $N_{vi}N_{vii}$ and O levels appear together as a hump in the photometric recording by Robinson and are only barely visible on the photographic plate.
Middle: ESCA spectrum recorded in Uppsala 1972 by monochromatized $AlK\alpha$ excitation. The magnification of this spectrum is 600 times that of Robinson.
Lower Part: ESCA spectrum of a gold foil with a fingerprint on the surface. The electron lines are entirely due to the lingerprint whereas the gold lines are missing.

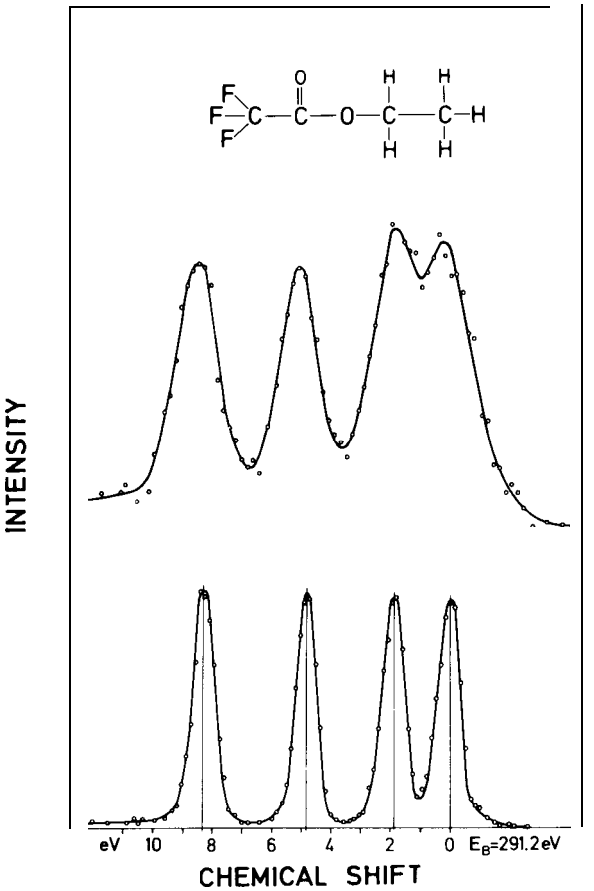


Fig.3. The ESCA shifts of the Cls in ethyl trifluoroacetate. Upper spectrum without and lower with X-ray monochromatization 184, 85).

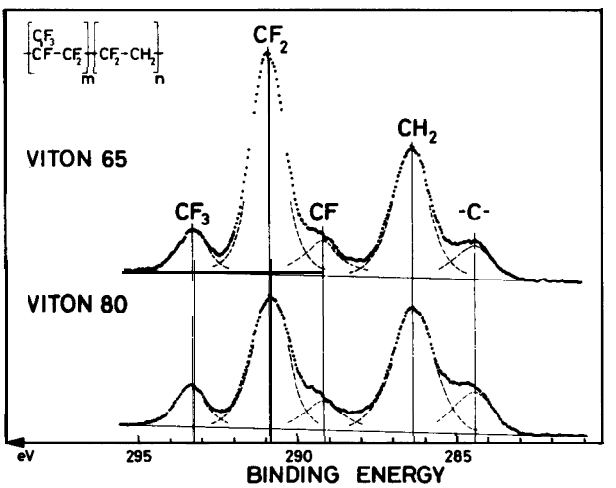


Fig.4. ESCA spectra of Viton 65 and Viton 80 polymers.

stance that relaxation energies for a series of similar electronic systems vary only marginally. This can be described by the division of the relaxation energy into two contributions, one connected to the atomic contraction at ionization, the other to the 'flow' of charge from the rest of the molecule [107, 108]. The atomic part, which is very nearly constant for one specific element, is the dominating contribution to the relaxation energy. The 'flow' part varies generally marginally for free molecules of similar structure, leading to constant relaxation energies. There are cases, however, where the 'flow' part can significantly change from one situation to another. One example is when a molecule is adsorbed on a metal surface. In such a situation the flow of conduction electrons from the metal substrate will contribute to the relaxation of the core hole. This can increase the relaxation energy by several eV [109, 110]. Other cases are pure metals and alloys where the conduction electrons are responsible for the screening of the hole. [111 - 116]. These are treated in theses work by N. Mårtensson [117] and R. Nyholm [118].

In view of the interesting applications which the chemical shift effect offered for chemistry and the fact that at that time we had found that electron spectroscopy was applicable for the analysis of all elements in the Periodic System, we coined the acronym ESCA, Electron Spectroscopy for Chemical Analysis. If one is particularly interested in conduction bands for metals or alloys (Fig. 5

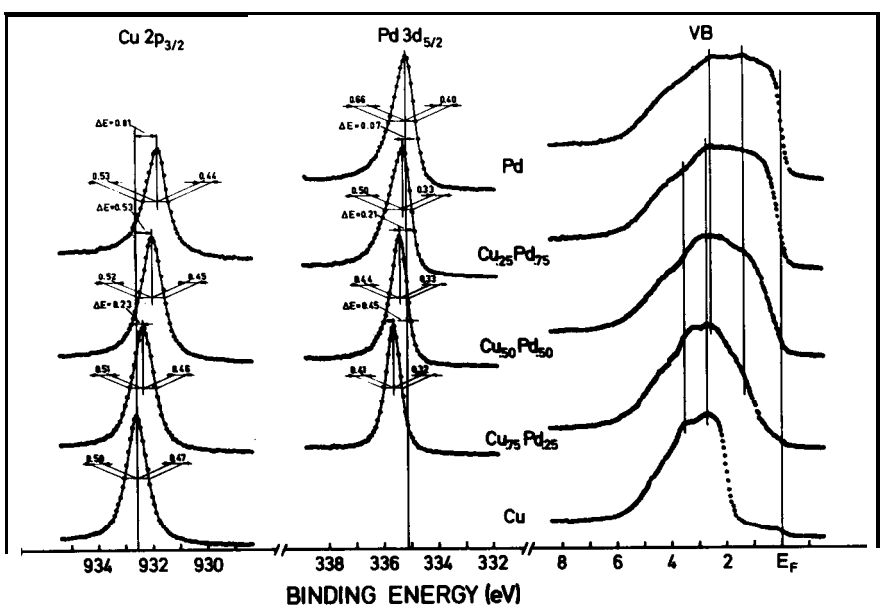


Fig. 5. Core and valence electron spectra (excited by monochromatized AlKα-radiation) of some Cu_xPd_{1-x} alloys, including the pure constituents. The binding energies undergo positive chemical shifts with increasing Cu-content. The asymmetries of the lines are due to creation of soft electron-hole pairs at the Fermi edge upon core ionization. The magnitude of the asymmetry is thus related to the (local) density of states at the Fermi level. The Pd lines are seen to become more symmetric as the Cu content increases (Pd local density of states decreases).

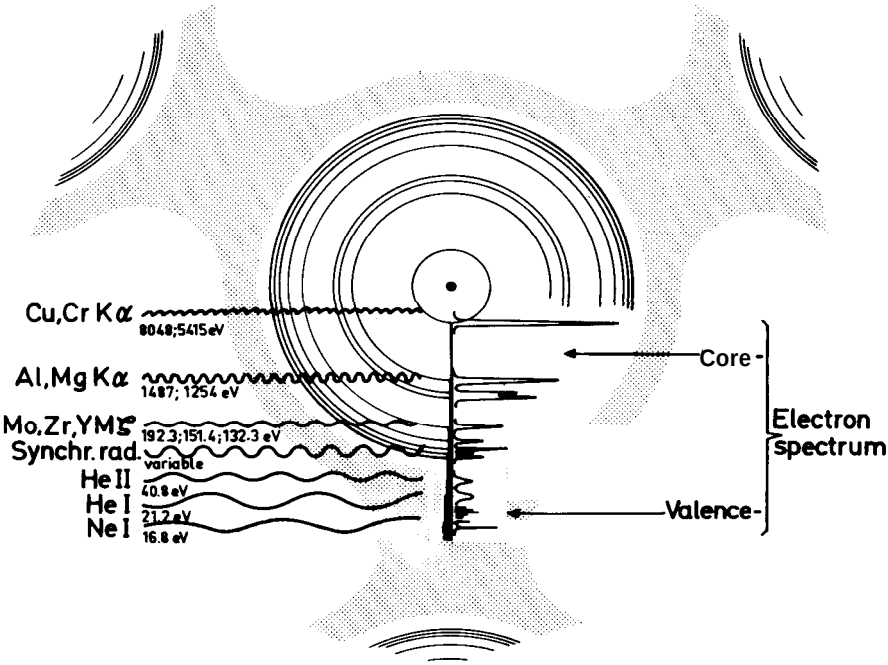


Fig.6. Regions of binding energies accessible with different photon sources
Solid circles: localized, atom-like orbitals.
Shaded area: more or less delocalized, molecular orbitals.

[119]) or valence electron structures of solid material in general, or free molecules, more detailed notations can be preferred. One useful distinction is between core and valence electron spectra (Fig. 6). Obviously, a further ground of classification is due to the different origin of the photoelectron and Auger electron lines, which both always occur in ESCA, as mentioned before. The corresponding chemical shift effect for the Auger electron lines we established soon afterwards [69] in the case of $Na_2S_2O_3$. Further studies [111, 112, 120-1331 have shown that the two shift effects are complementary. The combination of the two shifts provides insight into the mechanism of relaxation in the photoionization process. Auger electron spectra are given in Fig. 7 for a clean Mg metal, and when it is partly and finally fully oxidized to MgO [86].

Apart from the ordinary core electron lines and the Auger electron lines from the various shells, all characteristic of each element, the electron spectra also contain additional features. Satellites situated close to (-10 eV) the main core lines at the low energy sides are often observed with intensities around 10% of the latter. Fig. 8 [114] shows the electron spectrum of gaseous Hg. Inserted are the satellites to the $N_{VI}N_{VII}$ lines. Strong satellites were first observed [68] in our spectra in the KLL Auger electron spectrum of potassium in some compounds. Satellites have been found to occur frequently for core lines and occasionally they can even have intensities comparable to the main lines, e.g. paranitroaniline [134-136], transition metal compounds [137-146] and var-

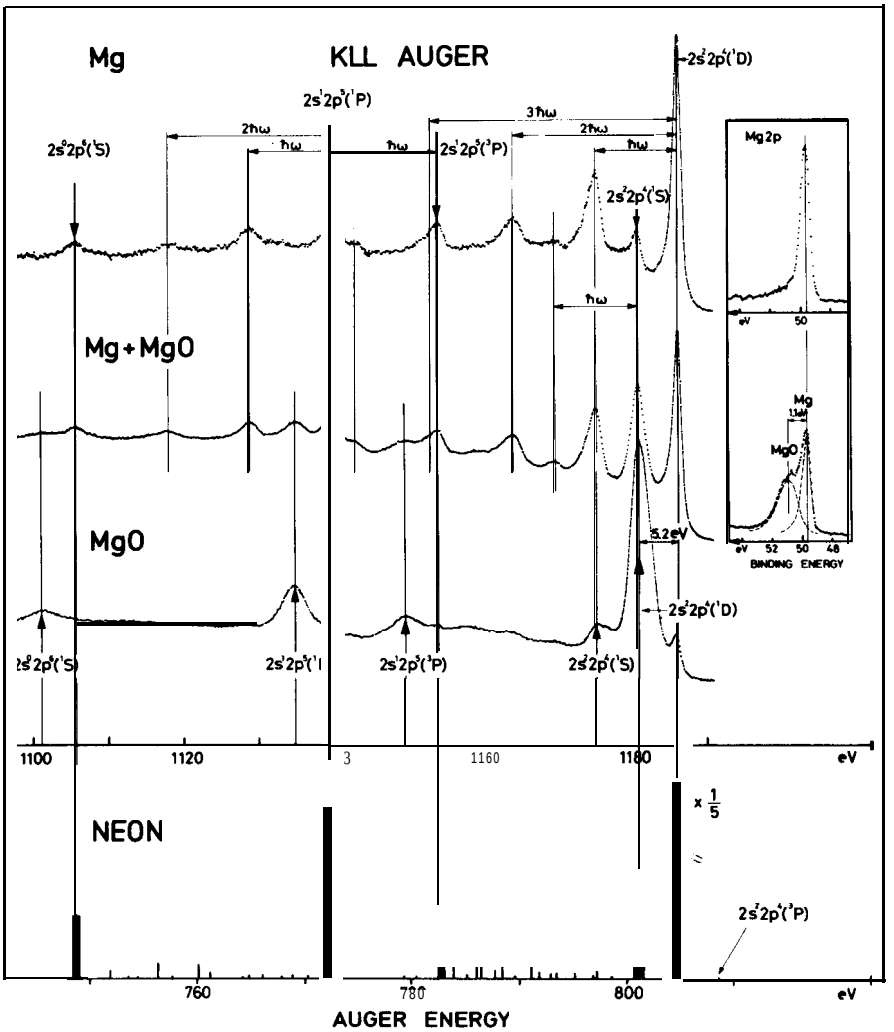


Fig. 7. MgKLL Auger electron spectra at different stages ofoxidation as obtained in ESCA. Upper spectrum is from a clean metal surface, lower spectrum from the oxidized metal (with only a trace of metal) and the middle spectrum is from an intermediate oxidation. Volume plasmon lines are observed. For comparison, the positions of the NeKLL Auger electron lines are given below, as recorded in the ESCA instrument by means of elcctron beam excitation.

ious adsorbed molecules on surfaces [148, 149]. Since these electrons can be visualized as being emitted from excited states, the satellites were given the name “shake-up” lines.

Molecules like O_2 or NO contain unpaired electrons and are therefore paramagnetic. Large classes of solid materials have similar properties. In such cases core electron spectra show typical features called spin-, multiplet-, or exchange splitting. We first observed this phenomenon [150] in oxygen when

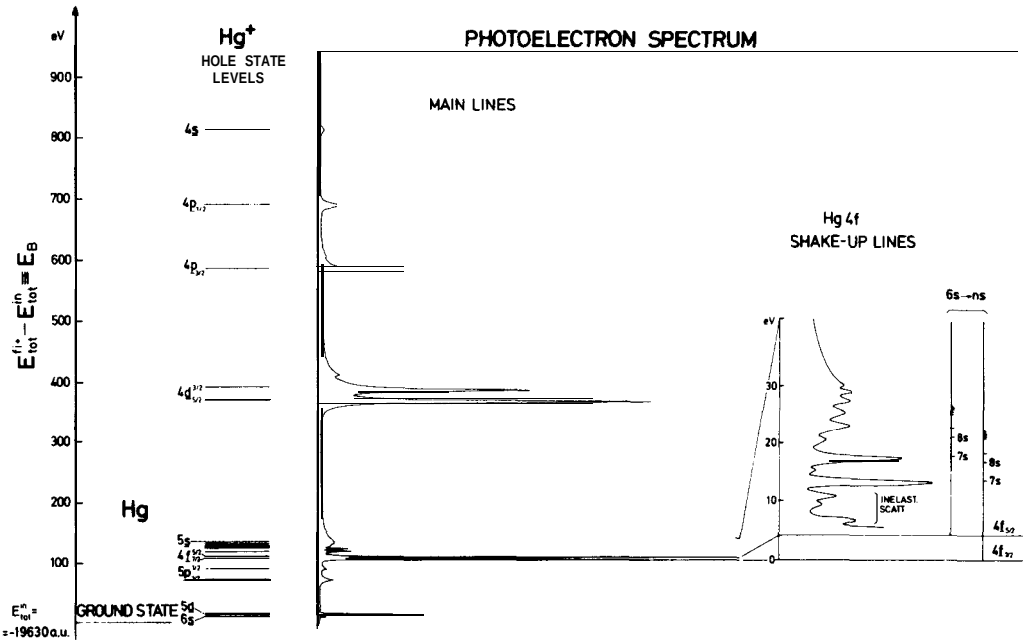


Fig. 8. Relation between the hole-state level system and observed photoelectron spectrum for the mercury atom. The figure illustrates that the main lint-s are connected with states of the ion where (in a one-electron picture) an atomic orbital has been removed from the neutral ground state. The energy region close to the 4f-lines has been expanded (far right) to show that the additional satellite lines observed (shake-up lines) are due to excitations (6s→ns) above the 4f hole ground states. (Note that the intensities of the 4f-lines have been truncated to fall into the scale of the figure.) (From refs. 114 and 260.)

air was introduced into the gas cell in our ESCA instrument (Fig. 9). The 1s line of O_2 is split in the intensity ratio of 2:1. This spin splitting is due to the exchange interaction between the remaining 1s electron and the two unpaired electrons in the $\pi_g 2p$ orbital, which are responsible for the paramagnetism of this gas. The resulting spin can be either 1/2 or 3/2. The corresponding electrostatic exchange energies can be calculated and correspond well with the measured splitting of 1.11 eV [151]. Apart from oxygen and nitrogen, argon and CO_2 can also be seen in air in spite of the low abundances of these gases. A statistical treatment of the data even exhibits the presence of neon (0.001%).

Other particular features in the spectra occur in the valence electron region, i.e. at binding energies extending from *zero* binding energy to say 50 eV. Our first study of this entire region concerned ionic crystals like the alkali halides [152].

In a later study [153] (1970) of a single crystal of NaCl we discovered the phenomenon of *ESCA diffraction. We* investigated the angular distribution of emitted Auger electrons from the NaKLL (^{1}Dp) transition and the photoelec-

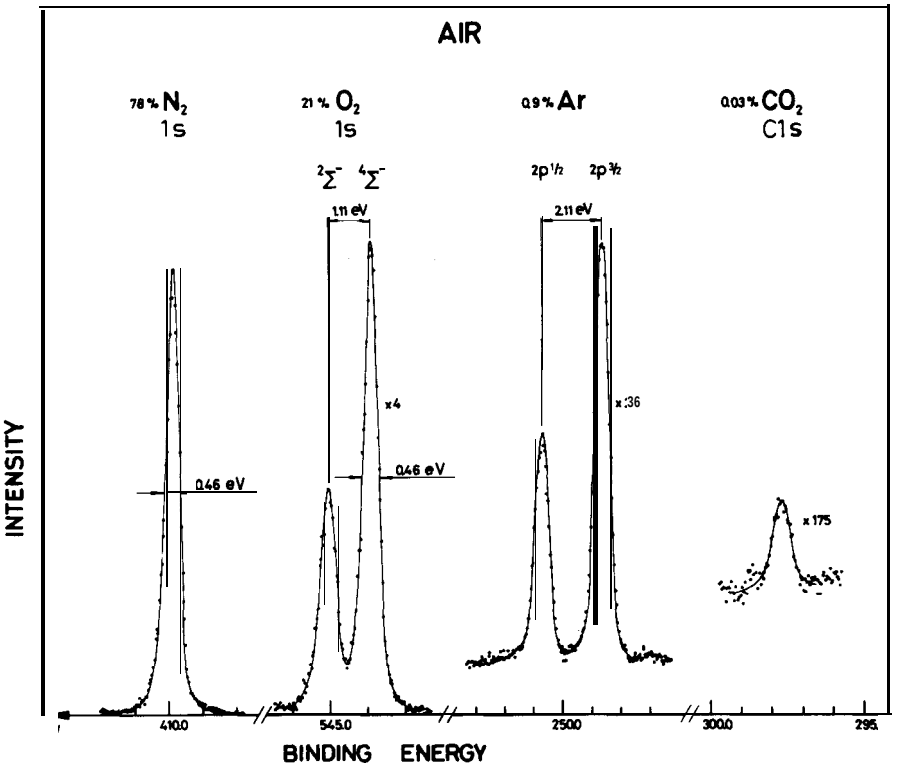


Fig. 9. Electron spectrum of air. The O1s is split into two components due to 'spin' or 'multiplet' splitting, Excitation was performed by means of monochromatized AlKα (Ahv = 0.2eV) radiation.

trons from the Nals, Na2s, C13$p_{3/2}$ and the C13p levels, the latter being the outermost valence orbital of the crystal. For excitation both AlKα and MgKα were used. The crystal could from outside be set at different angular positions relative to the emission direction of the electrons, which in turn was defined by the slit system of the ESCA instrument. For comparison, the angular distributions from polycrystalline samples were also recorded. In all cases typical diffraction patterns were found. In the control experiments on the polycrystalline samples there were no such patterns. Fig.10 shows two of the diffraction patterns recorded. Subsequent measurements [154-158] on other single crystals have shown agreement with the above investigation.

ESCA diffraction has more recently been applied to surface studies giving interesting information on the geometry of adsorbed molecules [159-161] on single crystals. This field is under development and should have a promising future in surface science.

In X-ray diffraction there is an incoming photon wave and an outgoing diffracted photon wave. In electron diffraction there is an incoming electron wave and an outgoing diffracted electron wave. In ESCA diffraction there is an

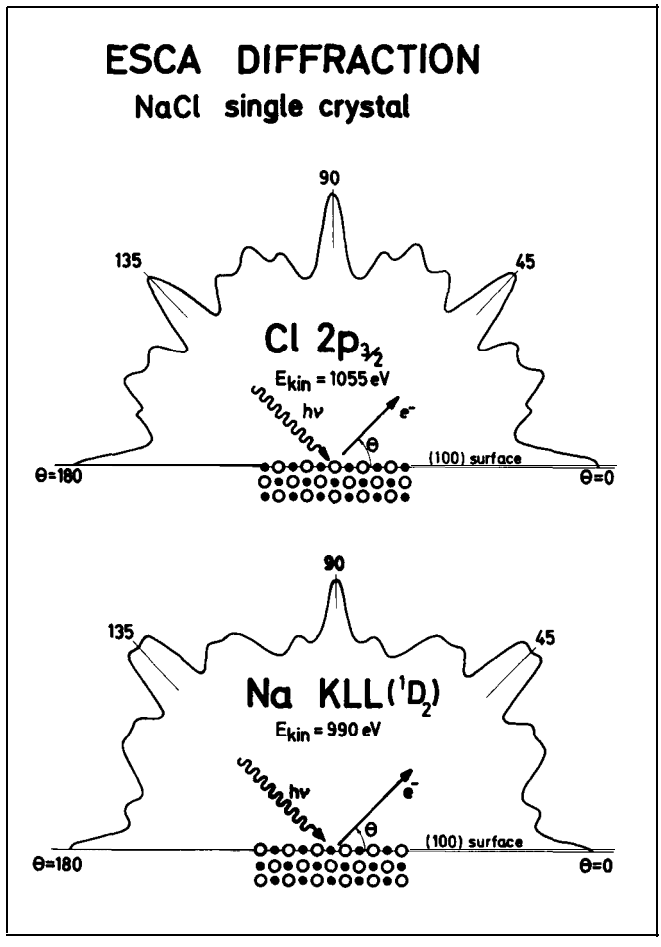


Fig. IO. Angular distributions of $Cl2p_{3/2}$ photoelectrons (MgKα) and NaKLL ($^1D_2$) Auger electrons from a NaCl single crystal.

incoming photon wave and an outgoing diffracted electron wave with different energies. These are three distinctly different physical phenomena which obviously require both different experimental equipment to observe and different theoretical treatments to evaluate. With more suitably built instruments for this purpose and with the addition of stronger X-ray sources and synchrotron radiation [161-164] the development can proceed further.

In order to study gases and vapours from liquids, we first introduced a freezing technique [165] to condense the gases onto the specimen plate. In this way we obtained the valence or molecular orbital spectrum of solidified benzene [166]. Soon afterwards we found that we could study the gaseous phase just as well by introducing differential pumping in the instrument. Acetone was our first study with this technique for gases, i.e. for free molecules, revealing two well separated Cls core lines, one for the ketocarbon and one for the methyl carbon in the intensity ratio of 1:2 [167].

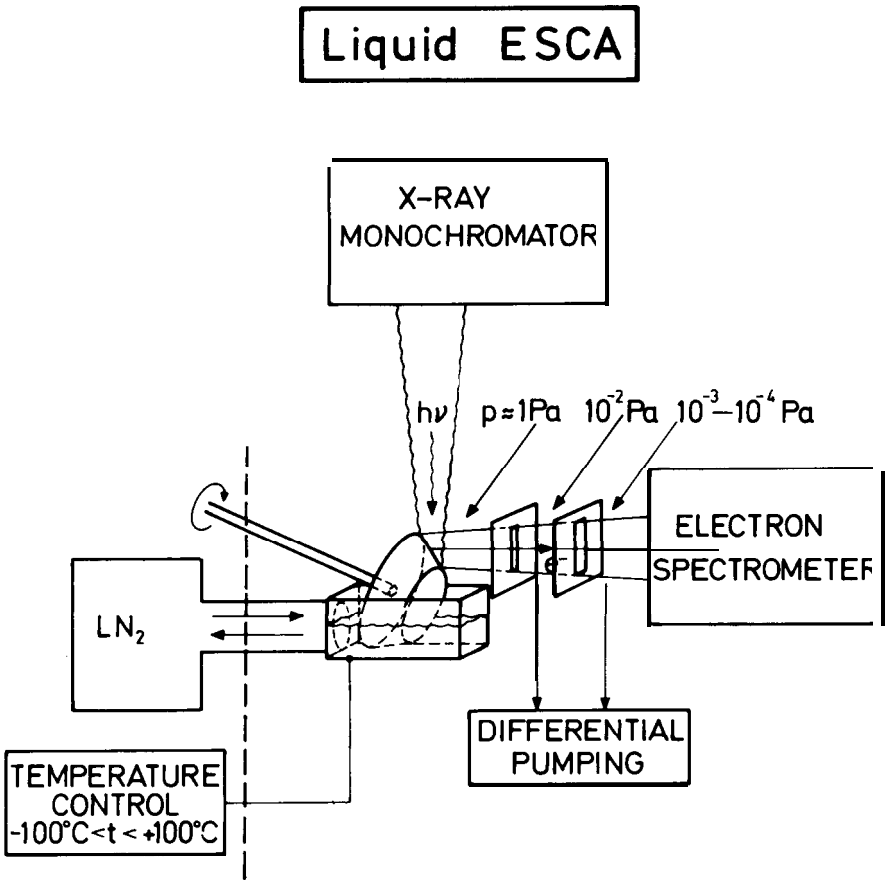


Fig. 11. Schematic diagram of the liquid-sample arrangement

Since solids, surfaces, gases and vapours from liquids were all found to be suitable samples in electron spectroscopy the question arose whether also liquids could be studied. This turned out to be quite possible and several satisfactory methods have been developed in our laboratory [168-173]. The early methods and applications are described in theses by H. Siegbahn [174], L. Asplund [175] and P. Kelfve [176]. Recently a new, more convenient arrangement (H. Siegbahn) has been developed which is shown in Fig. 11 [173]. A small trundle is rotating in the sample cell in which the liquid is introduced. A slit transmits the exciting radiation, e.g. X-radiation and the expelled electrons from the continuously wetted trundle can leave the house through a slit where differential pumping reduces the gas pressure. Cooling of the sample has been introduced which has enabled a vast increase of the number of liquids that can be studied. Fig. 12 shows part of a recent [173] spectrum of ethanol as a solvent in which iodine and sodium iodide are dissolved. One observes here a well resolved spin-orbit doublet of iodine $3d_{3/2}$

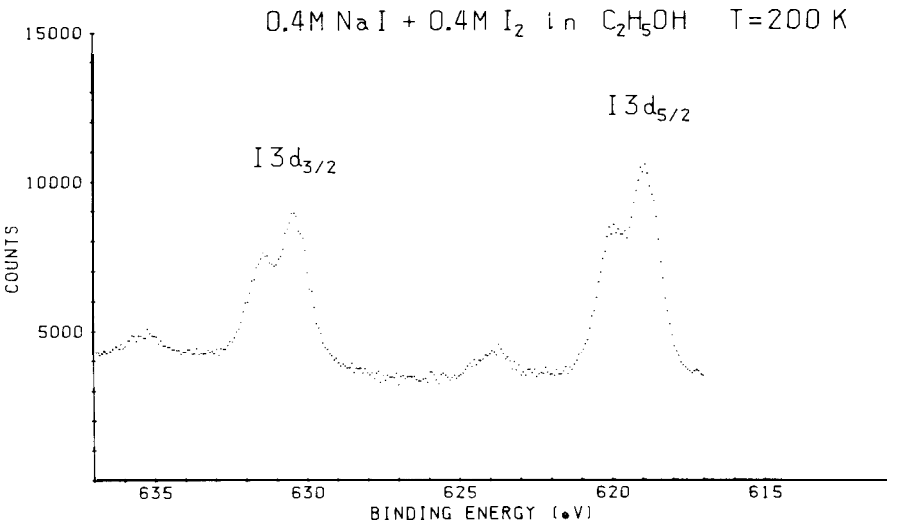


Fig. 12. I3d spectrum from a solution of NaI (0.4M)+I_2(0.4M)in ethanol obtained at T = 200K. The doublet for each spin-orbit component is due to ionization of the central atom (lower peak) and outer atoms (higher peak) of the I_3^- ion. The extra peaks at the high-binding-energy sides of each spin-orbit component are interpreted as shake-up structures.

and $3d_{5/2}$. Each of these electron lines is chemically split in the ratio of ~1:2. The interpretation is that I_3^- has been formed in the solution. The centrally located iodine has the highest binding energy. The correct intensity ratio of 1:2 is obtained when the shake-up satellites are ascribed to the two externally situated iodine atoms, a conclusion which is in agreement with what we have found for similar configurations in other electron spectra. Liquid ethanol itself is shown in Fig. 13. Here one observes the oxygen 1s core line, the chemically split carbon 1s line and the valence electron spectrum. The field of liquids is presently in a state of rapid development.

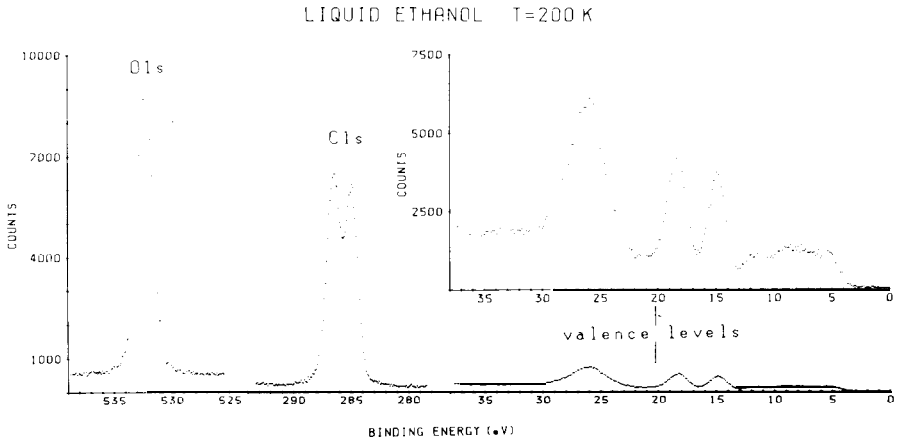


Fig. 13. ESCA spectrum of liquid ethanol obtained at T = 200K.

In the valence region for free molecules it was possible to achieve much improved resolution if UV light, especially the He resonance radiation at 21 eV was used for excitation. Development work in this field was performed by D. W. Turner [177-182] and W. C. Price [183, 184] and their coworkers in England. The conduction bands of metals could be studied by a corresponding technique using ultra-high vacuum (UHV) which was done by W. E. Spicer and coworkers [185-188] in the USA.

In my laboratory a large electrostatic sector focusing instrument was designed in the early part of the sixties for exciting electron spectra in the gaseous phase by VUV radiation and also by electrons. High resolution valence electron spectra were thus obtained and furthermore Auger and autoionization spectra of rare gases and organic molecules could be investigated at a resolution which enabled vibrational structures to appear also in the latter type of spectra. Studies of angular distributions were initiated by using polarized radiation. This was produced by VUV-polarizers which we developed in my laboratory. Much of the above work is described in theses by T. Bergmark [189], L. Karlsson [190], R. Jadrny [191] and L. Mattsson [192]. The Auger electron spectroscopy was further explored in more recent publications [128, 132, 193-196].

The source of excitation was for a time confined to either the soft X-ray region or the UV region with a gap between them from -50 eV (HeII) to 1250 eV ($MgK\alpha$). Fig. 14 shows the valence spectrum of SF_6 excited by HeI, He11 and $AlK\alpha$ [197, 198]. Some intermediate X-ray lines were later on added [199-205], such as $YM\zeta$ at 132 eV but the main step ofdevelopment was the introduction during the seventies of the variable synchrotron radiation [e.g. 206-210] which partly bridged the gap. The previous strong distinction between X and UV excited electron spectra is therefore not so easy to maintain any more unless one is emphasizing the particular technique at hand for exciting the spectra. This is naturally not a trivial point for most researchers, however, and excellent work can be done with one or the other technique alone or in combination.

In 1967 we had gone through most of the basic features of the spectroscopy, designed several new spectrometers (electrostatic double focusing ones included), developed new radiation sources in the soft X-ray and UV region, made theoretical investigations of the process of electronic relaxation at ionization and applied the spectroscopy to a variety of different research fields. We then decided to present the new spectroscopy in a more consistent and complete way than we had done before. At the end of that year our book "ESCA-Atomic, Molecular and Solid State Structure Studied by Means of Electron Spectroscopy" appeared [211]. Two years later we published a second book [212], this time on "ESCA, Applied to Free Molecules". At that time several instrumental firms started to develop commercial instruments. I took part in one of these developments at Hewlett-Packard in Palo Alto during a leave of abscence from my laboratory in 1968. I spent that year at the Lawrence Berkeley Laboratory with which we had had a long cooperation both in nuclear spectroscopy and then in ESCA. The Hewlett-Packard instrument [213] was designed to include

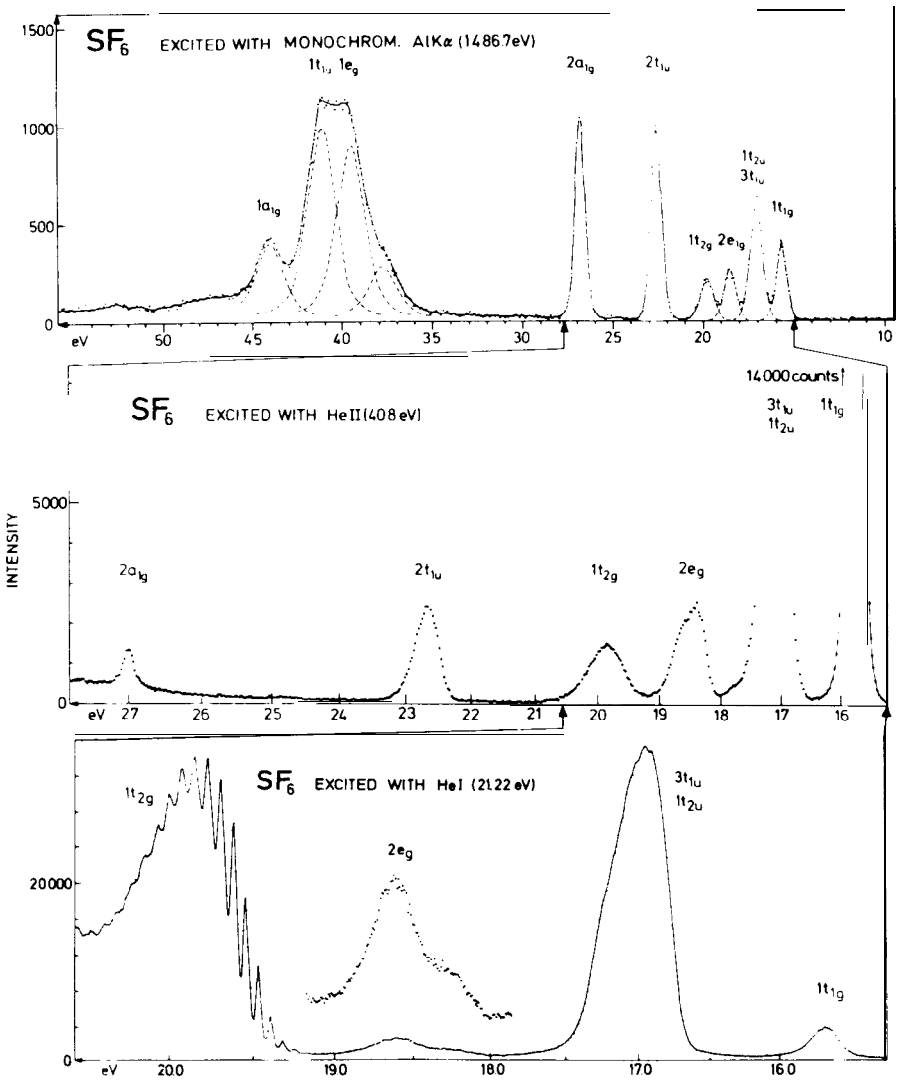


Fig. 14. Valence electron spectra of the SF_6 molecule excited with different photon energies (AlKα-, HeII and HeI-radiations). This figure illustrates the complementary nature of the various excitation sources. The AlKα-excited spectrum enables a recording of the full valence region (including the innermost orbitals), which is not possible with the lower photon energies. The higher resolution in the spectra excited with the He resonance radiations allows the study of finer details of each of the outer electron bands. Note also the strong variations in the relative intensities of the bands as a function of photon energy. This can be used as an aid in assigning the spectrum / 197,198/.

a monochromator for the AlKa radiation consisting of three spherically bent quartz crystals and a retarding electrostatic lens system to match the dispersion of these crystals to the electron spectrometer.

The spherically bent quartz crystal monochromator, having the property of being double focusing, was invented in my laboratory in 1958 [214] in quite another connection, namely in low angle scattering of X-rays against latex and other particles of biological interest [215]. The combination of double focusing both in the X-ray monochromator and in the electron analyzer has turned out to be essential for the further progress in ESCA. Other important technical developments have been the introduction of swiftly rotating water cooled anodes (U. Gelius), multidetector systems by means of electron channel plates and computerization of the instruments (E. Basilier).

In 1972 my coworker Ulrik Gelius and I made a new instrument [85, 216] with all these components included, in particular designed for the studies of gases. With the improved resolution of this instrument new structures could be resolved. One of principal interest was the discovery of the vibrational fine structure of core lines [82, 217]. Fig. 15 shows the line profile of the Cls in CH_4It turns out that this line can be separated into three components caused by the symmetric vibration when photoionization occurs in the 1s level of the central carbon atom. When the photoelectron leaves the methane molecule the latter shrinks about 0.05 A. The minimum of the new potential curve for the ion will consequently be displaced by the corresponding amount and Franck-Condon transitions which take place will then give rise to the observed vibra-

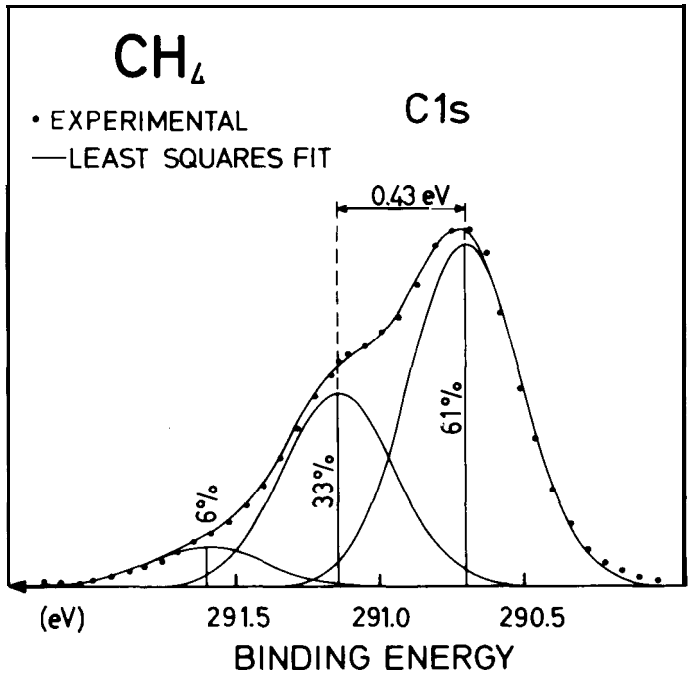


Fig. 15. Vibrational structure of the core electron line (Cls in CH_4. The line structure can bc quantitatively explained as a consequence of the shrinkage of the equilibrium distances upon core electron mission /217/.

tional fine structure of the electron line and with the intensities given by the Franck-Condon factors. This finding can be correlated with the simultaneously made discovery in our laboratory of vibrational fine structures in soft X-ray emission lines [218-233]. This development is further described in theses work by L. O. Werme [234], J. Nordgren [235] and H. Ågren [236]. Combined, these results show that vibrations occur in these molecules during X-ray emission both in the initial and the final states.

The above high resolution instrument designed together with U. Gelius was planned to be a prototype instrument for a new generation of advanced instruments which have now been constructed in a recently built laboratory for electron spectroscopy in Uppsala [237]. These have just been finished and are the sixth generation in the sequence from my laboratory since 1954. Two of the new instruments are designed for molecular studies and the third for surface

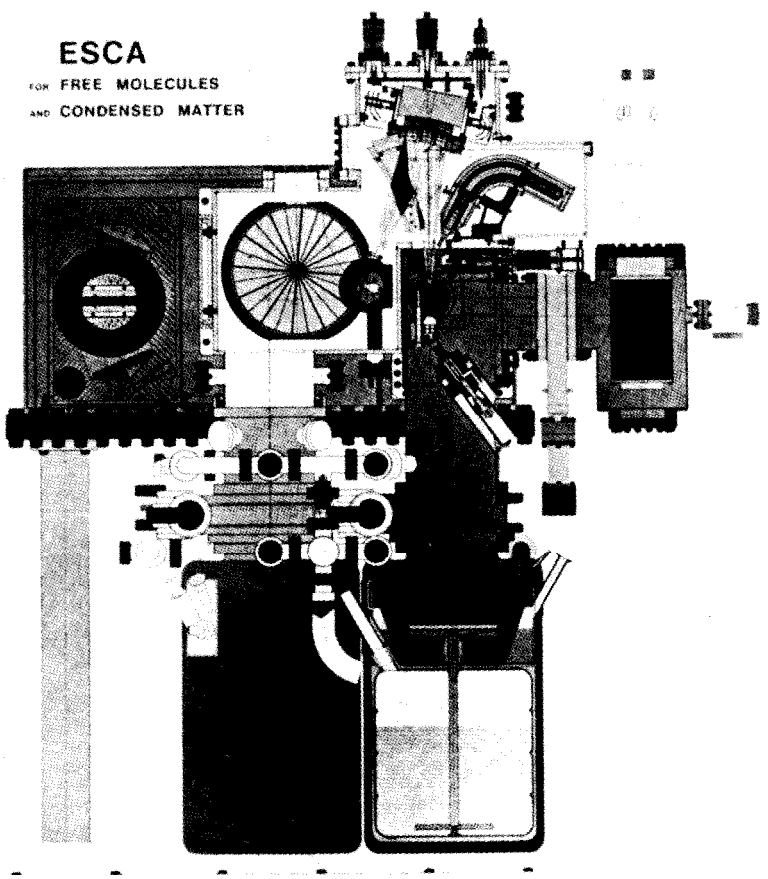


Fig. 16. Side view of the new ESCA instrument for free molecules and condensed matter. The instrument is UHV compatible and includes four diffcrent excitation modes (Monochromatized AlKα; Monochromatized and polarized UV; electron impact; monochromatized electron impact).

studies. The spherical electrostatic analyzer (R=36 cm) is provided with an electrostatic lens system due to B. Wannberg [238]. The modes of excitation included in the instruments are: monochromatic $AlK\alpha$ radiation ($\triangle h\upsilon=0.2$ eV) at 1486.6 eV; a UV light source with a grating providing selected UV lines between 10 eV to -50 eV; an electron monochromator of variable energy with an energy homogeneity of $\lesssim$ 10 meV and an additional electron gun for Auger electron excitation, also variable in energy (Fig. 16). A polarizer for the UV light at different wavelengths can alternatively be used in angular distribution studies.

The new instrument has been put into operation. As an illustration of its improved qualities Fig. 17 shows a new investigation of the previously recorded spectrum of methane (compare with Fig. 15) under increased resolution [261]. The vibrational structure is now well resolved. A convolution of the recorded spectrum using the window curve of the spectrometer and a computer program yields a remaining spectrum consisting of three narrow lines, the widths (- 100 meV) of which can be measured with good accuracy. For calibration purposes the Ar 2 $P_{1/2,\ 3/2}$ lines are simultaneously recorded. This investigation approaches an accuracy of a few meV in the determination of binding energies around 300 eV, i.e. close to 1: 10^5.

The brief account I have given above concerns the work which was done in my laboratory in the development of electron spectroscopy. During the seventies several reviews and books on electron spectroscopy have been written and for a complete account the reader has to go to such sources [239-258]. From my own laboratory two new books have just been completed authored by Hans Siegbahn and Leif Karlsson [259, 260], which cover developments mainly after 1970 and present current experimental and theoretical aspects on electron spectroscopy.

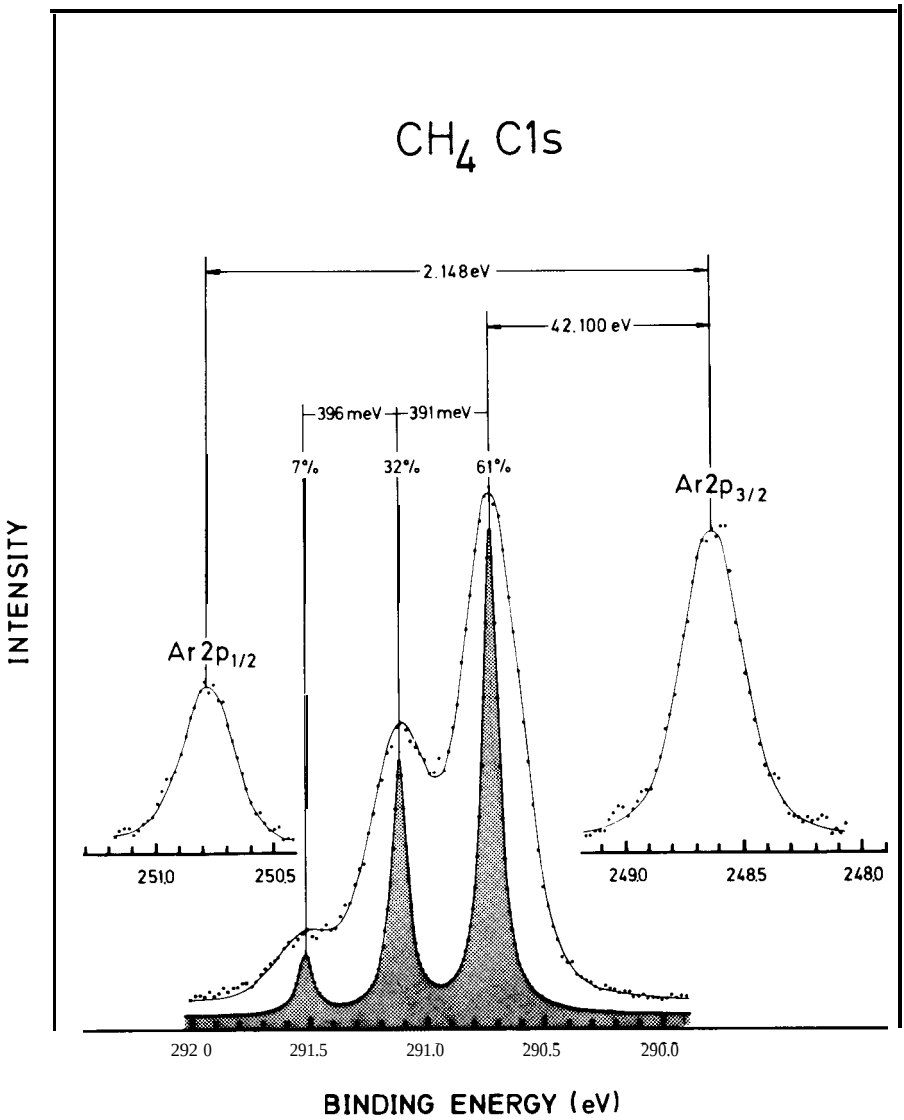


Fig. 17. New study of the methane Cls core vibrational structure (compare Fig. 15) by means of the instrument acc. to Fig. 16. The structure is resolved and deconvoluted into three narrow lines which yield the binding energies and widths of the components to a high degree of accuracy. Argon is used as a calibration gas. mixed with methane.

REFERENCE LIST

I. Siegbahn K.. Studies in P-Spectroscopy (Thesis), Arkiv f. Mat. Astr., Fys. *30A*, 1 (1944).
2. Siegbahn K., and Svartholm N., Nature 157, 872 (1946).
3. Svartholm N., and Siegbahn K., Arkiv f. Mat. Astr., Fys. *33A*, 21 (1946).
4. Hedgran A., Siegbahn K., and Svartholm N., Proc. Phys. Soc. *A63*, 960 (1950).
5. Siegbahn K., in Alpha-. Beta- and Gamma-Ray Spectroscopy, ref. 25, p. 79.
6. Hofstadter R., Les Prix Nobel, 1961.
7. Lindström G., Nuclear Resonance Absorption Applied to Precise Measurements of Nuclear Magnetic Moments and the Establishment of an Absolute Energy Scale in P-Spectroscopy (Thesis). Ark. Fysik 4, 1 (1951).
8. Hedgran A., Precision Measurements of Nuclear Gamma-Radiation by Techniques of Beta-Spectroscopy (Thesis), Almqvist & Wiksell. Uppsala 1952.
9. Bergström I., The Isomers of Krypton and Xenon: An Investigation with Eletromagnetically Separated Radioactive Isotopes (Thesis). Ark. Fysik 5. 191 (1952).
10. Thulin S., Studies in Nuclear Spectroscopy) with Electromagnetically Separated Gaseous Isotopes (Thesis), Almqvist & Wiksell, Uppsala 1954.
11. Gerholm T.R., Coincidence Spectroscopy; Studies of Nuclear Magnetic Dipole Transition Probabilities and of Positron Annihilation Lifetimes. Design and Construction of an Electron-Electron Coincidence Spectrometer and Its Application to the Investigation of Some Complex Nuclear Disintegrations (Thesis). Almqvist & Wiksell. Uppsala 1956.
12. Lindqvist I'., Angular Correlation Measurements: Studies of Relative Transition Probabilitirs of Gamma Radiation and of Internal Compton Effect. Experiments on Nuclear g-Factors (Thesis), Almqvist & Wiksell, Uppsala 1957.
13. Bäckström G.. Levels and Transitions of Atomic Nuclei: Experiments and Experimental Methods (Thesis), Almqvist & Wiksell, Uppsala 1960.
14. Pettersson B.-G., Experimental Studies of Electron-Gamma Directional Correlations (Thesis), Almqvist & Wiksell, Uppsala 196 1.
15. Karlsson E., Studies of Excitation Modes of Atomic Nuclei by Means of Magnetic Moment Determinations and General Decay Investigations (Thesis). Acta Universitatis Upsaliensis 11 , 1962.
16. Thun J.E., Studies of Nuclear Properties of Decay Modes of Radioactive Nuclei; Directional Correlation and Coincidence Investigations (Thesis). Acta Universitatis Upsaliensis 15, 1962.
17. Lindskog J., Studies of Nuclear Properties by Coincidence Spectrometer Methods; Decay Schemes and Transition Probabilities (Thesis), Acta Universitatis Upsaliensis 22. 1963.
18. Matthias E., Experimental and Theoretical Investigations of Perturbed Angular Correlations (Thesis), Almqvist & Wiksell, Uppsala 1963.
19. Schneider W., Applications of Digital Computer Techniques to Theoretical Investigations and the Analysis of Experimental Data in Nuclear Spectroscopy (Thesis), Almqvist & Wiksell, Uppsala 1963.
20. Bergman O., Experimental Studies of Nuclear States and Transitions (Thesis), Acta Universitatis Upsaliensis 41, 1964.
21. Sundström T, Electromagnetic Transitions in Atomic Nuclei: Experimental Studies of Transition Probabilities and Decay Schemes (Thesis). Acta Universitatis Upsaliensis 47, 1964.
22 Klcinheinz P., Studies of Internal Conversion Transitions in the Decay of Excited Nuclear States (Thesis), Almqvist & Wiksell, Uppsala 1965.
23. Bäcklin A., Nuclear Levels Excited in Ncutron Capture Reactions and Radioactive Decays (Thesis), Acta Universitatis Upsaliensis 96, 1967.
24. Siegbahn K. (Editor), Beta- and Gamma Ray Spectroscopy, North-Holland Publ. Co., Amsterdam, 1955.
25. Siegbahn K. (Editor), Alpha-. Beta and Gamma Ray Spectroscopy, North-Holland Publ. Co., Amsterdam, 1965.
26. Jenkin J.G., Leckey R.C.G. and Liesegang J., J. Electron Spectrosc. 12, I (1977).

27. Carlson T.A. (Editor), X-ray Photoelectron Spectroscopy, Benchmark Papers in Phys. Chem. and Chem. Phys. Vol. 2, Dowden, Hutchinson and Ross (1978).
28. Siegbahn K., Conf. of the Swedish Nat. Committee for Physics, Stockholm, 1952, Ark. Fys. 7, *86* (1954).
29. Siegbahn K., Conf. of the Swedish Nat. Committee for Physics, Stockholm, 1953, Ark. Fys. 8, 19 (1954).
30. Siegbahn K., Introductory Talk at the Int. Conf. on Beta and Gamma Radioactivity, Amsterdam, 1952; Physica 18, 1043 (1952).
31. Siegbahn K. and Edvarson K., Nucl. Phys. 1, 137 (1956).
32. Sokolowski E., Edvarson K. and Siegbahn K., Nucl. Phys. 1, 160 (1956).
33. Nordling C., Siegbahn K., Sokolowski E. and Wapstra A.H., Nucl. Phys. I, 326 (1956).
34. Nordling C., Sokolowski E. and Siegbahn K., Phys. Rev. 105, 1676 (1957).
35. Sokolowski E., Nordling C. and Siegbahn K., Ark. Fysik. 12, 301 (1957).
36. Robinson H., Phil. Mag. 50, 241 (1925).
37. Siegbahn K., Nordling C. and Sokolowski E., Proc. Rehovoth Conf. Nucl. Structure, North-Holland Publ. Co., Amsterdam, 1957, p. 291.
38. Sokolowski E., Nordling C. and Siegbahn K., Phys. Rev. 110, 776 (1958).
39. Nordling C., Sokolowski E. and Siegbahn K., Arkiv Fysik 13, 282 (1958).
40. Sokolowski E., Nordling C. and Siegbahn K., Arkiv Fysik 13, 288 (1958).
41. Nordling C., Sokolowski E. and Siegbahn K., Arkiv Fysik 13, 483 (1958).
42. Sokolowski E. and Nordling C., Arkiv Fysik 14, 557 (1959).
43. Sokolowski E., Arkiv Fysik 15, 1 (1959).
44. Nordling C., Arkiv Fysik 15, 241 (1959).
45. Nordling C., Arkiv Fysik *1.5, 397* (1959).
46. Nordling C. and Hagström S., Arkiv Fysik 15, 431 (1959).
47. BergvaIl P. and Hagström S., Arkiv Fysik 16, 485 (1960).
48. Nordling C. and Hagström S., Arkiv Fysik 16, 515 (1960).
49. Bergvall P. and Hagström S., Arkiv Fysik 17, 61 (1960).
50. Bergvall P., Hörnfeldt 0. and Nordling C., Arkiv Fysik 17, 113 (1960).
51. Nordling C., Hagström S. and Siegbahn K., Arkiv Fysik 22, 428 (1962).
52. Fahlman A., Hörnfeldt 0. and Nordling C., Arkiv Fysik 23, 75 (1962).
53. Hagström S., Hörnfeldt O., Nordling C. and Siegbahn K. Arkiv Fysik 23, 145 (1962).
54. Hagström S., Nordling C. and Siegbahn K., "Alpha-, Beta- and Gamma-Ray Spectroscopy", Appendix 2. Ed.Siegbahn K., North-Holland Publ. Co., Amsterdam 1965.
55. Hagström S. and Karlsson S.-E., Arkiv Fysik 26, 252 (1964).
56. Hagström S. and Karlsson S.-E., Arkiv Fysik 26, 451 (1964).
57. Fahlman A. and Hagström S., Arkiv Fysik 27, 69 (1964).
58. Andersson I. and Hagström S., Arkiv Fysik 27, 161 (1964).
59. Nordling C. and Hagström S., Z. Phys. 178, 418 (1964).
60. Fahlman A., Hamrin K., Nordberg R., Nordling C. and Siegbahn K., Phys. Rev. Lett. 14, 127 (1965).
61. Fahlman A., Hamrin K., Nordberg R., Nordling C., Siegbahn K. and Holm L.W., Phys. Lett. 19,643 (1966).
62. Hörnfeldt 0. and Fahlman A., Arkiv Fysik 22, 412 (1962).
63. Hörnfeldt O., Fahlman A. and Nordling C., Arkiv Fysik 23, 155 (1962).
64. Hörnfeldt O., Arkiv Fysik 23, 235 (1962).
65. Hagström S., Z. Phys. *178, 82 (1964).*
66. Fahlman A. and Nordling C. Arkiv Fysik 26, 248 (1964).
67. Fahlman A., Nordberg R., Nordling C. and Siegbahn K., Z. Phys. 192, 476 (1966).
68. Fahlman A., Hamrin K., Axelson G., Nordling C. and Siegbahn K., Z. Phys. 192,484 (1966).
69. Fahlman A., Hamrin K., Nordberg R., Nordling C. and Siegbahn K., Phys. Lett. 20, 159 (1966).
70. Sokolowski E., Investigations of Inner Electron Shells by Spectroscopic Studies of Photo- and Auger Electrons (Thesis), Almqvist & Wiksell, Uppsala 1959.

71. Nordling C., Experimental Studies of Electron Binding Energies and Auger Spectra (Thesis), Almqvist & Wiksell, Uppsala 1959.
72. Bergvall P., Investigations of Atomic and *Nuclear* Energy Levels by Crystal Diffraction and Photo Electron Spectroscopy (Thesis), Almqvist & Wiksell, Uppsala 1960.
73. Hörnfeldt O., Studies of Atomic Level Energies and Auger Spectra (Thesis), Almqvist & Wiksell, Uppsala 1962.
74. Hagström S., Studies of Some Atomic Properties by Electron Spectroscopy (Thesis), Almqvist & Wiksell, Uppsala 1964.
75. Fahlman A., Electron Spectroscopy of Atoms and Molecules (Thesis), Acta Universitatis Upsaliensis 73, 1966.
76. Hagström S., Nordling C. and Siegbahn K., Phys. Lett. 9, 235 (1964).
77. Karlsson S.-E., Precision Electron Spectroscopy for Studies of Nuclear and Atomic Energy Levels (Thesis), Acta Universitatis Upsaliensis 99, 1967.
78. Nordberg R., ESCA Studies of Atoms and Molecules (Thesis), Acta Universitatis Upsaliensis 118, 1968.
79. Hamrin K., ESCA Applied to Solids and Gases (Thesis), Acta Universitatis Upsaliensis 151, 1970.
80. Hedman J,, ESCA Studies of Electron Levels and Bands in Atoms, Molecules and Solids (Thesis), Acta Universitatis Upsaliensis 196, 1972.
81. Johansson G., Chemical Bonding and Electronic Structure of Molecules and Solids Studied by ESCA (Thesis), Acta Universitatis Upsaliensis 205, 1972.
82. Gelius U., Molecular Spectroscopy by Means of ESCA; Experimental and Theoretical Studies (Thesis), Acta Universitatis Upsaliensis 242, 1973.
83. Lindberg B., Some Aspects of the Sulphur-Oxygen Bond in Organic Sulphur Groups (Thesis), Acta Universitatis Upsaliensis 173, 1970.
84. ESCA I, ref. 211, p. 21.
85. Gelius U., Basilier E., Svensson S., Bergmark T. and Siegbahn K., J. Electron Spectrosc. 2, 405 (1974).
86. Siegbahn K., ibid. 5, 3 (1974).
87. Clark D.T., Dilks A., Shuttleworth S. and Thomas H.R., ibid., 14, 247 (1978).
88. Clark D.T. and Dilks A., J. Polymer Sci. 16, 911 (1978).
89. Clark D.T., Phys. Scr. 16. 307 (1977).
90. Clark D.T., in Handbook of X-ray and UV Photoelectron Spectroscopy (Ed. D. Briggs) Heyden 1977, p. 211.
91. ESCA I, ref. 211 , sect 111-9, p. 63.
92. Goscinski O., Howat G. and Åberg T., J. Phys. B8, 11 (1975).
93. Pickup B. and Goscinski O., Mol. Phys. 26: 1013 (1973).
94. Goscinski O., Hehenberger M., Roos B. and Siegbahn P., Chem. Phys. Lett. 33, 427, (1975).
95. ESCA II, ref. 212. Sect. 5.4.
96. Gelius U., Hedén P.-F., Hedman J., Lindberg B.J., Manne R.. Nordberg R., Nordling C. and Siegbahn K., Phys. Scr. 2, 70 (I 970).
97. Lindberg B..J., Hamrin K., Johansson G., Gelius U., Fahlman A., Nordling C. and Siegbahn K., Phys. Scr. I, 286 (1970).
98. Allison D.A., Johansson G., Allan CJ., Gelius U., Siegbahn H., Allison J. and Siegbahn K., J. Electron Spectrosc. *1, 269* (1972/73).
99. Shirley D.A., Adv. Chem. Phys. 23, 85 (1973).
100. Clark D.T., in Electron Emission Spectroscopy. (Eds. Dekeyser et al.) D. Reidel Publ. Co. (1973).
101. Davis D.W., Banna MS. and Shirley D.A., J. Chem. Phys. 60, 237 (1974).
102. Wyatt J.F., Hillier I.H., Saunders V.R., Connor J.A. and Barber M., J. Chem. Phys. 54, 5311 (1971).
103. Jolly W.L. and Perry W.B., Inorg. Chem. *13, 2686* (1974).
104. Jolly W.L. and Perry W.B., J. Amer. Chem. Soc. 95, 5442 (1974).
105. Carver, J. C., Gray R. C. and Hercules, D. M., J. Amer. Chem. Soc. 96, 6851 (1974).
106. Perry, W. B. and Jolly, W. L., Inorg. Chem. *13,* 1211 (1974).

107. Gelius, U., Johansson, G. Siegbahn, H. Allan, C. J., Allison, D. A., Allison J. and Sicgbahn, K., J, Electron Spectrosc. *1, 285* (1972173).
108. Siegbahn. H., Mcdeiros R. and Goscinski, O. ibid. 8, 149 (1976).
109. Bagus P. S. and Hermann, K., Solid State Comm. 20, 5 (1976).
110. Ellis, D. E., Baerends, E. J., Adachi H. and Averill, F. W., Surf. Sci. 64, 649 (1977).
111. Ley, L. Kowalczyk, S. P.. McFeely, F. R., Pollak R. A., and Shirley, D. A. Phys. Rev. B8, 2392 (1973).
112. Kowalrzyk. S. P., Pollak. R. A.. McFeely, F. R., Lcy L. and Shirley, D. A., ibid. *B8, 2387* (1973).
113. Shirley, D. A., Martin. R. L., Kowalczyk. S. P.. McFeely F. R. and Ley, I,.. Phys. Rev. *B15,* 544 (1977).
114. Svensson, S.. MårtenssonN. , Basilicr, E., Malmquist, P. Å., Gelius U. and Siegbahn, K., J. Electron Spcctrosc. 9, 51 (1976).
115. Jen, J. S. and Thomas, T. D.. Phys. Rev. *B13, 5284 (1976).*
116. Johansson B. and Mårtensson, N.. ibid. B21. 4427 (1980).
117. Mårtensson, N., Atomic, Molecular and Solid State Effects in Photoelectron Spectroscopy (Thesis), Acta Universitatis Upsalicnsis 579. Uppsala 1980.
118. Nyholm. R., Electronic Structure and Photoionization Studied by Electron Spectroscopy (Thesis). Acta Universitatis Upsaliensis 581, Uppsala 1980.
119. Mårtensson, N., Nyholm, R., Calén. H.. Hedman, J. and Johansson. B. Uppsala University, Institute of Physics Report, UUIP-1008 (July 1980).
120. Kowalczyk, S. P., Pollak. R. A., McFerly. F. R. and Shirley, D. A., Phys. Rev. *B8, 3583* (1973).
121. Kowalczyk. S. P.. Ley. L., McFeely. F. R., Pollak, R. A. and Shirley, D. A.. Phys Rev. *B9,* 381 (1974).
122. Wagner. C. D. and Biloen, P. Surf. Sci. 35. 82 (1973).
123. Matthew, J. A. D.. Surf. Sci. 40, 451 (1973).
124. Utriainen, J., Linkoaho. M. and Åberg, T.. in Proc. Int. Symp. X-ray Sp. and El. Str. of Mat. *1, 382* (1973).
125. Asplund, L., Kelfve, P., Siegbahn. H.. Goscinski. 0.. Fellncr-Feldegg, H., Hamrin, K., Blomster. B. and Siegbahn. K., Chem. Phys. Lett. 40, 353 (1976).
126. Keski-Rahkoncn. O. and Krause. M. O., J. Electron Spectrosc. 9, 371 (1976).
127. Sieghahn. H. and Goscinski. 0.. Phys. Scr. 13. 225 (1976).
128. Asplund, L, , Kelfve, P., Blomstcr, B., Siegbahn. H.. Siegbahn, K., Lozes, R. L. and Wahlgren U. I., Phys. Scr. *16. 273 (1977).*
129. Kim, K. S., Gaarcnstroom. S. W. and Winograd, N.. Chem. Phys. Lett. *41, 503* (1976).
130. Wagner, C. D., Far. Disc. Chem. Soc. 60, 291 (1975).
131. Wagner. C. D., in Handbook of X-ray and Ultraviolet Photoelectron Spectroscopy (Ed. D. Briggs) Chapt. 7 (Heyden. 1977).
132. Kelfvc, P.. Blomster, B., Siegbahn, H., Siegbahn. K., Sanhueza, E. and Goscinski O., Phys. *Scr. 21, 75* (1980).
133. Thomas. T. D., J. Electron Spectrosc. 20. 117 (1980).
134. ESCA I, ref. 21 1. p. 118.
135. Pignataro, S. and Distefano. G., J, Electron Spectrosc, 2, 171 (1973). Z. Naturforsch. 30a, 815 (1975).
136. Ågren, H., Roos, B., Bagus, P. S., Gelius. U., Malmquist, P. Å., Svensson, S., Maripuu, R. and Siegbahn, K., Uppsala University, Institute of Physics Report, UUIP-1057 (1982).
137. Yin, L., Adler, I., Tsang, T., Mantienzo, L. J. and Grim, S. O., Chem. Phys. Lett. 24, 81 (1974).
138. Pignataro, S., Foffani, A. and Distefano, G., Chem. Phys. Lett. 20, 350 (1973).
139. Johansson, L. Y., Larsson, R.. Blomquist, J.. Cederström, C., Grapengiesser, S., Hclgeson, U., Moberg, L. C. and Sundhom, M., Chem. Phys. Lett. 24, 508 (1974).
140. Barber, M., Connor, J, A. and Hillier, I. H., Chem. Phys. Lett. 9, 570 (1971).
141. Hüfner, S. and Wertheim, G. K., Phys. Rev. *B7. 5086* (1973).
142. Rosencwaig, A., Wertheim, G. K. and Guggenheim, H. J.. Phys. Rev. Lett. 27, 479 (1971).

143. Rosencwaig, A. and Wertheim, G. K., J. Electron Spectrosc. 1, 493 (1973).
144. Frost, D. C., McDowell, C. A. and Woolsey, I. S., Chem. Phys Lett. *17, 320* (1972).
145. Kim, K. S., Chem. Phys. Lett. 26, 234 (1974).
146. Carlson, T. A., Carver, J. C., Saethre, L. J., Santibanez, F. G. and Vernon, G. A., J. Electron Spectrosc, 5, 247 (1974).
147. Fuggle, J. C., Umbach, E., Menzel, D., Wandelt, K. and Brundle, C. R., Solid State Comm. 27,65 (1978).
148. Lang, N. D. and Williams, A. R., Phys. Rev. *B16, 2408* (1977).
149. Schönhammer, K. and Gunnarsson. 0.. Solid State Comm. 23. 691 (1978) and 26. 399 (1978) and Z., Phys. *B30, 297* (1978).
150. Hedman, J., Hedén. P. F.. Nordling, C. and Siegbahn. K., Phys. Lett. A29. 178 (1969).
1.51. ESCA II. ref 212, section 5.1, p 56.
152. ESCA 1, ref. 21 I, section IV, p. 72.
153. Siegbahn, K.. Gelius, U.. Siegbahn. H and Olsson. E., Phys. Lett. 32A. 221 (1970) and Phys. *Scr. I. 272* (1970).
1.54. Fadley, C. S. and Bergström S. Å. I,.. Phys Lett. 35A, 373 (1971).
155. Baird. R.J., Fadley. C. S. and Wagner. I,. F.. Phy. Rev. B15. 666 (1977).
136. Zehner, D.M., Noonan, J. R. and Jenkins. L. H.. Phys. Lett. 62A. 267 (1977).
157. Briggs. D.. Marbrow. R. A. and Lambert. R. M . Solid State Comm. 25. 40 (1978).
158. Erickson. N. E.. Phys. Scr. *16, 462* (1977).
159. Kono, S.. Goldberg, S. M.. Hall, N. F. T. and Fadley. C S.. Phys. Rev. Lett. 41, 1831 (1978).
160. Petersson, L.. G.. Kono. S.. Hall. N. F T.- . Fadley C S. and Pendry, J B.. Phys. Rev. Lett 42, 1545 (1979).
161. Woodruff, D. P., Norman. D.. Holland. B. W., Smith. N. W., Farrell. H. H. and Traum. M.. M.. Phys. Rev. Lett. *41,* 1 130 (1978).
t 62. Kevan. S. D., Rosenblatt. D.. H . Denley. D. R.. Lee. B.-C:. and Shirley. D. A., Phys. Rev. B20. 4133 (1979); Phys. Rev. Lett. 41. 156.5 (1978).
163. McGovern, I. T.. Eberhardt. W. and Plummer. E. W.. Sol. St. Comm. 32, 963 (1979).
164. Guillot. C., Jugnet. Y., l,asailly, Y., Lecante, J., Spanjaard. D. and Duc. T. M. (private communication. 1980).
165. Bergmark. T . Magnusson. N. and Siegbahn. K.. Ark. Fys. 37. 3.53 (1968).
166. ESCA I. ref. 211 , p 20.
167. ESCA I. ref. 211, p. 21.
168. Siegbahn. H. and Siegbahn. K., J. Electron Spretrosc. 2. 319 (1973).
169. Sieghahn. H.. Asplund, L., Kelfve. P.. Hamrin. K.. Karlsson. I,. and Siegbahn. K. ,J Electron Spectrosc. 5, 1059 (1974).
170. Siegbahn. H.. Asplund. L.. Kelfve. P. and Siegbahn. K., J. Electron Spectrosc. 7. 411 (1975).
171. Fellner-Feldegg. H.. Siegbahn. H.. Asplund. I... Kellfve. P. and Siegbahn. K., ,J. Electron Spectrosc. 7. 421 (197.5).
172. Lindberg. B Asplund. L., Fellner-Feldegg. H.. Kelfve. I'.. Siegbahn. H. and Siegbahn K.. Chem. Phys. Lett. .79, 8 (1976).
173. Siegbahn. H.. Svensson, S. and Lundholm. M., ,J. Electron Spectrosc. 24, 205 (1981). The first considerations of continuously renewable liquid samples in vacuum for ESCA and experiments on liquid beams were reported at the Asilomar conference 1971: K. Siegbahn: Perspectives and Problems in Electron Spectroscopy. (Editor: D. Shirley), North Holland, 1972. Further accounts for these developments arc found in ref. /168/-/172/. Different suggested alternatives for handling the renewable samples (liquid beams, wetted rotating wires and rotating disks) were presented at the Namur Conference 1974, K. Siegbahn: Electron Spectroscopy-An Outlook (secs 3 and 4, figs 5-10). Editors: R. Caudano and J. Verbist, Elsevier, 1974. References to recent other works on liquids are found in H. Siegbahn ct. al. ref. /173/ (R. Ballard et. al. see ref. ibid. 6 and 7: P. Delahay et. al. ref. ibid. 8-10; K. Burger et. al. ref. ibid. 11 and 12; S. Avanzino and W. ,Jolly ref ibid. 13).
174. Siegbahn, H., ESCA Studies of Electronic Structure and Photoionization in Gases, Liquids and Solids (Thesis), Acta Universitatis Upsaliensis 277, Uppsala 1974.

175. Asplund, L., Electron Spectroscopy of Liquids and Gases (Thesis), Acta Universitatis Upsaliensis 409, Uppsala 1977.
176. Kelfve, P., Electronic Structure of Free Atoms, Molecules and Liquids Studied by Electron Spectroscopy (Thesis), Acta Universitatis Upsaliensis 483, Uppsala 1978.
177. Turner, D. W. and Al-Joboury, M. I., J. Chem. Phys. 37, 3007 (1962).
178. Turner, D. W. and Al-Joboury, M. I., Chem. Soc., London, 5141 (1963).
179. Al-Joboury, M. I., May, D. P. and Turner, D. W., J. Chem. Soc., 616,4434, 6350 (1965).
180. Radwan, T. N. and Turner D. W., J. Chem. Soc., 85 (1966).
181. Al-Joboury, M. I. and Turner, D. W., J. Chem. Soc., 373 (1967).
182. Turner, D. W., Baker, C., Baker, A. D. and Brundle, C. R., Molecular Photoelectron Spectroscopy, Wiley-Interscence, London (1970).
183. Price, W. C., Endeavour 26, 78 (1967).
184. Lempka, H. J., Passmore, T. R. and Price, W. C., Proc. Roy. Soc. *A304, 53* (1968).
185. Berglund, C. N. and Spicer, W. E., Phys. Rev. 136, A1030; A1044 (1964).
186. Spicer, W. E., J. Appl. Phys. 37, 947 (1966).
187. Blodgett, A. J., Jr., and Spicer, W. E., Phys. Rev. 146, 390 (1966).
188. Blodgett, A. J., Jr. and Spicer, W. E. Phys. Rev. 158, 514 (1967).
189. Bergmark, T., Electronic Structure of Atoms and Molecules Studied by Means of Electron Spectroscopy (Thesis), Acta Universitatis Upsaliensis 274, 1974.
190. Karlsson, L., High Resolution Studies of the Valence Electron Structure of Atoms and Molecules by Means of Electron Spectroscopy (Thesis), Acta Universitatis Upsaliensis 429, 1977.
191. Jadmy, R., Electronic Structure Studied by Electron Spectroscopy and Angular Distributions (Thesis), Acta Universitatis Upsaliensis 463, 1978.
192. Mattsson, L., High Resolution Valence Electron Spectroscopy. Development of a VUV-Polarizer for Angular Distributions (Thesis), Acta Universitatis Upsaliensis 562, 1980.
193. Asplund, L., Kelfve, P., Blomster, B., Siegbahn, H. and Siegbahn, K., Phys. Scr. 16. 268 (1977).
194. Siegbahn, H., Asplund, L. and Kelfve, P., Chem. Phys. Lett. 35, 330 (1975).
195. Siegbahn, H., in Excited States in Quantum Chemistry (Eds. Nicolaides C. A. and Beck, D. R.), Plenum (1978) p. 273.
196. Darko, T., Siegbahn, H. and Kelfve, P., Chem. Phys. Lett. 82, 475 (1981).
197. Karlsson, L., Mattsson, L., Jadmy, R., Bergmark, T. and Siegbahn, K., Phys. Scr. 14, 230 (1976).
198. Gelius, U., J. Electron Spectrosc. 5, 985 (1974).
199. Krause, M. O., Chem. Phys. Lett. IO, 65 (1971).
200. Wuilleumier, F. and Krause, M. O., Phys. Rev. A10, 242 (1974).
201. Banna, M. S. and Shirley, D. A., Chem. Phys. Lett. 33, 441 (1975), J. Electron Spectrosc. 8, 23 (1976); ibid. 8, 255 (1976).
202. Nilsson, R., Nyholm, R., Berndtsson, A., Hedman,,J. and Nordling, C.. J. Electron Spectrosc. *9, 337* (1976).
203. Berndtsson, A., Nyholm, R. Nilsson, R. Hedman J. and Nordling, C. ibid. 13, 131 (1978).
204. Cavell, R. G. and Allison, D. A., Chem. Phys. Lett. 36, 514 (1975).
205. Allison, D. A. and Cavell, R. G., J. Chem. Phys. 68, 593 (1978).
206. **Kunz,** C. (Ed.), Synchrotron Radiation, Techniques and Applications, Topics in Current Physics 10, Springer-Verlag (1979).
207. Synchrotron Radiation Instrumentation and New Developments. Nuclear Instr. Meth. 152. (Conf. Vol., Eds: Wuilleumier, P. and Y. Farge) 1978.
208. Synchrotron Radiation Instrumentation. Nuclear Instr. Meth. *172* (Conf. Vol.. Eds: Ederer. D. L. and West,J. B.) 1980.
209. Synchrotron Radiation Facilities, ibid. 177 Conf. Vol., Ed. Howells. 51. R.) 1980.
210. Synchrotron Radiation Research, Eds: Winick H. and Doniach, S., Plenum (1980).
211. Siegbahn, K., Nordling, C., Fahlman, A., Nordberg, R., Hamrin, K., Hedman, J., Johansson, G., Bergmark, T., Karlsson, S.-E., Lindgren., I. and Lindberg, B., ESCA - Atomic, Molecu-

lar and Solid State Structure Studied by Means of Electron Spectroscopy, Nova Acta Regiae Soc. Sci. Upsaliensis, Ser. IV, Vol. 20 (1967).
212. Siegbahn, K., Nordling, C., Johansson, G., Hedman, J., Hedén, P. F., Hamrin, K., Gelius, U., Bergmark, T., Werme, L. O., Manne R. and Baer Y., ESCA Applied to Free Molecules, North-Holland Publ. Co., Amsterdam-London, 1969.
213. Siegbahn, K., Hammond, D., Fellner-Feldegg, H. and Barnett, E. F., Science 176, 245 (1972).
214. Wassberg. G. and Siegbahn, K., Ark. Fys. 14, 1 (1958).
215. Hagström, S. and Siegbahn, K., J. Ultrastructure Research 3, 401 (1960).
216. Gelius, U. and Siegbahn, K., Far. Disc. Chem. Soc. 54, 257 (1972).
217. Gelius, U., Svensson, S., Siegbahn H., Basilier, E., Faxälv, Å, and Siegbahn, K., Chem. Phys. Lett. 28, 1 (1974).
218. Siegbahn, K. in Electron Spectroscopy (Ed. D.A. Shirley), Proc. Intern. Conf. Asilomar, California, USA, 1971. North-Holland, Amsterdam, 1972.
219. Siegbahn, K., Werme, L.O., Grennberg, B., Lindeberg, S. and Nordling, C. Uppsala University Institute of Physics Report, UUIP-749 (Oct. 1971).
220. Siegbahn, K., Werme, L.O., Grennberg, B., Nordgren, J. and Nordling, C, Phys. Lett *41A* (2), 111 (1972).
221. Werme, L.O., Grennberg, B., Nordgren, J., Nordling, C. and Siegbahn, K., Phys. Lett *41A* (2), 113 (1972).
222. Werme, L.O., Grennberg, B., Nordgren, J., Nordling, C. and Siegbahn, K., Phys. Rev. Lett. 30 (12), 523 (1973).
223. Werme, L.O., Grennberg, B., Nordgren, J., Nordling, C. and Siegbahn, K., Nature 242 (5348), 453 (1973).
224. Werme, L.O., Grennberg, B., Nordgren, J., Nordling, C. and Siegbahn, K., J. Electron Spectrosc. 2, 435 (1973).
225. Werme, L.O., Nordgren, J., Nordling, C. and Siegbahn, K., C.R. Acad Sci. Paris, *B-119, 279* (1974).
226. Nordgren, J,, Werme, L.O., Ågren, H., Nordling, C. and Siegbahn, K., J. Phys. *B8(1),* L18 (1975).
227. Werme, L.O., Nordgren, J., Ågren, H., Nordling, C. and Siegbahn, K., Z. Physik 272, 131 (1975).
228. Nordgren, J., Ågren, H., Werme, L.O., Nordling, C. and Siegbahn, K., J, Phys. *B9 (2), 295* (1976).
229. Nordgren, J., Ågren, H., Selander, L., Nordling, C. and Siegbahn, K., paper presented at the Nordic Symposium on Atomic and Molecular Transition Probabilities in Lund, March 28-29, 1977.
230. Nordgren, J., Ågren, H., Nordling, C. and Siegbahn, K., Ann. Acad. Reg. Scientiarum Upsaliensis *21, 23* (1978).
231. Ågren, H., Nordgren, J., Selander, L., Nordling, C. and Siegbahn, K., Phys. Scr. *18, 499* (1978).
232. Nordgren, J., Ågren, H., Pettersson, L., Selander, L., Griep, S., Nordling, C. and Siegbahn, K., Phys. Scr. 20, 623 (1979).
233. Ågren, H., Selander, L., Nordgren, J., Nordling, C., Siegbahn, K. and Müller, J., Chem. Phys. 37, 161 (1979).
234. Werme, L.O., Electron and X-ray Spectroscopic Studies of Free Molecules (Thesis), Acta Universitatis Upsaliensis 241, 1973.
235. Nordgren, J., X-ray Emission Spectra of Free Molecules (Thesis) Acta Universitatis Upsaliensis 418, 1977.
236. Ågren, H., Decay and Relaxation of Core Hole States in Molecules (Thesis), Uppsala 1979.
237. Siegbahn, K. in Electron Spectroscopy, Eds. Hedman, J. and Siegbahn, K., Phys. Scr. *16,* 167, (1977).
238. Wannberg, B. and Sköllermo, A., J. Electron Spectrosc. 10, 45 (1977).
239. Phil. Trans. Roy. Soc. (London) 268, l-175 (1970).
240. Shirley, D.A. (Ed.) Electron Spectroscopy, North-Holland, Amsterdam (1972).
241. Faraday Disc. Chem. Soc. 54 (1972).

242. Dekeyser, W., et al. (Eds.) Electron Emission Spectroscopy, Reidel (1973)
243. Caudano, R. and Verbist, J. (Eds.) Electron Spectroscopy, Elsevier Amsterdam (1974)
244. Electron Spectroscopy of Solids and Surfaces, Far. Disc. Chem. Soc. 60 (1975).
245. Proc. Int. Symp. Electron Spectroscopy, Uppsala (May 1977), (Eds. Hedman, J, and Siegbahn, K.) Phys. Scr. *16 (5-6)* (1977).
246. Proc. Australian Conf. on Electron Spectroscopy, J. Electron Spectrosc. 15 (1979).
247. Baker, A.D. and Betteridge, D., Photoelectron Spectroscopy, Pergamon Press (1972).
248. Eland, J.H.D., Photoelectron Spectroscopy, Butterworths, London (1974).
249. Carlson, T.A., Photoelectron and Auger Spectroscopy. Plenum Press, New York (1975).
250. Brundle, C.R. and Baker, A.D. (Eds.) Electron Spectroscopy. Academic Press Vol.1 (1977), Vol. 2 (1978), Vol. 3 (1979). Vol. 4 (1981).
2.51. Briggs, D. (Ed.), Handbook of X-ray and Ultraviolet Photoelectron Spectroscopy, Heyden (1977).
252. Rabalais, J.W., Principles of Ultraviolet Photoelectron Spectroscopy, Wiley-Interscience (1977).
253. Ghosh. P.K., A Whiff of Photoelectron Spectroscopy, Swan Printing Press. New Delhi (1978).
254. Photoemission and the Electronic Propertics of Surfaces (Eds. Feuerbacher, B.. Fitton, B. and Willis. R.F.) Wiley-Interscience (1978).
253. Photoemission in Solids (Eds. Cardona. M. and Ley. I..). Topics in Applied Physics Vol. 26. 27, Springer-Verlag (1978. 1979).
256. Roberts, M.W, and McKee, C.S., Chemistry of. the Metal-Gas Interface. Clarendon Press (1978).
257. Ballard. R.E.. Photoelectron Spectroscopy and Molecular Orbital Theory, Adam Hilger (1978).
258. Berkowitz. ,J., Photoabsorption, Photoionization and Photoelectron Spectroscopy. Academic Press (1979).
259. Siegbahn, H. and Karlsson. I.., Photoelectron Spectroscopy, Handbuch d. Physik. Vol.31 (Ed. Mehlhorn. W.) Springer-Verlag (1982).
260. Siegbahn, H. and Karlsson. I... Electron Spectrosropy for Atoms, Molecules and Condensed Matter, North-Holland, Publ. Co. (to be published).
261. Gelius, U., Helenelund. K., Hedman, S.. Asplund, L.. Finnström, B. and Siegbahn, K. (to he published).

Physics 1982

KENNETH G WILSON

for his theory for critical phenomena in connection with phase transitions

THE NOBEL PRIZE FOR PHYSICS

Speech by Professor STIG LUNDQVIST of the Royal Academy of Sciences.
Translation from the Swedish text

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen,

The development in physics is on the whole characterized by a close interaction between experiment and theory. New experimental discoveries lead often rapidly to the development of theoretical ideas and methods that predict new phenomena and thereby stimulate further important experimental progress. This close interaction between theory and experiment keeps the frontiers of physics moving forward very rapidly.

However, there have been a few important exceptions, where the experimental facts have been well known for a long time but where the fundamental theoretical understanding has been lacking and where the early theoretical models have been incomplete or even seriously in error. I mention here three classical examples from the physics of the twentieth century, namely superconductivity, critical phenomena and turbulence. Superconductivity was discovered in the beginning of this century, but in spite of great theoretical efforts by many famous physicists, it took about fifty years until a satisfactory theory was developed. The theory of superconductivity was awarded the Nobel Prize in physics exactly ten years ago. The critical phenomena occur at phase transitions, for example between liquid and gas. These phenomena were known even before the turn of the century, and some simple but incomplete theoretical models were developed at an early stage. In spite of considerable theoretical efforts over many decades, one had to wait until the early seventies for the solution. The problem was solved in an elegant and profound way by Kenneth Wilson, who developed the theory which has been awarded this year's Nobel Prize in physics. The third classical problem I mentioned, namely turbulence, has not yet been solved, and remains a challenge for the theoretical physicists.

From daily life we know that matter can exist in different phases and that transitions from one phase to another may occur if we change, for example, the temperature. A liquid goes over into gas phase when sufficiently heated, a metal melts at a certain temperature, a permanent magnet loses its magnetization above a certain critical temperature, just to give a few examples. Let us consider the transition between liquid and gas. When we come close to the critical point, there will appear fluctuations in the density of the liquid at all possible scales. These fluctuations take the forms of drops of liquids mixed with bubbles of gas. There will be drops and bubbles of all sizes from the size of a single molecule to the volume of the system. Exactly at the critical point the scale of the largest fluctuations becomes infinite, but the role of the smaller fluctuations can by no means be ignored. A proper theory for the critical phenomena must take into account the entire spectrum of length scales. In

most problems in physics one has to deal with only one length scale. This problem required the development of a new type of theory capable of describing phenomena at all possible length scales, for example, from the order of a centimeter down to less than one millionth of a centimeter.

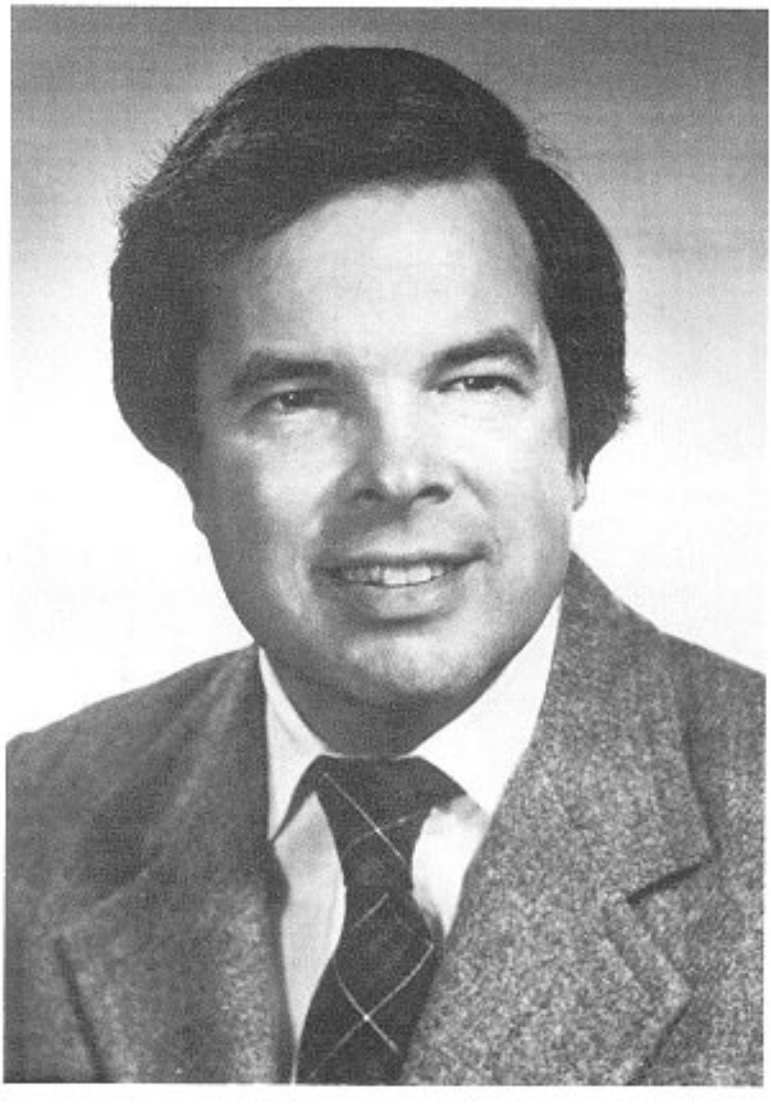
Wilson succeeded in an ingenious way to develop a method to solve the problem, published in two papers from 1971. A frontal attack on this problem is impossible, but he found a method to divide the problem into a sequence of simpler problems, in which each part can be solved. Wilson built his theory on an essential modification of a method in theoretical physics called renormalization group theory.

Wilson's theory gave a complete theoretical description of the behaviour close to the critical point and gave also methods to calculate numerically the crucial quantities. During the decade since he published his first papers we have seen a complete breakthrough of his ideas and methods. The Wilson theory is now also successfully applied to a variety of problems in other areas of physics.

Professor Wilson,

You are the first theoretical physicist to develop a general and tractable method, where widely different scales of length appear simultaneously. Your theory has given a complete solution to the classical problem of critical phenomena at phase transitions. Your new ideas and methods seem also to have a great potential to attack other important and up to now unsolved problems in physics.

I am very happy to have the privilege of expressing the warmest congratulations of the Royal Swedish Academy of Sciences. I now ask you to receive your Nobel Prize from the hands of His Majesty the King.



Kenneth G Wilson

KENNETH G. WILSON

I was born 1936 in Waltham, Massachusetts, the son of E. Bright Wilson Jr. and Emily Buckingham Wilson. My father was on the faculty in the Chemistry Department of Harvard University; my mother had one year of graduate work in physics before her marriage. My grandfather on my mother's side was a professor of mechanical engineering at the Massachusetts Institute of Technology; my other grandfather was a lawyer, and one time Speaker of the Tennessee House of Representatives.

My schooling took place in Wellesley, Woods Hole, Massachusetts (second, third/fourth grades in two years), Shady Hill School in Cambridge, Mass. (from fifth to eighth grade), ninth grade at the Magdalen College School in Oxford, England, and tenth and twelfth grades (skipping the eleventh) at the George School in eastern Pennsylvania. Before the year in England I had read about mathematics and physics in books supplied by my father and his friends. I learned the basic principle of calculus from *Mathematics and Imagination* by Kasner and Newman, and went of to work through a calculus text, until I got stuck in a chapter on involutes and evolutes. Around this time I decided to become a physicist. Later (before entering college) I remember working on symbolic logic with my father; he also tried, unsuccessfully, to teach me group theory. I found high school dull. In 1952 I entered Harvard. I majored in mathematics, but studied physics (both by intent), participated in the Putnam Mathematics competition, and ran the mile for the track team (and cross-country as well). I began research, working summers at the Woods Hole Oceanographic Institution, especially for Arnold Arons (then based at Amherst).

My graduate studies were carried out at the California Institute of Technology. I spent two years in the Kellogg Laboratory of nuclear physics, gaining experimental experience while taking theory courses; I then worked on a thesis for Murray Gell-Mann. While at Cal Tech I talked a lot with Jon Mathews, then a junior faculty member; he taught me how to use the Institute's computer; we also went on hikes together. I spent a summer at the General Atomic Company in San Diego working with Marshall Rosenbluth in plasma physics. Another summer Donald Groom (then a fellow graduate student) and I hiked the John Muir Trail in the Sierra Nevada from Yosemite Park to Mt. Whitney. After my third year I went off to Harvard to be a Junior Fellow while Gell-Mann went off to Paris. During the first year of the fellowship I went back to Cal Tech for a few months to finish my thesis. There was relatively little theoretical activity at Harvard at the time; I went often to M.I.T. to use their computer and eat lunch with the M. I. T. theory group, led by Francis Low.

In 1962 I went to CERN for a calendar year, first on my Junior Fellowship and then as a Ford Foundation fellow. Mostly, I worked but I found time to

join Henry Kendall and James Bjorken on a climb of Mt. Blanc. I spent January through August of 1963 touring Europe.

In September of 1963 I came to Cornell as an Assistant Professor. I received tenure as an Associate Professor in 1965, became Full Professor in 1971 and the James A. Weeks Professor in 1974. I came to Cornell in response to an unsolicited offer I received while at CERN; I accepted the offer because Cornell was a good university, was out in the country and was reputed to have a good folk dancing group, folk-dancing being a hobby I had taken up as a graduate student.

I have remained at Cornell ever since, except for leaves and summer visits: I spent the 1969- 1970 academic year at the Stanford Linear Accelerator Center, the spring of 1972 at the Institute for Advanced Study in Princeton, the fall of 1976 at the California Institute of Technology as a Fairchild Scholar, and the academic year 1979-80 at the IBM Zurich Laboratory.

In 1975 I met Alison Brown and in 1982 we were married. She works for Cornell Computer Services. Together with Douglas Von Houweling, then Director of Academic Computing and Geoffrey Chester of the Physics Department we initiated a computing support project based on a Floating Point Systems Array Processor. I helped write the initial Fortran Compiler for the Array Processor. Since that time I have (aside from using the array processor myself) been studying the role of large scale scientific computing in science and technology and the organizational problems connected with scientific computing. At the present time I am trying to win acceptance for a program of support for scientific computing in universities from industry and government.

I have benefitted enormously from the high quality and selfless cooperation of researchers at Cornell, in the elementary particle group and in materials research; for my research in the 1960's I was especially indebted to Michael Fisher and Ben Widom.

One other hobby of mine has been playing the oboe but I have not kept this up after 1969.

The home base for my research has been elementary particle theory, and I have made several contributions to this subject: a short distance expansion for operator products presented in an unpublished preprint in 1964 and a published paper in 1969; a discussion of how the renormalization group might apply to strong interactions, in which I discussed all possibilities except the one (asymptotic freedom) now believed to be correct; the formulation of the gauge theory in 1974 (discovered independently by Polyakov), and the discovery that the strong coupling limit of the lattice theory exhibits quark confinement. I am currently interested in trying to solve Quantum Chromodynamics (the theory of quarks) using a combination of renormalization group ideas and computer simulation.

I am also interested in trying to unlock the potential of the renormalization group approach in other areas of classical and modern physics. I have continued to work on statistical mechanics (specifically, the Monte Carlo Renormalization Group, applied to the three dimensional Ising model) as part of this effort.

(added in *1991)* : Wilson became the Director of the Center for Theory and Simulation in Science and Engineering (Cornell Theory Center) - one of five national supercomputer centers created by the National Science Foundation in 1985. In 1988, he moved to The Ohio State University's Department of Physics where he became the Hazel C. Youngberg Trustees Distinguished Professor. He is now heavily engaged in educational reform as a Co-Principal Investigator on Ohio's Project Discovery, one of the National Science Foundation's Statewide Systemic Initiatives.

He was elected to the National Academy of Sciences in 1975, the American Academy of Arts and Sciences in 1975, and the American Philosophical Society in 1984.

THE RENORMALIZATION GROUP AND CRITICAL PHENOMENA

Nobel lecture, 8 December 1982

by

KENNETH G. WILSON

Laboratory of Nuclear Studies, Cornell University,
Ithaca, New York 14853

1. *Introduction*

This paper has three parts. The first part is a simplified presentation of the basic ideas of the renormalization group and the ε expansion applied to critical phenomena, following roughly a summary exposition given in 1972[1]. The second part is an account of the history (as I remember it) of work leading up to the papers in I971-1972 on the renormalization group. Finally, some of the developments since 197 1 will be summarized, and an assessment for the future given.

II. *Many Length Scales and the Renormalization Group*

There are a number of problems in science which have, as a common characteristic, that complex microscopic behavior underlies macroscopic effects.

In simple cases the microscopic fluctuations average out when larger scales are considered, and the averaged quantities satisfy classical continuum equations. Hydrodynamics is a standard example of this where atomic fluctuations average out and the classical hydrodynamic equations emerge. Unfortunately, there is a much more difficult class of problems where fluctuations persist out to macroscopic wavelengths, and fluctuations on all intermediate length scales are important too.

In this last category are the problems of fully developed turbulent fluid flow, critical phenomena, and elementary particle physics. The problem of magnetic impurities in non-magnetic metals (the Kondo problem) turns out also to be in this category.

In fully developed turbulence in the atmosphere, global air circulation becomes unstable, leading to eddies on a scale of thousands of miles. These eddies break down into smaller eddies, which in turn break down, until chaotic motions on all length scales down to millimeters have been excited. On the scale of millimeters, viscosity damps the turbulent fluctuations and no smaller scales are important until atomic scales are reached.[2]

In quantum field theory, “elementary” particles like electrons, photons, protons and neutrons turn out to have composite internal structure on all size scales down to 0. At least this is the prediction ofquantum field theory. It is hard to make observations of this small distance structure directly; instead the particle scattering cross sections that experimentalists measure must be interpreted

using quantum field theory. Without the internal structure that appears in the theory, the predictions of quantum field theory would disagree with the experimental findings.[3]

A critical point is a special example of‘ a phase transition. Consider, for example, the water-steam transition. Suppose the water and steam are placed under pressure, always at the boiling temperature. At the critical point: a pressure of’218 Atm and temperature of 374°C,[4] the distinction between water and steam disappears, and the whole boiling phenomenon vanishes. The principal distinction between water and steam is that they have different densities. As the pressure and temperature approach their critical values, the difference in density between water and steam goes to zero. At the critical point one finds bubbles of steam and drops of water intermixed at all size scales from macroscopic, visible sizes down to atomic scales. Away from the critical point, surface tension makes small drops or bubbles unstable; but as water and steam become indistinguishable at the critical point, the surface tension between the two phases vanishes. In particular, drops and bubbles near micron sizes cause strong light scattering, called “critical opalescence”, and the water and steam become milky.

In the Kondo effect, electrons of all wavelengths from atomic wavelengths up to very much larger scales, all in the conduction band of a metal, interact with the magnetic moment of each impurity in the metal.[5]

Theorists have difficulties with these problems because they involve very many coupled degrees of freedom. It takes many variables to characterize a turbulent flow or the state of a fluid near the critical point. Analytic methods are most effective when functions of only one variable one degree of freedom) are involved. Some extremely clever transformations have enabled special cases of the problems mentioned above to be rewritten in terms of independent degrees of freedom which could be solved analytically. These special examples include Onsager’s solution of the two dimensional Ising model of a critical point, [6]the solution of Andrei and Wiegmann of the Kondo problem,[7] the solution of’ the Thirring model of a quantum field theory,[8] and the simple solutions of noninteracting quantum fields. ‘These are however only special cases; the entire problem of’ fully developed turbulence, many problems in critical phenomena and virtually all examples of strongly coupled quantum fields have defeated analytic techniques up till now.

Computers can extend the capabilities of‘ theorists, but even numerical computer methods are limited in the number of degrees of freedom that are practical. Normal methods of’ numerical integration fail beyond only 5 to 10 integration variables; partial differential equations likewise become extremely difficult beyond 3 or so independent variables. Monte Carlo and statistical averaging methods can treat some cases of’ thousands or even millions of variables but the slow convergence of these methods versus computing time used is a perpetual hassle. An atmospheric flow simulation covering all length scales of turbulence would require a grid with millimeter spacing covering thousands of miles horizontally and tens of miles vertically: the total number of grid points would be of order 10^{25} far beyond the capabilities of any present or conceivable computer.

The "renormalization group" approach is a strategy for dealing with problems involving many length scales. The strategy is to tackle the problem in steps, one step for each length scale. In the case ofcritical phenomena, the problem, technically, is to carry out statistical averages over thermal fluctuations on all size scales. The renormalization group approach is to integrate out the fluctuations in sequence starting with fluctuations on an atomic scale and then moving to successively larger scales until fluctuations on all scales have been averaged out,

To illustrate the renormalization group ideas the case of' critical phenomena will be discussed in more detail. First the mean field theory of Landau will be described, and important questions defined. The renormalization group will be presented as an improvment to Landau's theory.

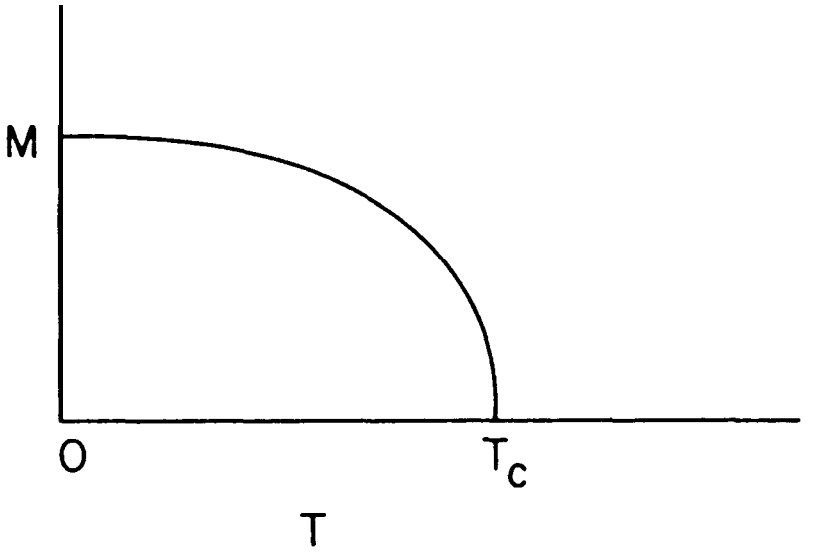
The Curie point of a ferromagnet will be used as a specific example of a critical point. Below the Curie temperature, an ideal ferromagnet exhibits spontaneous magnetization in the absence of' an external magnetic field; the direction of' the magnetization depends on the history of the magnet. Above the Curie temperature T_C, there is no spontaneous magnetization. Figure 1 shows a typical plot of' the spontaneous magnetization versus temperature. Just below the Curie temperature the magnetization is observed to behave as $(T_C - T)^{\beta}$, where β is an exponent somewhere near l/3 (in three dimensions).[9.10]

Magnetism is caused at the atomic level by unpaired electrons with magnetic moments, and in a ferromagnet, a pair of nearby electrons with moments aligned has a lower energy than if the moments are anti-aligned.[10] At high temperatures, thermal fluctuations prevent magnetic order. As the temperature is reduced towards the Curie temperature, alignment of one moment causes preferential alignment out to a considerable distance called the correlation length ξ. At the Curie **temperature, the** correlation length becomes infinite, marking the onset of preferential alignment of the entire system. Just above T_C, the correlation length is found to behave as $(T-T_c)^{-\nu}$, where ν is about 2/3 (in three dimensions).[11]

A simple statistical mechanical model of a ferromagnet involves a Hamiltonian which is a sum over nearest neighbor moment pairs with different energies for the aligned and antialigned case. In the simplest case, the moments are allowed only to be positive or negative along a fixed spatial axis; the resulting model is called the Ising model.[12]

The formal prescription for determining the properties of' this model is to compute the partition function Z, which is the sum of the Boltzmann factor $\exp(-H/kT)$ over all configurations of the magnetic moments, where k is Boltzmann's constant. The free energy F is proportional to the negative logarithm of Z.

The Boltzmann factor $\exp(-H/kT)$ is an analytic function of T near T_c, in fact for all T except T = 0. A sum of analytic functions is also analytic. Thus it is puzzling that magnets (including the Ising model) show complex non-analytic behavior at $T = T_c$. The true non-analytic behavior occurs only in the thermodynamic limit of a ferromagnet ofinfinite size; in this limit there are an infinite number of configurations and there are no analyticity theorems for the infinite sums appearing in this limit. However, it is difficult to understand how even an infinite sum can give highly non-analytic behavior. A major challenge has been to show how the non-analyticity develops.



0830383-001

Landau's proposal[7] was that if only configurations with a given magnetization density M are considered then the free energy is analytic in M. For small M, the form of the free energy (to fourth order in M) is (from the analyticity assumption)

$$F = V\{RM^2 + UM^4\} \tag{1}$$

where V is the volume of the magnet and R and U are temperature-dependent constants. (A constant term independent of M has been omitted). In the absence of an external magnetic field, the free energy cannot depend on the sign of M, hence only even powers of M occur. The true free energy is the minimum of F over all possible values of M. In Landau's theory, R is 0 at the critical temperature, and U must be positive so that the minimum of F occurs at M = 0 when at the critical temperature. The minimum of F continues to be at M = 0 if R is positive: this corresponds to temperatures above critical. If R is negative the minimum occurs for non-zero M, namely the M value satisfying

$$0 = \frac{\partial F}{\partial M} = (2RM + 4UM^3) \tag{2}$$

or

$$M = \sqrt{-R/(2U)} \tag{3}$$

This corresponds to temperatures below critical.

Along with the analyticity of the free energy in M, Landau assumed analyticity in T, namely that R and U are analytic functions of T. Near T_c this means that

to a first approximation, U is a constant and R (which vanishes at T,) is proportional to 'I'-'I', (It is assumed that dR/dT does not vanish at T_c). Then, below 'I', , the magnetization behaves as

$$M \propto (T_c - T)^{1/2} \tag{4}$$

i.e. the exponent β is 1/2 which disagrees with the evidence, experimental and theoretical, that β is about 1/3.[9]

Landau's theory allows for a slowly varying space-dependent magnetization. The free energy for this case takes the Landau-Ginzburg form[14]

$$F = \int d^3x \{[\nabla M(x)]^2 + RM^2(x) + UM^4(x) - B(x)M(x)\} \tag{5}$$

where B(x) is the external magnetic field. The gradient term is the leading term in an expansion involving arbitrarily many gradients as well as arbitrarily high powers of M. For slowly varying fields M(x) higher powers of gradients are small and are neglected. (Normally the $\nabla M^2(x)$ term has a constant coefficient - in this paper this coefficient is arbitrarily set to 1) . One use of this generalized free energy is to compute the correlation length ξ above T_c. For this purpose let B(x) be very small δ function localized at x = 0. The U term in F can be neglected, and the magnetization which minimize the free energy satisfies

$$-\nabla^2 M(x) + RM(x) = B\delta^3(x) \tag{6}$$

The solution M(x) is

$$M(x) \propto Be^{-\sqrt{R}|\vec{x}|}/|\vec{x}| \tag{7}$$

and the correlation length can be read off to be

$$\xi \propto 1/\sqrt{R} \tag{8}$$

Hence near T_c. ξ is predicted to behave as $(T - T_c)^{-1/2}$, which again disagrees with experimental and theoretical evidence.

The Landau theory assumes implicitly that analyticity is maintained as all space-dependent fluctuations are averaged out. The loss of analyticity arises only when averaging over the values of the overall average magnetization M. It is this overall averaging, over $e^{-F/kT}$, which leads to the rule that F must be minimized over M, and the subsequent non-analytic formula (4) for M. To be precise, if the volume of the magnet is finite, $e^{-F/kT}$ must be integrated over M, with analytic results. It is only in the thermodynamic limit $V \to \infty$ that the average of $e^{-F/kT}$ is constructed by minimizing F with respect to M. and the nonanalyticity of Eqn. (4) occurs.

'The Landau theory has the same physical motivation as hydrodynamics. Landau assumes that only fluctuations on an atomic scale matter. Once these have been averaged out the magnetization M(x) becomes a continuum, continous function which fluctuates only in response to external space-dependent stimuli. M(x) (or, if it is a constant, M) is then determined by a simple classical equation. Near the critical point the correlation function is itself the solution of the classical equation (6).

In a world with greater than four dimensions, the Landau picture is correct." Four dimensions is the dividing line - below four dimensions, fluctuations on all scales up to the correlation length are important and Landau theory breaks down,[16] as will be shown below. An earlier criterion by Ginzburg[15] also would predict that four dimensions is the dividing line.

The role of long wavelength fluctuations is very much easier to work out near four dimensions where their effects are small. This is the only case that will be discussed here. Only the effects ofwavelengths long compared to atomic scales will be discussed, and it will be assumed that only modest corrections to the Landau theory are required. For a more careful discussion see ref. 17.

Once the atomic scale fluctuations have been averaged out, the magnetization is a function M(x) on a continuum, as in Landau theory. However, long wavelength fluctuations are still present in M(x) - they have not been averaged out - and the allowed forms of 14(x) must be stated with care. To be precise, suppose fluctuations with wavelengths $< 2\pi L$ have been averaged out, where L is a length somewhat larger than atomic dimensions. Then M(x) can contain only Fourier modes with wavelengths $> 2\pi L$. This requirement written out, means

$$M(\vec{x}) = \int_{\vec{k}} e^{i\vec{k}\cdot\vec{x}} M_{\vec{k}} \tag{9}$$

where the integral over $\vec{k}$ means $(2\pi)^{-d}\int d^d k$, d is the number of space dimensions, and the limit on wavelengths means that the integration over E is restricted to values of $\vec{k}$ with $|\vec{k}| < L^{-1}$.

Averaging over long wavelength fluctuations now reduces to integrating over the variables $M_{\vec{k}}$, for all $|\vec{k}| < L^{-1}$. There are many such variables; normally this would lead to many coupled integrals to carry out, a hopeless task. Considerable simplifications will be made below in order to carry out these integrations.

We need an integrand for these integrations. The integrand is a constrained sum of the Boltzmann factor over all atomic configurations. The constraints an that all $M_{\vec{k}}$ for $|\vec{k}| < L^{-1}$ are held fixed. This is a generalization of the constrained sum in the Landau theory; the difference is that in the Landau theory only the average magnetization is held fixed. The result of the constrained sum will be written e^{-F}, similarly to Landau theory except for convenience the exponent is written F rather than F/kT (i.e. the factor 1 /kT is absorbed into an unconventional definition of F). The exponent F depends on the magnetization function M(x) of Eq. (9). We shall assume Landau's analysis is still valid for the form of F, namely F is given by Eq. (5). However, the importance of long wavelength fluctations means that the parameters R and U depend on L. Thus F should be denoted F_L:

$$F_L = \int_I d^d x \{(\nabla M)^2(x) + R_L M^2(x) + U_L M^4(x)\} \tag{10}$$

(in the absence of any external field) (in the simplified analysis presented here, the coefficient of $VM^2(x)$ is unchanged at 1). The assumption will be reviewed later.

The L dependence of R_L and U_L will be determined shortly. However, the breakdown of analyticity at the critical point is a simple consequence of this L dependence. The L dependence persists only out to the correlation length ξ: fluctuations with wavelengths $>\xi$ will be seen to be always negligible. Once all wavelengths of fluctuations out to $L \sim \xi$ have been integrated out, one can use the Landau theory; this means (roughly speaking) substituting R_ξ and U_ξ in the formulae (4) and (8) for the spontaneous magnetization and the correlation length. Since ξ is itself non-analytic in T at T = T, the dependence of R_ξ and U_ξ on ξ introduces new complexities at the critical point. Details will be discussed shortly.

In order to study the effects of fluctuations, only a single wavelength scale will be considered; this is the basic step in the renormalization group method. To be precise, consider only fluctuations with wavelengths lying in an infinitesimal interval L to $L+\delta L$. To average over these wavelengths of fluctuations one starts with the Boltzmann factor e^{-F_L} where the wavelengths between L and $L+\delta L$ are still present in M(x), and then averages over fluctuations in M(x) with wavelengths between L and $L+\delta L$. The result of these fluctuation averages is a free energy F_L+bl. for a magnetization function (which will be denoted $M_H(x)$) with wavelengths > $L+\delta L$ only. The Fourier components of $M_H(x)$ are the same $\vec{k}$ that appear in M(x) except that $|\vec{k}|$ is now restricted to be less than $1/(L+\delta L)$.

The next step is to count the number of integration variables $M_{\vec{k}}$ with $|\vec{k}|$ lying between $1/L$ and $1/(L+\delta L)$. To make this count it is necessary to consider a finite system in a volume V. Then the number of degrees of freedom with wavelengths between $2\pi L$ and $2\pi(L+\delta L)$ is given by the corresponding phase space volume, namely the product of k space and position space volumes. This product is (apart from constant factors like π, etc.) $L^{-(d+1)}V\delta L$.

It is convenient to choose the integration variables not to be the $M_{\vec{k}}$ themselves but linear combinations which correspond to localized wave packets instead of plane waves. That is, the difference $M_H(x)-M(x)$ should be expanded in a set of wave packet functions $\psi_n(x)$, each of which has momenta only in the range $1/L$ to $1/(L+\delta L)$, but which is localized in x space as much as possible. Since each function $\psi_n(x)$ must (by the uncertainty principle) fill unit volume in phase space, the position space volume for each $\psi_n(x)$ is

$$\delta V = L^{d+1}/\delta L \tag{11}$$

and there are $V/\delta V$ wavefunctions $\psi_n(x)$. We can write

$$M(x) = M_H(x)+\Sigma m_n \psi_n(x) \tag{12}$$

and the integrations to be performed are integrations over the coefficients m_n.

Because of the local nature of the Landau-Ginzburg free energy, it will be assumed that the overlap of the different wavefunctions ψ_n can be neglected. Then each m_n integration can be treated separately, and only a single such integration will be discussed here. For this single integration, the form of M(x) can be written

$$M(x) = M_H(x)+m\psi(x) \tag{13}$$

since only one term from the sum over n contributes within the spatial volume occupied by the wavefunction $\psi(x)$.

The other simplification that will be made is to treat $M_H(x)$ as if it were a constant over the volume occupied by $\psi(x)$. In other words the very long wavelengths in $M_H(x)$ are emphasized relative to wavelengths close to L.

The calculation to be performed is to compute

$$e^{-F_{L+\delta L}[M_H]} = \int_I^\infty dm\, e^{-F_L[M_H + m\psi]} \tag{14}$$

where $F_{L+\delta L}$ and F_L, involve integration only over the volume occupied by $\psi(x)$. In expanding out $F_L[M_H+m\psi]$ the following simplifications will be made. First, all terms linear in $\psi(x)$ are presumed to integrate to 0 in the x integration defining F_L. Terms of third order and higher in ψ are also neglected. The function $\psi(x)$ is presumed to be normalized so that

$$\int d^d x\, \psi^2(x) = 1 \tag{15}$$

and due to the limited range ofwavelengths in $\psi(x)$, there results

$$\int [\nabla\psi(x)]^2 d^d x \simeq 1/L^2 \tag{16}$$

The result of these simplifications is that the integral becomes

$$e^{-F_{L+\delta L}[M_H]} = e^{-F_L[M_H]} \int_{-\infty}^{\infty} dm \exp\{(R_L + \tfrac{1}{r^2})\, m^2 + 6U_L M_H^2 m^2\} \tag{17}$$

or

$$F_{L+\delta L}[M_H] = F_L\{M_H\} + \frac{1}{2}\ell n\left(\frac{1}{L^2} + R_L + 6U_L M_H^2\right) \tag{18}$$

The logarithm must be rewritten as an integral over the volume occupied by $\psi(x)$; this integral can then be extended to an integral over the entire volume V when the contributions from all other m_n integrations are included. Also the logarithm must be expanded in powers of M_H; only the M_H^2 and M_H^4 terms will be kept. Further it will be assumed that R_L, changes slowly with L. When L is at the correlation length ξ, $1/L^2$ and R_L, are equal (as already argued) so that for values of L intermediate between atomic sizes and the correlation length, R_L, is small compared to $1/L^2$. Expanding the logarithm in powers of R_L+GUt,Mi, to second order (to obtain an M_H^4 term) gives (of. Eq. (11)):

$$\begin{aligned}\frac{1}{2}\ell n\left(\frac{1}{L^2} + R_L + 6U_L M_H^2\right) &= \text{terms independent of } M_H \\ &+ (\delta V)(\delta L) L^{-d-1}\{3U_L M_H^2 L^2 - 9U_L^2 M_H^4 L^4 - 3R_L U_L M_H^2 L^4\}\end{aligned} \tag{19}$$

One can rewrite δV as an integral over the volume δV. There results the equations

$$R_{L+\delta L} = R_L + (3U_L L^{1-d} - 3R_L U_L L^{3-d})\delta L \tag{20}$$

$$U_{L+\delta L} = U_L - 9U_L^2 L^{3-d}\delta L \tag{21}$$

or

$$L\frac{dR_L}{dL} = 3L^{2-d}U_L - 3R_L U_L \cdot L^{4-d} \tag{22}$$

$$L\frac{dU_L}{dL} = -9U_L^2 L^{4-d} \tag{23}$$

These equations are valid only for $L < \xi$; for $L > \xi$ there is very little further change in R_L, or U_L, due to the switchover in the logarithm caused by the dominance of R_L rather than $1/L^2$. If d is greater than 4, it can be seen that R_L and U_L, are constant for large L, as expected in the Landau theory. For example, if one assumes R_L. and U_L. are constant for large L it is easily seen that integration of (22) and (23) only gives negative powers of L. For d <4 the solutions are not constant. Instead, U_L behaves for sufficiently large L as

$$U_{L} \simeq \frac{(4-d)}{9} L^{d-4} \tag{24}$$

(which is easily seen to be a solution of (23)), R_L. satisfies the equation

$$\frac{dR_L}{dL} + \frac{(4-d)}{3L} R_L = \frac{(4-d)}{3} L^{-3} \tag{25}$$

whose solution is

$$R_L = cL^{(d-4)/3} - \frac{(4-d)}{3} \frac{1}{2-(4-d)/3} L^{-2} \tag{26}$$

where c is related to the value of R_L. at some initial value of L. For large enough L, the L^{-2} term can be neglected.

The parameter c should be analytic in temperature, in fact proportional to T-T,. Hence, for large L

$$R_{L} \propto L^{(d-4)/3}(T-T_c) \tag{27}$$

which is analytic in T for fixed L. However the equation for ξ is

$$\xi \propto R_\xi^{-1/2} = (T-T_c)^{-1/2} \xi^{(4-d)/6} \tag{28}$$

Let

$$\varepsilon = 4\text{-}d \tag{29}$$

then the correlation length exponent is

$$\upsilon = \frac{1}{2}\frac{1}{1-\varepsilon/6} \tag{30}$$

which gives $\nu = 0.6$ in 3 dimensions. Similarly, the spontaneous magnetization below T_c behaves as $(R_\xi/U_\xi)^{1/2}$ giving

$$\beta = \frac{1}{2} - \frac{\varepsilon}{3}\frac{1}{1-\varepsilon/6} \tag{31}$$

These computations give an indication of how non-trivial values can be obtained for β and ν. *The* formulae derived here are not exact, due to the severe simplifications made, but at least they show that β and ν do not have to be 1/2 and in fact can have a complicated dependence on the dimension d.

A correct treatment is much more complex. Once $M_n(x)$ is not treated as a

constant, one could imagine expanding $M_H(x)$ in a Taylor's series about its value at some central location x_o relative to the location of the wavefunction $\psi(x)$, thus bringing in gradients of M_H. In addition, higher order terms in the expansion of the logarithm give higher powers of M_H. All this leads to a more complex form for the free energy functional F_L. with more gradient terms and more powers of M_H. The whole idea of the expansion in powers of M_H and powers of gradients can in fact be called into question. The fluctuations have an intrinsic size (i.e., m^2 has a size $\sim L^2$ as a consequence of the form of the integrand in Eq. 17) and it is not obvious that in the presence of these fluctuations, M is small. Since arbitrary wavelengths of fluctuations are important the function M is not sufficiently slowly varying tojustify an expansion in gradients either. This means that $F_L[M]$ could be an arbitrarily complicated function of M, an expression it is hard to write down, with thousands of parameters, instead of the simple Landau-Ginzburg form with only two parameters R_L, and U_L.

Fortunately, the problem simplifies near 4 dimensions, due to the small magnitude of U_L, which is proportional to $\varepsilon = 4\text{-}d$. All the complications neglected above arise only to second order or higher in an expansion in U_L which means second order or higher in ε. The computations described here are exact to order ε. See Ref. 17.

The renormalization group approach that was defined in 1971 embraces both practical approximations leading to actual computations and a formalism.[17] The full formalism cannot be discussed here but the central idea of "fixed points" can be illustrated.

As the fluctuations on each length scale are integrated out a new free energy functional $F_{L+\delta L}$ is generated from the previous functional F_L. This process is repeated many times. If F_L and $F_{L+\delta L}$ are expressed in dimensionless form, then one finds that the transformation leading from F_L to $F_{L+\delta L}$ is repeated in identical form many times. (The transformation group thus generated is called the "renormalization group"). As L becomes large the free energy F_L approaches a fixed point of the transformation, and thereby becomes independent of details of the system at the atomic level. This leads to an explanation of the universality[18],, of critical behavior for different kinds of systems at the atomic level. Liquid-gas transitions, magnetic transitions, alloy transitions, etc. all show the same critical exponents experimentally; theoretically this can be understood from the hypothesis that the same "fixed point" interaction describes all these systems.

To demonstrate the fixed point form of the free energy functional, it must be put into dimensionless form. Lengths need to be expressed in units of L, and M, R_L, and U_L rewritten in dimensionless form. These changes are easily determined: write

$$x = Ly \tag{32}$$

$$M(x) = L^{1-d/2}m(y) \tag{33}$$

$$R_L = 1/L^2 r_L \tag{34}$$

$$U_L = L^{d-4}u_L \tag{35}$$

$$F_L = \int d^d y\{(\nabla m)^2 + r_L m^2(y) + u_L m^4(y)\} \tag{36}$$

The asymptotic solution for the dimensionless parameters r_L, and u_L. is

$$r_L = cL^{2-\epsilon/3} - \frac{\epsilon}{3}\frac{1}{2-\epsilon/3} \tag{37}$$

$$u_L = \frac{\epsilon}{9} \tag{38}$$

Apart from the term in r_L, these dimensionless parameters are independent of L, denoting a free energy form which is also independent of L. The c term designates an instability of the fixed point, namely a departure from the fixed point which grows as L increases. The fixed point is reached only if the thermodynamic system is at the critical temperature for which c vanishes; any departure from the critical temperature triggers the instability.

For further analysis of the renormalization group formalism and its relation to general ideas about critical behavior, see e.g. ref. 17.

II I. *Some History Prior to 1971*

The first description of a critical point was the description of the liquid-vapor critical point developed by Van der Waals, [19] developed over a century ago following experiments of Andrews.[19] Then Weiss provided a description of the Curie point in a magnet."' Both the Van der Waals and Weiss theories were special cases of Landau's mean field theory." Even before 1900, experiments indicated discrepancies with mean Geld theory; in particular the experiments indicated that β was closer to $1/3$ than $1/2$.[19] In 1944, Onsager[6] published his famous solution to the two dimensional Ising model,[12] which explicity violated the mean field predictions. Onsager obtained $\nu = 1$ instead of the mean field prediction $\nu = 1/2$, for example. In the 1950's, Domb, Sykes, Fisher and others[21] studied simple models of critical phenomena in three dimensions with the help of high temperature series expansions carried to very high order, exacting critical point exponents by various extrapolation methods. They obtained exponents in disagreement with mean field theory but in reasonable agreement with experiment. Throughout the sixties a major experimental effort pinned down critical exponents and more generally provided a solid experimental basis for theoretical studies going beyond mean field theory. Experimentalists such as Voronel; Fairbanks, Buckingham, and Keller; Heller and Benedek; Ho and Litster, Kouvel and Rodbell, and Comly; Sengers; Lorentzen; Als-Nielsen and Dietrich; Birgeneau and Shirane; Rice; Chu; Teaney; Moldover; Wolf and Ahlers all contributed to this development, with M. Green, Fisher, Widom, and Kadanoff providing major coordination efforts." Theoretically, Widom[23] proposed a scaling law for the equation of state near the critical point that accommodated non-mean field exponents and predicted relations among them. The full set of scaling hypotheses were developed by Essam and Fisher, Domb and Hunter, Kadanoff, and Patashinskii, and Pokrovskii.[24] See also the inequalities of Rushbrooke[25] and Griffith.[26]

My own work began in quantum field theory, not statistical mechanics. A convenient starting point is the development ofrenormalization theory by Bethe, Schwinger, Tomonaga, Feynman, Dyson and others[27] in the late 1940's. The

first discussion of the "renormalization" group appeared in a paper by Stueckelberg and Petermann,[28] published in 1953.

In 1954 Murray Gell-Mann and Francis Low published a paper entitled "Quantum Electrodynamics at Small Distances"[29] which was the principal inspiration for my own work prior to Kadanoffs formulation[30] of the scaling hypothesis for critical phenomena in 1966.

Following the definition of Quantum Electrodynamics (QED) in the 1930's by Dirac, Fermi, Heisenberg, Pauli, Jordan, Wigner, et al.[27], the solution of QED was worked out as perturbation series in e_o, the "bare charge" of QED. The QED Lagrangian (or Hamiltonian) contains two parameters: e_o, and m_o, the latter being the "bare" mass of the electron. As stated in the introduction in QED the physical electron and photon have composite structure. In consequence of this structure the measured electric charge e and electron mass m are not identical to e_o and m,, but rather are given by perturbation expansions in powers of e_o. Only in lowest order does one find $e = e_o$, and $m = m_o$. Unfortunately, it was found in the 30's that higher order corrections in the series for e and m are all infinite, due to integrations over momentum that diverge in the large momentum (or small distance) limit."

In the late 1940's renormalization theory was developed, which showed that the divergences of Quantum Electrodynamics could all be eliminated if a change of parametrization was made from the Lagrangian parameters e_o and m_o to the measurable quantities e and m, and if at the same time the electron and electromagnetic fields appearing in the Lagrangian were rescaled to insure that observable matrix elements (especially of the electromagnetic field) are finite.[27]

There are many reparametrizations of Quantum Electrodynamics that eliminate the divergences but use different finite quantities than e and m to replace e_o and m_o. Stueckelberg and Petermann observed that transformation groups could be defined which relate different reparametrizations - they called these groups "groupes de normalization" which is translated "renormalization group". The Gell-Mann and Low paper,[29] one year later but independently, presented a much deeper study of the significance of the ambiguity in the choice of reparametrization and the renormalization group connecting the difference choices of reparametrization. Gell-Mann and Low emphasized that e, measured in classical experiments, is a property of the very long distance behavior of QED (for example it can be measured using pith balls separated by centimeters, whereas the natural scale of QED is the Gompton wavelength of the electron, $\sim 10^{-11}$ cm). Gell-Mann and Low showed that a family ofalternative parameters e_λ could be introduced, any one of which could be used in place of e to replace e_o. The parameter e_λ is related to the behavior of QED at an arbitrary momentum scale λ instead of at very low momenta for which e is appropriate.

The family of parameters e_λ introduced by Gell-Mann and Low interpolate between the physical charge e and the bare charge e , namely e is obtained as the low momentum $(\lambda \to 0)$ limit of e_λ and e_o is obtained as the high momentum $(\lambda \to \infty)$ limit of e_λ.

Gell-Mann and Low found that e_λ^2 obeys a differentia1 equation, of the form

$$\lambda^2 d(e_\lambda^2)/d(\lambda^2) = \psi(e_\lambda^2, m^2/\lambda^2) \quad \textbf{(39)}$$

where the φ function has a simple power series expansion with non-divergent coefficients independently of the value of λ, in fact as $\lambda \to \infty$ φ becomes a function of alone. This equation is the forerunner of my own renormalization group equations such as (22) and (23).

The main observation of Gell-Mann and Low was that despite the ordinary nature of the differential equation, Eq. (38), the solution was not ordinary, and in fact predicts that the physical charge e has divergences when expanded in powers of e_o, or vice versa. More generally, if e_λ is expanded in powers of $e_{\lambda'}$, the higher order coefficients contain powers of $\ell n(\lambda^2/\lambda'^2)$, and these coefficients diverge if either λ or λ' go to infinity, and are very large if λ^2/λ'^2 is either very large or very small.

Furthermore, Gell-Mann and Low argued that, as a consequence of Eqn. (38), e_o, must have a fixed value independently of the value of e; the fixed value of e_o could be either finite or infinite.

When I entered graduate school at California Institute of Technology, in 1956, the default for the most promising students was to enter elementary particle theory, the field in which Murray Gell-Mann, Richard Feynman, and Jon Mathews were all engaged. I rebelled briefly against this default, spending a summer at the General Atomic Corp. working for Marshall Rosenbluth on plasma physics and talking with S. Chandresekhar who was also at General Atomic for the summer. After about a month of work I was ordered to write up my results, as a result of which I swore to myself that I would choose a subject for research where it would take at least five years before I had anything worth writing about. Elementary particle theory seemed to offer the best prospects of meeting this criterion and I asked Murray for a problem to work on. He first suggested a topic in weak interactions of strongly interacting particles (K mesons, etc.) After a few months I got disgusted with trying to circumvent totally unknown consequences of strong interactions, and asked Murray to find me a problem dealing with strong interactions directly, since they seemed to be the bottleneck. Murray suggested I study K meson-nucleon scattering using the Low equation in the one meson approximation. I wasn't very impressed with the methods then in use to solve the Low equation, so I wound up fiddling with various methods to solve the simpler case of pion-nucleon scattering. Despite the fact that the one meson approximation was valid, if at all, only for low energies, I studied the high energy limit, and found that I could perform a "leading logarithms" sum very reminiscent of a very mysterious chapter in Bogoliubov and Shirkov's field theory text[31]; the chapter was on the renormalization group.

In 1960 I turned in a thesis to Cal Tech containing a mish-mash of curious calculations. I was already a Junior Fellow at Harvard. In 1962 I went to CERN for a year. During this period (1960-1963) I partly followed the fashions of the time. Fixed source meson theory (the basis for the Low equation) died, to be replaced by S matrix theory. I reinvented the "strip approximation" (Ter-Martirosyan had invented it first[32]) and studied the Amati-Fubini-Stanghellini

theory of multiple production.[33] I was attentive at seminars (the only period of my life when I was willing to stay fully awake in them) and I also pursued back waters such as the strong coupling approximation to fixed source meson theory?

By 1963 it was clear that the only subject I wanted to pursue was quantum field theory applied to strong interactions. I rejected S matrix theory because the equations of S matrix theory, even if one could write them down, were too complicated and inelegant to be a theory; in contrast the existence of a strong coupling approximation as well as a weak coupling approximation to fixed source meson theory helped me believe that quantum field theory might make sense. As far as strong interactions were concerned, all that one could say was that the theories one could write down, such as pseudoscalar meson theory, were obviously wrong. No one had any idea of a theory that could be correct. One could make these statements even though no one had the foggiest notion how to solve these theories in the strong coupling domain.

My very strong desire to work in quantum field did not seem likely to lead to quick publications; but I had already found out that I seemed to be able to get jobs even if I didn't publish anything so I did not worry about 'publish or perish' questions.

There was very little I could do in quantum field theory - there were very few people working in the subject, very few problems open for study. In the period 1963-1966 I had to clutch at straws. I thought about the "ξ-limiting" process of Lee and Yang." I spent a major effort disproving Ken Johnson's claims"' that he could define quantum electrodynamics for arbitrarily small e_o, in total contradiction to the result of Gell-Mann and Low. I listened to K. Hepp and others describe their results in axiomatic field theory[37]; I didn't understand what they said in detail but I got the message that I should think in position space rather than momentum space. I translated some of the work I had done on Feynman diagrams with some very large momenta (to disprove Ken Johnson's ideas) into position space and arrived at a short distance expansion for products of quantum field operators. I described a set of rules for this expansion in a preprint in 1964. I submitted the paper for publication; the referee suggested that the solution of the Thirring model might illustrate this expansion. Unfortunately, when I checked out the Thirring model, I found that while indeed there was a short distance expansion for the Thirring model,[38] my rules for how the coefficient functions behaved were all wrong, in the strong coupling domain. I put the preprint aside, awaiting resolution of the problem.

Having learned the fixed source meson theory as a graduate student, I continued to think about it. I applied my analysis of Feynman diagrams for some large momenta, to the fixed source model. I realized that the results I was getting became much clearer if I made a simplification of the fixed source model itself, in which the momentum space continuum was replaced by momentum slices.[39] That is, I rubbed out all momenta except well separated slices, e.g., $1 \leqslant |\mathbf{k}| \leqslant 2$, $\Lambda \leqslant |\mathbf{k}| \leqslant 2\Lambda, \Lambda^2 \leqslant |\mathbf{k}| \leqslant 2\Lambda^2, , \Lambda^n \leqslant |\mathbf{k}| \leqslant 2\Lambda^n$, etc. with A a large number.

This model could be solved by a perturbation theory very different from the methods previously used in field theory. The energy scales for each slice were

very different, namely of order L^n for the n^{th} slice. Hence the natural procedure was to treat the Hamiltonian for the largest momentum slice as the unperturbed Hamiltonian, and the terms for all lesser slices as the perturbation. In each slice the Hamiltonian contained both a free meson energy term and an interaction term, so this new perturbation method was neither a weak coupling nor a strong coupling perturbation.

I showed that the effect of this perturbation approach was that if one started with n momentum slices, and selected the ground state of the unperturbed Hamiltonian for the n^{th} slice, one wound up with an effective Hamiltonian for the remaining n-l slices. This new Hamiltonian was identical to the original Hamiltonian with only n-l slices kept, except that the meson-nucleon coupling constant g was renormalized (i.e., modified): the modification was a factor involving a non-trivial matrix element of the ground state of the n^{th}-slice Hamiltonian. [39]

This work was a real breakthrough for me. For the first time I had found a natural basis for renormalization group analysis: namely the solution and elimination of one momentum scale from the problem. There was still much to be done: but I was no longer grasping at straws. My ideas about renormalization were now reminiscent of Dyson's analysis of Quantum Electrodynamics."' Dyson argued that renormalization in Quantum Electrodynamics should be carried out by solving and eliminating high energies before solving low energies. I studied Dyson's papers carefully but was unable to make much use of his work."

Following this development, I thought very hard about the question "what is a field theory", using the ϕ^+ interaction of a scalar field (identical with the Landau-Ginzburg model of a critical point[14] discussed in my 1971 papers) as an example. Thoughout the '60's I taught quantum mechanics frequently, and I was very impressed by one's ability to understand simple quantum mechanical systems. The first step is a qualitative analysis minimizing the energy (defined by the Hamiltonian) using the uncertainty principle; the second step might be a variational calculation with wavefunctions constructed using the qualitative information from the first step; the final stage (for high accuracy) would be a numerical computation with a computer helping to achieve high precision. I felt that one ought to be able to understand a field theory the same way.

I realized that I had to think about the degrees of freedom that make up a field theory. The problem of solving the ϕ^4 theory was that kinetic term in the Hamiltonian (involving $(\nabla\phi)^2$) was diagonal only in terms of the Fourier components ϕ_k of the field, whereas the ϕ^+ term was diagonal only in terms of the field $\phi(x)$ itself. Therefore I looked for a compromise representation in which both the kinetic term and the interaction term would be at least roughly diagonal. I needed to expand the field $\phi(x)$ in terms of wavefunctions that would have minimum extent in both position space and momentum space, in other words wavefunctions occupying the minimum amount of volume in phase space. The uncertainty principle defines the lower bound for this volume, namely 1, in suitable units. I thought of phase space being divided up into blocks of unit volume. The momentum slice analysis indicated that momentum space should be marked off on a logarithmic scale, i.e. each momentum space volume should

correspond to a shell like the slices defined earlier, except that I couldn't leave out any momentum range so the shells had to be e.g...., $1 < kl < 2$, $2 < |k| < 4$, etc. By translational invariance the position space blocks would all be the same size for a given momentum shell, and would define a simple lattice of blocks. The position space blocks would have different sizes for different momentum shells.

When I tried to study this Hamiltonian I didn't get very far. It was clear that the low momentum terms should be a perturbation relative to the high momentum terms but the details of the perturbative treatment became too complicated. Also my analysis was too crude to identify the physics of highly relativistic particles which should be contained in the Hamiltonian of the field theory."

However, I learned from this picture of the Hamiltonian that the Hamiltonian would have to be cutoff at some large but finite value of momentum k in order to make any sense out of it, and that once it was cutoff, I basically had a lattice theory to deal with, the lattice corresponding roughly to the position space blocks for the largest momentum scale. More precisely, the sensible procedure for defining the lattice theory was to define phase space cells covering all of the cutoff momentum space, in which case there would be a single set of position space blocks, which in turn defined a position space lattice on which the field ϕ would be defined. I saw from this that to understand quantum field theories I would have to understand quantum field theories on a lattice.

In thinking and trying out ideas about "what is a field theory" I found it very helpful to demand that a correctly formulated field theory should be soluble by computer, the same way an ordinary differential equation can be solved on a computer, namely with arbitrary accuracy in return for sufficient computing power. It was clear, in the '60's, that no such computing power was available in practice; all that I was able to actually carry out were some simple exercises involving free fields on a finite lattice.

In the summer of 1966 I spent a long time at Aspen. While there I carried out a promise I had made to myself while a graduate student, namely I worked through Onsager's solution of the two dimensional Ising model. I read it in translation, studying the field theoretic form given in Lieb, Mattis and Schultz.'"

When I entered graduate school, I had carried out the instructions given to me by my father and had knocked on both Murray Gell-Mann's and Feynman's doors, and asked them what they were currently doing. Murray wrote down the partition function for the three dimensional Ising model and said it would be nice if I could solve it (at least that is how I remember the conversation). Feynman's answer was "nothing". Later, Jon Mathews explained some of Feynman's tricks for reproducing the solution for the two dimensional Ising model. I didn't follow what Jon was saying, but that was when I made my promise. Sometime before going to Aspen, I was present when Ben Widom presented his scaling equation of state,[23] in a seminar at Cornell. I was puzzled by the absence of any theoretical basis for the form Widom wrote down; I was at that time completely ignorant of the background in critical phenomena that made Widom's work an important development.

As I worked through the paper of Mattis, Lieb, and Schultz, I realized there

should be applications of my renormalization group ideas to critical phenomena, and discussed this with some of the solid state physicists also at Aspen. I was informed that I had been scooped by Leo Kadanoff and should look at his preprint.[30]

Kadanoff's idea was that near the critial point one could think of blocks of magnetic moments, for example containing 2x2x2 atoms per block, which would act like a single effective moment, and these effective moments would have a simple nearest neighbor interaction like simple models of the original system. The only change would be that the system would have an effective temperature and external magnetic field that might be distinct from the original. More generally the effective moments would exist on a lattice of arbitrary spacing L times the original atomic spacing; Kadanoffs idea was that there would be L-dependent temperature and field variables T_L. and h_L, and that T_{2L}, and h_{2L}. would be analytic functions of T_L. and h_L. At the critical point, T_L, and hi, would have fixed values independent of L. From this hypothesis Kadanoff was able to derive the scaling laws of Widom,[23] Fisher, etc.[24]

I now amalgamated my thinking about field theories on a lattice and critical phenomena. I learned about Euclidean (imaginary time) quantum field theory and the "transfer matrix" method for statistical mechanical models and found there was a close analogy between the two (see Ref. 17). I learned that for a field theory to be relativistic, the corresponding statistical mechanical theory had to have a large correlation length, i.e., be near a critical point. I studied Schiff's strong coupling approximation to the ϕ^4theory,[44] and found that he had ignored renormalization effects; when these were taken into account the strong coupling expansion was no longer so easy as he claimed. I thought about the implications of the scaling theory of Kadanoff, Widom et al. applied to quantum field theory, along with the scale invariance of the solution of the Thirring model" and the discussion of Kastrup and Mack of scale invariance in quantum field theory.[45] These ideas suggested that scale invariance would apply, at least at short distances, but that field operators would have non-trivial scale dimensions corresponding to the non-trivial exponents in critical phenomena. I redid my theory of short distance expansions based on these scaling ideas and published the result."' My theory did not seem to lit the main experimental ideas about short distance behavior (coming from Bjorken's and Feynman's analysis" of deep inelastic electron scattering) but I only felt confused about this problem and did not worry about it.

I returned to the fixed source theory and the momentum slice approximation. I made further simplifications on the model. Then I did the perturbative analysis more carefully. Since in real life the momentum slice separation factor A would be 2 instead of very large, the ratio 1 /A of successive energy scales would be l/2 rather than very small, and an all orders perturbative treatment was required in 1 /A. When the lower energy scales were treated to all orders relative to the highest energy scale, an infinitely complicated effective Hamiltonian was generated, with an infinite set of coupling constants. Each time an energy scale was eliminated through a perturbative treatment, a new infinitely complicated Hamiltonian was generated. Nevertheless, I found that for sufficiently large Λ I could

mathematically control rigorously the effective Hamiltonians that were generated; despite the infinite number of couplings I was able to prove that the higher orders of perturbation theory had only a small and boundable impact on the effective Hamiltonians, even after arbitrarily many iterations.[48]

This work showed me that a renormalization group transformation, whose purpose was to eliminate an energy scale or a length scale or whatever from a problem, could produce an effective interaction with arbitrarily many coupling constants, without being a disaster. The renormalization group formalism based on fixed points could still be correct, and furthermore one could hope that only a small finite number of these couplings would be important for the qualitative behavior of the transformations, with the remaining couplings being important only for quantitative computations. In other words the couplings should have an order of importance, and for any desired but given degree of accuracy only a finite subset of the couplings would be needed. In my model the order of importance was determined by orders in the expansion in powers of $1/\Lambda$. I realized however that in the framework of an interaction on a lattice, especially for Ising-type models, locality would provide a natural order of importance - in any finite lattice volume there are only a finite number of Ising spin interactions that can be defined. I decided that Kadanoffs emphasis on the nearest neighbor coupling of the Ising model[30] should be restated: the nearest neighbor coupling would be the most important coupling because it is the most localized coupling one can define, but other couplings would be present also in Kadanoff's effective "block spin" Hamiltonians. A reasonable truncation procedure on these couplings would be to consider a finite region, say 3^3 or 4^3 lattice sites in size, and consider only multispin couplings that could fit into these regions (plus translations and rotations of these couplings).

Previously all the renormalization group transformations I was familiar with involved a fixed number of couplings: in the Gell-Mann-Low case just the electric charge e_λ, in Kadanoffs case an effective temperature and external field. I had tried many ways to try to derive transformationsjust for these fixed number ofcouplings, without success. Liberated from this restriction, it turned out to be easy to define renormalization group transformations; the hard problem was to find approximations to these transformations which would be computable in practice. Indeed a number renormalization group transformations now exist (see Section IV and its references).

In the fall of 1970 Ben Widom asked me to address his statistical mechanics seminar on the renormalization group. He was particularly interested because Di Castro and Jona-Lisinio had proposed applying the field theoretic renormalization group formalism to critical phenomena,[49] but no one in Widom's group could understand Di Castro and Jona-Lasinio's paper. In the course oflecturing on the general ideas of fixed points and the like I realized I would have to provide a computable example, even if it was not accurate or reliable. I applied the phase space cell analysis to the Landau-Ginzburg model of the critical point and tried to simplify it to the point of a calculable equation, making no demands for accuracy but simply trying to preserve the essence of the phase space cell picture. The result was a recursion formula in the form of a nonlinear integral trans-

formation on a function of one variable, which I was able to solve by iterating the transformation on a computer.[50] I was able to compute numbers for exponents from the recursion formula at the same time that I could show (at least in part) that it had a fixed point and that the scaling theory of critical phenomena of Widom et al. followed from the fixed point formalism. Two papers of 1971 on the renormalization group presented this work.[50]

Some months later I was showing Michael Fisher some numerical results from the recursion formula, when we realized, together, that the nontrivial fixed point I was studying became trivial at four dimensions and ought to be easy to study in the vicinity of four dimensions. The dimension d appeared in a simple way as a parameter in the recursion formula and working out the details was straightforward; Michael and I published a letter[51] with the results. It was almost immediately evident that the same analysis could be applied to the full Landau-Ginzberg model without the approximations that went into the recursion formula. Since the simplifying principle was the presence of a small coefficient of the ϕ^+ term, a Feynman diagram expansion was in order. I used my field theoretic training to crank out the diagrams and my understanding of the renormalization group fixed point formalism to determine how to make use of the diagrams I computed. The results were published in a second letter in early 1972.[52] The consequent explosion of research is discussed in Part IV.

There were independent efforts on the same area taking place while I completed my work. The connection between critical phenomena and quantum field theory was recognized by Gribov and Migdal and Polyakov[53] and by axiomatic field theorists such as Symanzik T. T. Wu[55] worked on both field theory and the Ising model. Larkin and Khmelnitskii applied the field theoretic renormalization group of Gell-Mann and Low to critical phenomena in four dimensions and to the special case of uniaxial ferromagnets in three dimensions,[56] in both cases deriving logarithmic corrections to Landau's theory. Dyson formulated a somewhat artificial "hierarchical" model of a phase transition which was exactly solved by a one dimensional integral recursion formula.[54] This formula was almost identical to the one I wrote down later, in the 1971 paper. Anderson' worked out a simple but approximate procedure for eliminating momentum scales in the Kondo problem, anticipating my own work in the Kondo problem (see Sec. IV). Many solid state theorists were trying to apply diagrammatic expansions to critical phenomena, and Abe[58] and Scalapino and Ferrell[59] laid the basis for a diagrammatic treatment of models with a large number of degrees offreedom, for any dimension. (The limit of an infinite number ofdegrees of freedom had already been solved by Stanley[60]). Kadanoff was making extensive studies of the Ising model,[61] and discovered a short distance expansion for it similar to my own expansion for Geld theories. Fractional dimensions had been thought about before in critical phenomena.[62] Continuation of Feynman diagrams to non-integer dimensions was introduced into quantum Geld theory in order to provide a gauge invariant regularization procedure for non-abelian gauge theories:[63] this was done about simultaneously with its use to develop the ε expansion.

In the late '60's, Migdal and Polyakov[64] developed a "bootstrap" formulation

of critical phenomena based on a skeleton Feynman graph expansion, in which all parameters including the expansion parameter inself would be determined self-consistently. They were unable to solve the bootstrap equations because of their complexity, although after the ε expansion about four dimensions was discovered, Mack showed that the bootstrap could be solved to lowest order in ε[65]. If the 1971 renormalization group ideas had not been developed, the Migdal-Polyakov bootstrap would have been the most promising framework of its time for trying to further understand critical phenomena. However, the renormalization group methods have proved both easier to use and more versatile, and the bootstrap receives very little attention today.

In retrospect the bootstrap solved a problem I tried and failed to solve; namely how to derive the Gell-Mann-Low and Kadanoff dream of a fixed point involving only one or two couplings - there was only one coupling constant to be determined in the Migdal-Polyakov bootstrap. However, I found the bootstrap approach unacceptable because prior to the discovery of the ε expansion no formal argument was available to justify truncating the skeleton expansion to a finite number of terms. Also the skeleton diagrams were too complicated to test the truncation in practice by means of brute force computation of a large number of diagrams. Even today, as I review the problems that remain unsolved either by ε expansion or renormalization group methods, the problem of convergence of the skeleton expansion leaves me unenthusiastic about pursuing the bootstrap approach, although its convergence has never actually been tested. In the meantime, the Monte Carlo Renormalization group[66] has recently provided a framework for using small number of couplings in a reasonably effective and nonperturbative way: see Section IV.

I am not aware of any other independent work trying to understand the renormalization group from first principles as a means to solve field theory or critical phenomena one length scale at a time, or suggesting that the renormalization group should be formulated to allow arbitrarily many couplings to appear at intermediate stages of the analysis.

IV. *Results after 1971*

There was an explosion of activity after 1972 in both renormalization group and ε expansion studies. To review everything that has taken place since 1972 would be hopeless. I have listed a number of review papers and books which provide more detailed information at the end of this paper. Some principal results and some thoughts for the future will be outlined here. The "ε expansion" about four dimensions gave reasonable qualitative results for three dimensional systems. It enabled a much greater variety of details of critical behavior to be studied than was previously possible beyond the mean field level. The principal critical point is characterized by two parameters: the dimension d and the number of internal components n. Great efforts were made to map out critical behavior as a function of d and n. ε expansion and related small coupling expansions were carried to very high orders by Brézin, Le Guillou, Zinn-Justin,[67] and Nickel"" led to precise results for d ≃ 3.[69,70] The large n limit and l/n expansion was pursued further." A new expansion in 2+ε dimensions was

developed for n>2 by Polyakov.[72] For n = 1 there is an expansion in 1+e dimensions.[73] The full equation of state in the critical region was worked out in the ε expansion[74] and l/n expansion.[75] The special case n = 0 was shown by De Gennes to describe the excluded volume problem in polymer configuration problems and random walks.[76] Corrections to scaling were first considered by Wegner[77]. A recent reference is Aharony and Ahlers.[78]

Besides the careful study of the principal critical point other types of critical points and critical behavior were pursued. Tricritical phenomena were investigated by Riedel and Wegner,[79] where Landau theory was found to breakdown starting in three dimensions instead of four.[80] More general multicritical points have been analyzed."' Effects of dipolar forces,"' other long range forces,[83] cubic perturbations and anisotropies[84,85] were pursued. The problems of dynamics of critical behavior were extensively studied.[86] Liquid crystal transitions were studied by Halperin, Lubensky, and Ma.[87]

Great progress has been made in understanding special features of two dimensional critical points, even though two dimensions is too far from four for the ε expansion to be practical. The Mermin-Wagner theorem"" foreshadowed the complex character of two dimensional order in the presence of continuous symmetries. The number ofexactly soluble models generalizing the Ising model steadily increases.[89] Kosterlitz and Thouless[90] blazed the way for renormalization group applications in two dimensional systems, following earlier work by Berezinskii.[91] They analyzed the transition to topological order in the 2-dimensional xy model with its peculiar critical point adjoining a critical line at lower temperatures; for further work see José et al[92] and Fröhlich and Spence[93,94]. Kadanoff and Brown have given an overview of how a number of the two-dimensional models interrelate.[95] A subject of burning recent interest is the two-dimensional melting transition.[96] Among generalizations of the Ising model, the 3 and 4 state Potts model have received special attention. The three-state Potts model has only a first order transition in mean field theory and an expansion in 6-ε dimensions but has a second order transition in two dimensions.'" The four state Potts model has exceptional behavior in two dimensions (due to a "marginal variable"), which provides a severe challenge to approximate renormalization methods. Notable progress on this model has been made recently.[98]

A whole vast area of study concerns critical behavior or ordering in random systems, such as dilute magnets, spin glasses, and systems with random external fields. Random systems have qualitative characteristics of a normal system in two higher dimensions as was discovered by Lacour-Gayet and Toulouse[99] Imry, Ma, Grinstein, Aharony,[100] and Young"" and confirmed by Parisi and Sourlas[102] in a remarkable paper applying 'supersymmetry' ideas from quantum field theory.[103] The "replica method" heavily used in the study of random systems[104] involves an $n \to 0$ limit, where n is the number ofreplicas similar to the De Gennes $n \to 0$ limit defining random walks.[76] There are serious unanswered questions surrounding this limiting process. Another curious discovery is the existence of an $\varepsilon^{1/2}$ expansion found by Khmelnitskii and Grinstein and Luther.[105]

Further major areas for renormalization group applications have been in

percolation,[106] electron localization or conduction in random media, [107] the problems of structural transitions and "Lifshitz" critical points,"'" and the problem of interfaces between two phases.[109]

Much of the work on the **E** expansion involved purely Feynman graph techniques; the high order computations involved the Callan-Symanzik formulation[110] of Gell-Mann Low theory. The computations also depended on the special diagram computation techniques of Nickel[68] and approximate formulae for very large orders of perturbation theory first discussed by Lipatov.[111] In lowest order other diagrammatic techniques also worked, for example the Migdal-Polyakov bootstrap was solved to order ε by Mack.[65]

The modern renormalization group has also developed considerably, Wegner[77,112,113] strengthened the renormalization group formalism considerably. A number ofstudies, practical and formal"' were based on the approximate recursion formula introduced in 1971. Migdal and Kadanoff[115] developed an alternative approximate recursion formula (based on "bond moving" techniques). Real space renormalization group methods were initiated by Niemeijer and Van Leeuwen[116] and have been extensively developed since.[117, 118] The simplest real space transformation is Kadanoff's "spin decimation" transformation[119].[120] where roughly speaking some spins are held fixed while other spins are summed over, producing an effective interaction on the fixed spins.

The decimation method was very successful in two dimensions where the spins on alternative diagonals of a square lattice were held fixed.[120] Other real space formulations[116, 117] involved kernels defining block spin variables related to sums of spins in a block (the block could be a triangle, square, cube, a lattice site plus all its nearest neighbors, or whatever).

Many of the early applications of real space renormalization group methods gave haphazard results - sometimes spectacularly good, sometimes useless. One could not apply these methods to a totally new problem with any confidence of success. The trouble was the severe truncations usually applied to set up a practical calculation; interactions which in principle contained thousands of parameters were truncated to a handful of parameters. In addition, where hundreds of degrees of freedom should be summed over (or integrated over) to execute the real space transformation, a very much simplified computation would be substituted. A notable exception is the exactly soluble differential renormalization group transformation of Hilhorst, Schick and Van Leeuwen, which unfortunately can be derived only for a few two dimensional models.[121, 122.]

Two general methods have emerged which do not involve severe truncations and the related unreliability. First of all, I carried out a brute force calculation for the two dimensional Ising model using the Kadanoff decimation approach[119, 120] (as generalized by Kadanoff). Many interaction parameters (418) were kept and the spin sums were carried out over a very large finite lattice. The results were very accurate and completely confirmed my hypothesis that the local couplings of the shortest range were the most important. Most importantly the results could be an *optimization principle.* The fixed point of Kadanoff's decimation transformation depends on a single arbitrary parameter; it was possible to determine a best value for this parameter from internal consistency consider-

ations. Complex calculations with potentially serious errors always are most effective when an optimization principle is available and parameters exist to optimize on.[123] This research has never been followed up, as is often the case when large scale computing is involved. More recently, the Monte Carlo Renormalization group method,"" developed by Swendsen, myself, Shenker, and Tobochnik (see also Hilhorst and Van Leeuwen)"' has proved very accurate and may shortly overtake both the high temperature expansions and the ε expansion as the most accurate source of data on the three dimensional Ising model. The Monte Carlo Renormalization Group is currently most successful on two dimensional problems where computing requirements are less severe: it has been applied successfully to tricritical models and the four-state Potts model.[124] In contrast, the ε expansion is all but useless for two dimensional problems. Unfortunately, none of the real space methods as yet provide the detailed information about correlation functions and the like that are easily derived in the ε expansion.

A serious problem with the renormalization group transformations (real space or otherwise) is that there is no guarantee that they will exhibit fixed points. Bell and myself[125] and Wegner in a more general and elegant way[113] have shown that for some renormalization group transformations, iteration of a critical point does not lead to a fixed point, presumably yielding instead interactions with increasingly long range forces. There is no known principle for avoiding this possibility, and as Kadanoff has showed using his decimation procedure,[120] a simple approximation to a transformation can misleadingly give a fixed point even when the full transformation cannot. The treatment that I gave of the two dimensional Ising model has self consistency checks that signal immediately when long range forces outside the 418 interactions kept are becoming important. Nothing is known yet about how the absence of a fixed point would be manifested in the Monte Carlo renormalization group computations. Cautions about real space renormalization group methods have also been advanced by Griffiths et al.[126]

There is a murky connection between scaling ideas in critical phenomena and Mandelbrot's "fractals" theory - a theory of scaling of irregular geometrical structures (such as coastlines).[127]

Renormalization group methods have been applied to areas other than critical phenomena. The Kondo problem is one example. Early renormalization group work was by Anderson" and Fowler and Zawadowski.[128] I then carried out a very careful renormalization group analysis of the Kondo Hamiltonian,[129] producing effective Hamiltonians with many couplings for progressively smaller energy scales, following almost exactly the prescription I learned for fixed source meson theory. The result was the zero-temperature susceptibility to about 1 % accuracy, which was subsequently confirmed by Andrei and Wiegmann's[7] exact solution. Renormalization group methods have been applied to other Hamiltonian problems, mostly one dimensional.[130] In multidimensional systems and in many one dimensional systems, the effective Hamiltonians presently involve too many states to be manageable.

The renormalization group has played a key role in the development of Quan-

turn Chromodynamics - the current theory of quarks and nuclear forces. The original Gell-Mann Low theory[29] and the variant due to Callan and Symanzik[110] was used by Politzer, Gross and Wilczek[131] to show that nonabelian gauge theories are asymptotically free. This means that the short distance couplings are weak but increase as the length scale increases; it is now clear that this is the only sensible framework which can explain, qualitatively, the weak coupling that is evident in the analysis of deep inelastic electron scattering results (off protons and neutrons) and the strong coupling which is evident in the binding of quarks to form protons, neutrons, mesons, etc.[132] I should have anticipated the idea of asymptotic freedom[133] but did not do so. Unfortunately, it has been hard to study quantum chromodynamics in detail because of the effects of the strong binding of quarks at nuclear distances, which cannot be treated by diagrammatic methods. The development of the lattice gauge theory by Polyakov and myself[134] following pioneering work of Wegner[135] has made possible the use of a variety of lattice methods on the problems of quantum chromodynamics,[136] including strong coupling expansions, Monte Carlo simulations, and the Monte Carlo renormalization group methods.[67], [137] As computers become more powerful I expect there will be more emphasis on various modern renormalization group methods in these lattice studies, in order to take accurately into account the crossover from weak coupling at short distances to strong coupling at nuclear distances.

The study of unified theories ofstrong, weak and electromagnetic interactions makes heavy use of the renormalization group viewpoint. At laboratory energies the coupling strengths of the strong and electromagnetic interactions are too disparate to be unified easily. Instead, a unification energy scale is postulated at roughly 10^{15} GeV; in between renormalization group equations cause the strong and electroweak couplings to approach each other, making unification possible. Many grand unified theories posit important energy scales in the region between 1 and 10^{15} GeV. It is essential to think about these theories one energy scale at a time to help sort out the wide range of phenomena that are predicted in these theories. See Langacker[138] for a review. The study of grand unification has made it clear that Lagrangians describing laboratory energies are phenomenological rather than fundamental, and this continues to be the case through the grand unification scale, until scales are reached where quantum gravity is important. It has been evident for a long time that there should be applications of the renormalization group to turbulence, but not much success has been achieved yet. Feigenbaum[138] developed a renormalization group-like treatment of the conversion from order to chaos in some simple dynamical systems,"" and this work may have applications to the onset of turbulence. Feigenbaum's method is probably too specialized to be of broader use, but dynamical systems may be a good starting point for developing more broadly based renormalization group methods applicable to classical partial differential equations.[141]

In my view the extensive research that has already been carried out using the renormalization group and the ε expansion is only the beginning of the study of a much larger range of applications that will be discovered over the next twenty years (or perhaps the next century will be required). The quick successes

of the ε expansion are now past, and I believe progress now will depend rather on the more difficult, more painful exercises such as my own computations on the two dimensional Ising model and the Kondo problem,[120] or the Monte Carlo Renormalization group[66] computations. Often these highly quantitative, demanding computations will have to precede simpler qualitative analysis in order to be certain the many traps potentially awaiting any renormalization group analysis have been avoided.

Important potential areas of application include the theory of the chemical bond, where an effective interaction describing molecules at the bond level is desperately needed to replace current ab initio computations starting at the individual electron level."' A method for understanding high energy or large momentum transfer Quantum Chromodynamics (QCD) cross sections (including non-perturbative effects) is needed which will enable large QCD backgrounds to be computed accurately and subtracted away from experimental results intended to reveal smaller non-QCD effects. Practical areas like percolation, frost heaving, crack propagation in metals, and the metallurgical quench all involve very complex microscopic physics underlying macroscopic effects, and most likely yield a mixture of some problems exhibiting fluctuations on all length scales and other problems which become simpler classical problems without fluctuations in larger scales.

I conclude with some general references. Two semi-popular articles on the renormalization group are Wilson (1979) and Wilson (1975). Books include Domb and Green (1976); Pfeuty and Toulouse (1977); Ma (1976); Amit (1978); Patashinskii and Pokrovskii (1979)[143]; and Stanley (1971)[144.] Review articles and conference proceedings include Widom (1975)"'; Wilson and Kogut (1974); Wilson (1975); Fisher (1974)""; Wallace and Zia (1978); Greer and Moldover (1981); and Lévy et al. (1980).[147]

I thank the National Science Foundation for providing funding to me, first as a graduate student, then throughout most of my research career. The generous and long term commitment of the United States to basic research was essential to my own success. I thank my many colleagues at Cornell, especially Michael Fisher and Ben Widom, for encouragement and support. I am grateful for the opportunity to be a member of the international science community during two decades of extraordinary discoveries.

REFERENCES

1. Wilson, K. G. Physica 73, 119 (1974).
2. See, e.g. Rose H. A. and Sulem, P. L.J. Physique 39, 441 (1978).
3. See, c.g. Criegee L. and Knies, G. Phys. Rcpts. 83, 151 (1982).
4. See, CRC Handbook of chemistry and Physics (62nd Edition), Weast, R. C. rd. (CRC Press, Boca Raton, Florida, 1981), p. F-76.
5. See, e.g., Anderson, P. W . J. Phys. C3, 2346 (1970).
6. Onsager, L. Phys. Rev. 6.5, 117 (1944).
7. Andrei, N. Phys. Rev. Lett. 45, 379 (1980; Phys. Lett. 87A, 299 (1982); Andrei, N. and Lowcnstein, J, H. Phys. Rev. Lett. 46, 356 (1981): Wiegmann, P. B. Phys. Lett. 80A, 163 (1980); J. Phys. C14, 1463 (1981); Filyov V. M. Tzvrlik, A. M. and Weigmann, P. B. Phys. Lett. *81A,* 175 (1981).
8. Johnson', K. Nuovo Cimcnto 20, 773 (1961).
9. The experimcntal measurements on fluids (e.g., SF_6, He^3, and various organic fluids.) give $\beta = 0.32 \pm 0.02$, while current theoretical computations give $\beta = 0.325 \pm 0.005$: see Greer S. C and Moldover, M. R. Annual Rcview of Physical Chemistry 32, 233 (1981) for data and caveats.
10. For experimental reviews, see e.g. Heller, P. Repts. Prog. Phys. 311, 731 (1967): Kadanoff, I.. Götze, W. Hamblen, D. Hecht, R. Lewis, E. A. S. Palciauskas, V. V. Rayl. M. Swift. ,J. Aspres, D. Kane, ,J. *Static Phenomena* near *Critical Points: Theory: and Experiment,* Rev. Mod. Phys. 39, 395 (1967).
11. For the alloy transition in β brass see Als-Nielsen, J in Domb. C. and Green, M. S. eds.. *Phase Transitions and Critical Phenomena.* Vol. 5a (Academic Press, *New* York. 1978) p.87. See Ref. 10 for carlier reviews of other systems.
12. For a history of the Lcnz-Ising model see Brush. S. G Rev. Mod. Phys. 39, 883 (1967).
13. Landau. L. D. Phys. Zurn. Sowjetunion *11, 545* (1937); English translation: ter Haar, D. Men *of Physics: I.. D. Landau. Vol II.* II (Pergamon, Oxford. 1969).
14. Ginzburg. V. L. and Landau. L. D. Zh. Eksperim. i Teor. Fiz. 20. 1061 (1950). See also Schrieffer, J. Superconductivity [Benjamin. New York. 1964) p.19.
15. Ginzburg, V.L 1.. Fiz. Tverd. Tela 2 2031 (1960). English translation: Soviet Physics: Solid St. 2 1824 (1960).
16. Wilson, K. G. and Fisher. M. E. Phys. Rev. Letters 28. 240 (1972).
17. Wilson. K G and Kogut. J. Phys. Repts. 12C. 75. 7.5 (1974).
18. See, e.g., Guggcnhcim. E. A.J. Chem. Phys 13, 253 (1945): Griffiths. R. B. Phys. Rev. Letts. 24. 1479 (1970): Griffiths R. B. and Whcelcr, J. B. Phys. Rev. A2 1047 (1970): Kadanoff L. in *Phase Transitions and Critical Phenomena.* Domb C. and Green. M. S. Eds. Vol. *5a* (Academic. New York, 1978, p. I .).
19. See, e.g., DeBoer, J. Physica 7.7. 1 (1974): Klein, M.D. Physica 73, 28 (1971); Lcvelt.]. 51. H. Sengers, Physica 7.7, 73 (1974).
20. Weiss. P. J. Phys. 6, 661 (1907).
21. Sec. Domb. C. Proc. Roy. Soc. *A199,* 199 (1949) and, e.g. Fisher. M.E. E. Reports. Prog. Phys. 30, 615 (1967).
22. See. e.g., ref: 10 and Ahlers, G. Revs. Mod. Phys. 52. 489 (1980).
23. Wisdom, B. J. Chcm. Phys. 43, 3898 (1965).
24. Essam J. W. and Fisher, M. E. J. Chem. Phys. 38, 802 (1963): Domb C. and Hunter. D. L. Proc. Phys. Soc. 86, 1147 (1965) Kadanoff, L. P. Physics, 2. 263 (1966): Patashiniskii A. Z. and Prokrovskii. V. L. Zh. Eksper. Teor. Fiz. 50. 439 (1966); English translation: Soviet Physics-JETP 23,. 292 (1966).
25. Rushbrookc. G. S.. 1963. J. Chcm. Phys. 39. 842.
26. Griffiths, R. B.. 1965. Phys. Rev. Lett. 14, 623.
27. See the reprint collection Schwinger, J. Ed Quantum Electrodynamics.. Dover. New York. 1958).
28. Stueckelberg E. C. G. and Petermann. A Hclv. Phys. Acta 26. 499 (1953): see also Pctermann. A. Phys. Rcpts. 53, I57 11979).
29. Gell-Mann M. and Low. F. E. Phys. Rev. 95. 1300 (1954).

30. Kadanoff, L. P. Physics 2, 263 (1966).
31. Bogoliubov N. N. and Shirkov, D. V. Introduction to the *Theory of Quantized Fields* (Interscience; New York, 1959), Ch. VIII.
32. Ter-Martirosyan, K. A. Zh. Eksperim. Teor. Fiz. 39, 827 (1960); Soviet Physics-JETP 12, *575* (1961).
33. See, e.g. Wilson, K. G. Acta Physica Austriaca 17, 37 (1963-64) and references cited therein.
34. Wenzel, G. Helv. Phys. Acta 13, 269 (1940); 14, 633 (1941); see also Henley E. M. and Thirring, W. *Elementary Quantum Field Theory* (McGraw Hill, New York, 1962).
35. Lee T. D. and Yang, C. N. Phys. Rev. 128,885 (1962).
36. Johnson, K. Baker, M. and Willey, R. S. Phys. Rev. Lett. 11, 518 (1963). For a subsequent view see Baker, M. and Johnson, K., Phys. Rev. 183, 1292 (1969); ibid D3,2516,2541 (1971).
37. See e.g. Hepp, K. Acta Physica Austriaca 17, 85 (1963-64).
38. See e.g. Lowenstein, J. H. Commun. Math. Phys. 16, 265 (1970); Wilson, K. G. Phys. Rev. D2, 1473 (1970).
39. Wilson, K. G. Phys. Rev. 140, B445 (1965).
40. Dyson, F. J. Phys. Rev. 83, 608, 1207 (1951).
41. Mitter, P. K., and G. Valent, 1977, Phys. Lett. *B70, 65.*
42. See, e.g. Kogut J. and Susskind, L. Phys. Repts. 8C, 75 (1973) and referenccs cited therein.
43. Schultz, T. D. Mattis, D. C. and Lieb, E. H. Rev. Mod. Phys. 36, 856 (1964).
44. Schiff, L. I. Phys. Rev. 92, 766 (1953).
45. See, e.g. Mack, G. Nucl. Phys. *B5,* 499 (1968) and references cited therein.
46. Wilson, K. G. Phys. Rev. 179, 1499 (1969).
47. Bjorken, J. Phys. Rev. 179, 1547 (1969); Feynman, R. P. Phys. Rev. Lett. 23, 1415 (1969): for a review see Yan, T. M. Annual Review of Nuclear Science, 26, 199 (1976) and Feynman, R. P. *Photon-Hadron Interactions,* Benjamin, W. A. (Reading, Mass, 1972).
48. Wilson, K. G. Phys. Rev. D2, 1438 (1970).
49. Di Castro, C. and Jona-Lasinio, G. Phys. Letters *29A, 322* (1969).
50. Wilson, K. G Phys. Rev. *B4, 3174,* 3184 (1971).
51. Wilson, K. G. and Fisher, M. E. Phys. Rev. Lett. 28, 240 (1972).
52. Wilson. K. G. Phys. Rev. Lett. 28, 548 (1972).
53. Gribov, V. N. and Migdal, A. A. Zh. Eksp. Teor. Fiz. 5.5, 1498 (1968); English Translation: Soviet Physics-JETP 28, 784 (1968); Migdal, A. A ibid. 59, 1015 (1970); JETP 32, 552 (1971); Polyakov, A. M. ibid. 55, 1026 (1968); JETP 28, 533 (1969): Polyakov, A.M . ibid. 57, 271 (1969); JETP 30, 151 (1970); Polyakov, A. M. ibid. 59, 542 (1970); JETP 32. 296 (1971).
54. Symanzik, K.J. Math. Phys. 7, 510 (1966).
55. Wu, T. T. Phys. Rev. 149, 380 (1966); McCoy, B. M. and Wu, T. T. *The Two Dimensional Ising Model* (Harvard University Press, Cambridge, 1973); McCoy, B. M. Tracy, C. and Wu, 'I'. T. Phys. Lett. *MA, 283 (1977).*
56. Larkin, A. I. and Khmelnitskii, D. E. Zh. Eksp. Teor. Fiz. .56, 2087 (1969); English Translation: Soviet Physics JETP 29, 1 1123 (1969).
57. Dyson, F.J. Commun. Math. Phys. 12, 91 (1969); see also Baker. G. A. Jr., Phys. Rev. *B5, 2622 (1972).*
58. Abe, R. Progr. Theor. Phys. (Kyoto) 48, 1414 (1972); ibid. 49, 113 (1973); Abe, R. and Hikami, S. Phys. Letters *45A,* 11 *(1973)* Hikami, S. Progr. Theor. Phys. (Kyoto) 49, 1096 (1973).
59. Ferrell, R. A. and Scalapino, D.J. Phys. Rev. Letters 29, 413 (1972): Phys. Letters *41A.* 371 (1972).
60. Stanley, H. E. Phys. Rev. 176, 718 (1968).
61. Kadanoff L. P. Phys. Rev. Lett. 23, 1430 (1969); Phys. Rev. 188, 859 (1969).
62. See, e.g. Fisher, M. E. and Gaunt, D. S. Phys. Rev. 133, A224 (1964); Widom, B. Mol. Phys. *25, 657* (1973).
63. 't Hooft, G. and Veltman, M. Nucl. Phys. *B44. 189* (1972); Bollini, C. G and Giambiagi, J.J. Phys. Letts. *40B, 566 (1972);* Ashmore, J. F. Lettere Nuovo Cimento 4. 289 (1972).
64. Patashinskii, A. Z. and Pokrovskii, V. I.. Zh. Eksp. Teor. Fiz. 46, 994 (1964): English Translation: Soviet Physics, JETP /9, 677 (1964); Polyakov, A. M Zh. ETF. Pis. Red. *12, 538*

(1970); English Translation: JETP Letters 12, 381 (1970); Migdal, A. A. Phys. Letts. *37B, 386* (1971); Mack, G. and Symanzik, K. Common Math Phys. 27,247 (1972) and references cited therein.

65. Mack, G. in Strong Interaction Physics, Rühl, W. and Vanira, A. eds. (Springer-Verlag, Berlin, 1973), p.300.
66. Swendsen, R. H. Phys. Rev. Lett. 42, 859 (1979); Phys. Rev. *B20,* 2080 (1979); Wilson, K. G. Monte Carlo Calculation for the Lattice Gauge Theory, in *Recent Developments in Gauge Theories,* 't Hooft, G. et al. Eds. (Plenum Press, New York, 1980); Shenker, S. and Tobochnik, J. Phys. Rev. *B22, 4462 (* 1980).
67. See, e.g. the review of Zinn-Justin, J. Phys. Repts. 70, 109 (1981); see further Nickel, B. Physica *106A, 48* (1981).
68. Nickel, B. (unpublished), cited in Ref. 64 and prior papers.
69. Parisi, G., 1980, J. Stat. Phys. 23, 49.
70. Vladimirov, A. A., D. I. Kazakov, and O. Tarasov, 1979, Sov. Phys. JETP *50,* 521.
71. See, e.g. Ma, S. K. in Domb, C. and Green, *M. S.* eds. *Phase Transitions and Critical Phenomena,* Vol. 6 (Academic Press, London, 1976), p.250, and references cited therein.
72. Polyakov, A. M. Phys. Lett. *59B,* 79 (1975); Migdal, A. A., Zh. Eksp. Tear. Fiz. 69,. 1457 (1975); English translation: Soviet Physics-JETP 42, 743 (1976); Brkzin, E. Zinn-Justin, ,J. and LeGuillou, J, C. Phys. Rev. *B14,* 4976, (1976); Bhattacharjee, J. K. Cardy, J. L. and Scalapino, D. J. Phys. Rev. *B25,* 1681 (1982).
73. Wallace, D. J. and Zia, R. K. Phys. Rev. Lett. 35, 1399 (1979); Widom, B. (ref. 59).
74. Brézin, E. Wallace, D. and Wilson, K. G. Phys. Rev. Letters, 29, 591 (1972); Phys. Rev. *B7,* 232 (1973); Avdeeva, G. M. and Migdal, A. A. Zh. ETF. Pis. Red. 16, 253 (1972); English Translation: Soviet Physics-JETP Letters 16, 178 (1972); Avdeeva, G. M. Zh. Eksp. Tear. Fiz. 64, 741 (1973), English Translation: Soviet Physics-JETP 37, 377 (1973).
75. Brézin, E. and Wallace, D. J. Phys. Rev. *B7, 1967 (1973);* for a review see Ma S.-K. in Domb, C. and Green, M. S. Eds., Phase Transitions and Critical Phenomena, Vol. 6 (Academic Press, London, 1976) p.250.
76. DeGennes, P. G. Phys. Letters *38A, 339* (1972); des Cloiseaux, J. J. Phys. (Paris) 36, 281 (1975).
77. Wegner, F. J., 1972a, Phys. Rev. *B5, 4529.*
78. Aharony, A. and G. Ahlers, 1980, Phys. Rev. Lett. 44, 782.
79. Riedel, E. K. and Wegner, F. J. Phys. Rev. Lett. 29, 349 (1972); Phys. Rev. *B7, 248* (1973).
80. Stephen, M. J., E. Abrahams, and J. P. Straley, 1975, Phys. Rev. *B12, 256.*
81. See, e.g., Fisher, M. E. in A.I.P. Conference Proceedings #24, Magnetism *and Magnetic Materials* 1974, p. 273.
82. Fisher, M. E. and Aharony, A. Phys. Rev. Lett. 30, 559 (1973); Aharony, A. and Fisher, M. E. Phys. Rev. *B8,* 3323 (1973); Aharony, A. Phys. Rev. *B8, 3342, 3349, 3358, 3363* (1973); Aharony, A. Phys. Letters *44A,* 313 (1973).
83. Suzuki, M. Phys. Letters, *42A, 5* (1972); Suzuki, M. Progr. Theor. Phys. (Kyoto) 49, 424, 1106, 1440 (1973); Fisher, M. E. Ma, S. and Nickel, B. G. Phys. Rev. Letters 29, 917 (1972); Baker Jr., G. A. and Golner, G. R. Phys. Rev. Letters 31, 22 (1973); Suzuki, M. Yamazaki, Y. and Igarishi, G. Phys. Letters *42A,* 313 (1972); Sak, J. Phys. Rev. *B8, 281 (1973).*
84. Pfeuty, P. and Fisher, M. E. Phys. Rev. *B6,* 1889 (1972); Wallace, D. J. J. Phys. C6, 1390 (1973); Ketley, L. J. and Wallace, D. J. J. Phys. *A6,* 1667 (1973); Aharony, A. Phys. Rev. Letters 31, 1494 (1973); Suzuki, M. Progr. Theor. Phys. (Kyoto) 49, 1451 (1973); Liu, I.. Phys. Rev. Letters 31, 459 (1973); Grover, M. K. Phys. Letters *44A,* 253 (1973); Chang, T. S. and Stanley, H. E. Phys. Rev. *B8, 4435 (1973);* Aharony, A. Phys. Rev. *B8, 4270* (1973); A recent reference is Blankschtein, D. and Mukamel, D. Phys. Rev. *B25, 6939* (1982).
85. Wegner, F. J., 1972b, Phys. Rev. *B6,* 1891.
86. Halperin, B. I. Hohenberg, P. C. and Ma, S.-K. Phys. Rev. Letters 29, 1548 (1972); Suzuki, M. and Igarishi, G. Progr. Theor. Phys. (Kyoto) 49, 1070 (1973): Suzuki, M. Phys. Letters *43A, 245* (1973); Suzuki, M. Progr. Theor. Phys. (Kyoto) 50, 1767 (1973); for a review see Hohenberg, P. C. and Halperin, B. I. Revs. Mod. Phys. 49, 435 (1977); for a recent Monte Carlo Renormalization Group method see Tobochnik, J, Sarker, S. and Cordery, R. Phys.

Rev. Lett, 46, 1417, (1981). Other recent references are Ahlers, G. Hohenberg, P. C. and Kornblit, A. Phys. Rev. *B25*, 3136 (1982); Heilig, S. J. Luscombe, J. Mazrnko, G. F. Oguz, E. and Valls, 0. T. Phys. Rev. *B25, 7003 (1982).*

87. Halperin B. I., T. C. Lubensky, and S.-k. Ma, 1974, Phys. Rev. Letters, 32, 242.
88. Mermin, N. D. and Wagner, H. Phys. Rev. Lett. 17, 1133 (1966); Mermin, N. D. J. Math. Phys. 8, 1061 (1967); Hohenberg, P. C. Phys. Rev. 158, 383 (1967).
89. See, e.g., Baxter, R. *Exactly Solved Models in Statistical Mechanics,* (Academic Press, New York, 1982).
90. Kosterlitz,J. M. and Thouless, D. J. J. Phys. C6, 1181 (1973); Koster1itz.J. M. ,J. Phys. C7, 1046 (1974); for a review see Kosterlitz, J. M. and Thouless, D. J. Prog. Low Temp. Phys. 7, 373 (1978).
91. Bcrezinskii, V. L. Zh. Eksp. Tear. Fiz. 59, 907 (1970); (Soviet Physics-JETP 32,493 (1971)); Berezinskii, V. I.. Zh. Eksp. Tear. Fiz. *61,* 1144 (1971) (Soviet Physics-JETP 34,610 (1972)).
92. José, J. V., L. P. Kadanoff, S. Kirkpatrick, and D. R. Nelson, 1977, Phys. Rev. *B16, 1217.*
93. Frölich, J., and T. Spencer, 1981, Phys. Rev. Lett. 46. 1006.
94. Fröhlich, J., and T. Spencer, 1981, Commun. Math. Phys. *81, 527.*
95. Kadanoff, L. P. and Brown. A. C. Annals Phys. (N.Y.) 121, 318 (1979).
96. Nelson, D. R. and Halperin, B. I. Phys. Rev. *B19, 2457 (1979);* Young, A. P. Phys. Rev. *B19,* 1855 (1979).
97. Baxter, R. J. J. Phys. C6, 2445 (1973) rigorouus 2-d solution); see e.g., Banavar, J. R. Grest, G. S and Jasnow, D. Phys. Rev. *B25, 4639 (1982)* and references cited therin (d>2).
98. Nienhuis, B. Berker, A. N. Riedel, E. K. and Schick, M. Phys. Rev. Lett. 43. 737 (1979); Swendsen, R. H. Andelman, D. and Berker, A. N. Phys. Rev. *B24,6732* (1982) and references cited therein.
99. Lacour-Gayet, P., and G. Toulouse, 1974, J. Phys. 35. 425.
100. Imry, Y. and Ma, S.-K. Phys. Rev. Lett. 35, 1399 (1975); Grinstein, G. Phys. Rev. Lctt. 37, 944 (1976); Aharony, A. Imry, Y. and Ma, S.-K. Phys. Rev. Lett. 37. 1367 (1976); a recent reference is: Mukamel, D. and Grinstein, G. Phys. Rev. *B25, 381 (1982).*
101. Young, A. P., 1977, J. Phys. *C10,* L257.
102. Parisi, G. and Sourlas, N. Phys. Rev. Lctt. 43, 744 (1979).
103. Parisi, G., and N. Sourlas, 1981, Phys. Rev. Lett. 46. 871.
104. Edwards, S. F. and Anderson, P. W. J. Phys. F5, 965 (1975).
105. Khmelnitskii, D. E. Zh. Eksp. Teor. Fiz. 68, 1960 (1975), (Sov. Phys.-JETP *41. 981 (1976)):* Grinstein, G. and Luther, A. Phys. Rev. *B13, 1329 (1976).*
106. See, e.g., the review of Essam, J. W. Repts. Prog. Phys. 43. 833 (1980): a recent reference is Lobb, C. J. and Karasek, K. R. Phys. Rev. *B25, 492 (1982).*
107. Nagaoka, Y., and H. Fukuyama, editors, 1982, *Anderson Localization* Springer Series in Solid-State Science (Springer-Berlin), Vol. 39.
108. See, e.g., the review of Bruce, A. D. Adv. Phys. 29, 111 (1980); a recent paper is Grinstein. G. and Jayaprakash, C. Phys. Rev. *B25, 523* (1982).
109. See Ref. 68 and references cited therein.
110. Callan, C. G. Phys. Rev. D2, 1541 (1970); Symanzik, K. Commun. Math. Phys. 18. 227 (1970); Symanzik, K. Springer Tracts in Modern Physics 57. 222 (1971).
1 I 111. Lipatov, L. N. Zh. Eksp. Teor. Fiz. 72, 411 (1977); English translation, Soviet Physics-JETP 45, 216 (1977).
Brezin, E. Le Guillou; J, C. and Zinn-Justin, J. Phys. Rev. *D15. 1544,* 1558 (1977).
112. Wegner, F. J., 1974, J. Phys. C7, 2098.
113. *See, e.g.* F. Wegner, in *C.* Domb and *M. S.* Green, eds., *Phase Transitions and Critical Phenomena.* (Academic Press, London, 1976) p. 7.
114. See, e.g., Golner, G. Phys. Rev. *B8,* 3419 (1973); Langer, J. S. and Bar-on, M. Ann Phys. (N.Y.) 78, 421 (1973); Bleher, P. M. and Ya. Ya. G. Sinai. Commun. Math. Phys. 45, 247 (1975).
115. Migdal, A. A. Zh. Eksp. Tear. Fiz. 69, 810, 1457 (1975); English Translations: Soviet Physics-JETP 42, 413, 743 (1975), for a review and more references, see Kadanoff L. P. Revs. Mod. Phys. 49,267 (1977).

116. See, e.g., Niemeijer, Th. and Van Leeuwen. J. M. J. in Domb, C. and Green, M. S. eds., Phase Transitions and Critical Phenomena, Vol. 6 (Academic Press, London, 1976), p. 425.
117. See Ref. 80 and e.g., Riedel, E. K. Physica *106A*, 110 (1981) and references cited therein.
118. Burkhardt. T. W., and J. M. J. Van Leeuwen, 1982, *Real-Space Renormalization* (Springer, Berlin)
119. Kadanoff, I.. P., and A. Houghton, 1975, Phys. Rev. *B11, 377.*
120. See Wilson, K. G. Rev. Mod. Phys. 47, 773 (1975) esp. Sect. IV.
121. See Hilhorst H.J. and Van Leeuwen., J. M. J. Physica *106A,* 301 (1981).
122. Hilhorst. H. J., M. Schick, and J. M. J. Van Lceuwen, 1978. Phys. Rev. Lctt. 40, 1605.
123. I thank Yves Parlange for reminding me of this.
124. See, e.g. Swendsen, R. H. Andelman, D. and Berker, A. N. Phys. Rev. *B24, 6732* (1982): Landau, D. P. and Swendscn R. H. Phys. Rev. Lett. 46, 1437 (1981).
125. Bell, T. L. and Wilson, K. G. Phys., Rev. *B10, 3935 (1974).*
126. See, e.g. Griffiths, R. B. Physica *106A, 59* (1981).
127. See, e.g., Mandelbrot, Benoit B. *The Fractal Geometry of Nature* (Freeman, W. H. San Francisco, 1982);
128. Fowler, M. and Zawadowski, A. Solid State Commun. 9, 471 (1971).
129. Wilson, K. G. Revs. Mod. Phys. 47, 773 (1975); Krishna-Murthy, H. R. Wilkins, J. N. and Wilson, K. G. Phys. Rev. Letters 35, 1101 (1975); Phys. Rev. *B21,* 1044, 1104 (1980).
130. See. e.g. Drcll. S. D. Weinstein, M. and Yankielowicz, S. Phys. Rev. *D17, 1769 (1977);* Jullien. R. Fields. ,J. N. and Doniach. S. Phys. Rev. *Bl6. 4889* (1977); recent references are Hanke, W. and Hirsch, J. E. Phys. Rev. *B25,* 6748 (1982); Penson, K. A. Jullien, R. and Pfeuty, P. Phys. Rev. *B25, 1837* (1982) and references cited therein.
131. Politzcr, H. Phys. Rev. Lett. 30, 1346 (1973); Gross, D. and Wilczek, F. Phys. Rev. Lett. 30, 1343 (1973).
132. Ser. e.g. Altarelli, G. Phys. Repts. *81,* 1 (1982).
133. Wilson, K. G. Phys. Rev. *D3,* 1818 (1971).
134. Polyakov A. M. (unpublished); Wilson, K. G. Phys. Rev. *D10. 2455* (1974).
135. Wegner. F.J. Math. Phys. *12, 2259* (1971).
136. See the review of Bander, M. Phys. Repts. 75, 205 (1981).
137. Swendsen, R. H.. *1982,* in *Real-Space Renormalization,* edited by J.W Burkhardt and J. M. J. Van Leeuwen, (Springer. Berlin).
138. Langacker P., 1981. Phys. Rep. 72, 183.
139. Feigenbaum, M.J.J, Stat. Phys. *19, 25* (1978).
140. Sec. e.g. Eckmann. J. P. Rev. Mod. Phys. 53. 643 (1981): Ott, E. Rev. Mod. Phys. 53. 655 (1981).
141. See e.g. Coppersmith, S. and Fisher, D. Bell Laboratories preprint.
142. See. e.g., Mulliken. R. S. Ann. Rev. Phys. Chem. 29, 1 (1978); Löwdin, P. 0. Adv. Quantum Chem. *12,* 263 (1980); Hirst. D. M. Adv. Chem. Phys. 50,517 (1982); Bartlett, R. J. Ann. Rev. Phys. Chem. 32, 359 (1981): Case, D. A. Ann. Rev. Phys. Chem. 33. 151 (1982).
143. Patashinskii, A. Z.. and V. L. Pokrovskii, 1979. *Fluctuating Theory of Phase Transitions* (Perga-"Ion)
144. Stanley, H. E., 1972, *Introduction to Phase Transitions and* Critical *Phenomena* (Oxford University, London)
145. Widom, B., 1975 in *Fundamental Problems in Statistical Mechanics.* edited by E. C. G. Cohen (North-Holland) Vol. III, p. 1.
146. Fisher, 51. E., 1964, J. Math. Phys. 5, 944.
147. Levy, M., J. C. Le Guillou, and J. Zinn-,Justin, editors, *1980, Phase Transitions* (Plenum, New York).

Books Related to the Renormalization Group

Domb, C. and Green, M. *S.* Eds., *Phase Transitions and Critica/ Phenomena* (Academic Press, London) esp. Vol. 6 (1976);
Pfeuty, P. and Toulouse, G. *Introduction to the Renormalization Group and to Critical Phenomena* (Wiley, Sew York, 1977):

Ma, S.-K. *Modern Theory of Critical Phenomena* (Benjamin/Cummings, Reading, Mass., 1976);
Amit, D. J. *Field Theory, the Renormalization Group, and Critical Phenomena* (McGraw-Hill, New York, 1978).

Popular Articles (sort of]
Wilson, K. G. Sci. Am. 241, 158 (August, 1979);
Wilson, K. G. Adv. Math 16, 176 (1975).

Review Articles
Greer, S. C. and Moldover, M. R. Ann. Rev. Phys. Chem. 32, 233 (1981): Critical Behavior of Fluids and many references to prior reviews.
Bander, M. ref. 116: Quark Confinement
Zion-Justin, J. (Ref. 67): Precise Computation of critical Exponents from the ε expansion;
Baret, J. F. Prog. Surf. Membrane Sci. 14, 292 (1981);
Cadenhead, D. A. and Danielli, J. F. Eds., (Academic, New York, 1981): A practical area involving phase transitions.
Wallace, D. J. and Zia, R. K. Repts. Progr. Phys. 41, 1 (1978).

Physics 1983

SUBRAHMANYAN CHANDRASEKHAR

for his theoretical studies of the physical processes of importance to the structure and evolution of the stars

and

WILLIAM A FOWLER

for his theoretical and experimental studies of the nuclear reactions of importance in the formation of the chemical elements in the universe

THE NOBEL PRIZE FOR PHYSICS

Speech by Professor SVEN JOHANSSON of the Royal Academy of Sciences
Translation from the Swedish text

Your Majesties, Your Royal Highnesses, Ladies and Gentlemen,

Astrophysics is one of the areas in physics which has developed most rapidly during recent years. Through satellite technology it has become possible to study the different physical processes which are taking place in stars and other astronomical objects. Space has become a new and exciting laboratory for the physicist. It is true that experiments, in the proper sense of the word, cannot be carried out, but one may observe phenomena which can never be observed in terrestrial laboratories. In space we find matter in the most extreme forms; stars at immensely high temperatures and with enormously high densities, and particles and radiation with an energy which we cannot reach, even with our largest accelerators.

The common theme for this year's prize in physics is the evolution of the stars. From the moment of their birth out of interstellar matter until their extinction, the stars exhibit many physical processes of great interest. In order to put this year's prize in perspective, it is perhaps appropriate to give a short description of the evolution of the stars.

Stars are formed from the gas and dust clouds which are present in galaxies. Under the influence of gravity, this matter condenses and contracts to form a star. During these processes energy is released which leads to a rise in the temperature of the newly formed star. Eventually, the temperature becomes so high that nuclear reactions are initiated inside the star. Hydrogen, which is the primary constituent, burns to form helium. During this process pressure builds up which prevents further contraction, the star stabilizes, and may continue to exist for millions or thousands of millions of years. When the supply of hydrogen has been used up, other nuclear reactions come into play, especially in more massive stars, and heavier elements are thus formed. A particularly effective type of nuclear reaction is the successive addition of neutrons. Finally, the star is, to a large extent, composed of heavier elements, mainly iron and neighbouring elements, and the supply of nuclear fuel is exhausted. When the star has evolved this far it can no longer withstand the pressure of its own gravitational force and collapses, the product of collapse depending on the mass of the star.

For lighter stars with a mass roughly equal to that of the Sun, the collapse results in a so-called white dwarf. The star is so named because of its reduction in size, leading to an increase in its density to about 10 tons per cubic centimetre. The mechanism for the collapse is that the electron shell structure is crushed, so that the star consists of atomic nuclei in an electron gas.

For somewhat heavier stars, the collapse can lead to an explosion, the visible

result being a supernova. This is accompanied by a short-lived but intense neutron flux which leads to the formation of the heaviest elements. In these heavy stars the collapse can go even further, the atomic nuclei and the electrons combining to form neutrons. This results in a so-called neutron star which has the enormously high density of 100 million tons per cubic centimetre. A star with a mass of 1 to 2 times that of the Sun may be compressed so that the radius is only about 10 km. A neutron star is essentially a sphere of neutrons in a fluid form surrounded by a solid crust which is very much harder than steel.

The collapse of still heavier stars can lead to an even more exotic object, a black hole. Here the gravitational force is so strong that all matter which is sucked into the hole loses its identity, and is compressed into an infinitely small volume, i.e. a mathematical point. Not even light, emitted from within the black hole, may escape into the outside world, hence the name, black hole. The existence of a black hole may be revealed through the radiation which is emitted by matter which, when being sucked into it, undergoes a considerable increase in temperature before finally disappearing. Certain strange objects called quasars may possibly be a black hole in the centre of a galaxy.

It should now be clear that during their evolution stars exhibit many different physical processes of fundamental importance. Many scientists have studied the problems involved with these processes, but especially important contributions have been made by Subrahmanyan Chandrasekhar and William Fowler.

Chandrasekhar's work is particularly many-sided and covers many aspects of the evolution of stars. An important part of his work is a study concerning the problems of stability in different phases of their evolution. In recent years he has studied relativistic effects, which become important because of the extreme conditions which arise during the later stages of the star's development. One of Chandrasekhar's most well known contributions is his study of the structure of white dwarfs. Even if some of these studies are from his earlier years, they have become topical again through advances in the fields of astronomy and space research.

Fowler's work deals with the nuclear reactions which take place during the evolution of stars. Apart from generating the energy which is emitted, they are important because they lead to the production of the chemical elements from the starting material, which mainly consists of the lightest element, hydrogen. Not only has Fowler carried out a great deal of experimental work on nuclear reactions of interest in the astrophysical context, but has also worked on this problem from a theoretical point of view. In the 1950's, together with a number of colleagues, he developed a complete theory for the formation of the chemical elements in the Universe. This theory is still the basis of our knowledge in this area, and the latest advances in nuclear physics and space research have further shown this theory to be correct.

Professor Chandrasekhar and Professor Fowler,

Your pioneering work has laid the foundation for important developments in astrophysics and you have both been the source of inspiration for other scientists working in this field. The remarkable achievements of astronomy and

space research in recent years have vindicated your ideas and demonstrated their importance.

It is my privilege and pleasure to convey to you the warmest congratulations of the Royal Swedish Academy of Sciences. May I now ask you to come forward and receive your prize from the hands of His Majesty the King.

S. Chandrasekhar

SUBRAHMANYAN CHANDRASEKHAR

I was born in Lahore (then a part of British India) on the 19th of October 1910, as the first son and the third child of a family of four sons and six daughters. My father, Chandrasekhara Subrahmanya Ayyar, an officer in Government Service in the Indian Audits and Accounts Department, was then in Lahore as the Deputy Auditor General of the Northwestern Railways. My mother, Sita *(neé* Balakrishnan) was a woman of high intellectual attainments (she translated into Tamil, for example, Henrik Ibsen's A *Doll House),* was passionately devoted to her children, and was intensely ambitious for them.

My early education, till I was twelve, was at home by my parents and by private tuition. In 1918, my father was transferred to Madras where the family was permanently established at that time.

In Madras, I attended the Hindu High School, Triplicane, during the years 1922-25. My university education (1925-30) was at the Presidency College. I took *my* bachelor's degree, B.Sc. (Hon.), in physics in June 1930. In July of that year, I was awarded a Government of India scholarship for graduate studies in Cambridge, England. In Cambridge, I became a research student under the supervision of Professor R. H. Fowler (who was also responsible for my admission to Trinity College). On the advice of Professor P. A. M. Dirac, I spent the third of my three undergraduate years at the Institut for Teoretisk Fysik in Copenhagen.

I took my Ph.D. degree at Cambridge in the summer of 1933. In the following October, I was elected to a Prize Fellowship at Trinity College for the period 1933-37. During my Fellowship years at Trinity, I formed lasting friendships with several, including Sir Arthur Eddington and Professor E. A. Milne.

While on a short visit to Harvard University (in Cambridge, Massachusetts), at the invitation of the then Director, Dr. Harlow Shapley, during the winter months (January-March) of 1936, I was offered a position as a Research Associate at the University of Chicago by Dr. Otto Struve and President Robert Maynard Hutchins. I joined the faculty of the University of Chicago in January 1937. And I have remained at this University ever since.

During my last two years (1928-30) at the Presidency College in Madras, I formed a friendship with Lalitha Doraiswamy, one year my junior. This friendship matured; and we were married (in India) in September 1936 prior to my joining the University of Chicago. In the sharing of our lives during the past forty-seven years, Lalitha's patient understanding, support, and encouragement have been the central facts of my life.

After the early preparatory years, my scientific work has followed a certain

pattern motivated, principally, by a quest after perspectives. In practise, this quest has consisted in my choosing (after some trials and tribulations) a certain area which appears amenable to cultivation and compatible with my taste, abilities, and temperament. And when after some years of study, I feel that I have accumulated a sufficient body of knowledge and achieved a view of my own, I have the urge to present my point of view, ab initio, in a coherent account with order, form, and structure.

There have been seven such periods in my life: stellar structure, including the theory of white dwarfs (1929- 1939); stellar dynamics, including the theory of Brownian motion (1938- 1943); the theory of radiative transfer, including the theory ofstellar atmospheres and the quantum theory of the negative ion of hydrogen and the theory of planetary atmospheres, including the theory of the illumination and the polarization of the sunlit sky (1943- 1950); hydrodynamic and hydromagnetic stability, including the theory of the Rayleigh-Bernard convection (1952- 1961); the equilibrium and the stability of ellipsoidal figures of equilibrium, partly in collaboration with Norman R. Lebovitz (1961- 1968); the general theory of relativity and relativistic astrophysics (1962- 1971); and the mathematical theory of black holes (1974- 1983). The monographs which resulted from these several periods are:

1. An Introduction to the Study of Stellar Structure (1939, University of Chicago Press; reprinted by Dover Publications, Inc., 1967).

2a. Principles of Stellar Dynamics (1943, University of Chicago Press; reprinted by Dover Publications, Inc., 1960).

2b. 'Stochastic Problems in Physics and Astronomy', *Reviews of Modern Physics,* 15, l-89 (1943); reprinted in *Selected Papers on Noise and Stochastic Processes* by Nelson Wax, Dover Publications, Inc., 1954.

3. Radiative Transfer (1950, Clarendon Press, Oxford; reprinted by Dover Publications, Inc., 1960).

4. Hydrodynamic and Hydromagnetic Stability (1961, Clarendon Press, Oxford; reprinted by Dover Publications, Inc., 1981).

5. Ellipsoidal Figures of Equilibrium (1968; Yale University Press).

6. The Mathematical Theory of Black Holes (1983, Clarendon Press, Oxford).

However, the work which appears to be singled out in the citation for the award of the Nobel Prize is included in the following papers:

'The highly collapsed configurations of a stellar mass', *Mon. Not. Roy. Astron.* Soc., 91, 456-66 (1931).

'The maximum mass of ideal white dwarfs', *Astrophys. J., 74,* 81-2 (1931).

'The density of white dwarf stars', *Phil. Mag.,* 11, 592-96 (1931).

'Some remarks on the state of matter in the interior of stars', Z.f. *Astrophysik,* 5, 321-27 (1932).

'The physical state of matter in the interior of stars', *Observatory, 57, 93-9* (1934).

'Stellar configurations with degenerate cores', *Observatory, 57, 373-77* (1934).

'The highly collapsed configurations of a stellar mass' (second paper), *Mon. Not. Roy.* Astron. Soc., 95, 207-25 (1935).

'Stellar configurations with degenerate cores', *Mon. Not. Roy, Astron. Soc., 95,* 226-60 (1935).

'Stellar configurations with degenerate cores' (second paper), *Mon. Not. Roy. Astron. Soc., 95, 676-93* (1935).

'The pressure in the interior of a star', *Mon. Not. Roy. Astron. Soc., 96, 644-47* (1936).

'On the maximum possible central radiation pressure in a star of a given mass', *Observatory, 59, 47-8* (1936).

'Dynamical instability of gaseous masses approaching the Schwarzschild limit in general relativity', *Phys. Rev. Lett.,* 12, 114- 16 (1964); Erratum, *Phys. Rev. Lett.,* 12, 437-38 (1964).

'The dynamical instability of the white-dwarf configurations approaching the limiting mass' (with Robert F. Tooper), *Astrophys. J.,* 139, 1396-98 (1964).

'The dynamical instability of gaseous masses approaching the Schwarzschild limit in general relativity', *Astrophys. J.,* 140, 417-33 (1964).

'Solutions of two problems in the theory ofgravitational radiation', *Phys. Rev. Lett.,* 24, 611- 15 (1970); Erratum, *Phys. Rev. Lett., 24, 762* (1970).

'The effect of graviational radiation on the secular stability of the Maclaurin spheroid', *Astrophys.J.,* 161, 561-69 (1970).

ON STARS, THEIR EVOLUTION AND THEIR STABILITY

Nobel lecture, 8 December, 1983

by

SUBRAHMANYAN CHANDRASEKHAR

The University of Chicago, Chicago, Illinois 60637, USA

1. *Introduction*

When we think of atoms, we have a clear picture in our minds: a central nucleus and a swarm of electrons surrounding it. We conceive them as small objects of sizes measured in Angstroms ($\sim 10^{-8}$ cm); and we know that some hundred different species of them exist. This picture is, of course, quantified and made precise in modern quantum theory. And the success of the entire theory may be traced to two basic facts: *first,* the Bohr radius of the ground state of the hydrogen atom, namely,

$$\frac{h^2}{4\pi^2 me^2} \sim 0.5 \times 10^{-8} \text{ cm}, \tag{1}$$

where h is Planck's constant, m is the mass of the electron and e is its charge, provides a correct measure of atomic dimensions; and *second,* the reciprocal of *Sommerfeld's fine-structure constant,*

$$\frac{hc}{2\pi e^2} \sim 137, \tag{2}$$

gives the maximum positive charge of the central nucleus that will allow a stable electron-orbit around it. This maximum charge for the central nucleus arises from the effects of special relativity on the motions of the orbiting electrons.

We now ask: can we understand the basic facts concerning stars as simply as we understand atoms in terms of the two combinations of natural constants (1) and (2). In this lecture, I shall attempt to show that in a limited sense we can.

The most important fact concerning a star is its mass. It is measured in units of the mass of the sun, $\odot$, which is 2×10^{33} gm: stars with masses very much less than, or very much more than, the mass of the sun are relatively infrequent. The current theories of stellar structure and stellar evolution derive their successes largely from the fact that the following combination of the dimensions of a mass provides a correct measure of stellar masses:

$$\left(\frac{hc}{G}\right)^{3/2} \frac{1}{H^2} \simeq 29.2\,\odot, \tag{3}$$

where G is the constant of gravitation and H is the mass of the hydrogen atom. In the first half of the lecture, I shall essentially be concerned with the question: how does this come about?

2. The role of radiation pressure

A central fact concerning normal stars is the role which radiation pressure plays as a factor in their hydrostatic equilibrium. Precisely the equation governing the hydrostatic equilibrium of a star is

$$\frac{dP}{dr} = -\frac{GM(r)}{r^2}\rho, \tag{4}$$

where P denotes the total pressure, p the density, and $M(r)$ is the mass interior to a sphere of radius r. There are two contributions to the total pressure P : that due to the material and that due to the radiation. On the assumption that the matter is in the state of a perfect gas in the classical Maxwellian sense, the material or the gas pressure is given by

$$p_{\text{gas}} = \frac{k}{\mu H}\rho T, \tag{5}$$

where T is the absolute temperature, k is the Boltzmann constant, and μ is the mean molecular weight (which under normal stellar conditions is ~ 1.0). The pressure due to radiation is given by

$$p_{\text{rad}} = \frac{1}{3}\alpha T^4, \tag{6}$$

where α denotes Stefan's radiation-constant. Consequently, if radiation contributes a fraction (1–β) to the total pressure, we may write

$$P = \frac{1}{1-\beta}\frac{1}{3}\alpha T^4 = \frac{1}{\beta}\frac{k}{\mu H}\rho T. \tag{7}$$

To bring out explicitly the role of the radiation pressure in the equilibrium of a star, we may eliminate the temperature, T, from the foregoing equations and express P in terms of p and β instead of in terms of p and T. We find:

$$T = \left(\frac{k}{\mu H}\frac{3}{\alpha}\frac{1-\beta}{\beta}\right)^{1/3}\rho^{1/3} \tag{8}$$

and

$$P = \left[\left(\frac{k}{\mu H}\right)^4\frac{3}{\alpha}\frac{1-\beta}{\beta^4}\right]^{1/3}\rho^{4/3} = C(\beta)\rho^{4/3}\ (\text{say}). \tag{9}$$

The importance of this ratio, (1–β), for the theory of stellar structure was first emphasized by Eddington. Indeed, he related it, in a famous passage in his book on *The Internal Constitution of the Stars,* to the 'happening of the stars'.[1] A more rational version of Eddington's argument which, at the same time, isolates the combination (3) of the natural constants is the following:

There is a general theorem[2] which states that the pressure, P_c, at the centre of a star of a mass M in hydrostatic equilibrium in which the density, $p(r)$, at a point at a radial distance, r, from the centre does not exceed the mean density, $\bar{\rho}(r)$, interior to the same point r, must satisfy the inequality,

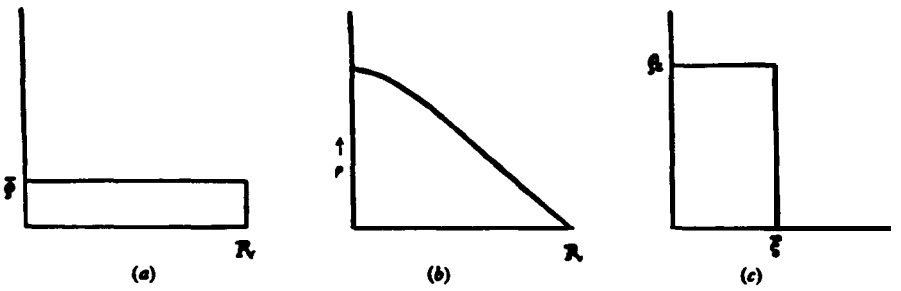


Fig. 1. A comparison of an inhomogeneous distribution of density in a star (b) with the two homogeneous configurations with the constant density equal to the mean density (a) and equal to the density at the centre (c).

$$\frac{1}{2}G\left(\frac{4}{3}\pi\right)^{1/3}\bar{\rho}^{4/3}M^{2/3} \leqslant \rho_c \leqslant \frac{1}{2}G\left(\frac{4}{3}\pi\right)^{1/3}\rho_c^{4/3}M^{2/3}, \tag{10}$$

where $\bar{\rho}$ denotes the mean density of the star and ρ_c its density at the centre. The content of the theorem is no more than the assertion that the actual pressure at the centre of a star must be intermediate between those at the centres of the two configurations ofuniform density, one at a density equal to the mean density of the star, and the other at a density equal to the density p_c at the centre (see Fig. 1). If the inequality (10) should be violated then there must, in general, be some regions in which adverse density gradients must prevail; and this implies instability. In other words, we may consider conformity with the inequality (10) as equivalent to the condition for the stable existence of stars.

The right-hand side of the inequality (10) together with P given by equation (9), yields, for the stable existence of stars, the condition,

$$\left[\left(\frac{k}{\mu H}\right)^4\frac{3}{a}\frac{1-\beta_c}{\beta_c^4}\right]^{1/3} \leqslant \left(\frac{\pi}{6}\right)^{1/3}GM^{2/3}, \tag{11}$$

or, equivalently,

$$M \geqslant \left(\frac{6}{\pi}\right)^{1/2}\left[\left(\frac{k}{\mu H}\right)^4\frac{3}{a}\frac{1-\beta_c}{\beta_c^4}\right]^{1/2}\frac{1}{G^{3/2}}, \tag{12}$$

where in the foregoing inequalities, β_c is a value of β at the centre of the star. Now Stefan's constant, a, by virtue of Planck's law, has the value

$$a = \frac{8\pi^5k^4}{15h^3c^3}. \tag{13}$$

Inserting this value a in the inequality (12) we obtain

$$\mu^2M\left(\frac{\beta_c^4}{1-\beta_c}\right)^{1/2} \geqslant \frac{(135)^{1/2}}{2\pi^3}\left(\frac{hc}{G}\right)^{3/2}\frac{1}{H^2} = 0.1873\left(\frac{hc}{G}\right)^{3/2}\frac{1}{H^2}. \tag{14}$$

We observe that the inequality (14) has isolated the combination (3) of

natural constants of the dimensions of a mass; by inserting its numerical value given in equation (3) we obtain the inequality,

$$\mu^2 M \left(\frac{\beta_c^4}{1-\beta_c}\right)^{1/2} \geq 5.48\odot. \tag{15}$$

This inequality provides an upper limit to $(1-\beta_c)$ for a star of a given mass. Thus,

$$1-\beta_c \leqslant 1-\beta_*, \tag{16}$$

where $(1-\beta_*)$ is uniquely determined by the mass M of the star and the mean molecular weight, μ, by the quartic equation,

$$\mu^2 M = 5.48 \left(\frac{1-\beta_*}{\beta_*^4}\right)^{1/2} \odot. \tag{17}$$

In Table 1, we list the values of $1-\beta_*$ for several values of $\mu^2 M$. From this table it follows in particular, that for a star of solar mass with a mean molecular weight equal to 1, the radiation pressure at the centre cannot exceed 3 percent of the total pressure.

Table 1

The maximum radiation pressure, $(1-\beta_*)$, at the centre of a star of a given mass, M.

$1-\beta_*$	$M\mu^2/\odot$	$1-\beta_*$	$M\mu^2/\odot$
0.01	0.56	0.50	15.49
.03	1.01	.60	26.52
.10	2.14	.70	50.92
.20	3.83	.80	122.5
.30	6.12	.85	224.4
0.40	9.62	0.90	519.6

What do we conclude from the foregoing calculation? We conclude that to the extent equation (17) is at the base of the equilibrium of actual stars, to that extent the combination of natural constants (3), providing a mass of proper magnitude for the measurement of stellar masses, is at the base of a physical theory of stellar structure.

3. *Do stars have enough energy to cool?*

The same combination of natural constants (3) emerged soon afterward in a much more fundamental context of resolving a paradox Eddington had formulated in the form of an aphorism: 'a star will need energy to cool.' The paradox arose while considering the ultimate fate of a gaseous star in the light of the then new knowledge that white-dwarf stars, such as the companion of Sirius, exist, which have mean densities in the range 10^5-10^7 gm cm^{-3}. As Eddington stated[3]

> I do not see how a star which has once got into this compressed state is ever going to get out of it.. It would seem that the star will be in an awkward predicament when its supply of subatomic energy fails.

The paradox posed by Eddington was reformulated in clearer physical terms by R. H. Fowler.[4] His formulation was the following:

> The stellar material, in the white-dwarf state, will have radiated so much energy that it has less energy than the same matter in normal atoms expanded at the absolute zero of temperature. If part of it were removed from the star and the pressure taken off, what could it do?

Quantitatively, Fowler's question arises in this way.

An estimate of the electrostatic energy, E_V, per unit volume of an assembly of atoms, of atomic number Z, ionized down to bare nuclei, is given by

$$E_v = 1.32 \times 10^{11} Z^2 \rho^{4/3}, \tag{18}$$

while the kinetic energy ofthermal motions, E_{kin}, per unit volume of free particles in the form of a perfect gas of density, ρ, and temperature, T, is given by

$$E_{\text{kin}} = \frac{3}{2}\frac{k}{\mu H}\rho\, T = \frac{1.24 \times 10^8}{\mu}\rho\, T. \tag{19}$$

Now if such matter were released of the pressure to which it is subject, it can resume a state of ordinary normal atoms only if

$$E_{\text{kin}} > E_V, \tag{20}$$

or, according to equations (18) and (19), only if

$$p < \left(0.94 \times 10^{-3} \frac{T}{\mu Z^2}\right)^{\frac{3}{2}}. \tag{21}$$

This inequality will be clearly violated if the density is sufficiently high. This is the essence of Eddington's paradox as formulated by Fowler. And Fowler resolved this paradox in 1926 in a paper' entitled 'Dense Matter' - one of the great landmark papers in the realm ofstellar structure: in it the notions of Fermi statistics and of electron degeneracy are introduced for the first time.

4. Fowler's resolution of Eddington's paradox; the degeneracy of the electrons in white-dwarf stars

In a completely degenerate electron gas all available parts of the phase space, with momenta less than a certain 'threshold' value p_0 - the Fermi 'threshold' - are occupied consistently with the Pauli exclusion-principle i.e., with two electrons per 'cell' of volume h^3 of the six-dimensional phase space. Therefore,

if $n(p)\,dp$ denotes the number of electrons, per unit volume, between p and $p+dp$, then the assumption-of complete degeneracy is equivalent to the assertion,

$$\left.\begin{aligned} n(p) &= \frac{8\pi}{h^3}p^2 \qquad (p \leqslant p_0), \\ &= 0 \qquad\quad (p > p_0). \end{aligned}\right\} \tag{22}$$

old momentum p_0, is determined by

The value of the threshold momentum p_0, is determined by the normalization condition

$$n = \int_0^{p_0} n(p)\,dp = \frac{8\pi}{3h^3}p_0^3, \tag{23}$$

where n denotes the total number of electrons per unit volume.

For the distribution given by (22), the pressure P and the kinetic energy E_{kin} of the electrons (per unit volume), are given by

$$P = \frac{8\pi}{3h^3}\int_0^{p_0} p^3 v_p\,dp \tag{24}$$

and

$$E_{\text{kin}} = \frac{8\pi}{h^3}\int_0^{p_0} p^2\,T_p\,dp, \tag{25}$$

where v_p and T_p are the velocity and the kinetic energy of an electron having a momentum p.

If we set

$$v_p = p/m \text{ and } T_p = p^2/2m, \tag{26}$$

appropriate for non-relativistic mechanics, in equations (24) and (25), we find

$$P = \frac{8\pi}{15h^3m}p_0^5 = \frac{1}{20}\left(\frac{3}{\pi}\right)^{2/3}\frac{h^2}{m}\,n^{5/3} \tag{27}$$

and

$$E_{\text{kin}} = \frac{8\pi}{10h^3m}p_0^5 = \frac{3}{40}\left(\frac{3}{\pi}\right)^{2/3}\frac{h^2}{m}\,n^{5/3}. \tag{28}$$

Fowler's resolution of Eddington's paradox consists in this: at the temperatures and densities that may be expected to prevail in the interiors of the white-dwarf stars, the electrons will be highly degenerate and E_{kin} must be evaluated in accordance with equation (28) and not in accordance with equation (19); and equation (28) gives,

$$E_{\text{kin}} = 1.39\times10^{13}\,(\rho/\mu)^{5/3}. \tag{29}$$

Comparing now the two estimates (18) and (29), we see that, for matter of the density occurring in the white dwarfs, namely $\rho \approx 10^5$ gm cm^{-3}, the total kinetic energy is about two to four times the negative potential-energy; and Eddington's

paradox does not arise. Fowler concluded his paper with the following highly perceptive statement:

> The black-dwarf material is best likened to a single gigantic molecule in its lowest quantum state. On the Fermi-Dirac statistics, its high density can be achieved in one and only one way, in virtue of a correspondingly great energy content. But this energy can no more be expended in radiation than the energy of a normal atom or molecule. The only difference between black-dwarf matter and a normal molecule is that the molecule can exist in a free state while the black-dwarf matter can only so exist under very high external pressure.

5. *The theory of the white-dwarf stars; the limiting mass*

The internal energy (= 3 *P/2)* of a degenerate electron gas that is associated with a pressure *P* is *zero-point energy;* and the essential content of Fowler's paper is that this zero-point energy is so great that we may expect a star to eventually settle down to a state in which all of its energy is of this kind. Fowler's argument can be more explicitly formulated in the following manner.[5]

According to the expression for the pressure given by equation (27), we have the relation,

$$P = K_1\rho^{5/3} \text{ where } K_1 = \frac{1}{20}\left(\frac{3}{\pi}\right)^{2/3}\frac{h^2}{m(\mu_e H)^{5/3}}, \tag{30}$$

where μ_e is the mean molecular weight per electron. An equilibrium configuration in which the pressure, *P,* and the density ρ, are related in the manner,

$$P = K\rho^{1+1/n}, \tag{31}$$

is an *Emden polytrope* of index n. The degenerate configurations built on the equation of state (30) are therefore polytropes of index 3/2; and the theory of polytropes immediately provides the relation,

$$K_1 = 0.4242\ (GM^{1/3}\ R) \tag{32}$$

or, numerically, for K_1 given by equation (30),

$$\log_{10}(R/R_\odot) = -\frac{1}{3}\log_{10}(M/\odot) - \frac{5}{3}\log_{10}\mu_e - 1.397. \tag{33}$$

For a mass equal to the solar mass and μ_e = 2, the relation (33) predicts R = 1.26 x $10^{-2}R_\odot$ and a mean density of 7.0 x 10^5 g m / c m^3. These values are precisely of the order of the radii and mean densities encountered in white-dwarf stars. Moreover, according to equations (32) and (33), the radius of the white-dwarf configuration is inversely proportional to the cube root of the mass. On this account, finite equilibrium configurations are predicted for all masses. And it came to be accepted that the white-dwarfs represent the last stages in the evolution of all stars.

But it soon became clear that the foregoing simple theory based on Fowler's premises required modifications. For the electrons, at their threshold energies at the centres of the degenerate stars, begin to have velocities comparable to that of light as the mass increases. Thus, already for a degenerate star of solar mass (with $\mu_e = 2$) the central density (which is about six times the mean density) is 4.19×10^6 gm / cm^3; and this density corresponds to a threshold momentum $p_0 = 1.29\, mc$ and a velocity which is 0.63 c. Consequently, the equation of state must be modified to take into account the effects of special relativity. And this is easily done by inserting in equations (24) and (25) the relations,

$$v_p = \frac{p}{m\,(1+p^2/m^2c^2)^{1/2}} \text{ and } T_p = mc^2\left[(1+p^2/m^2c^2)^{1/2}-1\right], \tag{34}$$

in place of the non-relativistic relations (26). We find that the resulting equation of state can be expressed, parametrically, in the form

$$P = A f(x) \text{ and } \rho = Bx^3, \tag{35}$$

where

$$A = \frac{\pi m^4 c^5}{3h^3},\ B = \frac{8\pi m^3 c^3 \mu_e H}{3h^3} \tag{36}$$

and

$$f(x) = x(x^2+1)^{1/2}\,(2x^2-3)+3\sinh^{-1} x. \tag{37}$$

And similarly

$$E_{\text{kin}} = Ag(x), \tag{38}$$

where

$$g(x) = 8x^3\left[(x^2+1)^{1/2}-1\right]-f(x). \tag{39}$$

According to equations (35) and (36), the pressure approximates the relation (30) for low enough electron concentrations $(x \ll 1)$; but for increasing electron concentrations $(x \gg 1)$, the pressure tends to[6]

$$P = \frac{1}{8}\left(\frac{3}{\pi}\right)^{1/3} hc\, n^{4/3}. \tag{40}$$

This limiting form of relation can be obtained very simply by setting $v_p = c$ in equation (24); then

$$P = \frac{8\pi c}{3h^3}\int_0^{p_0} p^3\, dp = \frac{2\pi c}{3h^3}p_0^4; \tag{41}$$

and the elimination of p_0 with the aid of equation (23) directly leads to equation (40).

While the modification of the equation of state required by the special

theory of relativity appears harmless enough, it has, as we shall presently show, a dramatic effect on the predicted mass-radius relation for degenerate configurations.

The relation between P and ρ corresponding to the limiting form (41) is

$$P = K_2 \rho^{4/3} \text{ where } K_2 = \frac{1}{8}\left(\frac{3}{\pi}\right)^{1/3} \frac{hc}{(\mu_e H)^{4/3}}. \tag{42}$$

In this limit, the configuration is an Emden polytrope of index 3. And it is well known that when the polytropic index is 3, the mass of the resulting equilibrium configuration is uniquely determined by the constant of proportionality, K_2, in the pressure-density relation. We have accordingly,

$$M_{\text{limit}} = 4\pi\left(\frac{K_2}{\pi G}\right)^{3/2}(2.018) = 0.197\left(\frac{hc}{G}\right)^{3/2}\frac{1}{(\mu_e H)^2} = 5.76\mu_e^{-2}\odot. \tag{43}$$

(In equation (43), 2.018 is a numerical constant derived from the explicit solution of the Lane-Emden equation for $n = 3$.)

It is clear from general considerations' that *the exact mass-radius relation for the degenerate configurations must provide an upper limit to the mass of such configurations given by equation (43); and further, that the mean density of the configuration must tend to infinity,* while the radius tends to zero, and $M \to M_{\text{limit}}$. These conditions, straightforward as they are, can be established directly by considering the equilibrium of configurations built on the exact equation of state given by equations (35) - (37). It is found that the equation governing the equilibrium of such configurations can be reduced to the form[8,9]

$$\frac{1}{\eta^2}\frac{d}{d\eta}\left(\eta^2\frac{d\phi}{d\eta}\right) = -\left(\phi^2 - \frac{1}{y_0^2}\right)^{3/2}, \tag{44}$$

where

$$y_0^2 = x_0^2 + 1, \tag{45}$$

and mcx_0 denotes the threshold momentum of the electrons at the centre of the configuration and η measures the radial distance in the unit

$$\left(\frac{2A}{\pi G}\right)^{1/2}\frac{1}{By_0} = l_1 y_0^{-1} \text{ (say)} \tag{46}$$

By integrating equation (44), with suitable boundary conditions and for various initially prescribed values of y_0, we can derive the exact mass-radius relation, as well as the other equilibrium properties, of the degenerate configurations. The principal results of such calculations are illustrated in Figures 2 and 3.

The important conclusions which follow from the foregoing considerations are: *first*, there is an upper limit, M_{limit}, to the mass of stars which can become degenerate configurations, as the last stage in their evolution; and *second*, that stars with $M > M_{\text{limit}}$ must have end states which cannot be predicted from the considerations we have presented so far. And finally, we observe that the

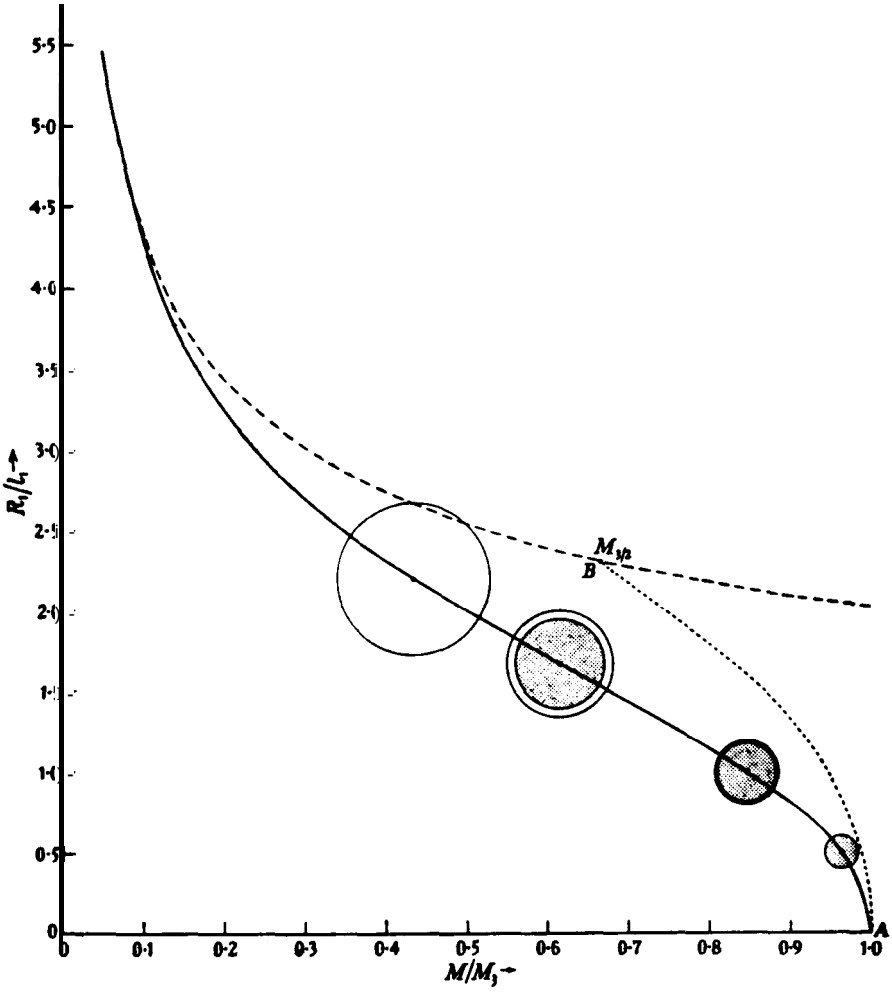


Fig. 2. The full-line curve represents the exact (mass-radius)-relation (l_1, is defined in equation (46) and M_3 denotes the limiting mass). This curve tends asymptotically to the ---- curve appropriate to the low-mass degenerate configurations, approximated by polytropes of index 3/2. The regions of the configurations which may be considered as relativistic ($\varrho > (K_1/K_2)^3$) are shown shaded. (From Chandrasekhar, S., Mon. *Not. Roy. Astr. Soc., 95, 207* (1935).)

combination of the natural constant (3) now emerges in the fundamental context of M_{limit} given by equation (43): its significance for the theory of stellar structure and stellar evolution can no longer be doubted.

6. Under what conditions can normal stars develop degenerate cores?

Once the upper limit to the mass of completely degenerate configurations had been established, the question that required to be resolved was how to relate its existence to the evolution of stars from their gaseous state. If a star has a mass less than M_{limit} the assumption that it will eventually evolvetowards the completely degenerate state appears reasonable. But what if its mass is greater than $M_{limit?}$

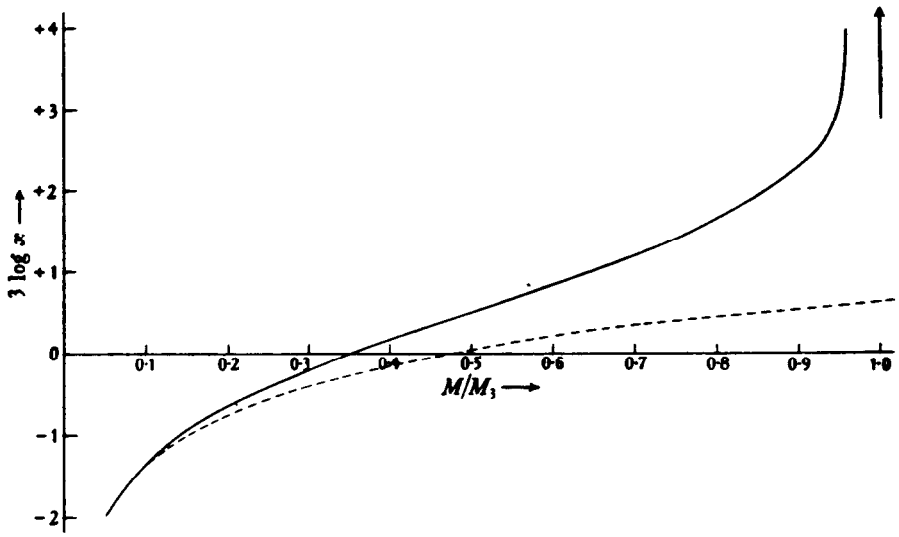


Fig. 3. The full-line curve represents the exact (mass-density)-relation for the highly collapsed configurations. This curve tends asymptotically to the dotted curve as $M \to 0$. (From Chandrasekhar, S., *Mon. Not. Roy. Astr. Soc., 95, 207* (1935).)

Clues as to what might ensue were sought in terms of the equations and inequalities of §§2 and 3.[10,11]

The first question that had to be resolved concerns the circumstances under which a star, initially gaseous, will develop degenerate cores. From the physical side, the question, when departures from the perfect-gas equation of state (5) will set in and the effects of electron degeneracy will be manifested, can be readily answered.

Suppose, for example, that we continually and steadily increase the density, at constant temperature, of an assembly of free electrons and atomic nuclei, in a highly ionized state and initially in the form of a perfect gas governed by the equation of state (5). At first the electron pressure will increase linearly with p; but soon departures will set in and eventually the density will increase in accordance with the equation of state that describes the fully degenerate electron-gas (see Fig. 4). The remarkable fact is that this limiting form of the equation of state is independent of temperature.

However, to examine the circumstances when, during the course of evolution, a star will develop degenerate cores, it is more convenient to express the electron pressure (as given by the classical perfect-gas equation of state) in terms of p and β_e defined in the manner (cf. equation (7)).

$$p_e = \frac{k}{\mu_e H}\rho T = \frac{\beta_e}{1-\beta_e}\frac{1}{3}\alpha T^4, \tag{47}$$

where p_e now denotes the electron pressure. Then, analogous to equation (9), we can write

$$p_e = \left[\left(\frac{k}{\mu_e H}\right)^4 \frac{3}{\alpha}\frac{1-\beta_e}{\beta_e}\right]^{1/3} \rho^{4/3}. \tag{48}$$

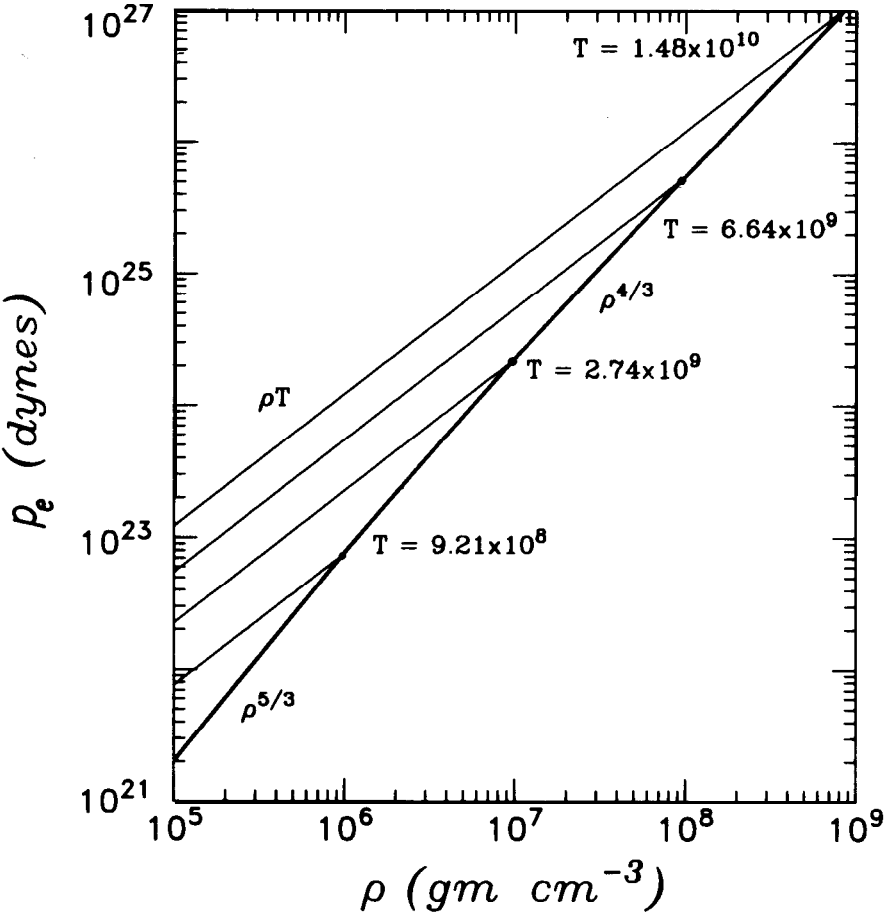


Fig. 4. Illustrating how by increasing the density at constant temperature degeneracy always sets in.

Comparing this with equation (42), we conclude that if

$$\left[\left(\frac{k}{\mu_e H}\right)^4 \frac{3}{\alpha}\frac{1-\beta_e}{\beta_e}\right]^{1/3} > K_2 = \frac{1}{8}\left(\frac{3}{\pi}\right)^{1/3}\frac{hc}{(\mu_e H)^{4/3}}, \tag{49}$$

the pressure p_e given by the classical perfect-gas equation of state will be greater than that given by the equation if degeneracy were to prevail, not only for the prescribed ρ and T, but for *all* p and T *having the same* β_e.

Inserting for a its value given in equation (13), we find that the inequality (49) reduces to

$$\frac{960}{\pi^4}\frac{1-\beta_e}{\beta_e} > 1, \tag{50}$$

or equivalently,

$$1-\beta_e > 0.0921 = 1-\beta_\omega \text{ (say)}. \tag{51}$$

(See Fig. 5)

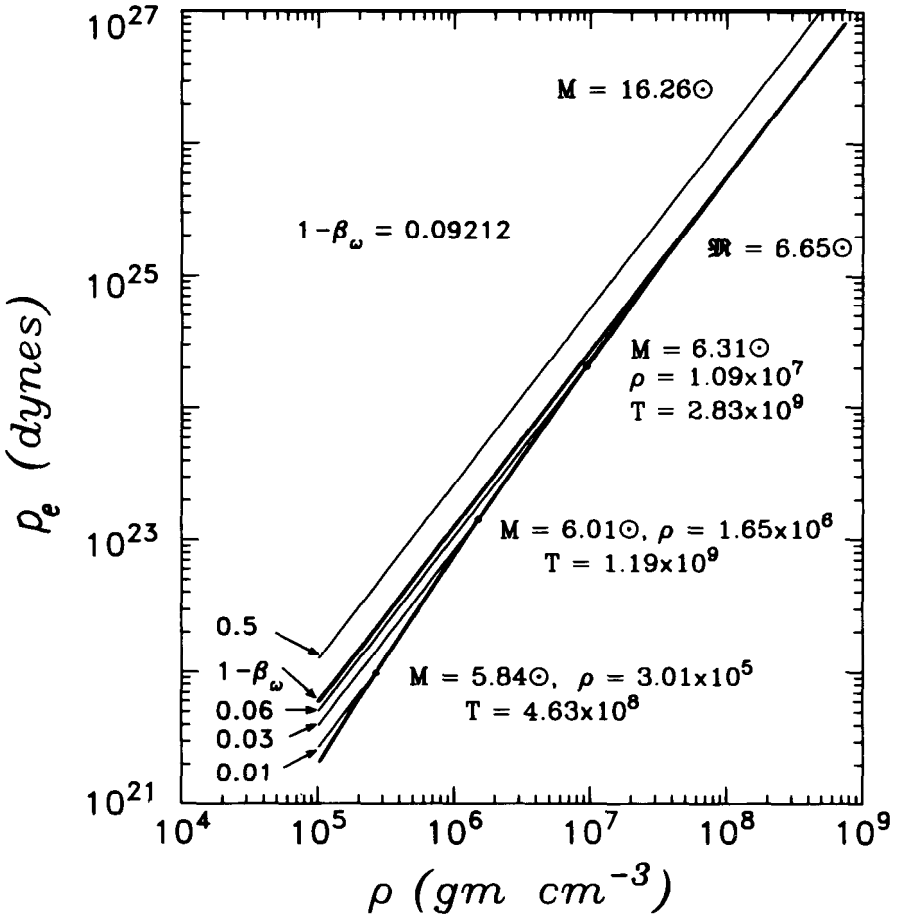


Fig. 5. Illustrating the onset of degeneracy for increasing density at constant β. Notice that there are no intersections for β> 0.09212. In the figure, 1–β is converted into mass of a star built on the standard model.

For our present purposes, the principal content of the inequality (51) is the criterion that for a star to develop degeneracy, it is necessary that the radiation pressure be less than 9.2 percent of (p_e+p_{rad}). This last inference is so central to all current schemes of stellar evolution that the directness and the simplicity of the early arguments are worth repeating.

The two principal elements of the early arguments were these: first, that radiation pressure becomes increasingly dominant as the mass of the star increases; and *second,* that the degeneracy of electrons is possible only so long as the radiation pressure is not a significant fraction of the total pressure - indeed, as we have seen, it must not exceed 9.2 percent of (p_e+p_{rad}). The second of these elements in the arguments is a direct and an elementary consequence of the physics of degeneracy; but the first requires some amplification.

That radiation pressure must play an increasingly dominant role as the mass of the star increases is one of the earliest results in the study of stellar structure that was established by Eddington. A quantitative expression for this fact is

given by Eddington's *standard model* which lay at the base of early studies summarized in his *The Internal Constitution of the Stars.*

On the standard model, the fraction β (= gas pressure/total pressure) is a constant through a star. On this assumption, the star is a polytrope of index 3 as is apparent from equation (9); and, in consequence, we have the relation (cf. equation (43))

$$M = 4\pi \left[\frac{C(\beta)}{\pi G}\right]^{3/2} (2.018) \tag{52}$$

where $C(\beta)$ is defined in equation (9). Equation (52) provides a quartic equation for β analogous to equation (17) for β^*. Equation (52) for $\beta = \beta_\omega$ gives

$$M = 0.197\, \beta_\omega^{-3/2} \left(\frac{hc}{G}\right)^{3/2} \frac{1}{(\mu H)^2} = 6.65\, \mu^{-2} \odot = \boldsymbol{M} \text{ (say)}. \tag{53}$$

On the standard model, then stars with masses exceeding M will have radiation pressures which exceed 9.2 percent of the total pressure. Consequently stars with $M > M$ cannot, at any stage during the course of their evolution, develop degeneracy in their interiors. Therefore, for such stars an eventual white-dwarf state is not possible unless they are able to eject a substantial fraction of their mass.

The standard model is, of course, only a model. Nevertheless, except under special circumstances, briefly noted below, experience has confirmed the standard model, namely that the evolution of stars of masses exceeding 7-8 $\odot$ must proceed along lines very different from those of less massive stars. These conclusions, which were arrived at some fifty years ago, appeared then so convincing that assertions such as these were made with confidence:

> Given an enclosure containing electrons and atomic nuclei (total charge zero) what happens if we go on compressing the material indefinitely? (1932)[10]
>
> The life history of a star of small mass must be essentially different from the life history of a star of large mass. For a star of small mass the natural white-dwarf stage is an initial step towards complete extinction. A star of large mass cannot pass into the white-dwarfstage and one is left speculating on other possibilities. (1934)[8]

And these statements have retained their validity.

While the evolution of the massive stars was thus left uncertain, there was no such uncertainty regarding the final states of stars of sufficiently low mass." The reason is that by virtue, again, of the inequality (10), the maximum central pressure attainable in a star must be less than that provided by the degenerate equation of state, so long as

$$\frac{1}{2} G \left(\frac{4}{3}\pi\right)^{1/3} M^{2/3} < K_2 = \frac{1}{8}\left(\frac{3}{\pi}\right)^{1/3} \frac{hc}{(\mu_e H)^{4/3}} \tag{54}$$

or, equivalently

$$M < \frac{3}{16\pi}\left(\frac{hc}{G}\right)^{3/2} \frac{1}{(\mu_e H)^2} = 1.74\, \mu_e^{-2} \odot. \tag{55}$$

We conclude that there can be no surprises in the evolution of stars of mass less than 0.43 $\odot$ (if $\mu_e = 2$). The end stage in the evolution of such stars can only be that of the white dwarfs. (Parenthetically, we may note here that the inequality (55) implies that the so-called 'mini' black-holes of mass $\sim 10^{15}$ g m cannot naturally be formed in the present astronomical universe.)

7. Some brief remarks on recent progress in the evolution of massive stars and the onset of gravitational collapse

It became clear, already from the early considerations, that the inability of the massive stars to became white dwarfs must result in the development of much more extreme conditions in their interiors and, eventually, in the onset of gravitational collapse attended by the super-nova phenomenon. But the precise manner in which all this will happen has been difficult to ascertain in spite of great effort by several competent groups of investigators. The facts which must be taken into account appear to be the following.*

In the first instance, the density and the temperature will steadily increase without the inhibiting effect of degeneracy since for the massive stars considered $1-\beta_e > 1-\beta_\omega$. On this account, 'nuclear ignition' of carbon, say, will take place which will be attended by the emission of neutrinos. This emission of neutrinos will effect a cooling and a lowering of $(1-\beta_e)$; but it will still be in excess of $1-\beta_\omega$. The important point here is that the emission of neutrinos acts selectively in the central regions and is the cause of the lowering of $(1-\beta_e)$ in these regions. The density and the temperature will continue to increase till the next ignition of neon takes place followed by further emission of neutrinos and a further lowering of $(1-\beta_e)$. This succession of nuclear ignitions and lowering of $(1-\beta_e)$ will continue till $1-\beta_e < 1-\beta_\omega$ and a relativistically degenerate core with a mass approximately that of the limiting mass $(=1.4\odot$ for $\mu_e = 2)$ forms at the centre. By this stage, or soon afterwards, instability of some sort is expected to set in (see following §8) followed by gravitational collapse and the phenomenon of the super-nova (of type II). In some instances, what was originally the highly relativistic degenerate core of approximately 1.4 $\odot$, will be left behind as a neutron star. That this happens sometimes is confirmed by the fact that in those cases for which reliable estimates of the masses of pulsars exist, they are consistently close to 1.4 $\odot$. However, in other instances - perhaps, in the majority of the instances - what is left behind, after all 'the dust has settled', will have masses in excess of that allowed for stable neutron stars; and in these instances black holes will form.

In the case of less massive stars $(M \sim 6\text{-}8\ \odot)$ the degenerate cores, which are initially formed, are not highly relativistic. But the mass of core increases with the further burning of the nuclear fuel at the interface of the core and the mantle; and when the core reaches the limiting mass, an explosion occurs following instability; and it is believed that this is the cause underlying super-nova phenomenon of type I.

* I am grateful to Professor D. Arnett for guiding me through the recent literature and giving me advice in the writing of this section.

From the foregoing brief description of what may happen during the late stages in the evolution of massive stars, it is clear that the problems one encounters are of exceptional complexity, in which a great variety of physical factors compete. This is clearly not the occasion for me to enter into a detailed discussion of these various questions. Besides, Professor Fowler may address himself to some of these matters in his lecture that is to follow.

8. Instabilities of relativistic origin: (I) The vibrational instability of spherical stars

I now turn to the consideration of certain types of stellar instabilities which are derived from the effects of general relativity and which have no counterparts in the Newtonian framework. It will appear that these new types of instabilities of relativistic origin may have essential roles to play in discussions pertaining to gravitational collapse and the late stages in the evolution of massive stars.

We shall consider first the stability of spherical stars for purely radial perturbations. The criterion for such stability follows directly from the linearized equations governing the spherically symmetric radial oscillations of stars. In the framework of the Newtonian theory of gravitation, the stability for radial perturbations depends only on an average value of the adiabatic exponent, Γ_1, which is the ratio of the fractional Lagrangian changes in the pressure and in the density experienced by a fluid element following the motion; thus,

$$\Delta P/P = \Gamma_1 \, \Delta\rho/\rho. \tag{56}$$

And the Newtonian criterion for stability is

$$\bar{\Gamma}_1 = \int_0^M \Gamma_1(r) \, P(r) \, dM(r) \div \int_0^M P(r) \, dM(r) \gtrsim \frac{4}{3}. \tag{57}$$

If $\bar{\Gamma}_1 < 4/3$, *dynamical instability* of a global character will ensue with an e-folding time measured by the time taken by a sound wave to travel from the centre to the surface.

When one examines the same problem in the framework of the general theory of relativity, one finds [12] that, again, the stability depends on an average value of Γ_1; but contrary to the Newtonian result, the stability now depends on the radius of the star as well. Thus, one finds that no matter how high $\bar{\Gamma}_1$ may be, instability will set in provided the radius is less than a certain determinate multiple of the *Schwarzschild radius,*

$$R_s = 2 \, GM/c^2. \tag{58}$$

Thus, if for the sake of simplicity, we assume that Γ_1 is a constant through the star and equal to 5/3, then the star will become dynamically unstable for radial perturbations, if $R < 2.4 \, R_s$. And further, if $r_1 \to \infty$, instability will set in for all $R < (9/8) \, R_s$. *The radius* $(9/8) \, R_s$ *defines, in fact, the minimum radius which any gravitating mass, in hydrostatic equilibrium, can have in the framework of general relativity.* This important result is implicit in a fundamental paper by Karl Schwarzschild published in 1916. (Schwarzschild actually proved that for a star in which the energy density is uniform, $R > (9/8)R_s$.)

In one sense, the most important consequence of this instability of relativistic origin is that if Γ_1 (again assumed to be a constant for the sake of simplicity) differs from and is greater than 4/3 only by a small positive constant, then the instability will set in for a radius R which is a large multiple of R_s; and, therefore, under circumstances when the effects of general relativity, on the structure of the equilibrium configuration itself, are hardly relevant. Indeed, it follows[13] from the equations governing radial oscillations of a star, in a first post-Newtonian approximation to the general theory of relativity, that instability for radial perturbations will set in for all

$$R < \frac{K}{\Gamma_1 - 4/3}\frac{2GM}{c^2}, \tag{59}$$

where K is a constant which depends on the *entire*' march of density and pressure in the equilibrium configuration in the Newtonian frame-work. Thus, for a polytrope of index n, the value of the constant is given by

$$K = \frac{5-n}{18}\left[\frac{2(11-n)}{(n+1)\,\xi_1^4\,|\theta_1'|^3}\int_0^{\xi_1}\theta\left(\frac{d\theta}{d\xi}\right)^2\xi^2\,d\xi + 1\right], \tag{60}$$

where θ is the Lane-Emden function in its standard normalization ($\theta = 1$ at $\xi = 0$), ξ is the dimensionless radial coordinate, ξ_1 defines the boundary of the polytrope (where $\theta = 0$) and θ_1' is the derivative of θ at ξ_1.

Table 2

Values of the constant K in the inequality (59) for various polytropic indices, n.

n	K	n	K
0	0.452381	3.25	1.28503
1.0	.565382	3.5	1.49953
1.5	.645063	4.0	2.25338
2.0	.751296	4.5	4.5303
2.5	.900302	4.9	22.906
3.0	1.12447	4.95	45.94

In Table 2, we list the values of K for different polytropic indices. It should be particularly noted that K increases without limit for $n \to 5$ and the configuration becomes increasingly centrally condensed.** Thus, already for $n = 4.95$ (for which polytropic index $\rho_c = 8.09 \times 10^6\bar{\rho}$), *K-46.* In other words, for the highly centrally condensed massive stars (for which Γ_1 may differ from 4/3 by as little

* It is for this reason that we describe the instability as *global.*

** Since this was written, it has been possible to show (Chandrasekhar and Lebovitz *13a)* that for $n \to 5$, the asymptotic behaviour of K is given by

$$K \to 2.3056/(5-n).$$

and, further, that along the polytropic sequence, the criterion for instability (59) can be expressed alternatively in the form

$$R < 0.2264\,\frac{2GM}{c^2}\,\frac{1}{\Gamma_1 - 4/3}\left(\frac{\rho_c}{\bar{\rho}}\right)^{1/3} \qquad (\rho_c > 10^6\bar{\rho})$$

as 0.01); the instability of relativistic origin will set in, already, when its radius falls below $5 \times 10^3 R_s$. Clearly this relativistic instability must be considered in the contexts of these problems.

A further application of the result described in the preceding paragraph is to degenerate configurations near the limiting mass'*. Since the electrons in these highly relativistic configurations have velocities close to the velocity of light, the effective value of Γ_1 will be very close to 4/3 and the post-Newtonian relativistic instability will set in for a mass slightly less than that of the limiting mass. On account of the instability for radial oscillations setting in for a mass less than M_{limit}, the period of oscillation, along the sequence of the degenerate configurations, must have a minimum. This minimum can be estimated to be about two seconds (see Fig. 6). Since pulsars, when they were discovered, were known to have periods much less than this minimum value, the possibility of their being degenerate configurations near the limiting mass was ruled out; and this was one of the deciding factors in favour of the pulsars being neutron stars. (But by a strange irony, for reasons we have briefly explained in § 7, pulsars which have resulted from super-nova explosions have masses close to 1.4 $\odot$!)

Finally, we may note that the radial instability of relativistic origin is the underlying cause for the *existence* of a maximum mass for stability: it is a direct consequence of the equations governing hydrostatic equilibrium in general relativity. (For a complete investigation on the periods of radial oscillation of neutron stars for various admissible equations of state, see a recent paper by Detweiler and Lindblom[15].)

9. Instabilities of relativistic origin: (2) The secular instability of rotating stars derived from the emission of gravitational radiation by non-axisymmetric modes of oscillation

I now turn to a different type of instability which the general theory of relativity predicts for rotating configurations. This new type of instability[16] has its origin in the fact that the general theory of relativity builds into rotating masses a dissipative mechanism derived from the possibility of the emission of gravitational radiation by non-axisymmetric modes of oscillation. It appears that this instability limits the periods of rotation of pulsars. But first, I shall explain the nature and the origin of this type of instability.

It is well known that a possible sequence of equilibrium figures of rotating homogeneous masses is the Maclaurin sequence of oblate spheroids[17]. When one examines the second harmonic oscillations of the Maclaurin spheroid, in a frame of reference rotating with its angular velocity, one finds that for two of these modes, whose dependence on the azimuthal angle is given by $e^{2i\varphi}$, the characteristic frequencies of oscillation, σ, depend on the eccentricity e in the manner illustrated in Figure 7. It will be observed that one of these modes becomes neutral (i.e., $\sigma = 0$) when $e = 0.813$ and that the two modes coalesce when $e = 0.953$ and become complex conjugates of one another beyond this

* By reason of the dominance of the radiation pressure in these massive stars and of β being very close to zero.

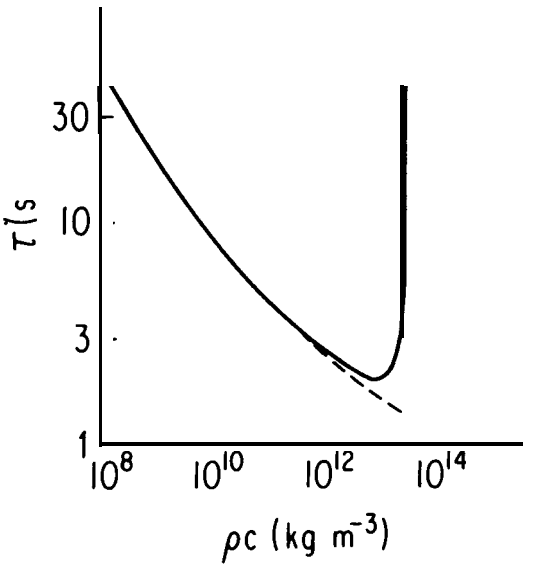


Fig. 6. The variation of the period of radial oscillation along the completely degenerate configurations. Notice that the period tends to infinity for a mass close to the limiting mass. There is consequently a minimum period of oscillation along these configurations; and the minimum period is approximately 2 seconds. (From J. Skilling, *Pulsating Stars* (Plenum Press, New York, 1968), p. 59.)

point. Accordingly, the Maclaurin spheroid becomes *dynamically unstable* at the latter point (first isolated by Riemann). On the other hand, the origin of the neutral mode at e = 0.813 is that at this point a new equilibrium sequence of triaxial ellipsoids-the ellipsoids of Jacobi-bifurcate. On this latter account, Lord Kelvin conjectured in 1883 that

> if there be any viscosity, however slight . . . the equilibrium beyond e = 0.81 cannot be secularly stable.

Kelvin's reasoning was this: viscosity dissipates energy but not angular momentum. And since for equal angular momenta, the Jacobi ellipsoid has a lower energy content than the Maclaurin spheroid, one may expect that the action of viscosity will be to dissipate the excess energy of the Maclaurin spheroid and transform it into the Jacobi ellipsoid with the lower energy. A detailed calculation[18] of the effect of viscous dissipation on the two modes of oscillation, illustrated in Figure 7, does confirm Lord Kelvin's conjecture. It is found that viscous dissipation makes the mode, which becomes neutral at e= 0.813, unstable beyond this point with an e-folding time which depends inversely on the magnitude of the kinematic viscosity and which further decreases monotonically to zero at the point, e = 0.953 where the dynamical instability sets in.

Since the emission of gravitational radiation dissipates *both* energy and angular momentum, it does *not* induce instability in the Jacobi mode; instead it

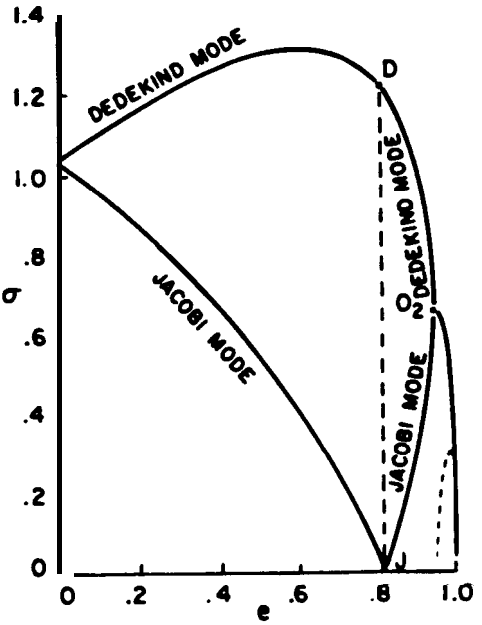


Fig. 7. The characteristic frequencies (in the unit $(\pi G\varrho)^{1/2}$) of the two even modes of second-harmonic oscillation of the Maclaurin sphcriod. The Jacobi sequence bifurcates from the Maclaurin sequence by the mode that is neutral ($\sigma = 0$) at e = 0.813; and the Dcdekind sequence bifurcates by the alternative mode at D. At O_2, (e = 0.9529) the Maclaurin spheroid becomes dynamically unstable. The real and the imaginary parts of the frequency, beyond O_2 are shown by the full line and the dashed curves, respectively. Viscous dissipation induces instability in the branch of the Jacobi mode; and radiation-reaction induces instability in the branch DO, of the Dedekind mode.

induces instability in the *alternative* mode at the same eccentricity. In the first instance this may appear surprising; but the situation we encounter here clarifies some important issues.

If instead of analyzing the normal modes in the rotating frame, we had analyzed them in the inertial frame, we should have found that the mode which becomes unstable by radiation reaction at e = 0.813, is in fact neutral at this point. And the neutrality of *this* mode in the inertial frame corresponds to the fact that the neutral deformation at this point is associated with the bifurcation (at this point) of a new triaxial sequence-the sequence of the Dedekind ellipsoids. These Dedekind ellipsoids, while they are congruent to the Jacobi ellipsoids, they differ from them in that they are at rest in the inertial frame and owe their triaxial figures to internal vortical motions. An important conclusion that would appear to follow from these facts is that in the framework of general relativity we can expect secular instability, derived from radiation reaction, to arise from a Dedekind mode of deformation (which is quasi-stationary in the inertial frame) rather than the Jacobi mode (which is quasi-stationary in the rotating frame).

A further fact concerning the secular instability induced by radiation reaction, discovered subsequently by Friedman[19] and by Comins[20], is that the

modes belonging to higher values of m (= 3, 4, , ,) become unstable at smaller eccentricities though the e-folding times for the instability becomes rapidly longer. Nevertheless it appears from some preliminary calculations of Friedman[21] that it is the secular instability derived from modes belonging to $m = 3$ (or 4) that limit the periods of rotation of the pulsars.

It is clear from the foregoing discussions that the two types of instabilities of relativistic origin we have considered are destined to play significant roles in the contexts we have considered.

10. *The mathematical theory of black holes*

So far, I have considered only the restrictions on the last stages of stellar evolution that follow from the existence of an upper limit to the mass of completely degenerate configurations and from the instabilities of relativistic origin. From these and related considerations, the conclusion is inescapable that black holes will form as one of the natural end products of stellar evolution of massive stars; and further that they must exist in large numbers in the present astronomical universe. In this last section I want to consider very briefly what the general theory of relativity has to say about them. But first, I must define precisely what a black hole is.

A black hole partitions the three-dimensional space into two regions: an inner region which is bounded by a smooth two-dimensional surface called the *event horizon;* and an outer region, external to the event horizon, which is asymptotically flat; and it is required (as a part of the definition) that no point in the inner region can communicate with any point of the outer region. This incommunicability is guaranteed by the impossibility of any light signal, originating in the inner region, crossing the event horizon. The requirement of asymptotic flatness of the outer region is equivalent to the requirement that the black hole is isolated in space and that far from the event horizon the space-time approaches the customary space-time of terrestrial physics.

In the general theory of relativity, we must seek solutions of Einstein's vacuum equations compatible with the two requirements I have stated. It is a startling fact that compatible with these very simple and necessary requirements, the general theory of relativity allows for stationary (i.e., time-independent) black-holes exactly a single, unique, two-parameter family of solutions. This is the Kerr family, in which the two parameters are the mass of the black hole and the angular momentum of the black hole. What is even more remarkable, the metric describing these solutions is simple and can be explicitly written down.

I do not know if the full import of what I have said is clear. Let me explain.

Black holes are macroscopic objects with masses varying from a few solar masses to millions of solar masses. To the extent they may be considered as stationary and isolated, to that extent, they are all, every single one of them, described *exactly* by the Kerr solution. This is the only instance we have of an exact description of a macroscopic object. Macroscopic objects, as we see them all around us, are governed by a variety of forces, derived from a variety of approximations to a variety of physical theories. In contrast, the only elements

in the construction of black holes are our basic concepts of space and time. They are, thus, almost by definition, the most perfect macroscopic objects there are in the universe. And since the general theory of relativity provides a single unique two-parameter family of solutions for their descriptions, they are the simplest objects as well.

Turning to the physical properties of the black holes, we can study them best by examining their reaction to external perturbations such as the incidence of waves of different sorts. Such studies reveal an analytic richness of the Kerr space-time which one could hardly have expected. This is not the occasion to elaborate on these technical matters[22]. Let it suffice to say that contrary to every prior expectation, all the standard equations of mathematical physics can be solved exactly in the Kerr space-time. And the solutions predict a variety and range of physical phenomena which black holes must exhibit in their interaction with the world outside.

The mathematical theory of black holes is a subject of immense complexity; but its study has convinced me of the basic truth of the ancient mottoes,

The simple is the seal of the true

and

Beauty is the splendour of truth.

REFERENCES

1. Eddington, A. S., *The* Internal Constitution of *the* Stars (Cambridge University Press, England, 1926), p. 16.
2. Chandrasekhar, S., Mon. *Not. Roy. Astr. Soc., 96, 644* (1936).
3. Eddington, A. *S., The Internal Constitution of the Stars* (Cambridge University Press, England, 1926), p. 172.
4. Fowler, R. H., *Mon. Not. Roy. Astr. Soc., 87, 114* (1926).
5. Chandrasekhar, S., *Phil. Mag.,* 11, 592 (1931).
6. Chandrasekhar, S., *Astrophys.J., 74,* 81 (1931).
7. Chandrasekhar, S., Mon. *Not. Roy. Astr. Soc.,* 91, 456 (1931).
8. Chandrasekhar, S., *Observatory, 57, 373* (1934).
9. Chandrasekhar, S., *Mon. Not. Roy. Astr. Soc., 95, 207 (1935).*
10. Chandrasekhar, S., *Z.f. Astrophysik, 5,* 321 (1932).
11. Chandrasekhar, S., *Observatory, 57, 93 (1934).*
12. Chandrasekhar, S., *Astrophys. J.,* 140, 417 (1964); see also *Phys. Rev. Lett.,* 12, 114 and 437 (1964).
13. Chandrasekhar, S., *Astrophys.J.,* 142, 1519 (1965).

13a Chandrasekhar, S. and Lebovitz, N. R., *Mon. Not. Roy. Astr. Soc., 207, 13 P (1984).*

14. Chandrasekhar, S. and Tooper, R. F., *Astrophys .J.,* 139, 1396 (1964).
15. Detweiler, S. and Lindblom, L., *Astrophys.J. Supp., 53, 93 (1983).*
16. Chandrasekhar, S., *Astrophys. J.,* 161, 561 (1970); see also *Phys. Rev. Lett., 24, 611 I* and 762 (1970).
17. For an account of these matters pertaining to the classical ellipsoids see Chandrasekhar, S., *Ellipsoidal Figures of Equilibrium* (Yale University Press, New Haven, 1968).
18. Chandrasekhar, S., *Ellipsoidal Figures* of *Equilibrium* (Yale University Press, New Haven, 1968), Chap. 5, § 37.
19. Friedman, J. L., *Comm. Math. Phys.,* 62, 247 (1978); see also Friedman, J. L. and Schutz, B. F., *Astrophys.J., 222,* 281 (1977).
20. Comins, N., Mon. *Not. Roy. Astr. Soc., 189, 233* and 255 (1979).
21. Friedman, J. L., *Phys. Rev. Lett.,* 51, 11 (1983).
22. The author's investigations on the mathematical theory of black holes, continued over the years 1974- 1983, are summarized in his last book *The Mathematica/ Theory of Black Holes* (Clarendon Press, Oxford, 1983).

The reader may wish to consult the following additional references:

1. Chandrasekhar, S., 'Edward Arthur Milnc: his part in the development of modern astrophysics,' *Quuart.J. Roy. Astr. Soc., 21,* 93-107 (1980).
2. Chandrasekhar, *S., Eddington; The Most Distinguished Astrophysicist* ***of*** *His Time* (Cambridge University Press, 1983).

WILLIAM ALFRED FOWLER

I was born in 1911 in Pittsburgh, Pennsylvania, the son of John MacLeod Fowler and Jennie Summers Watson Fowler. My parents had two other children, my younger brother, Arthur Watson Fowler and my still younger sister, Nelda Fowler Wood. My paternal grandfather, William Fowler, was a coal miner in Slammannan, near Falkirk, Scotland who emigrated to Pittsburgh to find work as a coal miner around 1880. My maternal grandfather, Alfred Watson, was a grocer. He emigrated to Pittsburgh, also around 1880, from Taniokey, near Clare in County Armagh, Northern Ireland. His parents taught in the National School, the local grammar school for children, in Taniokey, for sixty years. The family lived in the central part of the school building; my great grandfather taught the boys in one wing of the building and my great grandmother taught the girls in the other wing. The school is still there and I have been to see it.

I was raised in Lima, Ohio, from the age of two when my father, an accountant, was transferred to Lima from Pittsburgh. Each summer during my childhood the family went back to Pittsburgh during my father's vacation from work. He was an ardent sportsman and through him I became (and still am) a loyal fan of the Pittsburgh Pirates in the National Baseball League and of the Pittsburgh Steelers in the National Football League.

Lima was a railroad center served by the Pennsylvania, Erie, Nickel Plate and Baltimore & Ohio railroads. It was also the home of the Lima Locomotive Works which built steam locomotives. My brother, Arthur Watson Fowler, a mechanical engineer, worked for Lima Locomotive all his life until his retirement. After 1960 the company produced power shovels and construction cranes. As a boy I spent many hours in the switch yards of the Pennsylvania Railroad not far from my family home. It is no wonder that I go around the world seeking passenger trains still pulled by steam locomotives. In 1973 I travelled the Trans Siberian Railroad from Khabarovsk to Moscow because, among other reasons, the train was powered by steam for almost 2500 kilometers from Khabarovsk to Chita. It's not powered by steam but now I can afford to ride on the new Orient Express. It is also no wonder that on my 60th birthday my colleagues and former students presented me in Cambridge, England, with a working model, 3 l/4" gauge (l/16 standard size) British Tank Engine. I operated it frequently on the elevated track of the Cambridge and District Model Engineering Society. It is my pride and joy. I have named it *Prince Hal.*

I attended Horace Mann Grade School and Lima Central High School. A few of my high school teachers are still alive and I met them at my 50th class

reunion in 1979. I was President of the Senior Class of 1929. My teachers encouraged and fostered my interest in engineering and science but also insisted that I take four years of Latin rather than French or German. My family home was located across the street from the extensive playgrounds of Horace Mann School. There were baseball diamonds, tennis courts, a running track and a football field. During my high school days I played on the Central High School football team and won my letter as a senior. Horace Mann was Central's home football field. During my college days I served as Recreational Director of the Horace Mann playground during the summer. Not far from my home was Baxter's Woods with a running creek and swimming hole. What a wonderful environment it all was for my boyhood!

On graduation from school I enrolled at the Ohio State University in Columbus, Ohio, in ceramic engineering. I had won a prize for an essay on the production of Portland cement and ceramic engineering seemed a natural choice for me. Fortunately all engineering students took the same courses including physics and mathematics. I became fascinated with physics and when I learned from Professor Alpheus Smith, head of the Physics Department, that there was a new degree offered in Engineering Physics I enrolled in that option at the start of my sophomore year. So also did Leonard I. Schiff, who became a very great theoretical physicist. We were lifelong friends until his death a few years ago.

My parents were not affluent and my summer salary as recreation director did not cover my expenses at Ohio State. For my meals I waited table, washed the dishes and stoked the furnaces at the Phi Sigma Sigma Sorority. I worked Saturdays cutting and selling ham and cheese in an outside stall at the Central Market in Columbus. Early in the morning we put up the stall and unloaded the hams and cheeses from the wholesaler's truck; late at night we cleaned up and took down the stall. For eighteen hours work I was paid five dollars. I did scrape enough money together to join a social fraternity, Tau Kappa Epsilon. In my junior year I was elected to the engineering honorary society, Tau Beta Pi, and in my senior year I was elected President of the Ohio State Chapter.

My professors at Ohio State solidified my interest in experimental physics. Willard Bennett permitted me to do an undergraduate thesis on the "Focussing of Electron Beams" in his laboratory. From him I learned how different a working laboratory is from a student laboratory. The answers are not known! John Byrne permitted me to work after school hours in the electronic laboratory of the Electrical Engineering Department. I studied the characteristics of the Pentode! It was the best of worlds-the thrills of making real measurements in physics along with practical training in engineering.

On graduation from Ohio State I came to Caltech and became a graduate student under Charles Christian Lauritsen-physicist, engineer, architect and violinist-in the W. K. Kellogg Radiation Laboratory. Kellogg was constructed to Lauritsen's architectural plans by funds obtained from the American corn flakes king by Robert Andrews Millikan. Lauritsen was a native of Denmark and in common with many Scandinavians he loved the songs of Carl Michael Bellman, the 18th century Swedish poet-musician, He tried to teach me to sing

Bellman's drinking songs with a good Swedish accent but I failed miserably except in spirit or should I say spirits. 'Del Delsasso dubbed me Willy and it stuck'.

Charlie Lauritsen was the greatest influence in my life. He supervised my doctoral thesis on "Radioactive Elements of Low Atomic Number" in which we discovered *mirror nuclei* and showed that the nuclear forces are charge symmetric-the same between two protons as between two neutrons when charged particle Coulomb forces are excluded. He taught me many practical things-how to repair motors, plumbing, and electrical wiring. Most of all he taught me how to *do* physics and how to enjoy it. I also learned from my fellow graduate students Richard Crane and Lewis Delsasso. Charlie's son, Tommy Lauritsen, did his doctoral work under us and the three of us worked together as a team for over thirty-live years. We were primarily experimentalists. In the early days Robert Oppenheimer taught us the theoretical implications of our results. Richard Tolman taught us not to rush into the publication of premature results in those days of intense competition between nuclear laboratories.

Hans Bethe's announcement of the CN-cycle in 1939 changed our lives. We were studying the nuclear reactions of protons with the isotopes of carbon and nitrogen in the laboratory, the very reactions in the CN-cycle. World War II intervened. The Kellogg Laboratory was engaged in defense research throughout the war. I spent three months in the South Pacific during 1944 as a civilian with simulated military rank. I saw at first hand the heroism of soldiers and seamen and the horrors they endured.

Just before the war I married Ardiane Foy Olmsted whose family came to California over the plains and mountains of the western United States in the Gold Rush around 1850. We are the parents of two daughters, Mary Emily and Martha Summers, whom we refer to as our biblical characters. Martha and her husband, Robert Schoenemann, are the parents of our grandson, Spruce William Schoenemann. They live in Pawlet, a small village in Vermont-the Green Mountain State.

After the war the Lauritsens and I restored Kellogg as a nuclear laboratory and decided to concentrate on nuclear reactions which take place in stars. We called it Nuclear Astrophysics. Before the war Hans Staub and William Stephens had confirmed that there was no stable nucleus at mass 5. After the war Alvin Tollestrup, Charlie Lauritsen and I confirmed that there was no stable nucleus at mass 8. These mass gaps spelled the doom of George Gamow's brilliant idea that all nuclei heavier than helium (mass 4) could be built by neutron addition one mass unit at a time in his *big bang.* Edwin Salpeter of Cornell came to Kellogg in the summer of 1951 and showed that the fusion of three helium nuclei of mass four into the carbon nucleus of mass twelve could probably occur in Red Giant stars but not in the *big bang.* In 1953 Fred Hoyle induced Ward Whaling in Kellogg to perform an experiment which quantitatively confirmed the fusion process under the temperature and density conditions which Hoyle, Martin Schwarzschild and Allan Sandage had shown occur in Red Giants.

Fred Hoyle was the second great influence in my life. The grand concept of

nucleosynthesis in stars was first definitely established by Hoyle in 1946. After Whaling's confirmation of Hoyle's ideas I became a believer and in 1954/1955 spent a sabbatical year in Cambridge, England, as a Fulbright Scholar in order to work with Hoyle. There Geoffrey and Margaret Burbidge joined us. In 1956 the Burbidges and Hoyle came to Kellogg and in 1957 our joint efforts culminated in the publication of "Synthesis of the Elements in Stars" in which we showed that all of the elements from carbon to uranium could be produced by nuclear processes in stars starting with the hydrogen and helium produced in the *big bang*. This paper has come to be known from the last initials of the authors as B^2FH. A. G. W. Cameron single-handedly came forward with the same broad ideas at the same time.

Fred Hoyle became the Plumian Professor at Cambridge, was knighted by the Queen and founded the Institute of Theoretical Astronomy in Cambridge in 1966. I spent many happy summers at the Institute until Hoyle's retirement to Cumbria in the Lake District of England. Fred taught me more than astrophysics. He introduced me to English cricket, rugby and association football (we call it soccer). He took me to the Scottish Highlands and taught me how to read an ordnance map as well as how to enjoy climbing the 3000 ft peaks called Munros. I still go climbing somewhere in the British Isles every summer. It keeps me tit and renews my soul.

If has been a long row to hoe. Experimental measurements of the cross section of hundreds of nuclear reactions and their conversion into stellar reaction rates are essential if nucleosynthesis in stars is to be quantitatively confirmed. The Kellogg Laboratory has played a leading role for many years in this effort. I am fortunate that the Nobel Prize was awarded from team work. It is impossible to credit all my colleagues. In experimental nuclear astrophysics Charles Barnes and Ralph Kavanagh have played leading roles. So did Thomas Tombrello and Ward Whaling until they found other fields of interest and promise. In addition Robert Christy and Steven Koonin in theoretical nuclear physics, Jesse Greenstein in observational and theoretical astronomy and Gerald Wasserburg in precision geochemistry on meteoritic and lunar samples have played essential roles. Of my 50 graduate students who have contributed to the field I must single out Donald D. Clayton. His graduate student Stanford Woosley is my grand student and his student Rick Wallace is my great grand student. Nuclear Astrophysics continues to be an active and exciting field. This is clearly evident in my 70th birthday festschrift, "Essays in Nuclear Astrophysics" in which the Cambridge University Press presents the research studies of my colleagues and former students around the world as of 1982.

It is appropriate to conclude, without elaboration, with some details of my life outside the laboratory:

Awarded Medal for Merit by President Harry Truman, 1948
Elected member of the National Academy of Sciences, 1956
Awarded Barnard Medal for Meritorious Service to Science, 1965
Member of the National Science Board, 1968-74
Member of the Space Science Board, 1970-73, 1977-80
Designated Benjamin Franklin Fellow of the Royal Society of Arts, 1970

Awarded the G. Unger Vetlesen Prize, 1973
Awarded National Medal of Science by President Gerald Ford, 1974
Designated Associate of the Royal Astronomical Society, 1975
Elected President of the American Physical Society, 1976
Designated an Honorary Member of the Mark Twain Society, 1976
Awarded Eddington Medal of the Royal Astronomical Society, 1978
Awarded Bruce Gold Medal, Astronomical Society of the Pacific, 1979
Elected to the Society of American Baseball Research, 1980-
Honorary degrees from University of Chicago, 1976, Ohio State University, 1978, University of Liege 1981, Observatory of Paris 1981 and Denison University 1982.

(added in 1991) : My 80th birthday celebration was held August 11 to 14, 1991 as a Nuclear Astrophysics Symposium, which was one part of the Caltech Centennial Year events. Again my colleagues and former students participated along with other experts in the field of nuclear astrophysics.

Ardiane Fowler died in May 1988. In December 1989 I married Mary Dutcher, a descendant of the Dutch founders of New Amsterdam, now New York. She had taught grade school for many years on Long Island and had not previously been married. We reside in the two-story, New England style white frame house, which I purchased in 1958. It is only a ten-minute walk from Caltech. I am retired from teaching so my only routine trips to the Insitute are on Wednesdays for the Astronomy Seminar, Thursdays for the Physics Colloquium and Fridays for the Kellogg Nuclear Physics Seminar. Mary Dutcher Fowler has painted all her life and she now attends a painting school in Pasadena. We keep busy by taking long walks on many weekends and in general try to stay out of trouble.

Honorary degrees
Arizona State University, 1985
Georgetown University, 1986
University of Massachusetts, 1987
Williams College, 1988
Gustavus Adolphus College, 1991

Honours
Nobel Prize for Physics, 1983
Sullivant Medal, The Ohio State University, 1985
First recipient of the William A. Fowler Award for Excellence and Distinguished Accomplishments in Physics, Ohio Section, American Physical Society, 1986
Legion d'Honneur awarded by President Mitterand of France, 1989
Member of Lima City Schools Distinguished Alumni Hall of Fame, 1990
Member of Ohio Sci. & Tech. Hall of Fame, 1991

EXPERIMENTAL AND THEORETICAL NUCLEAR ASTROPHYSICS; THE QUEST FOR THE ORIGIN OF THE ELEMENTS

Nobel lecture, 8 December, 1983

by

WILLIAM A. FOWLER

W. K. Kellogg Radiation Laboratory
California Institute of Technology, Pasadena, California 91125

Ad astra per aspera et per ludum

I. *Introduction*

We live on planet Earth warmed by the rays of a nearby star we call the Sun. The energy in those rays of sunlight comes initially from the nuclear fusion of hydrogen into helium deep in the solar interior. Eddington told us this in 1920 and Hans Bethe developed the detailed nuclear processes involved in the fusion in 1939. For this he was awarded the Nobel Prize in Physics in 1967.

All life on earth, including our own, depends on sunlight and thus on nuclear processes in the solar interior. But the sun did not produce the chemical elements which are found in the earth and in our bodies. The first two elements and their stable isotopes, hydrogen and helium, emerged from the first few minutes of the early high temperature, high density stage of the expanding Universe, the so-called “big bang”. A small amount of lithium, the third element in the periodic table, was also produced in the big bang, but the remainder of the lithium and all of beryllium, element four, and boron, element live, are thought to have been produced by the spallation of still heavier elements by the cosmic radiation in the interstellar medium between stars. These elements are in general very rare in keeping with this explanation of their origin as reviewed in detail by Audouze and Reeves (1).

Where did the heavier elements originate? The generally accepted answer is that all of the heavier elements from carbon, element six, up to long-lived radioactive uranium, element ninety-two, were produced by nuclear processes in the interior of stars in our own Galaxy. The stars we see at the present time in what we call the *Milky Way* are located in a spiral arm of our Galaxy. In Sweden *you* call it *Vintergatan,* the Winter Street. We see with our eyes only a small fraction of the one hundred billion stars in the Galaxy. Astronomers cover almost the full range of the electromagnetic spectrum and thus can observe many more Galactic stars and even individual stars in other galaxies.

The stars which synthesized the heavy elements in the solar system were

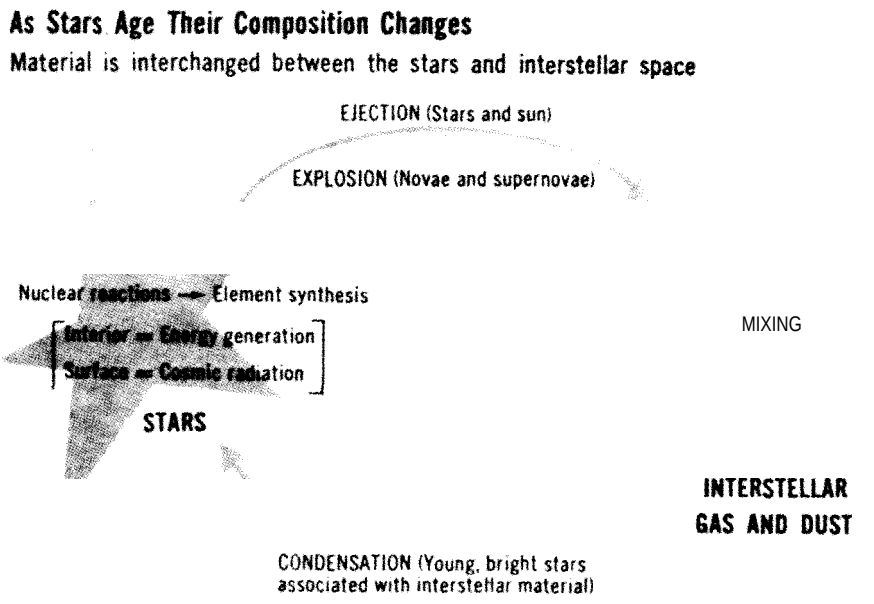


Figure 2. Synthesis of the elements in stars.

formed or born, evolved or aged, and eventually ejected the ashes of their nuclear fires into the interstellar medium over the lifetime of the Galaxy *before* the solar system itself formed four and one-half billion years ago.

The lifetime of the Galaxy is thought to be more than ten billion years but less than twenty billion years. In any case the Galaxy is much older than the solar system. The ejection of the nuclear ashes or newly formed elements took place by slow mass loss during the old age of the star, called the *giant* stage of stellar evolution, or during the relatively frequent outbursts which astronomers call novae, or during the final spectacular stellar explosions called *supernovae.* Supernovae can be considered to be the death of stars. White dwarfs or neutron stars or black holes which result from stellar evolution may represent a form of stellar purgatory.

In any case the sun and the earth and all the other planets in the solar system condensed under gravitational and rotational forces from a gaseous solar nebula in the interstellar medium consisting of "big bang" hydrogen and helium mixed with the heavier elements synthesized in earlier generations of Galactic stars. All of this is illustrated in Figure 1.

This idea can be generalized to successive generations of stars in the Galaxy with the result that the heavy element content of the interstellar medium and the stars which form from it increases with time. The oldest stars in the Galactic halo, that is, those we believe to have formed first, are found to have heavy element abundances less than one percent of the heavy element abundance of the solar system. The oldest stars in the Galactic disk have approximately ten percent. Only the less massive stars among those first formed can

have survived to the present as so-called Population II stars. Their small concentration of heavy elements may have been produced in a still earlier but more massive generation of stars, Population III, which rapidly exhausted their fuels and survived for only a very short lifetime. Stars formed in the disk of the Galaxy over its lifetime are referred to as Population I stars.

We speak of this element building as nucleosynthesis in stars. It can be generalized to other galaxies such as our twin, the Andromeda Nebula, and so this mechanism can be said to be a universal one. Astronomical observations on other galaxies have contributed much to our understanding of nucleosynthesis in stars.

We refer to the basic physics of energy generation and element synthesis in stars as Nuclear Astrophysics. It is a benign application of nuclear physics in contrast to military reactors and bombs. For the nuclear physicist this contrast is a personal and professional paradox. However, there is one thing of which I am certain. The science which explains the origin of sunlight must not be used to raise a dust cloud which will black out that sunlight from our planet.

As for all physics the field of Nuclear Astrophysics involves experimental and theoretical activities on the part of its practitioners and hence the first part of the title of this lecture. This lecture will emphasize nuclear experimental results and the theoretical analysis of those results almost but not entirely to the exclusion of other theoretical aspects. It will not in any way do justice to the observational activities of astronomers and cosmochemists which are necessary to complete the cycle: experiment, theory, observation. Nor will it do justice to the calculations by many theoretical astrophysicists of the results of nucleosynthesis of the elements and their isotopes under astrophysical conditions during the many stages of stellar evolution.

My deepest personal interest is in experimental data, in the analysis of the data and in the proper use of the data in theoretical stellar models. I continue to be encouraged in this regard by this one-hundred and nine year old quotation from Mark Twain:

> *There is something fascinating about science. One gets such wholesale returns of conjecture out of such a trifling investment of fact.*
>
> - Life on the Mississippi 1874

For me Twain's remark is a challenge to the experimentalist. The experimentalist must try to eliminate the word "trifling" through his endeavors in uncovering the facts of nature.

Experimental research and theoretical research are often very hard work. Fortunately this is lightened by the fun of doing physics and in obtaining results which bring a personal feeling of intellectual satisfaction. To my mind the hard work and the resulting intellectual fun transcend in a way the benefits which may accrue to society through subsequent technological applications. Please understand - I do not belittle these applications but I am unable to overlook the fact that they are a two-edged sword. My subject matter resulted from the hard work of a nuclear astrophysicist which when successful brought him joy and satisfaction. It was hard work but it was fun. Thus I have chosen

the subtitle for this lecture - “Ad astra per aspera et per ludum” which can be freely translated - “To the stars through hard work and fun.” This is in keeping with my paraphrase of the biblical quotation from Matthew “Man shall not live by work alone.”

With that in the record let us next ask what are the goals of Nuclear Astrophysics? First of all, Nuclear Astrophysics attempts to understand energy generation in the sun and other stars at all stages of stellar evolution. Energy generation by nuclear processes requires the transmutation of nuclei into new nuclei with lower mass. The small decrease in mass is multiplied by the velocity of light squared as Einstein taught us and a relatively large amount of energy is released.

Thus the first goal is closely related to the second goal that attempts to understand the nuclear processes which produced under various astrophysical circumstances the relative abundances of the elements and their isotopes in nature; whence the second part of the little of this lecture. Figure 2 shows a schematic curve of atomic abundances as a function of atomic weight. The data for this curve was first systemized from a plethora of terrestrial, meteoritic, solar and stellar data by Hans Suess and Harold Urey (2) and the available data has been periodically updated by A. G. W. Cameron (3). Major contributions to the experimental measurement of atomic transition rates needed to determine solar and stellar abundances have been made by my colleague, Ward Whaling (4). References (3) and (4) occur in a book *Essays in* Nuclear *Astrophysics* which reviews the field up to 1982. In the words of one of America’s baseball immortals, Casey Stengel, “You can always look it up.”

The curve in Figure 2 is frequently referred to as “universal” or “cosmic” but in reality it primarily represents relative atomic abundances in the solar system and in Main Sequence stars similar in mass and age to the sun. In current usage the curve is described succinctly as “solar”. It is beyond the scope of this lecture to elaborate on the difficult, beautiful research in astronomy and cosmochemistry which determined this curve. How this curve serves as a goal can be simply put. In the sequel it will be noticed that calculations of atomic abundances produced under astronomical circumstances at various postulated stellar sites are almost invariably reduced to ratios relative to “solar” abundances.

II. *Early Research on Element Synthesis*

George Gamow and his collaborators, R. A. Alpher and R. C. Herman (5), attempted to synthesize all of the elements during the big bang using a nonequilibrium theory involving neutron (n) capture with gamma-ray (γ) emission and electron (e) beta-decay by successively heavier nuclei. The synthesis proceeded in steps of one mass unit at a time since the neutron has approximately unit mass on the mass scale used in all the physical sciences. As they emphasized, this theory meets grave difficulties beyond mass 4 (^{4}He) because no stable nuclei exist at atomic mass 5 and 8. Enrico Fermi and Anthony Turkevich attempted valiantly to bridge these “mass gaps” without success and permitted Alpher and Herman to publish the results of their

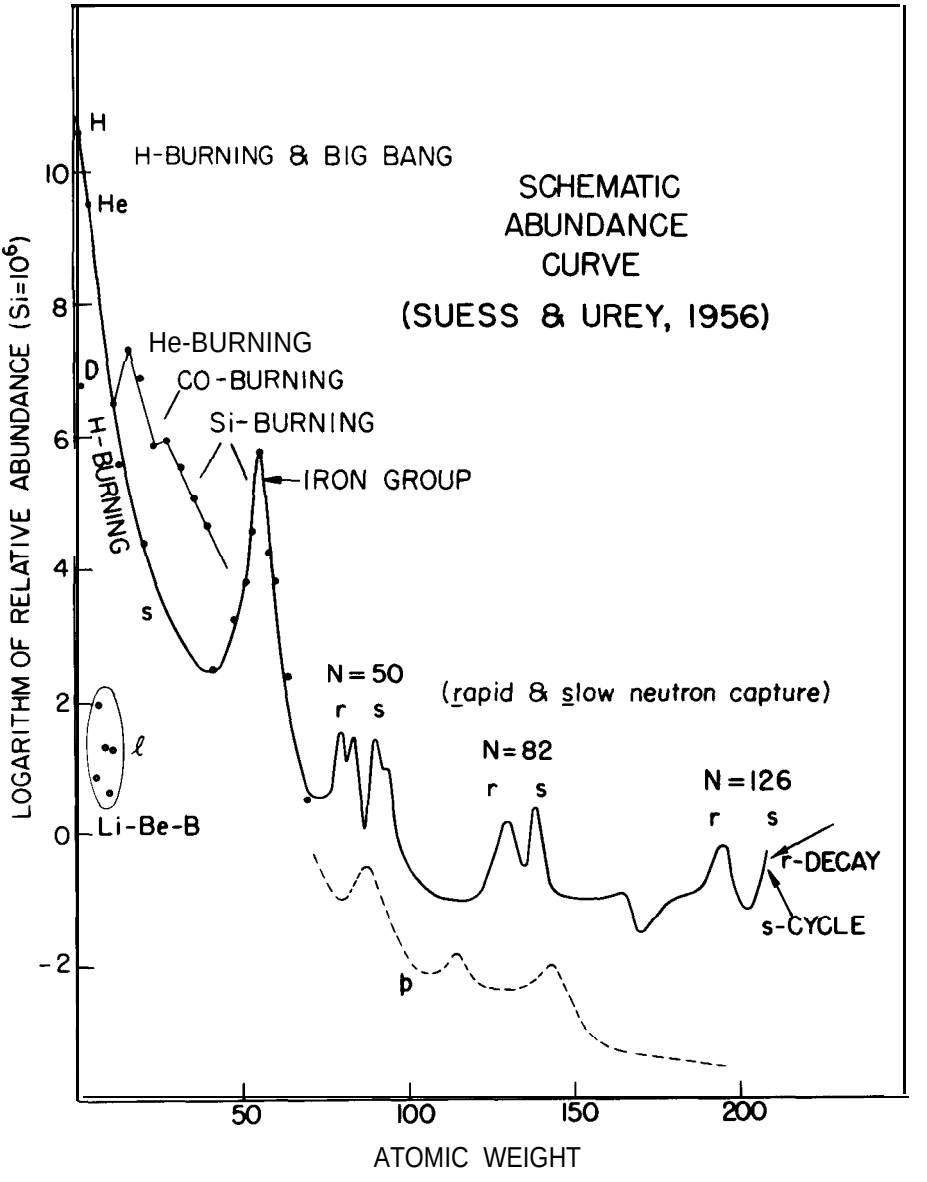


Figure 2. Schematic curve of atomic abundances relative to Si = 10^6 versus atomic weight for the sun and similar Main Sequence stars.

attempts. Seventeen years later Wagoner, Fowler, and Hoyle (6) armed with nuclear reaction data accumulated over the intervening years succeeded only in producing ^{7}Li at a mass fraction of at most 10^{-8} compared to hydrogen plus helium for acceptable model universes. All heavier elements totaled less than 10^{-11} by mass. Wagoner, Fowler, and Hoyle (6) did succeed in producing ^{2}D, ^{3}He, ^{4}He, and ^{7}Li in amounts in reasonable agreement with observations at the time. More recent observations and calculations are frequently used to

place constraints on models of the expanding universe and in general favor open models in which the expansion continues indefinitely. In other words there is not enough ordinary matter to close the universe. However, if neutrinos have only 10^{-5} the mass of the electron, they close the universe.

It was in connection with the "mass gaps" that the W. K. Kellogg Radiation Laboratory first became involved, albeit unwittingly, in astrophysical and cosmological phenomena. Before proceeding it is appropriate at this point to discuss briefly the origins of the Kellogg Radiation Laboratory where I have worked for 50 years. The laboratory was designed and the construction supervised by Charles Christian Lauritsen in 1930 through 1931. Robert Andrews Millikan, the head of Calteach, acquired the necessary funds from Will Keith Kellogg, the American "corn flakes king." The Laboratory was built to study the physics of 1 MeV X-rays and the application of those X-rays in the treatment of cancer. In 1932 Cockcroft and Walton discovered that nuclei could be disintegrated by protons (p), the nuclei of the light hydrogen atom ^{1}H, accelerated to energies well under 1 MeV. Lauritsen immediately converted one of his X-ray tubes into a positive ion accelerator (they were powered by alternating current transformers!) and began research in nuclear physics. Robert Oppenheimer and Richard Tolman were instrumental in convincing Millikan that Lauritsen was doing the right thing. Oppenheimer played an active role in the theoretical interpretation of the experimental results obtained in the Kellogg Laboratory in the early crucial years.

Lauritsen supervised my doctoral research from 1933- 1936 and I worked closely with him until his death. It was he who taught me that physics was both hard work and fun. He was a native of Denmark and was an accomplished violinist as well as physicist, architect and engineer. He loved the works of Carl Michael Bellman, the famous Swedish poet-musician of the 18th century, and played and sang Bellman for his students. It is well known that many of Bellman's works were drinking songs. That made it all the better.

We must now return to the first involvement of the Kellogg Radiation Laboratory in the mass gap at mass 5. In 1939, in Kellogg, Hans Staub and William Stephens (7) detected resonance scattering by ^{4}He of neutrons with orbital angular momentum equal to one in units of $\hbar$ (p-wave) and energy somewhat less than 1 MeV as shown in Figure 3. This confirmed previous reaction studies by Williams, Shepherd, and Haxby (8) and showed that the ground state of ^{5}He is unstable. As fast as ^{5}He is made it disintegrates! The same was later shown to be true for ^{5}Li, the other candidate nucleus at mass 5. The Pauli exclusion principle dictates for fermions that the third neutron in ^{5}He must have at least unit angular momentum and not zero as permitted for the first two neutrons with antiparallel spins. The attractive nuclear force cannot match the outward centrifugal force in classical terminology. Still later, in the Kellogg Radiation Laboratory, Tollestrup, Fowler, and Lauritsen (9) confirmed, with improved precision, the discovery of Hemmendinger (10) that the ground state of ^{8}Be is unstable. They (9) found the energy of the ^{8}Be break-up to be 89±5 keV compared to the currently accepted value of 91.89±0.05 keV! The Pauli exclusion principle is again at work in the instability of ^{8}Be. As

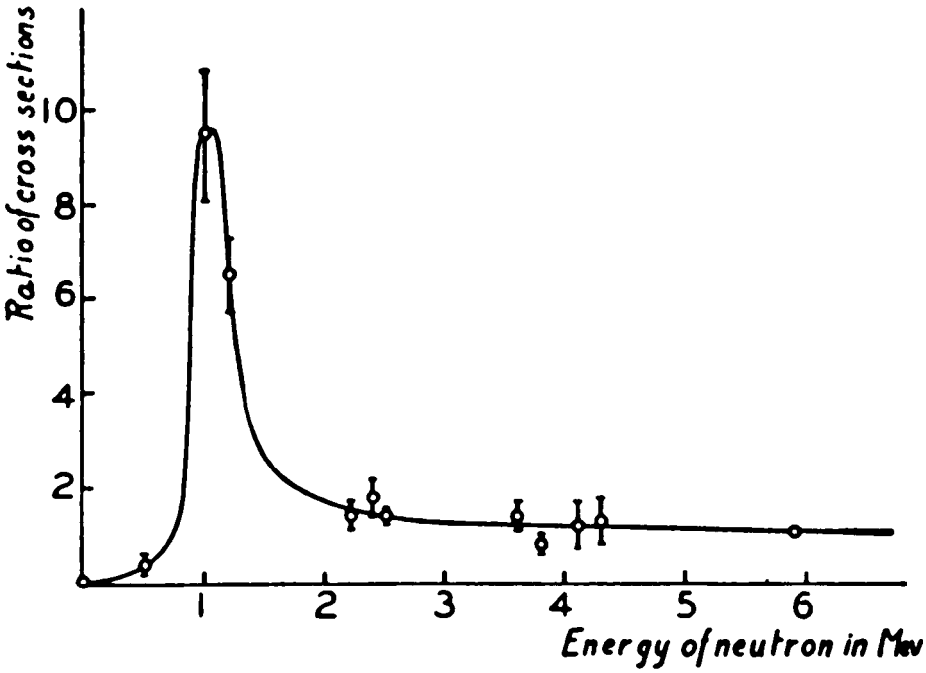


Figure 3. The ratio of the backward scattering cross section of helium to hydrogen as a function of the laboratory energy in MeV of the incident neutron.

fast as ^{8}Be is made it disintegrates into two ^{4}He-nuclei. The latter may be bosons but they consist of fermions. The mass gaps at 5 and 8 spelled the doom of Gamow's hopes that all nuclear species could be produced in the big bang one unit of mass at a time.

The eventual commitment of the Kellogg Radiation Laboratory to Nuclear Astrophysics came about in 1939 when Bethe (11) brought forward the operation of the CN-cycle as one mode of the fusion of hydrogen into helium in stars (since oxygen has been found to be involved the cycle is now known as the CNO-cycle). Charles Lauritsen, his son Thomas Lauritsen, and I were measuring the cross sections of the proton bombardment of the isotopes of carbon and nitrogen which constitute the CN-cycle. Bethe's paper (11) told us that we were studying in the laboratory processes which are occurring in the sun and other stars. It made a lasting impression on us. World War II intervened but in 1946 on returning the laboratory to nuclear experimental research, Lauritsen decided to continue in low-energy, *classical* nuclear physics with emphasis in the study of nuclear reactions thought to take place in stars. In this he was strongly supported by Ira Bowen, a Caltech Professor of Physics who had just been appointed Director of the Mt. Wilson Observatory, by Lee DuBridge, the new President of Caltech, by Carl Anderson, Nobel Prize winner 1936, and by Jesse Greenstein, newly appointed to establish research in astronomy at Caltech. In Kellogg, Lauritsen did not follow the fashionable trend to higher and higher energies which has continued to this day. He did support Robert Bacher and others in establishing high energy physics at Caltech.

Although Bethe (11) in 1939 and others still earlier had previously discussed energy generation by nuclear processes in stars the grand concept of nucleosynthesis in stars was first definitely established by Fred Hoyle (12). In two classic papers the basic ideas of the concept were presented within the framework of stellar structure and evolution with the use of the then known nuclear data.

Again the Kellogg Laboratory played a role. Before his second paper Hoyle was puzzled by the slow rate of the formation of '*C-nuclei from the fusion $(3\alpha \rightarrow {}^{12}\mathrm{C})$ of three alpha-particles (a) or ^{4}He-nuclei in Red Giant Stars. Hoyle was puzzled because his own work with Schwarzschild (13) and previous work of Sandage and Schwarzschild (14) had convinced him that helium burning through $3\alpha \rightarrow {}^{12}\mathrm{C}$ should commence in Red Giants just above 10^8K rather than at 2×10^8K as required by the reaction rate calculation of Salpeter (15). Salpeter made his calculation while a visitor at the Kellogg Laboratory during the summer of 1951 and used the Kellogg value (9) for the energy of ^{8}Be in excess of two ^{4}He to determine the resonant rate for the process $(2\alpha \leftrightarrow {}^8\mathrm{Be})$ which takes into account both the formation and decay of the ^{8}Be. However, in calculating the next step, ${}^8\mathrm{Be} + \alpha \rightarrow {}^{12}\mathrm{C} + \gamma$, Salpeter had treated the radiative fusion as nonresonant.

Hoyle realized that this step would be speeded up by many orders of magnitude, thus reducing the temperatures for its onset, if there existed an excited state of ^{12}C with energy 0.3 MeV in excess of ^{8}Be + α at rest and with the angular momentum and parity (0^+, 1^-, 2^+, 3^-, ...) dictated by the selection rules for these quantities. Hoyle came to the Kellogg Laboratory early in 1953 and questioned the staff about the possible existence of his proposed excited state. To make a long story short Ward Whaling and his visiting associates and graduate students (16) decided to go into the laboratory and search for the state using the ^{14}N (d, α) ^{12}C-reaction. They found it to be located almost exactly where Hoyle had predicted. It is now known to be at 7.654 MeV excitation in ^{12}C or 0.2875 MeV above ^{8}Be + *a* and 0.3794 MeV above *3a*. Cook, Fowler, Lauritsen, and Lauritsen (17) then produced the state in the decay of radioactive ^{12}B and showed it could break up into *3a* and thus by reciprocity could be formed from 3α. They argued that the spin and parity of the state must be 0^+ as is now known to be the case.

The $3\alpha \rightarrow {}^{12}\mathrm{C}$ fusion in Red Giants jumps the mass gaps at 5 and 8. This process could never occur under big bang conditions. By the time the ^{4}He was produced in the early expanding Universe the subsequent density and temperature were too low for the helium fusion to carbon to occur. In contrast, in Red Giants, after hydrogen conversion to helium during the Main Sequence stage, gravitational contraction of the helium core raises the density and temperature to values where helium fusion is ignited. Hoyle and Whaling showed that conditions in Red Giant stars are just right.

Fusion processes can be referred to as nuclear burning in the same way we speak of chemical burning. Helium burning in Red Giants succeeds hydrogen burning in Main Sequence stars and is in turn succeeded by carbon, neon, oxygen, and silicon burning to reach to the elements near iron and somewhat

beyond in the periodic table. With these nuclei of intermediate mass as seeds, subsequent processes similar to Gamow's involving neutron capture at a slow rate (s-process) or at a rapid rate (r-process) continued the synthesis beyond ^{209}Bi, the last stable nucleus, up through short lived radioactive nuclei to long lived ^{232}Th, ^{235}U, and ^{238}U the parents of the natural radioactive series. This last requires the r-process which actually builds beyond mass 238 to radioactive nuclei which decay back to ^{232}Th, ^{235}U, and ^{238}U rapidly at the cessation of the process.

The need for two neutron capture processes was provided by Suess and Urey (2). With the adroit use of relative isotopic abundances for elements with several isotopes they demonstrated the existence of the double peaks (r and s) in Figure 2. It was immediately clear that these peaks were associated with neutron shell filling at the magic neutron numbers N = 50,82, and 126 in the nuclear shell model of Hans Jensen and Maria Goeppert-Mayer who won the Nobel Prize in Physics just twenty years ago.

In the s-process the nuclei involved have low capture cross-sections at shell closure and thus large abundances to maintain the s-process flow. In the r-process it is the proton-deficient radioactive progenitors of the stable nuclei which are involved. Low capture cross-sections and small beta-decay rates at shell closure lead to large abundances but after subsequent radioactive decay these large abundances appear at lower A values than for the s-process since z is less and thus A = N + Z is less. In Hoyle's classic papers (12) stellar nucleosynthesis up to the iron group elements was attained by charged particle reactions. Rapidly rising Coulomb barriers for charged particles curtailed further synthesis. Suess and Urey (2) made the breakthrough which led to the extension of nucleosynthesis in stars by neutrons unhindered by Coulomb barriers all the way to ^{238}U.

The complete run of the synthesis of the elements in stars was incorporated into a paper by Burbidge, Burbidge, Fowler, and Hoyle (18), commonly referred to as B^{2}FH, and was independently developed by Cameron (19). Notable contributions to the astronomical aspects of the problem were made by Jesse Greenstein (20) and by many other observational astronomers. Since that time Nuclear Astrophysics has developed into a full-fledged scientific activity including the exciting discoveries of isotopic anomalies in meteorites by my colleagues Gerald Wasserburg, Dimitri Papanastassiou and Samuel Epstein and many other cosmochemists. What follows will highlight a few of the many experiments and theoretical researches under way at the present time or carried out in the past few years. This account will emphasize research activities in the Kellogg Laboratory because they are closest to my interest and knowledge. However, copious references to the work of other laboratories and institutions are cited in the hope that the reader will obtain a broad view of current experimental and theoretical studies in Nuclear Astrophysics.

This account cannot discuss the details of the nucleosynthesis of all the elements and their isotopes which would, for a given nuclear species, involve discussing all the reactions producing that nucleus and all those which destroy it. The reader will find some of these details for ^{12}C, ^{16}O and ^{55}Mn.

It will be noted that the measured cross sections for the reactions are customarily very small at the lowest energies of measurement, for $^{12}C(\alpha,\gamma)^{16}O$ even less than one nanobarn (10^{-33} cm^2) near 1.4 MeV. This means that experimental Nuclear Astrophysics requires accelerators with large currents of well focussed, monoenergic ion beams, thin targets of high purity and stability, detectors of high sensitivity and energy resolution and experimentalists with great tolerance for the long running times required and with patience in accumulating data of statistical significance. Classical Rutherfordian measurements of nuclear cross sections are required in experimental nuclear astrophysics and the results are in turn essential to our understanding of the physics of nuclei.

A comment on nuclear reaction notation is necessary at this point. In the reaction $^{12}C(\alpha,\gamma)^{16}O$ discussed in the previous paragraph ^{12}C is the laboratory target nucleus, *a* is the incident nucleus (^4He) accelerated in the laboratory, γ is the photon produced and detected in the laboratory, and ^{16}O is the residual nucleus which can also be detected if it is desireable to do so. If ^{12}C is accelerated against a gas target of ^4He and the ^{16}O-products are detected but not the gamma rays then the laboratory notation is $^4He(^{12}C,^{16}O)y$. The stars could not care less. In stars all the particles are moving and only the center-of-momentum system is important for the determination of stellar reaction rates. In $^{12}C(\alpha,n)^{15}O(e^+\nu)^{15}N$, *n* is the neutron promptly produced and detected and e^+ is the beta-delayed positron which can also be detected. The neutrino emitted with the position is designated by ν.

As an aside at this point I am proud to recall that I first spoke to the Royal Swedish Academy of Sciences on "Nuclear Reactions in Stars" on January 26, 1955. It does not seem so long ago and some of you in the audience heard that talk!

III. Stellar Reaction Rates from Laboratory Cross Sections

Thermonuclear reaction rates in stars are customarily expressed as N_A <a υ> reactions per second per (mole cm^{-3}) where N_A = 6.022 x 10^{23} mole^{-1} is Avogadro's number and $\langle\sigma\upsilon\rangle$ is the Maxwell-Boltzmann average as a function of temperature for the product of the reaction cross section, σ, in cm^2, and the relative velocity of the reactants, *v* in cm sec^{-1}. Multiplication of $\langle\sigma\upsilon\rangle$ by the product of the number densities per cm^3 of the two reactants is necessary to obtain rates in reactions per second per cm^3. N_A is incorporated so that mass fractions can be used as described in detail in Fowler, Caughlan and Zimmerman (2 1). These authors also describe procedures for reactions involving more than two reactants and give analytical expressions for reactions mainly involving γ, e, *n*, p and *a* with nuclei having atomic mass number $A \leq 30$. Bose-Einstein statistics for γ have been necessarily incorporated but the extension to Fermi-Dirac statistics for degenerate e, n and p and the extension to Bose-Einstein statistics for *a* are not included. Factors for calculating reverse reaction rates are given.

Early work on the evaluation of stellar reaction rates from experimental laboratory cross sections was reviewed in Bethe's Nobel Lecture (11). Fowler,

DEFINITION OF TIME S-FACTOR (BETHE 1967) AS A FUNCTION OF REACTION ENERGY(E)

$$\sigma(E) = \pi\lambda\!\!\!^{-2} \times P \times \text{INTRINSIC NUCLEAR FACTOR}$$

$$\pi\lambda\!\!\!^{-2} \propto E^{-1} \qquad \lambda\!\!\!^{-} = \text{DE BROGLIE WAVE LENGTH}/2\pi$$

$$P(E) = \text{GAMOW PENETRATION FACTOR}$$

$$\propto \exp(-E_G^{\frac{1}{2}}/E^{\frac{1}{2}}) \qquad E_G \approx Z_0^2 Z_1^2 A \quad \text{MeV}$$

$$S(E) \equiv E\sigma(E)\ \exp(+E_G^{\frac{1}{2}}/E^{\frac{1}{2}})$$

$$S(E) \begin{cases} \text{PERMITS MORE PRECISE EXTRAPOLATION FROM} \\ \text{LOWEST ENERGY MEASUREMENTS IN LABORATORY} \\ \text{TO VERY LOW EFFECTIVE STELLAR ENERGIES} \end{cases}$$

Table 1.

Caughlan and Zimmerman (21) have provided detailed numerical and analytical procedures for converting laboratory cross sections into stellar reaction rates. It is first of all necessary to accommodate the rapid variation of the nuclear cross sections at low energies which are relevant in astrophysical circumstances. For neutron induced reactions this is accomplished by defining a cross-section S-factor equal to the cross section (σ) multiplied by the interaction velocity (v) in order to eliminate the usual v^{-1} singularity in the cross section at low velocities and low energies.

For reactions induced by charged particles such as protons, alpha particles or the heavier ^{12}C, ^{16}O... nuclei it is necessary to accommodate the decrease by many orders of magnitude from the lowest laboratory measurements to the energies of astrophysical relevance. This is done in the way first suggested by E. E. Salpeter (22) and emphasized by the second of references Bethe (11). Table 1 shows how a relatively slowly varying S-factor can be defined by eliminating the rapidly varying term in the Gamow penetration factor governing transmission through the Coulomb barrier. The cross section is usually expressed in barns (10^{-24}c m^2) and the energy in MeV (1.602×10^{-6} erg) so the S-factor is expressed in MeV-barns although keV-barns is sometimes used. In Table 1, the two charge numbers and the reduced mass in atomic mass units of the interacting nuclei are designated by Z_0, Z_1, and *A*. Table 2 then shows how stellar reaction rates can be calculated as an average over the Maxwell-Boltzmann distribution for both nonresonant and resonant cross sections. In Table 2 the effective stellar reaction energy is given numerically by

STELLAR REACTION RATES AS
FUNCTIONS OF TEMPERATURE(T)

$$\langle \sigma v \rangle_{MB} = f(T) \propto T^{-3/2} \int S(E) \exp(-E_G^{1/2}/E^{1/2} - E/kT)\, dE$$

MB ≡ AVERAGE OVER MAXWELL-BOLTZMANN DISTRIBUTION

MAXIMUM IN INTEGRAND OCCURS AT E_r AND AT

E_o ≡ EFFECTIVE STELLAR REACTION ENERGY $\propto E_G^{1/3} T^{2/3}$

NONRESONANT RATE

$$\langle \sigma v \rangle_{nr} \propto S(E_o) T^{-2/3} \exp(-3E_o/kT) \qquad E_o/kT \propto T^{-1/3}$$

RESONANT RATE

$$\langle \sigma v \rangle_r \propto S(E_r) T^{-3/2} \exp(-E_r/kT)$$

E_r = ENERGY AT RESONANCE

Table 2.

$E_0 = 0.122(Z_0^2 Z_1^2 A)^{1/3} T_9^{2/3}$ MeV where T_9 is the temperature in units of 10^9K. Expressions for reaction rates derived from theoretical statistical model calculations are given by Woosley, Fowler, Holmes, and Zimmerman (23).

It is true that the extrapolation from the cross sections measured at the lowest laboratory energies to the cross sections at the effective stellar energy can often involve a decrease by many orders of magnitude. However the elimination of the Gamow penetration factor, which causes this decrease, is based on the solution of the Schroedinger equation for the Coulomb wave functions in which one can have considerable confidence. The main uncertainty lies in the variation of the S-factor with energy which depends primarily on the value chosen for the radius at which formation of a compound nucleus between two interacting nuclei or nucleons occurs as discussed long ago in reference (18). The radii used by my colleagues and me in recent work are given in reference (23). There is, in addition, the uncertainty in the *intrinsic nuclear factor* of Table 1 which can only be eliminated by recourse to laboratory experiments. The effect of a resonance in the compound nucleus just below or just above the threshold for a given reaction can often be ascertained by determination of the properties of the resonance in other reactions in which it is involved and which are easier to study.

IV. *Hydrogen Burning in Main Sequence Stars and the Solar Neutrino Problem*

Hydrogen burning in Main Sequence stars has contributed at the present time only about 20 percent more helium than that which resulted from the big bang.

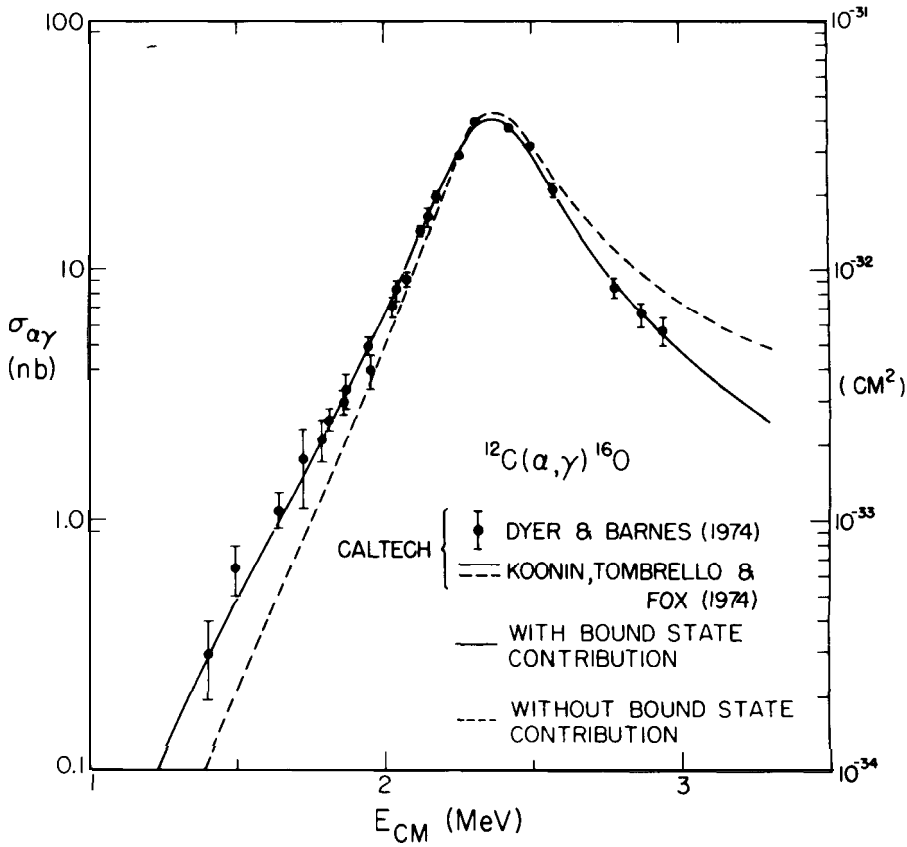


Figure 4. The cross section in nanobarns *(nb)* versus center-of-momentum energy in Mev for $^{12}C(a, \gamma)^{16}O$ measured by Dyer and Barnes (35) and compared with theoretical calculations by Koonin, Tombrello and Fox (see 35).

However, hydrogen burning in the sun has posed a problem for many years. In 1938 Bethe and Critchfield (24) proposed the proton-proton or pp-chain as one mechanism for hydrogen burning in stars. From many cross-section measurements in Kellogg and elsewhere it is now known to be the mechanism which operates in the sun rather than the CNO-cycle.

Our knowledge of the weak nuclear interaction (beta decay, neutrino emission and absorption, etc.) tells us that two neutrinos are emitted when four hydrogen nuclei are converted into helium nuclei. Detailed elaboration of the pp-chain by Fowler (25) and Cameron (26) showed that a small fraction of these neutrinos, those from the decay of ^{7}Be and ^{8}B, should be energetic enough to be detectable through interaction with the nucleus ^{37}Cl to form radioactive ^{37}Ar, a method of neutrino detection suggested by Pontecorvo (27) and Alvarez (28). Raymond Davis (29) and his collaborators have attempted for more than 25 years to detect these energetic neutrinos employing a 380,000 liter tank of perchloroethylene ($C_2{}^{35}Cl_3{}^{37}Cl_1$) located one mile deep in the Homestake Gold Mine in Lead, South Dakota. They find only about one quarter of the number expected on the basis of the model dependent calculations of Bahcall et al. (30).

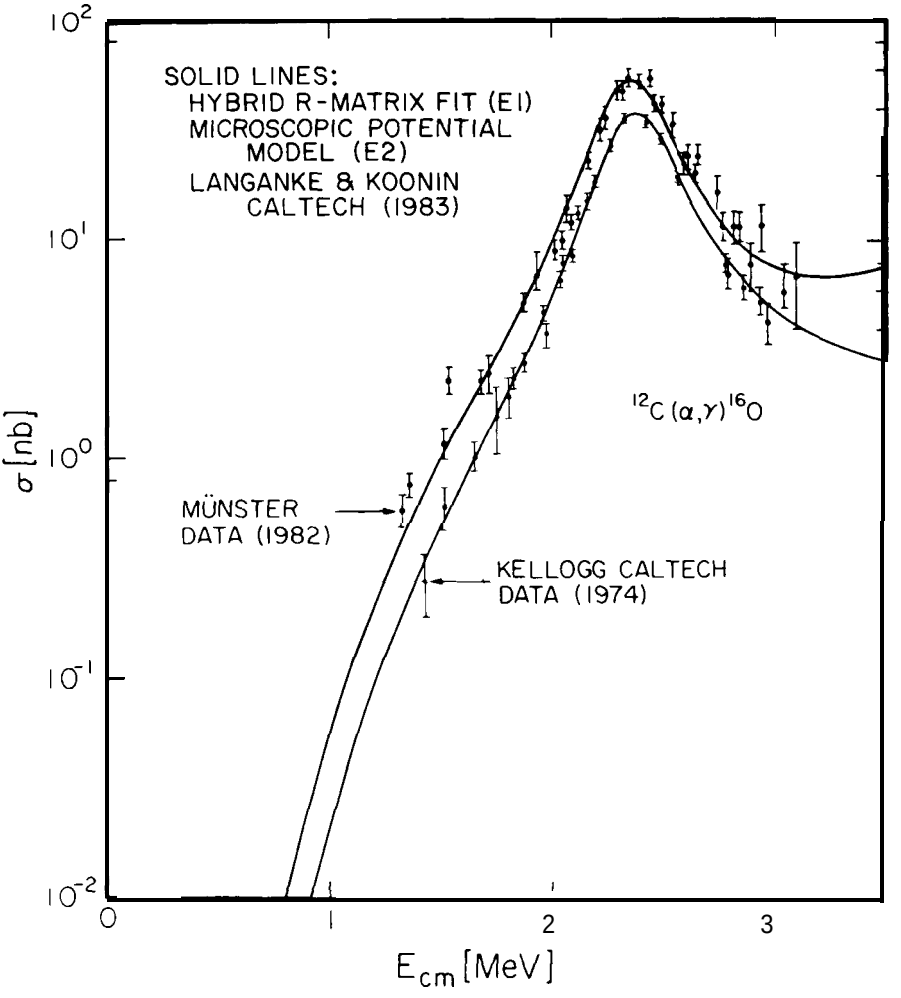


Figure 5. The cross section in nanobarns *(nb)* versus center-of-momentum energy in MeV for $^{12}C(\alpha,\gamma)^{16}O$. The Miinster data was obtained by Kettner *et al.* (36) and the Kellogg Caltech data was obtained by Dyer and Barnes (35). The solid lines are theoretical calculations made by Langanke and Koonin (34).

Something is wrong-either the standard solar models are incorrect, the relevant nuclear cross sections are in error, or the electron-type neutrinos produced in the sun are converted in part into undetectable muon neutrinos or tauon neutrinos on the way from the sun to the earth. There indeed have been controversies about the nuclear cross sections which have been for the most part resolved as reviewed in Robertson *et al.* and Osborne *et al.* (31) and Skelton and Kavanagh (32).

It is generally agreed that the next step is to build a detector which will detect the much larger and model independent flux of low energy neutrinos from the sun through neutrino absorption by the nucleus ^{71}Ga to form radioactive ^{71}Ge. This will require 30 to 50 tons of gallium at a cost (for 50 tons) of approximately 25 million dollars or 200 million Swedish crowns. An interna-

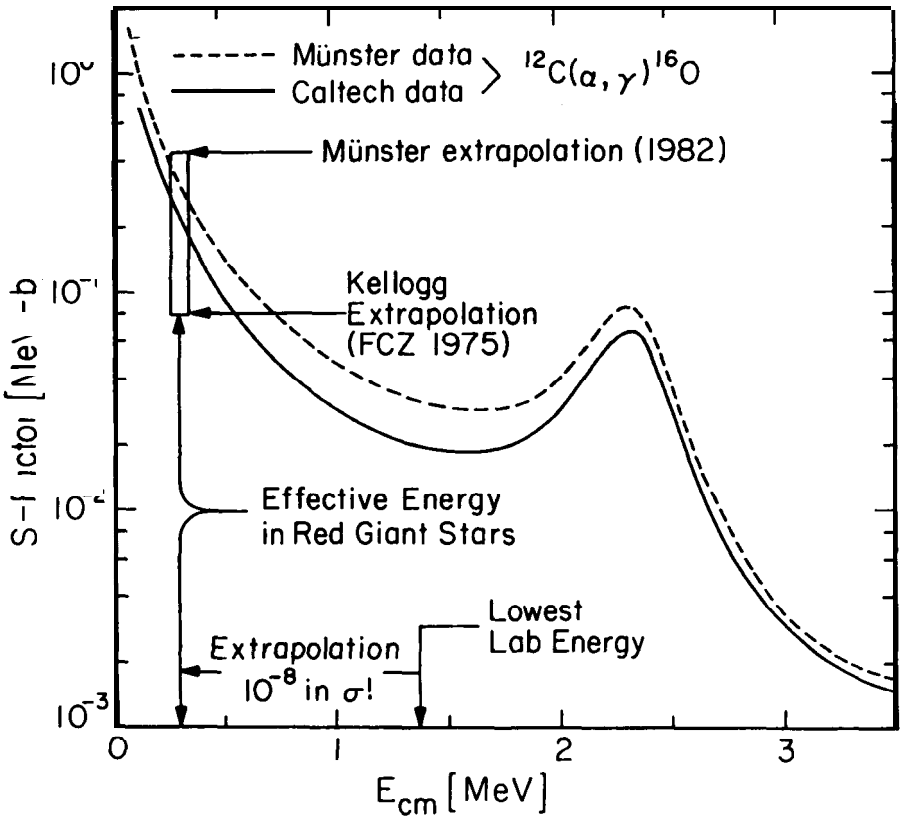


Figure 6. The cross section factor, S in MeV-barns, versus center-of-momentum energy in MeV for $^{12}C(\alpha,\gamma)^{16}O$. The dashed and solid curves are the theoretical extrapolations of the Miinster and Kellogg Caltech data by Langanke and Koonin (34).

tional effort is being made to obtain the necessary amount of gallium. We are back at square one in Nuclear Astrophysics. Until the solar neutrino problem is resolved the basic principles underlying the operation of nuclear processes in stars are in question. A gallium detector should go a long way toward resolving the problem.

The Homestake detector must be maintained in low level operation until the chlorine and gallium detectors can be operated at full level simultaneously. Otherwise endless conjecture concerning time variations in the solar neutrino flux will ensue. Morever the results of the gallium observations may uncover information that has been overlooked in the past chlorine observations. In the meantime bromine could be profitably substituted for chlorine in the Homestake detector. The chlorine could eventually be resubstituted.

The CNO-cycle operates at the higher temperatures which occur during hydrogen burning in Main Sequence stars somewhat more massive than the sun. This is the case because the CNO-cycle reaction rates rise more rapidly with temperature than do those of the pp-chain. The cycle is important because ^{13}C, ^{14}N, ^{15}N, ^{17}O, and ^{18}O are produced from ^{12}C and ^{16}O as seeds. The role of these nuclei as sources of neutrons during helium burning is discussed in Section V.

V. *The Synthesis of* ^{12}C *and* ^{16}O *and Neutron Production in Helium Burning*

The human body is 65% oxygen by mass and 18% carbon with the remainder mostly hydrogen. Oxygen (0.85%) and carbon (0.39%) are the most abundant elements heavier than helium in the sun and similar Main Sequence stars. It is little wonder that the determination of the ratio ^{12}C /^{16}O produced in helium burning is a problem of paramount importance in Nuclear Astrophysics. This ratio depends in a fairly complicated manner on the density, temperature and duration of helium burning but it depends directly on the relative rates of the 3a $\rightarrow$ ^{12}C process and the ^{12}C(c~,y)'~0 process. If 3a $\rightarrow$ ^{12}C is much faster than ^{12}C(CZ,~)'~~ then no ^{16}O is produced in helium burning. If the reverse is true then no ^{12}C is produced. For the most part the subsequent reaction $^{16}O(\alpha,\gamma)^{20}Ne$ is slow enough to be neglected.

There is general agreement about the rate of the 3a $\rightarrow$ ^{12}C process as reviewed by Barnes (33). However, there is a lively controversy at the present time about the laboratory cross section for $^{12}C(\alpha,\gamma)^{16}O$ and about its theoretical extrapolation to the low energies at which the reaction effectively operates. The situation is depicted in Figures 4, 5 and 6 taken with some modification from Langanke and Koonin (34), Dyer and Barnes (35) and Kettner *et al. (36).* The Caltech data obtained in the Kellogg Laboratory is shown as the experimental points in Figure 4 taken from Dyer and Barnes (35) who compared their results with theoretical calculations by Koonin, Tombrello and Fox (see 35). The Münster data is shown as the experimental points in Figure 5 taken from Kettner *et al.* (36) in comparison with the data of Dyer and Barnes (35). The theoretical curves which yield the best tit to the two sets of data are from Langanke and Koonin (34).

The crux of the situation is made evident in Figure 6 which shows the extrapolations of the Caltech and Miinster cross section factors from the lowest measured laboratory energies (~1.4MeV) to the effective energy ~0.3MeV, at T = 1.8 x 10^8K, a representative temperature for helium burning in Red Giant stars. The extrapolation in cross sections covers a range of 10^{-8}! The rise in the cross section factor is due to the contributions of two bound states in the ^{16}O nucleus just below the $^{12}C(\alpha,\gamma)^{16}O$ threshold as clearly indicated in Figure 4. It is these contributions plus differences in the laboratory data which produce the current uncertainty in the extrapolated S-factor. Note that Langanke and Koonin (34) increase the 1975 extrapolation of the Caltech data by Fowler, Caughlan, and Zimmerman (21) by a factor of 2.7 and lower the 1982 extrapolation of the Miinster data by 23%. There remains a factor of 1.6 between their extrapolation of the Miinster data and of the Caltech data. There is a lesson in all of this. The semiempirical extrapolation of their data by the experimentalists, Dyer and Barnes (35), was only 30% lower than that of Langanke and Koonin (34) and their quoted uncertainty extended to the value of Langanke and Koonin (34). Caughlan *et al.* (21) will tabulate the analysis of the Caltech data by Langanke and Koonin (34).

With so much riding on the outcome it will come as no surprise that both laboratories are engaged in extending their measurements to lower energies with higher precision. In the discussion of quasistatic silicon burning in what

follows it will be found that the abundances produced in that stage of nucleosynthesis depend in part on the ratio of ^{12}C to ^{16}O produced in helium burning and that the different extrapolations shown in Figure 6 are in the range crucial to the ultimate outcome of silicon burning. These remarks do not apply to explosive nucleosynthesis.

Recently the ratio of ^{12}C to ^{16}O produced under the special conditions of helium flashes during the asymptotic giant phase of evolution has become of great interest. The hot blue star PG 1159-035 has been found to undergo nonradial pulsations with periods of 460 and 540 seconds and others not yet accurately determined. The star is obviously highly evolved having lost its hydrogen atmosphere, leaving only a hot dwarf of about 0.6 solar masses behind. Theoretical analysis of the pulsations by Starrfield et al. and Becker (37) requires substantial amounts of oxygen in the pulsation-driving regions where the oxygen is alternately ionized and deionized. Carbon is completely ionized in these regions and only diminishes the pulsation amplitude. It is not yet clear that sufficient oxygen is produced in helium flashes which certainly involve $3\alpha + ^{12}C$ but may not last long enough for $^{12}C(a,y)\ ^{16}O$ to be involved. The problem may not lie in the nuclear reaction rates according to references (37). We shall see!

In what follows in this paper β^+-decay is designated by $(e^+\nu)$ since both a positron (e') and a neutrino *(Y)* are *emitted.* Similarly β^--decay will be designated by $(e^-\bar{\nu})$ since both an electron (e^-) and antineutrino $(\bar{\nu})$ are *emitted.* Electron capture (often indicated by E) will be designated by (e^-,ν), the comma indicating that an electron is captured and a neutrino emitted. The notations $(e^+,\bar{\nu}),(\nu,e^-)$ and $(\bar{\nu},e^+)$ should now be obvious.

Neutrons are produced when helium burning occurs under circumstances in which the CNO-cycle has been operative in the previous hydrogen burning. When the cycle does not go to completion copious quantities of ^{13}C are produced in the sequence of reactions $^{12}C(P,Y)^{13}Ne(\sim+\sim)^{13}C$. In subsequent helium burning, neutrons are produced by $^{13}C(a,n)^{16}O$. When the cycle goes to completion the main product (>95%) is ^{14}N In subsequent helium burning, ^{18}O and ^{22}Ne are produced in the sequence of reactions $^{14}N(n,y'j^{18}F(e+\nu)^{18}O(\alpha,\gamma)^{22}Ne$ and these nuclei in turn produce neutrons through $^{1}O(\alpha,n)\ ^{21}Ne(\alpha,n)^{24}Mg$ and $^{22}Ne(\alpha,n)^{25}Mg$. However, the astrophysical circumstances and sites under which the neutrons produce heavy elements through the s-process and the r-process are, even today, matters of some controversy and much study (See Section XI).

VI. *Carbon, Neon, Oxygen, and Silicon Burning*

The advanced burning processes discussed in this section involve the network of reactions shown in Figure 7. Because of the high temperature at which this network can operate, radioactive nuclei can live long enough to serve as live reaction targets. In addition excited states of even the stable nuclei are populated and also serve as targets. The determination of the nuclear cross sections and stellar rates of the approximately 1000 reactions in the network has

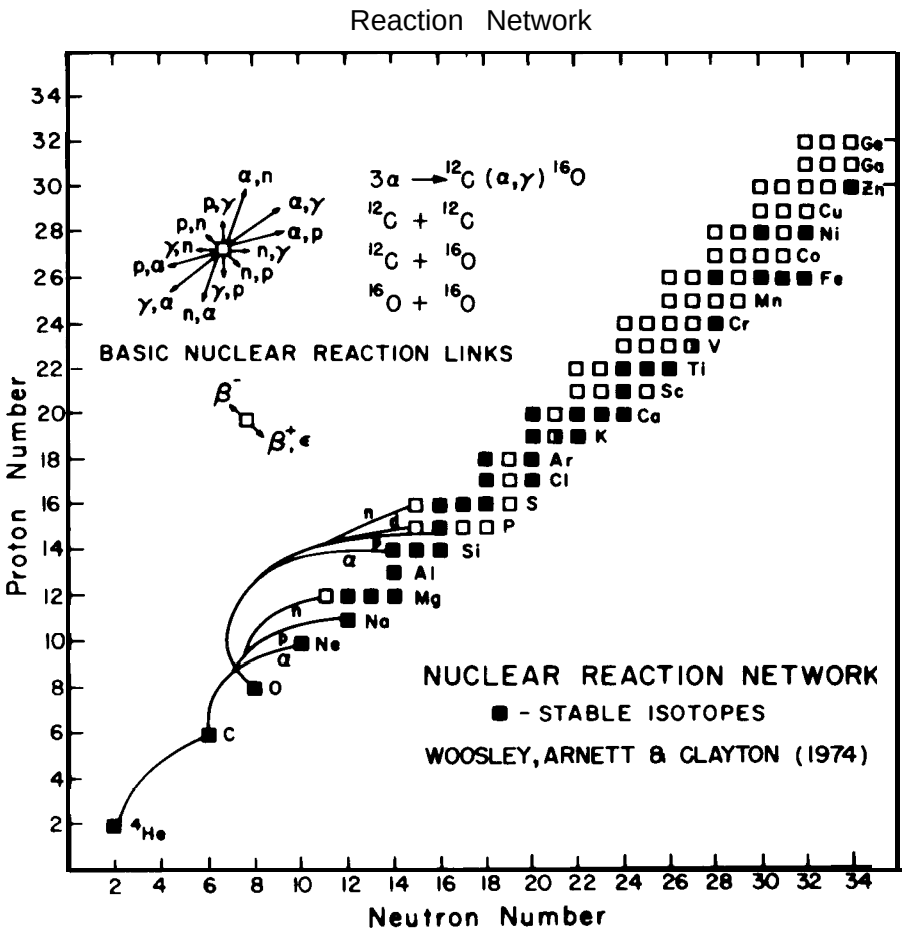


Figure 7. The reaction network for nucleosynthesis involving the most important stable and radioactive nuclei with $N = 2$ to 34 and $Z = 2$ to 32. Stable nuclei are indicated by solid squares. Radioactive nuclei are indicated by open squares. Excited states of both are involved in the reaction network.

involved and will continue to involve extensive experimental and theoretical effort.

The following discussion applies to massive enough stars such that electron degeneracy does not set in as nuclear evolution proceeds through the various burning stages discussed in this section. In less massive stars electron-degeneracy can terminate further nuclear evolution at certain stages with catastrophic results leading to the disruption of the stellar system. The reader will find Figure 8, especially 8(a), instructive in following the discussion in this section. Figure 8 is taken from Woosley and Weaver (38) and a much more detailed recent version is shown in Figure 9 from Weaver, Woosley and Fuller (39). Figure 8(a) applies to the preexplosive stage of a young (Population I) star of 25 solar masses and shows the result of various nuclear burnings in the following mass zones: (1)> l0M,, convective envelope with the results of some CNO-burning; (2) 7-10 $M_{\odot}$, products mainly of H-burning; (3)$6.5-7M_{\odot}$,

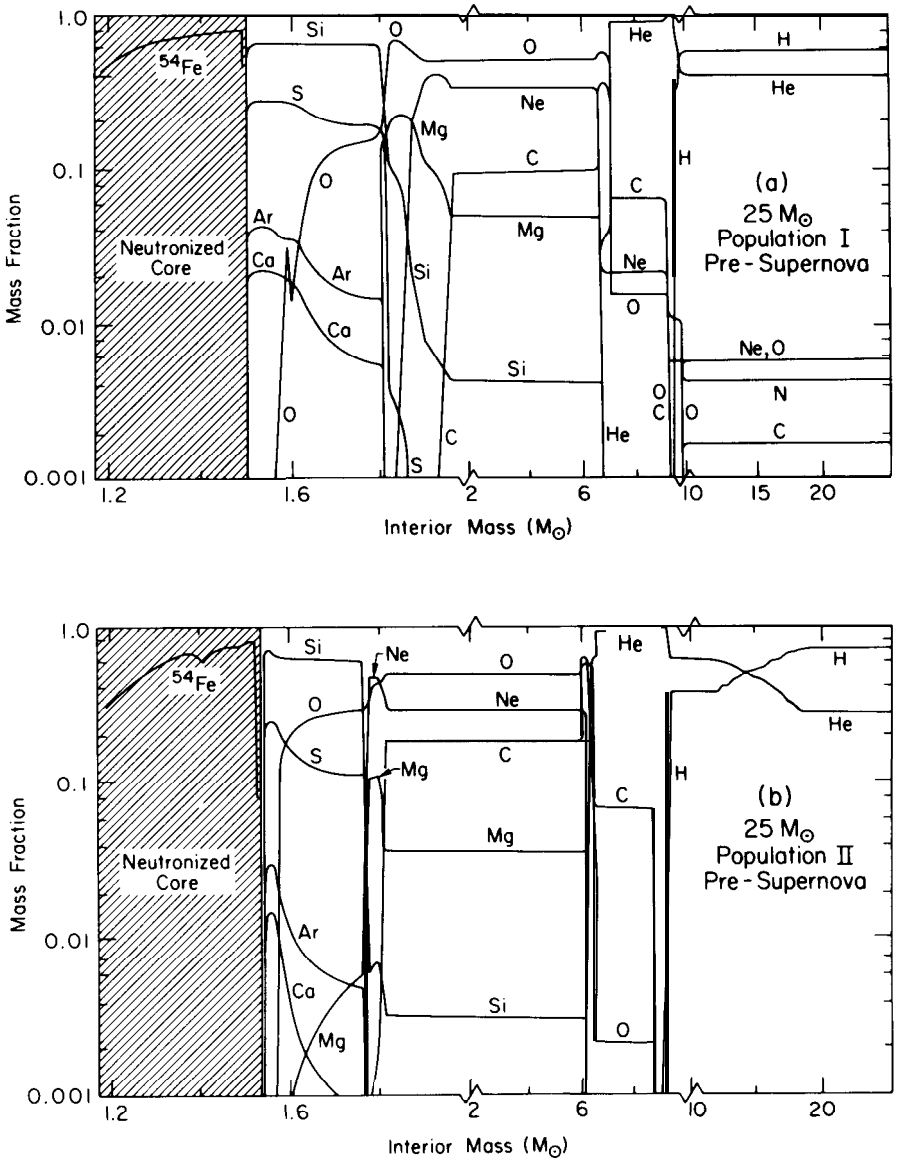


Figure 8. Pre-supernova abundances by mass fraction versus increasing interior mass in solar masses, $M_{\odot}$, measured from zero at the stellar center to 25 $M_{\odot}$, the total stellar mass from Woosley and Weaver (38). (a) Population I star. (b) Population II star.

products of He-burning; (4) $1.9-6.5M_{\odot}$ products of C-burning; (5) 1.8-1.9Mo products of Ne-burning; (6) 1.5-1.8M,, products of O-burning; (7)< $1.5M_{\odot}$, the products of S-burning in the partially neutronized core are not shown in detail but consist mainly of ^{54}Fe as well as substantial amounts of other neutron-rich nuclei such as ^{48}Ca, ^{50}Ti, ^{54}Cr and ^{58}Fe. ^{54}Fe, ^{48}Ca and ^{50}Ti have $N = 28$, for which a neutron subshell is closed. Both Figures 8(a) and 8(b) have been evaluated shortly after photodisintegration has initiated core collapse which will then be subsequently sustained by the reduction of the outward

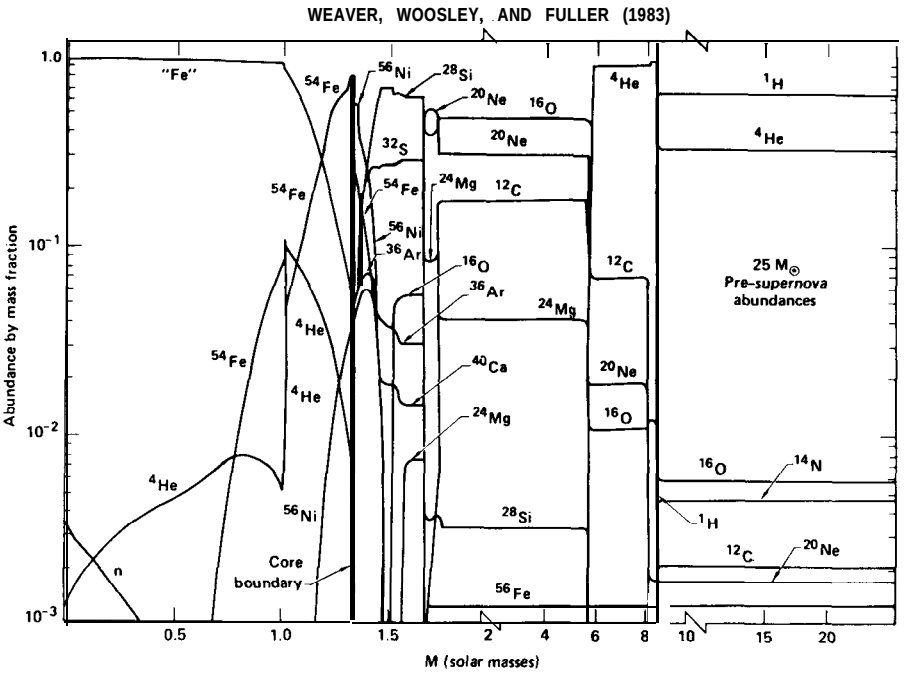


Figure 9. Pre-supernova abundances by mass fraction versus increasing interior mass for a Population I star with total mass equal to 25 $M_{\odot}$ from Weaver, Woosley and Fuller (39).

pressure through electron-capture and the resulting almost complete neutronization of the core.

It must be realized that the various burning stages took place originally over the central regions of the star and finally in a shell surrounding that region. Subsequent stages modify the inner part of the previous burning stage. For example, in the 25 solar mass Population I star of Figure 8(a), C-burning took place in the central 6.5 solar masses of the star but the inner 1.9 solar masses were modified by subsequent Ne-, 0- and Si-burning.

Helium burning produces a stellar core consisting mainly of ^{12}C and ^{16}O. After core contraction the temperature and density rise until carbon burning through $^{12}C + ^{12}C$ fusion is ignited. The S-factor for the total reaction rate shown in Figure 10 has been taken from page 213 of reference (33) and is based on measurements in a number of laboratories. The extrapolation to the low energies of astrophysical relevance is uncertain as Figure 10 makes clear and more experimental and theoretical studies are urgently needed. At the lowest bombarding energy, 2.4 MeV, the cross section is $\sim 10^{-8}$ barns. For a representative burning temperature of 6×10^{8} K the effective energy is E_0 = 1.7 MeV and the extrapolated cross section is $\sim 10^{-13}$ barns. The main product of carbon burning is ^{20}Ne produced primarily in the $^{12}C(^{12}C,\alpha)^{20}Ne$ reaction. The reactions $^{12}C\ (^{12}C, p)^{23}Na$ and $^{12}C(^{12}C, n)^{23}Mg(e^{+}\nu)^{23}Na$ also occur as well as many secondary reactions such as $^{23}Na(p,\alpha)^{20}Ne$. When the ^{12}C is exhausted, ^{20}Ne and ^{16}O are the major remaining constituents. As the temperature rises,

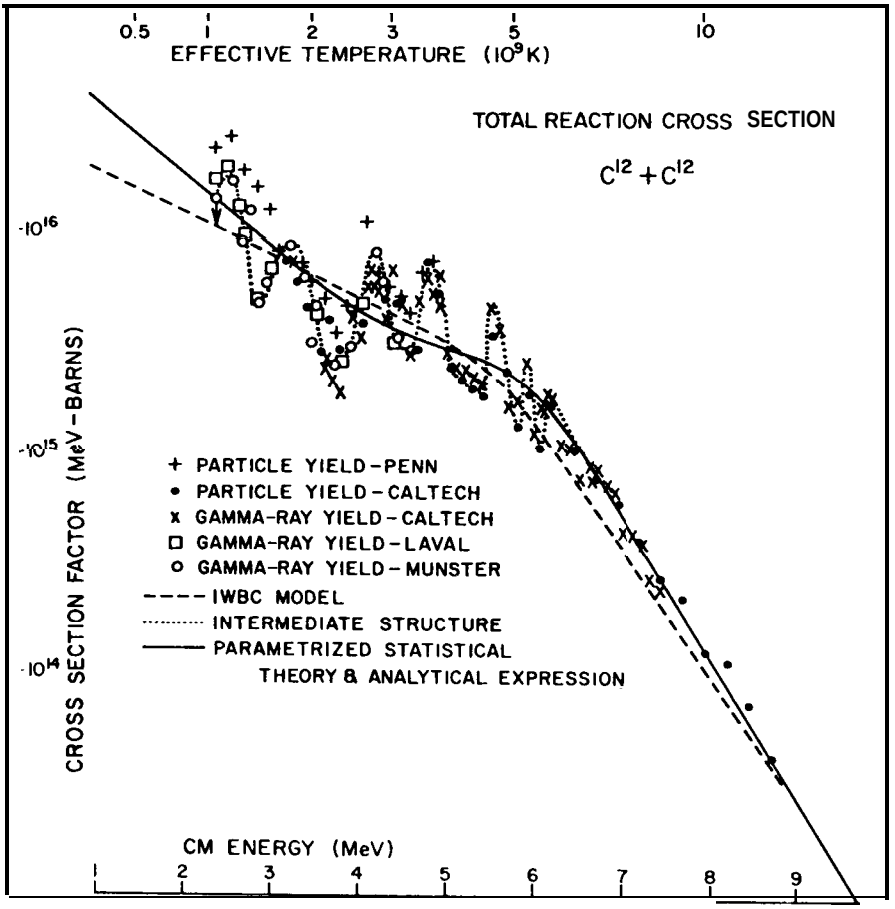


Figure 10. The total cross-section factor in MeV-barns versus center-of-momentum energy in MeV for the fusion of ^{12}C and ^{12}C. The experimental data from several laboratories are shown along with schematic intermediate structure in the dotted curve. Two parametrized adjustments to the data, ignoring intermediate structure, are shown in the dashed and solid curves.

from further gravitational contraction, the ^{20}Ne is destroyed by photodisintegration, $^{20}Ne(\gamma,\alpha)^{16}O$. This occurs because the alpha-particle in ^{20}Ne is bound to its closed-shell partner, ^{16}O, by only 4.731 MeV. In ^{16}O, for example, the binding of an alpha-particle is 7.162 MeV.

The next stage is oxygen burning through $^{16}O + ^{16}O$ fusion. The S-factor for the total reaction rate is shown in Figure 11 and is based entirely on data obtained in the Kellogg Laboratory at Caltech. The work of Hulke, Rolfs, and Trautvetter (40) using gamma-ray detection is in fair agreement with the gamma-ray measurements at Caltech. As in the case of $^{12}C + ^{12}C$ the extrapolation to the low energies of astrophysical relevance is uncertain although only one of many possible extrapolations is shown in Figure 11. The main product of oxygen burning is ^{28}Si through the primary reaction $^{16}O(^{16}O,a)^{28}Si$ and a number of secondary reactions. Under some conditions neutron induced reactions lead to the synthesis of significant quantities of ^{30}Si. Oxygen burning can result in nuclei with a small but important excess of neutrons over protons.

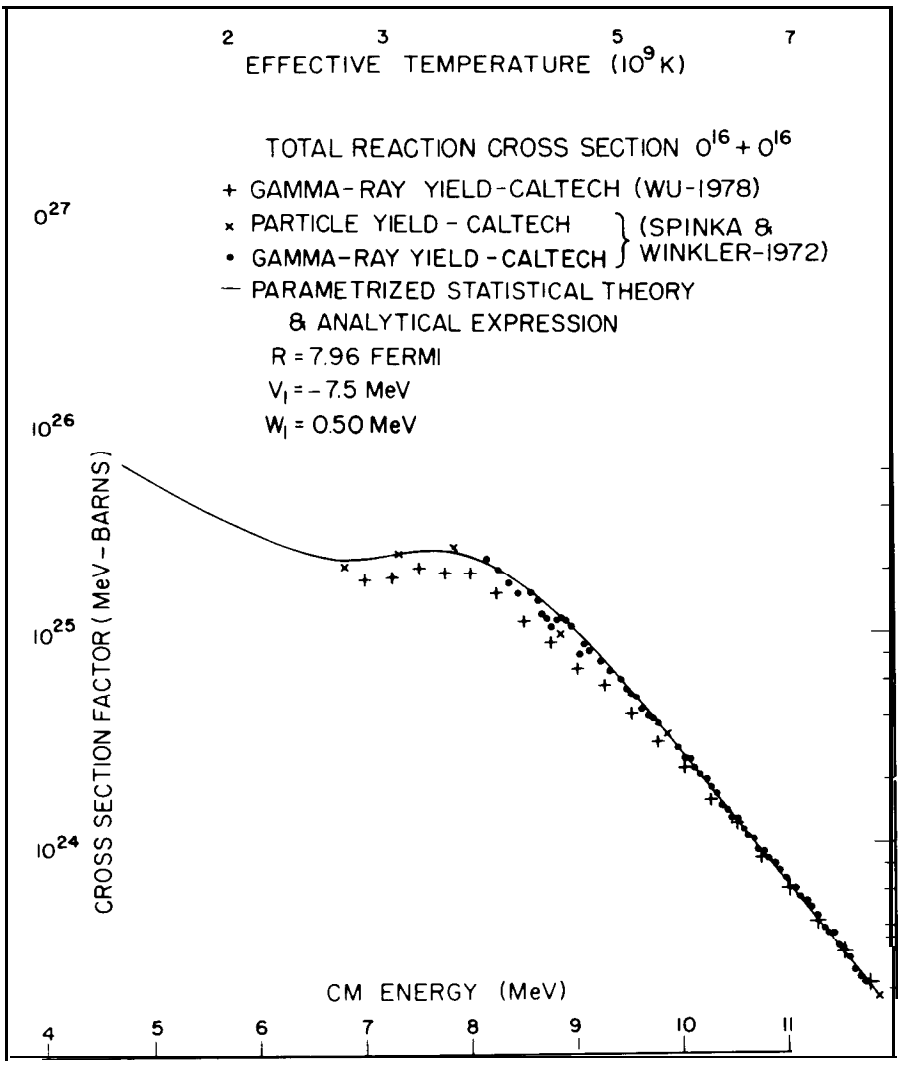


Figure II. The total cross-section factor in MeV-barns versus center-of-momentum energy in MeV for the fusion of $^{16}0 + {}^{16}O$. The experimental data from several measurements at Caltech are shown and compared with a parametrized theoretical adjustment in the solid curve.

The onset of Si-burning signals a marked change in the nature of the *fusion* process. The Coulomb barrier between two ^{28}Si nuclei is too great for fusion to produce the compound nucleus, ^{56}Ni, directly at the ambient temperatures (T_9 = 3 to 5) and densities ($\varrho = 10^5$ to 10^9 g cm^{+3}). However, the ^{28}Si and subsequent products are easily photodisintegrated by (γ,α),(γ,n) and (γ,p)-reactions. As Si-burning proceeds more and more ^{28}Si is reduced to nucleons and alpha particles which can be captured by the remaining ^{28}Si nuclei to build through the network in Figure 7 up to the iron group nuclei. The main product in explosive Si-burning is ^{56}Ni which transforms eventually through two beta-decays to ^{56}Fe.

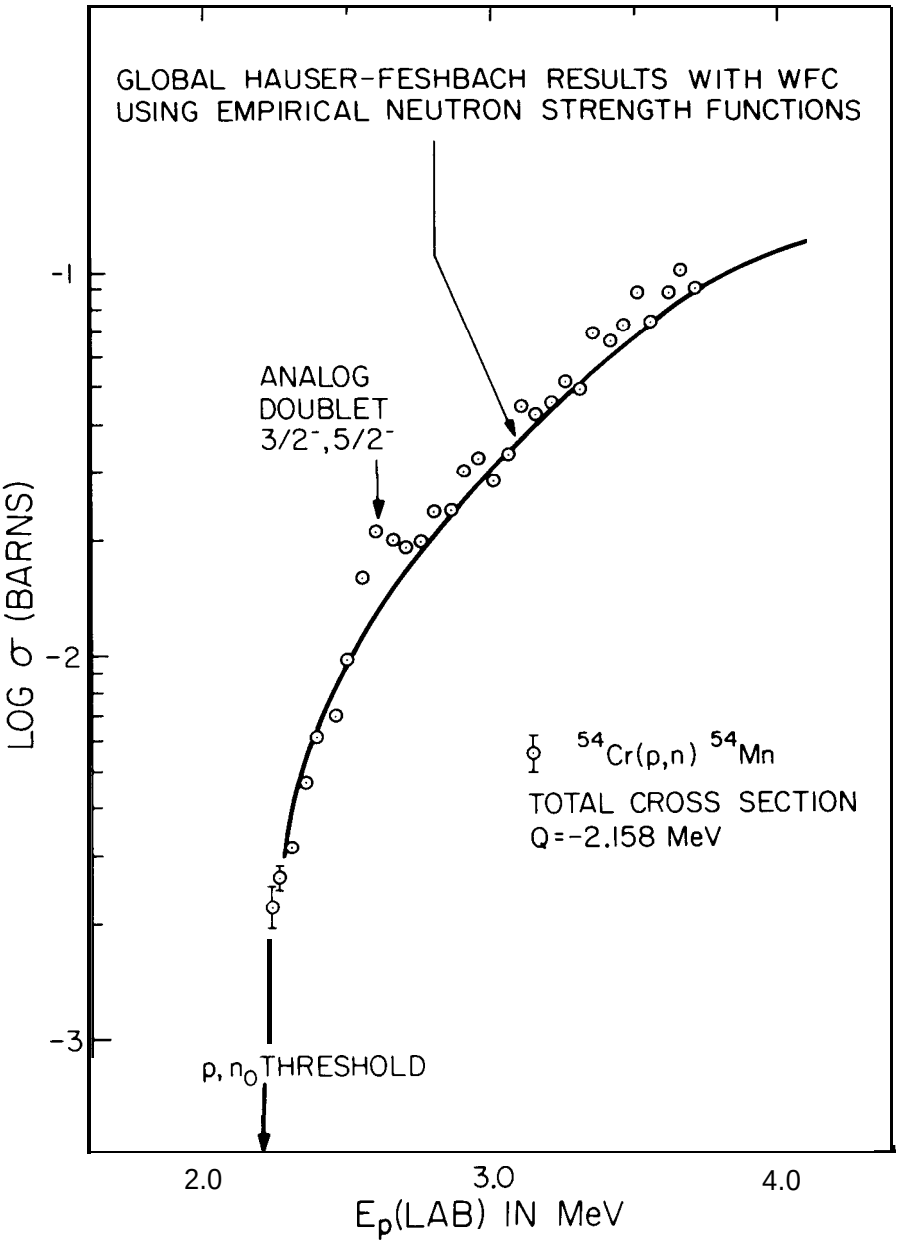


Figure 12. The total cross section in barns integrated over all outgoing angles versus laboratory proton energy in MeV for the reaction $^{54}Cr(p,n)^{54}Mn$. The data of Zysking et al. (42) are compared with unnormalized global Hauser-Feshbach calculations made by Woosley et al. (23).

In quasistatic Si-burning the weak interactions are fast enough that ^{54}Fe, with two more neutrons than protons, is the main product. Because of the important role played by alpha particles (α) and because of the inexorable trend to equilibrium (e) involving nuclei near mass 56, which have the largest binding energies per nucleon of all nuclear species, B^{2}FH (18) broke down, what is now called Si-burning, into their a-process and e-process. Quasi-

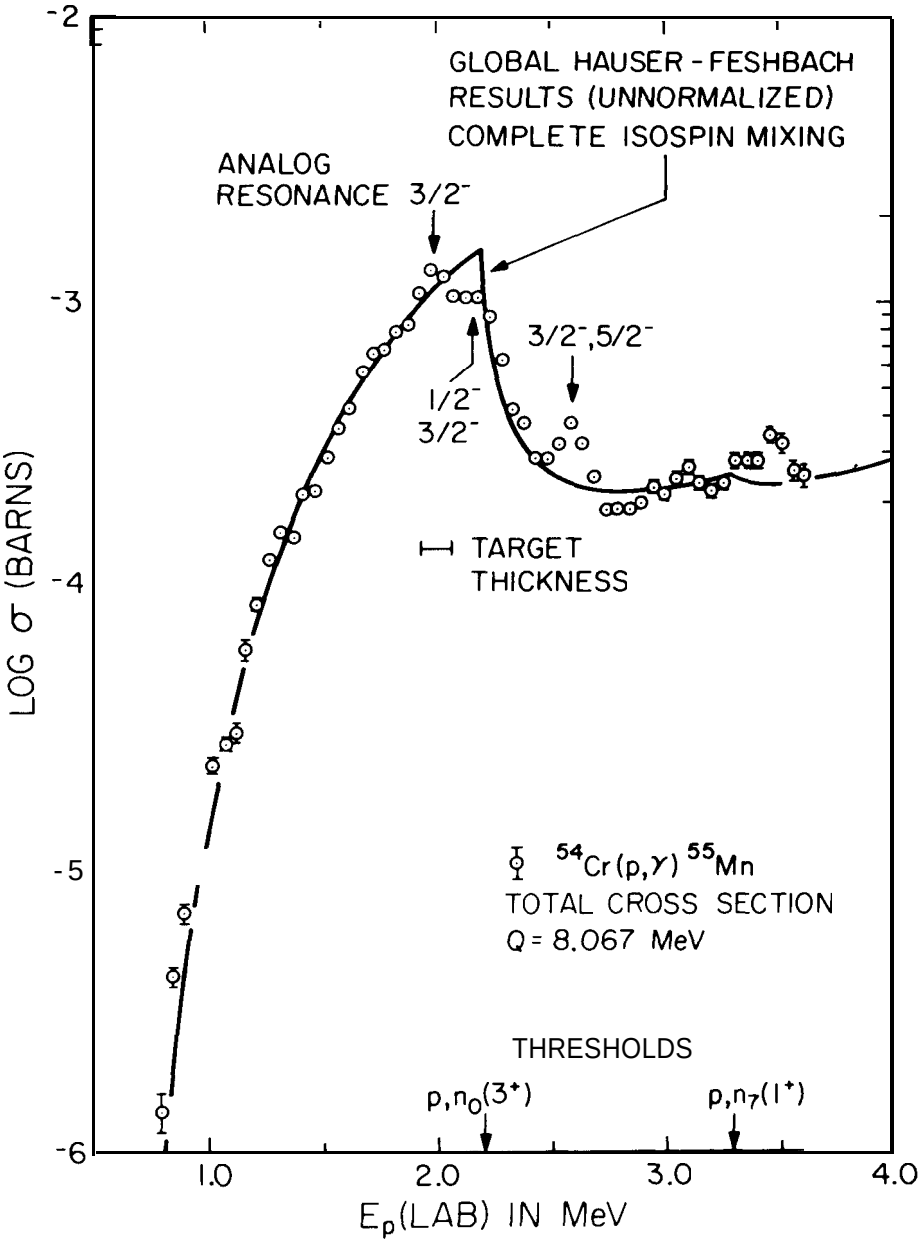


Figure 13. The total cross section in barns integrated over all outgoing angles versus laboratory proton energy in MeV for the reaction $^{54}Cr(p, \gamma)^{55}Mn$. The data of Zyskind et al. (42) are compared with unnormalized global Hauser-Feshbach calculations made by Woosley et al. (23).

equilibrium calculations for Si-burning were made by Bodansky, Clayton and Fowler (41) who cite the original papers in which the basic ideas of Si-burning were developed. Modern computers permit detailed network flow calculations to be made as discussed in references (38) and (39).

The extensive laboratory studies of Si-burning reactions are reviewed in reference (33). Figures 12 and 13 adapted from Zyskind et al. (42) show the

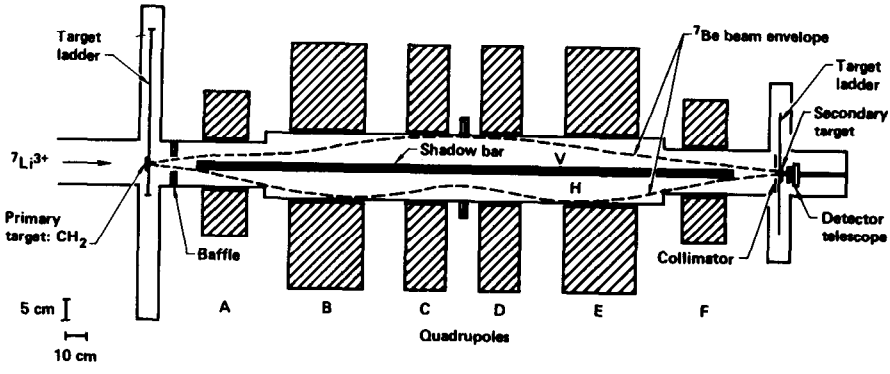


Figure 14. Radioactive beam transport system developed by Haight *et al.* (44).

laboratory excitation curves for ^{54}Cr *(p,n)* ^{54}Mn and ^{54}Cr *(p,g)* ^{55}Mn as examples. The neutrons produced in the first of these reactions will increase the number of neutrons available in Si-burning but will not contribute directly to the synthesis of ^{55}Mn as does the second reaction. In fact, above its threshold at 2.158 MeV the *(p,n)* -reaction competes strongly with the (p,γ)-reaction, which is of primary interest, and produces the pronounced *competition cusp* in the excitation curve in Figure 13. Competition in the disintegration of the compound nucleus produced in nuclear reactions was stressed very early by Niels Bohr so perhaps the cusps should be called Bohr Cusps. They arise from the same basic cause but are not the long known *Wigner* Cusps. It will be clear from Figure 13 that the rate of the $^{54}Cr(p,\gamma)^{55}Mn$ reaction at very high temperatures will be an order of magnitude lower because of the cusp than would otherwise be the case.

The element manganese has only one isotope, ^{55}Mn. The manganese in nature is produced in quasistatic Si-burning most probably through the $^{54}Cr(p,\gamma)^{55}Mn$-reaction just discussed in the previous paragraph. The reaction network extends to ^{54}Cr and then on through ^{55}Mn. $^{51}V(\alpha,\gamma)^{55}Mn$ and $^{52}V\alpha n)^{55}Mn$ may also contribute especially in explosive Si-burning. The overall synthesis of ^{55}Mn involves a balance in its production and destruction. In quasistatic Si-burning the reactions which destroy ^{55}Mn are most probably $^{55}Mn(p,\gamma)^{56}Fe$ and $^{55}Mn(pn)^{55}Fe$, which are discussed and illustrated in Mitchell and Sargood (43). $^{55}Mn(\alpha,\gamma)^{59}Co$, $^{55}Mn(\alpha,p)^{58}Fe$, and $^{55}Mn(\alpha,n)^{58}Co$ may also destroy some ^{55}Mn in explosive Si-burning. In the figures discussed in Section VIII it will be noted that calculations of the overall synthesis of ^{55}Mn yield values in fairly close agreement with the abundance of this nucleus in the solar system. Unfortunately the same can not be said about many other nuclei.

The laboratory measurements on Si-burning reactions have covered only about 20% of the reactions in the network of Figure 7 involving stable nuclei as targets. Direct measurements on short lived radioactive nuclei and the excited states of all nuclei are impossible at the present time. In this connection the production of radioactive ion-beams holds great promise for the future. Richard Boyd and Haight et al. (44) have pioneered in the development of this

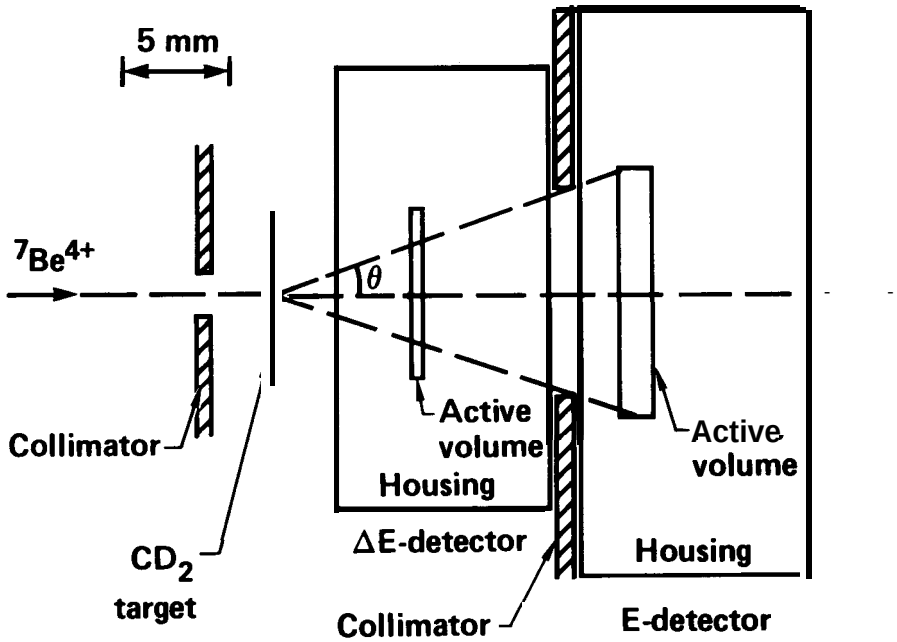


Figure 1.5. Detail of the target and detector in the radioactive beam transport system developed by Haight *et* al. (44).

technique. It will also be possible to study with this technique the reaction rates of the fairly long-lived isomeric excited states of stable nuclei. Figures 14 and 15 show the beam transport system developed by Haight *et* al. (44) which has produced accelerated beams of ^{7}Be and ^{13}N and successfully determined the cross section of the reaction ^{2}H(^{7}Be,^{8}B)*n* to be 59±11 millibarns for 16.9 MeV ^{7}Be-ions. The equivalent center-of-momentum energy for the ^{7}Be(*d*,*n*)^{8}B reaction is 3.8 MeV. It is my view that continued development and application of radioactive ion-beam techniques could bring the most exciting results in laboratory Nuclear Astrophysics in the next decade. For example the rate of the ^{13}N(*p*,γ)^{14}O reaction, which will be studied as ^{1}H(^{13}N,γ)^{14}O, is crucial to the operation of the so-called fast CN-cycle.

In any case it has been clear for some time that experimental results on Si-burning reactions must be systematized and supplemented by comprehensive theory. Fortunately theoretical average cross sections will suffice in many cases. This is because the stellar reaction rates integrate the cross sections over the Maxwell-Boltzmann distribution. For most Si-burning reactions resonances in the cross section are closely spaced and even overlapping and the integration covers a wide enough range of energies that the detailed structure in the cross sections is automatically averaged out. The statistical model of nuclear reactions developed by Hauser and Feshbach (45), which yields average cross sections, is ideal for the purpose. Accordingly Holmes, Woosley, Fowler and Zimmerman (46) undertook the task of developing a *global, parametrized* Hauser-Feshbach theory and computer program for use in Nuclear Astrophysics. Reference (23) is an extension of this work. The free parameters are the

STATISTICAL MODEL CALCULATIONS VS MEASUREMENTS (I)
RATIO OF REACTION RATE (GROUND STATE OF TARGET) FROM WOOSLEY, FOWLER, HOLMES & ZIMMERMAN (AD & ND TABLES 22, 371, 1978) TO REACTION RATES FROM EXPERIMENTAL YIELD MEASUREMENTS (1970-1982) AT BOMBAY, CALTECH, COLORADO, KENTUCKY, MELBOURNE & TORONTO

REACTION	$T_9 = T/10^9$ K				
	1	2	3	4	5
$^{23}Na(p,n)^{23}Mg$	1.4	1.2	1.1	1.1	1.0
$^{25}Mg(p,\gamma)^{26}Al$	1.2	1.1	1.0	0.9	0.8
$^{25}Mg(p,n)^{25}Al$	1.1	1.0	0.9	0.8	0.8
$^{27}Al(p,\gamma)^{28}Si$	3.7	2.1	1.5	1.3	1.1
$^{27}Al(p,n)^{27}Si$	1.8	1.4	1.3	1.3	1.2
	0.9	0.9	0.9	1.0	1.0
$^{28}Si(p,\gamma)^{29}P$		1.2	1.3	1.2	0.9
$^{29}Si(p,\gamma)^{30}P$		1.0	1.6	1.6	1.5
$^{39}K(p,\gamma)^{40}Ca$	15	4.5	3.0	2.6	2.5
$^{41}K(p,\gamma)^{42}Ca$	0.5	0.5	0.5	0.4	0.4
$^{41}K(p,n)^{41}Ca$	0.8	1.0	1.1	1.2	1.3
$^{40}Ca(p,\gamma)^{41}Sc$				0.1	0.2
$^{42}Ca(p,\gamma)^{43}Sc$	1.3	1.4	1.4	1.4	1.3
	0.0	1.1	1.3	1.4	1.4

Table 3.

radius, depth and compensating reflection factor of the black-body, square-well equivalent of the Woods-Saxon potential characteristic of the interaction between n, p and *a* with nuclei having $Z \geq 8$. Two free parameters must also be incorporated to adjust the intensity of electric and magnetic dipole transitions for gamma radiation. Weak interaction rates must also be specified and ways and means for doing this will be discussed later in Section VII.

The parameters originally chosen for *n, p* and α-reactions were taken from earlier work of Michaud and Fowler (47) who depended heavily on studies by Vogt (see 47). These parameters and those chosen for electromagnetic and weak interactions have survived comparison of the theory with a plethora of laboratory measurements. More sophisticated programs have been developed which use experimental neutron strength functions instead of that from the equivalent square well or which use realistic Woods-Saxon potentials for all interactions as done by Mann (48). In addition marked improvement in the

correspondence between theory and experiment is found when width-fluctuation corrections are made as described in Zyskind et al. (49).

It is well known that the free parameters can always be adjusted to lit the cross sections and reaction rates of any one particular nuclear reaction. This is not done in a *global* program. The parameters are in principle determined by the best least squares fit to all reactions for which experimental results are available. For example see the figure, p. 307, in reference (46). It is on this basis that some confidence can be had in predictions in those cases where experimental results are unavailable.

The original program, references (46) and (23), has produced reaction rates either in numerical or analytical form as a function of temperature. Ready comparison with integrations of laboratory cross sections for target ground states are possible. Using the same *global* parameters which apply to reactions involving the ground states of stable nuclei the theoretical program calculates rates for the ground states of radioactive nuclei and for the excited states of both stable and radioactive nuclei. Summing over the statistically weighted contributions of the ground and known excited states or theoretical level density functions yields the stellar reaction rate for the equilibrated statistical population of the nuclear states. After summing, division by the partition function of the target nucleus is necessary. Analytical parametrized expressions for the partition functions of nuclei with $8 \leqslant Z \leqslant 36$ are given in Table IIA of reference (23) as a function of temperature over the range $0 \leqslant T \leqslant 10^{10}\mathrm{K}$.

Sargood (50) has compared experimental results from a number of laboratories for protons and alpha particles reacting with 80 target nuclei which are, of course, in their ground states with the theoretical predictions of reference (23). Ratios of statistical model calculations to laboratory measurements for 12 cases are shown in Table 3 for temperatures in the range from 1 to 5×10^{9}K. The double entry for $^{27}\mathrm{Al}(p,n)^{27}\mathrm{Si}$ signifies ratios of theory to measurements made in two different laboratories. It is fair to note that the theoretical calculations match the experimental results within 50% with a few marked exceptions. In American vernacular "You win some and you lose some". For the rather light targets in Table 3, especially at low temperature, the global mean rates can be in error whenever more and stronger resonances or fewer and weaker resonances than expected on average occur in the excitation curve of the reaction at low energies.

Sargood (50) has also compared the ratio of the stellar rate of a reaction with target nuclei in a thermal distribution of ground and excited states with the rate for all target nuclei in their ground state. The latter is of course determined from laboratory measurements. A number of cases are tabulated for $T = \mathbf{5\times10^{9}}$K in Table 4. In many cases, notably for reactions producing gamma rays, the ratio of stellar to laboratory rates is close to unity. In other cases the ratios can be high by several orders of magnitude. This can occur for a number of reasons. It frequently occurs when the ground state can interact only through partial waves of high angular momentum resulting in small penetration factors and thus small cross sections and rates. This makes clear a basic assumption in the prediction of stellar rates: a statistical theory which does well

STELLAR/LABORATORY REACTION RATES

$\langle \sigma v \rangle^* / \langle \sigma v \rangle^\circ$

TEMPERATURE = 5×10^9 K

D. G. Sargood, Australian Journal of Physics (1983)

Woosley, Fowler, Holmes & Zimmerman, At. Data & Nucl. Data Tables, 22, 371 (1978)

Target nucleus	Reaction								
	(n,γ)	(n,p)	(n,α)	(p,γ)	(p,n)	(p,α)	(α,γ)	(α,n)	(α,p)
^{20}Ne	0.959	12.2	4.98	0.954	34.1	6.86	0.907	4.90	1.29
^{21}Ne	0.808	6.15	1.13	0.818	1.78	1.95	0.943	0.985	1.37
^{22}Ne	0.917	159	22.1	0.895	5.11	2.72	0.968	0.996	2.46
^{23}Na	0.897	4.95	9.70	0.890	2.17	0.944	0.826	1.30	0.918
^{24}Mg	0.939	20.4	7.30	0.924	120	15.0	0.835	4.70	1.04
^{25}Mg	0.905	5.05	3.18	0.862	3.48	5.02	0.958	0.973	1.10
^{26}Mg	0.968	71.4	53.8	0.958	8.05	4.91	0.974	1.00	1.41
^{27}Al	0.934	4.12	10.9	0.913	3.22	1.14	0.905	1.13	0.972
^{28}Si	0.976	6.51	7.26	0.950	140	23.5	0.933	3.55	1.02
^{29}Si	0.943	8.67	3.34	0.907	3.18	50.1	0.927	0.964	1.18
^{30}Si	0.989	09.4	28.6	0.982	2.99	6.63	0.973	1.01	1.09
^{31}P	0.972	2.63	18.4	0.901,	3.77	1.11	0.969	1.70	0.978
^{32}S	0.988	2.33	1.57	0.980	90.1	7.35	0.975	3.79	1.00
^{33}S	0.943	1.46	1.06	0.920	4.73	3.24	0.916	0.995	1.01
^{34}S	1.00	25.0	13.1	0.979	8.02	2.02	0.964	1.05	1.02
^{36}S	0.996	428	95.9	1.00	1.00	1.02	0.995	1.00	1.68
^{35}Cl	0.972	1.19	3.06	0.948	4.48	1.05	0.945	1.23	0.992
^{37}Cl	0.994	26.0	13.7	0.987	1.00	1.00	0.985	1.00	0.995

Table 4

predicting ground state results is assumed to do equally well in predicting excited state results. This assumption is frequently not valid. Bahcall and Fowler (51) have shown that in a few cases laboratory measurements on inelastic scattering involving excited states can be used indirectly to determine reaction cross sections for those states.

Ward and Fowler (52) have investigated in detail the circumstances under which long lived isomeric states do not come into equilibrium with ground states. When this occurs it is necessary to incorporate into network calculations the stellar rates for both the isomeric and ground state. An example of great interest is the nucleus ^{26}Al. The ground state has spin and parity, $J^\pi = 5^+$ and isospin, $T = 0$, and has a mean lifetime for positron emission to ^{26}Mg of 10^6 years. The isomeric state at 0.228 MeV has $J^\pi = 0$, $T = 1$ and mean lifetime 9.2 seconds. Ward and Fowler (52) show that the isomeric state effectively does not come into equilibrium with the ground state for $T < 4 \times 10^8$K. At these low temperatures both the isomeric state and the ground state of ^{28}Al must be included in the network of Figure 7.

VII. *Astrophysical Weak-Interaction Rates*

Weak nuclear interactions play an important role in astrophysical processes in conjunction with the strong nuclear interactions as indicated in Figure 7. Only through the weak interaction can the overall proton number and neutron number of nuclear matter change during stellar evolution, collapse, and explo-

sion. The formation of a neutron star requires that protons in ordinary stellar matter capture electrons. Gravitational collapse of a Type II supernova core is retarded as long as electrons remain to exert outward pressure.

Many years of theoretical and experimental work on weak-interaction rates in the Kellogg Laboratory and elsewhere have culminated in the calculation and tabulation by Fuller, Fowler and Newman (53) of electron and positron emission rates and continuum electron and positron capture rates, as well as the associated neutrino energy loss rates for free nucleons and 226 nuclei with mass numbers between A = 21 and 60. Extension to higher and lower values for A is now underway.

These calculations depended heavily on experimental determinations in Kellogg by Wilson, Kavanagh and Mann (54) of Gamow-Teller elements for 87 discrete transitions in intermediate-mass nuclei. The majority of the experimental matrix elements for both Fermi and Gamow-Teller discrete transitions as well as nuclear level data were taken from the exhaustive tabulation of Lederer and Shirley (55). Unmeasured matrix elements for allowed transitions were assigned a mean value as described in the second of references (53). These mean values were $| M_F |^2 = .062$ and $| M_{GT} |^2 = .039$ corresponding to $\log ft = 5$, where f is the phase space factor and t is the half-life for the transition. Nuclear physicists traditionally think in terms of log ft-values in connection with weak interaction rates.

Simple shell model arguments were employed to estimate Gamow-Teller sum rules and collective state resonance excitation energies. These estimates have been shown to be high by -50% fair approximations for y-nuclei and P-nuclei by recent high resolution measurements on p,n-reactions and ^3T,^{3}He-reactions by Goodman *et al.* and Ajzenberg-Selove *et al.* respectively (56). Here $T^<$, with $T = | N-Z |$ represents, for example, ^{56}Fe with $T = 2$ in $^{56}\mathrm{Fe}(e^-,\nu)^{56}\mathrm{Mn}$ or $^{56}\mathrm{Fe}(n,p)^{56}\mathrm{Mn}$. Similarly $T^>$ designates ^{56}Mn with $T = 3$. The work described in references (53) emphasizes the great need for additional results for P-nuclei using the n,p-reaction as well as the ^3T,^{3}He-reaction from which matrix elements for electron capture can be obtained.

Moment method shell model calculations of Gamow-Teller strength functions have been performed by S.D. Bloom and G. M. Fuller (57) with the Lawrence Livermore National Laboratory's vector shell model code for the ground states and first excited states of ^{56}Fe, ^{60}Fe, and ^{64}Fe. These detailed calculations confirm the general trends in Garnow-Teller strength distributions used in the approximations of references (53).

The discrete state contribution to the rates, dominated by experimental information and the Fermi transitions, determines the weak nuclear rates in the regime of temperature and densities characteristic of the quasistatic phases of presupernova stellar evolution. At the higher temperatures and densities characteristic of the supernova collapse phase, which is of such great current interest as discussed in detail in Brown, Bethe and Baym (58), the electron-capture rates are dominated by the Gamow-Teller collective resonance contribution.

The detailed nature and the difficulty of the theoretical aspects of the

combined atomic, nuclear, plasma, and hydrodynamic physics problems in Type II supernova implosion and explosion were brought home to us by Hans Bethe during his stay in our laboratory as a Caltech Fairchild Scholar early in 1982. His visit plus long-distance interaction with his collaborators resulted in the preparation of two seminal papers, Bethe, Yahil, and Brown (59) and Bethe, Brown, Cooperstein, and Wilson (60).

Current ideas on the nuclear equation of state predict that early in the collapse of the iron core of a massive star the nuclei present will become so neutron rich that allowed electron capture on protons in the nuclei is blocked. Allowed electron capture, for which $\Delta l = 0$, is not permitted when neutrons have filled the subshells having orbital angular momentum, *1,* equal to that of the subshells occupied by the protons.

This neutron shell blocking phenomenon, and several unblocking mechanisms-operative at high temperature and density, including forbidden electron capture, have been studied in terms of the simple shell model by Fuller (61). Though the unblocking mechanisms are sensitive to details of the equation of state, typical conditions result in a considerable reduction of the electron capture rates on heavy nuclei leading to significant dependence on electron capture by the small number of free protons and a decrease in the overall neutronization rate.

The results of one-zone collapse calculations which have been made by Fuller (61) suggest that the effect of neutron shell blocking is to produce a larger core lepton fraction (leptons per baryon) at neutrino trapping. In keeping with the Chandrasekhar relation that core mass is proportional to the square of the lepton fraction this leads to a larger *final-core* mass and hence a stronger post-bounce shock. On the other hand the incorporation of the new electron capture rates during precollapse Si-burning reduces the lepton fraction and leads to a smaller initial-core mass and thus to a smaller amount of material (initial-core mass minus *final-core* mass) in which the post-bounce shock can be dissipated. The dissipation of the shock is thus reduced. This is discussed in detail in reference (39).

Recent work on the weak-interaction has concentrated on making the previously calculated reaction rates as efficient as possible for users of the published tables and the computer tapes which are made available on request. The stellar weak interaction rates of nuclei arc in general very sensitive functions of temperature and density. Their temperature dependence arises from thermal excitation of parent excited states and from the lepton distribution functions in the integrands of the decay and continuum capture phase space factors.

For electron and positron emission, most of the temperature dependence is due to thermal population of parent excited states at all but the lowest temperatures and highest densities. In general, only a few transitions will contribute to these decay rates and hence the variation of the rates with temperature is usually not so large that rates cannot be accurately interpolated in temperature and density with the standard grids provided in references (53). The density dependence of these decay rates is minimal. In the case of electron emission, however, there may be considerable density dependence due to Pauli blocking

for electrons where the density is high and the temperature is low. This does not present much of a problem for practical interpolation since the electron-emission rate is usually very small under these conditions.

The temperature and density dependence of continuum electron and positron capture is much more serious problem. In addition to temperature sensitivity introduced through thermal population of parent excited states, there are considerable effects from the lepton distribution functions in the integrands of the continuum-capture phase-space factors. This sensitivity of the capture rates means that interpolation in temperature and density on the standard grid to obtain a rate can be difficult, requiring a high-order interpolation routine and a relatively large amount of computer time for an accurate value. This is especially true for electron capture processes with threshold above zero energy.

We have found that the interpolation problem can be greatly eased by defining a simple continuum-capture phase-space integral, based on the parent-ground-state to daughter-ground-state transition Q-value, and then dividing this by the tabulated rates (53) at each temperature and density grid point to obtain a table of effective ft-values; these turn out to be much less dependent on temperature and density. This procedure requires a formulation of the capture phase-space factors which is simple enough to use many times in the inner loop of stellar evolution nucleosynthesis computer programs. Such a formulation in terms of standard Fermi integrals has been found, along with approximations for the requisite Fermi integrals. When the chemical potential (Fermi energy) which appears in the Fermi integrals goes through zero these approximations have continuous values and continuous derivatives.

We have recently found expressions for the reverse reactions to e^-, e^+-capture, (i.e., $\nu,\bar{\nu}$-capture) and for $\nu,\bar{\nu}$blocking of the direct reactions when $\nu,\bar{\nu}$-states are partially or completely filled. These reverse reactions and the blocking are important during supernova core collapse when neutrinos and antineutrinos eventually become trapped, leading to equilibrium between the two directions of capture. General analytic expressions have been derived and approximated with computer-usable equations. All of these new results described in the previous paragraphs will be published in Fuller, Fowler, and Newman (62) and new tapes including $\nu,\bar{\nu}$-capture will be made available to users on request.

VIII. *Calculated abundances for $A \lesssim 60$ with Brief Comments on Explosive Nucleosynthesis*

Armed with the available strong and weak nuclear reaction rates which apply to the advanced stages of stellar evolution, theoretical astrophysicists have attempted to derive the elemental and isotopic abundances produced in quasi-static presupernova nucleosynthesis and in explosive nucleosynthesis occurring during supernova outbursts.

The various stages of preexplosive nucleosynthesis have been discussed in Sections IV, V and VI and it is fair to say that there is reasonably general agreement on nucleosynthesis during these stages. On the other hand explosive nucleosynthesis is still an unsettled matter, subject to intensive study at the

present time as reviewed for example, in Woosley, Axelrod, and Weaver (63).

The abundance produced in explosive nucleosynthesis must of necessity depend on the detailed nature of supernova explosions. Ideas concerning the nature of Type I and Type II supernova explosions were published many years ago by Hoyle and Fowler (64) and Fowler and Hoyle (65). It was suggested that Type I supernovae of small mass were precipitated by the onset of explosive carbon burning under conditions of electron-degeneracy where pressure is approximately independent of temperature. Carbon burning raises the temperature to the point where the electrons are no longer degenerate and explosive disruption of the star results. For Type II supernovae of larger mass it was suggested that Si-burning produced iron-group nuclei which have the maximum binding energies of all nuclei so that nuclear energy is no longer available. Subsequent photodisintegration and electron-capture in the stellar core leads to core implosion and ignition of explosive nucleosynthesis in the infalling inner mantle which still contains nuclear fuel. These ideas have "survived" but, to say the least, with considerable modification over the years as indicated in the excellent review by Wheeler (66). Modern views on Type II supernovae are given in references (38), (39), (58), (59), and (60) and on Type I supernovae in Nomoto (67).

We can return to the nuclear abundance problem by reference to Figure 16 taken from reference (38), which shows the distribution of the final abundances by mass fraction in the supernova ejecta of a 25 $M_\odot$ Population I star. The presupernova distribution is that shown in Figure B(a). The modification in the abundances for the mass zones interior to $2.4\ M_\odot$ under the very apparent. The mass exterior to $2.4\ M_\odot$ ejected with little or no modification in nuclear abundances. The supernova explosion was simulated by arbitrarily assuming that the order of 10^{51} ergs was delivered to the ejected material by the shock generated in the bounce or rebound of the collapsing and hardening core.

Integration over the mass zones of Figure 16 for 1 over those of Figure 8(a) for enabled Woosley and Weaver (38) to calculate the isotopic abundances ejected into the interstellar medium by their ding Population I simulated supernova. The results relative to solar abundances (the reader should refer to the last paragraph of Section I) are shown in Figure 17 taken from reference (38). The relative ratios are normalized to unity for ^{16}O for which the overproduction ratio was 14, that is, for each gram of ^{16}O originally in the star, 14 grams were ejected. This overproduction in a single supernova can be expected to have produced the heavy element abundances in the interstellar medium just prior to formation of the solar system given the fact that supernovae occur approximately every one hundred years in the Galaxy. The ultimate theoretical calculations will yield a constant overproduction factor of the order of 10.

The results shown in Figure 17 are disappointing if one expects the ejecta of and Population I supernovae to match solar system abundances with a relatively constant overproduction factor. The dip in abundances from sulfur to chromium is readily apparent. Woosley and Weaver (38) point out that calculations must be made for other stellar masses and properly integrated over the

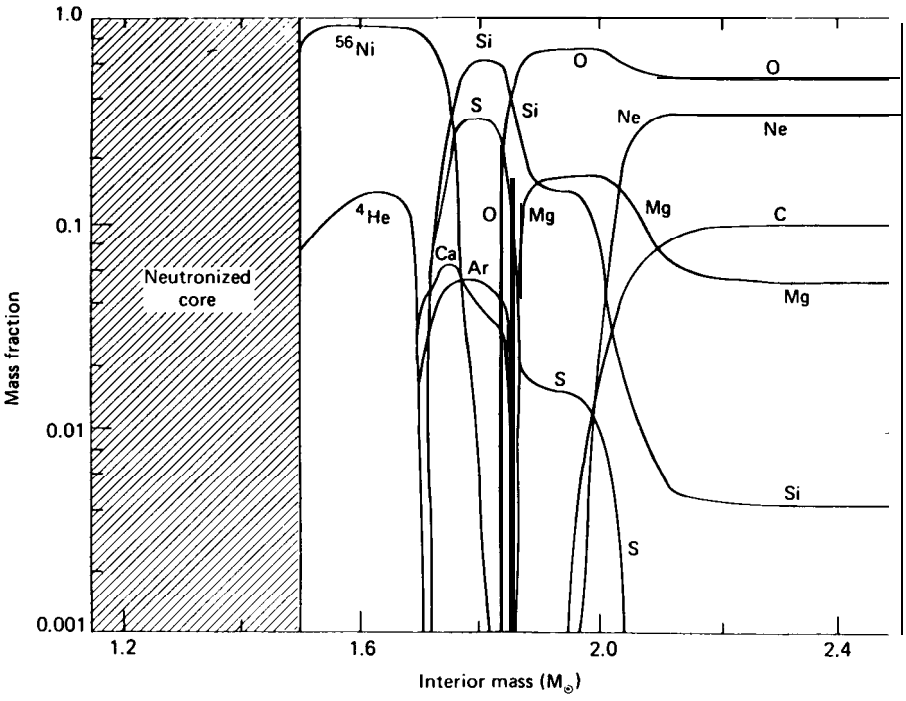


Figure 16. Final abundances by mass fraction versus increasing interior mass in solar masses, $M_\odot$, in Type II supernova ejecta from a Population I star with total mass equal to $25M_\odot$ from Woosley and Weaver (38).

mass distribution for stellar formation which varies roughly inversely proportional to mass. Woosley, Axelrod and Weaver (63) discuss their expectations of the abundances produced in stellar explosions for stars in the mass range $10M_\odot$ to $10^6M_\odot$. They show that a $200M_\odot$ Population III star produces abundant quantities of sulfur, argon, and calcium which possibly compensate for the dip in figure 17. Population III stars are massive stars in the range $100M_\odot<M<300M_\odot$ which are thought to have formed from hydrogen and helium early in the history of the Galaxy and evolved very rapidly. Since their heavy element abundance was zero they have no counterparts in presently forming Population I stars as well as no counterparts among old, low mass Population II stars.

Other authors have suggested a number of solutions to the problem depicted in Figure 17. Nomoto, Thielemann and Wheeler (68) have calculated the abundances produced in carbon deflagration models of Type I supernovae. By adding equal contributions from Type I and Type II supernovae they obtain Figure 18 which can be considered somewhat more satisfactory than Figure 17. On the other hand Arnett and Thielemann (69) have recalculated quasistatic presupernova nucleosynthesis for $M\approx20M_\odot$ using a value for the $^{12}C\ (\alpha,\gamma)^{16}O$ rate equal to three times that given in references listed under (21). This would seem to be justified by the recent analysis of $^{12}C\ (\alpha,\gamma)\ ^{16}O$ data in reference (34) as discussed in Section V. They then assume that explosive nucleosynthesis will

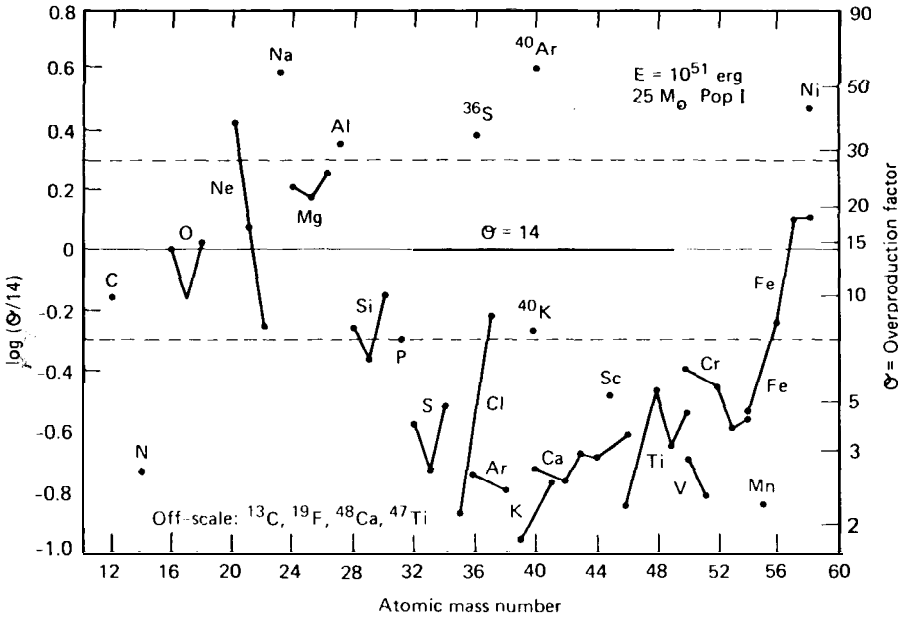


Figure 17. Overabundance (ϑ) relative to 14 times solar abundances versus atomic mass number for nucleosynthesis resulting from a Type II, Population I supernova with total mass equal to 25 $M_{\odot}$ from Woosley and Weaver (38).

not substantially modify their quasistatic abundances and obtain the results shown in Figure 19. The average overproduction ratio is roughly 14 and deviations are in general within a factor of two of this value. However, their assumption of minor modification during explosion and ejection is questionable.

I feel that the results discussed in this section and those obtained by numerous other authors show promise of an eventual satisfactory answer to the question where and how did the elements from carbon to nickel originate. We shall see!

IX. *Isotopic Anomalies in Meteorites and Observational Evidence for Ongoing Nucleosynthesis*

Almost a decade ago it became clear that nucleosynthesis occurred in the Galaxy up to the time of formation of the solar system or at least up until several million years before the formation. For slightly over one year it has been clear that nucleosynthesis continues up to the present time or at least within several million years of the present. The decay of radioactive ^{26}Al($\tilde{\tau}$= 1.04 x 10^6 years) is the key to these statements which bring great satisfaction to most experimentalists, theorists, and observers in Nuclear Astrophysics. For the record it must be admitted that the word "clear" is subject to certain reservations in the minds of some investigators but as a believer, "clear" is clear to me.

Isotopic anomalies in meteorites produced by the decay of shortlived radioactive nuclei were first demonstrated in 1960 by Reynolds (70) who found large enrichments of ^{129}Xe in the Richardson meteorite. Jeffery and Reynolds (71)

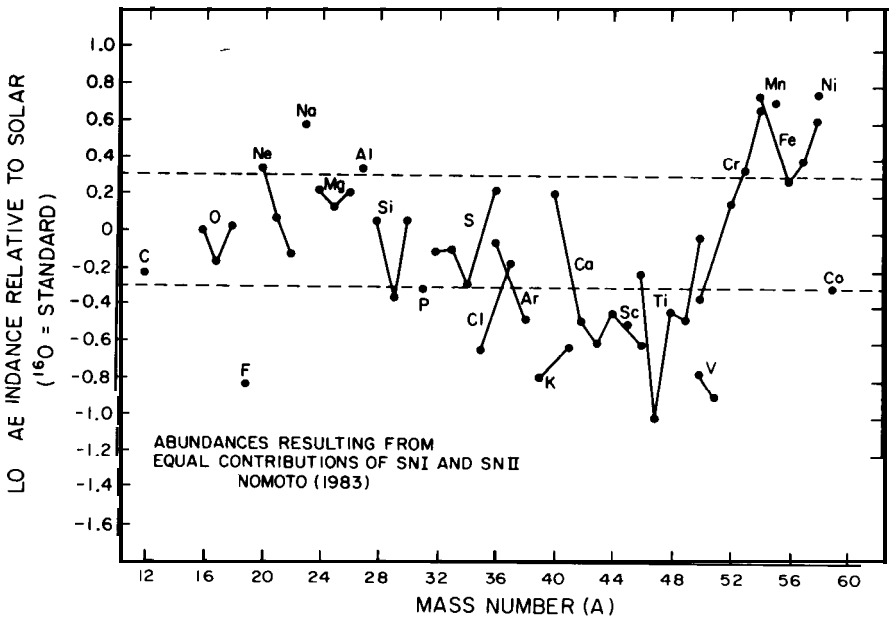


Figure 18. Abundances relative to solar with the abundance of ^{16}O taken as standard produced by equal contributions from typical Type I and Type II supernovae from Nomoto, Thielemann and Wheeler (68).

demonstrated in 1961 that the excess ^{129}Xe correlated with ^{127}I in the meteorite and thus showed that the ^{129}Xe resulted from the decay *in situ* of ^{129}I(; = 23 x 10^6 years). Quantitative results indicated that ^{129}I /^{127}I= 10^{-4} at the time of meteorite formation. On the assumption that ^{129}I and ^{127}I are produced in roughly equal abundances in nucleosynthesis (most probably in the r-process) over a period of ~ 10^{10} years in the Galaxy prior to formation of the solar system and taking into account that only the ^{129}I produced over a period of the order of its lifetime survives, Wasserburg, Fowler, and Hoyle (72) suggested that a period of free decay of the order of 10^8 years or more occurred between the last nucleosynthetic event which produced ^{129}I and its incorporation in meteorites in the solar system. There remains evidence for such a period in some cases, notably ^{244}Pu, but probably not in the history of the nucleosynthetic events which produced ^{129}I and other "short"-lived radioactive nuclei such as ^{26}Al and ^{107}Pd(i= 9.4 x 10^6 years).

The substantiated meteoritic anomalies in ^{26}Mg from ^{26}Al, in ^{107}Ag from ^{107}Pd, in ^{129}Xe from ^{129}I and in the heavy isotopes of Xe from the fission of ^{244}Pu($\bar{\tau}$ = 117 x 10^6 years; fission tracks also observed) as well as searches in the future for anomalies in ^{41}K from ^{41}Ca(t = 0.14 x 10^6 years), in ^{60}Ni from ^{60}Fe($\bar{\tau}$ = 0.43 x 10^6 years), in ^{53}Cr from ^{53}Mn($\bar{\tau}$ = 5.3 x 10^6 years), and in ^{142}Nd from ^{146}Sm(i = 149 x 10^6 years; a-decay) are discussed exhaustively by my colleagues Wasserburg and Papanastassiou (73). They espouse *in situ* decay for the observations to date but my former student D. D. Clayton (74) argues that the anomalies occur in interstellar grains preserved in the meteorites and originally produced by condensation in the expanding and cooling envelopes of

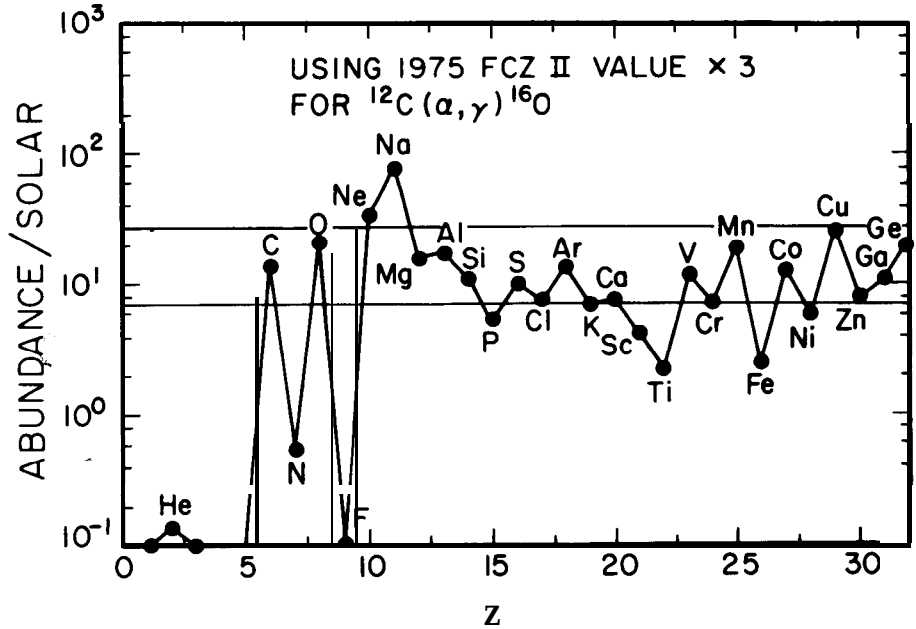


Figure 19. Overabundance yields relative to solar versus atomic number, Z, resulting from the explosion of a Type II supernova with mass approximately equal to $20M_{\odot}$ from Arnett and Thielemann (69). The horizontal lines are a factor two higher and lower than the average overabundance equal to 14. It is assumed that the pre-supernova abundances were not modified during the supernova explosion. The reaction rate for $^{12}C(\alpha,\gamma)^{16}O$ of Fowler, Caughlan and Zimmerman (21) was multiplied by a factor of 3 in accordance with the theoretical analysis by Langanke and Koonin (34).

supernovae and novae. Wasserburg and Papanastassiou (73) write on p. 90 "There is, as yet, no compelling evidence for the presence of preserved presolar grains in the solar system. All of the samples so far investigated appear to have melted or condensed from a gas, and to have chemically reacted to form new phases." With mixed emotions I accept this.

Before turning to some elaboration of the $^{26}A1/^{26}Mg$ case it is appropriate to return to a discussion of the free decay interval mentioned above. It is the lack of a detectable anomalies in ^{235}U from the decay of $^{247}Cm(\tau = 23 \times 10_6$ years) in meteorites as shown by Chen and Wasserburg (75) coupled with the demonstrated occurrence of heavy Xe anomalies from the fission of $^{244}Pu(\bar{\tau} = 117 \times 10^6$ years) as discussed for example by 'Burnett, Stapanian and Jones (76) which demands a free decay interval of the order of several times 10^8 years. This interval is measured from the "last" r-process nucleosynthesis event (supernova?) which produced the actinides, Th, U, Pu, Cm, and beyond, up to the "last" nucleosynthesis events (novae?, supernovae with short-run r-processes?) which produced the short-lived nuclei $^{26}A1$, ^{107}Pd, and ^{129}I before the formation of the solar system. The fact that the anomalies produced by these short-lived nuclei relative to normal abundances all are of the order of 10^{-4} despite a wide range in their mean lifetimes (1.04 to 23 x 10^6 years) indicates that this anomaly range must be the result of inhomogeneous mixing of exotic materials with much larger quantities of normal solar system materials over a short time

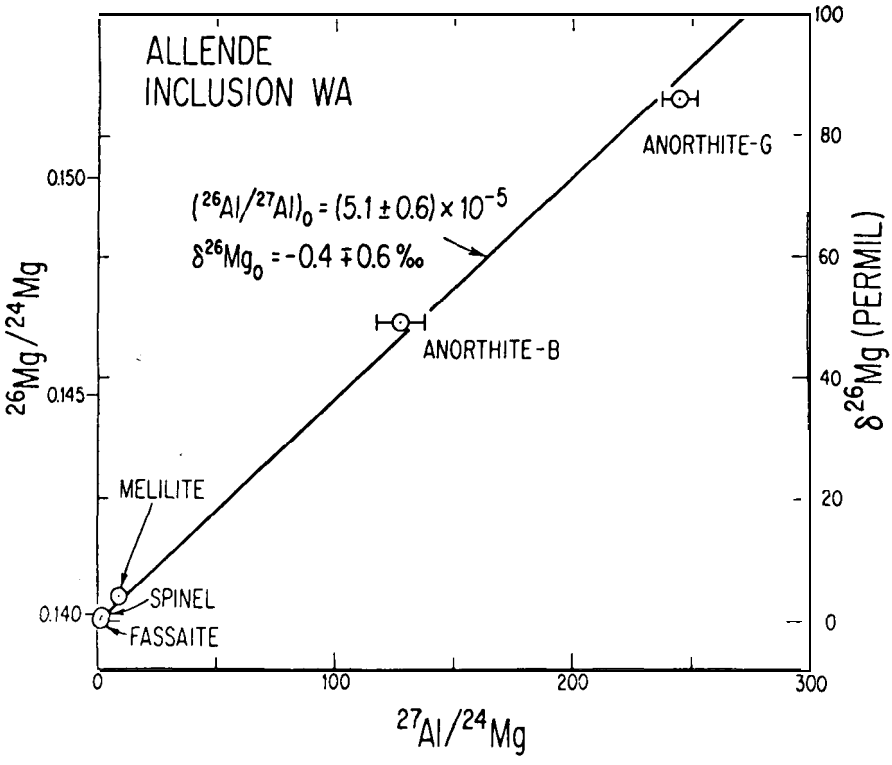


Figure 20. Evidence for the *in situ* decay of ^{26}A1 in various minerals in inclusion WA of the Allende meteorite from Lee, Papanastassiou and Wasserburg (77). The linear relation between ^{26}Mg/^{24}Mg and ^{27}Al/^{24}Mg implies that ^{26}Al/^{27}Al = (5.1 ± 0.6) × 10^{-5} at the time of information of the inclusion with ^{26}Al considered to react chemically in the same manner as ^{27}Al.

rather than the result of free decay. The challenges presented by this conclusion are manifold. Figure 14 of reference (73) shows the time scale for the formation of dust, rain, and hailstones in the early solar system and for the aggregation into chunks and eventually the terrestrial planets. The solar nebula was almost but not completely mixed when it collapsed to form the solar system. From ^{26}Al it becomes clear that the mixing time down to an inhomogenity of only one part in 10^3 (see what follows) was the order of 10^6 years.

Evidence that ^{26}A1 was alive in interstellar material in the solar nebula which condensed and aggregated to form the parent body (planet in the asteroid belt?) of the Allende meteorite is shown in Figure 20 taken with some modification from Lee, Papanastassiou, and Wasserburg (77). The Allende meteorite fell near Pueblito de Allende in Mexico on February 8, 1969 and is a carbonaceous chondrite, a type of meteorite thought to contain the most primitive material in the solar system *unaltered since its original solidification.*

Figure 20 depicts the results for ^{26}Mg/^{24}Mg versus ^{27}A1/^{24}Mg in different mineral phases (spinel etc.) from a Ca-Al-rich inclusion called WA obtained from a chondrule found in Allende. It will be clear that excess ^{26}Mg correlates linearly with the amount of ^{27}Al in the mineral phases. Since ^{26}Al is chemically identical with ^{27}Al, it can be inferred that phases rich in ^{27}Al were initially rich

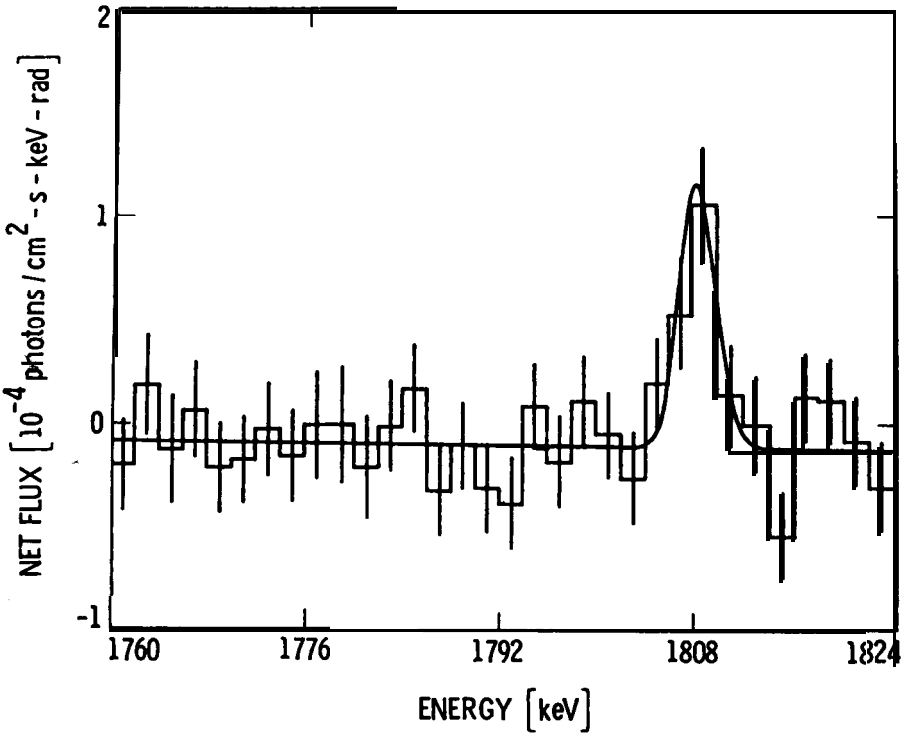


Figure 21. The High Energy Astrophysical Observatory (HEAO 3) data on gamma rays in the energy range 1760 to 1824 keV emitted from the Galactic equatorial plane from Mahoney et al. (78). The line at 1809 keV is attributed to the decay of radioactive ^{26}Al(t = 1.04 x 10^6 years) to the excited state of ^{26}Mg at this energy.

in ^{26}Al which subsequently decayed *in situ* to produce excess ^{26}Mg. ^{26}Al was alive with abundance 5 x 10^{-5} that of ^{27}Al in one part of the solar nebula when the WA inclusion aggregated during the earliest stages of the formation of the solar system. The unaltered inclusion survived for 4.5 billion years to tell its story. Other inclusions in Allende and other meteorites yield ^{26}Al/^{27}Al from zero up to ~10^{-3} with 10^{-4} a representative value. The reader is referred to reference (73) for the rich details of the story and the important and significance of non-accelerator-based contributions to Nuclear Astrophysics.

Evidence that ^{26}Al is alive in the interstellar medium today is shown in Figure 21 from Mahoney, Ling, Wheaton and Jacobson (78), my colleagues at Caltech's Jet Propulsion Laboratory (JPL). Figure 21 shows the gamma-ray spectrum observed in the range 1760 to 1824 keV by instruments aboard the High Energy Astronomical Observatory, HEAO 3, which searched for diffuse gamma-ray emission from the Galactic equatorial plane.

The discrete line at 1809 keV, detected with a significance of nearly five standard deviations, is without doubt due to the transition from the first excited state at 1809 keV in ^{26}M$_g$ to its ground state. Radioactive ^{26}Al decays by ^{26}Al$(e^+\nu)^{26}$Mg (γ) ^{26}Mg to this state and thence to the ground state of ^{26}Mg. This gamma-ray transition shows clearly that ^{26}A1 is alive in the interstellar medium in the Galactic equatorial plane today. Given the mean life-time (1.04

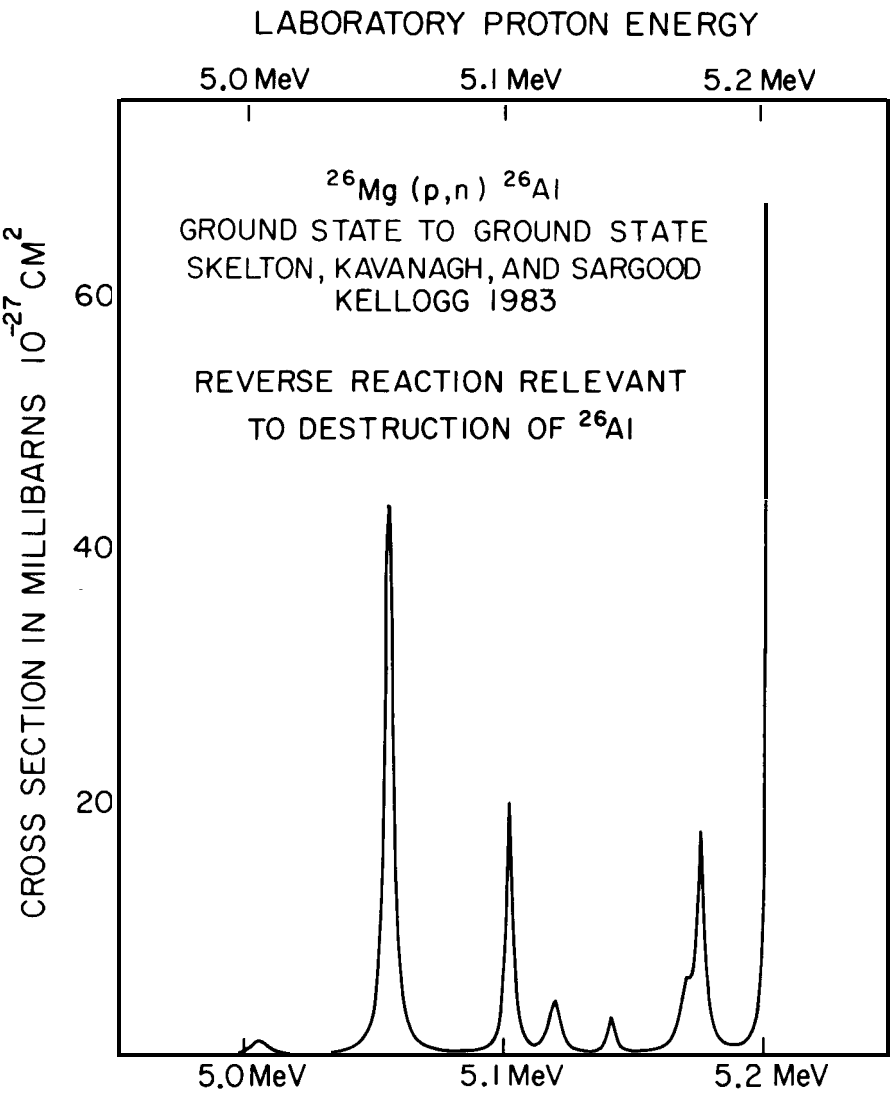


Figure 22. The cross section in millibarns versus laboratory proton energy for the ground-state to ground-state reaction ^{26}Mg(p, n)^{26}Al from Skelton, Kavanagh and Sargood (79).

x 10^6 years) of ^{26}Al, this shows that ^{26}Al has been produced no longer than several million years ago and is probably being produced continuously. It is no great extrapolation to argue that nucleosynthesis in general continues in the Galaxy at the present time. Quantitatively the observations indicate that ^{26}Al/^{27}Al ~ 10^{-5} in the interstellar medium. This value averages over the Galactic plane interior to the sun at the present time. This average value was probably much the same when the solar system formed but the variations in ^{26}A1/^{27}A1 for various meteoritic inclusions show that there were wide variations in the solar nebula about this value ranging from zero to 10^{-3}.

The question immediately arises, what is the site of the synthesis of the ^{26}Al? Since the preparation of reference (52) I have been convinced that ^{26}A1 could

not be by synthesized in supernovae at high temperatures where neutrons are copiously produced because of the expectation of a large cross section for $^{26}Al(n, p)^{26}Mg$. This expectation has been borne out by the measurements on the reverse reaction $^{26}Mg(p,n)^{26}A1$ in the Kellogg Laboratory by Skelton, Kavanagh and Sargood (79). Figure 22 is taken from Figure 1 a of these authors and shows the great beauty of high resolution measurements in experimental Nuclear Astrophysics. Until the "Al-targets just recently available can be bombarded with neutrons it is necessary to supplement the laboratory measurements on $^{26}Mg(p,n)^{26}A1$, perforce involving the ground state of ^{26}Mg, with theoretical calculations involving excited states, reference (23), in order to calculate the stellar rate for $^{26}Al(n,p)^{26}Mg$. There is little doubt that this rate is very large indeed.

In references (74) and (78) and in Arnould et al. (80) it is suggested that $^{26}A1$ is produced in novae. This is quite reasonable on the basis of nucleosynthesis in novae as discussed in Truran (81). In current models for novae hydrogen from a binary companion is accreted by a white dwarf until a thermal runaway involving the fast CN cycle occurs. Similarly a fast MgAl cycle may occur with production of $^{26}Al/^{27}Al \geqslant 1$ as shown in Figure 9 of reference (52). The recent experimental measurements cited in reference (52) substantiate this conclusion. Clayton (74) argues that the estimated 40 novae occurring annually in the Galactic disk can produce the observed $^{26}Al/^{27}Al$ ratio of the order of 10^{-5} on average. He assumes that each nova ejects $10^{-4}M_0$ of material containing a mass fraction of $^{26}A1$ equal to 3×10^{-4}.

Another possible source of $^{26}A1$ is spallation induced by irradiation of protoplanetary material by high energy protons from the young sun as it settled on the main sequence. This possibility was discussed very early by Fowler, Greenstein, and Hoyle (82) who also attempted to produce D, Li, Be, and B in this way, requiring such large primary proton and secondary neutron fluxes that many features of the abundance curve in the solar system would have been changed substantially. A more reasonable version of the scenario was presented by Lee (83) but without notable success. I find it difficult to believe that an early irradiation produced the anomalies in meteorites. The $^{26}A1$ in the interstellar medium today certainly cannot have been produced in this way.

Anomalies have been found in meteorites in the abundances compared with normal solar system material of the stable isotopes of may elements: 0, Ne, Mg, Ca, Ti, Kr, Sr, Xe, Ba, Nd, and Sm. The possibility that the oxygen anomalies are non-nuclear in origin has been raised by Thiemens and Heidenreich (84) but the anomalies in the remaining elements are generally attributed to nuclear processes.

One example is a neutron-capture/beta-decay $(n\beta)$ process studied by Sandler, Koonin, and Fowler (85). The seed nuclei consisted of all of the elements from Si to Cr with normal solar system abundances. When this process operates at neutron densities $\approx 10^{7}$ mole cm^{-3} and exposure times of $\approx 10^{3}$ s, small admixtures $(\leqslant 10^{-4})$ of the exotic material produced are sufficient to account for most of the Ca and Ti isotopic anomalies found in the Allende meteorite inclusion EK- 1-4-I by Niederer, Papanastassiou, and Wasserburg

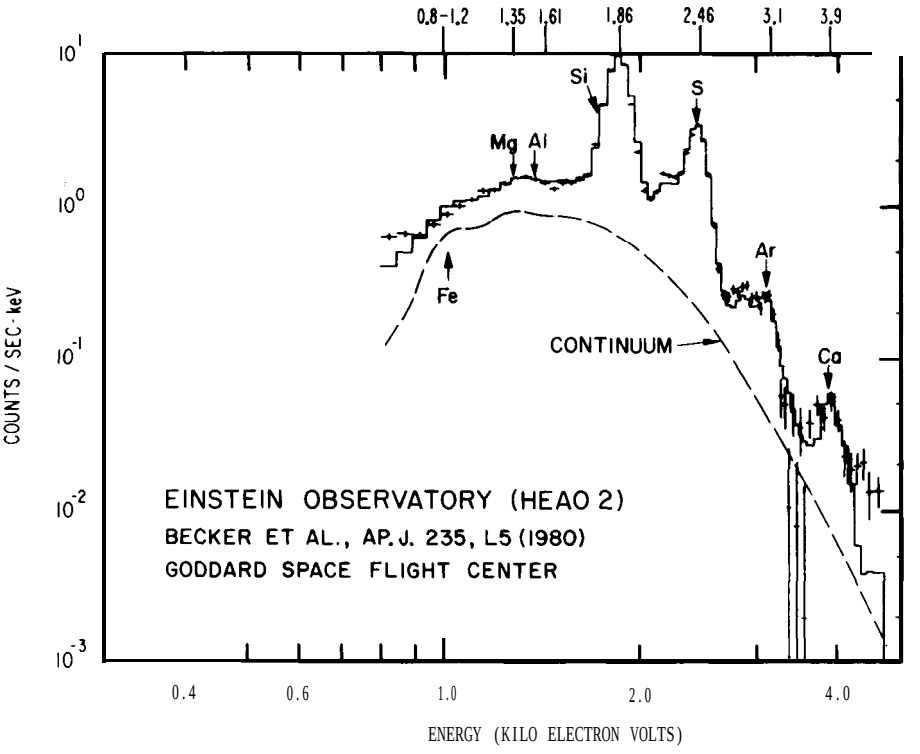


Figure 23. The Einstein Observatory (HEAO 2) data on the X-ray spectrum of Tycho Brahe's Supernova Remnant from Becker et al. (89).

(86). The anomalies in stable isotope abundances are of the same order as those for short lived radioactive nuclei and strongly support the view that the solar nebula was inhomogeneous and not completely mixed with regions containing exotic materials up to 10^{-4} or more of normal material.

Agreement for the ^{46}Ca and ^{49}Ti anomalies in EK-1-4-1 was obtained only by increasing the theoretical Hauser-Feshbach cross sections for $^{46}\mathrm{K}(n,\gamma)$ and $^{49}\mathrm{Ca}(n,\gamma)$ by a factor of 10 on the basis of probable thermal resonances just above threshold in the compound nuclei ^{47}K and ^{50}Ca respectively. In a CERN report which subsequently became available Huck et al. (87) reported an excited state in ^{50}C a just 0.16 MeV above the $^{49}\mathrm{Ca}(n,\gamma)$ threshold which can be produced by s-wave capture and fulfills the requirements of reference (85).

Reference (85) suggests that the $\approx 10^3$s exposure time scale is determined by the mean life-time of $^{13}\mathrm{N}(\bar{\tau} = 862$ s), produced through $^{12}\mathrm{C}(p,\gamma)^{13}$N by a jet of hydrogen suddenly introduced into the helium burning shell of a Red Giant star where a substantial amount of ^{12}C has been produced by the 3 $\alpha \rightarrow {}^{12}\mathrm{C}$ process. The beta decay $^{13}\mathrm{N}(e^{+}\nu)^{13}\mathrm{C}$ is followed by $^{13}C(\alpha,n)^{16}$O as the source of the neutrons. All of this is very interesting, if true. More to the point reference (85) predicts the anomalies to be expected in the isotopes of chromium. Attempts to measure these anomalies are underway now by Wasserburg and his colleagues. Again, we shall see!

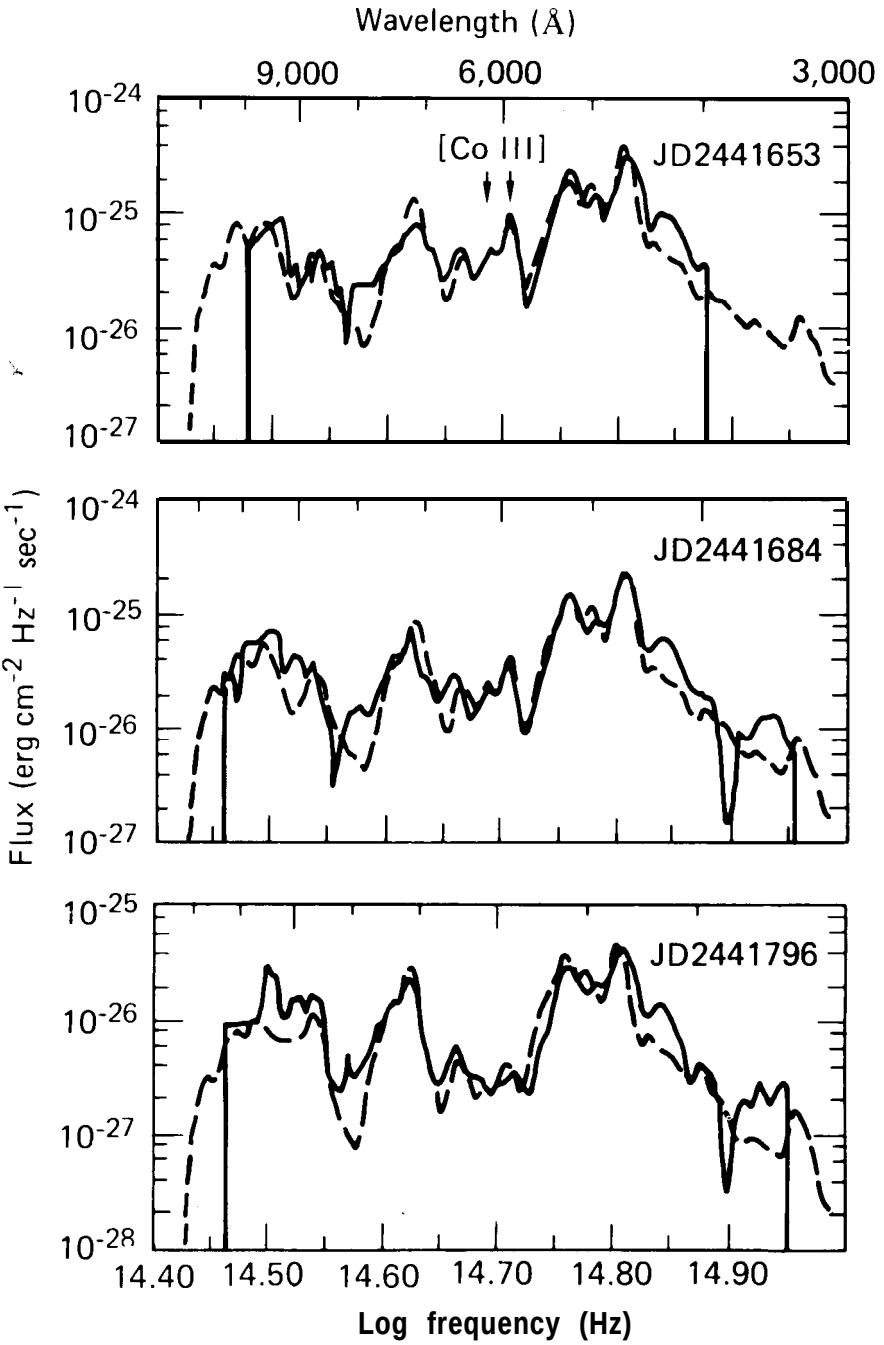


Figure 24. Analysis by Axelrod (91) yielding two emission lines from Co III in the observations on SN 1972e obtained by Kirshner and Oke (92). The observations were made 233, 264 and 376 days after JD2441420, assigned as the initial day of the supernova explosion. The mean lifetime of ^{56}Co is 114 days; the Co III lines appear to decay in keeping with their emission from radioactive ^{56}Co.

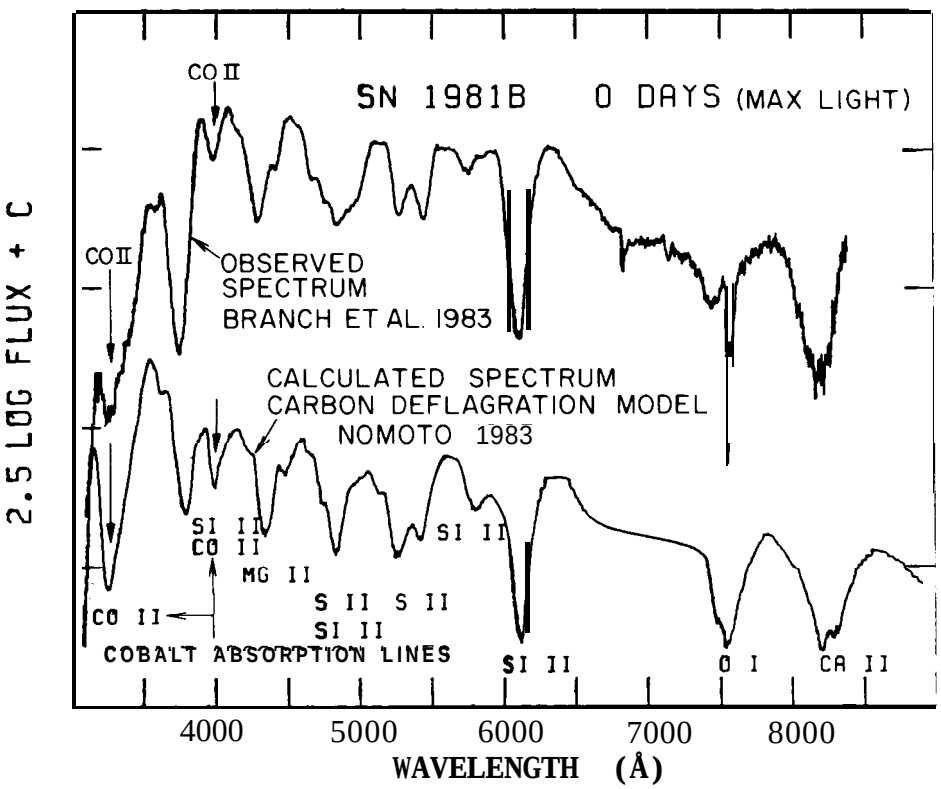


Figure 2.5. Top: Analysis by Branch et al. (93) of their absorption spectrum of SN 1981b at maximum light showing evidence for Co II absorption features. Bottom: Comparison with the calculated spectrum expected from the carbon deflagration model for Type I supernova according to the calculations of Nomoto (67).

X. Observational Evidence for Nucleosynthesis in Supernovae

Over the years there has been considerable controversy concerning elemental abundance observations in the optical wave-length range on Galactic supernovae remnants. To my mind the most convincing evidence for nucleosynthesis in supernovae has been provided by Chevalier and Kirshner (88) who obtained quantitative spectral information for several of the fast-moving knots in the supernova remnant Cassiopeia A (approximately dated 1659 but a supernova event was not observed at that time). The knots are considered to be material ejected from various layers of the original star in a highly asymmetric, non-spherical explosion. In one knot, labelled KB33, the following ratios *relative to solar, designated by brackets* were observed: [S/O] = 61, [Ar/O] = 55, [Ca/O] = 59. It is abundantly clear that oxygen burning to the silicon group elements in the layer in which KB33 originated has depleted oxygen and enhanced the silicon group elements. Other knots and other features designated as filaments show different abundance patterns, albeit, not so easily interpreted. The moral for supernova modelers is that spherically symmetric supernova explosions may be the easiest to calculate but are not to be taken as

realistic. Admittedly they have a good answer: it is expensive enough to compute spherically symmetric models. OK, OK!

Most striking of all has been the payoff from the NASA investment in the High Energy Astronomy Observatory (HEAO 2), now called the Einstein Observatory. From this satellite Becker et *al.* (89) observed the X-ray spectrum in the range. 1 to 4 keV of Tycho Brahe's supernova remnant (1572) shown in Figure 23. An X-ray spectrum is much simpler than an optical spectrum. For me it is wonderful that satellite observations show the K-level X-rays expected from Si, S, Ar, and Ca just where the Handbook of Chemistry and Physics says they ought to be! Such observations are not all that easy in a terrestrial laboratory. Shull (90) has used a single-velocity, non-ionization-equilibrium model of a supernova blast wave to calculate abundances in Tycho's remnant *relative to solar, designated by brackets* and finds: [Si] = 7.6, [S] = 6.5, [Ar] = 3.2 and [Ca] = 2.6. With considerably greater uncertainty he gives [Mg] = 2.0 and [Fe] = 2.1. He finds different enhancements in Kepler's remnant (1604) and in Cassiopeia A. One more lesson for the modelers: no two supernovae are alike. Nucleosynthesis in supernovae depends on their initial mass, rotation, mass loss during the Red Giant stage, the degree of symmetry during explosion, initial heavy element content, and probably other factors. These details aside it seems clear that supernovae produce enhancements in elemental abundances up to iron and probably beyond. Detection of the much rarer elements beyond iron will require more sensitive X-ray detectors operating at higher energies. The nuclear debris of supernovae eventually enriches the interstellar medium from which succeeding generations of stars are formed. It becomes increasingly clear that novae also enrich the interstellar medium. Sorting out these two contributions poses interesting problems in ongoing research in all aspects of Nuclear Astrophysics.

Explosive silicon burning in the shell just outside a collapsing supernova core primarily produces ^{56}Ni as shown in Figure 16. It is generally believed that the initial energy source for the light curves of Type I supernovae is electron capture by ^{56}Ni ($\bar{\tau}$ = 8.80 days) to the excited state of ^{56}Co at 1.720 MeV with subsequent gamma ray cascades to the ground state. These gamma rays are absorbed and provide energy to the ejected envelope. The subsequent source of energy is the electron capture and positron emission by ^{56}Co ($\bar{\tau}$ = 114 days) to a number of excited states of ^{56}Fe with gamma ray cascades to the stable ground state of ^{56}Fe. Both the positrons and the gamma rays heat the ejected material. If ^{56}Co is an energy source there should be spectral evidence for cobalt in newly discovered Type I supernovae since its lifetime is long enough for detailed observations to be possible after the initial discovery.

The cobalt has been observed! Axelrod (91) studied the optical spectra of SN1972e obtained by Kirshner and Oke (92). The spectra obtained at 233, 264 and 376 days after Julian day 2441420, assigned as the initial day of the explosive event, are shown in Figure 24. Axelrod assigned the two emission lines near 6000Å (log v = 14.7) to Co III. They are clearly evident at 233 and 264 days, but are only marginally evident at 376 days (- $\bar{\tau}$) later. The lines decay in reasonable agreement with the mean lifetime of ^{56}Co.

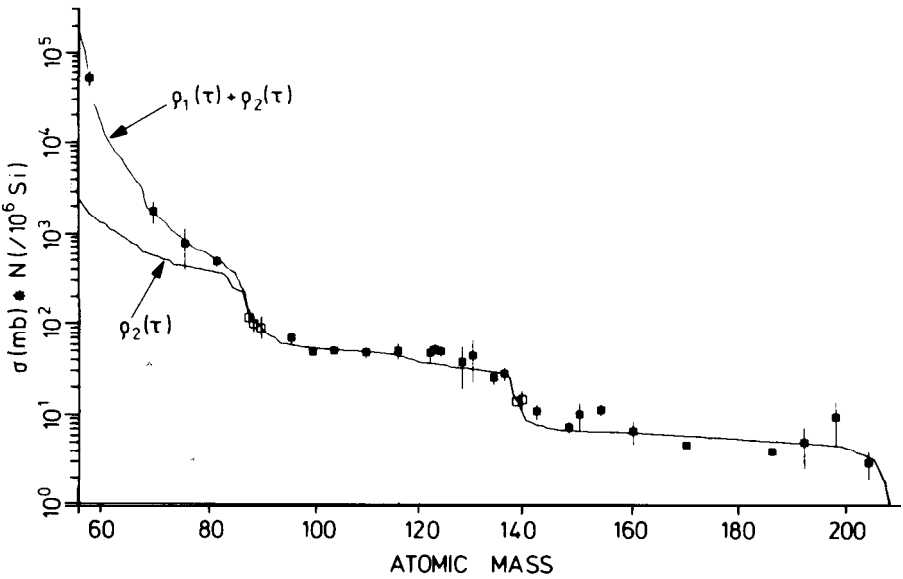


Figure 26. Neutron capture cross section at 30 keV in millibarns multiplied by solar system abundances relative to Si = 10^6 versus atomic mass for nuclei produced in the s-process from Almeida and Käppeler (99). Theoretical calculations are shown for a single exponential distribution $p_1(\tau)$ in neutron exposure, τ, and for two such distributions, $p_1(\tau) + p_2(\tau)$.

Branch *et al.* (93) have studied absorption spectra during the first hundred days of SN1981b. Their results at maximum light are shown in the top curve of Figure 25. Using the carbon deflagration model for Type I supernovae of Nomoto (67), Branch (94) has calculated the spectrum shown in the lower curve of Figure 25. Deep absorption lines of Co II are clearly evident near 3300Å and 4000Å.

It is my conclusion that there is substantial evidence for nucleosynthesis in supernovae of elements produced in oxygen and silicon burning. The role of neutron capture processes in supernovae will be discussed in the next section.

XI. *Neutron capture processes in nucleosynthesis*

In Section I the need for two neutron capture processes for nucleosynthesis beyond $A \gtrsim 60$ was discussed in terms of early historical developments in Nuclear Astrophysics. These two processes were designated the s-process for neutron capture slow (s) compared to electron beta-decay and the r-process for neutron capture rapid (r) compared to electron beta-decay in the process networks.

For a given element the heavier isotopes are frequently bypassed in the s-process and are produced only in the r-process; thus the designation r-only. Lighter isotopes are frequently shielded by more neutron rich stable isobars in the r-process and are produced only in the s-process; thus the designation s-only. The lightest isotopes are frequently very rare because they are not produced in either the s-process or the r-process and are thought to be produced in what is called the p-process. The p-process involves positron produc-

tion and capture, proton-capture, neutron-photoproduction and/or (*p, n)*-reactions and will not be discussed further. The reader is referred to Audouze and Vauclair (95). The results of the s-process, the r-process and the p-process are frequently illustrated by reference to the ten stable isotopes of tin. The reader is referred to Figures 10 and 11 of the first reference in Fowler (96).

It is fair to say that the s-process has the clearest phenomenological basis of all processes of nucleosynthesis. This is primarily the result of the correlation of s-process abundances first delineated by Seeger, Fowler and Clayton (97) with the beautiful series of measurements on neturon capture cross sections in the 1 to 100 keV range by the Oak Ridge National Laboratory group under Macklin and Gibbons (98).

This correlation is illustrated in Figure 26 which shows the product of neutron capture cross sections (σ) at 30 KeV multiplied by s-process abundances (N) as a function of atomic mass for s-only nuclei and those produced predominately by the s-process. It is not difficult to understand in first order approximation that the product σN should be constant in the s-process synthesis. A nucleus with a small (large) neutron capture cross section must have a large (small) abundance to maintain continuity in the s-capture path. Figure 26 demonstrates this in the *plateaus* shown from $A = 90$ to 140 and from $A = 140$ to 206. The anomalous behavior below $A = 80$ is discussed in Almeida and Käppeler (99) from which Figure 26 is taken.

Nuclear shell structure introduces the *precipices* shown in Figure 26 at A~84, A~138 and A~208 which correspond to the s-process abundance peaks in Figure 2. At these values for *A* the neutron numbers are "magic," N= 50, 82, and 126. The cross sections for neutron capture into new neutron shells are very small at these magic numbers. With a finite supply of neutrons it follows that the σN product must drop to a new plateau just as observed. Quantitative explanations of this effect have been given by Ulrich (100) and by Clayton and Ward (101).

What is the site of the s-process and what is the source of the neutrons? A very convincing answer has been given by Iben (102) that the site is the He burning shell of a pulsating Red Giant with the $^{22}Ne(\alpha,n)^{25}Mg$ reaction as the neutron source. Critical discussions have been given by Almeida and Käppeler (99) and by Truran (103). The latter reference reserves the possibility that the $^{13}C(\alpha,n)^{16}O$ reaction is the neutron source.

We turn now to the r-process. This process has been customarily treated by the *waiting point* method of B^{2}FH (18). Under explosive conditions a large flux of neutrons drives nuclear seeds to the neutron rich side of the valley of stability where, depending on the temperature, the (n,γ)-reaction and the (γ,n)-reaction reach equality. The nuclei wait at this point until electron beta-decay transforms neutrons in the nuclei into protons whence further neutron capture can occur. At the cessation of the r-process the neutron rich nuclei decay to their stable isobars. In first order this means that the abundance of an r-process nucleus multiplied by the electron beta-decay rate of its neutron rich r-process isobar progenitor will be roughly constant. At magic neutron numbers in the neutron rich progenitors, beta-decay must perforce open the closed neutron

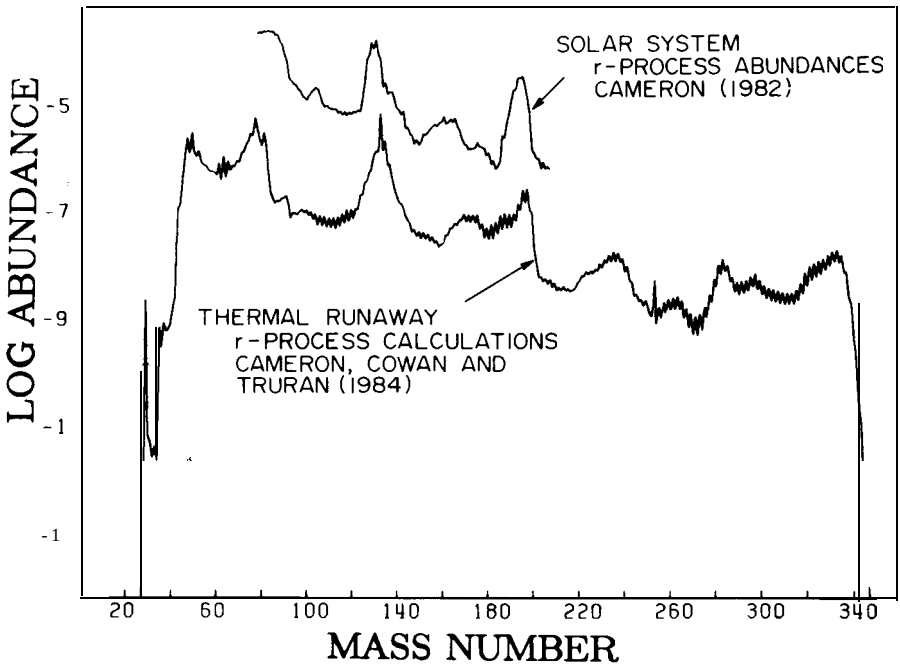


Figure 27. Abundances produced in the r-process versus atomic mass number in the thermal runaway model (lower curve) of Cameron, Cowan and Truran (107) compared with the solar system r-process abundances (upper curve) of Cameron (3).

shell in transforming a neutron into a proton and there the rate will be relatively small. Accordingly the abundance of progenitors with N = 50, 82 and 126 will be large. The associated number of protons will be less than in the corresponding s-process nuclei with a magic number of neutrons. It follows then that the stable daughter isobars will have smaller mass numbers and this is indeed the case, the r-process abundance peaks occurring at A~80, A~130 and A~ 195, all below the corresponding s-process peaks as illustrated in Figure 2.

A phenomenological correlation of r-process abundances with beta-decay rates made by Becker and Fowler (104) and a detailed illustration of this correlation between solar system r-process abundances and theory is given in Figure 13 of the first of references (96). It is too phenomenological to satisfy critical nuclear astrophysicists. They wish to know the site of the high neutron fluxes demanded for r-process nucleosynthesis and the details of the r-process path through nuclei far from the line of beta-stability.

There is also general belief at the present time that the waiting point approximation is a poor one and must be replaced by dynamical r-process flow calculations taking into account explicit $(n,\gamma),(\gamma,n)$ and beta-decay rates with time varying temperature and neutron flux. Schramm (105) has discussed such calculations in some detail and has emphasized that nonequilibrium effects are particularly important during the freeze-out at the end of the r-process when the **temperature drops** and the neutron flux falls to zero. Simple dynamical calcula-

tions have been made by Blake and Schramm (106) for a process they designated as the n-process and Sandler, Fowler, and Koonin (85) for their *n*β-process discussed in Section IX. The most ambitious calculations have been made by Cameron, Cowan, and Truran (107). This paper gives references to their previous herculean efforts in dynamical r-process calculations. An example of their results are shown in Figure 27. They emphasize that they have not been able to find a plausible astrophysical scenario for the initial ambient conditions required for Figure 27. In spite of this I am convinced that they are on the right track to an eventual understanding of the dynamics and site of the r-process.

Many suggestions have been made for possible sites of the r-process almost all in supernovae explosions where the basic requirement of a large neutron flux of short duration is met. These suggestions are reviewed in Schramm (105) and Truran (103). To my mind the helium core thermal runaway r-process of Cameron, Cowan, and Truran (107) is the most promising. These authors do not rule out the $^{22}Ne(\alpha,n)^{25}Mg$ reactions as the source of the neutrons but their detailed results shown in Figure 27 are based on the $^{13}C(\alpha,n)^{16}O$ reaction as the source. They start with a star formed from material with the same heavy element abundance distribution as in the solar system but with smaller total amount. They assume a significant amount of ^{13}C in the helium core of the star after hydrogen burning. This ^{13}C was produced previously by the introduction of hydrogen into the core which had already burned half of its helium into ^{12}C. For Figure 27 they assume a ^{13}C abundance of 14.3 percent by mass, density equal to 10^{6} gm cm^{-3}, and an initial temperature of 1.6 x 10^{8} K which is raised by the initial slow ^{13}C burning to an eventual maximum of 3.6 x 10^{8} K. The electrons in the core are initially degenerate but the rise in temperature lifts the degeneracy producing a thermal runaway with expansion and subsequent cooling of the core. This event is the second helium-flash episode in the history of the core and, if it occurs, only a small amount of the r-process material produced need escape into the interstellar medium to contribute the r-process abundance in solar system material. It is my belief that a realistic astrophysical site for the thermal runaway, perhaps with different initial conditions, will be found. I rest the case.

XI I. Nucleocosmochronology

Armed with their r-process calculations of the abundances for the long lived parents of the natural radioactive series ^{232}Th, ^{235}U, and ^{238}U and with the then current solar system abundances of these nuclei B^{2}FH(18) were able to determine the duration of r-process nucleosynthesis from its beginning in the first stars in the Galaxy up to the last events before the formation of the solar system. The general idea was originally suggested by Rutherford (108). B^{2}FH (18) made a major advance in taking into account the contributions to the abundance of the long lived *eon glasses* from the decay of their short lived progenitors also produced in the r-process. The parents of the natural radioactive series are indeed excellent eon glasses with their mean lifetimes: ^{232}Th, 20.0 x 10^{9} years; ^{238}U, 6.51 x 10^{9} years; ^{235}U 1.03 x 10^{9} years. The analogy with

hour glasses is fairly good; the sand in the top of the *hour glass* is the radioactive parent, that in the bottom is the daughter. The analogy fails in that in the eon glasses "sand" is being added or removed, top and bottom, by nucleosynthesis (production in stars) and astration (destruction in stars). Properly expressed differential equations can compensate for this failure.

The abundances used were those observed in meteorites assumed to be closed systems since their formation, taken to have occurred 4.55 billion years ago. It was necessary to correct for free decay during this period in order to obtain abundances for comparison with calculations based on r-process production plus decay over the duration Galactic nucleosynthesis before the meteorites became closed systems. Fortunately ratios of abundances, $^{232}Th/^{238}U$ and $^{235}U/^{238}U$, sufficed since absolute abundances could not, and still cannot, be calculated with the necessary precision. The calculations required only the elemental ratio, Th/U, in meteorites since the isotopic ratio, $^{235}U/^{238}U$ was assumed to be the same for meteoritic and terrestrial samples. The Apollo Program has added lunar data to the meteoritic and terrestrial in recent years.

B^{2}FH (18) considered a number of possible models, one of which assumed r-process nucleosynthesis *uniform* in time and an arbitrary time interval between the last r-process contribution to the material of the solar nebula and the closure of the meteorite systems. A zero value for this time interval indicated that the production of uranium started 18 billion years ago. When this time interval was taken to be 0.5 billion years, the production started 11.5 billion years ago. These values are in remarkable, if coincidental, concordance with current values.

It is appropriate to point out at this point that nucleocosmochronology yields, with additional assumptions, an estimate for the age of the expanding Universe completely independent of red shift-distance observations in astronomy on distant galaxies. The assumptions referred to in the previous sentence are that the r-process started soon, less than a billion years, after the formation of the Galaxy and that the Galaxy formed soon, less than a billion years, after the "big bang" origin of the Universe. Adding a billion years or so to the start of r-process nucleosynthesis yields an independent value, based on radioactivity, for the age or time back to the origin of the expanding Universe.

Much has transpired over the recent years in the field of nucleocosmochronology. I have kept my hand in most recently in Fowler (109). Exponentially decreasing nucleosynthesis with the time constant in the negative exponent a free parameter to be determined by the observed abundance ratios along with the duration of nucleosynthesis was introduced by Fowler and Hoyle (110). For the time constant in the denominator of the exponent set equal to infinity, uniform synthesis results. When it is set equal to zero, a single spike of synthesis. results. With two observed ratios, two free parameters in a model can be determined. As time went on the ratios $^{129}I/^{127}I$ and $^{244}Pu/^{238}U$ with $\tau(^{129}I) = 0.023 \times 10^9$ years and $\tau(^{244}Pu) = 0.117 \times 10^9$ years were added to nucleocosmochronology to permit the determination of two additional free parameters, the arbitrary time interval of B^{2}FH (18) previously discussed and the

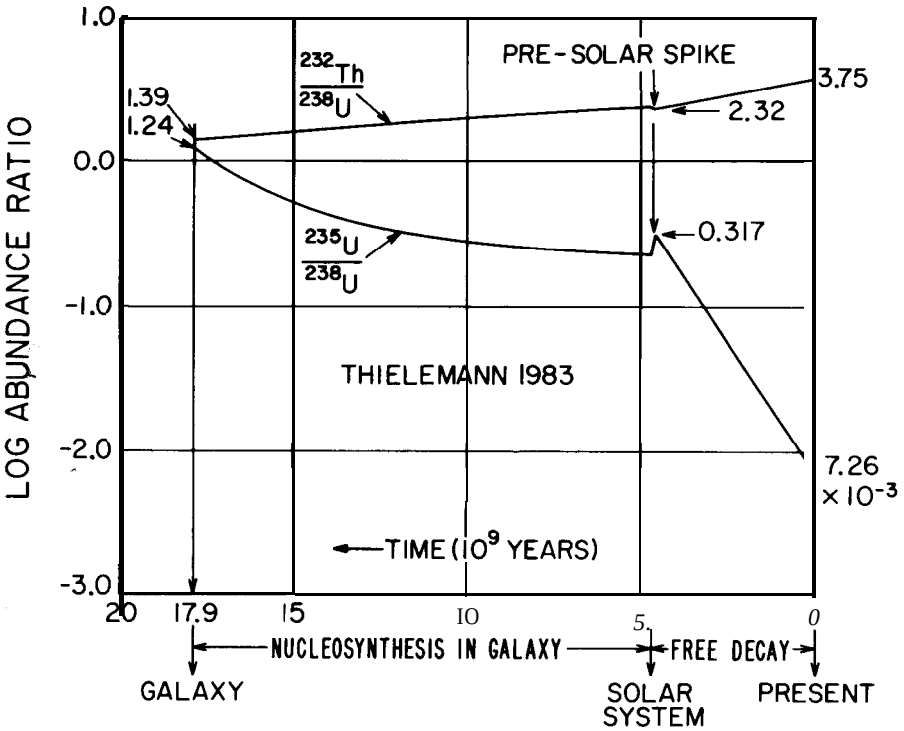


Figure 28. The abundance ratios for $^{232}Th/^{238}U$ and for $^{235}U/^{238}U$ produced by theoretical r-process nucleosynthesis over the lifetime of the Galaxy prior to the information of the solar system from Thielemann, Metzinger and Klapdor (113). The free decays over the lifetime of the solar system to reach the present values for these ratios, $(^{232}Th/^{238}U)_0 = 3.75$ and $(^{235}U/^{238}U)_0 = 7.26 \times 10^{3}$ are also shown. The production ratios in each r-process event was theoretically calculated to be 1.39 for $^{232}Th/^{238}U$ and 1.24 for $^{235}U/^{238}U$. Compare with Figure 10 in first reference under Fowler (109).

fraction of r-process nucleosynthesis produced in a last gasp "spike" at the end of the exponential time dependence.

Sophisticated models of Galactic evolution were introduced by Tinsley (111). A method for model independent determinations of the mean age of nuclear chronometers at the time of solar system formation was developed by Schramm and Wasserburg (112). In this method the mean age is one-half the duration for uniform synthesis and is equal to the actual time of single spike nucleosynthesis. This indicates that one can expect no more than a range of a factor of two in the time back to the beginning of nucleosynthesis in widely different models for its time variation. These developments are reviewed in Schramm (105).

The most recent calculations are those of Thielemann, Metzinger and Klapdor (113). Their results, revised by his own most recent calculations, are shown in Figure 28 prepared by F.-K. Thielemann. The pre-solar spike and its time of occurrence before the meteorites became closed systems depend primarily on the *minute glasses,* ^{129}I and ^{244}Pu. The *eon glasses,* $^{232}Th/^{238}U$ $^{235}U/^{238}U$ $^{235}U/^{238}U$, indicate that r-process nucleosynthesis in the Galaxy started 17.9 billion years

ago with uncertainties of +2 billion years and -4 billion years according to reference (113). This value is to be compared with my value of 10.5±2.3 billion years ago given in Fowler (109). Inputs of production and final abundance ratios have changed in (113)! Thielemann and I are now recomputing the new value for the duration using an initial spike in Galactic synthesis plus uniform synthesis thereafter. It should be noted that 1 to 2 billion years must be added to the age of the Galaxy to obtain the age of the Universe.

Reference (113) indicates that the age of the expanding Universe is 19 billion years give or take several billion years. This is to be compared to the Hubble time or reciprocal of Hubble's constant given by Sandage and Tammann (114) as 19.5 ± 3 billion years. However, the Hubble time is equal to the age of the expanding Universe only for a completely open Universe with mean matter density much less than the critical density for closure which can be calculated from the value' for the Hubble time just given to be 5×10^{-30} gm cm^{-3}. The observed visible matter in galaxies is estimated to be ten percent of this which reduces the age of the Universe to 16.5 billions years. Invisible matter, neutrinos, black holes, etc. may add to the gravitational forces which decrease the velocity of expansion and may thus decrease the age to that corresponding to critical density which is 13.0 billion years. The new concept of the inflationary universe yields exactly the critical density and thus support the value of 13 billion years. If the expansion velocity was greater in the past, the time to the present radius of the Universe is correspondingly less. Moreover, there are those who obtain results for the Hubble time equal to about one-half that of Sandage and Tammann (114) as reviewed in van den Bergh (115). There is much to be done on all fronts!

A completely independent nuclear chronology involving the radiogenic ^{187}Os produced during Galactic nucleosynthesis by the decay of ^{187}Re($i = 65 \times 10^9$ years) was suggested by Clayton (116). Schramm (105) discusses still other chronometric pairs. Clayton's suggestion involves the s-process even though the ^{187}Re is produced in the r-process. It requires that the abundance of ^{187}Re, the parent, be compared to that of its daughter, ^{187}Os, when the s-only production of this daughter nucleus is subtracted from its total solar system abundance. This was to be done by comparing the neutron capture cross section of ^{187}Os with that of its neighboring s-only isotope ^{186}Os which does not have a longlived radioactive parent and using the $N\sigma$ = constant rule for the s-process.

Fowler (117) threw a monkey wrench into the works by pointing out that ^{187}Os has a low-lying excited state at 9.75 keV which. is practically fully populated at $kT = 30$ keV corresponding to the temperature $T = 3.5 \times 10^8$K, at which the s-process is customarily assumed to occur. Moreover with spin, $J = 3/2$, this state has twice the statistical weight $(2J + 1)$ of the ground state with spin, $J = 1/2$. Measurements of the ground state neutron capture cross section yields only one-third of what one needs to know.

All of this has led to a series of beautiful and difficult measurements for neutron induced reactions on the isotopes of osmium. Winters and Macklin (118) found the Maxwell-Boltzmann average ground state (laboratory) cross-

section ratio for ^{186}Os(n, γ) relative to that for ^{187}Os(n, γ) to be 0.478 ± 0.022 at kT= *30* keV with a slow dependence on temperature. This ratio must be multiplied by a theoretical factor to correct the ^{187}Os cross section in the denominator of the cross-section ratio for that of its excited state. The larger the theoretical ^{187}Os excited state capture, the smaller this factor. Woosley and Fowler (119) used Hauser-Feshbach theory to give an estimate for this factor in the range 0.8 to 1.10 which is little comfort in view of the fact that it multiplies one number comparable to the number from which it must be subtracted. These factors translated into a time for the beginning of the r-process in the Galaxy in the range 14 to 19 billion years. In desperation I suggested privately that inelastic neutron scattering off the ground state of ^{187}Os to its excited state at 9.75 keV might yield information on the properties of the excited state. Measurements by Macklin et al. (120) and Hershberger *et al.* (12 1) determined these inelastic neutron scattering cross sections which yielded inherent support of the lower value of the Woosley and Fowler (119) factor and thus a greater value for the time back to the beginning of r-process nucleosynthesis in the 18 to 20 billion year range. It has to be admitted that this is concordant with the latest value from the Th/U-nucleocosmochronology.

Once again in desperation I privately suggested that measurement of the neutron capture cross section on the ground state of ^{189}Os might be helpful. In ^{189}Os the ground state has the same spin and Nilsson numbers as the excited state of ^{187}Os and the excited state corresponds to the ground state of ^{187}Os. Measurements by Browne and Berman (122) are available but are now being checked by an Oak Ridge National Laboratory, Denison University, and University of Kentucky consortium.

It will be clear that the lifetime of ^{187}Re comes directly into the calculations under discussion. There has been some discrepancy in the past between lifetimes measured geochemically and those measured directly by counting the electrons emitted in the 2.6 keVdecay ^{187}Re ($e^-\nu$) ^{187}Os. Direct measurement yields only the lifetime for electronemission to the continuum while geochemistry yields the lifetime for electron-emission both to the continuum and to bound states in ^{187}Os. The entire matter is treated in considerable theoretical detail by Williams, Fowler, and Koonin (123) who found that the bound-state decay is negligible and that the direct measurements by Payne and Drever (124), which agree with the geochemical measurements of Hirt et al. (125), are correct.

There is also the vexing problem of the possible decrease in the effective lifetime of ^{187}Re in the Galactic environment. The ^{187}Re included in the material of the interstellar medium which forms new stars is subject to destruction by the s-process (astration) as well as being produced by the r-process. This decreases the effective lifetime of the ^{187}Re and all chronometric times based on the Re/Os chronology. This problem is discussed in elaborate detail by Yokoi, Takahashi and Arnould (126). The time back to the beginning of r-process nucleosynthesis could be as low as 12 billion years. It appropriates to end this last section before the concluding section with considerable uncertainity in nucleocosmochronology indicating that, as in all nuclear astrophysics,

there is much exciting experimental and theoretical work to be done for many years to come. Amen!

XIII. *Conclusion*

In spite of the past and current researches in experimental and theoretical Nuclear Astrophysics, illustrated in what I have just shown you, the ultimate goal of the field has not been attained. Hoyle's grand concept of element synthesis in the stars will not be truly established until we attain a deeper and more precise understanding of many nuclear processes operating in astrophysical environments. Hard work must continue on all aspects of the cycle: experiment, theory, observation. It is not just a matter of filling in the details. There are puzzles and problems in each part of the cycle which challenge the basic ideas underlying nucleosynthesis in stars. Not to worry-that is what makes the field active, exciting and fun! It is a great source of satisfaction to me that the Kellogg Laboratory continues to play a leading role in experimental and theoretical Nuclear Astrophysics.

And now permit me to pass along one final thought in concluding my lecture. My major theme has been that all of the heavy elements from carbon to uranium have been synthesized in stars. Let me remind you that your bodies consist for the most part of these heavy elements. Apart from hydrogen you are 65 percent oxygen and 18 percent carbon with smaller percentages of nitrogen, sodium, magnesium, phosphorus, sulphur, chlorine, potassium, and traces of still heavier elements. Thus it is possible to say that you and your neighbor and I, each one of us and all of us, are truly and literally a little bit of stardust.

Charles Christian Lauritsen taught me a Swedish toast. I conclude with this toast to my Swedish friends: *"Din skål, min skål. alla vackra flickors skål. Skål!"*

ACKNOWLEDGEMENTS

My work in Nuclear Astrophysics has involved collaborative team work with many people and I am especially grateful to Fay Ajzenberg-Selove, Jean Audouze, C. A. Barnes, E. M Burbidge, G. R. Burbidge, G. R. Caughlan, R. F. Christy, D. D. Clayton, G. M. Fuller, J. L. Greenstein, Fred Hoyle, Jean Humblet, R. W. Kavanagh, S. E. Koonin, C. C. Lauritsen, Thomas Lauritsen, D. N. Schramm, T. A. Tombrello, R. V. Wagoner, G. J. Wasserburg, Ward Whaling, S. E. Woosley, and B. A. Zimmerman.

For aid and helpful cooperation in all aspects of my scientific work, especially in the preparation of publications, I am grateful to Evaline Gibbs, Jan Rasmussen, Kim Stapp, Marty Watson and Elisabeth Wood. I acknowledge support for my research over the years by the Office of Naval Research (1946 to 1970) and by the National Science Foundation (1968 to present).

REFERENCES

1. J. Audouze and H. Reeves, ***p, 355, Essays in Nuclear Astrophysics*** edited by C. A. Barnes, D. D. Clayton, and D. N. Schramm (Cambridge University Press 1983).
2. H. E. Suess and H. C. Urey, Revs. Mod. Phys. 28,53 (1956).
3. A. G. ***W.*** Cameron, ***p. 23, Essays in Nuclear Astrophysics,*** edited by C. A. Barnes, D. D. Clayton and D. N. Schramm (Cambridge University Press 1982).
4. W. Whaling, p, 65, ***Essays in*** Nuclear ***Astrophysics,*** edited by C. A. Barnes, D. D. Clayton, and D. N. Schramm (Cambridge University Press 1982).
5. R. A. Alpher and R. C. Herman, Revs. Mod. Phys. 22, 153 (1950).
6. R. V. Wagoner, W. A. Fowler, and F. Hoyle, Ap. J, 148, 3 (1967).
7. H. Staub and W. E. Stephens, Phys. Rev 55, 131 (1939).
8. J. H. Williams, W. G. Shepherd, and R. 0. Haxby, Phys. Rev. 52, 390 (1937).
9. A. V. Tollestrup, W. A. Fowler, and C. C. Lauritsen, Phys. Rev 76, 428 (1949).
10. A. Hemmendinger, Phys. Rev. 73, 806 (1948); Phys. Rev. 74, 1267 (1949).
11. H. A. Bethe, Phys. Rev. 55, 434 (1939); H. A. Bethe, Les ***Prix Nobel 1967*** (Almquist & Wikseli International, Stockholm).
12. F. Hoyle, Mon. Not. R. Astron. Soc. 106, 343 (1946); Ap. J. Suppl. 1, 121 (1954). For a discussion of earlier ideas, including suggestions and retractions, see Chap. XII in S. Chandrasekhar, ***Stellar Structure*** 1939 (University of Chicago Press, Chicago).
13. F. Hoyle and M. Schwarzschild, Ap. J. Suppl. 2, 1 (1955).
14. A. R. Sandage and M. Schwarzschild, Ap. J. 116, 463 (1952). In particular see last paragraph on p. 475.
15. E. E. Salpeter, Ap. J. 115, 326 (1952).
16. D. N. F. Dunbar, R. E. Pixley, W. A. Wenzel, and W. Whaling, Phys. Rev. 92,649 (1953).
17. C. W. Cook, W. A. Fowler, C. C. Lauritsen, and T. Lauritsen, Phys. Rev. 107, 508 (1957).
18. E. M. Burbidge, G. R. Burbidge, W. A. Fowler, and F. Hoyle, Rev. Mod. Phys. 29, 547 (1957). Hereafter referred to as B^2FH (18). Also see F. Hoyle, W. A. Fowler, E. M. Burbidge, and G. R. Burbidge, Science 124,611 (1956).
19. A. G. W. Cameron, Publ. Astron. Soc. of Pac., 69. 201 (1957).
20. J. L. Greenstein, Chapter 10, ***Modern Physics for the Engineer*** edited by L. N. Ridenour (McGraw-Hill ***1954);*** p, ***45, Essays in Nuclear Astrophysics,*** edited by C. A. Barnes, D. D. Clayton, and D. N. Schramm (Cambridge University Press 1982). Of key importance was the discovery of technicium lines in S-stars by P.W. Merrill, Science, 115, 484 (1952).
21. W. A. Fowler, G. R. Caughlan, and B. A. Zimmerman, Ann. Rev. Astron. and Astroph. 5, 525 (1967); 13, 69 (1975). Also see M. .J. Harris ***et al. ibid.*** 21, 165 (1983) and G. R. Caughlan et al. Atomic Data and Nuclear Data Tables (1984), accepted for publication.
22. E. E. Salpeter, Phys. Rev. 88, 547 (1957); 97, 1237 (1955).
23. S. E. Woosley, W. A. Fowler, J, A. Holmes, and B. A. Zimmerman, Atomic Data and Nuclear Data Tables 22, 371 (1978).
24. H. A. Bethe and C. L. Critchfield, Phys. Rev. 54, 248 (1938).
25. W. A. Fowler, Ap. J. 127, 551 (1958).
26. A. G. W. Cameron, Ann. Rev. Nucl. Sci. 8, 249 (1958).
27. B. Pontecorvo, Chalk River Laboratory Report PD-205 (1946).
28. L. W. Alvarez, University of California Radiation Laboratory Report UCRL-328 (1949).
29. R. Davis, Jr., p. 2, ***Science Underground,*** edited by M. M. Nieto *et al.* (American Institute of Physics, New York 1983).
30. J. N. Bahcall, W. F. Huebner. S. H. Lubow, P. D. Parker, and R. K. Ulrich, Rev. Mod. Phys. 54, 767 (1982).
31. R. G. H. Robertson, P. Dyer, T. J. Bowles, Ronald E. Brown, Nelson Jarmie, C. J. Maggiore, and S. M. Austin, Phys. Rev. C27, 11 (1983); J. I.. Osborne, C. A. Barnes, R. W. Kavanagh, R. M. Kremer, G. J. Mathews and J. L. Zyskind, Phys, Rev. Letters 48, 1664 (1982).
32. R. T. Skelton and R. W. Kavanagh, Nuclear Physics A414, 141 (1984).
33. ***C.*** A. Barnes, p, 193, ***Essays in Nuclear Astrophysics,*** edited by C. A. Barnes, D. P. Clayton, and D. N. Schramm (Cambridge University Press 1982).

34. K. Langanke and S. E. Koonin, Nucl. Phys. A410, 334 (1983) and private communication (1983).
35. P. Dyer and C. A. Barnes, Nucl. Phys. A233, 495 (1974); S. E. Koonin, T. A. Tombrello, and G. Fox, Nucl. Phys. A220, 221 (1974).
36. K U. Kettner, H. W. Becker, L. Buchmann, J. Gorres, H. Kräwinkel, C. Rolfs, P. Schmalbrock, H. P. Trautvetter, and A. Vlieks, Zeits. f. Physik A308, 73 (1982).
37. S. G. Starrfield, A. N. Cox, S. W. Hodson, and W. D. Pesnell, Ap. J. 268, L27 (1983); S. A. Becker, private communication (1983).
38. S. E. Woosley and T. A. Weaver, p. 381, Essays in Nuclear ***Astrophysics,*** edited by C. A. Barnes, D. D. Clayton, and D. N. Schramm (Cambridge University Press 1982).
39. T. A. Weaver, S. E. Woosley and G. M. Fuller, ***Proceedings of the*** Conference on ***Numerical Astrophysics,*** edited by R. Bowers, J. Centrella, J. LcBlanc, **and** M. LeBlanc (Science Books International 1983).
40. G. Hulke, C. Rolfs, and H. P. Trautvetter, Zeits, f. Physik A297, 161 (1980).
41. D. Bodansky, D. D. Clayton, and W. A. Fowler, Ap. J. Suppl. 16, 299 (1968).
42. J. L. Zyskind, J. M. Davidson, M. T. Esat, M. H. Shapiro, and R. H. Spear, Nucl. Phys. A301, 179 (1978).
43. L. W. Mitchell and D. G. Sargood, Aust. J. Phys. 36, 1 (1983).
44. R. ***N.*** Boyd, ***Proceedings of the Workshop on Radioactive ion Beams and Small Cross Section Measurements,*** (The Ohio State University Press, Columbus 1981); R. C. Haight, G. J. Mathews, R. M. White, L. A. Avilés, and S. E. Woodward, Nuclear Instruments and Methods 212, 245 (1983).
45. W. Hauser and H. Feshbach, Phys. Rev. 78, 366 (1952).
46. J. A. Holmes, S. E. Woosley, W. A. Fowler, and B. A. Zimmerman, Atomic Data and Nuclear Data Tables 18, 305 (1976).
47. G. Michaud and W. A. Fowler, Phys. Rev. C2, 2041 (1970); also see E. W. Vogt, ***Advances in Nuclear Physics,*** 1, 26 1 (1969).
48. F.M. Mann, Hanford Engineering and Development Laboratory internal report HEDI,-TME-7680 (1976, unpublished).
49. J. L. Zyskind, C. A. Barnes, J. M. Davidson, W. A. Fowler, R. E. Marrs, and M. H. Shapiro, Nucl. Phys. A343,295 (1980).
50. D. G. Sargood, Physics Reports 93,61 (1982); Aust. J. Phys. 36, 583 (1983).
51. N. A. Bahcall and W. A. Fowler, Ap. J. 157,645 (1969).
52. R. A. Ward and W. A. Fowler, Ap. J. 238,266 (1980). For recent experimental data on the production of ^{26}Al, through ^{25}Mg(p, γ)^{26}Al, see A. E. Champagne, A. J. Howard, and P. D. Parker, Ap. J. 269, 686 (1983). For recent experimental data on the destruction of ^{26}Al, through ^{2h}Al (p, γ)^{27}Si, see L.. Buchmann, M. Hilgemaier, A. Krauss, A. Redder, C. Rolfs, and H. P. Trautvetter (in publication, 1984).
53. G. M. Fuller, W. A. Fowler, and M. J. Newman, Ap. .J. Suppl. 42, 447 (1980); Ap. J. 252, 715 (1982); Ap. J. Suppl. 48, 279 (1982).
54. H. S. Wilson, R. W. Kavanagh, and F. M. Mann, Phys. Rev. C22, 1696 (1980).
55. C. M. Lederer and V. S. Shirley, Editors, ***Table of Isotopes: Seventh Edition*** (John Wiley & Sons, Inc., New York 1978).
56. C. D. Goodman, C. A. Goulding, M. B. Greenfield, J. Rapaport, D. E. Bainum, C. C. Foster, W. G. Love, and F. Petrovich, Phys. Rev. Lett. 44, 1755 (1980); F. Ajzenberg-Selove, R. E. **Brown, E.** R. Flynn, and J. W. Sunier, Phys. Rev. Lett. (1984), in publication.
57. S. D. Bloom and G. M. Fuller (in preparation, 1984).
58. G. E. Brown, H. A. Bethe and G. Baym, Nucl. Phys. A375, 481 (1982).
59. H. A. Bethe, A. Yahil, and G. E. Brown, Ap. J. Letters 262, L7 (1982).
60. H. A. Bethe, G. E. Brown, J. Cooperstein and J. R. Wilson, Nucl. Phys. A403, 625 (1983).
61. G.M. Fuller, Ap. J. 252, 741 (1982).
62. G. M. Fuller, W. A. Fowler, and M. J. Newman (in preparation, 1984).
63. S. E. Woosley, T. S. Axelrod, and T. A. Weaver, ***Stellar Nucleosynthesis,*** edited by C. Chiosi and A. Renzini (Dordrecht: Reidel 1984).
64. F. Hoyle and W. A. Fowler, Ap. J. 132, 565 (1960).

65. W. A. Fowler and F. Hoyle, Ap. J. Suppl. 9, 201 (1964).
66. J. C. Wheeler, Rep. Prog. Phys. 44, 85 (1981).
67. K. Nomoto, A. J. 253, 798 (1982); Ap. J. 257, 780 (1982); ***Stellar Nucleosynthesis,*** edited by C. Chiosi and A. Renzini (Dordrecht: Reidel 1984).
68. K. Nomoto, F.-K. Thielemann, and J, C. Wheeler, Ap. J, (1984), in publication.
69. W. D. Arnett and F.-K. Thielemann, ***Stellar Nucleosynthesis,*** edited by C. Chiosi and A. Renzini (Dordrecht: Reidel 1984).
70. J. H. Reynolds, Phys. Rev. Lett. 4, 8 (1960).
71. P. M. Jeffery and J. H. Reynolds, J. Geophys. Res. 66, 3582 (1961).
72. G. J. Wasserburg, W. A. Fowler, and F. Hoyle, Phys. Rev. Lett. 4, 112 (1960).
73. G. J. Wasserburg and D. A. Papanastassiou, p. 77, ***Essays in*** Nuclear ***Astrophysics,*** edited by C. A. Barnes, D. D. Clayton, and D. N. Schramm (Cambridge University Press 1982).
74. D. D. Clayton, Ap. J. 199, 765 (1975); Space Sci. Rev. 24, 147 (1979); Ap. J. 268, 381 (1983); Ap. J. (1984), in publication; also see D. D. Clayton and F. Hoyle, Ap. J. Lett. 187, Ll0l (1974); Ap. J. 203, 490 (1976).
75. J. H. Chen and G. J. Wasserburg, Earth Planet Sci. Lett. 52, 1 (1981).
76. D. S. Burnett, M. I. Stapanian, and J. H. Jones, p. 144, ***Essays in*** Nuclear ***Astrophysics,*** edited by C. A. Barnes, D. D. Clayton and D. N. Schramm (Cambridge University Press 1982).
77. Typhoon Lee, D. A. Papanastassiou and G. J. Wasserburg, Ap. J. Lett. 211, L107 (1977).
78. W. A. Mahoney, J. C. Ling, W. A. Wheaton, and A. S. Jacobson, Ap. J. (1984), in publication; also see W. A. Mahony, J. C. Ling, A. S. Jacobson, and R. E. Lingenfelter, Ap. J. 262, 742 (1982).
79. R. T. Skelton, R. W. Kavanagh, and D. G. Sargood, Ap. J. 271, 404 (1983).
80. M. Arnould, H. Nørgaard, F.-K. Thielemann, and W. Hillebrandt, Ap. J. 237, 931 (1980).
81. J. W. Truran, p. 467, ***Essays in Nuclear Astrophysics,*** edited by C. A. Barnes, D. D. Clayton and D. N. Schramm (Cambridge University Press 1982).
82. W. A. Fowler, J, L. Greenstein, and F. Hoyle, Geophys. J. 6, 148 (1962).
83. Typhoon Lee, Ap. J. 224, 217 (1978).
84. M. H: Thiemens and J. E. Heidenreich, Science 219, 1073 (1983).
85. D. G. Sandler, S. E. Koonin, and W. A. Fowler, Ap. J. 259,908 (1982).
86. F. R. Niederer, D. A. Papanastassiou and G. J. Wasserburg, Ap. J. Lett. 228, L93 (1979).
87. A. Huck, G. Klotz, A. Knipper, C. Miéhé, C. Richard-Serre, and G. Walter, CERN 81-09, ***378*** (1981).
88. R. A. Chevalier and R. P. Kirshner, Ap. J. 233, 154 (1979).
89. R. H. Becker, S. S. Holt, B. W. Smith, N. E. White, E. A. Boldt, R. F. Mushotsky, and P. J. Serlemitsos, Ap. J. Letters 234, L73 (1979). For overabundances in oxygen and neon in the x-ray spectra of Puppis A see C. R. Canizares and P. F. Winkler, Ap. J. Letters 246, L33 (1981).
90. .J. M. Shull, Ap. J. 262, 308 (1982) and private communication (1983).
91. T. S. Axelrod, ***Late Time Optical Spectra*** *from* ***the*** ^{56}Ni ***Model for Type*** 1 ***Supenovae,*** Thesis, University of California, Berkeley, UCRL-52994, (1980).
92. R. P. Kirshner and J. B. Oke, Ap. J. 200, 574 (1975).
93. D. Branch, C. H. Lacy, M. L. McCall, P.G. Sutherland, A. Uomoto, J, C. Wheeler, and B. J. Wills, Ap. J. 270, 123 (1983).
94. D. Branch, Proceedings of Yerkes Observatory Conference on "Challenges and New Developments in Nucleosynthesis," edited by W. D. Arnett (University of Chicago Press 1984).
95. J, Audouze and S. Vauclair, p. 92, ***An*** *Introduction to* ***Nuclear Astrophysics,*** (D. Reidel, Dordrecht 1980).
96. W. A. Fowler, Proc. Nat. Acad. Sciences ***52, 524 (1964); Nuclear Astrophysics,*** (American Philosophical Society, Philadelphia 1967).
97. P. A. Seeger, W. A. Fowler, and D. D. Clayton, AP. J. Suppl. 11, 121 (1965).
98. R. L. Macklin and J. H. Gibbons, Rev. Mod Phys. 37, 166 (1965). Also see B. J. Allen, R. L. Macklin, and J. H. Gibbons, Adv. Nucl. Phys. 4, 205 (1971).
99. J. Almeida and F. Kappeler, Ap. J. 265, 417 (1983).

100. R. K. Ulrich, p. 139, ***Explosive Nucleosynthesis,*** edited by D. N. Schramm and W. D. Arnett (University of Texas Press, Austin 1973).
101. D. D. Clayton and R. A. Ward, AP. J. 193, 397 (1974).
102. I. Iben, Jr., Ap. J. 196, 525 (1975).
103. J. W. Truran, p. ***95. International Physics Conference Series No.*** 64, (The Institute of Physics, London 1983).
104. R. A. Becker and W. A. Fowler, Phys. Rev. 115, 1410 (1959).
105. D. N. Schramm, p. ***325, Essays in*** Nuclear ***Astrophysics,*** edited by C. A. Barnes, D. D. Clayton, and D. N. Schramm (Cambridge University Press 1982).
106. J. B Blake and D. N. Schramm, Ap. J. 209,846 (1976).
107. A. G. W. Cameron, J. J. Cowan, and J. W. Truran, Proceedings of Yerkes Observatory Conference on "Challenges, and New Developments in Nucleosynthesis," edited by W. D. Arnett (University of Chicago Press 1984).
108. E. Rutherford, Nature 123, 313 (1929).
109. W. A. Fowler, p. 61, ***Proceedings of the Welch Foundation Conferences on Chemical Research, XXI, Cosmochemistry,*** edited by W. D. Milligan (Robert A. Welch Foundation, Houston 1977); ***also see W.*** A. Fowler, p. ***67, Cosmology, Fusion and Other Matters,*** edited by F. Reines (Colorado Associated University Press 1972).
110. W. A. Fowler and F. Hoyle, Ann. Phys. 10, 280 (1960).
111. B. M.Tinsley, Ap. J. 198, 145 (1975).
112, D. N. Schramm and G. J. Wasserburg, Ap. J. 163, 75 (1970).
113. F.-K. Thielemann, J. Metzinger, and H. V. Klapdor Z. Phys. A309, 301 (1983) and private communication.
114. A. Sandage and G. A. Tammann, Ap. J. 256, 339 (1982).
115. S. van den Bergh, Nature 229, 297 (1982).
116. D. D. Clayton, Ap. J. 139, 637 (1964).
117. W. A. Fowler, Bulletin American Astronomical Society, 4, 412 (1972).
118. R. R. Winters and R. L. Macklin, Phys. Rev. C25, 208 (1982).
119. S. E. Woosley and W. A. Fowler, Ap. J. 233, 411 (1979).
120. R. L. Macklin, R. R. Winters, N. W. Hill, and J. A. Harvey, Ap. J. 274, 408 (1983).
121. R. L. Hershberger, R. L. Macklin, M. Balakrishnan, N. W. Hill and M. T. McEllistrem, Phys. Rev, C28, 2249 (1983).
122. J. C. Browne and B. L. Berman, Phys. Rev. C23, 1434 (1981).
123. R. D. Williams, W. A. Fowler, and S. E. Koonin, Ap. J. (1984), in publication.
124. J, A. Payne, Thesis: ***An Investigation of the Beta Decay of Rhenium to Osmium, Using High Temperature Proportional Counters,*** University of Glasgow (1965); R. W. P. Drever, private communication (1983).
125. B. Hirt, G. R. Tilton, W. Herr, and W. Hoffmeister, ***Earth Scicnce and Meteorites,*** edited by J. Geiss and E. D. Goldberg (North Holland Press, Amsterdam 1963).
126. K. Yokoi, K. Takahashi, and M. Arnould, Astron. Astrophys. 117, 65 (1983).