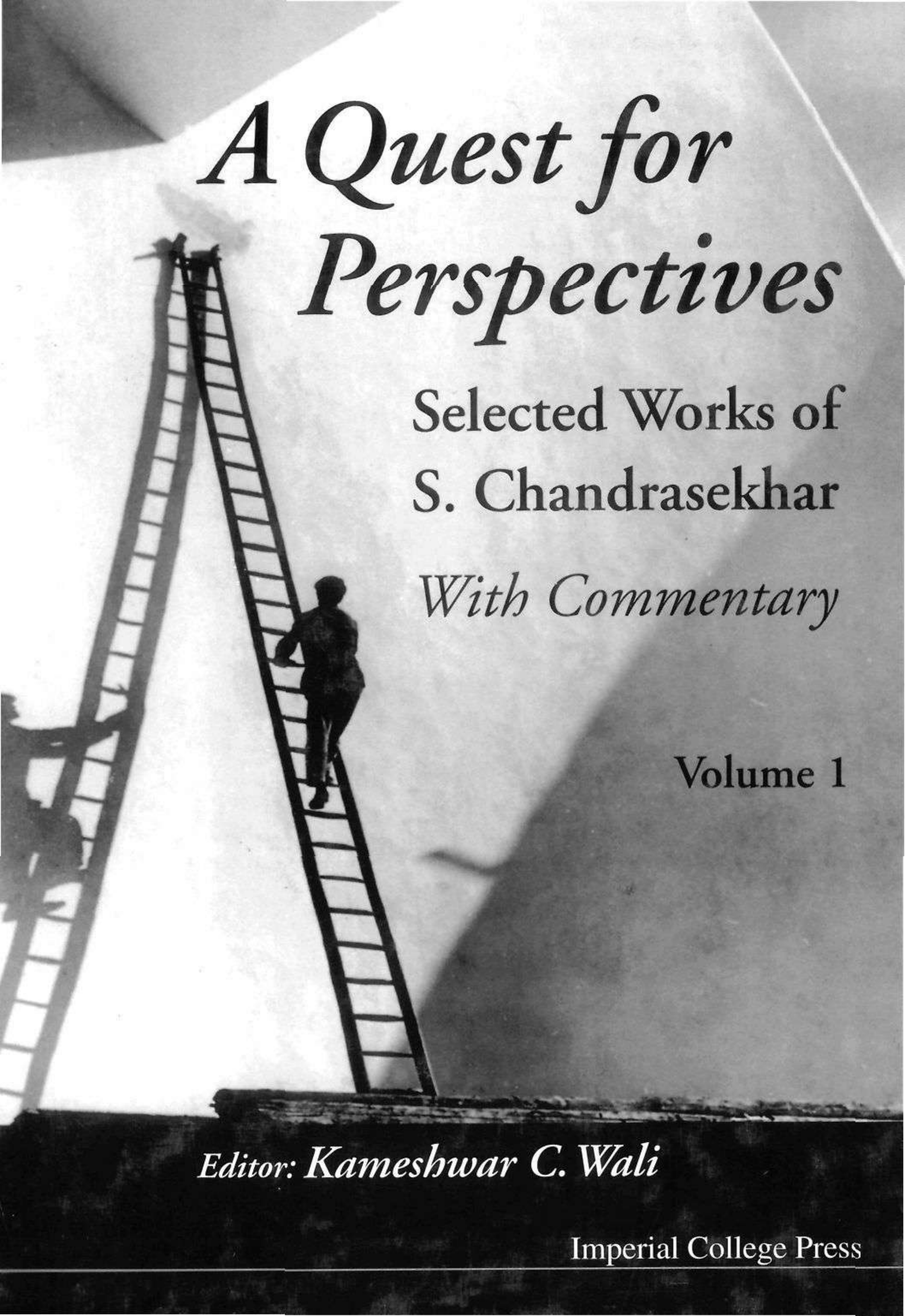


A Quest for Perspectives

Selected Works of S. Chandrasekhar

With Commentary

Volume 1

Editor: Kameshwar C. Wali

Imperial College Press

A Quest for Perspectives

Selected Works of S. Chandrasekhar

With Commentary

Volume 1

This page is intentionally left blank

A Quest for Perspectives

Selected Works of S. Chandrasekhar

With Commentary

Volume 1

Editor

Kameshwar C. Wali

Syracuse University

ICP

Imperial College Press

Published by

Imperial College Press
57 Shelton Street
Covent Garden
London WC2H 9HE

Distributed by

World Scientific Publishing Co. Pte. Ltd.
P O Box 128, Farrer Road, Singapore 912805
USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

The editor and the publisher would like to thank the following organisations and publishers for their assistance and their permission to reproduce the articles found in this volume:

Academic Press (*J. Math. Anal. Appl.*), American Academy of Arts and Sciences, American Mathematical Society, American Philosophical Society, American Physical Society (*Phys. Rev. Lett.*, *Rev. Mod. Phys.*), Blackwell Science (*Mon. Not. R. Astron. Soc.*), Current Science Association, Elsevier Science Publishers B. V., Graduate Institute for Applied Mathematics (*J. Rat. Mech. Anal.*), Kluwer Academic Publishers, Macmillan Magazines (*Nature*), New York Academy of Sciences, Plenum Press, Royal Society, Springer-Verlag (*Z. Astrophys.*), University of Chicago Press (*Astrophys. J.*)

Cover picture: *An Individual's View of the Individual (Man on the Ladder)*, by Piero Borello. Courtesy of the artist.

A QUEST FOR PERSPECTIVES
Selected Works of S. Chandrasekhar (with Commentary)
Volume I

ISBN 1-86094-201-6 (set)
ISBN 1-86094-208-3 (pbk) (set)
ISBN 1-86094-283-0
ISBN 1-86094-284-9 (pbk)

Printed in Singapore by Uto-Print

To

Lalitha Chandrasekhar

This page is intentionally left blank

Contents

This page is intentionally left blank

Preface

It is well known that S. Chandrasekhar (known simply as Chandra in the scientific world) followed a unique style of research. This is best described in his own words in the autobiographical account published along with his Nobel Lecture in 1983:

> After the early preparatory years, my scientific work has followed a certain pattern motivated principally by a quest after perspectives. In practice, 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, ability, 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 an urge to present my point of view *ab initio*, in a coherent account with order, form, and structure.

Thus, Chandra's research, motivated principally by *a quest after perspectives*, covered a wide range of investigations that comprised: (1) stellar structure, including the theory of white dwarfs (1929–39); (2) stellar dynamics, including the theory of Brownian motion (1938–43); (3) the theory of radiative transfer, the theory of the illumination and the polarization of the sunlit sky, and the quantum theory of the negative ion of hydrogen (1943–50); (4) hydrodynamic and hydromagnetic stability (1952–61); (5) the equilibrium and the stability of ellipsoidal figures of equilibrium (1961–68); (6) the general theory of relativity and relativistic astrophysics (1962–71); (7) the mathematical theory of black holes (1974–83); (8) colliding wave, the nonradial oscillations of stars in general relativity and the writing of Newton's *Principia* (1983–95). During each of these periods lasting several years, Chandra produced a series of papers, and in most cases the series ended with a monograph on the subject that represented his "matured outlook, and the subject matter organized in a coherent framework."

Because of these books and monographs, one may question the need for a volume or volumes of collected original papers. Indeed, Chandra himself, when approached by the University of Chicago Press, with the idea of publishing a collection of his papers, thought it would be superfluous to do so. Eventually, when he was persuaded, he sought, in the selection process, the help of Martin Schwarzschild, Robert Mullikin, H.C. van de Hulst, Norman Lebovitz, Kip Thorne, and Basilis Xanthopoulos, who were familiar with one or another major area of research in which he had worked. The selection of the papers was based on two criteria: first, none of the selected papers should have already been included in any of his published books, and second, the papers of possible historical interest should be given preference. The University of Chicago

Press has published six volumes (and a seventh one after his death) containing a major body of Chandra's original papers and articles.

These volumes containing published papers in their original form are of extreme importance to researchers and historians of science. They are an invaluable source for studying the subject as it developed, uninfluenced by Chandra's own perspective and his attempt to provide an integrated and coherent account in the monographs. Bearing this in mind, I have collected together in this single volume a subset of original papers from each of the aforementioned periods of Chandra's research. The first nine parts contain technical articles, while the last part is devoted to some articles of a popular and historical nature. In selecting articles for this anthology, my principal aim has been to include those papers that are to a great extent self-contained and laid the groundwork for subsequent detailed investigations in a given area and a given problem. I have also made it a point to include many of Chandra's articles that are based on his talks at various conferences. They are masterly, complete-with-historical-background, concurrent work, and his own work put in a proper perspective.

While the seven volumes of *Selected Papers* published by the University of Chicago Press are indispensable to libraries and research institutions, I hope the present anthology of selected papers will make Chandra's original papers more affordable and accessible to individual researchers. It gives only a glimpse, perhaps, at the vast arena of theoretical physics and astrophysics that Chandra dominated with his monumental contributions. The readers may benefit by two other books that will supplement this anthology. These are: (1) *Black Holes and Relativistic Stars* (edited by Robert M. Wald, University of Chicago Press, 1998), the proceedings of a "working scientific symposium" on the theory of black holes and relativistic stars that was the preoccupation of Chandra during the last phase of his life; (2) *From White Dwarfs to Black Holes: The Legacy of S. Chandrasekhar* (edited by G. Srinivasan, University of Chicago Press, 1999). The articles in these two volumes provide a grand tour of the colossal scientific edifice Chandra has left behind. While presenting a summary of the important contributions to a particular area, the articles describe also the impact of Chandra's work on the further development of the subject.

Finally, I would like to thank Ms Emily Davis for her help in preparing this anthology. My thanks are also due to Senior Editor H.T. Leong and his colleagues at World Scientific for the excellence they have brought to this book.

Kameshwar C. Wali
November 30, 2000

I. Early Years; The Theory of White Dwarfs and Stellar Interiors

The papers in this section span the years 1929–40, beginning with the very first paper Chandra published, when he was still an undergraduate at the Presidency College, Madras, India. This paper deals with the effect of the then new Fermi–Dirac quantum statistics on Compton scattering.

Chandra was introduced to the new quantum mechanics by Arnold Sommerfeld during the latter's visit to Madras in the fall of 1928. The new quantum mechanics, which had stunned Europe, had not yet made its way to India. Sommerfeld was invited to speak to the science students and Chandra, who was among them, made arrangements to see him the following day in his hotel room. Chandra had mastered the atomic theory as laid out in Sommerfeld's classic book on the old quantum theory, *Atomic Structure and Spectral Lines.* He approached him with the brash confidence of a young undergraduate to impress the master with his knowledge as well as his intense desire to pursue a research career in physics. But Sommerfeld shocked him by telling him that the old quantum theory in his book was no longer of any use. It had been replaced by the revolutionary new quantum mechanics due to Schrodinger, Heisenberg, Dirac, and others. While Chandra had also studied on his own the classical Maxwell–Boltzmann statistics, Sommerfeld told him that too had undergone a fundamental change in the light of the new quantum mechanics. Seeing a crestfallen young student in front of him, Sommerfeld offered him the galley proofs of his as-yet-unpublished paper that contained an account of the new Fermi–Dirac quantum statistics and its application to the electron theory of metals.

Chandra would later characterize this encounter as the "single most important event" in his scientific career. For someone with less determination and passion for study than Chandra, such an encounter would have had disastrous consequences. But Chandra immediately launched into a serious study of the new developments in atomic theory. From Sommerfeld's paper, he learned enough about Fermi–Dirac statistics to write, within a few months, his first paper, titled "The Compton Scattering and the New Statistics." He sent it to Ralph H. Fowler in Cambridge, England, requesting him to communicate it for publication in the *Proceedings of the Royal Society.* Chandra had chosen Fowler since he had come across, by pure chance, soon after his encounter with Sommerfeld, Fowler's paper on the theory of collapsed configuration of stars, namely the white dwarfs. Published in the *Monthly Notices of the Royal Astronomical Society,* it contained still another application of Fermi–Dirac statistics to the stellar matter in the form of degenerate electrons in white dwarfs. So for Chandra, at the time, Fowler was someone who knew Fermi–Dirac statistics and consequently someone who could understand his paper and help its publication. The paper was indeed published in the *Proceedings.* However, this chance circumstance was to have a profound influence on Chandra's future scientific career. The following year, when he was offered unexpectedly a Government of India scholarship to continue his research in England after his graduation, he didn't have to think hard before choosing Cambridge University, and Fowler as his doctoral adviser.

He began to study the theory of white dwarfs and was able immediately to extend Fowler's work by combining Fowler's ideas with Eddington's polytropic considerations for a star. Papers 2 to 5 contain Chandra's most significant work in his early years, leading to the celebrated discovery of the Chandrasekhar limit on the mass of a star that could become a white dwarf and its broader implications for the problem of stellar evolution. Paper 2 contains a brief account of the discovery of the limiting mass he made on his long voyage from India to England in 1930. A point of historical interest: Paper 3, written soon thereafter, contains a more detailed account. Chandra had withheld from submitting it for publication in the British journals because of E.A. Milne's dissent and his influence on what was published and what was not. It is this paper that concludes with the often quoted statement "Great progress in the analysis of stellar structure is not possible before we can answer the following fundamental question: Given an enclosure containing electrons and atomic nuclei (total charge zero), what happens if we go on compressing the material indefinitely?" In order to avoid confrontation with Milne, Chandra waited till he was in Copenhagen at the Niels Bohr Institute, before sending it for publication in *Zeitschrift fuer Astrophysik* at Potsdam, Germany. But, as luck would have it, Milne was in Potsdam and he was asked to referee the paper. Milne advised against publishing the paper and wrote to Chandra saying, "Unfortunately I have been unable to recommend acceptance, as the paper contains a mistake in principle, and in any case it would only do harm to your reputation if it were published."

Papers 4, 5 and 7 are the seminal papers on the theory of white dwarfs. During the fall of 1934, Chandra worked out the complete theory using the exact relativistic equation of state describing degenerate matter. With extensive numerical work, he established beyond any doubt the validity of the limiting mass condition, reiterating the question he had posed in his earlier paper. A star with mass greater than the limiting mass would not reach an equilibrium state of a white dwarf in the course of its evolution. It will continue to collapse. Chandra presented his results at the January 1935 meeting of the Royal Astronomical Society. His findings raised challenging, fundamental questions. What happens to the more massive stars as they continue to collapse? Are there terminal stages of stars other than that of white dwarfs? Paper 6 contains what transpired after Chandra presented his paper (no. 5). Instead of getting appreciation and recognition for a fundamental discovery, Chandra unexpectedly faced what amounted to a public humiliation. Because no sooner had he presented his paper than Sir Arthur Eddington tore apart Chandra's dramatically stated conclusions by attacking the very concept of relativistic degeneracy on which Chandra's work depended. Characterizing it as leading to stellar buffoonery and *reductio ad absurdum* behavior of a star, he made it look as though Chandra had made a simple conceptual error and gotten it all wrong. While he himself missed the great opportunity of being the first one to recognize the possibility of the existence of black holes and other terminal stages of a star, Eddington's attack had a profound influence on Chandra's personal life as well as the progress of astronomy. The irony of this encounter and the consequences of this unexpected occurrence are more fully described elsewhere.

After this historic meeting and the unexpected encounter, Chandra put aside his work on white dwarfs and went on to study stellar dynamics and other problems. But, four years later, in 1939, he was invited to an international meeting in Paris devoted to the special topics of novae and white dwarfs. Chandra and Eddington both gave talks. Paper 8 contains Chandra's paper and the discussion afterward at this meeting. Chandra took the opportunity to state again his conclusions regarding the limiting mass based on relativistic degeneracy. Included in this paper is some work he had done on rotating white dwarfs. He had withheld it from publication because of the controversy. Paper 9 is drawn from the *Proceedings of the American Philosophical Society*. It contains a description of the general methods to determine the physical conditions in stellar interiors and the theory of white dwarfs at the end of the decade of the 1930s.

The last paper in this part is of special interest, in that it is totally out of line with the general body of Chandra's work. It is a very brief communication published in *Nature* on certain numerical coincidences based on purely dimensional arguments. Saying that he was hesitating to publish this on account of the conviction that "purely dimensional arguments" do not lead one very far, he presents a formula in terms of the well-known fundamental constants, h (Planck's constant), c (velocity of light), G (Newton's constant), and H (mass of the proton). The formula, for certain powers of the combination of these constants, yields cosmological constants such as masses of stellar magnitude, the number of particles in the universe, and the mass of the Milky Way.

I. Early Years; The Theory of White Dwarfs and Stellar Interiors

The Compton Scattering and the New Statistics.

By S. Chandrasekhar, The Presidency College, Madras.

(Communicated by R. H. Fowler, F.R.S.—Received June 20, 1929.)

1. *Introduction.*

Great success has been achieved by Sommerfeld in the electron theory of metals by assuming that there are free electrons in them which obey the Fermi-Dirac statistics. It has been assumed in the case of univalent metals that on the average one electron per atom is free. In general, however, the valency electrons can be considered as free.* These free electrons will take part in the Compton scattering. The analysis of such a Compton effect reduces to the analysis of the collisions between radiation quanta and an electron gas. The general features of such a scattering was first considered by Dirac.† But he has assumed a Maxwellian distribution for the electrons which will not be applicable to the case under consideration, because the electrons in a conductor being degenerate do not obey the Maxwell's law, but the Fermian distribution.

In considering such a process we take it that the conservation of momentum and energy principles are satisfied for each particular collision just as in Compton's theory—only we are here dealing with moving electrons instead of stationary electrons which Compton considers. Thus electrons of different momenta components will produce different Compton shifts, and the intensity of any particular shift will depend on the number of electrons in that state. Thus we have to average for the radiation falling on an assembly of electrons whose momenta are distributed according to the Fermi-Dirac law.

The above is just a natural extension of Compton's theory. In this connection mention should be made of Jauncey's‡ theory of bound electrons whose arguments are essentially what we have put forward in the previous paragraph. But his theory has not been quite satisfactory because he has not assumed any definite distribution of the electrons.

* Rosenfeld, 'Naturw.,' p. 49 (1929).
† 'M.N.R.A.S.,' vol. 85, p. 825 (1925).
‡ 'Phys. Rev.,' vol. 25, p. 723 (1925).

2. *Compton scattering with Moving Electrons.*

Let m_x, m_y, m_z be the momentum of the scattering electron and g_x, g_y, g_z those of the quantum, g_t, m_t represent the masses of the electron and the quantum multiplied by the velocity of light c. If we take polar co-ordinates

$$g_x = h\nu \cos \theta/c\ ;\ g_y = h\nu \sin \theta \cos \phi/c\ ;\ g_z = \sin \theta \sin \phi\ \ g_t = h\nu/c. \qquad (1)$$

Then the conservation of momentum and energy gives

$$(m_u, g_u) - (m_u, g_u') = (g_u, g_u'). \qquad (2)$$

The above equation gives the frequency of the scattered quantum in terms of the initial momentum of the electron and the incident quantum, and the directions of the incident and scattered quanta.

Equation (2) reduces to

$$m_t - m_x \cos \theta' - m_y \sin \theta' = \frac{\nu}{\nu'}(m_t - m_x) - \frac{h\nu}{c}(1 - \cos \theta'), \qquad (3)$$

if we assume that the directions of the incident quantum is along the x axis and that of the scattered quantum in the xy plane. Here θ' is simply the angle of scattering.

3. *The Spectral-intensity Distribution Function.*

Before considering the case of scattering of monochromatic X-radiation, we will consider first the more general case when the incident radiation is continuous. Suppose we have such a pencil of radiation confined to a small solid angle $d\omega$ and let I_ν be the intensity per unit frequency range. Let this radiation be incident on an assembly of dn electrons of momentum m_x, m_y, m_z. Let the intensity of radiation scattered in the solid angle $d\omega'$ and frequency range ν' and $\nu' + d\nu'$ be given by

$$R(\nu')\, d\nu'\, d\omega'. \qquad (4)$$

Then it has been shown by Dirac (*loc. cit.*, equation (8)) that

$$R(\nu') = \frac{h^2}{m^2c^3} \,.\, dn \,.\, I_\nu d\omega \frac{\nu' F(a, b)}{\nu m_t}. \qquad (5)$$

Here ν' is to be regarded as a function of g_x', g_y', g_z' and m_x, m_y, m_z, being that frequency of the incident quantum which will be scattered by an (m_x, m_y, m_z) electron into the frequency range ν' to $\nu' + d\nu'$.

In the above equation $F(a, b)$ is a function which depends on the scattering law adopted and a and b the two invariants connected with the scattering process which as well as the initial momentum m_x, m_y, m_z of the electron and g_x, g_y, g_z of the quantum specify the collision.

Now for dn in equation (5), we have to put the Fermi-expression

$$dn = \frac{V}{h^3} \cdot G \cdot \frac{dm_x\, dm_y\, dm_z}{\exp(\Sigma m_x^2/2mkT)/A + 1}, \tag{6}$$

and integrate with respect to m_x, m_y, m_z. In the above equation A is the constant appearing in the Fermi-Dirac statistics. It has different values according as we consider a degenerate or a non-degenerate gas. When the system is non-degenerate A is a small positive quantity and then has the value

$$A = nh^3 \cdot (2\pi mkT)^{-3/2}/G. \tag{7}$$

A degenerate system corresponds to A being a large quantity and in that case

$$\log A = \left(\frac{3n}{4\pi G}\right)^{2/3} \cdot \frac{h^2}{2mkT}. \tag{8}$$

Then by equation (6)

$$R(\nu') = \frac{h^2}{m^2c^3} \cdot \frac{V}{h^3} \cdot G \iiint_{-\infty}^{\infty} \frac{I_\nu d\omega \nu' F}{\nu m_t} \cdot \frac{dm_x\, dm_y\, dm_z}{\exp(\Sigma m_x^2/2mkT)/A + 1} \tag{9}$$

$$= \frac{h^2}{m^2c^3} \cdot \frac{V}{h^3} \cdot G \int_0^\infty I_\nu\, d\omega\, \psi(\nu, \nu')\, d\nu. \tag{10}$$

Where

$$\psi(\nu, \nu') = \iint_{-\infty}^{\infty} \frac{\nu' F}{\nu m_t} \cdot \frac{dm_y\, dm_z}{\exp(\Sigma m_x^2/2mkT)/A + 1} \Big/ \frac{\partial \nu}{\partial m_x}. \tag{11}$$

Where m_x and $\partial\nu/\partial m_x$ are to be evaluated in terms of m_y, m_z and ν by means of equation (3)

$$(6) \quad \frac{\partial m_x}{\partial \nu} = \frac{mc}{\nu'(1 - \cos\theta')} - \frac{h}{c}, \tag{12}$$

and

$$m_x^2 + m_y^2 = \frac{\beta}{\gamma^2}\left[m_y - \frac{K \sin\theta'}{\beta}\right]^2 + \frac{K^2}{\gamma^2}\left(1 - \frac{\sin^2\theta'}{\beta}\right), \tag{13}$$

where

$$\beta = 1 - 2\nu\cos\theta'/\nu' + (\nu/\nu')^2, \quad \gamma = \nu/\nu' - \cos\theta',$$

$$K = -mc(\nu/\nu' - 1) + h\nu(1 - \cos\theta')/c. \tag{14}$$

Then

$$\psi(\nu, \nu') = \iint_{-\infty}^{\infty} \frac{F\left[\dfrac{1}{\nu(1 - \cos\theta')} - \dfrac{h\nu'}{mc^2\nu}\right] dm_y\, dm_z}{\dfrac{1}{A}\exp\left\{\dfrac{\dfrac{\beta}{\gamma^2}\left[m_y - \dfrac{K\sin\theta'}{\beta}\right]^2 + \dfrac{K^2}{\gamma^2} \cdot \left[1 - \dfrac{\sin^2\theta'}{\beta}\right] + m_z^2}{2mkT}\right\} + 1}. \tag{15}$$

Suppose now that the radiation is monochromatic, then the number of quanta scattered between the frequency range ν' and $\nu' + d\nu'$ into the solid angle $d\omega'$ is given by

$$\mathrm{R}(\nu') = \frac{h^2}{m^2c^3} \cdot \frac{\mathrm{V}}{h^3} \cdot \mathrm{G} \cdot \mathrm{I}_\nu \, d\nu \, dw \, \psi(\nu, \nu'). \tag{16}$$

So that the (spectral) distribution of intensity about the primary frequency ν is given by $\psi(\nu, \nu')$, which we shall now evaluate.

$$\psi(\nu, \nu') = 4\mathrm{F}_0 \left[\frac{1}{\nu(1 - \cos\theta')} - \frac{h\nu'}{mc^2\nu}\right] \iint_0^\infty \frac{dm_y \, dm_z}{\frac{1}{\mathrm{B}} \exp\left(\frac{\beta}{\gamma^2} \cdot \frac{m'^2_y + m_z^2}{2mk\mathrm{T}}\right) + 1}, \tag{17}$$

where

$$\left.\begin{aligned} m'_y &= m_y - \frac{\mathrm{K} \sin\theta'}{\beta} \\ \frac{1}{\mathrm{B}} &= \frac{1}{\mathrm{A}} \cdot \exp\left\{\frac{\mathrm{K}^2}{\gamma^2 \cdot 2mk\mathrm{T}} \cdot \left(1 - \frac{\sin^2\theta'}{\beta}\right)\right\} \end{aligned}\right\}. \tag{18}$$

If we introduce the new variables

$$y = m'^2_y \, \beta/\gamma^2 \cdot 2mk\mathrm{T}, \quad z = m_z^2/2mk\mathrm{T}.$$

Then

$$\begin{aligned} \psi(\nu, \nu') &= 2\mathrm{F}_0 \left[\frac{1}{\nu(1 - \cos\theta')} - \frac{h\nu'}{mc^2\nu}\right] \frac{\gamma \cdot mk\mathrm{T}}{\beta^{\frac{1}{2}}} \cdot \iint_0^\infty \frac{y^{-\frac{1}{2}} \cdot z^{-\frac{1}{2}} \cdot dy \, dz}{e^{y+z}/\mathrm{B} + 1} \\ &= 2\mathrm{F}_0 \left[\frac{1}{\nu(1 - \cos\theta')} - \frac{h\nu'}{mc^2\nu}\right] \cdot \frac{\gamma \cdot mk\mathrm{T}}{\beta^{\frac{1}{2}}} \cdot \mathrm{U}_0, \end{aligned} \tag{19}$$

where U_0 is the special case of the general Sommerfeld integral

$$\mathrm{U}_\rho = \frac{1}{\Gamma(\rho + 1)} \cdot \int_0^\infty \frac{u^\rho \, du}{e^u/\mathrm{B} + 1}, \tag{20}$$

which gives for $\rho = 0$*

$$\mathrm{U}_0 = \pi \log(\mathrm{B} + 1). \tag{21}$$

Hence we get our intensity distribution function

$$\psi(\nu, \nu') = 2\mathrm{F}_0 \left[\frac{1}{\nu(1 - \cos\theta')} - \frac{h\nu'}{mc^2\nu}\right] \cdot \frac{\gamma \cdot mk\mathrm{T}\pi}{\beta^{\frac{1}{2}}} \cdot \log(\mathrm{B} + 1). \tag{22}$$

Case I.—If B is large and positive we get

$$\begin{aligned} \psi(\nu, \nu') = 2\mathrm{F}_0 &\left[\frac{1}{\nu(1 - \cos\theta')} \cdot \frac{h\nu'}{mc^2\nu}\right] \cdot \frac{\gamma \cdot mk\mathrm{T}\pi}{\beta^{\frac{1}{2}}} \\ &\times \left[\log \mathrm{A} - \frac{\mathrm{K}^2}{\gamma^2 \cdot 2mk\mathrm{T}} \left(1 - \frac{\sin^2\theta'}{\beta}\right)\right], \end{aligned} \tag{23}$$

the value of log A being given by (8).

* Sommerfeld, 'Z. Physik,' vol. 47, p. 1 (1928), equation (31A).

An approximation of the above equation to an order of accuracy where the Compton-shift is neglected is

$$\psi(\nu, \nu') = \frac{\sqrt{2}\, F_0 \pi m k T}{\nu (1 - \cos\theta')^{\frac{1}{2}}} \cdot \left[\log A - \frac{(\nu' - \nu)^2}{a\nu^2}\right]. \tag{24}$$

where $a = 4kT(1 - \cos\theta')/mc^2$.

Case II.—If B is small due to the smallness of A we get, to the same order of accuracy as (24), the equation

$$\psi(\nu, \nu') = \frac{n}{h^3 \cdot G} \cdot (2\pi m k T)^{-3/2} \cdot \frac{\sqrt{2} F_0 \pi m k T}{\nu (1 - \cos\theta')^{\frac{1}{2}}} \cdot e^{-(\nu' - \nu)^2/a\nu^2}, \tag{25}$$

the one given by Dirac (*loc. cit.*, equation (13)).

Equation (25) gives an exponential distribution of intensity about the primary frequency for the scattered radiation. But equations (23) and (24) indicate that the distribution of intensity of the radiation scattered by a degenerate electron gas does not follow an exponential law but gives a parabolic distribution. This perhaps explains the rather broad structure of the Compton modified radiation.*

4. *The Compton effect.*

It is natural that the distribution of intensity predicted by equation (23) places the maximum peak of intensity at a place where the Compton's theory for a free-stationary electron predicts a line. Remembering that in any case $\nu/\nu' = 1$ the maximum frequency will be at a modified frequency where $K = 0$ where

$$K = -mc(\nu/\nu' - 1) + h\nu(1 - \cos\theta')/c = 0,$$

i.e., where

$$\lambda' - \lambda = h(1 - \cos\theta')/mc, \tag{26}$$

i.e., on an intensity-frequency graph the maximum occurs at a place corresponding to the Compton shift.

5. *The Effect of Temperature.*

If we consider the Compton scattering by an electron-gas, the distribution function of which depends on temperature, it would naturally be expected

* Mr. J. W. Du Mond in a private communication to the author from the California Institute of Technology, Pasadena, has kindly pointed out that the above is the characteristic of the Compton-radiation from conductors. His paper in the May issue of the 'Physical Review' (vol. 33, p. 643) gives experimental details and theoretical calculations as well. He has independently derived the parabolic structure.

that the spectral intensity distribution function in the Compton scattering would also depend on temperature and Dirac's classical expression (25) does indicate this by the explicit appearance of the temperature factor in $\psi(\nu, \nu')$. But if we substitute the value of log A given by (8) in (23) we get

$$\psi(\nu, \nu') = 2F_0\left[\frac{1}{\nu(1-\cos\theta')} - \frac{h\nu'}{mc^2\nu}\right] \cdot \frac{\gamma \, . \, \pi}{\beta^{\frac{1}{2}}} \times \left[\tfrac{1}{3}\left(\frac{6}{\pi}n\right)^{2/3} h^2 - \frac{K^2}{\gamma^2} \cdot \left(1 - \frac{\sin^2\theta'}{\beta}\right)\right], \quad (27)$$

where all the temperature factors have cancelled out. Thus Compton-scattering by an *electron-gas* will not be influenced by temperature. Further the Compton scattering by the bound electrons will also not be influenced by the ranges of temperature available in the laboratory. Thus it appears that the *total* Compton scattering will not be affected by temperature.*

6. *The Effect of a Magnetic-field.*

We will consider the scattering by the conduction electrons only. When the scatterer is placed in a magnetic-field the distribution function for the electrons changes, and in that case the number of electrons in the momentum range m_x, m_y, m_z and $m_x + dm_x$, $m_y + dm_y$, $m_z + dm_z$ is given by the Pauli's expression†

$$dn = \frac{V}{h^3} \cdot \frac{dm_x \, . \, dm_y \, . \, dm_z}{\exp\left(\frac{\varepsilon_m}{kT} + \frac{\Sigma m_x^2}{2m_0 kT}\right)\Big/ A + 1}, \quad (28)$$

where $\varepsilon_m = mg\,\mu_0 H$; where $\mu_0 = -eh/4\pi m_0 c$ = a Bohr magneton, g = the Lande factor, H = the field strength.

For the summation over all the values of the quantum number m from $-j$ to $+j$ we have the relations

$$\left.\begin{aligned} \sum_{m=-j}^{+j} \varepsilon_m &= 0 \\ \sum_{m=-j}^{+j} \varepsilon_m^2 &= \tfrac{1}{3} G \mu^2 H^2 \end{aligned}\right\}. \quad (29)$$

To derive the spectral-intensity distribution function we have to substitute

* Since the writing of the above a report by Jauncey and Bowers has appeared ('Bull. Amer. Phys. Soc.,' vol. 4, p. 26 (1929)) giving experimental observations which support the above conclusions.

† 'Z. Physik,' vol. 41, p. 81 (1927).

(28) instead of (6) in equation (5) and carry out the integration as before. The final result as one can easily see is

$$\psi(\nu, \nu') = 2F_0 \left| \frac{1}{\nu(1-\cos\theta')} - \frac{h\nu'}{mc^2\nu} \right| \cdot \frac{\gamma \cdot mkT \cdot \pi}{\beta^{\frac{1}{2}}}$$

$$\times \left[\log \Lambda + \sum_{m=-j}^{+j} \frac{-\epsilon_m}{kT} - \frac{K^2}{2\gamma^2 \cdot mkT}\left(1 - \frac{\sin^2\theta'}{\beta}\right)\right], \quad (30)$$

which on account of (29) becomes identical with (23). Thus it appears on Pauli's theory of the paramagnetism of an electron-gas that the scattering of such an assembly should not be influenced by the presence of a magnetic field. In this connection mention should be made of an experimental observation of Bothe* where he tried the influence of a magnetic field. The scatterer he used was paraffin, and he tried up the field strengths of the order of 20,000 Γ. But he could detect no influence.

Summary.

In this paper the Compton scattering by an electron-gas on the Fermi-Dirac statistics is considered. The theory predicts a distribution of spectral intensity not exponentially falling off about the maximum but *parabolically*. It places the peak of maximum intensity at a place where the Compton relation $\lambda' - \lambda = h(1 - \cos\theta')/mc$ is satisfied. Further, the theory indicates that there should be no influence of temperature or magnetic field in Compton scattering.

In conclusion the author wishes to express his thanks to Dr. R. H. Fowler, F.R.S., and Mr. N. F. Mott for kindly going through the manuscript and suggesting improvements.

* 'Z. Physik,' vol. 41, p. 872 (1927).

THE MAXIMUM MASS OF IDEAL WHITE DWARFS

By S. CHANDRASEKHAR

ABSTRACT

The theory of the *polytropic gas spheres* in conjunction with the equation of state of a *relativistically degenerate electron-gas* leads to a *unique value for the mass of a star* built on this model. This mass ($=0.91\odot$) is interpreted as representing the upper limit to the mass of an ideal white dwarf.

In a paper appearing in the *Philosophical Magazine*,[1] the author has considered the density of white dwarfs from the point of view of the theory of the polytropic gas spheres, in conjunction with the degenerate non-relativistic form of the Fermi-Dirac statistics. The expression obtained for the density was

$$\rho=2.162\times 10^6\times\left(\frac{M}{\odot}\right)^2 , \tag{1}$$

where $M/\odot$ equals the mass of the star in units of the sun. This formula was found to give a much better agreement with facts than the theory of E. C. Stoner,[2] based also on Fermi-Dirac statistics but on uniform distribution of density in the star which is not quite justifiable.

In this note it is proposed to inquire as to what we are able to get when we use the relativistic form of the Fermi-Dirac statistics for the degenerate case (an approximation applicable if the number of electrons per cubic centimeter is $> 6\times 10^{29}$). The pressure of such a gas is given by (which can be shown to be rigorously true)

$$P=\frac{1}{8}\left(\frac{3}{\pi}\right)^{\frac{1}{3}}\cdot hc\cdot n^{4/3} , \tag{2}$$

where h equals Planck's constant, c equals velocity of light; and as

$$n=\frac{\rho}{\mu H(1+f)} , \tag{3}$$

[1] 11, No. 70, 592, 1931.

[2] *Philosophical Magazine*, 7, 63, 1929.

μ equals the molecular weight, 2.5, for a fully ionized material, H equals the mass of hydrogen atom, and f equals the ratio of number of ions to number of electrons, a factor usually negligible. Or, putting in the numerical values,

$$P = K\rho^{4/3} , \tag{4}$$

where K equals 3.619×10^{14}. We can now immediately apply the theory of polytropic gas spheres for the equation of state given by (4), where for the exponent γ we have

$$\gamma = \frac{4}{3} \text{ or } 1 + \frac{1}{n} = \frac{4}{3} \text{ or } n = 3 .$$

We have therefore the relation[1]

$$\left(\frac{GM}{M'}\right)^2 = \frac{(4K)^3}{4\pi G} ,$$

or

$$\begin{aligned} M &= 1.822 \times 10^{33} , \\ &= .91 \odot \text{(nearly)} . \end{aligned} \tag{5}$$

As we have derived this mass for the star under ideal conditions of extreme degeneracy, we can regard 1.822×10^{33} as the maximum mass of an ideal white dwarf. This can be compared with the earlier estimate of Stoner[2]

$$M_{\text{max}} = 2.2 \times 10^{33} , \tag{6}$$

based again on uniform density distribution. The "agreement" between the accurate working out, based on the theory of the polytropes, and the cruder form of the theory is rather surprising in view of the fact that in the corresponding non-relativistic case the deviations were rather serious.

Trinity College
Cambridge
November 12, 1930

[1] A. S. Eddington, *Internal Constitution of Stars*, p. 83, eq. (57.3.)

[2] *Philosophical Magazine*, 9, 944, 1930.

Some Remarks on the State of Matter in the Interior of Stars.

By **S. Chandrasekhar** (Copenhagen).

With 3 figures. (Received September 28, 1932.)

It is shown that for *all* stars for which the radiation-pressure is greater than a tenth of the total pressure, an appeal to the FERMI-DIRAC statistics to avoid the central singularity which arises in the discussions of the centrally condensed and the collapsed stars cannot be made. The bearing of this result on the possible state of matter in the interior of stars is indicated.

Since the publication of MILNE's memoir on the "Analysis of Stellar Structure" in the Monthly Notices for November 1930[1]) a great deal of work has been done to consider "composite" stellar models. But the following simple considerations seem to have escaped notice and it seems worth while to state them explicitly.

§ 1. The Surfaces of Demarcation. As we approach the centre of a centrally-condensed or a collapsed star, we change over to the equation of state $p = K_1 \varrho^{5/3}$ if the perfect gas law breaks down. If the perfect gas-law breaks down at all, the actual transition from the perfect-gas envelope to the degenerate core must occupy a certain zone, but we could for the sake of convenience consider a definite surface of demarcation defined as the surface at which the two equations of state give the same gas-pressure.

Now in the perfect gas envelope the *total* pressure is given by

$$P = \left[\left(\frac{k}{\mu}\right)^4 \frac{3}{a} \left(\frac{1-\beta}{\beta^4}\right)\right]^{1/3} \varrho^{4/3}, \tag{1}$$

where

$$\beta = 1 - \frac{\varkappa L}{4\pi c G M}, \tag{2}$$

$\varkappa$ = the opacity coefficient, L = luminosity in ergs cm^{-3}, M = mass in grams, k = BOLTZMANN's Constant, μ = molecular weight $= \alpha m_{\mathrm{H}}$ (say), m_{H} = mass of the hydrogen atom. Since for the standard model the gas pressure p is given by

$$p = \beta P, \tag{3}$$

we have

$$p = C \varrho^{4/3}, \tag{4}$$

where

$$C = \left[\left(\frac{k}{\mu}\right)^4 \frac{3}{a} \frac{1-\beta}{\beta}\right]^{1/3} = \frac{2.692 \cdot 10^{15}}{\alpha^{4/3}} \left[\frac{1-\beta}{\beta}\right]^{1/3}. \tag{4'}$$

[1]) Refered to as l. c.

The equation of state in the degenerate zone is

$$p = K_1 \varrho^{5/3}, \tag{5}$$

where

$$K_1 = \frac{1}{20}\left(\frac{3}{\pi}\right)^{2/3} \frac{h^2}{m\,\mu^{5/3}} = \frac{9.890 \cdot 10^{12}}{\alpha^{5/3}}. \tag{6}$$

At the first surface of demarcation which we will call S_1, the density ϱ_1 is given by

$$C\,\varrho_1^{4/3} = K_1\,\varrho_1^{5/3},$$

or

$$\varrho_1 = \left(\frac{C}{K_1}\right)^3. \tag{7}$$

Now, it is well known that the equation of state (5) changes over again into the relativistic-degenerate-equation of state

$$p = K_2 \varrho^{4/3}, \tag{8}$$

where

$$K_2 = \frac{h\,c}{8\,\mu^{4/3}}\left(\frac{3}{\pi}\right)^{1/3} = \frac{1.228 \cdot 10^{15}}{\alpha^{4/3}}. \tag{8'}$$

Hence if *circumstances permit* we have to consider a second surface of demarcation, S_2, where the density ϱ_2 is given by

$$\varrho_2 = \left(\frac{K_2}{K_1}\right)^3. \tag{9}$$

Hence we have *two* surfaces of demarcation if and only if

$$\varrho_2 > \varrho_1,$$

or

$$\left(\frac{K_2}{K_1}\right)^3 > \left(\frac{C}{K_1}\right)^3, \tag{10}$$

i. e. only when [cf. equations (4'), (6), (8')]

$$\frac{1-\beta}{\beta} < \frac{h^3\,c^3\,a}{512\,\pi\,k^4} = 0.1015,$$

or

$$\beta > 0{,}9079. \tag{11}$$

It may be remarked in passing that the above value for β is independent of the assumed molecular weight. *It depends only on the mass, luminosity and opacity in the gaseous envelope.* It is also independent of whether we consider the same opacity for the degenerate zone and the gaseous envelope, or different opacities in the two regions.

§ 2. The meaning of the fundamental inequality (11) is made clear by the following.

In the following graph I plot $\log p$ against $\log \varrho$.

For numerical calculations I use $\alpha = 2$. The straight line ABK represents the equation of state $p = K_1 \varrho^{5/3}$ and BC the equation of state $p = K_2 \varrho^{4/3}$. These two intersect at B where the density is that which corresponds to the *second* surface of demarcation, namely ϱ_2. ABC gives roughly the equation of state of a degenerate gas.

Let us consider a star for which $\beta = 0.98$. By (4) we get

$$\log p = 14.455 + {}^4/_3 \log \varrho. \tag{12}$$

DE represents this equation. It intersects the degenerate equation of state AB, C at E. The point E corresponds to the first surface of demarcation S_1. Hence for all stars for which $\beta = 0.98$, we first traverse a perfect gas envelope with an equation of state represented by DE. Then we traverse a degenerate zone corresponding to EB and finally (if we have not yet reached the centre) a relativistically degenerate zone.

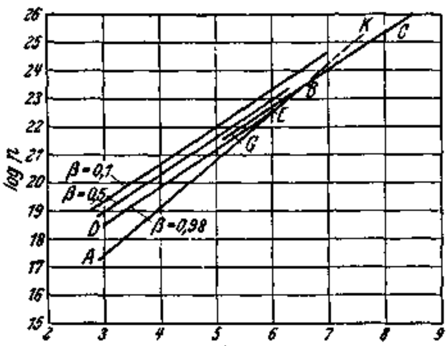


Fig. 1.

Now, if $\beta = 0.9079$, then GB represents the perfect gas equation of state and the degenerate zone reduces to a single layer, and the relativistically degenerate zone is described equally well by the perfect gas equation.

Now if $\beta < 0.9079$[1]) the perfect gas equation of state has *no* intersections with ABC and this means that however high the density may become the temperature rises sufficiently rapidly to prevent the matter from becoming degenerate.

In this connection it will have to be remembered that considerations of relativity do not affect the equation of state of a perfect gas. $p = NkT$, *is true independent of relativity.*

§ 3. Centrally-Condensed Stars. Now, for each mass M there is a unique luminosity L_0 — the "EDDINGTON luminosity" which makes the star

[1]) The radiation pressure is *greater* than a tenth of the total pressure if $\beta < 0.9079$.

a perfect gas sphere, with a polytropic index 3. This L_0 characteristizes a unique β_0 which is in fact related to M by means of EDDINGTON's quartic equation:

$$1 - \beta = 0.003\,09 \left(\frac{M}{\odot}\right)^2 \alpha^4 \beta^4 . \tag{13}$$

Now from the definition of a centrally-condensed and a collapsed star, it is clear that

$$\left.\begin{aligned} \beta_{c \cdot c} &< \beta_0 \\ \beta_{col.} &> \beta_0 . \end{aligned}\right\} \tag{14}$$

Consider first the mass $\mathfrak{M}$ for which $\beta_0 = 0.9079$. By (13) we have

$$\mathfrak{M}/\odot = 6{,}623\, \alpha^{-2}. \tag{15}$$

If we assume $\alpha = 2$,

$$\mathfrak{M}/\odot = 1.656. \tag{15'}$$

Now consider a centrally-condensed star of mass M *greater* than (or equal to) $\mathfrak{M}$. Then we obviously have

$$_M\beta_0 < {}_{\mathfrak{M}}\beta_0 = 0.908.$$

$$_M\beta_{c \cdot c} < {}_{\mathfrak{M}}\beta_0 < 0{,}908. \tag{16}$$

Hence, we have the result that for *all centrally condensed stars of mass greater tham $\mathfrak{M}$, the perfect gas equation of state does not break down, however high the density may become, and the matter does not become degenerate. An appeal to the Fermi-Dirac statistics to avoid the central singularity cannot be made.*

Since however we cannot allow the infinite density which the centrally condensed solution of EMDEN's differential equation — index 3 — allows at the centre and in the absence of our knowledge of any equation of state governing the perfect gas other than that of degenerate matter, our only way out of the singularity is to assume that there exists a maximum density ϱ_{max} which matter is capable of. We have therefore to consider the "fit" of a gaseous envelope of the centrally condensed type on to a homogeneous core at the maximum density of matter. *If we insist on the density to be continuous at the interface* the equation of "fit" is found to be[1])

$$\frac{1}{3}\, \xi' \Theta'^3 = - \left(\frac{d\Theta}{d\xi}\right)_{\xi = \xi'}, \tag{17}$$

[1]) S. CHANDRASEKHAR, M. N. 91, 456, 1931, equation (47).

where the polytropic equation describing the gaseous part of the star is

$$\frac{1}{\xi^2}\frac{d}{d\xi}\left(\xi^2\frac{d\Theta}{d\xi}\right) = -\Theta^3, \tag{17'}$$

where ξ' is the value of ξ at which $\varrho_{\max}$ begins. In (17') the meaning of Θ and ξ are the following:

$$\varrho = \lambda_3\Theta^3, \quad r = \xi\left[\frac{C}{\pi G\beta}\right]^{-1/3}\lambda_3^{-1/3}. \tag{17''}$$

(λ_3 is a *homology constant*). But (17) has *no* solutions if Θ is of the Emden's or of the centrally-condensed type. Hence the acceptance of a $\varrho_{\max}$ does not

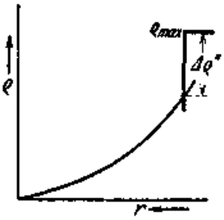


Fig. 2.

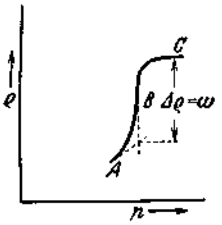


Fig. 3.

help us out of the difficulty if we insist on the density to be continuous at the interface. The procedure then to construct an equilibrium configuration would be to proceed along the centrally condensed solution until the mean density $\varrho_m(r)$ of the *surviving mass* $M(r)$ equals $\varrho_{\max}$ which will occur at a determinate $r = r''$ (say) where

$$M(r'') = {}^4/_3\,\pi r''^3\,\varrho_{\max}; \tag{18}$$

we then replace the material inside $r = r''$ by a sphere of incompressible matter at the density $\varrho_{\max}$. At r'' there will be a discontinuity of density (see Fig. 2).

Now the form of Θ as $\xi \to 0$ for a centrally condensed solution is (Milne, l. c.):

$$\Theta \sim \frac{1/\sqrt{2}}{\xi\,[\log(D/\xi)]^{1/2}}, \tag{19}$$

where D is a constant. D is *fixed* by the condition that the analytic continuation of (19) passes through $\xi = 1$ and $\Theta = 0$ and satisfies here the requisite boundary condition, namely

$$M = -\frac{4}{\pi^{1/2}}\left(\frac{C}{G\beta}\right)^{3/2}\left(\xi^2\frac{d\Theta}{d\xi}\right)_0. \tag{20[1]}$$

[1]) C is given by equation (4').

Hence we get the result that D is a function of L, M and $\varkappa$ only and hence fixed. Since D is fixed by the boundary condition, it follows that the value of ξ'' at which $\Theta(\xi'')$ becomes equal to $\Theta_{\max}$ (where cf. equation (17''))

$$\Theta_{\max} = \varrho_{\max}^{1/3}\, \lambda_3^{-1/3}, \tag{21}$$

is fixed as a function of λ_3. In other words the discontinuity in Θ, $\Delta\Theta''$ at the interface ξ'' is a single-valued function of L, M, $\varkappa$ and λ or

$$\Delta\Theta'' = F(L, M, \varkappa;\ \lambda_3) \tag{21'}$$

or by (21)

$$\Delta\varrho'' = f(L, M, \varkappa;\ \lambda_3) \tag{22}$$

where $\Delta\varrho''$ is the discontinuity of density at the interface.

But it has been suggested by LANDAU[1]) (among others) that the maximum density of matter will arise *after* some kind of *overcompressibility, the incompressibility setting in later* (see Fig. 3).

Further it has been suggested that 1) the pressure at which the overcompressibility sets in must be a physical property of the atomic nuclei and the electrons in the enclosure, and 2) the form of the curve ABC is again an intrinsic physical property of matter. If we idealise the situation of Fig. 3, we see that *$\Delta\varrho$ ought to be a physical property of matter.* Let this $\Delta\varrho$ be ω. Then by (22) we have to so choose the homology constant λ_3, that $\Delta\varrho''$ equals ω:

$$f(L, M, \varkappa;\ \lambda_3) = \omega. \tag{23}$$

This fixes λ_3 and hence by (17'') fixes r_0 — the radius of the configuration. Hence *we are able to obtain equilibrium configurations for arbitrary mass, and arbitrary luminosity, the radius however bieng determinate in each case.*

§ 4. In the above section we have tried to construct the equilibrium configurations for all centrally-condensed stars of mass greater than $\mathfrak{M}$[2]), and found that the introduction of a homogeneous core at the maximum density of matter ($\varrho_{\max}$) with a discontinuity of density at the interface was necessary. We may now ask about the equilibrium configurations for centrally condensed stars with $\beta > 0.908$. Now the star has clearly a degenerate zone (see Fig. 1). A little consideration shows that if we come along a centrally-condensed solution in the perfect gas part of the star then at the interface S_1 (cf. § 1) we are compelled to choose a centrally-condensed solution for the polytropic equation of index "$^3/_2$"

[1]) I am indebted to Dr. STRÖMGREN for advice on these matters.

[2]) Or more generally, centrally-condensed stars with $\beta < 0.908$.

to describe the non-relativistic degenerate part of the star[1]); also at the second surface demarcation S_2 we are again forced to choose a centrally-condensed solution for the potytropic equation of index "3". Hence in this case also we are unable to avoid the central singularity by appealing to the FERMI-DIRAC statistics alone. The star must have a homogeneous core with a discontinuity of density at the interface. The considerations of the previous section apply and we see that the centrally-condensed stars $\beta > 0.908$ differ from the centrally-condensed stars with $\beta < 0.908$ only in this, that while in the former type of stars we have to traverse a degenerate zone before reaching the homogeneous core, in the latter type, the stellar material continues to be a perfect gas till we reach the homogeneous core. Thus we find that *all centrally-condensed stars (on the standard model) must have a homogeneous core at the centre with a discontinuity of density at the interface.*

§ 5. Collapsed-Stars. Just a few remarks about collapsed stars may be permitted. A detailed analysis of *highly-collapsed* stars has been given elsewhere (CHANDRASEKHAR, l. c.).

Consider a collapsed star of mass greater than $\mathfrak{M}$ and let further $\beta_0 < \beta_{\text{col.}} < 0.9078$. In other words the "collapse" has not proceeded sufficiently far to increase β beyond 0.9078. In such a case the collapse can occur only on a homogeneous core. But if the collapse proceeds sufficiently far, such that $\beta_{\text{col.}} > 0.9078$ in spite of β_0 being less than 0,9078, the star will then possess a degenerate zone as well.

Conclusion: We may conclude that great progress in the analysis of stellar structure is not possible before we can answer the following fundamental question:

Given an enclosure containing electrons and atomic nuclei, (total charge zero) what happens if we go on compressing the material indefinitely?

[1]) This is also true if we ascribe different opacities to the gaseous and the degenerate part of the star.

THE HIGHLY COLLAPSED CONFIGURATIONS OF A STELLAR MASS. (SECOND PAPER.)

S. Chandrasekhar, Ph.D.

1. A study of the equilibrium of degenerate gas spheres has a twofold significance in the analysis of stellar structure, namely, in providing an approach to a proper theory of white dwarfs, and also, we shall see, in providing a certain limiting sequence of configurations to which all stars must tend eventually. A beginning in the study of these configurations was made by the author in a previous communication,* where for convenience the equation of state of degenerate matter was taken to correspond to one or other of the two limiting forms $p = K_1\rho^{5/3}$ or $p = K_2\rho^{4/3}$ according as the density was less than or greater than a certain density ρ' where

$$\rho' = (K_2/K_1)^3,$$

ρ' itself being such that both the equations of state yield the same calculated value for the pressure. Actually in the analysis a certain small temperature gradient was allowed for. Working on the standard model it was assumed that the ratio β of the gas pressure to the total pressure was a constant, but by hypothesis ("highly collapsed") β was taken to be very nearly unity. On these assumptions it followed that stars of mass less than a certain specified $M_{3/2}$ (see I, § 6, page 462) were complete Emden polytropes with index $n = 3/2$, and further that configurations of greater mass must be *composite*, *i.e.* must have inner regions where degeneracy is predominantly relativistic. Lastly, and this was the most important conclusion reached, these composite configurations have a *natural limit*: On the standard model a completely relativistically degenerate configuration has a mass given by (*cf.* I, equation (36))

$$M = -\frac{4}{\pi^{1/2}}\left(\frac{K_2}{G}\right)^{3/2}\left(\xi^2\frac{d\theta_3}{d\xi}\right)_1 . \beta^{-3/2} = M_3\beta^{-3/2} \text{ (say)}, \qquad (1)†$$

where θ_3 is the Emden function with index $n = 3$. These configurations have zero radius (*cf.* the remarks in I following the equations (45), (46), page 463).‡

* *M.N.*, 91, 456, 1931 (referred to as I). See also the earlier papers of the author in *Phil. Mag.*, 11, 592, 1931, and *Astrophysical Journal*, 64, 92.

† In I we denoted by M_3 what we have now defined as $M_3\beta^{-3/2}$. It is convenient to separate out the term involving β from the purely "mass factor."

‡ In I this "singularity" was formally avoided by introducing a state of "maximum density" for matter, but now we shall not introduce any such hypothetical states, mainly for the reason that it appears from general considerations that when the central density is high enough for marked deviations from the known gas laws (degenerate or otherwise) to occur the configurations then would have such small radii that they would cease to have any practical importance in astrophysics.

Apart from the above results of a general character, the analysis in I did not lead to any further quantitative results. To obtain by the methods of I anything more exact would have meant very considerable numerical work to "fit" an appropriate solution of Emden's equation with index $n=3/2$ (to describe the outer ordinarily degenerate envelope) with an *Emden function* of index 3 (to describe the inner relativistically degenerate core). It would be very much more satisfactory to take the exact equation describing the degenerate state and treat the whole degenerate parts of a star on the same footing instead of as in I, further subdividing it to correspond to one or other of the two limiting forms of the equation describing the degenerate state. By a very remarkable coincidence the differential equation (governing the structure of a degenerate gas sphere in hydrostatic equilibrium) based on the exact equation of state takes an extremely simple form. We show, in fact, that the structure of the configuration is governed by a solution of the differential equation,

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

It is to be noticed that there is only one parameter occurring in the equation, and a single system of integrations should suffice to obtain a clear insight into these configurations. Equation (2) has a formal similarity with Emden's equation. Indeed, we shall show that under certain circumstances ϕ can be expressed in terms of the Emden functions with appropriate indices. It is the derivation of the above equation that has led to the developments summarised in this and the following paper. In this paper we shall establish this equation and provide tables of solutions. In the analysis we shall omit all references to radiation pressure, *i.e.* this paper strictly deals with configurations having $\beta=1$. The introduction of radiation in these configurations involves quite delicate considerations, and all these find a proper treatment in the paper following this one.

2. *The Differential Equation governing the Structure of Degenerate Matter in Gravitational Equilibrium.*—The pressure-density relation for a degenerate gas can be written parametrically as follows:—

$$\left.\begin{aligned} p&=\frac{\pi m^4c^5}{3h^3}[x(2x^2-3)(x^2+1)^{1/2}+3\sinh^{-1}x],\\ \rho&=\frac{8\pi m^3c^3\mu H}{3h^3}x^3,\end{aligned}\right\} \tag{3}$$

where m = mass of the electron, c = velocity of light, h = Planck's constant, H = mass of the proton, μ = molecular weight. Equation (3) is established in Appendix I to this paper, where also $f(x)$ is tabulated. We rewrite (3) as

$$p=A_2f(x)\,;\qquad \rho=Bx^3, \tag{4}$$

* This equation was given without proof in the author's preliminary note in the *Observatory*, 57, 373, 1934.

where
$$A_2 = \frac{\pi m^4 c^5}{3h^3}; \qquad B = \frac{8\pi m^3 c^3 \mu H}{3h^3}, \qquad \left.\begin{matrix} \\ f(x) = x(2x^2-3)(x^2+1)^{1/2} + 3\sinh^{-1} x. \end{matrix}\right\} \tag{5}$$

The equations of equilibrium are, as usual,
$$\left.\begin{aligned} \frac{dp}{dr} &= -\frac{GM(r)}{r^2}\rho, \\ \frac{dM(r)}{dr} &= 4\pi\rho r^2. \end{aligned}\right\} \tag{6}$$

From (6) we have
$$\frac{1}{r^2}\frac{d}{dr}\left(\frac{r^2}{\rho}\frac{dp}{dr}\right) = -4\pi G\rho. \tag{7}$$

Substitute for p and ρ from (4). We have
$$\frac{A_2}{B}\frac{1}{r^2}\frac{d}{dr}\left(\frac{r^2}{x^3}\frac{df(x)}{dr}\right) = -4\pi G B x^3. \tag{8}$$

From the definition of $f(x)$ in (5) we easily verify that
$$\frac{df(x)}{dr} = \frac{8x^4}{(x^2+1)^{1/2}}\frac{dx}{dr}, \tag{9}$$
or
$$\frac{1}{x^3}\frac{df(x)}{dr} = \frac{8x}{(x^2+1)^{1/2}}\frac{dx}{dr} = 8\frac{d\sqrt{x^2+1}}{dr}. \tag{10}$$

Hence (8) can be rewritten as
$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\sqrt{x^2+1}}{dr}\right) = -\frac{\pi G B^2}{2A_2}x^3. \tag{11}$$

Put
$$y^2 = x^2 + 1. \tag{12}$$

Then
$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dy}{dr}\right) = -\frac{\pi G B^2}{2A_2}(y^2-1)^{3/2}. \tag{13}$$

Let x take the value x_0 at the centre.

Further, let y_0 be the corresponding value of y at the centre. Introduce the new variables η and ϕ defined as follows :—
$$r = a\eta; \qquad y = y_0\phi, \tag{14}$$
where
$$\left.\begin{aligned} a &= \left(\frac{2A_2}{\pi G}\right)^{1/2}\frac{1}{By_0}, \\ y_0^2 &= x_0^2 + 1. \end{aligned}\right\} \tag{15}$$

Our differential equation finally takes the form
$$\frac{1}{\eta^2}\frac{d}{d\eta}\left(\eta^2\frac{d\phi}{d\eta}\right) = -\left(\phi^2 - \frac{1}{y_0^2}\right)^{3/2}. \tag{16}$$

By (14) we have to seek a solution of (16) such that ϕ takes the value unity at the origin. Further, from symmetry the derivative of ϕ must

14

vanish at the origin. The *boundary* is defined at the point where the density vanishes, and this by (12) means that if η_1 specifies the boundary

$$\phi(\eta_1) = \frac{1}{y_0}. \tag{17}$$

3. From our definitions of the various quantities we find that

$$\rho = \rho_0 \frac{y_0^3}{(y_0^2 - 1)^{3/2}} \left(\phi^2 - \frac{1}{y_0^2}\right)^{3/2}, \tag{18}$$

where
$$\rho_0 = Bx_0^3 = B(y_0^2 - 1)^{3/2} \tag{18'}$$

specifies the central density. Also we may notice that the scale of length a introduced in (15) has in terms of the physical quantities the form

$$a = \frac{1}{4\pi m \mu H y_0} \left(\frac{3h^3}{2cG}\right)^{1/2}, \tag{19}$$

or putting in numerical values

$$a = \frac{7{\cdot}720 \times 10^8}{\mu y_0} = l_1 y_0^{-1} \text{ cm. (say)}. \tag{20}$$

4. *The Potential.*—The function ϕ itself has a physical meaning. If V is the inner gravitational potential, then from general theory we have

$$\frac{dV}{dr} = \frac{1}{\rho} \frac{dP}{dr}. \tag{21}$$

From (5) and (10) we see that

$$\frac{dV}{dr} = \frac{8A_2}{B} y_0 \frac{d\phi}{dr}, \tag{22}$$

or integrating
$$V = \frac{8A_2}{B} y_0 \phi + \text{constant}. \tag{23}$$

If we choose the arbitrary zero of the potential on the boundary of the configuration we have by (17) that the "constant" in (23) is $(8A_2/B)$. Hence finally

$$V = \frac{8A_2}{B} y_0 \left(\phi - \frac{1}{y_0}\right). \tag{24}$$

5. *The Mass Relation.*—The mass of the material enclosed up to a point η is clearly

$$M(\eta) = 4\pi \int_0^\eta \rho r^2 dr = 4\pi a^3 \int_0^\eta \rho \eta^2 d\eta. \tag{25}$$

By (18),
$$M(\eta) = 4\pi \rho_0 \frac{a^3 y_0^3}{(y_0^2 - 1)^{3/2}} \int_0^\eta \left(\phi^2 - \frac{1}{y_0^2}\right)^{3/2} \eta^2 d\eta, \tag{26}$$

or using our differential equation (16)

$$M(\eta) = -4\pi \rho_0 \frac{a^3 y_0^3}{(y_0^2 - 1)^{3/2}} \int_0^\eta \frac{d}{d\eta}\left(\eta^2 \frac{d\phi}{d\eta}\right) d\eta. \tag{27}$$

Remembering that ρ_0 is given by (18) we have explicitly

$$M(\eta) = -4\pi\left(\frac{2A_2}{\pi G}\right)^{3/2}\frac{1}{B^2}\eta^2\frac{d\phi}{d\eta}. \tag{28}$$

The mass of the whole configuration is therefore

$$M = -4\pi\left(\frac{2A_2}{\pi G}\right)^{3/2}\frac{1}{B^2}\left(\eta^2\frac{d\phi}{d\eta}\right)_{\eta=\eta_1}. \tag{29}$$

We notice that in (28) and (29) y_0 does not *explicitly* occur. It is of course implicitly present inasmuch as in the differential equation defining ϕ, y_0 occurs.

6. *The Relation between the Mean and the Central Density.*—Let $\bar{\rho}(\eta)$ be the mean density of the material inside η. Then

$$M(\eta) = \tfrac{4}{3}\pi\alpha^3\eta^3\bar{\rho}(\eta). \tag{30}$$

Comparing (28) and (30), we have

$$\tfrac{1}{3}\eta^3\bar{\rho}(\eta) = -\rho_0\frac{y_0^3}{(y_0^2-1)^{3/2}}\eta^2\frac{d\phi}{d\eta}, \tag{31}$$

or

$$\frac{\bar{\rho}(\eta)}{\rho_0} = -3\frac{y_0^3}{(y_0^2-1)^{3/2}}\frac{1}{\eta}\frac{d\phi}{d\eta}. \tag{32}$$

From (32) we deduce that *the relation between the mean and the central density of the whole configuration is*

$$\rho_0 = -\bar{\rho}\left(1-\frac{1}{y_0^2}\right)^{3/2}\frac{\eta_1}{3\phi'(\eta_1)} \tag{33}$$

(ϕ' denoting the derivative)—a relation analogous to the corresponding relation in the theory of polytropes.

7. *An Approximation for Configurations with Small Central Densities.*—When the central density is small we should have the law $p = K_1{}^{5/3}$ holding approximately, and the corresponding configurations must have structures which can approximately be represented by an Emden polytrope with index $n = 3/2$. We establish this on our differential equation in the following way:—

Now by definition $y_0^2 = x_0^2 + 1$, and we need a first-order approximation when x_0^2 is small. *We shall neglect all quantities of order x_0^4 or higher.* Then

$$y_0 = 1 + \tfrac{1}{2}x_0^2. \tag{34}$$

Put

$$\phi^2 - \frac{1}{y_0^2} = \theta. \tag{35}$$

In our approximation we have

$$\phi = 1 - \tfrac{1}{2}(x_0^2 - \theta). \tag{36}$$

At the origin ϕ takes the value unity. Hence

$$\theta(0) = x_0^2. \tag{37}$$

From (16) we derive the following differential equation for θ :—

$$\frac{1}{2}\frac{d^2\theta}{d\eta^2}+\frac{1}{\eta}\frac{d\theta}{d\eta}=-\theta^{3/2}. \tag{38}$$

Put

$$\xi=2^{1/2}\eta. \tag{39}$$

Then

$$\frac{1}{\xi^2}\frac{d}{d\xi}\left(\xi^2\frac{d\theta}{d\xi}\right)=-\theta^{3/2}, \tag{40}$$

which is Emden's equation with index $n=3/2$, but *the solution we need is not the Emden function in the usual normalisation* * *with* $\theta=1$ *at* $\xi=0$. By (37) our θ takes the value x_0^2 at the origin. Denote by $\theta_{3/2}$ the Emden function. Now it is a property of the differential equation (40) that if θ is any solution then $C^4\theta(C\xi)$ is also a solution where C is any arbitrary constant. Hence if we put

$$C=x_0^{1/2}, \tag{41}$$

and take for θ, $\theta_{3/2}$, we would obtain the solution we need. Hence

$$\theta=x_0^2\theta_{3/2}(x_0^{1/2}\xi)=x_0^2\theta_{3/2}(\sqrt{2x_0}\eta). \tag{42}$$

By (37) then

$$\phi=1-\tfrac{1}{2}x_0^2\{1-\theta_{3/2}(\sqrt{2x_0}\eta)\}+O(x_0^4), \tag{43}$$

which relates ϕ with $\theta_{3/2}$. From (43) we see that for these configurations the boundary η_1 must be such that

$$(\theta_{3/2}\sqrt{2x_0}\eta_1)=0. \tag{44}$$

Let $\xi_1(\theta_{3/2})$ be the boundary of the *Emden function*. Then from (44) we deduce that

$$\eta_1=\frac{\xi_1(\theta_{3/2})}{\sqrt{2x_0}}. \tag{45}$$

From (45) we see that as $y_0\to 1$, $x_0\to 0$, $\eta_1\to\infty$. The radius tends to infinity with the same singularity.

Again from (43) we have

$$\frac{d\phi}{d\eta}=\tfrac{1}{2}x_0^2\sqrt{2x_0}\frac{d\theta_{3/2}(\xi)}{d\xi}. \tag{46}$$

Combining (45) and (46) we have a relation we shall need later :

$$\left(\eta^2\frac{d\phi}{d\eta}\right)_1=\left(\frac{x_0}{2}\right)^{3/2}\left(\xi^2\frac{d\theta_{3/2}}{d\xi}\right)_1. \tag{47}$$

Further,

$$\left(\frac{1}{\eta}\frac{d\phi}{d\eta}\right)_1=x_0^3\left(\frac{1}{\xi}\frac{d\theta_{3/2}}{d\xi}\right)_1. \tag{48}$$

We shall find the above expressions useful when we come to discuss "highly"

* In the sequel by "*Emden function*" *we shall always mean the one which takes the value unity at the origin*. We shall denote the *Emden function* with index n by θ_n.

collapsed configurations ($(1-\beta)$ finite but small), but now we verify that the scheme is consistent. From (48) and (33) we have

$$\rho_0 = -\bar{\rho}\left(\frac{\xi}{3\theta'_{3/2}}\right)_1, \tag{49}$$

which is precisely the formula for an Emden polytrope with index $n = 3/2$. Again from (29) and (47)

$$M = -4\pi\left(\frac{2A_2}{\pi G}\right)^{3/2}\frac{1}{B^2}\left(\frac{x_0}{2}\right)^{3/2}\left(\xi^2\frac{d\theta_{3/2}}{d\xi}\right)_1. \tag{50}$$

To compare the above with the formula derived on the law $p = K_1\rho^{5/3}$ we note that the degenerate constant K_1, given by

$$K_1 = \frac{1}{20}\left(\frac{3}{\pi}\right)^{2/3}\frac{h^2}{m(\mu H)^{5/3}}, \tag{51}$$

is related to our A_1 and B by the relation

$$K_1 = \frac{8}{5}\frac{A_2}{B^{5/3}}. \tag{52}$$

Combining (50) and (52) and setting λ_2 to denote the central density ($= Bx_0^3$) we find that

$$M = -4\pi\left(\frac{5K_1}{8\pi G}\right)^{3/2}\lambda_2^{1/2}\left(\xi^2\frac{d\theta_{3/2}}{d\xi}\right)_1, \tag{53}$$

which is the usual formula since on the law $p = K\rho^{1+\frac{1}{n}}$ the polytropic relation is

$$M = -4\pi\left(\frac{(n+1)K}{4\pi G}\right)^{3/2}\lambda_2^{\frac{3-n}{2n}}\left(\xi^2\frac{d\theta_n}{d\xi}\right)_1. \tag{53'}$$

8. *The Limiting Mass.*—From our differential equation (16) we see that

$$\phi \to \theta_3 \quad \text{as} \quad y_0 \to \infty. \tag{54}$$

But from (20) we see that at the same time the radius tends to zero. From (28) then

$$M \to -4\pi\left(\frac{2A_2}{\pi G}\right)^{3/2}\frac{1}{B^2}\left(\xi^2\frac{d\theta_3}{d\xi}\right)_1. \tag{55}$$

To see that we have now simply recovered our earlier result in I (equation (36)) we have only to notice that the relativistic degenerate constant K_2, defined by

$$K_2 = \left(\frac{3}{\pi}\right)^{1/3}\frac{hc}{8(\mu H)^{4/3}}, \tag{56}$$

is related to our A_2 and B by the relation

$$K_2 = \frac{2A_2}{B^{4/3}}. \tag{57}$$

9. As mentioned in § 1, we shall denote by M_3 the mass

$$M_3 = 4\pi\left(\frac{2A_2}{\pi G}\right)^{3/2}\frac{1}{B^2}\omega_3^0, \tag{58}$$

where following Milne we have introduced the quantity $\omega_3{}^0$ defined by

$$\omega_3{}^0 = -\left(\xi^2 \frac{d\theta_3}{d\xi}\right)_1. \tag{59}$$

If we define correspondingly that

$$\Omega(y_0) = -\left(\eta^2 \frac{d\phi}{d\eta}\right)_{\eta=\eta_1} \tag{60}$$

for our function ϕ, then the mass relation can be written as

$$M(y_0)\omega_3{}^0 = M_3\Omega(y_0). \tag{61}$$

As the mass of the configuration increases monotonically with increasing y_0, we have the useful inequality

$$\Omega(y_0) > \omega_3{}^0 \qquad (y_0 \text{ finite}). \tag{62}$$

Finally we may note that the insertion of numerical values in our formula for M_3 yields

$$M_3 = 5{\cdot}728\mu^{-2} \times \odot, \tag{63}$$

where $\odot$ represents the mass of the Sun.

10. *The General Results.*—In the previous sections, §§ 7, 8, 9, we have merely related our present treatment with the results obtained in I on the basis of the polytropic theory. Those results appear as simple limiting cases. However, the exact treatment on the basis of our differential equation

$$\frac{1}{\eta^2}\frac{d}{d\eta}\left(\eta^2 \frac{d\phi}{d\eta}\right) = -\left(\phi^2 - \frac{1}{y_0{}^2}\right)^{3/2} \tag{64}$$

at the same time provides much more quantitative information. The boundary conditions

$$\phi = 1, \qquad \frac{d\phi}{d\eta} = 0 \quad \text{at} \quad \eta = 0, \tag{65}$$

combined with a particular value for y_0, would determine ϕ completely, and therefore the mass of the configuration as well. The equation (64) does not admit of a "homology constant," and hence *each mass has a density distribution characteristic of itself which cannot be inferred from the density distribution in a configuration of a different mass.* This difference between our configurations governed by (64) and polytropes has, as we shall see, an important bearing in the theory of general stellar models considered in the following paper.

Each specified value for y_0 determines uniquely the mass M, the radius R_1 and the ratio of the mean to the central density. We have (collecting together our earlier results):

$$M/M_3 = \Omega(y_0)/\omega_3{}^0, \tag{66}$$

$$R_1/l_1 = \eta_1/y_0, \tag{67}$$

$$\rho_0/B = (y_0{}^2 - 1)^{3/2}, \tag{68}$$

$$\bar{\rho}/\rho_0 = -\frac{1}{\left(1-\frac{1}{y_0^2}\right)^{3/2}}\frac{3}{\eta_1}\left(\frac{d\phi}{d\eta}\right)_1. \tag{69}$$

In (67) we have introduced a new unit of length ($l_1 = \alpha y_0$),

$$l_1 = \frac{1}{4\pi m\mu H}\left(\frac{3h^3}{2cG}\right) = 7\cdot720\mu^{-1}\times 10^8 \text{ cm.}, \tag{67'}$$

and which therefore does not involve factors in y_0. Further, the physical variables determining the structure of the configuration are:

$$\rho = \rho_0\frac{1}{\left(1-\frac{1}{y_0^2}\right)^{3/2}}\left(\phi^2-\frac{1}{y_0^2}\right)^{3/2}, \tag{70}$$

$$\bar{\rho} = -\rho_0\frac{1}{\left(1-\frac{1}{y_0^2}\right)^{3/2}}\frac{3}{\eta}\frac{d\phi}{d\eta}, \tag{71}$$

$$M(\eta) \propto -\eta^2\frac{d\phi}{d\eta}. \tag{72}$$

11. In § 10 we have reduced the problem of the structure of degenerate gas spheres to a study of our functions ϕ for different initially prescribed values for the parameter y_0. The integration has been numerically effected for the following ten different values of the parameter :—

$$1/y_0^2 = 0\cdot8,\ 0\cdot6,\ 0\cdot5,\ 0\cdot4,\ 0\cdot3,\ 0\cdot2,\ 0\cdot1,\ 0\cdot05,\ 0\cdot02,\ 0\cdot01. \tag{73}$$

The following expansion for ϕ near the origin may be noted here for further reference :—

$$\phi = 1 - \frac{q^3}{6}\xi^2 + \frac{q^4}{40}\xi^4 - \frac{q^5(5q^2+14)}{7\,!}\xi^6 + \frac{q^6(339q^2+280)}{3\times 9\,!}\xi^8 - \frac{q^7(1425q^4+11436q^2+4256)}{5\times 11\,!}\xi^{10} + \cdots \tag{74}$$

where

$$q^2 = 1 - \frac{1}{y_0^2}. \tag{75}^*$$

The important quantities of interest are the boundary quantities occurring in equations (66), (67), (69). These are tabulated in Table I for the different values of y_0.

12. From the figures of Table I it is easy to calculate the mass in units of M_3, the radius in units of l_1 and the central density ($=x_0^3$) in units of B

* When $y_0 \to \infty$, $q \to 1$ and the series (74) goes over into the expansion for Emden θ_3 near the origin (cf. *British Association Tables*, 2, Introduction, equation on top of page v).

($=9{\cdot}8848 \times 10^5\ \mu$ grams cm.$^{-3}$). These express the chief physical characteristics of these configurations in the "natural system" of units occurring in the theory of these configurations. In Table III they are converted into the more conventional system of units expressing the radius and the density in C.G.S. units and the mass in units of the Sun. The actual figures tabulated are for $\mu = 1$. The figures for other values of μ can be obtained by multiplying M by μ^{-2}, R_1 by μ^{-1} and ρ by μ. To see the order of magnitudes involved here it is of interest to point out that the mass $4{\cdot}852\odot\mu^{-2}$ has a radius only slightly over the radius of the Earth (radius of the Earth 6×10^8 cm. compared to $7{\cdot}7 \times 10^8$ cm. for the radius of $4{\cdot}852\odot$). The mass $0{\cdot}957 M_3$ has a radius considerably less than the radius of the Earth.

TABLE I

$\frac{1}{y_0^2}$	η_1	$-\eta_1^2\phi'(\eta_1)$	$\rho_0/\bar{\rho}$
0	6·8968	2·0182	54·182
·01	5·3571	1·9321	26·203
·02	4·9857	1·8652	21·486
·05	4·4601	1·7096	16·018
·1	4·0690	1·5186	12·626
·2	3·7271	1·2430	9·9348
·3	3·5803	1·0337	8·6673
·4	3·5245	0·8598	7·8886
·5	3·5330	0·7070	7·3505
·6	3·6038	0·5679	6·9504
·8	4·0446	0·3091	6·3814
1	∞	0	5·9907

TABLE II

The Physical Characteristics of Degenerate Spheres in the "Natural" Units

$\frac{1}{y_0^2}$	M/M_3	R_1/l_1	ρ_0/B
0	1	0	∞
·01	0·95733	0·53571	985·038
·02	0·92419	0·70508	343
·05	0·84709	0·99732	82·8191
·1	0·75243	1·28674	27
·2	0·61589	1·66682	8
·3	0·51218	1·96102	3·56423
·4	0·42600	2·22908	1·83711
·5	0·35033	2·49818	1
·6	0·28137	2·79148	0·54433
·8	0·15316	3·61760	0·125
1·0	0	∞	0

Table III

The Physical Characteristics of Degenerate Spheres in the Usual Units

(Calculations are for $\mu = 1$. For other values μ, M should be multiplied by μ^{-2}, R_1 by μ^{-1}, ρ_c by μ)

$\frac{1}{y_0^2}$	$M/\odot$	ρ_0 in grm./cm.$^{-3}$	ρ_{mean} in grm./cm.$^{-3}$	Radius in cm.
0	5·728	∞	∞	0
·01	5·484	$9{\cdot}737 \times 10^8$	$4{\cdot}716 \times 10^7$	$4{\cdot}136 \times 10^8$
·02	5·294	$3{\cdot}391 \times 10^8$	$1{\cdot}578 \times 10^7$	$5{\cdot}443 \times 10^8$
·05	4·852	$8{\cdot}187 \times 10^7$	$5{\cdot}111 \times 10^6$	$7{\cdot}699 \times 10^8$
·1	4·310	$2{\cdot}669 \times 10^7$	$2{\cdot}114 \times 10^6$	$9{\cdot}936 \times 10^8$
·2	3·528	$7{\cdot}908 \times 10^6$	$7{\cdot}960 \times 10^5$	$1{\cdot}287 \times 10^9$
·3	2·934	$3{\cdot}523 \times 10^6$	$4{\cdot}065 \times 10^5$	$1{\cdot}514 \times 10^9$
·4	2·440	$1{\cdot}816 \times 10^6$	$2{\cdot}302 \times 10^5$	$1{\cdot}721 \times 10^9$
·5	2·007	$9{\cdot}885 \times 10^5$	$1{\cdot}345 \times 10^5$	$1{\cdot}929 \times 10^9$
·6	1·612	$5{\cdot}381 \times 10^5$	$7{\cdot}741 \times 10^4$	$2{\cdot}155 \times 10^9$
·8	0·877	$1{\cdot}236 \times 10^5$	$1{\cdot}936 \times 10^4$	$2{\cdot}793 \times 10^9$
1·0	0	0	0	∞

Now if we define that matter is "relativistically degenerate" for densities greater than $\rho'(=(K_2/K_1)^3)$, then we can from our results easily find the masses which are characterised by central regions of "relativistic degeneracy." The value of x corresponding to ρ' is readily seen to be 1·25. Hence

$$\frac{1}{y_0'^2} = \frac{1}{x'^2 + 1} = 0{\cdot}39024. \tag{76}$$

From fig. 1 we now see that for $M \leqslant 0{\cdot}43 M_3$ there are no regions which are "relativistically degenerate" on this convention. For $M > 0{\cdot}43 M_3$ there are regions in which $x > x'(=1{\cdot}25)$, and the fraction of the whole radius inside which $x > x'$ rapidly increases to unity. In the mass-radius curve we can therefore draw circles about each point with radii proportional to the actual radii of the corresponding configurations, and draw inside each a concentric circle to represent the "relativistic" region. This has been done in fig. 2 at a few points. We see that even for $M = 0{\cdot}75 M_3$ there is barely a "fringe" of ordinarily degenerate regions. This diagram clearly illustrates a general principle that degeneracy never usually sets in without being relativistic.

13. *Comparison with the Results on Emden Polytrope $n = 3/2$.*—It is of interest to see in how far the results of the above exact treatment differ from what one would obtain on the law $p = K_1 \rho^{5/3}$. We have already shown in § 7 that one gets these Emden configurations as limiting cases for zero density and therefore for small masses (expressed in units of M_3). Our comparison here therefore amounts to a comparison of the results based on an exact treatment of the equation (64) with the limiting form for $y_0 \to 1$ extrapolated for all masses. For this purpose it is convenient to rewrite the formulæ for the case of the polytrope $n = 3/2$ in the following way.

From (45) and (50) we have now

$$R_1 = \frac{l_1 \xi_1(\theta_{3/2})}{\sqrt{2x_0}}, \tag{77}$$

$$M/M_3 = \left(\frac{x_0}{2}\right)^{3/2} \frac{1}{\omega_3^0} \left(\xi^2 \frac{d\theta_{3/2}}{d\xi}\right)_1. \tag{77'}$$

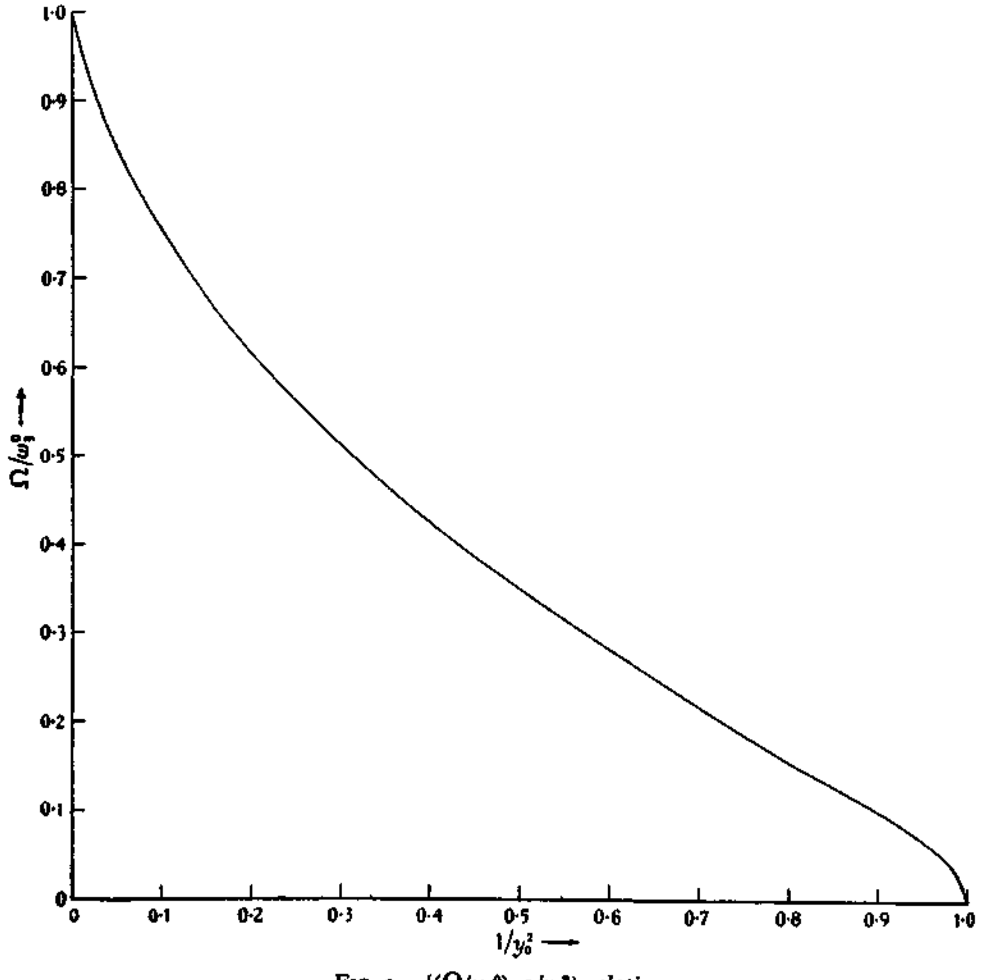


FIG. 1.—$\{(\Omega/\omega_3^0),\ 1/y_0^2\}$-*relation.*

From (77) and (77′) we have on eliminating x_0

$$2R_1 = \left(\frac{\omega_{3/2}{}^0 M_3}{\omega_3{}^0 M}\right)^{1/3} . \, l_1, \tag{78}$$

where following Milne we have introduced the "invariant" $\omega_{3/2}{}^0$ defined by

$$\omega_{3/2}{}^0 = -\left(\xi^5 \frac{d\theta_{3/2}}{d\xi}\right)_1 = 132 \cdot 3843. \tag{79}$$

It is of interest to notice that the two invariants $\omega_3{}^0$ and $\omega_{3/2}{}^0$ of the Emden equation with the indices $n=3$ and $3/2$ occur in (78) in a "symmetrical way." Numerically (78) is found to be

$$R_1 = 2{\cdot}01647\left(\frac{M_3}{M}\right)^{1/3} . l_1. \tag{80}$$

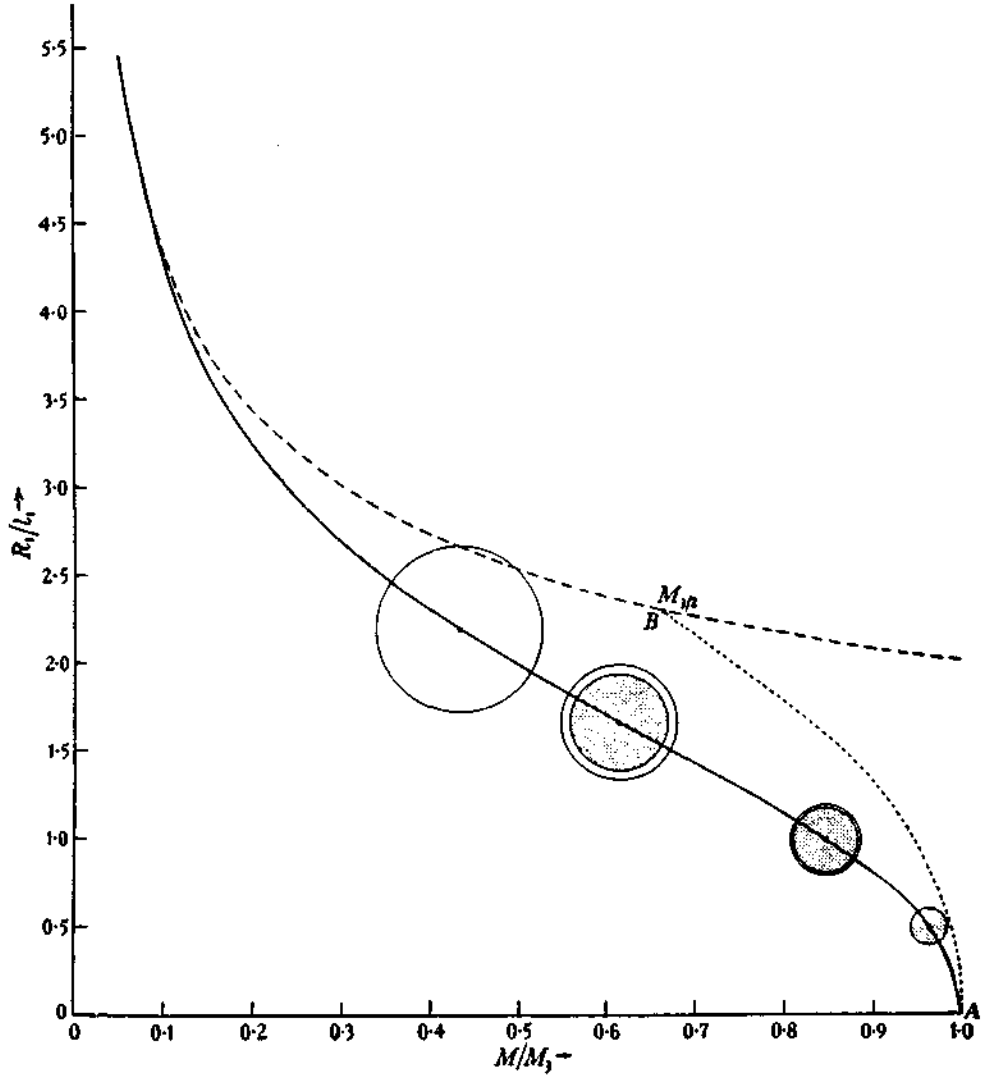


FIG. 2.—*The full line curve represents the exact (mass-radius)-relation for the highly collapsed configurations. This curve tends asymptotically to the - - - - curve as $M \to 0$.*

(80) expresses the *mass-radius* relation for the polytropic limit, the radius and the mass expressed in the same units as the quantities in Table II. Similarly the mass-central density relation now reads

$$x_0{}^3 = 4{\cdot}42381(M/M_3)^2. \tag{81}$$

The results calculated on the basis of (80) and (81) for the same masses as in Table II are summarised in Table IV. The corresponding curves are shown dotted in figs. 2 and 3.

TABLE IV

M/M_3	R_1/l_1	x_0^3
1	2·0165	4·4238
0·9573	2·0459	4·0538
0·9242	2·0700	3·7780
0·8471	2·1311	3·1739
0·7524	2·2174	2·5042
0·6159	2·3701	1·6778
0·5122	2·5203	1·1603
0·4260	2·6801	0·8027
0·3503	2·8606	0·5429
0·2814	3·0772	0·3502
0·1532	3·7691	0·1038

One notices clearly from these two curves how marked the deviations from the limiting curves become even for quite small masses. Thus for $M = 0{\cdot}15M_3$ the central density predicted by our exact treatment is about 25 per cent. greater and the radius about 5 per cent. smaller. The relativistic effects are therefore quite significant even for small masses. They certainly cannot be ignored for masses greater than $0{\cdot}2M_3$. Of course the extrapolation of the $n = 3/2$ configurations for masses (in units of M_3) approaching unity is quite misleading. These completely collapsed configurations have a natural limit, and our exact treatment now shows how this limit is reached.

It is of interest to compare the full-line curve in fig. 2 representing our exact (mass-radius) curve with what one would obtain by the methods of I, where the degenerate spheres of mass greater than a certain limit $M_{3/2}$ were considered as "composite configurations." The mass $M_{3/2}$ was defined as one in which the Emden polytrope with $n = 3/2$ * would have a central density $\rho'(=(K_2/K_1)^3)$. In our present notation we have by (81)

$$M_{3/2} = \sqrt{\frac{(1{\cdot}25)^3}{4{\cdot}42381}}\,.\,M_3 = 0{\cdot}66446M_3. \tag{82}$$

This particular point is marked as B in fig. 2 on the - - - - curve. A treatment of the composite configurations by the methods of I would have led to some kind of curve like the dotted one in fig. 2 conjecturally drawn. But fortunately it is now not necessary to go into the very elaborate numerical work that would have been involved to fix the part BA by the methods of I. By a single system of integrations we have now fixed the exact nature of the (mass-radius) curve for these completely collapsed configurations.

* The equation of state being $p = K_1\rho^{5/3}$.

14. *The Relative Density Distributions in the Different Configurations.*— Our main diagram (fig. 4) now illustrates the relative density distributions in the configurations studied. Here we have plotted (ρ/ρ_0) against (η/η_1) for the different masses for which we have numerical results. The two limiting density distributions specified by Emden, θ_3 and $\theta_{3/2}$, are also shown (dotted) in the same figure. Fig. 4, which is *the* principal outcome of our studies, presents a set of ten out of a continuous family of density distributions

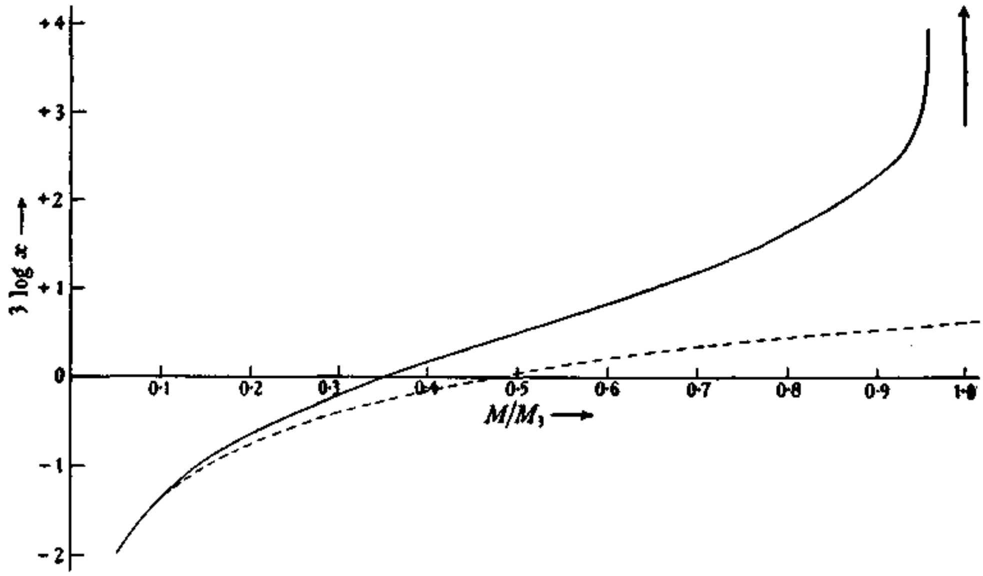


FIG. 3.—*The full line curve represents the exact (mass, log ρ_0)-relation for the highly collapsed configurations. This curve tends asymptotically to the dotted curve as $M \to 0$.*

covering the range specified by the polytropic distributions of indices 3/2 and 3.

15. *Concluding Remarks.*—In this paper we have strictly confined ourselves to the case "$\beta = 1$." But in stellar problem the radiation pressure (even if small) necessarily plays a deciding rôle, and the question as to in what sense we have to understand the completely degenerate spheres studied here as representing "the limiting sequence of configurations to which all stars must tend eventually" can be answered only by introducing radiation in these configurations. To do this properly we have first to develop adequate methods to treat composite configurations consisting of degenerate cores (of the structures studied here) surrounded by gaseous envelopes. These and related problems are studied in the following paper (p. 226).

16. *Manuscript Copy of Tables.*—The functions ϕ and their derivatives ϕ' (to six and five significant figures respectively) have been computed by the author for the values of $1/y_0^2$ specified in (73). In addition to ϕ and ϕ' the auxiliary functions ρ/ρ_0, $\rho_0/\bar{\rho}$, $-\eta^2\phi'$ and two other functions U and V (defined in equation (91) of the following paper) have also been tabulated. The auxiliary functions were calculated correct to five significant figures. All

the functions were tabulated for steps of 0·1 for the argument η. A manuscript copy of these tables has been deposited in the Library of the Society.*

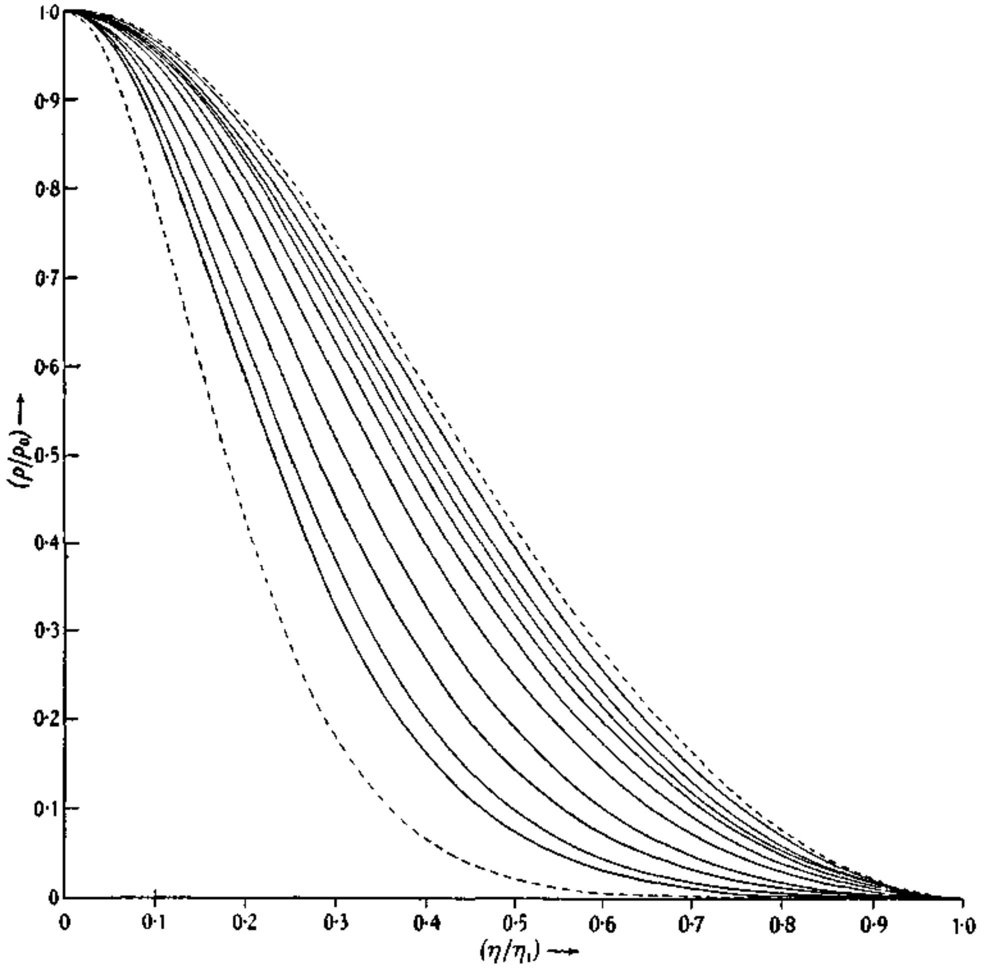


FIG. 4.—*The relative density distributions in the highly collapsed configurations. The upper dotted curve corresponds to the polytropic distribution $n=3/2$ and the lower dotted curve to the polytropic distribution $n=3$. The inner curves represent the density distributions for $1/y_0^2$ = 0·8, 0·6, 0·5, 0·4, 0·3, 0·2, 0·1, ·05, ·02, ·01 respectively.*

APPENDIX

The Equation of State for a Degenerate Gas.—The equation has been derived by Stoner (among others),† but we shall give a simpler derivation of the same.

In a completely degenerate electron assembly all the electrons have momenta less than a certain "threshold" value p_0, and in the region of the

* Dr. Chandrasekhar's Tables can be consulted by Fellows on application to the Assistant Secretary (Editors).

† *M.N.*, 92, 444, 1931.

available phase space of volume $\frac{4}{3}\pi p_0{}^3 V$ every cell of volume h^3 contains just two electrons. Clearly then we have

$$n = \frac{8\pi}{h^3}\int_0^{p_0} p^2 dp, \tag{1*}$$

$$\mathfrak{E} = \frac{8\pi V}{h^3}\int_0^{p_0} E p^2 dp, \tag{2}$$

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

where n is the number of electrons per unit volume in the assembly of volume V, $\mathfrak{E}$ the total energy and E the kinetic energy of a free electron. We have now denoted the pressure by P instead of by "p" as in the text of the paper to avoid confusion with the momentum, which has to be denoted by "p." From (1) and from (2) and (3) we have respectively

$$p_0{}^3 = \frac{3h^3 n}{8\pi}; \qquad P = \frac{8\pi}{3h^3}E(p_0)p_0{}^3 - \frac{\mathfrak{E}}{V}. \tag{4}$$

Equations (1) to (4) are quite general. Now in the relativistic mechanics we have

$$E = mc^2\left\{\left(1 + \frac{p^2}{m^2c^2}\right)^{1/2} - 1\right\}, \tag{5}$$

or

$$p^2 = \frac{E(E + 2mc^2)}{c^2}. \tag{5'}$$

Using (5′) in (3) we have, after some minor transformations, that

$$P = \frac{8\pi m^4 c^5}{3h^3}\int_0^{\theta_0} \sinh^4 \theta d\theta, \tag{6}$$

where

$$\sinh \theta = p/mc; \qquad \sinh \theta_0 = p_0/mc. \tag{7†}$$

(7) yields at once that

$$P = \frac{8\pi m^4 c^5}{3h^3}\left[\frac{\sinh^3 \theta \cosh \theta}{4} - \frac{3}{16}\sinh 2\theta + \frac{3}{8}\theta\right]_{\theta=\theta_0}. \tag{8}$$

Writing x for (p_0/mc) we have

$$P = \frac{\pi m^4 c^5}{3h^3}\left[x(2x^2 - 3)(x^2 + 1)^{1/2} + 3\sinh^{-1} x\right], \tag{9}$$

* This equation follows directly from the expression for the number of waves associated with electrons whose energies lie between E and $E + dE$ given by Dirac (*P.R.S.*, **112**, 660, 1926, his unnumbered equation on p. 671). Actually Dirac obtains this result using the Klein-Gordon relativistic wave equation. That the same result would follow from Dirac's relativistic wave equation (on neglecting the states of kinetic energy—which is permissible when no external perturbations are present) is clear from J. von Neumann, *Z. f. Physik*, **48**, 868, 1928.

† θ here introduced will not be confused with the Emden function.

$$\rho = n\mu H = \frac{8\pi m^3 c^3 \mu H}{3h^3} x^3, \tag{10}$$

which are the equations quoted in the text. Our derivation now shows "why" we are able to reduce the differential equation for degenerate gas spheres in gravitational equilibrium to such a simple form. The "reason" is that we have such an elementary integral for P as in (6).

The function $f(x)$ on the right-hand side of (9) has the following asymptotic forms :—

$$f(x) \sim \tfrac{8}{5}x^5 - \tfrac{4}{7}x^7 + \tfrac{1}{3}x^9 - \tfrac{5}{22}x^{11} + \ . \ . \ . \quad x \to 0, \tag{11}$$

$$f(x) \sim 2x^4 - 3x^2 + \ . \ . \ . \quad x \to \infty\,. \tag{12}$$

Finally we notice that

$$\frac{f(x)}{2x^4} < 1 \quad \text{for all finite } x. \tag{13}$$

The inequality in (13) is a *strict* one. If only the first terms in the expansions (11) and (12) are retained, we can easily eliminate x from (9) and (10) for these limiting cases and obtain, as we should expect, that

$$P = K_1\rho^{5/3} \quad (x \to 0)\,; \qquad P = K_2\rho^{4/3} \quad (x \to \infty\,), \tag{14}$$

with

$$K_1 = \frac{1}{20}\left(\frac{3}{\pi}\right)^{2/3} \frac{h^2}{m(\mu H)^{5/3}}\,; \qquad K_2 = \left(\frac{3}{\pi}\right)^{1/3} \frac{hc}{8(\mu H)^{4/3}}. \tag{15*}$$

If we write our "equation of state" (9) and (10) parametrically as (changing to "p" to denote pressure),

$$p = A_2 f(x)\,; \qquad \rho = Bx^3, \tag{16}$$

we find, on putting in the numerical values for the constants, that (in C.G.S. units)

$$A_2 = 6{\cdot}0406 \times 10^{22}\,; \qquad B = 9{\cdot}8848 \times 10^5\mu, \tag{17}$$

or

$$\left.\begin{aligned} \log\rho &= \ \ 5{\cdot}9950 + 3\log x + \log\mu, \\ \log p &= 22{\cdot}7811 + \log f(x) \end{aligned}\right\}. \tag{18}$$

Stoner has previously made some calculations concerning the (p, ρ) relation for a degenerate gas, but for the study in the following paper more accurate tables for $f(x)$ were needed. Accordingly the whole computation was re-

* The law $P = K_2\rho^{4/3}$ was first used by the author in his paper on "Highly Collapsed Configurations," etc. (*M.N.*, 91, 456, 1931). This law has also been derived by E. C. Stoner (*M.N.*, 92, 444, 1932), T. E. Sterne (*M.N.*, 93, 764, 1933), and is also implicitly contained in J. Frenkel (*Z. f. Physik*, 50, 234, 1928). The law has also been used by L. Landau (*Physik. Zeits. d. Soviet Union*, 1, 285, 1932). It may also be pointed out that the law $P = K_2\rho^{4/3}$ is implicit in certain equations in a paper by F. Juttner (*Z. f. Physik*, 47, 542, 1928, equations in §§ 13, 17 ; our equation (6) above is a limiting form of Juttner's integral $Q(\alpha, \gamma\,; +1)$). This last work of Juttner is related to his earlier work on the relativistic theory of an ideal classical gas, for a convenient summary of which see W. Pauli, *Relativitätstheorie* (Leipzig, Teubner), § 49.

done and the results are tabulated in Table V. I am indebted to Dr. Comrie and Mr. Sadler for the loan of a manuscript copy of a seven-figure table for $\sinh^{-1} x$, which was valuable in the computations of $f(x)$.

TABLE V

x	$f(x)$	$f(x)/2x^4$
0	0	0
0·2	0·000505	0·15785
0·4	0·015527	·30325
0·6	0·111126	·42873
0·8	0·435865	·53206
1·0	1·229907	·61495
1·2	2·82298	·68070
1·4	5·62991	·73276
1·6	10·14696	·77415
1·8	16·94969	·80731
2·0	26·69159	·83411
2·2	40·10347	·85598
2·4	57·99311	·87398
2·6	81·24509	·88894
2·8	110·8207	·90149
3·0	147·7578	·91209
3·5	279·8113	·93232
4·0	484·5644	·94641
4·5	784·5271	·95659
5·0	1205·2069	·96417
6·0	2525·739	·97444
7·0	4710·192	·98088
8·0	8070·587	·98518
9·0	$1·296694 \times 10^4$	·98818
10·0	$1·980725 \times 10^4$	·99036
20·0	$3·192093 \times 10^5$	·99753
30·0	$1·618212 \times 10^6$	·99890
40·0	$5·116812 \times 10^6$	·99938
50·0	$1·249501 \times 10^7$	·99960
60·0	$2·591280 \times 10^7$	·99972
70·0	$4·801018 \times 10^7$	·99980
80·0	$8·190727 \times 10^7$	·99984
90·0	$13·12039 \times 10^7$	·99988
100·0	$19·9980 \times 10^7$	·99990

Trinity College, Cambridge :
1935 *January* 1.

STELLAR CONFIGURATIONS WITH DEGENERATE CORES.

S. Chandrasekhar, Ph.D.

1. When Professor Milne began his investigations on stellar structure the following problem in specific relation to the standard model was in the forefront of his studies. On the hypothesis of a perfect gas * the mass M of the configuration is a single-valued function of β, defining the constant ratio of the gas to the total pressure in the configuration. The relation in question is of course Eddington's quartic equation. Call the appropriate β, β_M. *Has the star equilibrium configurations when* $\beta \neq \beta_M$*?*

Professor Milne himself supplied the first part of the answer. If, for a prescribed mass M, $(1-\beta)$ were greater than $(1-\beta_M)$, then the outer parts of the configuration must be described by *centrally condensed* † singularity possessing solutions of Emden's equation with index 3; on the other hand, if $(1-\beta)$ were less than $(1-\beta_M)$, then the outer parts must be described by *collapsed solutions* of Emden's equation.

It is of course clear that if we agree to describe the outer parts of a configuration by singularity possessing solutions of the differential equations involved, then we must assume that somewhere in the inner regions the perfect gas laws break down. Hence the possibility of the usefulness or otherwise of these centrally condensed and collapsed solutions of Emden's equation ($n=3$) depends essentially upon whether the physics of an ionised gas predicts marked deviations from the perfect gas law under any circumstances. Professor Milne therefore drew attention to the fact that such deviations were predicted by atomic physics, and further pointed out that according to a suggestion originally due to R. H. Fowler such deviations were realised in Nature under the observed white-dwarf conditions.

The above restatement of Milne's problem is of importance in our present discussion, and emphasises that the physics of the situation is the most important consideration. What is meant can be exemplified as follows. Suppose, for instance, that for some prescribed values of β the physical conditions are always of a character that the ideal gas laws do not break down. Then of course the question of using the singular solutions for such configurations does not arise. The general point to realise in this context is that we cannot infer from the existence of singular solutions of the differential equations involved that the gas laws must break down in the inner regions. It is thus more important to examine whether by following these singular solutions we do ever reach physical circumstances where consistent with our knowledge of the equations of state of an ionised gas we have, in fact,

* By *perfect gas* we shall always mean a gas ideal in the classical sense; *i.e.* the corresponding equation of state is $p=(k/\mu H)\rho T$.

† We assume that the reader is generally familiar with the arrangement of the solutions of Emden's differential equation. Otherwise reference should be made to the researches of R. H. Fowler and others.

marked deviations from the ideal gas laws. If there are no deviations at densities, however high (along these singular solutions), then we have simply to abandon the use of them for those particular configurations.

2. The above remarks elaborating Milne's original point of view are necessary, because there exists a large class of stellar configurations for which the use of the singular solutions has to be abandoned precisely for the reasons explained towards the end of the last paragraph. Thus it has already been shown by the author that for stellar configurations in which the radiation pressure continues to be always greater than about a tenth of the total pressure, degeneracy does not set in anywhere.* On the standard model this result has the consequence that for all stars of mass greater than a certain mass $\mathfrak{M}$, Milne's problem has the trivial answer that *there exists no equilibrium configuration which is characterised by a* $\beta \neq \beta_M$ *for* $M \geqslant \mathfrak{M}$, *since consistent with the physics of degenerate matter stars of mass greater than or equal to* $\mathfrak{M}$ *are necessarily wholly gaseous.* This circumstance, however, does not minimise the importance of Milne's problem, for it *has* a non-trivial solution when the configuration has a mass less than $\mathfrak{M}$, and the value of Milne's fundamental problem lies in this, that without his general formulation it could hardly have been possible to analyse the structure of stellar configurations of mass less than $\mathfrak{M}$.

In this paper a first attempt is made to solve Milne's problem consistent with the exact equation of state for degenerate matter. The analysis has been made possible by the derivation in the previous paper (referred to as II)† of the exact differential equation to describe degenerate matter in gravitational equilibrium. The present paper falls into two distinct parts. In the first part we develop certain consequences based on general principles. A systematic treatment of the composite configurations consisting of a gaseous envelope surrounding a degenerate core is undertaken in the second part of the paper. Towards the end the bearing of the results of these studies on the wider problems of stellar evolution is briefly commented upon.

I. *General Considerations*

3. *Equation of State of a Perfect Gas and Degeneracy Conditions.*—Consider material at density ρ and temperature T. The gas pressure according to the perfect gas law is

$$p = \left(\frac{k}{\mu H}\right)\rho T, \tag{1}$$

where k is the Boltzmann constant, μ the molecular weight and H the mass of the proton. At temperature T the radiation pressure p' is, by Stefan's law,

$$p' = \tfrac{1}{3}aT^4. \tag{2}$$

* *Zeit. für Astrophysik*, 5, 321, 1932. The arguments in this paper do not make it sufficiently clear that the result is of a very general character. They are better stated in the author's article in the *Observatory*, 67, 93, 1934—especially the remarks in the footnotes of p. 95.

† The earlier paper (*M.N.*, 91, 456, 1931) will be referred to as I.

Let P denote the total pressure, and let the gas pressure p be a fraction β of the total pressure.* Then

$$P = p + p' = \frac{1}{\beta}p = \frac{1}{1-\beta}p'. \tag{3}$$

Eliminating T in (1) we have

$$p = \left[\left(\frac{k}{\mu H}\right)^4 \frac{3}{a}\frac{1-\beta}{\beta}\right]^{1/3} \rho^{4/3}. \tag{4}$$

Instead of (4) we shall introduce a parametric representation as in II, equations (4), (5). We set

$$\rho = Bx^3; \qquad B = \frac{8\pi m^3 c^3 \mu H}{3h^3}. \tag{5}$$

Substituting (5) in (4) we have

$$p = A_2\left(\frac{512\pi k^4}{h^3c^3a}\frac{1-\beta}{\beta}\right)^{1/3} . \, 2x^4, \tag{6}$$

where A_2 as defined in II is given by

$$A_2 = \frac{\pi m^4 c^5}{3h^3}. \tag{7}$$

In (6) we substitute for a the theoretical expression

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

and obtain the simple expression

$$p = A_2\left(\frac{960}{\pi^4}\frac{1-\beta}{\beta}\right)^{1/3} . \, 2x^4 = 2A_1 x^4 \quad \text{(say)}. \tag{9}$$

It has of course to be understood that equation (9) is merely another form for (1).

Now for material at density ρ and temperature T we can also formally calculate the pressure which would be given by the degenerate formula. With the same definition for ρ as in (6) we have

$$p_{\text{deg}} = A_2 f(x), \tag{10}$$

where as in II, equation (5),

$$f(x) = x(2x^2-3)(x^2+1)^{1/2} + 3\sinh^{-1} x. \tag{11}$$

It is clear that degeneracy for a system with a specified β would set in if for some finite value for x the pressures given by the two formulæ (9) and (10) were equal, *i.e.* if the equation

$$\left(\frac{960}{\pi^4}\frac{1-\beta}{\beta}\right)^{1/3} . \, 2x^4 = f(x) \tag{12}$$

* No implication of β being a *constant* is made here. We are simply defining a parameter to describe material at ρ and T.

is soluble in x. If a solution for (12) exists, then for values of x much smaller than the root of (12) the pressure would be given by (9), and for values of x much larger than the root of (12) the pressure would be given by (10). The criterion for degeneracy then is the following: For a given ρ and T *formally* calculate on the perfect gas law the value for β (as we have done in equation (3)), and with this value for β seek a solution for (12); if a solution exists, and if the value of x at the specified density ρ is much greater than the root, then the system is degenerate; if a solution exists, and if the value of x at the density ρ is much smaller than the root, then the system is a perfect gas in the classical sense. If (12) has no solution for a prescribed β, then the system is *a fortiori* not degenerate. Now we know that (see Appendix to II)

$$\frac{f(x)}{2x^4} < 1 \quad \text{for all finite } x, \tag{13}$$

and

$$\frac{f(x)}{2x^4} \to 1 \quad \text{as} \quad x \to \infty \quad (i.e.\ \rho \to \infty). \tag{13'}$$

Hence if

$$\left(\frac{960}{\pi^4}\frac{1-\beta}{\beta}\right) \geqslant 1, \tag{14}$$

then the system at temperature T and density ρ is necessarily not degenerate. Let β_ω be such that the relation (14) is an equality, *i.e.*

$$\frac{1-\beta_\omega}{\beta_\omega} = \frac{\pi^4}{960} = 0{\cdot}1014678\ \ldots, \tag{15}$$

or

$$1-\beta_\omega = 0{\cdot}09212\ \ldots; \qquad \beta_\omega = 0{\cdot}90788.\ \ldots \tag{16}$$

If for material at density ρ and temperature T the fraction $(1-\beta)$ calculated by means of the equations (1), (2) and (3) is greater than $1-\beta_\omega$, then the system is definitely not degenerate.

For values of β greater than β_ω (12) admits of a solution in finite x, and this value of x would define a convenient measure as to at what densities degeneracy would set in for a system with this prescribed β. We shall adopt in these circumstances the following *approximation for the real equation of state of an ionized gas*:—

$$\left.\begin{aligned} p &= A_2 f(x), \qquad x \geqslant x', \\ \text{and} \qquad p &= 2A_1 x^4, \qquad x \leqslant x', \end{aligned}\right\} \tag{17}$$

x' being such that

$$\left(\frac{960}{\pi^4}\frac{1-\beta}{\beta}\right)^{1/3} = \frac{f(x')}{2x'^4}. \tag{18}$$

When the density is exactly Bx'^3 we shall say that degeneracy is "just beginning to develop."

In Table I the solutions of (18) for different x's are tabulated and graphically illustrated in fig. 1. From this graph we can directly read off the values of x below which we can regard the material as a perfect gas for a calculated β.

4. *The Equations for the Gaseous Envelope.*—These have been set up by various authors from Eddington onwards, but we shall briefly give the

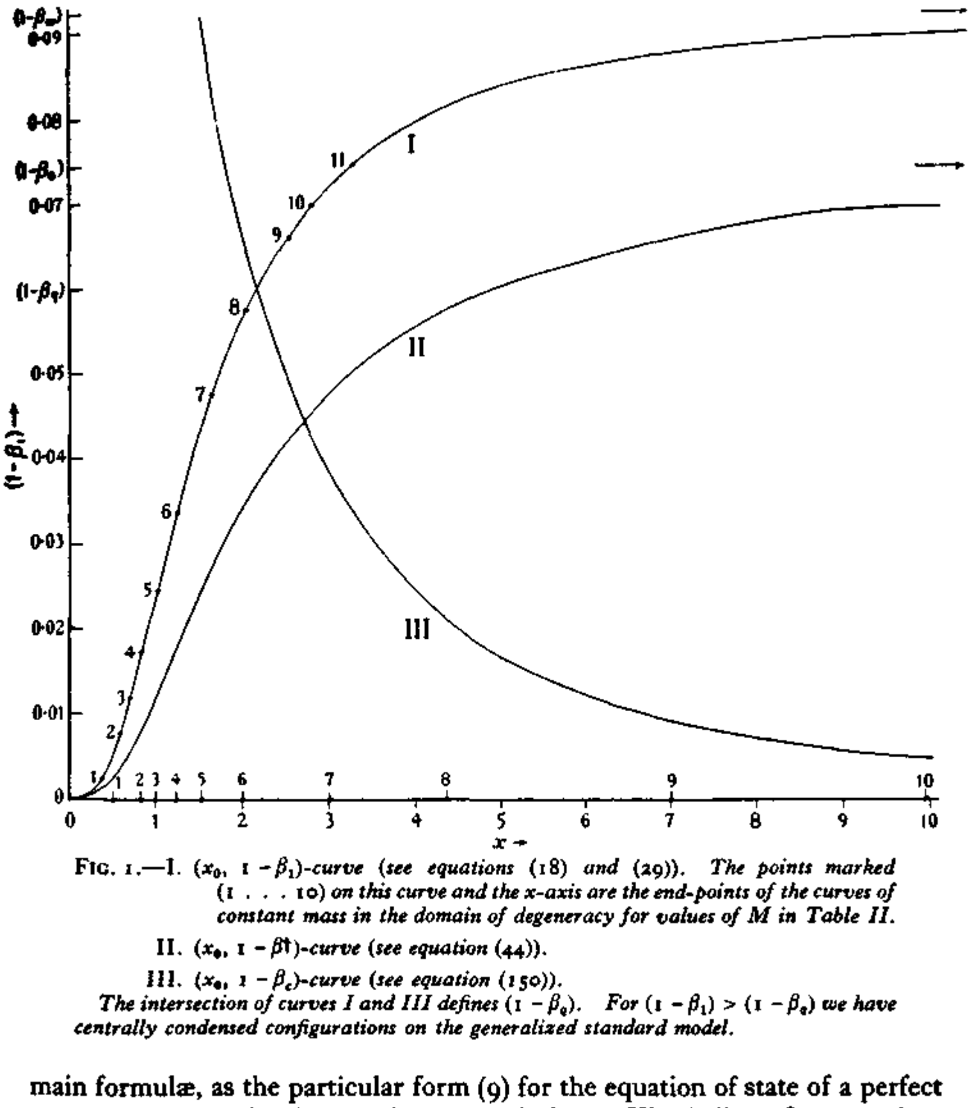


FIG. 1.—I. (x_0, $1-\beta_1$)-*curve (see equations* (18) *and* (29)). *The points marked* (1 . . . 10) *on this curve and the x-axis are the end-points of the curves of constant mass in the domain of degeneracy for values of M in Table II.*

II. (x_*, $1-\beta\dagger$)-*curve (see equation* (44)).

III. (x_*, $1-\beta_c$)-*curve (see equation* (150)).

The intersection of curves I and III defines $(1-\beta_\omega)$. *For* $(1-\beta_1) > (1-\beta_\omega)$ *we have centrally condensed configurations on the generalized standard model.*

main formulæ, as the particular form (9) for the equation of state of a perfect gas we are now using has not been used before. We shall confine ourselves to the standard model. In this model (as also in the "generalised model") for the gaseous portions the ratio $\beta_1 : (1-\beta_1)$ between the gas and the radiation pressure is a constant. Also in a well-known notation

$$\beta_1 = 1 - \frac{(\kappa\eta)_1 L}{4\pi c G M}. \tag{19}$$

The parametric representation for the equation of state in equations (5) and (9) can be used, since "β" is now a constant. The reduction to Emden's equation

$$\frac{1}{\xi^2}\frac{d}{d\xi}\left(\xi^2\frac{d\theta}{d\xi}\right) = -\theta^3, \tag{20}$$

proceeds in the usual way. In our notation the reduction is effected by the following substitutions :—

$$\left.\begin{aligned} x = \lambda_1\theta\,; \qquad r = \alpha_1\xi, \\ \alpha_1 = \left(\frac{2A_1}{\pi G\beta_1}\right)^{1/2}\frac{1}{B\lambda_1}. \end{aligned}\right\} \tag{21}$$

We find also that

$$M(\xi) = -4\pi\left(\frac{2A_1}{\pi G\beta_1}\right)^{3/2}\frac{1}{B^2}\xi^2\frac{d\theta}{d\xi}. \tag{22}$$

Finally it might be recalled that

$$A_1 = A_2\left(\frac{960}{\pi^4}\,\frac{1-\beta_1}{\beta_1}\right)^{1/3}. \tag{23}$$

5. *The Specification of the "Domain of Degeneracy" in the Milne Diagram.*—Equation (22) applied to the boundary gives

$$M = -4\pi\left(\frac{2A_1}{\pi G\beta_1}\right)^{3/2}\frac{1}{B^2}\left(\xi^2\frac{d\theta}{d\xi}\right)_1. \tag{24}$$

If the configuration is wholly gaseous, then we have

$$M = 4\pi\left(\frac{2A_2}{\pi G}\right)^{3/2}\frac{1}{B^2}\left(\frac{960}{\pi^4}\,\frac{1-\beta_1}{\beta_1{}^4}\right)^{1/2}\omega_3{}^0, \tag{25}^*$$

where

$$\omega_3{}^0 = -\left(\xi^2\frac{d\theta_3}{d\xi}\right)_1, \tag{26}$$

and θ_3 is now the *Emden function*. Equation (25) is of course Eddington's quartic equation (in a very different notation), and makes M a function of β_1 only and independent of the central density and hence the radius. In a diagram, therefore, in which we plot the radius R of the configuration against $(1-\beta_1)$ the curves of constant mass are lines parallel to the R axis. Milne's problem formulated in § 1 is to ask *whether the curves of constant mass in any part of this plane are distorted by the physical possibility of degeneracy at the centre.* As Milne was the first to use such a plot, I shall refer to any system of curves of constant mass in the $(R,\ 1-\beta_1)$ plane as a *Milne diagram.*†

Now for a given mass M equation (25) determines a β_1. Start with this mass having an infinite radius and imagine it being slowly contracted. At

* A simpler form for (25) is given in equation (55).

† In his paper in *M.N.*, 91, 4, 1931, Professor Milne has used such a plot in his fig. 4 (p. 47) for the first time. However, his conjectural drawings here (and elsewhere) have to be very considerably modified on the basis of our present analysis.

first the configuration will be so rarefied that it will be wholly gaseous, and the path of the representative point in the $(R, 1-\beta_1)$ plane will be along the line parallel to the R axis through $\beta=\beta_1$. How far is this process of contraction possible? From our arguments in § 3 this is theoretically possible to an unlimited extent if $(1-\beta_1)>(1-\beta_\omega)$. Denote by $\mathfrak{M}$ the mass of the configuration which has a $\beta_1=\beta_\omega$. By (25)

$$\mathfrak{M}=4\pi\left(\frac{2A_2}{\pi G}\right)^{3/2}\frac{1}{B^2}\left(\frac{960}{\pi^4}\frac{1-\beta_\omega}{\beta_\omega^4}\right)^{1/2}\omega_3^0. \tag{27}$$

For configurations with mass greater than $\mathfrak{M}$ the appropriate $(1-\beta_1)$ is greater than $(1-\beta_\omega)$, and hence the "representative point" will travel down an "Eddington line" (by which we mean the line parallel to the R axis through $\beta_1(M)$ on the $(1-\beta_1)$ axis), however far the contraction may proceed. But the situation is different when the mass of the configuration is less than $\mathfrak{M}$. These have a "$(1-\beta_1)$" $<(1-\beta_\omega)$, and hence a stage must come when the configuration should begin to develop central regions of degeneracy. On the scheme of approximation (17) we can now easily see how far we can continue the contraction before degeneracy sets in.

Let the central density be ρ_0. Then

$$\rho_0=Bx_0^3. \tag{28}$$

Degeneracy would "just begin to develop" at the centre for a value of $x=x_0$ such that

$$\frac{f(x_0)}{2x_0^4}=\left(\frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1}\right)^{1/3}. \tag{29}$$

For this configuration the mean density $\bar{\rho}$ is simply

$$\bar{\rho}=-3\left(\frac{1}{\xi}\frac{d\theta_3}{d\xi}\right)_1 Bx_0^3. \tag{30}$$

((30) is just the usual formula expressing the relation between the mean and the central density of a polytrope.) The radius R_0 of the configuration is therefore given by

$$\tfrac{4}{3}\pi R_0^3=\frac{\text{Mass}}{\text{Mean density}}. \tag{31}$$

Substituting in the above the expressions (25) and (29) we obtain

$$R_0=\left(\frac{2A_2}{\pi G}\right)^{1/2}\left(\frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1^4}\right)^{1/6}\frac{1}{Bx_0}\xi_1(\theta_3), \tag{32}$$

where $\xi_1(=6{\cdot}897\ldots)$ defines the boundary of the Emden function θ_3.

Define a unit of length l by (see equation (19), II)

$$l=\left(\frac{2A_2}{\pi G}\right)^{1/2}\frac{\xi_1}{B}=\frac{7{\cdot}720\times 6{\cdot}897\times 10^8}{\mu},$$

or

$$l=5{\cdot}324\times 10^9\mu^{-1}. \tag{33}$$

From (32) then

$$\frac{R_0}{l}=\left(\frac{960}{\pi^4}\,\frac{1-\beta_1}{\beta_1{}^4}\right)^{1/6}\frac{1}{x_0}, \tag{34}$$

where x_0 is again determined from (29), using which we can rewrite (34) more conveniently as

$$\frac{R_0}{l}=\left(\frac{f(x_0)}{2x_0{}^4}\,\frac{1}{\beta_1}\right)^{1/2}\frac{1}{x_0}. \tag{35}$$

It is a fairly simple matter to calculate from (29) and (35) corresponding pairs of values for R_0 and β_1. These are tabulated in Table I. This $(R_0, 1-\beta_1)$ curve can therefore be drawn (see fig. 3). The region bounded by this curve and the two axes then defines the *domain of degeneracy*, meaning that it is only in this region that the Eddington lines are distorted.

TABLE I

x	$1-\beta_1$	R_0/l	$M/\mathfrak{M}$	x	$1-\beta_1$	R_0/l	$M/\mathfrak{M}$
0	0	∞	0	2·8	0·06919	0·3515	0·8245
0·2	0·00040	1·9868	0·0543	3·0	·07149	·3304	·8422
0·4	·00282	1·3787	·1451	3·5	·07598	·2870	·8767
0·6	·00793	1·0956	·2458	4·0	·07920	·2535	·9014
0·8	·01505	0·9187	·3435	4·5	·08158	·2268	·9195
1·0	·02305	·7934	·4320	5·0	·08337	·2051	·9332
1·2	·03101	·6985	·5093	6·0	·08583	·1721	·9520
1·4	·03839	·6235	·5754	7·0	·08739	·1481	·9639
1·6	·04495	·5627	·6313	8·0	·08844	·1299	·9719
1·8	·05068	·5123	·6784	9·0	·08918	·1157	·9776
2·0	·05561	·4699	·7180	10·0	·08972	·1043	·9817
2·2	·05983	·4337	·7515	20·0	·09150	·0524	·9953
2·4	·06344	·4025	·7798	30·0	·09185	·0350	·9979
2·6	·06653	·3753	·8039	∞	·09212	0	1

From (29) and (35) we have that, as $\beta_1 \to \beta_\omega$,

$$x_0 \to \infty, \qquad R_0 \to 0. \tag{36}$$

Hence, as we should expect, the $(R_0, 1-\beta_1)$ curve intersects the $(1-\beta_1)$ axis, where $\beta_1 = \beta_\omega$. One can further prove that the curve in fact cuts the $(1-\beta_1)$ axis *vertically*, but we omit the proof.

Again from (29) we see that as $\beta_1 \to 1$, $x_0 \to 0$. For small x_0 we have in fact (see equation (11) in the Appendix of the previous paper)

$$\frac{f(x_0)}{2x_0{}^4} \sim \frac{4}{5}x_0, \quad (x_0 \to 0). \tag{37}$$

Hence from (35) as $\beta_1 \to 1$,

$$\frac{R_0}{l} \sim \left(\frac{4}{5}\right)^{1/2}\frac{1}{x_0{}^{1/2}}. \tag{38}$$

6. Finally we notice here a "mass relation" which we subsequently need. Required the mass $M(x_0)$ of the configuration which intersects the $(R_0, 1-\beta_1)$ curve when the central density corresponds to some specified x_0. From (25) and (29) we have

$$M(x_0) = -4\pi\left(\frac{2A_2}{\pi G}\right)^{3/2}\frac{1}{B^2}\left(\frac{f(x_0)}{2x_0^4}\frac{1}{\beta_1}\right)^{3/2}\omega_3^0. \tag{39}$$

Comparing this with (27) we have

$$M(x_0) = \mathfrak{M}\left(\frac{f(x_0)}{2x_0^4}\frac{\beta_\omega}{\beta_1}\right)^{3/2}. \tag{40}$$

From (40) we again see that, as $\beta_1 \to \beta_\omega$,

$$x_0 \to \infty, \qquad \frac{f(x_0)}{2x_0^4} \to 1, \qquad M \to \mathfrak{M} \tag{41}$$

—as it should (see fig. 2, where $M(x_0)$ is plotted against x_0, and also the mass of a completely collapsed configuration against x_0, which corresponds to its central density).

7. *The Nature of the Curves of Constant Mass for $M \leqslant M_3$ in the Domain of Degeneracy.*—In § 6 we have shown at what stage a configuration of mass less than $\mathfrak{M}$ (contracting from infinite extension) begins to develop degeneracy at the centre. This happens when the appropriate Eddington line for the specified mass intersects the $(R_0, 1-\beta_1)$ curve. If the contraction continues further the configuration will begin to develop finite degenerate cores, and the major problem is to see how the curves of constant mass run inside the domain of degeneracy. To do this properly we should develop methods to treat composite configurations, but before we proceed to do this we can obtain some features from general principles.

Now in the last paper we have already made an analysis of *completely collapsed configurations*. Each mass (less than M_3) has a certain uniquely determined radius. Thus if the mass under consideration has a central density corresponding to $y=y_0$, then the radius R_1 is given

$$R_1 = a\eta_1 = \left(\frac{2A_2}{\pi G}\right)^{1/2}\frac{\eta_1}{By_0}, \tag{42}$$

where η_1 is the boundary of the corresponding function ϕ satisfying the differential equation (16) II. In terms of the unit of length l (equation (33)) we have

$$\frac{R_1}{l} = \frac{1}{y_0}\frac{\eta_1(\phi(y_0))}{\xi_1(\theta_3)}. \tag{43}$$

These completely collapsed configurations correspond to $\beta_1 = 1$. Hence we know from (43) the point at which the curves of constant mass for $M \leqslant M_3$ must intersect the R axis. Also for any mass we can calculate the value β_1 has in the wholly gaseous state. From equation (29) II and equation (25) we have the relation

$$\left(\frac{960}{\pi^4}\frac{1-\beta^\dagger}{\beta^{\dagger 4}}\right)^{1/2} = \frac{\Omega(y_0)}{\omega_3^0} = \frac{M}{M_3}, \tag{44}$$

where as in II, equation (60),

$$\Omega(y_0) = -\eta^2\left(\frac{d\phi}{d\eta}\right)_1, \tag{45}$$

and $\beta^\dagger$ is the value of β_1 for a wholly gaseous configuration, which in its completely collapsed state has a central density corresponding to $y=y_0$.

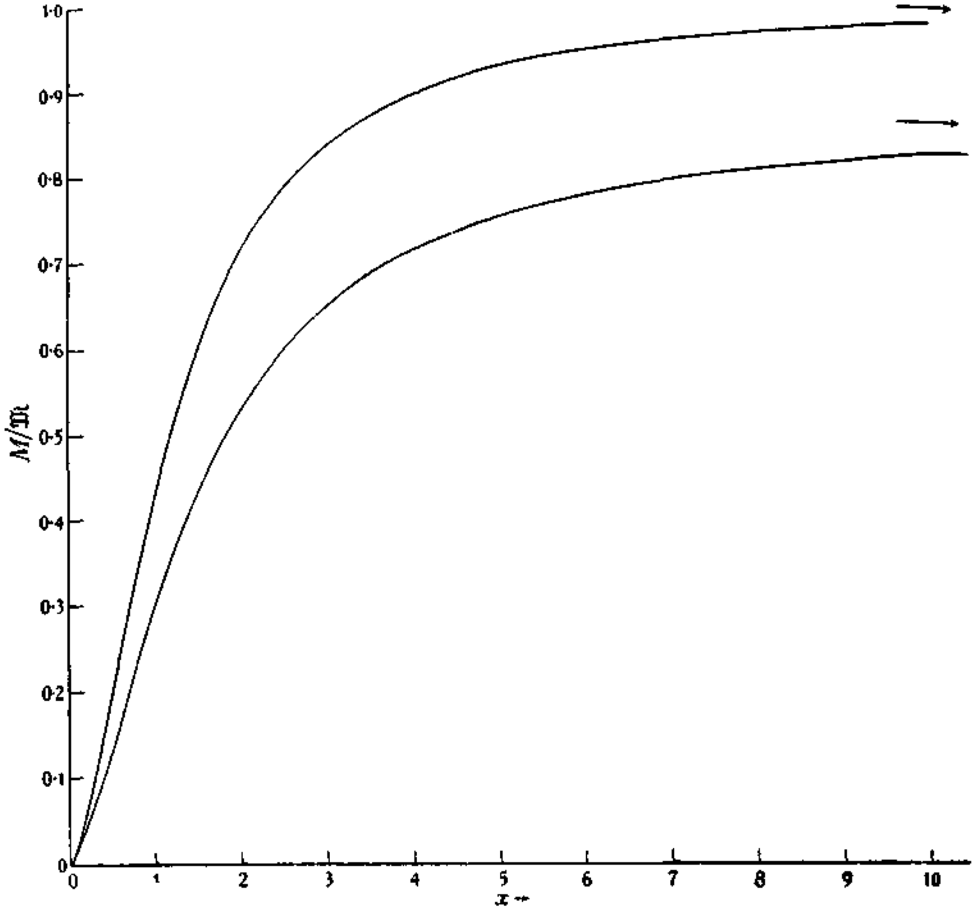


FIG. 2.—*The upper curve illustrates the* (M, x_0)*-relation for configurations in which degeneracy is just developing at the centre. The lower curve illustrates the* (M, x_0)*-relation for completely collapsed configurations.*

Now the line through $\beta^\dagger$ parallel to the R axis will intersect the $(R_0, 1-\beta_1)$ curve at $(R_0(M(y_0)), 1-\beta^\dagger)$. In the domain of degeneracy the continuation of the curve must in some way connect the point $(R_0(M(y_0)), 1-\beta^\dagger)$ and the point R_1 on the R axis, where

$$R_1 = \frac{1}{y_0(M)} \cdot \frac{\eta_1(\phi(y_0(M)))}{\xi_1(\theta_3)}. \tag{46}$$

In the previous paper we have obtained the numerical values for η_1, Ω, etc., for ten different values of y_0, and for the corresponding configurations the values of R_1 and $(1-\beta^\dagger)$ can be numerically evaluated. The results are given in Table II. In fig. 1 the curve $(x_0, 1-\beta^\dagger)$ is drawn. From this curve we can directly read off the value of the central density $(=Bx_0^3)$ in the completely collapsed state for a configuration which in the wholly gaseous state has a given value for $(1-\beta_1)$.

TABLE II

$\frac{1}{y_0^2}$	M/M_3	$(1-\beta^\dagger)\times 10$	R_1/l
0	1	0·7446	0
·01	0·95733	0·6966	·07767
·02	0·92419	0·6596	·10223
·05	0·84709	0·5746	·14460
·1	0·75243	0·4732	·18657
·2	0·61589	0·3358	·24168
·3	0·51218	0·2414	·28434
·4	0·42600	0·1718	·32320
·5	0·35033	0·1187	·36222
·6	0·28137	0·0779	·40475
·8	0·15316	0·0236	·52453
1·0	0	0	∞

We have thus fixed the "end-points" for the curves of constant mass with $M < M_3$ in the domain of degeneracy. The corresponding pairs of points on the $(R_0, 1-\beta_1)$ curve and the R axis are shown in figs. 1 and 3 (the points marked 5 to 15 on the R axis and also on the $(R_0, 1-\beta_1)$ curve). The corresponding points are also marked in fig. 1 on the $(x_0, 1-\beta_1)$ curve and the x axis. Finally we may notice that from (44) and (34) we have

$$\frac{R_0}{l}=\left(\frac{\Omega(y_0)}{\omega_3{}^0}\right)^{1/3}\frac{1}{x_0(\beta^\dagger)}=\left(\frac{M}{M_3}\right)^{1/3}\frac{1}{x_0(\beta^\dagger)}. \tag{47}$$

From (47) and (38) we now have on eliminating x_0 that

$$\frac{R_0}{l}\sim 0{\cdot}8\left(\frac{M_3}{M}\right)^{1/3}, \quad (x_0\to 0,\ \beta_1\to 1). \tag{48}$$

We have already obtained in II, § 13 (equation (80)), a similar relation for the radii of completely collapsed configurations of small mass. In our present units we have

$$\frac{R_1}{l}\sim 0{\cdot}29238\left(\frac{M_3}{M}\right)^{1/3}, \quad (x_0\to 0). \tag{49}$$

For a given $M(<<M_3)$ we have

$$\frac{R_0}{R_1}\to 2{\cdot}7362, \quad (M\to 0). \tag{50}$$

Thus for small masses ("small" when expressed in units of M_3) there is a contraction by a factor of about 3 the configuration in passing from a state in which degeneracy is just beginning to develop at the centre to the completely collapsed state. In the gaseous state the central density is about 54 times the mean density, while in the completely collapsed state the central density is only 6 times the mean density. The net result, however, is an increase in the central density by about a factor 2 (more exactly 2·2650) in passing to the completely collapsed state. These results are only approximately true even for a star of mass $=0{\cdot}1M_3$.

8. *Some Relations for the Mass* M_3.—From the arguments in § 7 it is clear that the curve of constant mass for $M=M_3$ passes through the origin of our system of axes. We shall denote by β_0 the value of β_1 which M_3 has in the wholly gaseous state. From (44) we have

$$\frac{960}{\pi^4}\frac{1-\beta_0}{\beta_0^{\,4}}=1. \tag{51}$$

The numerical solution of (51) is found to be

$$1-\beta_0=0{\cdot}07446\,;\qquad \beta_0=0{\cdot}92554. \tag{51$'$}$$

The asymptote to the $(x_0,\ 1-\beta^\dagger)$ curve drawn in fig. 1 is the line parallel to the x axis through $(1-\beta_0)$. From (40) and (51) we have the following relation between M_3 and $\mathfrak{M}$:—

$$M_3=\mathfrak{M}\beta_\omega^{3/2}=0{\cdot}86505\mathfrak{M}. \tag{52}^*$$

9. *The Nature of the Curves of Constant Mass for* $M>M_3$ *in the Domain of Degeneracy*.—In §§ 7, 8 we have fixed the "end-points" for the curves of constant mass for configurations with mass less than or equal to M_3, and saw further that the appropriate curve for M_3 must pass through the origin. The question now arises, What happens for configurations with $\mathfrak{M}\geqslant M>M_3$? The answer to this question can be given if in the composite configurations "$\kappa\eta$" has the same value in the core as in the gaseous envelope outside. For it was proved in I that on the standard model the completely relativistic configuration has a mass

$$M=M_3\beta^{-3/2}, \tag{53}$$

and is of zero radius (*cf.* I, p. 463; the result is restated in II, § 1). Hence *the curves of constant mass for* $M>M_3$ *must cross the* $(1-\beta_1)$ *axis at a point* $(1-\beta^*)$, *say, such that*

$$M=M_3\beta^{*-3/2}. \tag{54}$$

Let us denote by $\beta^\dagger$ the value of β_1 in the wholly gaseous state. There is a simple relation between $\beta^\dagger$ and β^*. To establish this we first notice that equation (25) can be rewritten in the form

$$M=M_3\left(\frac{960}{\pi^4}\frac{1-\beta^\dagger}{\beta^{\dagger 4}}\right)^{1/2}. \tag{55}$$

* This relation without proof was given in the author's preliminary note in *The Observatory*, 57, 373, 1934.

Equating (54) and (55) we have the relation

$$\beta^* = \left(\frac{\pi^4}{960}\frac{\beta^{\dagger 4}}{1-\beta^\dagger}\right)^{1/3}. \tag{56}$$

We will derive (56) again from our analysis of composite configurations, but we now verify that the relation (56) is precisely what would make our scheme consistent. Thus when $\beta^\dagger = \beta_0$, *i.e.* when $M = M_3$, we have from (50) that $\beta^* = 1$; in other words, the appropriate curve for M_3 must pass through the origin, which in fact it does. Again, when $\beta^\dagger = \beta_\omega$, then (50) yields that $\beta^* = \beta_\omega$; the appropriate curve for $\mathfrak{M}$ is therefore the full line through $(1-\beta_\omega)$ parallel to the R axis, as we should have expected.

The following table gives a set of corresponding pairs of values for β^* and $\beta^\dagger$ (see also fig. 3, where the corresponding pairs of points on the $(R_0, 1-\beta_1)$ curve and the $(1-\beta_1)$ axis are marked 1, 2, 3, 4) :—

Table III

$1-\beta^\dagger$	$1-\beta^*$	$M/\mathfrak{M}$.
0·09212	0·09212	1
0·090	0·08220	0·9838
0·085	0·05768	0·9457
0·080	0·03143	0·9075
0·075	0·00319	0·8692
0·07446	0	0·8651

The above results are of course true only on the usual standard model. The question as to what happens on the *generalised standard model* for configurations of mass greater than M_3 cannot be satisfactorily answered without a proper treatment of composite configurations, to which we proceed now.

II. *The Analysis of Composite Configurations*

In the treatment of these composite configurations on the generalised standard model we shall adopt with suitable modifications the methods which have been developed by Milne.* The main results are summarised in § 21.

10. On the generalised standard model we have the pressure integral (Milne, *loc. cit.*, equation (7))

$$p = \tfrac{1}{3}aT^4\left(\frac{4\pi cGM}{\kappa\eta L} - 1\right) + D, \tag{57}$$

in any interval in which "$\kappa\eta$" is constant.† D changes discontinuously

* E. A. Milne, *M.N.*, **92**, 610, 1932 (referred to as *loc. cit.*). See also T. G. Cowling, *M.N.*, **91**, 472, 1931.

† The "η" used here will not be confused with our variable η, defining the radius vector. "$\kappa\eta$" (in this combination) occurs only in §§ 10, 11.

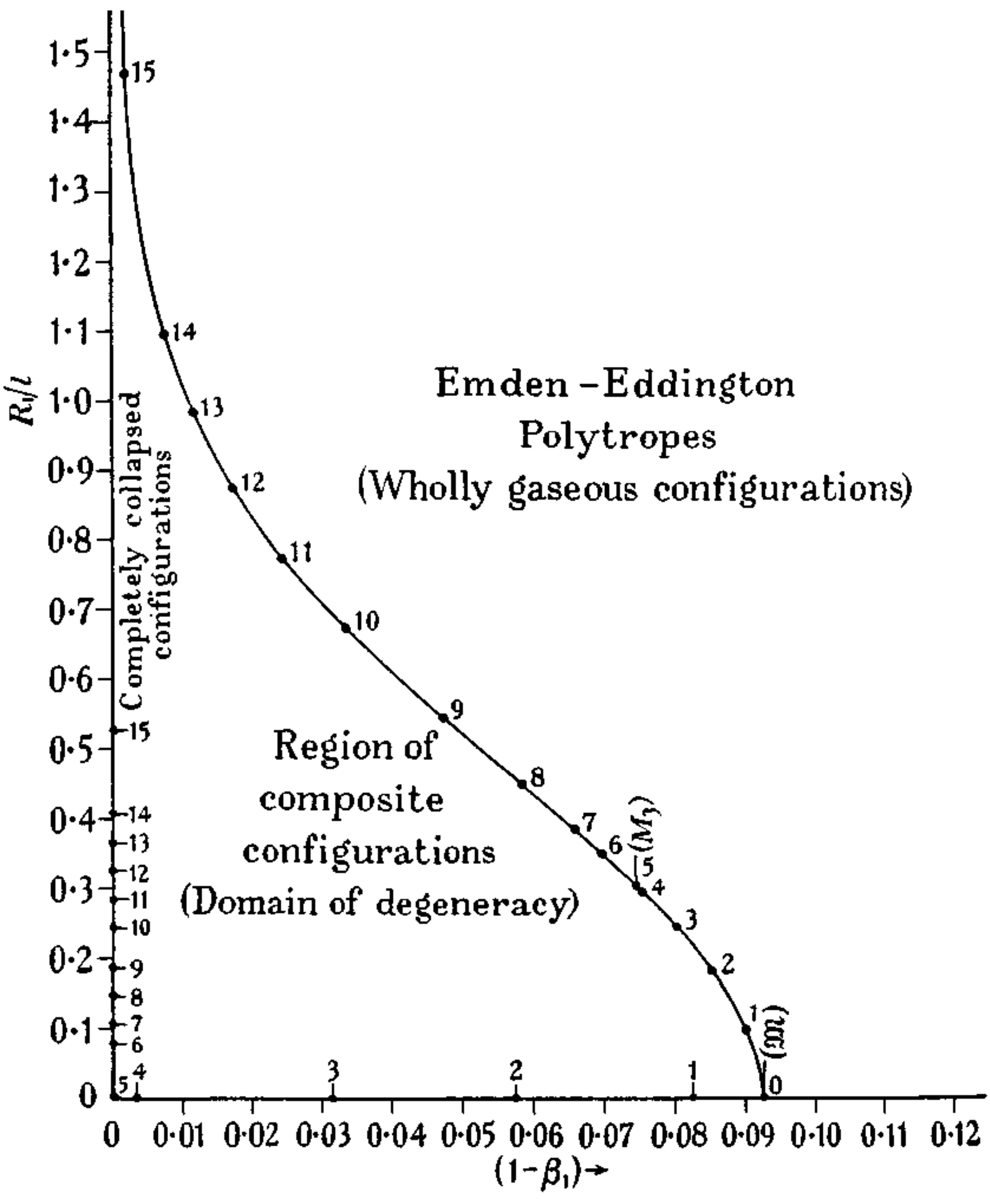


FIG. 3.—*The curve running from* $1-\beta_1=\cdot 092$ . . . *to infinity along the* R_1 *axis is the* $(R_0, 1-\beta_1)$*-curve (see equation (35)). The points marked* (5 . . . 15) *on the* $(R_0, 1-\beta_1)$*-curve and the* R_1 *axis are the end-points (in the domain of degeneracy) for the curves of constant mass for those values of M tabulated in Table II. The points marked* (1 . . . 4) *on the* $(R_0, 1-\beta_1)$*-curve and on the* $(1-\beta_1)$*-axis are the corresponding end-points for some curves of constant mass in the domain of degeneracy on the usual standard model* $(\beta_1=\beta_2)$. *(See Table III.)*

from one constant value to another when "$\kappa\eta$" changes. We shall assume that "$\kappa\eta$" assumes different values in the degenerate core and the gaseous envelope.

Let D_1 be the value of D in the gaseous envelope and D_2 in the degenerate core. In the gaseous envelope the usual external boundary condition $T=0$ when $\rho=0$ leads to D_1 being zero. From (57), then in the gaseous envelope we have the constancy of the ratio of the gas to the radiation pressure. The reduction to Emden's equation follows in the usual way. The relevant equations in our notation have already been given in § 4, equations (21) and (22).

11. *Interfacial Conditions.*—Let the interface occur at a density $\rho=\rho'$ and a temperature $T=T'$ such that at this density and temperature the pressures given by the degenerate and the perfect gas formulæ are the same. Hence if $x=x'$ at the interface, then

$$\frac{f(x')}{2x'^4}=\left(\frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1}\right)^{1/3}, \tag{58}$$

where as in § 4

$$\beta_1=\left(1-\frac{(\kappa\eta)_1 L}{4\pi cGM}\right). \tag{59}$$

Again from (57) we have

$$\frac{1-\beta_1}{\beta_1}p'=\frac{1-\beta_2}{\beta_2}p'-D_2\frac{1-\beta_2}{\beta_2}, \tag{60}$$

where

$$\beta_2=\left(1-\frac{(\kappa\eta)_2 L}{4\pi cGM}\right), \tag{61}$$

and D_2 is the value of "D" in the degenerate core, which must be non-zero if $(\kappa\eta)_2\neq(\kappa\eta)_1$. In (60) we have used p' to denote the gas pressure at the interface. Solving (60) we have

$$D_2=\frac{(\beta_1-\beta_2)}{\beta_1(1-\beta_2)}p'. \tag{62}$$

This value of D_2 ensures the continuity of T at the interface and we need no longer consider the temperature.

12. *The Equations for the Degenerate Core.*—From (57) we now have

$$P=p+p_r=\frac{1}{\beta_2}A_2 f(x)-D_2\frac{1-\beta_2}{\beta_2}, \tag{63}$$

where D_2 is defined by (62), A_2 and $f(x)$ having the same meanings as hitherto.

The reduction to our differential equation (16), II, follows at once. We introduce the variables

$$r=\alpha_2\eta\,; \qquad y=(x^2+1)^{1/2}=y_0\phi, \tag{64}$$

where

$$\left.\begin{aligned}\alpha_2&=\left(\frac{2A_2}{\pi G\beta_2}\right)^{1/2}\frac{1}{By_0},\\ y_0^2&=x_0^2+1,\end{aligned}\right\} \tag{65}$$

and find as before that ϕ satisfies the differential equation

$$\frac{1}{\eta^2}\frac{d}{d\eta}\left(\eta^2\frac{d\phi}{d\eta}\right) = -\left(\phi^2 - \frac{1}{y_0^2}\right)^{3/2}. \tag{66}$$

(It has to be especially noticed that if the mass of a composite configuration is specified then y_0 is one of the quantities that has to be determined and has therefore to be regarded as an unknown.) We have the mass relation

$$M(\eta) = -4\pi\left(\frac{2A_2}{\pi G\beta_2}\right)^{3/2}\frac{1}{B^2}\eta^2\frac{d\phi}{d\eta}. \tag{67}$$

13. *The Equations of Fit.*—Let the interface occur at $\xi = \xi'$ and $\eta = \eta'$. We then have two sets of equations for the absolute value of the radius at which degeneracy sets in (*i.e.* at the interface), and also for the mass of the degenerate core:

$$\left.\begin{aligned} x' &= y_0\left(\phi^2(\eta') - \frac{1}{y_0^2}\right)^{1/2}, \\ r' &= \left(\frac{2A_2}{\pi G\beta_2}\right)^{1/2}\frac{1}{By_0}\eta', \\ M(r') &= -4\pi\left(\frac{2A_2}{\pi G\beta_2}\right)^{3/2}\frac{1}{B^2}\left(\eta^2\frac{d\phi}{d\eta}\right)_{\eta=\eta'}, \end{aligned}\right\} \tag{68}$$

and

$$\left.\begin{aligned} x' &= \lambda_1\theta(\xi'), \\ r' &= \left(\frac{2A_1}{\pi G\beta_1}\right)^{1/2}\frac{1}{B\lambda_1}\xi', \\ M(r') &= -4\pi\left(\frac{2A_1}{\pi G\beta_1}\right)^{3/2}\frac{1}{B^2}\left(\xi^2\frac{d\theta}{d\xi}\right)_{\xi=\xi'}, \end{aligned}\right\} \tag{69}$$

where x' is such that

$$\frac{f(x')}{2x'^4} = \left(\frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1}\right)^{1/3} = \frac{A_1}{A_2}. \tag{70}$$

Equating the different expressions for x', r' and $M(r')$ we have

$$\lambda_1\theta(\xi') = y_0\left(\phi^2(\eta') - \frac{1}{y_0^2}\right)^{1/2}, \tag{71}$$

$$\left(\frac{A_1\beta_2}{A_2\beta_1}\right)^{1/2}\frac{\xi'}{\lambda_1} = \frac{\eta'}{y_0}, \tag{72}$$

$$\left(\frac{A_1\beta_2}{A_2\beta_1}\right)^{3/2}\left(\xi^2\frac{d\theta}{d\xi}\right)_{\xi=\xi'} = \left(\eta^2\frac{d\phi}{d\eta}\right)_{\eta=\eta'}. \tag{73}$$

From (71) and (72) we obtain

$$\left(\frac{A_1\beta_2}{A_2\beta_1}\right)^{1/2}\xi'\theta(\xi') = \eta'\left(\phi^2(\eta') - \frac{1}{y_0^2}\right)^{1/2}. \tag{74}$$

We can eliminate the "homology constant" λ_1 of the Emden equation

16

completely from the discussion by employing an argument due to Milne (*loc. cit.*, p. 617) :—

If $g(\xi)$ is a solution of Emden's equation ($n=3$) vanishing at $\xi=\xi_0$, then $\theta=Cg(C\xi)$ is also a solution which vanishes at ξ_1, where $C\xi_1=\xi_0$. Put $C\xi'=a$. Then

$$\left(\xi^2\frac{d\theta}{d\xi}\right)_{\xi=\xi'}=a^2g'(a)\,;\qquad \xi'\theta(\xi')=ag(a). \tag{75}$$

(A prime to the g's denotes the derivative.) Substituting these in (73) and (74) we have

$$\left(\frac{A_1\beta_2}{A_2\beta_1}\right)^{1/2}ag(a)=\eta'\left(\phi^2(\eta')-\frac{1}{y_0^2}\right)^{1/2}, \tag{76}$$

$$\left(\frac{A_1\beta_2}{A_2\beta_1}\right)^{3/2}a^2g'(a)=\left(\eta^2\frac{d\phi}{d\eta}\right)_{\eta=\eta'}. \tag{77}$$

From the above we obtain the more convenient set:

$$\frac{a[g(a)]^3}{g'(a)}=\frac{\eta'\left(\phi^2(\eta')-\frac{1}{y_0^2}\right)^{3/2}}{\phi'(\eta')}, \tag{78}$$

$$\frac{ag'(a)}{g(a)}=\left(\frac{A_2\beta_1}{A_1\beta_2}\right)\frac{\eta'\phi'(\eta')}{\left(\phi^2(\eta')-\frac{1}{y_0^2}\right)^{1/2}}. \tag{79}$$

The above two equations combined with the following two,

$$x'=y_0\left(\phi^2(\eta')-\frac{1}{y_0^2}\right)^{1/2}, \tag{80}$$

$$\frac{f(x')}{2x'^4}-\left(\frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1}\right)^{1/3}=\frac{A_1}{A_2}, \tag{81}$$

form our complete system of the equations of fit. In his investigations on the "fitting" of one polytropic distribution to another Milne was able to reduce all the equations of fit to just a pair of equations. The circumstance which gave rise to this simplification is the fact that polytropic distributions admit of a constant of "homology" which can always be eliminated (just as we have eliminated λ_1). But our differential equation for ϕ does not admit of a constant of homology,* and we cannot (in principle) simplify the equations (78) to (81) further. We shall presently discuss these equations of fit and outline methods for solving them, but we shall first tabulate the formulæ for the physical characteristics of these configurations.

14. *The Physical Characteristics of the Configurations.*

Radius of the Core—

$$r'=\left(\frac{2A_2}{\pi G}\right)^{1/2}\frac{1}{Bx'}\left(\frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1^4}\right)^{1/6}ag(a). \tag{82}$$

* We have drawn attention to this circumstance in II, § 10.

Mass of the Core—

$$\frac{M(r')}{M(r)}=\frac{a^2g'(a)}{\xi_0^2g'(\xi_0)}. \tag{83}$$

Mean Density of the Core—

$$\rho_m\,(\text{core}) = -3Bx'^3\frac{a^2g'(a)}{[ag(a)]^3}. \tag{84}$$

(Mean Density of Core)/(Interfacial Density)—

$$\frac{\rho_m\,(\text{core})}{\rho'} = -3\frac{a^2g'(a)}{[ag(a)]^3}. \tag{85}$$

Radius of the Configuration—

$$R_1=\left(\frac{2A_2}{\pi G}\right)^{1/2}\frac{1}{Bx'}\left(\frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1^4}\right)^{1/6}\xi_0 g(a). \tag{86}$$

Mean Density of the Configuration—

$$\rho_m = -3Bx'^3\frac{\xi_0^2g'(\xi_0)}{[\xi_0 g(a)]^3}. \tag{87}$$

Effective Temperature—

$$T_e=\left(\frac{L}{ac}\right)^{1/4}\left(\frac{G}{2A_2}\right)^{1/4}\left(\frac{\pi^4}{960}\frac{\beta_1^4}{1-\beta_1}\right)^{1/12}\frac{B^{1/2}x'^{1/2}}{[\xi_0 g(a)]^{1/2}}. \tag{88}$$

Central Density—

$$\rho_0 = B(y_0^2-1)^{1/2}. \tag{89}$$

Some of the preceding formulæ become indeterminate when the relative core radius tends to unity (these configurations are not necessarily completely collapsed), but they can be transformed to determinate forms by using the following relation easily obtained from the equations of fit :—

$$g(a)=\left(\frac{A_2\beta_1}{A_1\beta_2}\right)\frac{\eta'}{a}\left[\frac{1}{a^2g'(a)}\left(\eta^2\frac{d\phi}{d\eta}\right)_{\eta=\eta'}\right]^{1/3}\left(\phi^2(\eta')-\frac{1}{y_0^2}\right)^{1/2}. \tag{79'}$$

The physical variables may then be expressed in the following alternative forms :—

$$r'=\left(\frac{2A_2}{\pi G\beta_2}\right)^{1/2}\frac{1}{By_0}\eta'. \tag{82'}$$

$$R_1=\left(\frac{2A_2}{\pi G\beta_2}\right)^{1/2}\frac{1}{By_0}\left(\frac{\xi_0}{a}\right)\eta'. \tag{84'}$$

$$\rho_m=3By_0^3\frac{M}{M_3}\left(\frac{\beta_2}{\beta_1}\right)^{3/2}\left(\frac{a}{\xi_0}\right)^3\omega_3^0\,.\,\frac{1}{\eta'^3}. \tag{87'}$$

$$T_e=\left(\frac{L}{ac}\right)^{1/4}\left(\frac{G\beta_2}{2A_2}\right)^{1/4}\frac{B^{1/2}y_0^{1/2}}{\eta'^{1/2}}\left(\frac{a}{\xi_0}\right)^{1/2}. \tag{88'}$$

In all the above formulæ we have (following Milne) used only such functions as ξ/ξ_0, $\xi g(\xi)$, $-\xi^2g'(\xi)$, so as to be independent of the "normalisation" to the zero ξ_0.

15. *The General Method to Solve the Equations of Fit.*—Following Milne we define the auxiliary functions $u(\xi)$, $v(\xi)$ associated with a solution of Emden's equation ($n=3$) as follows :—

$$u(\xi) = -\frac{\xi\theta^3}{\theta'}; \qquad v(\xi) = -\frac{\xi\theta'}{\theta}. \tag{90}$$

We shall in a similar way define two auxiliary functions $U(\eta)$, $V(\eta)$ associated with our function ϕ as follows :—

$$U(\eta) = -\frac{\eta\left(\phi^2 - \frac{1}{y_0^2}\right)^{3/2}}{\phi'}; \qquad V(\eta) = -\frac{\eta\phi'}{\left(\phi^2 - \frac{1}{y_0^2}\right)^{1/2}}. \tag{91}$$

The equations of fit then take the form :

$$u(a) = U(\eta'), \tag{92}$$

$$v(a) = \left(\frac{\pi^4}{960}\frac{\beta_1}{1-\beta_1}\right)^{1/3}\left(\frac{\beta_1}{\beta_2}\right)V(\eta'), \tag{92'}$$

$$\frac{f(x')}{2x'^4} = \left(\frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1}\right)^{1/3}, \tag{93}$$

$$x' = y_0\left(\phi^2(\eta') - \frac{1}{y_0^2}\right)^{1/2}. \tag{94}$$

The functions U and V have been tabulated by the author, and Fairclough has tabulated u and v for some solutions of Emden's equation. The general procedure, then, to solve the equations of fit would be the following :—

Consider a sequence of configurations having the same initially prescribed central density, and therefore having a given prescribed value for y_0^2. This means that for this sequence of configurations the degenerate cores are all governed by the same function ϕ. Since the central density is known, and since further the interfacial density must be less than this, it follows that a sequence of configurations having the same central density must be characterised by values of $(1-\beta_1)$ less than a certain critical value $(1-\beta_1(y_0))$, say, such that

$$\frac{f(x_0)}{2x_0^4} = \left(\frac{960}{\pi^4}\frac{1-\beta_1(y_0)}{\beta_1(y_0)}\right)^{1/3}, \quad (x_0^2 = y_0^2 + 1) \tag{95}$$

(we shall sometimes use $(1-\beta_1(x_0))$ instead of $(1-\beta_1(y_0))$).

Now choose a value of β_1 greater than $\beta_1(y_0)$. Equation (93) can be solved for x'. This value of x' would determine the interfacial value of $\eta = \eta'$ from equation (94), and this value of η' would in turn determine $U(\eta')$ and $V(\eta')$. Equations (92) and (92′) finally determine the values of $u(a)$ and $v(a)$. The last stage in the solution of the equations of fit is to determine $\omega_3(=-\xi^2\theta')$ such that along this solution for some value of ξ, u and v have the values already determined. The solution will be *uniquely* determined in this way since u and v are one parameter families of curves (the parameter in fact being ω_3), and the problem simply reduces to one of finding the particular (u, v) curve which passes through a given point in that plane. Once the appropriate (u, v) curve has been determined, the value of "a" is determined at the same time from the "ladder of points" on the (u, v) curve labelling the value of ξ/ξ_0. The value of ω_3 determines also the mass of the configuration, which is given by (*cf.* equation (55))

$$M = M_3\left(\frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1^4}\right)^{1/2}\left(\frac{\omega_3}{\omega_3^0}\right). \tag{96}$$

The physical characteristics now follow from the equations of § 14.

Having constructed in the above way sequences of composite configurations, each sequence being characterised by a different particular value for the central density (but constant along each sequence), we can construct a system of (*mass*, $1-\beta_1$) and (*mass*, *radius*) curves, a pair of curves for each sequence. From these curves (or more accurately by using methods of interpolation) we can then construct the (*mass*, *radius*) curves for different assigned values for $(1-\beta_1)$. From this, the final step of drawing the curves of constant mass in the $(R, 1-\beta)$ diagram follows at once.

The number of solutions of Emden's equation $(n=3)$ that have been integrated so far do not yet provide sufficient material to carry through the above programme with sufficient accuracy, but even a purely analytical discussion of the equations of fit yields information about the different types of composite configurations that exist, and also some general features of the system of curves of constant mass in the $(R, 1-\beta_1)$ diagram.

16. *An Approximate Solution of the Equations of Fit for Highly Collapsed Configurations.*—If the interfacial density is so small that we can regard degeneracy to have set in unrelativistically, then the equations of fit can be solved to a first approximation if in addition the relative core radius is near unity. As the method of obtaining this approximate solution illustrates the general method outlined in § 14, we shall give a short derivation of the same.

Near the zero of ϕ we have the expansion

$$\phi = \frac{1}{y_0} + \frac{\Omega}{\eta_1}(\tau + \tau^2 + \tau^3 + \ . \ . \ .), \tag{97}$$

where

$$\tau = \frac{\eta_1 - \eta}{\eta_1}. \tag{98}$$

From the above we readily find that

$$\phi^2 - \frac{1}{y_0^2} \sim \frac{2\Omega}{y_0\eta_1}\tau, \qquad (\tau \to 0), \tag{99}$$

$$U \sim \left(\frac{2\eta_1}{y_0}\right)^{3/2}\Omega^{1/2}\tau^{3/2}, \qquad (\tau \to 0), \tag{100}$$

$$V \sim \left(\frac{2\eta_1}{y_0}\right)^{-1/2}\Omega^{1/2}\tau^{-1/2}, \quad (\tau \to 0). \tag{101}$$

From (100) and (101) we deduce the following asymptotic relation near $\tau = 0$:—

$$V \sim \frac{\Omega^{2/3}}{U^{1/3}}. \tag{102}$$

Now if degeneracy sets in unrelativistically, i.e. *if* $(1-\beta_1)$ *is very nearly zero,* equations (93) and (94) can be written as

$$\tfrac{4}{5}x' = \left(\frac{960}{\pi^4}\,\frac{1-\beta_1}{\beta_1}\right)^{1/3}, \tag{103}$$

$$x' = y_0\left(\phi^2(\eta') - \frac{1}{y_0^2}\right)^{1/2}. \tag{103'}$$

From (99) and (103) we now obtain (neglecting all quantities of order τ^2 and more)

$$\tau' = \left(\frac{5}{8}\right)^2 \frac{2\eta_1}{\Omega y_0}\left(\frac{960}{\pi^4}\,\frac{1-\beta_1}{\beta_1}\right)^{2/3}, \tag{104}$$

thus determining the place where the interface occurs. At $\tau = \tau'$ we have from (100) and (101) that

$$U(\tau') = \left(\frac{5}{8}\right)^3\left(\frac{2\eta_1}{y_0}\right)^3 \frac{1}{\Omega}\left(\frac{960}{\pi^4}\,\frac{1-\beta_1}{\beta_1}\right), \tag{105}$$

$$V(\tau') = \left(\frac{5}{8}\right)^{-1}\left(\frac{2\eta_1}{y_0}\right)^{-1}\Omega\left(\frac{\pi^4}{960}\,\frac{\beta_1}{1-\beta_1}\right)^{1/3}. \tag{106}$$

The equations of fit (92) and (92′) now determine $u(a)$ and $v(a)$. We have

$$u(a) = \left(\frac{5\eta_1}{4y_0}\right)^3 \frac{1}{\Omega}\left(\frac{960}{\pi^4}\,\frac{1-\beta_1}{\beta_1}\right), \tag{107}$$

$$v(a) = \left(\frac{4y_0}{5\eta_1}\right)\Omega\left(\frac{\pi^4}{960}\,\frac{\beta_1}{1-\beta_1}\right)^{2/3}\left(\frac{\beta_1}{\beta_2}\right). \tag{108}$$

Now for a solution of Emden's equation we have (*cf.* Milne, *loc. cit.*, p. 624)

$$u \sim \omega_3^2 t^3; \qquad v \sim t^{-1}, \tag{109}$$

where

$$t = \frac{\xi_0 - \xi}{\xi_0}. \tag{110}$$

Equations (107), (108) and (109) determine ω_3, and t' at the interface. We find that

$$t' = \left(\frac{5\eta'}{4y_0}\right)\frac{1}{\Omega}\left(\frac{960}{\pi^4}\,\frac{1-\beta_1}{\beta_1}\right)^{2/3}\left(\frac{\beta_2}{\beta_1}\right), \tag{111}$$

$$\omega_3 = \Omega\left(\frac{\pi^4}{960}\,\frac{\beta_1}{1-\beta_1}\right)^{1/2}\left(\frac{\beta_1}{\beta_2}\right)^{3/2}. \tag{112}$$

If we denote

$$\omega_3^2 = C^{-1} \tag{113}$$

—the "discriminant" in Milne's notation—we can rewrite (104) and (111 in the forms

$$t' = C^{2/3}\left(\frac{5}{8}\frac{\beta_1}{\beta_2}\right)\left(\frac{2\eta_1\Omega^{1/3}}{y_0}\right); \qquad \tau' = C^{2/3}\left(\frac{5}{8}\frac{\beta_1}{\beta_2}\right)^2\left(\frac{2\eta_1\Omega^{1/3}}{y_0}\right) \tag{114}$$

or

$$a = \xi_0\left[1 - C^{2/3}\left(\frac{5}{8}\frac{\beta_1}{\beta_2}\right)\left(\frac{2\eta_1\Omega^{1/3}}{y_0}\right)\right], \tag{115}$$

$$\eta' = \eta_1\left[1 - C^{2/3}\left(\frac{5}{8}\frac{\beta_1}{\beta_2}\right)^2\left(\frac{2\eta_1\Omega^{1/3}}{y_0}\right)\right]. \tag{115'}$$

The above solutions will be useful to evaluate the thicknesses of gaseous envelopes in white dwarfs, but these applications are reserved for future communications.

It is of interest to see that with *further restrictions* our result (115), (115′) goes over into certain of Milne's formulæ.

If the mass of the configuration is small (*expressed in units of* M_3), then we have (II, equations (45) and (48))

$$\eta_1 = \frac{\xi_1(\theta_{3/2})}{\sqrt{2x_0}}, \tag{116}$$

$$\Omega = \left(\frac{x_0}{2}\right)^{3/2}\left(-\xi^2\frac{d\theta_{3/2}}{d\xi}\right)_1. \tag{117}$$

Using these we see that

$$\left(\frac{2\eta_1}{y_0}\Omega^{1/3}\right)^3 \to \left(-\xi^5\frac{d\theta_{3/2}}{d\xi}\right)_1 = \omega_{3/2}, \tag{118}$$

the "invariant" for the Emden equation $n = 3/2$. With this substitution in (115), (115′) we have Milne's formulæ (*loc. cit.*, equations (38), (39)). But our derivation of his results shows that his formulæ are valid only under the following conditions: (1) $(1-\beta_1)$ *is very small*, and (2) *the central density of the configuration is also very small.* From our results in II, § 13, it follows then that his formulæ give a fair approximation only for M less than about a tenth of M_3, but even then it can be regarded only as an approximation of an approximation. If one considers formally the fitting of an Emden solution of $n = 3$ to an Emden function of index $n = 3/2$, then the corresponding formulæ (which are Milne's formulæ (38) and (39)) formally give solutions (a fact pointed out by Milne) also when $\beta_1 \sim 0$. Actually, however, if $\beta_1 \sim 0$ the second terms in the brackets of (115) and (115′) are small, but our earlier approximation (103) which has led to these formulæ will no longer be valid—indeed in our analysis there exist no composite configurations for $\beta_1 < \beta_\omega (= \cdot 908)$, and the formal solutions which Milne's formulæ (38) and (39) predict for $\beta \sim 0$ have clearly no physical meaning.*

17. We can deduce from our approximate formulæ of § 16 some results regarding how the curves of constant mass ($M \leqslant M_3$) intersect the x axis

* See, however, his remarks on top of p. 622 (*loc. cit.*).

in the $(x, 1-\beta_1)$ diagram. From (96) and (112) we have for the mass of the configuration

$$M = M_3\left(\frac{\Omega}{\omega_3^0}\right)\beta_2^{-3/2}. \tag{119}$$

But $M_3(\Omega/\omega_3^0)$ is the mass of the *completely* collapsed configuration (*cf.* II, equation (61)), which has the same central density as our composite configuration. Hence when the curves of constant mass intersect the x (or the R) axis we have the relation

$$M(0, x_0) = M(1-\beta_1, x_0)\beta_2^{-3/2}. \tag{120}$$

If $\beta_1 = \beta_2$ (*i.e.* in the usual standard model), we have

$$M(0, x_0) = M(1-\beta_1, x_0)\beta_1^{-3/2}. \tag{121}$$

From this one can readily show that *the curves of constant mass intersect the x axis in the $(x, 1-\beta_1)$ diagram with negative slope which tends to zero both when $M \to 0$ and $M \to M_3$. Hence there exists a mass-curve which intersects the x axis with a maximum negative slope.*

On the other hand, if $\beta_2 = 1$ (the extreme case in the generalised standard model), (121) leads to

$$M(0, x_0) = M(1-\beta_1, x_0). \tag{121'}$$

From (121') it follows that *curves of constant mass for all $M \leqslant M_3$ intersect the x axis in the $(x, 1-\beta_1)$ diagram at right angles.* From this it also follows that for $M = M_3$ the appropriate curve passes through the origin, cutting the R axis at right angles. We shall show in § 20 that for $M = M_3$ the whole segment of the $(1-\beta_1)$ axis from zero to $1-\beta_\omega$ is a part of its curve of constant mass.

18. *A General Discussion of the Equations of Fit and the Types of Configurations that Exist.*—The general disposition of the (u, v) curves for the solutions of Emden's equations is known from Milne's work and Fairclough's integrations.* The most important feature is that the Emden ω_3^0 curve divides the two families of curves, the collapsed curves ($\omega_3 > \omega_3^0$) all lying entirely above the ω_3^0 curve, and the centrally condensed curves all lying entirely below the ω_3^0 curve. The ω_3^0 curve itself runs from $(u = 3, v = 0)$ to $(u = 0, v = \infty)$. For all types for $u \sim 0$, $v \sim \infty$ the asymptotic form of the curve is (*cf.* equation (109))

$$v \sim \frac{\omega_3^{2/3}}{u^{1/3}}. \tag{122}$$

Further, the initial negative slope for the ω_3^0 curve is easily found to be 5/9.

Now the (U, V) curves on the other hand *all* run from $(U = 3, V = 0)$ to infinity along the V axis, the asymptotic form being (*cf.* equation (102))

$$V \sim \frac{\Omega^{2/3}}{U^{1/3}}. \tag{123}$$

* *M.N.*, 93, 40, 1932.

Also the initial negative slope is readily found from the relations (see equation (74), II)

$$U \sim 3\left(1 - \frac{q}{5}\eta^2\right), \tag{124}$$

$$V \sim \tfrac{1}{3}q^2\eta^2, \tag{125}$$

where as in II

$$q = \left(1 - \frac{1}{y_0^2}\right)^{1/2} = \frac{x_0}{(x_0^2+1)^{1/2}}. \tag{126}$$

We have then

$$\frac{dV}{dU} \sim -\frac{5}{9}q = -\frac{5}{9}\frac{x_0}{(x_0^2+1)^{1/2}}. \tag{127}$$

From (123) and (127) we see that all the (U, V) curves lie entirely below the Emden $\omega_3{}^0$ curve.

From the above relations we can infer the different types of configurations (*e.g.* collapsed, centrally condensed, quasi-diffuse) that will occur in our theory. We shall consider only two cases: (*a*) the usual standard model with $\beta_1 = \beta_2$, and (*b*) the extreme case of the generalised standard model with $\beta_2 = 1$. The intermediate cases can be treated similarly but will not be considered here.

19. *The Usual Standard Model* ($\beta_1 = \beta_2$).—We will denote by $\Gamma(\beta_1, y_0)$ the curve

$$\Gamma(\beta_1, y_0) = \left\{U, \left(\frac{\pi^4}{960}\frac{\beta_1}{1-\beta_1}\right)^{1/3} V\right\}. \tag{128}$$

We shall also denote it by $\Gamma(\beta_1, x_0)$, x_0 being the corresponding value for x when $y = y_0$. For a given y_0 the $\Gamma(\beta_1, y_0)$ curves are defined for only values of $(1 - \beta_1)$ less than a critical value $(1 - \beta_1(y_0))$ depending on y_0 alone. The initial negative slope of a $\Gamma(\beta_1, y_0)$ curve is

$$\frac{5}{9}\frac{2x'^4}{f(x')}\frac{x_0}{(x_0^2+1)^{1/2}}.$$

Now

$$\frac{2x'^4}{f(x')}\frac{x_0}{(x_0^2+1)^{1/2}} \geqslant \frac{2x_0^4}{f(x_0)}\frac{x_0}{(x_0^2+1)^{1/2}} > 1, \tag{129}$$

the second inequality in (129) being a *strict* one for all finite x_0. *Hence all the* $\Gamma(\beta_1, y_0)$ *curves initially start above the Emden curve* $\omega_3{}^0$.

This has an immediate consequence. Consider a mass M which has in the wholly gaseous state a value for $\beta_1 = \beta^\dagger$. The curve $\Gamma(\beta^\dagger, x_0(\beta^\dagger))$ (where $x_0(\beta^\dagger)$ is the value of x at which degeneracy sets in for $\beta_1 = \beta^\dagger$), like the other Γ curves, starts from the point (3, 0), initially going above the $\omega_3{}^0$ curve. For this configuration $u = 3$, $v = 0$ and $\omega_3 = \omega_3{}^0$. When the central density of the configuration slightly increases, the degenerate core is of small but finite dimensions, and the appropriate values for $u(a)$ and $v(a)$ must still be in the neighbourhood of (3, 0), but since the Γ curves all lie initially above the $\omega_3{}^0$ curve, it follows at once from the disposition of the (u, v) curves that the *configuration is necessarily collapsed*. If centrally condensed configurations exist there must be some with small but finite degenerate cores to ensure the continuity of the curves of constant mass in the $(x, 1-\beta_1)$ diagram, but as we have shown that when degeneracy develops into a core, the formation is one due to the collapse of the configuration. From this it follows that *on the standard model* $(\beta_1 = \beta_2)$ *there exist only collapsed configurations.*

Now the $\Gamma(\beta_1, y_0)$ curves do not in general ascend above *all* the collapsed curves, though all the Γ curves start initially below all the collapsed curves. For a $\Gamma(\beta_1, y_0)$ curve goes to infinity with the asymptotic relation

$$\left(\frac{\pi^4}{960}\frac{\beta_1}{1-\beta_1}\right)^{1/3} V \sim \left(\frac{\pi^4}{960}\frac{\beta_1}{1-\beta_1}\right)^{1/3} \frac{\Omega^{2/3}(y_0)}{U^{1/3}}. \tag{130}$$

Hence by (122) these curves ascend above only such (u, v) curves which satisfy the inequality

$$\omega_3^{2/3} \leqslant \left(\frac{\pi^4}{960}\frac{\beta_1}{1-\beta_1}\right)^{1/3} \Omega^{2/3}(y_0), \tag{131}$$

or

$$\omega_3 \leqslant \left(\frac{\pi^4}{960}\frac{\beta_1}{1-\beta_1}\right)^{1/2} \Omega(y_0). \tag{132}$$

When the equality sign in (132) occurs the two curves touch at infinity. From (96) we have

$$\omega_3 = \omega_3^0 \left(\frac{1-\beta^\dagger}{\beta^{\dagger 4}}\frac{\beta_1^4}{1-\beta_1}\right)^{1/2}, \tag{133}$$

where $\beta^\dagger$ has the meaning with which we have used it so far. From (132) and (133) it follows that

$$\omega_3^0 \leqslant \left(\frac{\pi^4}{960}\frac{\beta^{\dagger 4}}{1-\beta^\dagger}\right)^{1/2} \beta_1^{-3/2}\Omega(y_0). \tag{134}$$

Now if $(1-\beta^\dagger) < (1-\beta_0)$, *i.e.* when $M < M_3$,

$$\left(\frac{\pi^4}{960}\frac{\beta^{\dagger 4}}{1-\beta^\dagger}\right) > 1, \tag{135}$$

and there exists a $y_0 = y_0^*$ (say) such that

$$\Omega(y_0^*) = \left(\frac{960}{\pi^4}\frac{1-\beta^{\dagger 4}}{\beta^{\dagger 4}}\right)^{1/2} \omega_3^0. \tag{135'}$$

From (135) and (135′) it follows that for $M < M_3$ there exist collapsed configurations, and in fact also a completely collapsed state for any particular mass $M(< M_3)$ with a central density corresponding to y_0 determined by (135′). When $M = M_3$ (134) can still be satisfied with $\beta_1 = 1$, only when $y_0 = \infty$. The curve for M_3 passes through the origin in the $(R, 1-\beta_1)$ diagram, but in the $(x, 1-\beta_1)$ diagram the appropriate curve goes to infinity with the x axis as the asymptote.

Finally, when $M > M_3$ we have

$$\left(\frac{\pi^4}{960}\frac{\beta^{\dagger 4}}{1-\beta^\dagger}\right) < 1, \tag{136}$$

and since the maximum value of $\Omega(y_0)$ is ω_3^0, the inequality (134) can be satisfied only for such values of $\beta_1 \leqslant \beta^*$, where

$$\beta^* = \left(\frac{\pi^4}{960}\frac{\beta^{\dagger 4}}{1-\beta^\dagger}\right)^{1/3}. \tag{137}$$

When $\beta_1 = \beta^*$, we have from (133) that

$$\omega_3 = \omega_3^0\left(\frac{\pi^4}{960}\frac{\beta^*}{1-\beta^*}\right)^{1/2}. \tag{138}$$

From (138) it follows that this ω_3 curve touches the $\Gamma(\beta^*, \infty)$ curve at infinity, *i.e.* the configuration is one in which the relative core radius is unity. Hence *when $M > M_3$ configurations collapsed more than to the extent* (138) *do not exist*. The curves of constant mass therefore intersect the $(1-\beta_1)$ axis at the point $(1-\beta^*)$ in the $(R, 1-\beta_1)$ diagram. In the $(x, 1-\beta_1)$ diagram the curves of constant mass for $M > M_3$ run to infinity, having as the asymptote the line through $(1-\beta^*)$ parallel to the x axis. We have thus re-derived some of our earlier results in § 9 from a proper discussion of the equations of fit. Since, however, only collapsed configurations exist, it is now obvious that in the domain of degeneracy the curves of constant mass connect by some "direct" path the point $(R_0(M), 1-\beta^\dagger)$ on the $(R_0, 1-\beta_1)$ curve to a point on the R axis $(M \leqslant M_3)$ or a point on the $(1-\beta_1)$ axis $(M > M_3)$.

20. *The Generalised Standard Model* $(\beta_2 = 1)$.—We shall now define by $\Gamma'(1-\beta_1, y_0)$ the curve

$$\Gamma'(1-\beta_1, y_0) = \left\{U, \left(\frac{\pi^4}{960}\frac{\beta_1^4}{1-\beta_1}\right)^{1/3} V\right\}, \tag{139}$$

with $(1-\beta_1)$ of course less than $(1-\beta_1(y_0))$. The initial negative slope of these curves Γ' are

$$\frac{5}{9}\left(\frac{\pi^4}{960}\frac{\beta_1^4}{1-\beta_1}\right)^{1/3} q, \qquad \left(q = \left(1 - \frac{1}{y_0^2}\right)^{1/2}\right). \tag{140}$$

Further, for $U \sim 0$, $V \sim \infty$, we have the asymptotic relation

$$\left(\frac{\pi^4}{960}\frac{\beta_1^4}{1-\beta_1}\right)^{1/3} V \sim \left(\frac{\pi^4}{960}\frac{\beta_1^4}{1-\beta_1}\right)^{1/3}\frac{\Omega^{2/3}}{U^{1/3}}. \tag{141}$$

Now if $(1-\beta_1) > (1-\beta_0)$, then

$$\left(\frac{\pi^4}{960}\frac{\beta_1^4}{1-\beta_1}\right) < 1, \qquad (1-\beta_1) > (1-\beta_0). \tag{142}$$

Since further $\Omega < \omega_3^0$, $q < 1$ (y_0 finite), we have the result that the $\Gamma'(1-\beta_1, y_0)$ *curves for* $(1-\beta_1) > (1-\beta_0)$ *all lie entirely below the* ω_3^0 *curve*. Hence

On the generalised standard model with $\beta_2 = 1$ *there exist centrally condensed configurations and only centrally condensed configurations for* $(1-\beta_1) > (1-\beta_0)$.

Consider now the mass M_3. By (133) we have

$$\omega_3 = \omega_3^0\left(\frac{\pi^4}{960}\frac{\beta_1^4}{1-\beta_1}\right)^{1/2}, \tag{143}$$

since for $M = M_3$, $\beta^\dagger = \beta_0$. Hence the asymptotic forms for the (u, v) curves are

$$v \sim \left(\frac{\pi^4}{960}\frac{\beta_1^4}{1-\beta_1}\right)^{1/3}\frac{\omega_3^{0\,2/3}}{u^{1/3}}. \tag{144}$$

Comparing this with (141), we see that when $\Omega = \omega_3^0$ (in which case $(1-\beta_1)$ need be less than only $(1-\beta_\omega)$) the appropriate (u, v) curves for the different values $0 \leqslant (1-\beta_1) \leqslant (1-\beta_\omega)$ all touch the $\Gamma'(1-\beta_1, \infty)$ at ∞. Hence the segment $0 \leqslant (1-\beta_1) \leqslant (1-\beta_\omega)$ of the $(1-\beta_1)$ axis in the $(R, 1-\beta_1)$ diagram is a part of the curve of constant mass for $M = M_3$. Hence

The curve of constant mass for $M = M_3$ in the domain of degeneracy consists of a centrally condensed branch joining the point $(R_0(M_3), 1-\beta_0)$ (on the $(R_0, 1-\beta_1)$ curve) to the point $(1-\beta_\omega)$ on the $(1-\beta_1)$ axis and the segment $0 \leqslant (1-\beta_1) \leqslant (1-\beta_\omega)$ of the $(1-\beta_1)$ axis.

From the above it follows that:

Composite configurations with $M > M_3$ are all necessarily centrally condensed.

It is easy to see that for these the curves of constant mass in the domain of degeneracy consists of a simple connection between the point $(R_0(M), 1-\beta^\dagger)$ to the point $(0, 1-\beta_\omega)$. But the point configurations with $M_3 \leqslant M \leqslant \mathfrak{M}$ at $(0, 1-\beta_\omega)$ are not all identical when each one of them is considered as the proper limit of an appropriate sequence of configurations (along the respective curves of constant masses). We then find that they all have different relative core radii.

For these point configurations the equations of fit take the limiting forms

$$u(a) = u(b, \omega_3^0), \tag{145}$$

$$v(a) = \beta_\omega v(b, \omega_3^0), \tag{146}$$

i.e. the problem consists in merely fitting two regions, the inner core being governed by the Emden function ($n = 3$) and the outer envelope by a centrally condensed solution of Emden's equation ($n = 3$). The appropriate (u, v) curve for a mass $M_3 > M \geqslant \mathfrak{M}$ is selected by its value for ω_3, which is readily seen to be

$$\omega_3 = \omega_3^0\left(\frac{960}{\pi^4}\frac{1-\beta^\dagger}{\beta^{\dagger 4}}\right)^{1/2}\beta_\omega^{3/2} \; (< \omega_3^0). \tag{147}$$

The equations of fit have a unique solution, since for these configurations

$$\left(\frac{960}{\pi^4}\frac{1-\beta^\dagger}{\beta^{\dagger 4}}\right)^{1/3}\beta_\omega > \beta_\omega, \tag{148}$$

and consequently the appropriate centrally condensed solution ultimately ascends above the Emden curve. The two curves must therefore intersect, the values of a and b (a on the centrally condensed curve, b on the Emden curve) at the point of intersection being the required solution of the equations of fit. In particular when $\beta^\dagger = \beta_\omega$ one easily sees that the appropriate point of intersection is $(3, 0)$—in other words, for $M = \mathfrak{M}$ the relative core radius for the point configuration at $(0, 1-\beta_\omega)$ is zero. We have already seen that for $M = M_3$ the point configuration at $(0, 1-\beta_\omega)$ has a relative core radius of unity. Hence *the limiting relative core radii along the centrally condensed sequences of composite configurations (each sequence being characterised by a mass $M_3 \leqslant M \leqslant \mathfrak{M}$) varies from unity for M_3 to zero for $\mathfrak{M}$.*

We pass on now to consider the types of configurations that occur for $M < M_3$.

The initial negative slope for $\Gamma'(1-\beta_1, y_0)$ curves can also be written in the form (*cf.* equation (140)):

$$\S\beta_1 \frac{2x'^4}{f(x')} \frac{x_0}{(x_0^2+1)^{1/2}}. \tag{149}$$

Now by the inequality (129) we can always formally calculate a "$\beta_c(x_0)$" such that

$$\beta_c(x_0) = \frac{f(x_0)}{2x_0^4} \frac{(x_0^2+1)^{1/2}}{x_0}. \tag{150}$$

If now

$$\beta_1 \geqslant \beta_c(x_0), \tag{151}$$

then

$$\beta_1 \frac{2x'^4}{f(x')} \frac{x_0}{(x_0^2+1)^{1/2}} \geqslant \beta_1 \frac{2x_0^4}{f(x_0)} \frac{x_0}{(x_0^2+1)^{1/2}} \geqslant 1. \tag{152}$$

On the other hand, if $\beta_1 < \beta_c(x_0)$, then the second inequality in (152) would not necessarily follow. But for this to happen $(1-\beta_1)$ must satisfy simultaneously two inequalities, since for a given y_0, $(1-\beta_1)$ must be less than $(1-\beta_1(y_0))$ determined by (95). The two inequalities then are

$$\left.\begin{aligned}(1-\beta_1) &> 1-\beta_c(x_0),\\ (1-\beta_1) &\leqslant 1-\beta_1(y_0).\end{aligned}\right\} \tag{153}$$

The two inequalities will not in general be satisfied. In fig. 1 the curve $(1-\beta_c(x))$ is plotted against x. In the same figure we have also the curve $(1-\beta_1(y))$. The two curves are seen to intersect. Let $\beta_1 = \beta_q$ and $x = x_q$ at the point of intersection. Then

$$1-\beta_c(x_q) = 1-\beta_1(y_q), \qquad (y_q^2 = x_q^2+1). \tag{154}$$

Numerically it is found that

$$1-\beta_q = \cdot 060 \text{ (approximately)}; \qquad x_q = 2\cdot 2 \text{ (approximately)}. \tag{155}$$

We shall also define a mass M_q by the relation

$$M_q = M_3\left(\frac{960}{\pi^4}\frac{1-\beta_q}{\beta_q^4}\right)^{1/2}. \tag{156}$$

i.e. M_q is the mass which has the value β_q for β_1 in the wholly gaseous state. Numerically we find that

$$M_q = 0\cdot 87 M_3 \text{ (approximately)}. \tag{156'}$$

Now consider a mass less than M_q, then for these configurations $(1-\beta^\dagger) < (1-\beta_q)$. Consequently the curve $\Gamma'(1-\beta^\dagger, x_0(\beta^\dagger))$ starts initially above the ω_3^0 curve. From this it follows that the composite configurations of mass $M < M_q$ with small but finite degenerate cores are necessarily of the collapsed type. But a collapsed configuration has a $(1 - \beta_1) < (1 - \beta^\dagger) < (1 - \beta_q)$, and the corresponding Γ' curve starts with a still greater initial negative slope. In this way one sees that *all composite configurations with $M \leqslant M_q$ are collapsed configurations.*

When, however, $M > M_q$ the curve $\Gamma'(1 - \beta^\dagger, x_0(\beta^\dagger))$ lies entirely below the ω_3^0 curve, and so do the curves for values of x_0 slightly greater. Centrally condensed configurations are possible. *Hence for $M_q < M \leqslant M_3$, centrally condensed and quasi-diffuse configurations exist in addition to the usual collapsed configurations.*

The curves of constant mass ($M \leqslant M_3$) intersect the R axis in the $(R, 1 - \beta_1)$ diagram at the same points as in the usual standard model. Further, as we have already shown in § 17, all these curves intersect the x axis in the $(x, 1 - \beta_1)$ diagram at right angles.

21. *Summary of the Main Results of* §§ 16–20.—In the preceding sections we have obtained some general information regarding the types of composite configurations that exist, and it is convenient to summarise here the main results.

A. There exist no composite configurations for $M \geqslant \mathfrak{M}$, where

$$\mathfrak{M} = M_3 \beta_\omega^{-3/2},$$

and β_ω is such that

$$\frac{960}{\pi^4} \frac{1 - \beta_\omega}{\beta_\omega} = 1.$$

B. *The Usual Standard Model* ($\beta_1 = \beta_2$).—(*a*) All composite configurations ($M < \mathfrak{M}$) are all necessarily collapsed.

(*b*) (i) For a prescribed $M < M_3$ the sequence of equilibrium configurations has as its limit a completely collapsed configuration ($\beta_1 = 1$) with a central density Bx_0^3 related to M by

$$M = M_3(\Omega(y_0)/\omega_3^0),$$

where

$$\Omega(y_0) = -\left(\eta^2 \frac{d\phi}{d\eta}\right)_1; \qquad \omega_3^0 = -\left(\xi^2 \frac{d\theta_3}{d\xi}\right)_1,$$

θ_3 being the Emden function ($n = 3$) and ϕ satisfying the differential equation

$$\frac{1}{\eta^2} \frac{d}{d\eta}\left(\eta^2 \frac{d\phi}{d\eta}\right) = -\left(\phi^2 - \frac{1}{y_0^2}\right)^{3/2}, \qquad (y_0^2 = x_0^2 + 1).$$

(ii) For $M = M_3$ the limiting completely collapsed configuration is of zero radius, and its value of "β_1" in the wholly gaseous state (denoted by β_0) satisfies the relation

$$\frac{960}{\pi^4} \frac{1 - \beta_0}{\beta_0^4} = 1.$$

(iii) For $M_3 < M < \mathfrak{M}$ the sequence of collapsed configurations end with a finite non-zero value for "$1 - \beta_1$," the maximum value β^* of β_1 being related to $\beta^{\dagger}$ (the value β_1 has for the prescribed mass in its wholly gaseous state) by the equation

$$\beta^* = \left(\frac{\pi^4}{960} \frac{\beta^{\dagger 4}}{1-\beta^{\dagger}}\right)^{1/3}.$$

C. *The Generalised Standard Model* ($\beta_2 = 1$).—(i) There exists a mass $M_q (\sim 0{\cdot}87 M_3)$ such that all composite configurations with $M \leqslant M_q$ are necessarily collapsed.

(ii) For $M_q < M < M_3$ each mass has a sequence of centrally condensed configurations which passes continuously (through a quasi-diffuse configuration) into a sequence of collapsed configurations, ending in a completely collapsed configuration with a central density related to the mass as in B (i) above.

(iii) For $M = M_3$ the composite configurations of *finite* radii are all centrally condensed. The limiting configuration along the centrally condensed sequence is one of zero radius with $1-\beta_1 = 1-\beta_\omega$. The relative core radius tends to unity along the centrally condensed sequence as the limiting configuration at $(1-\beta_\omega)$ is approached. Finally a whole sequence of point configurations for $0 \leqslant 1-\beta_1 \leqslant 1-\beta_\omega$ exists for this mass.

(iv) For $M_3 < M < \mathfrak{M}$ the composite configurations are *all* centrally condensed. The centrally condensed sequences for the different masses all end with a point configuration at $1-\beta_1 = 1-\beta_\omega$. The relative core radii for the different masses along their respective centrally condensed sequences tend to different limits, decreasing monotonically from unity for $M = M_3$ to zero for $M = \mathfrak{M}$.

The general results on the generalised standard model ($\beta_2 = 1$) are therefore of the character illustrated in fig. 4. To obtain more precise information one would have to solve the equations of fit, a simple and direct method for which has been outlined in § 15. But as mentioned there, the number of quadratures for the singular solutions of Emden's equation ($n = 3$) that have been carried out so far (by Fairclough) do not yet provide sufficient data to solve the equations of fit by the method of § 15 to any reasonable degree of accuracy. There exist numerical quadratures for only four collapsed solutions and three centrally condensed solutions, and our method to solve the equations of fit depends on a process of interpolation among the (u, v) tables for different ω_3. It is, however, possible to avoid the above process of interpolation by adopting an indirect method to solve the equations of fit, but this indirect method requires very much more numerical work to be done on our ϕ functions. These detailed calculations are reserved for future communications, but it is fortunate that the very circumstance that the degenerate core is governed by our differential equation for ϕ has allowed us to infer quite detailed results (summarised above) regarding the types of composite configurations that exist, and conclude therefrom the general character of the results illustrated in figs. 3 and 4.

We pass on now to comment briefly on the bearing of our results on certain aspects of the general problem of stellar evolution and stellar structure.

22. Consider a stellar mass $M < M_3$. In §§ 10–21 we have examined how such a stellar mass tends to the completely collapsed configuration (with the same mass) when the luminosity tends to zero through a sequence

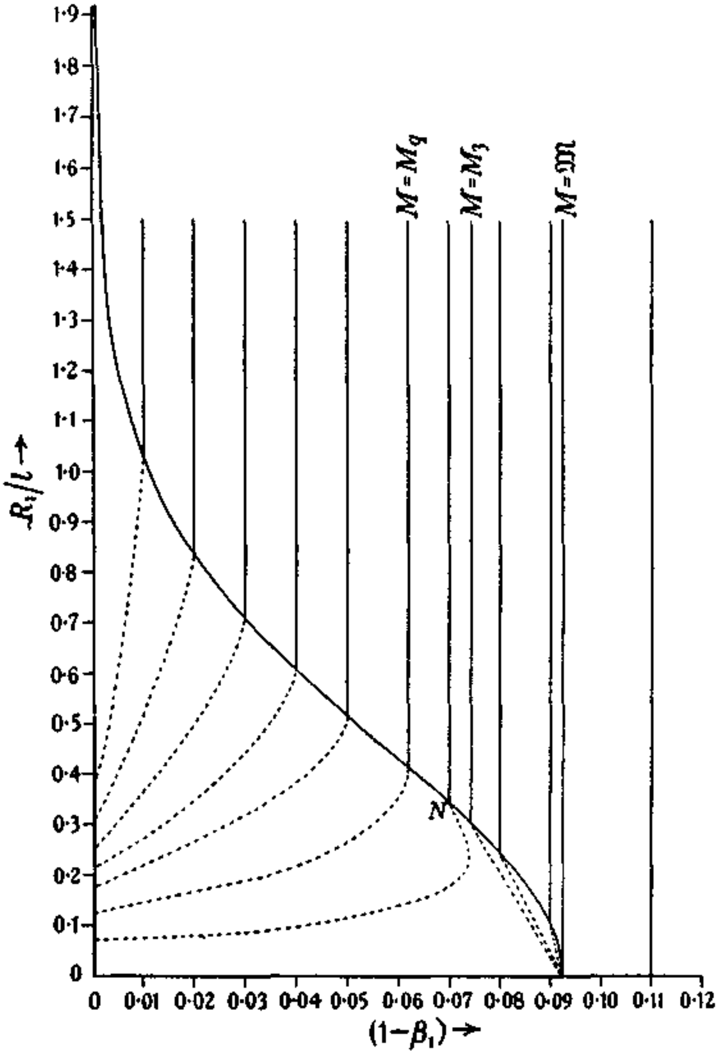


FIG. 4.—*The general nature of the curves of constant mass on the generalized standard model* ($\beta_2 = 1$). *The end-points of the various curves in the domain of degeneracy have been fixed exactly. The actual curves are shown dotted as the equations of fit have not been solved with any accuracy.*

of equilibrium configurations all conforming to the standard model. As it is clearly immaterial how the limit of zero luminosity is reached, it follows that *any* stellar mass $M < M_3$ must necessarily develop central regions of degeneracy when the luminosity decreases sufficiently, and should eventually become a "black dwarf"—to use a term due to R. H. Fowler—the structure of which will be governed by our differential equation for ϕ. Thus it is seen that all stellar masses less than M_3 will in the course of evolution pass through the white-dwarf stage, which itself is only a preliminary towards a state of complete degeneracy and darkness. It is more than probable that there exists a large number of such faint more or less completely collapsed configurations, though only a few white dwarfs are known. Such faint configurations would be difficult to detect both on account of their faintness and on account of their small dimensions. We have seen in II, § 12 (Table III), that $M = 5\cdot48\mu^{-2}\odot$ will in the completely collapsed state have a radius quite significantly smaller than the radius of the Earth.

Again we may expect that the fundamental distinction made on the standard model between masses less than and greater than $\mathfrak{M}$—in that for $M > \mathfrak{M}$ the equilibrium configurations are necessarily wholly gaseous (in the perfect gas sense)—is only a counterpart on the standard model of a more general result that all stars of sufficiently large mass are necessarily perfect gas configurations; for the increased dominance of radiation pressure for large stellar masses is quite a general result and, as we have already seen, the possibility of degeneracy is entirely excluded if only the radiation pressure is greater than a tenth of the total pressure throughout the entire mass. It should indeed be possible to define a "domain of degeneracy" on the basis of any model (as we have done in § 5 for the standard model) by drawing in the (radius-luminosity) diagram a curve similar to our $(R_0, 1-\beta_1)$ curve.* The result will always be that stars of mass greater than a certain limit will all be perfect gas configurations and will therefore conform to Eddington's mass-luminosity relation however much they may contract. As a result such stars can never pass *directly* into the white-dwarf stage. A possibility is that at some stage in the process of contraction the flux of radiation in the outer layers of the stars will become so large that a profuse ejection of matter will begin to take place. This ejection of atoms will continue till the mass of the star becomes small enough for central degeneracy to be possible. Once degeneracy sets in, the star will evolve along some white-dwarf sequence —possibly at the Kelvin-Helmholtz rate—to end again as a completely collapsed configuration.

One can perhaps look for a confirmation of these ideas in the fact that Wolf-Rayet stars, which are known to eject matter, must be massive if they are to conform to Eddington's mass-luminosity relation.† Thus Kosirev ‡ estimates that an average Wolf-Rayet star has a mass of about $10\odot$, and that further the annual loss of matter due to the ejection of atoms amounts to 10^{-6} of its mass.§ Once the mass becomes less than M_3 the luminosity would begin to decrease relatively more rapidly, the radial ejection would

* The author has already constructed such curves for some more general stellar models, but these extensions of the ideas developed in this paper will not be considered here.

† A massive star will be a perfect gas configuration, and we have therefore no reason to suspect that the Wolf-Rayet stars will not conform to Eddington's mass-luminosity relation.

‡ *M.N.*, 94, 430, 1934.

§ Some earlier estimates by Beals (*Pub. D.A.O.*, 4, No. 17, p. 297) gave rather lower rates, but in the author's opinion Beals's figures are definitely underestimates.

cease and the star collapse into a white dwarf. It is seen that these ideas fit consistently with the general indications that the nuclei of planetary nebula are white dwarfs and that the nebular envelope consists of matter ejected originally from the nuclear star. Our arguments in addition indicate that the masses of the nuclei of planetary nebula must be considerably less than that of an average Wolf-Rayet star. Thus, whether the star is of large or of small mass, the final stage in its evolution is always the white-dwarf stage, the only difference being that a star of large mass must first decrease its mass below M_3 by passing through the Wolf-Rayet stage. But in either case the completely collapsed configurations represent the "limiting sequence of stellar configurations to which all stars must tend eventually."

23. On the usual standard model we have shown that all the composite configurations are of the collapsed type. There is, however, a range of mass between M_3 and $\mathfrak{M}$ for which the configurations tend to zero radius with finite luminosity (which is less than the luminosity in the wholly gaseous state). But before that stage is reached there will occur the loss of mass due to the ejection of atoms, arising from the increased outward flux of radiation, and the final course of evolution for these masses must be more or less similar to that described for the perfect gas configurations in § 22, the only difference being that for the masses in the range $M_3 < M < \mathfrak{M}$ (or its equivalent on more general stellar models) the ejection would begin to take place when the luminosity is less than that predicted by Eddington's mass-luminosity relation.

24. *The Nova Phenomenon.*—We have shown that on the generalised standard model with $\beta_2 = 1$, for $M_3 > M > M_a$, there exist equilibrium configurations which are centrally condensed, and that for a prescribed mass the centrally condensed branch passes continuously into a sequence of collapsed configurations through a quasi-diffuse configuration. The quasi-diffuse configuration has a "$1 - \beta_1$" which is the same as that which characterises the wholly gaseous configuration. A possibility then for a configuration (with $M_a > M > M_3$) which, contracting from infinite extension, has reached a stage when degeneracy is just beginning to develop at the centre (*i.e.* the representative point in the $(R, 1 - \beta_1)$ plane is at a point like N in fig. 4) is that, instead of evolving along the centrally condensed branch, there occurs a discontinuous decrease of the external radius, the configuration passing into the quasi-diffuse state with a *finite* degenerate core with the same luminosity as it had at N. The consequence of such a decrease in radius would be the release of the difference of the total potential energies of the configuration at N and in its quasi-diffuse state vertically below N. It is possible that in this way the nova phenomenon (or more possibly the "super-nova phenomenon") occurs.

Ideas similar to those suggested above have been previously proposed by Milne. But it is perhaps necessary to draw attention to the differences between the suggestion made above and Milne's original ideas. *Firstly*, he regarded the "nova outburst" as occurring in the passage of a centrally condensed star into the quasi-diffuse state, *i.e.* the configuration is regarded as being initially in the centrally condensed branch, and the "nova outburst" as taking place when on *decreasing* luminosity the radius of the configuration *increases* till a point similar to N is reached. On the other hand, according to the suggestion made above the "nova outburst" is assumed to take place when a wholly gaseous configuration *contracting* from infinite extension passes directly into the collapsed sequence without ever passing into the centrally condensed branch. *Secondly*, on Milne's ideas all stars greater than a certain critical mass * can in principle become novæ, while our analysis shows that stars exhibiting the nova phenomenon must have masses in a comparatively small range.

25. *Have the Centrally Condensed Configurations that Exist for $M_a < M < M_3$ on the Generalised Standard Model any Relation to the Ordinary Stars?*—There is one initial difficulty in making any suggestions as to the relation of these centrally condensed configurations to the ordinary stars. For all these configurations have radii less than $0{\cdot}4l$ (see fig. 4), which is about $2 \times 10^9\mu^{-1}$ cm. This upper limit is itself very considerably smaller than the radii of ordinary stars. In connection with such a difficulty Milne has suggested † that such a radius discrepancy might be due to the use of the wrong boundary conditions ($T=0$, $\rho=0$), and that the use of the proper boundary conditions ($T=T_0$, $\rho=0$) might remove this discordance. Actually such investigations as have been made by Cowling ‡ and others do not favour this suggestion. But even if it should be possible to remove this discordance in a way which makes ordinary stars have small "cores" of the structure and dimensions similar to our centrally condensed configurations surrounded by extremely tenuous but extended envelopes, there would still be further difficulties to be overcome. The first is the opacity difficulty. In a general investigation Eddington § has shown that theories which postulate such extended envelopes would merely transfer the opacity difficulty from the core to the envelope, with indeed much larger factors in the discrepancy. It does not seem possible either, to remove this increased discordance by the hydrogen-abundance hypothesis. There is another difficulty as well. Milne has suggested in a different connection that if and when centrally condensed configurations are possible, the configuration having the maximum luminosity for the prescribed mass should exhibit the Cepheid phenomenon. If so, then on our analysis the Cepheids must have masses in a rather small range; but this is in very serious discordance with the results of observation. For these centrally condensed configurations, though they have luminosities greater than the corresponding gaseous configurations, are at most only 20 per cent. greater, and the mass-luminosity relation shows that the masses of the Cepheids can be almost anything. ‖ Thus on the whole the author is inclined to the view that the

* His critical mass is defined as one which has a "$1-\beta_1$" $=0{\cdot}2$ in the wholly gaseous state. One finds that this means $M=2{\cdot}1937M_3$, which is *greater* than our $\mathfrak{M}$.

† *Zeit. für Astrophysik*, 4, 75, 1932.

‡ T. G. Cowling, *Zeit. für Astrophysik*, 4, 331, 1932.

§ A. S. Eddington, *M.N.*, 91, 109, 1931. I am indebted to Sir A. S. Eddington for discussions on these and related matters.

‖ See Table xxv in *Internal Constitution of the Stars* (Cambridge).

centrally condensed configurations that occur in the theory of the generalised standard model have no relation to the ordinary stars. But the existence of a centrally condensed branch may have important bearings on other problems—for example in a possible explanation of the nova phenomenon as outlined in § 24.

Concluding Remarks.—In this paper the general problems of stellar structure as they present themselves on the standard model have been rediscussed, using the exact differential equation derived in the preceding paper to describe degenerate matter in gravitational equilibrium. Since we have restricted ourselves exclusively to the standard model, it is clear that only a first preliminary attack has been made on a much wider problem, of how the conclusions regarding stellar constitution and stellar evolution that have been drawn on the perfect gas hypothesis for the stars have to be modified by the physical possibility of degeneracy in stellar interiors. The methods that have been developed in this paper and the results obtained would have to be extended for more general stellar models before any very definite conclusions could be drawn.

The main results of the analysis are summarised in § 21 and figs. 1, 3 and 4; some general conclusions which follow are considered in §§ 22–25.

Trinity College, Cambridge:
1935 *January* 4.

Discussion of Papers 4 and 5 by A. S. Eddington and E. A. Milne

Dr. Chandrasekhar read a paper describing the research which he has recently carried out, an account of which has already appeared in *The Observatory*, **57**, 373, 1934, investigating the equilibrium of stellar configurations with degenerate cores. He takes the equation of state for degenerate matter in its exact form, that is to say, taking account of relativistic degeneracy. An important result of the work is that the life history of a star of small mass must be essentially different from that of a star of large mass. There exists a certain critical mass $\mathfrak{M}$. If the star's mass is greater than $\mathfrak{M}$ the star cannot have a degenerate core, but if the star's mass is less then $\mathfrak{M}$ it will tend, at the end of its life history, towards a completely collapsed state.

Prof. Milne. I have had an opportunity of seeing Dr. Chandrasekhar's paper. We have both been working on the same problem. I had intended to present a paper, written around Mr. Fairclough's latest numerical results, to this Meeting of the Society, but it has been unavoidably delayed. In many ways the methods pursued and the results obtained are the same as Dr. Chandrasekhar's. I have pursued a cruder method of analysis, but I believe that my method gives more insight into the fundamental physical postulates underlying the work, takes account of our ignorance of the behaviour of degenerate matter, and gives a more rational picture. A result common to our theory and Dr. Chandrasekhar's is that the more massive a star, the smaller its radius when completely collapsed. This has a bearing on the Russell diagram.

The President. Fellows will wish to return their thanks to Dr. Chandrasekhar. I now invite Sir Arthur Eddington to speak on his paper " Relativistic Degeneracy ".

Sir Arthur Eddington. Dr. Chandrasekhar has been referring to degeneracy. There are two expressions commonly used in this connection, " ordinary " degeneracy and " relativistic " degeneracy, and perhaps I had better begin by explaining the difference. They refer to formulæ expressing the electron pressure P in terms of the electron density σ. For ordinary degeneracy $P_e = K\sigma^{5/3}$. But it is generally supposed that this is only the limiting form at low densities of a more complicated relativistic formula, which shows P varying as something between $\sigma^{5/3}$ and $\sigma^{4/3}$, approximating to $\sigma^{4/3}$ at the highest densities. I do not know whether I shall escape from this meeting alive, but the point of my paper is that there is no such thing as relativistic degeneracy !

I would remark first that the relativistic formula has defeated the original intention of Prof. R. H. Fowler, who first applied the theory of degeneracy to astrophysics. When, in 1924, I suggested that owing to ionization we might have to deal with exceedingly dense matter in astronomy, I was troubled by a difficulty that there seemed to be no way in which a dense star could cool down. Apparently it had to go on radiating for ever, getting smaller and smaller. Soon afterwards Fermi-Dirac statistics were discovered, and Prof Fowler applied them to the problem and showed that they solved the difficulty : but now Dr. Chandrasekhar has revived it again. Fowler used the ordinary formula ; Chandrasekhar, using the relativistic formula which has been accepted for the last five years, shows that a star of mass greater than a certain limit $\mathfrak{M}$ remains a perfect gas and can never cool down. The star has to go on radiating and radiating and contracting and contracting until, I suppose, it gets down to a few km. radius, when gravity becomes strong enough to hold in the radiation, and the star can at last find peace.

Dr. Chandrasekhar had got this result before, but he has rubbed it in in his last paper ; and, when discussing it with him, I felt driven to the conclusion that this was almost a *reductio ad absurdum* of the relativistic degeneracy formula. Various accidents may intervene to save the star, but I want more protection than that. I think there should be a law of Nature to prevent a star from behaving in this absurd way !

If one takes the mathematical derivation of the relativistic degeneracy formula as given in astronomical papers, no fault is to be found. One has to look deeper into its physical foundations, and these are not above suspicion. The formula is based on a combination of relativity mechanics and non-relativity quantum theory, and I do not regard the offspring of such a union as born in lawful wedlock. I feel satisfied myself that the current formula is based on a partial relativity theory, and that if the theory is made complete the relativity corrections are compensated, so that we come back to the "ordinary" formula.

Suppose we are dealing with a cubic centimetre of material in the middle of a star. Ordinarily we analyse this into electrons, protons, etc., travelling about in all directions. In wave mechanics, the electrons are represented by waves. There are two kinds of waves, progressive and standing. In the ordinary analysis of matter into electrons one is dealing with progressive waves; but in the analysis which leads to the Exclusion Principle (used in deriving the degeneracy formula) the electron is represented by a standing wave. Now an electron represented by a standing wave is a quite different sort of entity from the electron represented by a progressive wave. The former is constantly changing its identity. I might compare the progressive wave with Professor Stratton and the standing wave with the President of the Royal Astronomical Society; only, to make the analogy a good one, the Society would have to change its President gradually and continuously, instead of suddenly every two years. The formulæ which apply to such a President would be different from the formulæ which apply to an ordinary individual; and this point has a definite bearing on the question. The electron represented by a progressive wave can be brought to rest by a Lorentz transformation, and it then becomes a standing wave. This transformation introduces a factor into the equation, which is not needed if the waves referred to are standing waves originally. My main point is that the Exclusion Principle presupposes analysis into standing waves, and this has been wrongly combined with formulæ which refer to progressive waves.

The President. The arguments of this paper will need to be very carefully weighed before we can discuss it. I ask you to return thanks to Sir Arthur Eddington.

CORRESPONDENCE.

To the Editors of 'The Observatory'.

The Configuration of Stellar Masses.

Gentlemen,—

In view of the fundamental character of the paper read by Sir Arthur Eddington at the meeting of the Royal Astronomical Society on 1935 January 11, perhaps I may be allowed to state that the basis of the calculation just completed by Mr. Norman Fairclough (referred to in my remarks at the meeting) is the equation of state $p=K\rho^{5/3}$ for a degenerate gas. For the sake of simplicity, and to have a well-defined case fully worked out, we had restricted attention to composite configurations for which "relativistic degeneracy", whether it exists or not, was ignored. Sir Arthur Eddington's investigations may now confer on our work a justification to which it is only accidentally entitled. The work consists in the carrying out of the programme sketched in *M. N. R. A. S.*, **92**, 610, 1932, and there left unfinished, namely the enumeration of *radii* for all possible values of L (luminosity) and M (mass) in the form of curves of r_1 (radius) against L for constant M. The work evidently coincides in aim, and partly in results, with the similar work by Dr. S. Chandrasekhar. The differences can more profitably be discussed when our papers are prepared for publication.

I am, Gentlemen,
Yours faithfully,

Oxford, 1935 Jan. 13. E. A. Milne.

STELLAR CONFIGURATIONS WITH DEGENERATE CORES. (SECOND PAPER.)

S. Chandrasekhar, Ph.D.

1. In a previous communication * the general problems of stellar structure as they present themselves on the standard model were rediscussed, using the exact relativistic equation of state to describe degenerate matter.† The method developed in I is, however, quite general and consists essentially in relating the completely degenerate gas spheres governed by the differential equation

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

---

* *M.N.*, **95**, 226–260, 1935. This paper will be referred to as I.

† In a recent paper, *M.N.*, **95**, 297, 1935, Eddington has questioned the validity of the relativistic equation of state for degenerate matter which is still generally accepted. There are, however, grounds for not abandoning the accepted form of the equation of state—the arguments are presented in the preceding paper by Dr. Christian Møller and the writer.

‡ This equation was established in the author's paper, *M.N.*, **95**, 207–226, 1935. This paper will be referred to as H.C. II. The earlier paper, *M.N.*, **91**, 456, 1931, will be referred to as H.C. I.

with the wholly gaseous configurations. Since on the standard model approximation for the gaseous configurations the ratio $(1-\beta_1)$ of the radiation pressure to the total pressure is a function of the mass only, the study of the curves of constant mass in the $(R,\ 1-\beta_1)$-diagram allows a convenient approach to the problem. In this diagram wholly gaseous configurations are represented by lines parallel to the R-axis, while the completely degenerate configurations are represented by points on the R-axis. The relation between these two sets of configurations was obtained by starting with a wholly gaseous configuration of prescribed mass and infinite extension and slowly contracting it and considering whether deviations from perfect gas laws towards degeneracy set in at all and if so when. In this way a *domain of degeneracy* in the $(R,\ 1-\beta_1)$-diagram was defined in which the configurations must be composite. To fix the precise nature of the curves of constant mass in the domain of degeneracy one requires further assumptions regarding the opacity of the degenerate core, but the problem of relating the degenerate spheres with the gaseous configurations was in principle solved.

2. But the discussion in I was incomplete in so far as the explicit appearance of the physically important parameter, namely, the luminosity L was suppressed by the use of $(1-\beta_1)$ as the main variable. To gain further physical insight it is necessary therefore to transform the discussion of the curves of constant mass in the $(R,\ 1-\beta_1)$-plane to a discussion of the curves of constant mass in the $(\log L,\ \log R)$-plane. This is done in Section I of this paper.

3. To complete the discussion we shall have to verify that the general results are not dependent on the very special nature of the model on which they have been obtained. As Jeans has more than once emphasised,* considerable caution is required in interpreting results based on stellar models which make gaseous configurations Emden polytropes of index 3. The more general analysis in which $(1-\beta_1)$ was allowed to vary through the configuration was provided by Jeans.† It follows from his analysis that *for fairly general stellar models the ratio $(1-\beta_c)$ of the radiation pressure to the total pressure at the centre of the configuration is a function of the mass only and is independent of the radius.* It is therefore clear that the whole discussion of I (especially that in Section I of that paper) can now be repeated on this more general analysis by considering the curves of constant mass in the $(R,\ 1-\beta_c)$-plane. In this plane the gaseous configurations are represented by lines parallel to the R-axis, and the relation between these "Emden-Jeans" polytropes to the completely degenerate configurations can be examined as before. This is done in Section II of this paper.

4. Lastly, in Section III various miscellaneous problems which arise are briefly considered.

* *Astronomy and Cosmogony* (Cambridge), chap. iii. Also *M.N.*, **85**, 201, 1925.

† *Astronomy and Cosmogony* (Cambridge), §§ 78–86, 88–92.

Section I.

5. *The Physical Variables.*—As shown in I, equation (55), Eddington's quartic equation can be written in the form

$$M = M_3\left(\frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1^4}\right)^{1/2}, \tag{2}$$

where

$$M_3 = -4\pi\left(\frac{2A_2}{\pi G}\right)^{3/2}\frac{1}{B^2}\left(\xi^2\frac{d\theta_3}{d\xi}\right)_1, \tag{3}$$

where A_2, B and the other symbols have the same meaning as in I. M_3 of course represents the upper limit to the mass of a completely degenerate configuration.

If R is the radius of the configuration, then one easily finds that the central density ρ_0 is given by

$$\rho_0 = B \cdot \frac{M}{M_3}\left(\frac{l}{R}\right)^3, \tag{4}$$

where l is the unit of length introduced in I, equation (33), namely,

$$l = \left(\frac{2A_2}{\pi G}\right)^{1/2}\frac{\xi_1(\theta_3)}{B}. \tag{5}$$

Again the central temperature T_0 of the configuration can be determined from the equation

$$T_0 = -\frac{\beta_1 \mu H}{4k}\frac{GM}{R(\xi\theta_3')_1}. \tag{6}^*$$

Using (2) and (3), (6) can be rewritten as

$$\frac{4kT_0}{mc^2} = \frac{Ml}{M_3 R}\beta_1. \tag{7}$$

If M^*, R^*, ρ^*, T^* denote the mass, the radius, the density and the temperature when expressed in units of M_3, l, B and $(mc^2/4k)$ respectively then we have

$$M^* = \left(\frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1^4}\right)^{1/2}; \quad \rho_0^* = \frac{M^*}{R^{*3}}; \quad T_0^* = \frac{M^*}{R^*}\beta_1. \tag{8}$$

Also we notice the relation

$$\frac{T_0^{*3}}{\rho_0^*} = \frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1}. \tag{9}$$

6. *Luminosity.*—We start with the equation †

$$L = \frac{4\pi c G M(1-\beta_1)}{\alpha\kappa_c}, \tag{10}$$

* See, for instance, Milne, *Handbuch der Astrophysik*, Band III/1, p. 209.

† A. S. Eddington, *Internal Constitution of the Stars* (Cambridge), p. 124 (equation 90.1).

where κ_c is the central opacity. We shall assume that

$$\kappa = \kappa_1 \frac{\rho^*}{T^{*7/2}}, \tag{11}$$

where κ_1 is the opacity at $\rho = B$ and $T = mc^2/4k$.

From (8), (9), (10) and (11) we derive that

$$L = \frac{4\pi c G M_3}{a\kappa_1}\left(\frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1^4}\right)^{7/4}\beta_1^{7/2}(1-\beta_1)R^{*-1/2}. \tag{12}$$

If we now introduce the unit of luminosity L_1 defined by

$$L_1 = \frac{4\pi c G M_3}{a\kappa_1}, \tag{13}$$

then (12) can be rewritten in the form

$$L^* = (M^*\beta_1)^{7/2}(1-\beta_1)R^{*-1/2}, \tag{14}$$

where L^* is used to denote the luminosity expressed in units of L_1. If we further introduce the quantity $L^*(\beta_1)$ defined by

$$L^*(\beta_1) = \left(\frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1^4}\right)^{7/4}\beta_1^{7/2}(1-\beta_1), \tag{15}$$

then we have from (14) that

$$\log L^* = \log L^*(\beta_1) - \tfrac{1}{2}\log R^*. \tag{16}$$

The first term on the right-hand side of (16) is a function of the mass only, and hence the curves of constant mass in the ($\log L^*$, $\log R^*$) diagram are straight lines.

7. *The Domain of Degeneracy in the* ($\log L^*$, $\log R^*$)*-Diagram.*—In I, § 5, we showed at what stage a configuration of a prescribed mass less than $\mathfrak{M}$ (contracting from infinite extension) would "just begin to develop degeneracy" * at the centre. This occurs when the radius R_0^* of the configuration is given by (I, equation (34))

$$R_0^* = \left(\frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1^4}\right)^{1/6}\frac{1}{x_0(\beta_1)}, \tag{17}$$

where $x_0(\beta_1)$ is such that

$$\frac{f(x_0)}{2x_0^4} = \left(\frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1}\right)^{1/3}. \tag{18}$$

This value of R_0^* substituted in (12) defines the corresponding luminosity L_0^*:

$$L_0^* = \left(\frac{960}{\pi^4}\frac{1-\beta_1}{\beta_1^4}\right)^{5/3}\beta_1^{7/2}(1-\beta_1)(x_0(\beta_1))^{-1/2}. \tag{19}$$

(17) and (19) together define a curve in the ($\log L^*$, $\log R^*$)-plane,

* What is here meant by "degeneracy just beginning to develop" is stated on p. 229 of my last paper (I).

corresponding to the $(R_0^*, 1-\beta_1)$-curve in the $(R^*, 1-\beta_1)$-plane. Thus for any given mass less than $\mathfrak{M}$ the intersection of the line

$$\log L^* = \log L^*(\beta^\dagger) - \tfrac{1}{2} \log R^* \tag{20}$$

with the $(\log L_0^*, \log R_0^*)$-curve defines the stage at which degeneracy would just begin to develop at the centre. In (20), $\beta^\dagger$ represents the value β_1 has for the prescribed mass in the wholly gaseous state. In Table I a set of corresponding pairs of values for $\log L_0^*$ and $\log R_0^*$ is given and the corresponding locus is shown in fig. 1. The part of the plane below this curve defines our ***domain of degeneracy*** in this plane.

Table I

x	$1-\beta_1$	$\log L_0^*$	$\log R_0^*$	x	$1-\beta_1$	$\log L_0^*$	$\log R_0^*$
0	0	$-\infty$	$+\infty$	3·0	·07149	$\bar{2}\cdot 9413$	$\bar{1}\cdot 5190$
0·2	·00039	$\bar{8}\cdot 2429$	0·2982	3·5	·07598	$\bar{1}\cdot 0520$	$\bar{1}\cdot 4579$
0·4	·00282	$\bar{6}\cdot 6625$	·1394	4·0	·07920	$\bar{1}\cdot 1339$	$\bar{1}\cdot 4039$
0·6	·00793	$\bar{5}\cdot 9546$	·0397	4·5	·08158	$\bar{1}\cdot 1969$	$\bar{1}\cdot 3556$
0·8	·01505	$\bar{4}\cdot 7689$	$\bar{1}\cdot 9632$	5·0	·08337	$\bar{1}\cdot 2479$	$\bar{1}\cdot 3120$
1·0	·02305	$\bar{3}\cdot 3221$	$\bar{1}\cdot 8995$	6·0	·08583	$\bar{1}\cdot 3249$	$\bar{1}\cdot 2357$
1·2	·03101	$\bar{3}\cdot 7164$	$\bar{1}\cdot 8441$	7·0	·08739	$\bar{1}\cdot 3817$	$\bar{1}\cdot 1706$
1·4	·03839	$\bar{2}\cdot 0076$	$\bar{1}\cdot 7949$	8·0	·08844	$\bar{1}\cdot 4261$	$\bar{1}\cdot 1138$
1·6	·04496	$\bar{2}\cdot 2290$	$\bar{1}\cdot 7503$	9·0	·08918	$\bar{1}\cdot 4624$	$\bar{1}\cdot 0625$
1·8	·05068	$\bar{2}\cdot 4016$	$\bar{1}\cdot 7095$	10·0	·08972	$\bar{1}\cdot 4931$	$\bar{1}\cdot 0182$
2·0	·05561	$\bar{2}\cdot 5391$	$\bar{1}\cdot 6720$	20·0	·09150	$\bar{1}\cdot 6691$	$\bar{2}\cdot 7193$
2·2	·05983	$\bar{2}\cdot 6506$	$\bar{1}\cdot 6372$	30·0	·09185	$\bar{1}\cdot 7620$	$\bar{2}\cdot 5435$
2·4	·06344	$\bar{2}\cdot 7427$	$\bar{1}\cdot 6048$	∞	·09212	∞	$-\infty$
2·6	·06653	$\bar{2}\cdot 8198$	$\bar{1}\cdot 5744$				
2·8	·06919	$\bar{2}\cdot 8852$	$\bar{1}\cdot 5459$				

It is of course clear that the $(\log L_0^*, \log R_0^*)$-curve asymptotically approaches the line

$$\log L^* = \log L^*(\beta_\omega) - \tfrac{1}{2} \log R^* = 1\cdot 0378 - \tfrac{1}{2} \log R^*, \tag{21}$$

where, as in I, equation (15), β_ω is such that

$$\frac{960}{\pi^4} \frac{1-\beta_\omega}{\beta_\omega} = 1. \tag{22}$$

That the $(\log L_0^*, \log R_0^*)$-curve asymptotically approaches the line (21) simply corresponds to the fact that $\mathfrak{M}$ represents the upper limit to the masses for which degeneracy can set in on contraction.

When $M << M_3$ one easily obtains the asymptotic relation (*cf.* I, equation 48)

$$L_0^* \sim \frac{\pi^4}{960}\left(\frac{4}{5}\right)^{33/2} R_0^{*-17}, \tag{23}$$

or

$$\log L_0^* = \bar{3}\cdot 4073 - 17 \log R_0^*, \quad (M^* \to 0). \tag{24}$$

The line (24) is also shown in fig. 1.

8. *The Nature of the Curves of Constant Mass in the Domain of Degeneracy.* —Consider a mass less than M_3. Then for this mass there exists an equili-

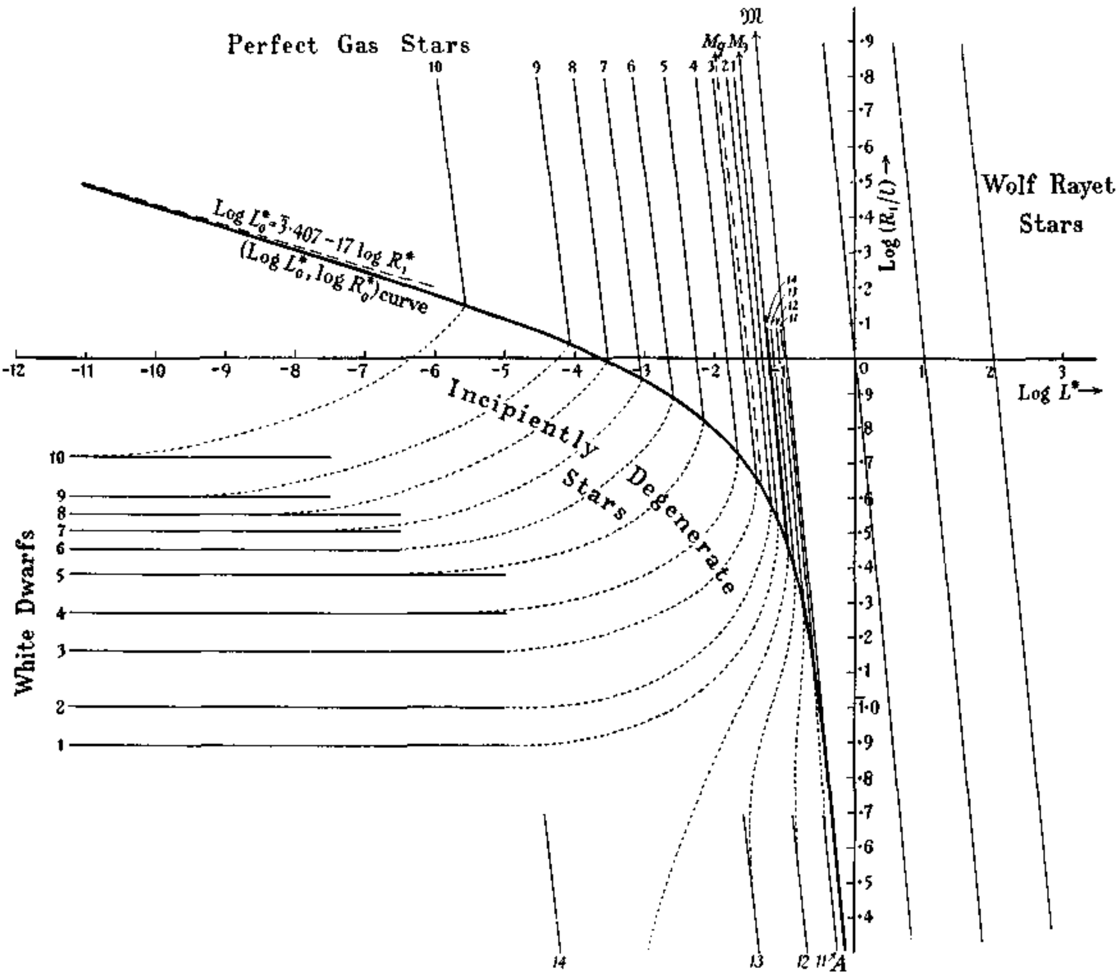


FIG. 1.—*The nature of the curves of constant mass in the (log L^*, log R^*)-plane on the usual standard model.*

For a general description of the results summarized in the above diagram, see 8 *and also* 23. *On the generalized standard model the system of the curves will look slightly different: thus the continuation of the curves of constant mass marked* (11 . . . 14) *in the perfect gas region will all tend asymptotically to the line* $\mathfrak{M}A$. *But the continuation of the curves of constant mass for those marked* (1 . . . 10) *in the perfect gas region must on all models eventually become asymptotic to the lines* (1 . . . 10) *in the domain of degeneracy (which is the region below the (log L_0^*, log R_0^*)-curve).*

brium configuration in which it is completely degenerate and has a radius R_1^* given by (*cf.* I, equation (46))

$$R_1^* = \frac{1}{y_0(M)} \frac{\eta_1(\phi(y_0(M)))}{\xi_1(\theta_3)}. \tag{25}$$

Hence if we start with this mass and contract it from infinite extension, then its luminosity increases according to (20) till this line intersects the ($\log L_0^*$, $\log R_0^*$) locus. On further contraction the configuration develops a degenerate core of finite dimensions and the luminosity must ultimately decrease, and as $L^* \to 0$ the curve must tend asymptotically to

$$\log R^* = \log R_1^* = \text{constant}. \tag{26}$$

In H.C. II we have obtained the values of η_1, M^*, etc. for ten different values of y_0, and for these configurations the values of $\log R_1^*$ and $\log L^*(\beta^\dagger)$ can be evaluated. The results of such calculations are given in Table II. The corresponding lines are shown in fig. 1 (the lines marked 1 to 10 in the domain of degeneracy and also in the perfect gas region).

The precise nature of the curves of constant mass in the domain of degeneracy will depend on the assumptions one makes regarding the opacity of the degenerate core. We shall indicate the qualitative results for the two extreme models ($\beta_1 = \beta_2$ and $\beta_2 = 1$) discussed in I, §§ 18–21.

(A) *The Usual Standard Model* ($\beta_1 = \beta_2$).—(i) For $M < M_3$, since the composite configurations are all of the collapsed type, it is clear that as soon as the configuration begins to develop degeneracy at the centre the luminosity should begin to increase less rapidly than $R^{-\frac{1}{2}}$. The curves of constant mass must therefore be of the nature shown by the dotted curves in fig. 1.

TABLE II

$\frac{1}{y_0^2}$	M/M_3	$\log L^*(\beta_1)$	$\log R_1^*$
0	1.0	$\bar{2}.7543$	$-\infty$
.01	.95733	$\bar{2}.6670$	$\bar{2}.8903$
.02	.92419	$\bar{2}.5957$	$\bar{1}.0096$
.05	.84709	$\bar{2}.4172$	$\bar{1}.1602$
.1	.75243	$\bar{2}.1690$	$\bar{1}.2708$
.2	.61589	$\bar{3}.7374$	$\bar{1}.3832$
.3	.51218	$\bar{3}.3286$	$\bar{1}.4538$
.4	.42600	$\bar{4}.9116$	$\bar{1}.5095$
.5	.35033	$\bar{4}.4620$	$\bar{1}.5590$
.6	.28137	$\bar{5}.9521$	$\bar{1}.6072$
.8	.15316	$\bar{6}.5173$	$\bar{1}.7198$

(ii) The curve for $M = M_3$ bends inwards as it enters the domain of degeneracy and goes to $(-\infty, -\infty)$.

(iii) For $M_3 < M < \mathfrak{M}$ the luminosity initially increases less rapidly than $R^{-1/2}$ on entering the domain of degeneracy, but should again ultimately increase as $R^{-1/2}$, since for these masses the curves of constant mass should be asymptotic to the line

$$\log L^* = \log L^*(\beta^*) - \tfrac{1}{2} \log R^*, \tag{27}$$

where β^* is related to $\beta^\dagger$ (the value β_1 has in the wholly gaseous state) by the equation (*cf.* I, equation (56)),

$$\beta^* = \left(\frac{\pi^4}{960}\frac{\beta^{\dagger 4}}{1-\beta^\dagger}\right)^{1/3}. \tag{28}$$

In fig. 1 this feature of the usual standard model is indicated.

(B) *The Generalized Standard Model* ($\beta_2 = 1$). — (i) For $M \leqslant M_q$ ($\sim 0{\cdot}87 M_3$) the composite configurations are all of the collapsed type and the qualitative nature of the curves of constant mass for these masses must therefore be of the same nature as A (i) above.

(ii) For $M_q < M < M_3$, the composite configurations are initially of the centrally condensed type, and hence for these masses the luminosity will begin to increase *more* rapidly than $R^{-1/2}$ on developing degeneracy at the centre. However, the luminosity must begin to decrease after attaining a certain maximum, since eventually the configurations must tend towards the completely degenerate state.

(iii) For $M_3 \leqslant M < \mathfrak{M}$ the luminosity increases more rapidly than $R^{-1/2}$ in the domain of degeneracy and the curves of constant mass for all these masses must asymptotically tend to the line (*cf.* equations (21), (22))

$$\log L^* = \log L^*(\beta_\omega) - \tfrac{1}{2}\log R^*. \tag{29}$$

9. So far we have restricted ourselves to the standard model approximation, and in relating the completely degenerate configurations with $M \leqslant M_3$ with the wholly gaseous configurations we have seen how such a stellar mass tends to the completely degenerate state when the luminosity tends to zero through a sequence of equilibrium configurations all conforming to the standard model. In addition to M_3 there appeared another mass $\mathfrak{M}$ which played an important rôle in the theory. $\mathfrak{M}$ was initially defined as one for which $\beta^\dagger = \beta_\omega$ (*cf.* I, equations (15), (29)), but the relation between M_3 and $\mathfrak{M}$, namely (I, equation (52)),

$$\mathfrak{M} = M_3 \beta_\omega^{-3/2}, \tag{30}$$

merely means that the existence of an upper limit M_3 to the mass of a completely degenerate configuration and the upper limit $\mathfrak{M}$ to the mass of a configuration for which degeneracy can at all set in on contraction are closely related to one another. Thus configurations in the mass range $M_3 \leqslant M \leqslant \mathfrak{M}$ bridge the gap between masses for which we have equilibrium configurations with zero luminosity and those which cannot develop degenerate cores however far the contraction may proceed.

The existence of a mass $\mathfrak{M}$ is by no means surprising, for, as we have already emphasised in I, § 22, the increased dominance of the radiation pressure for large stellar masses is quite a general result,* and the possibility of degeneracy is entirely excluded if only the radiation pressure is greater than a tenth of the total pressure throughout the entire mass. Hence it is clear that the general features that can be inferred from fig. 1 of this paper,

* An elementary proof of this result is given in the author's report in *Nordisk Astronomisk Tidskrift*, **16**, 37, 1935.

for instance, should to a large extent be independent of the model on the basis of which the discussion has been carried out. It is, however, of interest to verify that this is so by discussing the relation between the completely degenerate configurations and the wholly gaseous configurations on the basis of a more general scheme than what the standard model provides. As the analysis required for this verification is given in Section II below, we shall postpone to Section III some general considerations which arise from a closer examination of fig. 1.

Section II

10. The starting-point of our present discussion is provided by Jeans's investigations on gaseous configurations in which "$1-\beta_1$" is allowed to vary. We shall briefly recapitulate Jeans's analysis in our present notation.

It is clear, of course, that $(1-\beta_1)$ must decrease inwards but the precise law of variation will depend on various factors. If one assumes that

$$\frac{M}{L}\frac{L(r)}{M(r)} \propto T^{\delta}, \tag{31}$$

and that further the coefficient of opacity varies according to the law

$$\kappa \propto \frac{\rho}{T^{7/2}}, \tag{32}$$

then one can easily show that to a fair degree of approximation we have

$$\frac{\beta_1}{(1-\beta_1)^2} = \frac{\beta_c}{(1-\beta_c)}\left(\frac{T}{T_0}\right)^{(\frac{1}{2}-\delta)}, \tag{33}^*$$

where β_1 as usual defines the ratio of the gas pressure to the total pressure, and β_c the value of β_1 at the centre of the configuration where the temperature is assumed to be equal to T_0. Equation (33) is valid for layers not immediately near the surface. Further, we have quite generally that

$$\frac{\rho}{\rho_0} = \frac{\beta_1}{1-\beta_1}\frac{1-\beta_c}{\beta_c}\left(\frac{T}{T_0}\right)^3. \tag{34}$$

From (33) and (34) we deduce that

$$\frac{\beta_1}{(1-\beta_1)^2}\left(\frac{\beta_1}{1-\beta_1}\right)^{\frac{1}{3}(\frac{1}{2}-\delta)} = \frac{\beta_c}{(1-\beta_c)^2}\left(\frac{\beta_c}{1-\beta_c}\right)^{\frac{1}{3}(\frac{1}{2}-\delta)}\left(\frac{\rho}{\rho_0}\right)^{\frac{1}{3}(\frac{1}{2}-\delta)}. \tag{35}$$

From (35) we obtain that

$$[(1+\beta_1)+\tfrac{1}{3}(\tfrac{1}{2}-\delta)]\frac{d\beta_1}{\beta_1(1-\beta_1)} = \tfrac{1}{3}(\tfrac{1}{2}-\delta)\frac{d\rho}{\rho}. \tag{36}$$

Since, however, the total pressure P is given,

$$P = \left[\left(\frac{k}{\mu H}\right)^4 \frac{3}{a}\frac{1-\beta_1}{\beta_1^4}\right]^{1/3}\rho^{4/3}, \tag{37}$$

* This equation is due to Jeans and Woltjer.

we have (assuming μ constant *)

$$\frac{dP}{P} = -\frac{1}{3}\frac{4-3\beta_1}{\beta_1(1-\beta_1)}d\beta_1 + \frac{4}{3}\frac{d\rho}{\rho}\,; \tag{38}$$

or using (36),

$$\frac{dP}{P} = \frac{1}{3}\left[4 - \frac{(\frac{1}{2}-\delta)(4-3\beta_1)}{3(1+\beta_1)+(\frac{1}{2}-\delta)}\right]\frac{d\rho}{\rho}. \tag{39}$$

On the other hand, if

$$P = K\rho^{1+\frac{1}{n}}, \tag{40}$$

we should have

$$\frac{dP}{P} = \left(1+\frac{1}{n}\right)\frac{d\rho}{\rho}. \tag{41}$$

Comparing (39) and (41) we have for the "*effective polytropic index*" n the expression

$$n = 3 + (1-2\delta)\frac{4-3\beta_1}{1+3\beta_1+2\delta(1-\beta_1)}. \tag{42}$$

Equation (42) shows how the effective polytropic index n varies through the star. (With $\delta=\frac{1}{2}$, $n=3=$ constant, and we go back to the standard model.) However, Jeans considers that a fair approximation is obtained by regarding the whole configuration as a complete Emden polytrope with an index n_{β_c} given by

$$n_{\beta_c} = 3 + (1-2\delta)\frac{4-3\beta_c}{1+3\beta_c+2\delta(1-\beta_c)}. \tag{43}$$

11. If one assumes (43), then (37) can be rewritten as ($n=n_{\beta_c}$),

$$P = \left[\left(\frac{k}{\mu H}\right)^4 \frac{3}{a}\frac{1-\beta_c}{\beta_c^4}\right]^{1/3} \frac{1}{\rho_0^{(3-n)/3n}}\rho^{1+\frac{1}{n}}. \tag{44}$$

Or in terms of our A_2 and B we have

$$P = \frac{2A_2}{B^{4/3}}\left(\frac{960}{\pi^4}\frac{1-\beta_c}{\beta_c^4}\right)^{1/3}\frac{1}{\rho_0^{(3-n)/3n}}\rho^{1+\frac{1}{n}}. \tag{45}$$

The justification for (44) is simply that it gives the same initial variation of P with ρ at the centre as is jointly predicted by (35) and (37) taken together.

With (45) as the "equation of state" the structure of the configuration is completely specified by the Emden function $\theta_{n_{\beta_c}}$ with index n_{β_c}. We easily find that the mass of the configuration is given by

$$M = M_3\left(\frac{960}{\pi^4}\frac{1-\beta_c}{\beta_c^4}\right)^{1/2}\mathcal{J}_M(n_{\beta_c}), \tag{46}$$

where

$$\mathcal{J}_M(n) = \left(\frac{n+1}{4}\right)^{3/2}\frac{(\xi^2\theta_n{}')_1}{(\xi^2\theta_3{}')_1}. \tag{47}$$

* Variation of μ according to the law $\mu\alpha T^{\gamma}$ can easily be taken into account, but we shall not consider these refinements here.

If $\delta = 1/2$, $n_{\beta_c} = 3$ and $\mathcal{J}_M(3) = 1$, and (46) reduces to our earlier equation (2). Hence we can rewrite (46) as

$$M_{(\delta)}(\beta_c) = M_{(1/2)}(\beta_c) \cdot \mathcal{J}_M(n_{\beta_c}), \tag{48}$$

in an obvious notation.

12. *The Domain of Degeneracy in the* $(R,\ 1-\beta_c)$*-diagram.*—For a given mass M equations (43), (46) and (47) uniquely determine a β_c. In particular there will be a mass for which $(1-\beta_c) = (1-\beta_\omega)$. We shall denote this mass by $\mathfrak{M}_{(\delta)}$. By (46)

$$\mathfrak{M}_{(\delta)} = M_3 \beta_\omega^{-3/2} \mathcal{J}_M(n_{\beta_\omega}) = \mathfrak{M}_{(1/2)} \mathcal{J}_M(n_{\beta_\omega}), \tag{49}^*$$

where

$$n_{\beta_\omega} = 3 + (1-2\delta)\frac{4-3\beta_\omega}{1+3\beta_\omega+2\delta(1-\beta_\omega)}. \tag{50}$$

Arguing as in I, § 5, we now see that all configurations with $M > \mathfrak{M}_{(\delta)}$ are necessarily wholly gaseous, and that therefore for these masses the curves of constant mass in the $(R,\ 1-\beta_c)$-diagram are fully represented by the lines parallel to the R-axis. However, for $M < \mathfrak{M}_{(\delta)}$ degeneracy would begin to develop when the central density is given by

$$\rho_0 = Bx_0^3, \tag{51}$$

where x_0 is such that

$$\frac{f(x_0)}{2x_0^4} = \left(\frac{960}{\pi^4}\frac{1-\beta_c}{\beta_c}\right)^{1/3}. \tag{52}$$

The radius $R_0(\beta_c\,;\ \delta)$ of this configuration can be determined as in I, § 5, and we find that

$$\frac{R_0(\beta_c\,;\ \delta)}{l} = \left(\frac{n+1}{4}\right)^{1/2}\left(\frac{960}{\pi^4}\frac{1-\beta_c}{\beta_c^4}\right)^{1/6}\frac{1}{x_0}\frac{\xi_1(\theta_{n_{\beta_c}})}{\xi_1(\theta_3)}. \tag{53}$$

Comparing this with I, equation (34), we deduce that

$$R_0(\beta_c\,;\ \delta) = R_0(\beta_c\,;\ 1/2)\mathcal{J}_R(n_{\beta_c}), \tag{54}$$

where

$$\mathcal{J}_R(n) = \left(\frac{n+1}{4}\right)^{1/2}\frac{\xi_1(\theta_n)}{\xi_1(\theta_3)}. \tag{55}$$

(53) now defines a curve in the $(R,\ 1-\beta_c)$-plane. The region enclosed by the two axes and the $(R_0,\ 1-\beta_c)$-curve now defines our domain of degeneracy in this plane.

13. In our numerical work we shall confine ourselves exclusively to the case $\delta = 0$. Then we have

$$n_{\beta_c} = 3 + \frac{4-3\beta_c}{1+3\beta_c}. \tag{56}$$

The quantities $\mathcal{J}_R(n)$ and $\mathcal{J}_M(n)$ are known for certain values of n, and for the intermediate values recourse was made to methods of interpolation.

* It may be noticed here that $\mathfrak{M}_{(1/2)}$ is our original "$\mathfrak{M}$."

14. From (56) we now deduce that

$$n_{\beta_\omega}=3\cdot343\,;\quad \mathcal{J}_M(3\cdot343)=1\cdot080. \tag{57}$$

Hence

$$\mathfrak{M}_{(0)}=1\cdot080\mathfrak{M}_{(1/2)}=1\cdot249M_3, \tag{58}$$

or putting in numerical values

$$\mathfrak{M}_{(0)}=7\cdot153\mu^{-2}\odot. \tag{59}$$

15. In Table III we have tabulated for this model ($\delta=0$) the corresponding sets of values for x, $1-\beta_c$, R_0 and M. The (R_0, $1-\beta_c$)-curve is shown in fig. 2.

TABLE III

x	$1-\beta_c$	n_{β_c}	R_0/l	M/M_3
0	0	3·250	∞	0
0·2	·00039	3·250	2·381	·066
0·4	·00282	3·253	1·655	·178
0·6	·00793	3·257	1·320	·301
0·8	·01505	3·264	1·113	·421
1·0	·02305	3·272	·967	·531
1·2	·03101	3·280	·856	·627
1·4	·03839	3·287	·769	·710
1·6	·04496	3·294	·697	·780
1·8	·05068	3·299	·638	·839
2·0	·05561	3·304	·587	·889
2·2	·05983	3·309	·544	·931
2·4	·06344	3·312	·506	·967
2·6	·06653	3·316	·473	·998
2·8	·06919	3·318	·444	1·024
3·0	·07149	3·321	·418	1·046
3·5	·07598	3·326	·365	1·090
4·0	·07920	3·329	·323	1·122
4·5	·08158	3·331	·289	1·145
5·0	·08337	3·333	·262	1·163
6·0	·08583	3·336	·220	1·187
7·0	·08739	3·338	·190	1·202
8·0	·08844	3·339	·167	1·212
9·0	·08918	3·340	·149	1·220
10·0	·08972	3·340	·134	1·225
20·0	·09150	3·342	·067	1·242
30·0	·09185	3·342	·045	1·246
∞	·09212	3·343	0	1·249

16. *The Nature of the Curves of Constant Mass for $M \leqslant M_3$ in the Domain of Degeneracy.*—Let $\beta_c{}^\dagger$ be the value of β_c for a wholly gaseous configuration (of mass less than M_3) which in its completely collapsed state has a central density corresponding to $y=y_0$. Then by equation (46)

$$\frac{\Omega(y_0)}{\omega_3{}^0}=\left(\frac{960}{\pi^4}\,\frac{1-\beta_c{}^\dagger}{\beta_c{}^{\dagger 4}}\right)^{1/2}\mathcal{J}_M(n_{\beta_c}{}^\dagger). \tag{60}$$

Now the line through $(1-\beta_c^\dagger)$ parallel to the R-axis will intersect the $(R_0, 1-\beta_c)$-curve at $(R_0(M(y_0)), 1-\beta_c^\dagger)$. In the domain of degeneracy the continuation of the curve must in some way connect the point $(R_0(M(y_0)), 1-\beta_c^\dagger)$ and the point R_1 on the R-axis where

$$\frac{R_1}{l}=\frac{1}{y_0(M)}\frac{\eta_1(\phi(y_0(M)))}{\xi_1(\theta_3)}. \tag{61}$$

In I (Table II) we have already tabulated the values of R_1 for ten different values for y_0 and for these configurations $(1-\beta_c^\dagger)$ was obtained by interpolating among the figures given in Table III. The corresponding pairs of points on the $(R_0, 1-\beta_c)$-curve and the R-axis are shown in fig. 2. (The points marked 5 to 15 on the R-axis and also on the $(R_0, 1-\beta_c)$-curve.) It is of course clear that for $M=M_3$ the associated curve of constant mass must pass through the origin of our system of axes. If we denote by $\beta_c(0)$ the value of β_c which M_3 has in the wholly gaseous state then we should have by (60) that

$$\left(\frac{960}{\pi^4}\frac{1-\beta_c(0)}{\beta_c^4(0)}\right)^{1/2}\mathcal{J}_M(n_{\beta_c(0)})=1. \tag{62}$$

Numerically $(1-\beta_c(0))$ is found to be 0·0668.

17. *The Nature of the Curves of Constant Mass for $M_3 \leqslant M \leqslant \mathfrak{M}_{(\delta)}$ in the Domain of Degeneracy.*—The discussion of this case will naturally depend on the assumption one makes regarding the opacity in the degenerate core. We shall assume that " $(\kappa\eta)_2$ " is constant in the core. Then we have (*cf.* I, equation (63))

$$P=\frac{1}{\beta_2}A_2f(x)-D_2\frac{1-\beta_2}{\beta_2}, \tag{63}$$

where D_2 is a constant and

$$\beta_2=\left(1-\frac{(\kappa\eta)_2L}{4\pi cGM}\right). \tag{63'}$$

The reduction to our differential equation (1) for the degenerate core follows at once.

In our present scheme " β " is of course allowed to vary in the gaseous envelope, and we shall in the first instance consider the case where β_2 is just equal to the value of β_1 at the interface between the degenerate core and the gaseous envelope. For this case the discussion can be carried out as in I, § 9.

A completely relativistically degenerate configuration has a mass given by (H.C. I, page 463)

$$M=M_3\beta^{-3/2}, \tag{64}$$

and is of zero radius. Since we have the further relation

$$\mathfrak{M}_{(1/2)}=M_3\beta_\omega^{-3/2}, \tag{65}$$

it follows that the curve of constant mass for $\mathfrak{M}_{(1/2)}$ must connect the point $(R_0, 1-\beta_c^\dagger(\mathfrak{M}_{(1/2)}))$ on the $(R_0, 1-\beta_c)$-curve and the point $(0, 1-\beta_\omega)$ on the $(1-\beta_c)$-axis. It is therefore clear that for $M_3 \leqslant M \leqslant \mathfrak{M}_{(1/2)}$ the

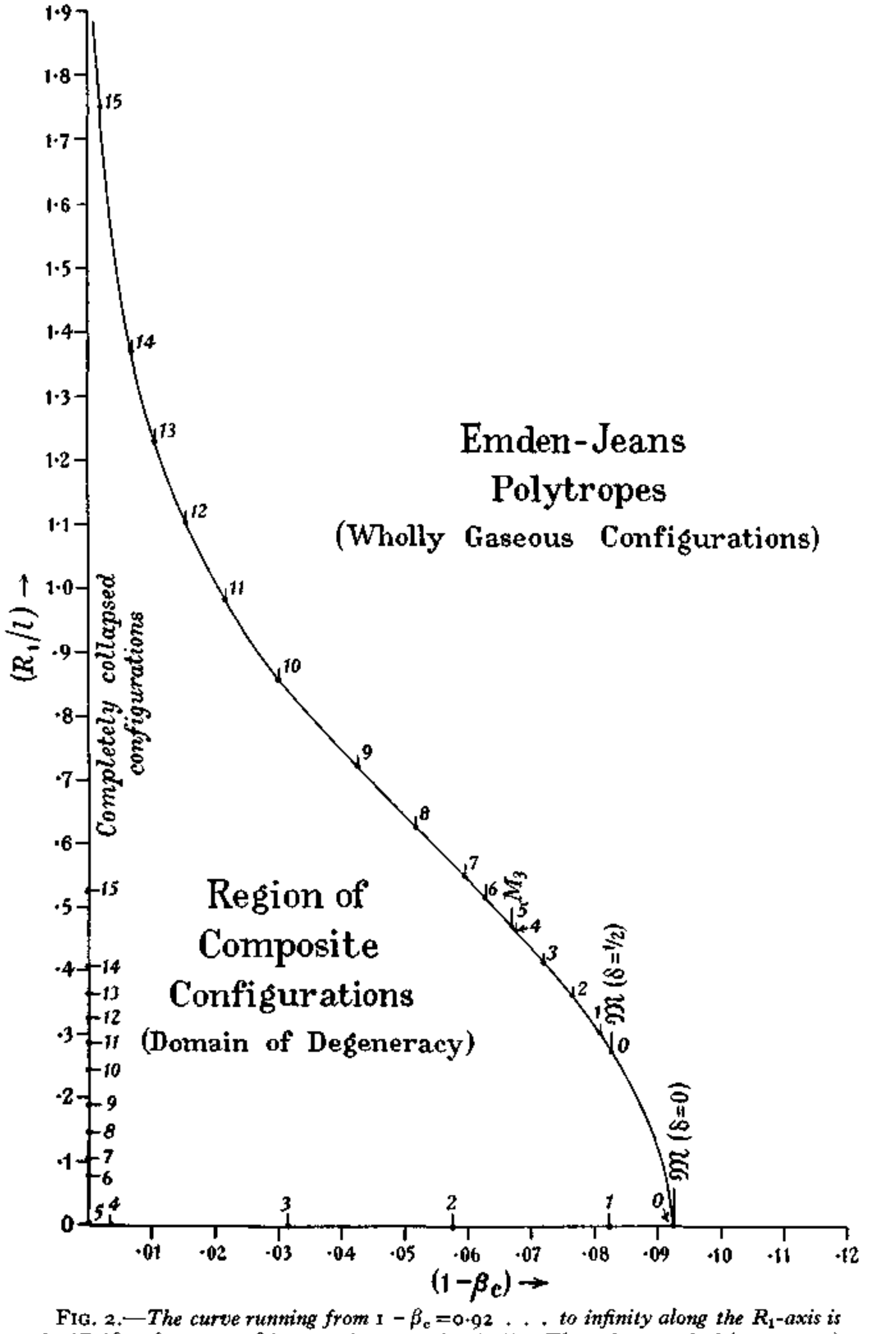


FIG. 2.—*The curve running from* $1-\beta_c=0{\cdot}92$ . . . *to infinity along the* R_1*-axis is the* $(R_0(\beta_c\,;\,\delta=0),\ 1-\beta_c)$*-curve (see equation (53)). The points marked* (5 . . . 15) *on the* $(R_0,\ 1-\beta_c)$*-curve and on the* R_1*-axis are the end points in the domain of degeneracy for the curves of constant mass for the values of* M *for which the* ϕ*-integrals are known (H.C. II). The points marked* (0, 1 . . . 4) *on the* $(R_0,\ 1-\beta_c)$*-curve and on the* $(1-\beta_1)$*-axis are the corresponding end points for some curves of constant mass in the domain of degeneracy on the model discussed in* 17 *(see equation (67)). The point* 0 *corresponds to* $\mathfrak{M}_{(\delta=\frac{1}{2})}$, i.e. *our earlier* "$\mathfrak{M}$."

curves of constant mass must cross the $(1-\beta_c)$-axis at a point $(1-\beta^*)$ such that

$$M = M_3 \beta^{*-3/2}. \tag{66}$$

If $\beta_c{}^\dagger$ denotes the value of β_c in the wholly gaseous state then by equating (66) and (46) we have

$$\beta^* = \left(\frac{\pi^4}{960}\frac{\beta_c{}^{\dagger 4}}{1-\beta_c{}^\dagger}\right)^{1/3}(\mathfrak{J}_M(n_{\beta_c}{}^\dagger))^{-2/3}. \tag{67}$$

Some corresponding pairs of points on the $(R_0, 1-\beta_c)$-curve and the $(1-\beta_c)$-axis are shown in fig. 2.

Finally, if $\mathfrak{M}_{(1/2)} \leqslant M \leqslant \mathfrak{M}_{(3)}$ it is immediately clear that the curves of constant mass consist simply of segments connecting the point $(R_0, 1-\beta_c{}^\dagger(M))$ on the $(R_0, 1-\beta_c)$-curve to the point $(0, 1-\beta_\omega)$ on the $(1-\beta_c)$-axis.

We thus see that on this model the curves of constant mass in the $(R_0, 1-\beta_c)$-plane combine in the same diagram some of the features of both figs. 3 and 4 of I, obtained on the basis of the two extreme cases of the generalized standard model.

18. It may finally be pointed out that if one assumed that the opacity of the degenerate core is zero then the general qualitative features of the system of the curves of constant mass in the $(R, 1-\beta_c)$-plane must be *exactly* the same as in I, fig. 4.

19. A complete discussion on the basis of Jeans's model will require a study of the composite configurations. The formal theory (which would run similar to I, §§ 11 to 15) can easily be sketched, but as such discussions are not of much interest without the necessary numerical work (which would be considerable) we shall not go into these details here. However, it is clear that the general results derived on the basis of the standard model are fully retained even in this more general analysis.

Section III

20. *The Wolf-Rayet Phenomenon.*—It has already been suggested in I that the Wolf-Rayet phenomenon of the radial ejection of matter may be indirectly due to the fact that the stars of mass greater than $\mathfrak{M}$ (or its equivalent $\mathfrak{M}_{(\delta)}$ on the more general stellar models discussed in Section II) cannot pass *directly* into the white-dwarf stage.* This suggestion is confirmed by observation in so far as general estimates do indicate that Wolf-Rayet stars are massive and dense. On the theoretical side the suggestion gains further support from the following argument:—

Consider a mass greater than $\mathfrak{M}_{(1/2)}$. On the standard model the star must necessarily be wholly gaseous, and we have in a certain system of "natural units" (*cf.* equation (14))

$$L^* = (M^*\beta_1)^{7/2}(1-\beta_1)R^{*-1/2}. \tag{68}$$

* This suggestion (in a rather different form) was independently made to the writer by Dr. W. H. McCrea, to whom the author had earlier communicated a preliminary statement of the main results in the form later published in *Observatory*, **57**, 373, 1934.

Further, the value of the surface gravity g is given by

$$g=\frac{GM}{R^2}. \tag{69}$$

From (68) and (69) we have for the ratio X between the integrated flux of radiation πF at the surface of the star to the value of gravity, the expression

$$X=\frac{\pi F}{g}=\frac{L}{4\pi cGM}=\frac{L_1}{4\pi cGM_3}M^{*5/2}\beta_1{}^{7/2}(1-\beta_1)R^{*-1/2}. \tag{70}$$

From our definition of L_1 in (13) we have

$$X=\left(\frac{1}{\alpha\kappa_1}\right)M^{*5/2}\beta_1{}^{7/2}(1-\beta_1)R^{*-1/2}. \tag{71}$$

From (71) we see that for a given mass ($>\mathfrak{M}$) the ratio X steadily increases with decreasing R and in fact tends to infinity. This suggests that at some stage in the process of contraction the radiation pressure (a measure of which is given by πF) must overbalance gravity. Ejection of matter must necessarily ensue. In drawing this inference it is of course realized that the deduction from (71) cannot be regarded as a rigorous proof in so far as in our analysis the equations of equilibrium have been integrated up to the boundary. But if one takes this last boundary condition seriously then one cannot also strictly speak of a "mass-luminosity-effective temperature" relation as the temperature has been made zero at the boundary. That this involves no real contradiction was shown in the early writings of Eddington, Jeans and Russell and more recently by Cowling. Bearing this in mind it is now clear that the fact that $X\to\infty$, as $R\to 0$ merely means that the approximations underlying the deduction of the mass-luminosity-radius relation (68) should cease to be valid at some stage. Our conclusion that the ejection of matter must ensue since $X\to\infty$, as $R\to 0$ is now seen to be equivalent to the suggestion that the Wolf-Rayet phenomenon should set in precisely in the region of the Russell diagram where the mass-luminosity-radius relation for the massive wholly gaseous configurations ceases to be valid on the perfect-gas hypothesis itself. It is of course necessary that the star should be massive ($M>\mathfrak{M}$ or its equivalent on more general stellar models) for otherwise we could not extrapolate (68) to high mean densities—degeneracy would have set in earlier for the less massive stars.

21. *The Hydrogen Content of the Massive Stars.*—In § 20 we have used the term "massive stars" to denote those with $M>\mathfrak{M}_{(\delta)}$. It was found on Jeans's model ($\delta=0$) that we have

$$\mathfrak{M}_{(0)}=7\cdot 15\mu^{-2}\odot. \tag{72}$$

To define $\mathfrak{M}_{(\delta)}$ more precisely we need to know the hydrogen content. Depending on the hydrogen content $\mathfrak{M}_{(\delta)}$ can be varied numerically by a factor 16 ($\mu=\frac{1}{2}$ to $\mu=2$), and it becomes necessary therefore to know at least the minimum hydrogen contents of stars as a function of their mass.

Fortunately we have for our guidance here Strömgren's systematic investigations of this problem.* From Strömgren's work it appears that the molecular weight of the massive B stars already tends towards the lower limit 0·5. Thus we can conclude that for our purposes a star of mass greater than about 25☉ can be regarded as "massive."

It is of interest to recall in this connection that in his "Interpretation of the Hertzsprung-Russell diagram" Strömgren says: "With an appreciable overcompressible nucleus the predicted luminosities would be appreciably larger than the observed, and increasing the hydrogen content—as is usually possible to remove the difference—is not possible in these cases, as the limit has already been reached. We conclude then that for the B stars in question there cannot be any appreciable nucleus." We now see that Strömgren's conclusions receive further indirect confirmation from our analysis.

22. *The Hydrogen Content of the White Dwarfs.*—The hydrogen content of the white dwarfs had been investigated earlier by various writers on the Emden polytrope $n=3/2$ approximation for them. In our H.C. II we have made an exact study of these completely degenerate configurations, and it is now possible to make a more reliable estimate for the appropriate molecular weights for the white dwarfs.

The necessary data required for this calculation are given in Table III of H.C. II. The following table is due to Strömgren :—

Table IV

μ for Sirius B

μ	1·91	1·74	1·58	1·44	1·29	0·95
T_{eff} (calculated)	18,300°	13,800°	11,300°	9700°	8300°	6200°

From the above table Strömgren concludes that the value of μ for Sirius B should be about 1·6, which means relatively low hydrogen content. The low hydrogen content of the white dwarfs has already been discussed by Strömgren (*loc. cit.*).

23. *Some Remarks on Figure* 1.—Figure 1 is of course the domain of the Hertzsprung-Russell diagram. From an examination of this diagram it is immediately clear that the white dwarfs are placed in their right positions in the Russell diagram. The two essential observational results concerning the white dwarfs, namely, their small mass and low luminosity, receive their natural explanations. The region of the diagram in which we should expect the Wolf-Rayet stars is also indicated. The region of the ordinary stars is indicated by "perfect gas stars." Presumably stars like Krueger 60 are representatives of the "incipiently degenerate" region of our diagram.

The above general conclusions, so far as they go, should clarify the present position regarding stellar structure.

* B. Strömgren, *Z. f. Astrophysik*, 7, 222, 1933.

24. *Deviations from Perfect Gas Laws arising from Causes other than Degeneracy.*—In our discussion we have so far considered only deviations from perfect gas laws which are due to degeneracy. However, Dirac's theory of the electron predicts a further different type of deviation from the perfect gas laws due to the production of electron pairs at very high temperatures. The bearing of this phenomenon on the theory of stellar structure has been examined in a preliminary communication by L. Rosenfeld and the writer.* As we have indicated in that letter, the deviations from the perfect gas laws arising from this cause are of quite negligible importance for stars with $M < \mathfrak{M}_{(\delta)}$; however, they become increasingly important for the very massive stars. The detailed results of this study will be published separately by Rosenfeld and the writer, but it may be mentioned here that it follows from that study that the production of electron pairs will be of importance in considerations of the structure of stars of masses about 80☉ and more. The existence of such very massive stars is indicated by the work of J. S. Plaskett, O. Struve, Bottlinger, Trumpler and others, and it seems very probable that the discussion of their structure will lead to some essentially new considerations in the studies on stellar structure.

Finally, it is necessary to point out in this connection that J. von Neumann has recently shown that the *very* ultimate equation of state for matter should *always* be

$$P = \frac{1}{3}c^2\rho. \tag{73}$$

The considerations of this new equation of state does not, however, introduce any essential modifications in our present scheme.

Concluding Remarks.—In two earlier papers (*M.N.*, **95**, 207–260, 1935) a first systematic attack was made on the problem of how the conclusions regarding stellar constitution and stellar evolution that have been drawn on the perfect gas hypothesis for the stars have to be modified by the physical possibility of degeneracy in stellar interiors. In this paper the discussion is carried one stage further. Firstly, the physical results have been made more explicit by considering the curves of constant mass in the $(\log L, \log R)$-diagram, which is essentially the domain of the Hertzsprung-Russell diagram. Secondly, the analysis has been extended to include other stellar models more general than the standard model. Thirdly, the bearing of the Wolf-Rayet phenomenon on the evolution of massive stars is examined a little more closely. Certain other miscellaneous questions have also been considered.

In conclusion, I wish to record here my best thanks to Dr. W. H. McCrea, Professor J. von Neumann, Dr. L. Rosenfeld and Dr. B. Strömgren for the encouraging interest they have taken in these studies and for many stimulating discussions.

Trinity College, Cambridge :
1935 *June* 7.

* *Nature*, **135**, 999, 1935.

Printed in Edinburgh by Neill & Co., Ltd., 1935 *August*

The White Dwarfs and Their Importance for Theories of Stellar Evolution

by S. CHANDRASEKHAR.

1. According to current ideas the structure of the White Dwarfs if to be understood in terms of the deviations from the perfect gas law ($p \propto \rho T$) which quantum statistics predicts. The application of the quantum theory to the statistical mechanics of an electron assembly indicates that the classical treatment of the same problem ignores two factors. The first is the consequences of the Pauli principle which requires the wave functions describing the system to be antisymmetrical in the coordinates (x, y, z, and spin) of the different electrons; the second is the effect of the variation of mass with velocity predicted by the special theory of relativity. Both of these effects can be taken into account and the application of standard methods leads to the following parametric form for the equation of state of an electron gas:

$$N = \frac{8\pi V m^3 c^3}{h^3} \int_0^\infty \frac{\sinh^2\theta \cosh\theta \, d\theta}{\frac{1}{\Lambda} e^{\theta mc^2 \cosh\theta} + 1} \tag{1}$$

$$P = \frac{8\pi m^4 c^5}{3 h^3} \int_0^\infty \frac{\sinh^4 \theta \, d\theta}{\frac{1}{\Lambda} e^{\theta mc^2 \cosh \theta} + 1} \tag{2}$$

where

$$\theta = 1/kT \tag{3}$$

In equations (1), (2), and (3), m, c, and h denote the mass of the electron, the velocity of light, and the Planck constant, respectively. Further these equations refer to the pressure P exerted by the electrons in an enclosure of volume V containing N electrons.

Equations (1) and (2) can be written in the forms :

$$N = V f_1(\Lambda, \theta)\,; \qquad P = f_2(\Lambda, \theta). \tag{4}$$

and the elimination of Λ will give the required equation of state. This elimination cannot be effected satisfactorily except in two limiting cases, namely when (i) $\Lambda << 1$ and (ii) $\Lambda >> 1$. For these two cases we have respectively,

$$N = PV \tag{5}$$

and

$$N = \frac{8\pi m^3c^3}{3h^3}\, x^3 = A\, x^3$$

$$P = \frac{\pi m^4 c^5}{3h^3}\, f(x) = B\, f(x) \tag{6}$$

where

$$f(x) = x\,(2x^2 - 3)\,(x^2 + 1)^{1/2} + 3\,\sinh^{-1} x \tag{6'}$$

and

$$A = \frac{8\,\pi m^3c^3}{3h^3}\,; \qquad B = \frac{\pi m^4c^5}{3h^3} \tag{6''}$$

These two limiting forms for the equation of state defined by (1) and (2) are said to correspond to the non-degenerate and degenerate cases respectively [1].

2. It is now generally agreed that the equation of state appropriate for the discussion of the structure of White Dwarfs is the degenerate form of the equation of state. The main argument for this conclusion comes from the application of the theory of stellar envelopes which shows that the gaseous outer regions of the White Dwarfs will constitute only a thin outer fringe. Consequently almost the entire mass of the White Dwarfs must be degenerate in the sense already indicated.

The equations governing the equilibrium of the White Dwarfs are therefore (in the standard notation) :

$$\frac{1}{r^2}\,\frac{d}{dr}\left(\frac{r^2}{\rho}\,\frac{dP}{dr}\right) = -\,4\,\pi\,G\,\rho \tag{7}$$

where P and ρ are related according to equation (6).
By the transformations

$$r = \left(\frac{2A}{\pi G}\right)^{1/2} \frac{1}{B y_0}\,\eta\,; \qquad y = y_0\,\Phi \tag{8}$$

where

$$y_0^2 = x_0^2 + 1 \tag{9}$$

equation (7) reduces to

$$\frac{1}{\eta^2}\frac{d}{d\eta}\left(\eta^2 \frac{d\Phi}{d\eta}\right) = -\left(\Phi^2 - \frac{1}{y_0^2}\right)^{3/2} \tag{10}$$

Equation (10) has to be solved with the boundary conditions

$$\Phi = 1\,; \quad \frac{d\Phi}{d\eta} = 0 \quad \text{at} \quad \eta = 0 \tag{11}$$

For each specified value of y_0 we have one such solution. The boundary is defined at the point where the density vanishes, and this by (6) and (9) means that if η_1 specified the boundary,

$$\Phi(\eta_1) = 1/y_0 \tag{12}$$

The integrations for the function Φ have been carried out for ten different values of y_0 and the physical characteristics of the resulting configuration are shown in Table 1.

The most important characteristic of these configurations is that they possess a natural limit, i. e. as

$$y^0 \rightarrow \infty \tag{13}$$

TABLE I.

THE PHYSICAL CHARACTERISTICS OF COMPLETELY DEGENERATE CONFIGURATIONS

$1/y_0^2$	$M/\odot$	ρ_0 in grams per cubic centimeter	ρ mean in grams per cubic centimeter	Radius in centimeters
0	5,75	∞	∞	0
0,01	5,51	$9{,}85 \times 10^8$	$3{,}70 \times 10^7$	$4{,}13 \times 10^8$
0,02	5,32	$3{,}37 \times 10^8$	$1{,}57 \times 10^7$	$5{,}44 \times 10^8$
0,05	4,87	$8{,}13 \times 10^7$	$5{,}08 \times 10^6$	$7{,}69 \times 10^8$
0,1	4,33	$2{,}65 \times 10^7$	$2{,}10 \times 10^6$	$9{,}92 \times 10^8$
0,2	3,54	$7{,}85 \times 10^6$	$7{,}9 \times 10^5$	$1{,}29 \times 10^9$
0,3	2,95	$3{,}50 \times 10^6$	$4{,}04 \times 10^5$	$1{,}51 \times 10^9$
0,4	2,45	$1{,}80 \times 10^6$	$2{,}29 \times 10^5$	$1{,}72 \times 10^9$
0,5	2,02	$9{,}82 \times 10^5$	$1{,}34 \times 10^5$	$1{,}93 \times 10^9$
0,6	1,62	$5{,}34 \times 10^5$	$7{,}7 \times 10^4$	$2{,}15 \times 10^9$
0,8	0,88	$1{,}23 \times 10^5$	$1{,}92 \times 10^4$	$2{,}79 \times 10^9$
1,0	0	0	0	∞

(The values given in this table differ slightly from the published values (S. Chandrasekhar, *M. N. R. A. S.*, 95, 208, 1935, Table III). The difference is due to the change in the accepted value of the fundamental physical constants. The calculations are for $\mu_e = 1$. For other values of μ_e, M should be multiplied by μ_e^{-2}, R by μ_e^{-1}, and ρ_0 by μ.)

and

$$\Phi \rightarrow \theta_3 \tag{14}$$

(where θ_3 is the Lane-Emden function of index 3), and the mass tends to a finite limit M_3. Numerically it is found that

$$M_3 = 5.75\ \mu_e^{-2} \odot \tag{15}$$

A glance at Table I shows that the mean density, the mass, and the radius of these degenerate configurations are all of the right order of magnitude to provide the basis for the theoretical discussion of the White Dwarfs. However, a really satisfactory test of the theory will be capable of providing an observational basis for the existence of a mass such that as we approach it the mean density increases several times, even for a slight increase in mass. The observational evidence which supports this theoretical prediction is discussed in Dr. Kuiper's report to this Colloquium.

3. The Mass $\mathfrak{M} = M_3\, \beta_\omega^{-3/2}$, $(\beta_\omega = 0.908 \ldots)$.

Closely connected with the existence of M_3 are the circumstances which enable us to find an upper limit to the mass of a stellar configuration which, consistent with the physics of degenerate matter, can be regarded as wholly degenerate. This limit arises in the following way :

Consider an electron assembly of N electrons in a volume V at temperature T. Then, on the basis of the perfect gas law, the electron pressure p_e would be given by

$$p_e = \left(\frac{N}{V}\right) kT \tag{16}$$

At temperature T we also have the radiation pressure of amount given by the Stefan-Boltzmann law :

$$p_r = 1/3 \quad a \quad T^4 \tag{17}$$

where the radiation constant a is given by

$$a = \frac{8\,\pi^5\, k^4}{15\ h^3 c^3} \tag{18}$$

Let us denote by P the total pressure $(= p_e + p_r)$ and introduce a parameter β_e, defined as follows :

$$P = p_r + p_e = \frac{1}{\beta_e} \, p_e = \frac{1}{1 - \beta_e} \, p_r \qquad (19)$$

Eliminating T between the relations (16) and (17), we find

$$p_e = \left[k^4 \frac{3}{a} \frac{1 - \beta_e}{\beta_e} \right]^{1/3} n^{4/3} \qquad (20)$$

where we have used n for (N/V). Let

$$n = \frac{8\pi \, m^3 c^3}{3h^3} \, x^3 \qquad (21)$$

Then, *equation* (20) *can be transformed into*

$$p_e = \frac{\pi m^4 c^5}{3h^3} \left(\frac{512 \, \pi \, k^4}{h^3 c^3 a} \, \frac{1 - \beta_e}{\beta_e} \right)^{1/3} 2x^4 \qquad (22)$$

or, using (18), we have

$$p_e = A \left(\frac{960}{\pi^4} \, \frac{1 - \beta_e}{\beta_e} \right)^{1/3} 2x^4 \qquad (23)$$

where A is defined as in equation (6″).

Now for an assembly having the same number N of electrons in the volume V we can formally calculate the electron pressure that would be given by the degenerate formula, namely

$$p_{deg} = A \, f(x) \qquad (24)$$

for f (x) as defined in (6) it is readily shown that

$$f(x) < 2x^4 \qquad (x < \infty) \qquad (25)$$

Hence, comparing (23) and (24), we have the result that if for a prescribed N and T, the value of β_e calculated on the basis of the perfect gas equation (16) be such that

$$\frac{960}{\pi^4} \, \frac{1 - \beta_e}{\beta_e} \geqslant 1 \qquad (26)$$

then under these circumstances matter can *never* become degenerate. Inequality (26) is readily seen to be equivalent to

$$\beta_e < 0.90788\ldots = \beta_\omega \ \text{(say)} \qquad (27)$$

Let us now "introduce" radiation into the completely degenerate configurations. If we consider a degenerate configuration built on the standard model, then

$$P = \beta_e^{-1} p_e \tag{28}$$

where p_e is given by (6). For such configuration we easily find that

$$M(\beta_e\,;\, y_0) = M(1\,;\, y_0)\,\beta_e^{-3/2} \tag{29}$$

in an obvious notation In particular

$$M(\beta_e\,;\,\infty) = M_3\,\beta_e^{-3/2} \tag{30}$$

From (30) it would appear at first sight that by allowing $\beta_e \to 0$ we can obtain degenerate configurations of any mass. This is, however, incorrect. For if $\beta_e < \beta_\omega$ then matter cannot be regarded as degenerate. Hence the maximum mass of the configuration which can be regarded as degenerate is given by

$$\mathfrak{M} = M_3\;\beta_\omega^{-3/2} \tag{31}$$

The result just stated is extremely general and can be proved as follows :

Consider a completely degenerate configuration of mass M slightly less than M_3. The density will everywhere be so great that we can increase the radiation pressure from zero to a value only slightly less than $(1\text{-}\beta_\omega)$ at each point of the configuration and still regard the matter as degenerate. The mass of the new configuration so obtained will be approximately $M\beta_\omega^{-3/2}$. When $M \to M_3$ the result becomes exact.

We have thus proved that the maximum mass of a stellar configuration which, consistent with the physics of degenerate matter, can be regarded as wholly degenerate, is $M_3\;\beta_\omega^{-3/2}$. Numerically,

$$\mathfrak{M} = 6.65\,\mu_e^{-2}\;\odot \tag{32}$$

3. Evolutionary Significance of M_3 and of $\mathfrak{M}$.

It should be clear now that a discussion of the role which the White Dwarfs are likely to play in any theory of stellar evolution must be necessarily linked with the evolutionary significance which we attach to the two critical masses M_3 and $\mathfrak{M}$.

For stars of mass less than M_3 we can tentatively assume that the completely degenerate state represents the last stage in the evolution of the stars — the stage of complete darkness and extinction. These completely degenerate configurations with $M < M_3$ are of course characterized by the finite radii.

For $M > M_3$ no such simple interpretation is possible. The problem that we are faced with can be stated as follows :

Consider a star of mass greater than $\mathfrak{M}$ and suppose that it has exhausted all its sources of subatomic energy — hydrogen in this connection. The star must then contract according to the Helmholtz-Kelvin time scale. Since degeneracy cannot set in, in the interior of such stars, continued and unrestricted contraction is possible, in theory.

However, we may expect instability of one kind or another (e. g. rotational) to set in long before, resulting in the " explosion " of the star into smaller fragments. It is also conceivable that the star may decrease its mass below M_3 by a process of continual ejection of matter. The Wolf Rayet phenomenon is suggestive in this connection.

For stars with masses in the range $M_3 < M < \mathfrak{M}$ there exist other possibilities. During the contractive stage, such stars are likely to develop degenerate cores. If the degenerate cores attain sufficiently high densities (as is possible for these stars) the protons and electrons will combine to form neutrons. This would cause a sudden diminution of pressure resulting in the collapse of the star onto a neutron core giving rise to an enormous liberation of gravitational energy. This may be the origin of the Supernova phenomenon.

The above remarks on the evolutionary significance of M_3 and $\mathfrak{M}$ are made with due reserve and no definiteness is claimed for them.

4. The Rotation of the White Dwarfs.

The effects of rotation on the structure of White Dwarfs are likely to be of considerable importance in connection with the remarks made in the preceding section. The theory of such rotationally distorted White Dwarfs can be developed on lines analogous to the authors's earlier investigations on the theory of distorted polytropes.

It can be shown that the structure of such rotating White Dwarfs will be governed by the differential equation

$$\frac{1}{\eta^2}\frac{\partial}{\partial\eta}\left(\eta^2\frac{\partial\Phi}{\partial\eta}\right)+\frac{1}{\eta^2}\frac{\partial}{\partial\mu}\left((1-\mu^2)\frac{\partial\Phi}{\partial\mu}\right)=\left(\Phi^2-\frac{1}{y_0^2}\right)^{3/2}+V \quad (33)$$

where

$$y=y_o\,\Phi\,;\quad y^2=x^2+1 \qquad (34)$$

(y_o referring to the central value) and $\mu = \cos\theta$ (θ being the polar angle). Further

$$v=\frac{\omega^2}{2\pi\,GBy_o^2} \qquad (35)$$

For small v (i. e. for small rotational velocities) the solution of (33) is found to be

$$\Phi=\Phi(y_o)+v\left\{\psi_o-\frac{5}{6}\frac{\eta_1^2\psi_2}{3\psi_2(\eta_1)+\eta_1\psi_2'(\eta_1)}P_2(\mu)\right\} \quad (36)$$

where η_1 refers to the boundary of $\Phi(y_0)$ and ψ_o and ψ_2 satisfy the differential equations

$$\frac{1}{\eta_2}\frac{d}{d\eta}\left(\eta^2\frac{d\psi_o}{d\eta}\right)=-3\Phi\left(\Phi^2-\frac{1}{y_o^2}\right)^{1/2}\psi_o+1 \quad (37)$$

$$\frac{1}{\eta_2}\frac{d}{d\eta}\left(\eta^2\frac{d\psi_2}{d\eta}\right)=\left\{-3\Phi\left(\Phi^2-\frac{1}{y_o^2}\right)^{1/2}+\frac{6}{\eta_2}\right\}\psi_2 \quad (38)$$

and P_2 is the second Legendre plynomial in μ. The equation of the boundary of the rotating configuration is given by

$$\eta=\eta_1+\frac{v}{|\Phi'|}\left[\psi_o(\eta_1)-\frac{5}{6}\frac{\eta_1^2\psi_2(\eta_1)P_2(\mu)}{3\psi_2(\eta_1)+\eta_1\psi'(\eta_1)}\right] \quad (39)$$

giving for the oblateness of the configuration the expression

$$\sigma=\frac{5}{4}\frac{v}{|\Phi'^1|}\frac{\eta_1\psi_2(\eta_1)}{3\psi_2(\eta_1)+\eta_1\psi_2'(\eta_1)} \qquad (40)$$

The discussion of these results must await the integration of the equations (37) and (38).

1. Detailed derivations of the formulae quoted will be found in the author's monograph « An Introduction to the Study of Stellar Structure », Chapter X. More recently more rigorous derivations of the same results have been given by D. van Dantzig.

Discussion of Dr. Chandrasekhar's Communication.

Professor Russell and Dr. Chandrasekhar consider a star whose mass is smaller than that of a White Dwarf, of the order of 0,20 or less. Such a star would become less luminous as it approached degeneracy, but the distribution of its density and temperature would be the same as for a White Dwarf. Its radius would be larger than for stars of larger mass, and, for the same surface temperature it would be brighter than a normal White Dwarf We could then explain the region mentioned by Dr. Kuiper where no objects are observed.

Sir Arthur Eddington states that White Dwarfs do not have any particular luminosity anyway until their radii reached rather large values ; the larger the luminosity, the greater the expenditure of energy, and the shorter the life.

Professor Russell agrees that the star would pass rapidly through such a stage, and would show a much larger surface temperature and bolometric magnitude, although its visual magnitude would not be very much larger.

Dr. Strömgren brings up the question of a star on the verge of becoming degenerate. Dr. Chandrasekhar says that one of his pupils and Dr. Stoner are working on the problem of what happens in a region in which the equation of state of degenerate matter approaches the equation of state of a perfect gas. No conclusions can be drawn yet, since the computations are incomplete. If degeneracy is just beginning to appear in the center, the radius is three times that of a completely degenerate star. Since the matter is almost wholly gaseous, the luminosity is high, and with such a small radius the star goes to a remote portion of the Hertzsprung-Russell diagram.

Sir Arthur wishes to know if Dr. Chandrasekhar's statement that the total pressure is the sum of the gaseous pressure and the degenerate pressure is true or if it is merely a simplification. Dr. Chandrasekhar shows that it is true.

Answering a question of Dr. Atkinson's, Dr. Chandrasekhar states that in an incompletely degenerate star the degenerate nucleus must have a mass less than $\mathfrak{M}$, but that the mass of the gaseous portion may be considerable. In practice, however, a degenerate nucleus cannot exist unless the total mass of the star is less than $\mathfrak{M}$. The matter has been studied in detail for various models.

Professor Russell requests a value of the density at which the equation of state for degeneracy differs appreciably from that of a perfect gas. Dr. Chandrasekhar says that the perfect gas laws would hold up to densities of 10^3 or 10^4; in Sirius B it is 1700, with a temperature of $10^{7\circ}$. Then, suggests Professor Russell, the perfect gas laws would hold for an ordinary red dwarf.

Sir Arthur Eddington and Dr. Chandrasekhar answer that their temperature is only $10^{6\circ}$, so that ionization would not be complete, and there would be corrections. In some cases other types of degeneracy exist where the ionization is high. Dr. Kuiper adds that for red dwarfs the density would be nearer 150 than 1000, as Dr. Chandrasekhar quoted, and that furthermore the temperatures are probably too low. Little is known for M dwarfs later than M2; the evidence of their temperatures indicates that as we go from M0 to M6 the density stays nearly constant, while the central temperatures decrease. Sir Arthur Eddington questions the adopted effective temperatures. In the case of Wolf 359, says Dr. Kuiper, the large color index of 5, measured as far as 8500 A, is due to the great intensity of the titanium bands. Professor Russell emphasizes the importance of observations on the spectral type and the color index of eclipsing variables in this connection, particularly those with high density, but Dr. Kuiper has some doubts on the feasibility of such observations, since there are only a few such stars with great enough luminosity to permit spectroscopic observation.

THE INTERNAL CONSTITUTION OF THE STARS

S. CHANDRASEKHAR

Yerkes Observatory, University of Chicago

(*Read February 17, 1939, in Symposium on Progress in Astrophysics*)

In this paper an attempt is made to describe the general methods that have been developed to determine the physical conditions in stellar interiors. In view of the complexity of the problem, it is of value to consider, in the first instance, only those methods which involve the minimum of assumptions. Three such methods have been developed in recent years. They are

I. The method of the integral theorems.
II. The method of the homologous transformations.
III. The method of stellar envelopes.

I. The Method of the Integral Theorems

In this method the fundamental assumption is made that stars are in hydrostatic equilibrium.[1] We should then have

$$\frac{dP}{dr} = -\frac{GM(r)}{r^2}\rho, \tag{1}$$

where P is the total pressure and ρ the density at a distance r from the center and $M(r)$ is the mass enclosed inside r. Since P is the total pressure we can write

$$P = p_r + p_g, \tag{2}$$

where p_g is the gas pressure and p_r the radiation pressure; *i.e.*,

$$p_r = \tfrac{1}{3}aT^4; \qquad p_g = f(\rho, T), \tag{3}$$

where a is the Stefan-Boltzmann constant, and $f(\rho, T)$ specifies the equation of state. If the assumption is made that stellar material behaves like a perfect gas then

$$p_g = \frac{k}{\mu H}\rho T, \tag{4}$$

where k is the Boltzmann constant, μ the mean molecular mass, and H the mass of the hydrogen atom.

[1] We thus exclude from our considerations rotating and variable stars.

From the geometry of the case we have in addition to (1)

$$\frac{dM(r)}{dr} = 4\pi r^2 \rho. \tag{5}$$

Using (5) we can rewrite (1) as

$$dP = -\frac{1}{4\pi}\frac{GM(r)dM(r)}{r^4}. \tag{6}$$

A direct integration of (6) yields [2]

$$P_c = \frac{1}{4\pi}\int_0^R \frac{GM(r)dM(r)}{r^4}, \tag{7}$$

where R is the radius of the star and P_c is the central pressure. From (7) we immediately infer that

$$P_c > \frac{G}{4\pi}\int_0^R \frac{M(r)dM(r)}{R^4} = \frac{1}{8\pi}\frac{GM^2}{R^4}. \tag{8}$$

Hence with no assumption except that stars are in hydrostatic equilibrium we can set a lower limit to the central pressure. Numerically (8) reduces to

$$P_c > 4.50 \times 10^8 \left(\frac{M}{\odot}\right)^2 \left(\frac{R_\odot}{R}\right)^4 \text{ atmospheres,} \tag{9}$$

where $\odot$ and $R_\odot$ refer to the mass and the radius of the sun.

We can similarly obtain a lower limit to the mean pressure defined by

$$M\bar{P} = \int_0^R PdM(r). \tag{10}$$

After an integration by parts (10) reduces to

$$M\bar{P} = -\int_0^R M(r)dP, \tag{11}$$

or using (6)

$$M\bar{P} = \frac{G}{4\pi}\int_0^R \frac{M^2(r)dM(r)}{r^4}. \tag{12}$$

From (12) we obtain without difficulty [3] that

$$\bar{P} > \frac{1}{12\pi}\frac{GM^2}{R^4} \tag{13}$$

[2] Using the boundary condition that $P = 0$ at $r = R$ the radius of the star.
[3] By replacing r by R and taking it outside the integral sign.

or numerically

$$\bar{P} > 3.0 \times 10^8 \left(\frac{M}{\odot}\right)^2 \left(\frac{R_\odot}{R}\right)^4 \text{ atmospheres.} \tag{14}$$

In other words, we can expect pressures of the order of 10^9 atmospheres in the stellar interiors.

We can get somewhat sharper inequalities if we supplement our assumption of hydrostatic equilibrium by another one, namely, that the mean density $\bar{\rho}(r)$ interior to r does not increase outward. This assumption implies that

$$\bar{\rho}(r) \geqslant \rho(r). \tag{15}$$

It will be noticed that we do not exclude completely the possibility of negative density gradients. We only insist that the actual density $\rho(r)$ at any point does not exceed the mean density $\bar{\rho}(r)$ interior to the point considered.

To establish inequalities for P_c etc., we consider the following expression:

$$I_{\sigma,\nu}(r) = \int_0^r \frac{GM^\sigma(r)dM(r)}{r^\nu}. \tag{16}$$

We further assume that

$$3(\sigma + 1) - \nu > 0. \tag{17}$$

By the definition of the mean density $\bar{\rho}(r)$ we have

$$\bar{\rho}(r) = M(r)/\tfrac{4}{3}\pi r^3. \tag{18}$$

From (18) we derive that

$$r^\nu = \left\{\left(\frac{3}{4\pi}\right)\frac{M(r)}{\bar{\rho}(r)}\right\}^{\nu/3}. \tag{19}$$

Substituting (19) in (16) we have

$$I_{\sigma,\nu}(r) = G\left(\frac{4\pi}{3}\right)^{\nu/3} \int_0^r \bar{\rho}^{\nu/3}(r) M^{(3\sigma-\nu)/3}(r)dM(r). \tag{20}$$

Since according to our assumption $\bar{\rho}(r)$ does not increase outward we get an upper bound for the integral on the right hand side of (20) by replacing $\bar{\rho}(r)$ by ρ_c and taking it outside the integral sign. In the same we get a lower bound by replacing $\bar{\rho}(r)$ by its value at r[4] and taking it outside the integral sign. We obtain in this way

$$\frac{3G}{3(\sigma + 1) - \nu}\left(\frac{4\pi}{3}\right)^{\nu/3} M^{(3\sigma+3-\nu)/3}(r)\bar{\rho}^{\nu/3}(r) \leqslant I_{\sigma,\nu}(r)$$
$$\leqslant \frac{3G}{3(\sigma + 1) - \nu}\left(\frac{4\pi}{3}\right)^{\nu/3} M^{(3\sigma+3-\nu)/3}(r)\rho_c^{\nu/3}. \tag{21}$$

[4] r here refers to the upper bound of the integral defining $I_{\sigma,\nu}(r)$.

(21) is a fundamental inequality from which several results of importance can be derived.

From (6) we derive that

$$P_c - P = \frac{1}{4\pi} I_{1,4}(r). \tag{22}$$

From (21) we now infer that

$$\frac{1}{2}\left(\frac{4\pi}{3}\right)^{1/3} GM^{2/3}(r)\bar{\rho}^{4/3}(r) \leqslant P_c - P \leqslant \frac{1}{2}\left(\frac{4\pi}{3}\right)^{1/3} GM^{2/3}(r)\rho_c^{4/3}. \tag{23}$$

If we put $r = R$ in the above inequality, we find

$$\frac{1}{2}\left(\frac{4\pi}{3}\right)^{1/3} GM^{2/3}\bar{\rho}^{4/3} \leqslant P_c \leqslant \frac{1}{2}\left(\frac{4\pi}{3}\right)^{1/3} GM^{2/3}\rho_c^{4/3}. \tag{24}$$

The left hand part of the above inequality can be rewritten as

$$P_c \geqslant \frac{3}{8\pi}\frac{GM^2}{R^4} \tag{25}$$

or numerically

$$P_c > 1.35 \times 10^9 \left(\frac{M}{\odot}\right)^2 \left(\frac{R_\odot}{R}\right)^4 \quad \text{atmospheres.} \tag{26}$$

(26) improves the earlier inequality (9) by a factor 3.

If we put $r = R$ in (21) we obtain

$$\frac{3}{3\sigma + 3 - \nu}\frac{GM^{\sigma+1}}{R^\nu} \leqslant I_{\sigma,\nu} \leqslant \frac{3}{3\sigma + 3 - \nu}\frac{GM^{\sigma+1}}{r_c^\nu}, \tag{27}$$

where r_c is defined by

$$\frac{4}{3}\pi r_c^3 \rho_c = M. \tag{28}$$

From (12) we see that

$$M\bar{P} = \frac{1}{4\pi} I_{2,4}. \tag{29}$$

Hence from (27) we derive that

$$\frac{3}{20\pi}\frac{GM^2}{R^4} \geqslant \bar{P} \geqslant \frac{3}{20\pi}\frac{GM^2}{r_c^4}. \tag{30}$$

The right hand part of the inequality (30) reduces to

$$\bar{P} > 5.4 \times 10^8 \left(\frac{M}{\odot}\right)^2 \left(\frac{R_\odot}{R}\right)^4 \quad \text{atmospheres.} \tag{31}$$

Remembering that the potential energy Ω and the mean value of gravity

$\bar{g}$ are given by

$$-\Omega = G\int_0^R \frac{M(r)dM(r)}{r} = I_{1,1}, \tag{32}$$

and

$$M\bar{g} = G\int \frac{M(r)dM(r)}{r^2} = I_{1,2}, \tag{33}$$

we have [again using (27)] that

$$\frac{3}{5}\frac{GM^2}{R} \leqslant -\Omega \leqslant \frac{3}{5}\frac{GM^2}{r_c}, \tag{34}$$

and

$$\frac{3}{4}\frac{GM}{R^2} \leqslant \bar{g} \leqslant \frac{3}{4}\frac{GM}{r_c^2}. \tag{35}$$

The inequality for Ω further enables us to set a lower limit to the mean temperatures of gaseous stars in which radiation pressure can be neglected.[5] We define the mean temperature $\bar{T}$ by

$$M\bar{T} = \int_0^R TdM(r). \tag{36}$$

If the radiation pressure can be neglected

$$P = p_g = \frac{k}{\mu H}\rho T. \tag{37}$$

Hence from (36) and (37) we have

$$M\bar{T} = \frac{\mu H}{k}\int_0^R \frac{P}{\rho}dM(r) \tag{38}$$

$$= \frac{\mu H}{k}\int_0^R PdV, \tag{38'}$$

where dV is the volume element. However, according to a well known theorem in potential theory

$$-\Omega = 3\int_0^R PdV. \tag{39}$$

Hence combining (34), (38′) and (39) we find

$$\frac{1}{5}\frac{\mu H}{k}\frac{GM}{R} \leqslant \bar{T} \leqslant \frac{1}{5}\frac{\mu H}{k}\frac{GM}{r_c}. \tag{40}$$

The left hand side of the above inequality reduces to

$$\bar{T} \geq 4.6 \times 10^6 \mu \frac{M}{\odot}\frac{R_\odot}{R} \quad \text{degrees.} \tag{41}$$

[5] As we shall see below this is justifiable for the stars of normal mass.

In other words we can expect temperatures of the order of a few million degrees in stellar interiors.

The physical content of the results (24), (27), (34), (35) and (40) is the following: We are given a certain equilibrium configuration of mass M and radius R with an arbitrary density distribution, arbitrary except for the conditions that $\bar{\rho}(r)$ does not increase outward. From the given configuration we can construct two other configurations of uniform density—one with a constant density equal to $\bar{\rho}$ and the other with a constant density equal to ρ_c (see Fig. 1). The radii of these two

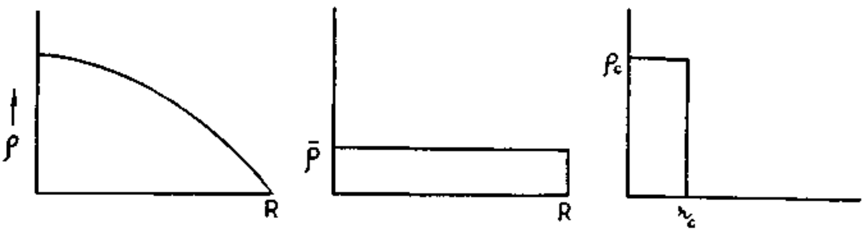


FIG. 1.

configurations are clearly R and r_c, respectively. What we have shown is that the physical variables characterizing the given equilibrium configuration, namely, P_c, $\bar{P}$, $-\Omega$, $\mathcal{g}$ and $\bar{T}$ (for the case of negligible radiation pressure) have values respectively less than those for the configuration of uniform density with $\rho = \rho_c$ and respectively greater than those for the configuration of uniform density with $\rho = \bar{\rho}$. Thus the given configuration is, in this sense, intermediate between the two configurations of uniform density with $\rho = \rho_c$ and $\rho = \bar{\rho}$ respectively.

Now one of the quantities which is of considerable importance in the discussion of the physical conditions in stellar interiors is the relative importance of the radiation pressure. This can be conveniently measured by the fraction $(1 - \beta)$ defined according to the relations:

$$\left.\begin{aligned}(1-\beta)P &= p_r = \frac{1}{3}aT^4,\\ \beta P &= p_g = \frac{k}{\mu H}\rho T.\end{aligned}\right\} \tag{42}$$

By a simple elimination of T between the equations (42) we obtain

$$P = \left[\left(\frac{k}{\mu H}\right)^4 \frac{3}{a}\frac{1-\beta}{\beta^4}\right]^{1/3} \rho^{4/3}. \tag{43}$$

Hence

$$P_c = \left[\left(\frac{k}{\mu_c H}\right)^4 \frac{3}{a}\frac{1-\beta_c}{\beta_c^4}\right]^{1/3} \rho_c^{4/3}. \tag{44}$$

From (24) and (44) we now have

$$\left[\left(\frac{k}{\mu_c H}\right)^4 \frac{3}{a} \frac{1-\beta_c}{\beta_c^4}\right]^{1/3} \leqslant \left(\frac{\pi}{6}\right)^{1/3} GM^{2/3}, \tag{45}$$

or

$$M \geqslant \left(\frac{6}{\pi}\right)^{1/2} \left[\left(\frac{k}{\mu_c H}\right)^4 \frac{3}{a} \frac{1-\beta_c}{\beta_c^4}\right]^{1/2} \frac{1}{G^{3/2}}. \tag{47}$$

Define $(1 - \beta^*)$ by the equation

$$M = \left(\frac{6}{\pi}\right)^{1/2} \left[\left(\frac{k}{\mu_c H}\right)^4 \frac{3}{a} \frac{1-\beta^*}{\beta^{*4}}\right]^{1/2} \frac{1}{G^{3/2}}. \tag{48}$$

Combining (47) and (48) we obtain

$$\frac{1-\beta^*}{\beta^{*4}} \geqslant \frac{1-\beta_c}{\beta_c^4}. \tag{49}$$

Hence we should have

$$1 - \beta_c \leqslant 1 - \beta^*. \tag{50}$$

The inequality (50) shows that for a gaseous star the value of $(1 - \beta)$ at the center cannot exceed an amount depending on the mass of the star only. Table 1 gives the value of $(1 - \beta^*)$ for different values of the mass M. As an example of the application of Table 1, we see that for the sun, $1 - \beta_c < 0.03$ while for Capella ($M = 4.18\odot$), $1 - \beta_c < 0.2$ assuming in both cases that $\mu_c = 1$. Table 1 clearly illustrates that for the normal stars the radiation pressure as a factor in the equation of hydrostatic equilibrium can be neglected.

TABLE 1

VALUES OF $(1 - \beta^*)$

$1-\beta^*$	$\frac{M}{\odot}\mu_c^2$	$1-\beta^*$	$\frac{M}{\odot}\mu_c^2$	$1-\beta^*$	$\frac{M}{\odot}\mu_c^2$
0.01	0.56	0.14	2.77	0.35	7.67
.02	0.81	.15	2.94	.40	9.62
.03	1.01	.16	3.11	.45	12.15
.04	1.19	.17	3.28	.50	15.49
.05	1.36	.18	3.46	.55	20.06
.06	1.52	.19	3.64	.60	26.52
.07	1.68	.20	3.83	.65	36.05
.08	1.83	.21	4.02	.70	50.92
.09	1.98	.22	4.22	.75	75.89
.10	2.14	.23	4.43	.80	122.5
.11	2.29	.24	4.65	.85	224.4
.12	2.45	.25	4.87	.90	519.6
0.13	2.61	0.30	6.12	1.00	∞

A problem of importance that has not been considered so far is the question of the minimum central temperature of stars. In solving this problem we assume (1) that the density decreases outward and (2) that the temperature decreases outward. In considering this problem we regard T as a known function of P and ρ. This problem of the minimum central temperature leads to the consideration of the following more general problem.

Given that ρ and a certain function $F(P, \rho)$ both do not increase outward, what is the minimum value of $F_c \equiv F(P_c, \rho_c)$ for an equilibrium configuration of known mass M and radius R? This problem can be solved if suitable restrictions are imposed on F. We can, in fact, prove the following.[6]

Let $F(P, \rho)$ be such that

$$2\frac{\partial \log F}{\partial \log P} + \frac{\partial \log F}{\partial \log \rho} > 0 \qquad (P, \rho > 0), \tag{51}$$

and

$$\frac{\partial F}{\partial P} > 0. \tag{52}$$

Then in any equilibrium configuration of prescribed mass and radius in which both F and ρ do not increase outward the minimum value of F_c is attained in the sequence of equilibrium configurations which consist of cores of constant F and homogeneous envelopes.

If we choose for F the form

$$F = P\rho^{\delta}, \tag{53}$$

then the condition (51) implies that

$$\delta > -2. \tag{54}$$

Hence an immediate consequence of the theorem stated is the following: In any equilibrium configuration of prescribed mass and radius in which both ρ and $K = P/\rho^{(n+1)/n}$, $(n > 1)$, do not increase outward the minimum value of K_c is attained in the sequence of equilibrium configurations which consist of polytropic cores of index n and homogeneous envelopes.[7] It will be noticed that if we put $n = \infty$ in this special theorem we have a means of finding the lower limit to the central temperatures of stars with negligible radiation pressure. For the general problem however, we should use the theorem in its more general form. We have

$$P = \frac{k}{\mu H}\rho T + \frac{1}{3}aT^4. \tag{55}$$

[6] The details of the proof the theorem stated will be published by the author elsewhere.
[7] This special theorem has been proved by the author, see *Ap. J.*, 87, 535, 1938.

Equation (55) when regarded as implicitly determining T as a function of P and ρ specifies the function F. It is easily verified that $T(P, \rho)$ satisfies the restrictions (51) and (52). Consequently, for a star of given mass and radius we have to construct the sequence of equilibrium configurations consisting of isothermal cores and homogeneous envelopes and determine the minimum central temperature along this sequence. The theorem now asserts that this would then give the absolute minimum for T_c under the restrictions imposed. The numerical work required to determine the minimum central temperatures for stars of different masses is rather long and tedious. Only the results will be quoted:

It is found that we can write

$$T_c \geqslant \frac{1}{2} Q(M) \frac{\mu H}{k} \frac{GM}{R}, \tag{56}$$

where Q is factor which depends in a complicated way on M and μ. The following table gives the values of $Q(M)$ for certain values of M.

TABLE 2

$\frac{M}{\odot}\mu^2$	$Q(M)$
3.4	0.61
4.5	0.595
5.75	0.575
7.1	0.554
19.4	0.421

(56) can be written also as:

$$T_c \geqslant 11.5\mu Q(M) \frac{M}{\odot} \frac{R_\odot}{R} \times 10^6 \text{ degrees.} \tag{57}$$

It is found that

$$Q \to 0.64 \qquad \text{as} \qquad M \to 0, \tag{58}$$

and

$$Q \to \beta^* \qquad \text{as} \qquad M \to \infty. \tag{59}$$

From Table 2 and the inequality (57) we see that for the sun $T_c > 7.4 \times 10^6$ degrees, while for Capella $T_c > 1.8 \times 10^6$ degrees. These results again show that we can expect temperatures of at least a few million degrees in stellar interiors. Though we have found only the minimum values it is clear that the actual values must be of the same order. They may differ from the minimum values by a factor of the order 10 but we cannot expect values of an entirely different order of magnitude.

There is one other application of the theorem we are considering which is of some importance. If we put $\delta = -4/3$ in (53) then the theorem will enable us to set a *lower* limit to $(1 - \beta_c)$ for stars in which ρ

and $(1 - \beta)$ do not increase outward. The analysis of the appropriate composite configurations leads to the surprising result that the minimum value of $(1 - \beta_c)$ is the constant value which we would ascribe to a star of the same mass if it were a complete polytrope of index $n = 3$. Now a stellar model which has played a conspicuous role in the development of the subject of stellar interiors is the so-called *standard model* in which $(1 - \beta)$ is *assumed* to be a constant. The use of this model has been criticized from several directions. For this reason it is important to realize that apart from other considerations the standard model has a definite value in so far as it has a maximal (or minimal as the case may be) characteristic. We shall return later to the physical meaning of the assumption that $(1 - \beta)$ does not increase outward.

If we now assume that stars are in radiative equilibrium then the *radiative temperature gradient* is determined by

$$\frac{dp_r}{dr} = -\frac{\kappa L(r)}{4\pi c r^2}\rho, \tag{60}$$

where κ is the stellar opacity coefficient and $L(r)$ is the amount of energy in ergs per second crossing the spherical surface of radius r. Further in (60) c is the velocity of light and the other symbols have their usual meanings. From (1) and (60) we derive that

$$\frac{dp_r}{dP} = \frac{L}{4\pi cGM}\kappa\eta, \tag{61}$$

where

$$\eta = \frac{L(r)/M(r)}{L/M} = \frac{\bar{\epsilon}(r)}{\bar{\epsilon}}. \tag{62}$$

The physical meaning of η is that it is the ratio of the average rate of liberation of energy $\bar{\epsilon}(r)$ interior to the point r to the corresponding average $\bar{\epsilon}$ for the whole star.

Integrating (61) from $r = r$ to $r = R$ and using the boundary condition $p_r = 0$ at $r = R$ we have

$$p_r = \frac{L}{4\pi cGM}\overline{\kappa\eta}(r)P, \tag{63}$$

where

$$\overline{\kappa\eta}(r) = \frac{1}{P}\int_R^r \kappa\eta \, dP. \tag{64}$$

Equation (63) can be written alternatively in the form

$$L = \frac{4\pi cGM(1 - \beta)}{\overline{\kappa\eta}(r)}. \tag{65}$$

If we put $r = R$ in the above equation we get a fundamental relation which enables the evaluation of L in terms of an average value of $\kappa\eta$:

$$L = \frac{4\pi cGM(1 - \beta_c)}{\overline{\kappa\eta}}. \tag{66}$$

From equation (65) we conclude that the condition $(1 - \beta)$ not increasing outward (which we used earlier in the discussion) is equivalent to the assumption of $\overline{\kappa\eta}(r)$ not increasing outward. Alternatively we should have

$$\kappa\eta(r) \geqslant [\overline{\kappa\eta}]_R{}^r. \tag{67}$$

It is clear from (67) that we can actually allow a decrease of $\kappa\eta$ (within limits) as we approach the center. In actual stellar configurations η might be expected to decrease outward, but this will not be generally true of κ. For this reason it is important to realize that the minimal characteristic of the standard model was proved under restrictions which do not exclude the possibility of $\kappa\eta$ actually decreasing outward [within the limits set by (67)].

Equation (66) enables us to set an upper limit to a mean value of the stellar opacity. For if we combine (66) with the inequality (50) we obtain

$$\overline{\kappa\eta} \leqslant \frac{4\pi cGM(1 - \beta^*)}{L}. \tag{68}$$

If we assume that $\epsilon(r)$ does not increase outward then η will not increase outward and the minimum value of η is unity. Hence

$$\overline{\kappa\eta} \geqslant \bar{\kappa}, \tag{69}$$

where

$$\bar{\kappa} = \frac{1}{P_c}\int_0^{P_c} \kappa dP. \tag{70}$$

Hence we have

$$\bar{\kappa} < \frac{4\pi cGM(1 - \beta^*)}{L}. \tag{71}$$

If we apply (71) to the case of Capella ($M = 4.18\odot$, $L = 120L_\odot$) we find that

$$\bar{\kappa}_{\text{Capella}} < 100 \text{ gm}^{-1} \text{ cm}^2. \tag{72}$$

It should be noticed that our method of averaging weights the central regions of the configuration very heavily and hence the upper limit (71) is essentially an upper limit to the opacity at the center of the configuration. The physical meaning of the inequality (71) is this: If for a star of given mass M and luminosity L, $\bar{\kappa}$ should be greater than the limit set by (71) then either the density or the rate of generation of energy ϵ or both must increase outward in some finite regions of the interior of a star.

We thus see that no special assumptions are required to establish the orders of magnitude of the physical variables in stellar interiors.[8]

II. The Method of the Homologous Transformations

We have already seen that if we restrict ourselves to the considerations of stars of masses less than five to six times the solar mass then we can neglect the radiation pressure in the equation of hydrostatic equilibrium, *i.e.*,

$$P \backsimeq \frac{k}{\mu H} \rho T, \tag{73}$$

or in a somewhat better approximation we can write

$$P = \frac{k}{\bar{\beta}\mu H} \rho T, \tag{74}$$

where $\bar{\beta}$ is a certain average value of β inside the star. We assume that $\bar{\beta}$ is very nearly unity. Our two equations of equilibrium are:

$$\frac{k}{\bar{\beta}\mu H} \frac{d}{dr} (\rho T) = - \frac{GM(r)}{r^2} \rho, \tag{75}$$

$$\frac{dM(r)}{dr} = 4\pi r^2 \rho. \tag{76}$$

To make the system of equations governing the equilibrium of the star complete we need an additional equation to determine the temperature gradient. If radiative equilibrium obtains then [*cf.* equation (60)]

$$\frac{d}{dr} \left(\frac{1}{3} a T^4 \right) = - \frac{\kappa L(r)}{4\pi c r^2} \rho, \tag{77}$$

where it is clear from definition that

$$L(r) = \int_0^r 4\pi r^2 \rho \epsilon dr. \tag{78}$$

In (78) ϵ is the rate of generation of energy in ergs per second per gram of the substance. An alternative form for (77) can be obtained by combining equations (61) and (65). We have

$$\frac{dp_r}{dP} = (1 - \beta) \frac{\kappa \eta}{\overline{\kappa\eta}(r)} = \frac{p_r}{P} \frac{\kappa \eta}{\overline{\kappa\eta}(r)}, \tag{79}$$

[8] The references to the literature will be found in the author's monograph *An Introduction to the Study of Stellar Structure* (Chicago, 1939).

or in a somewhat different form:

$$\frac{dT}{T} = \frac{1}{4} \frac{\kappa\eta}{\overline{\kappa\eta}(r)} \frac{dP}{P}. \tag{80}$$

Now, according to certain well known methods of argument [9] an existing temperature is stable or unstable according as it is less than or greater than the corresponding adiabatic gradient. For an enclosure containing matter and radiation the condition of adiabacy is given by

$$dQ = d(aVT^4 + c_V T) + \left(\frac{1}{3} aT^4 + \frac{k}{\mu H} \rho T \right) dV = 0, \tag{81}$$

where c_V is the specific heat of the gas at constant volume. Equation (81) can be rewritten in the form [10]

$$\frac{dT}{T} = \frac{\Gamma_2 - 1}{\Gamma_2} \frac{dP}{P}, \tag{82}$$

where

$$\Gamma_2 = 1 + \frac{(4 - 3\beta)(\gamma - 1)}{\beta^2 + 3(\gamma - 1)(1 - \beta)(4 + \beta)}, \tag{83}$$

where γ is the ratio of the specific heats ($= 5/3$) for the gas and β has the same meaning as hitherto. The condition for the stability of the radiative gradient is therefore

$$4 \frac{\Gamma_2 - 1}{\Gamma_2} \geqslant \frac{\kappa\eta}{\overline{\kappa\eta}(r)}. \tag{84}$$

The radiative gradient becomes " unstable " if $\kappa\eta/\overline{\kappa\eta}(r)$ exceeds the value on the right hand side of the above inequality. The following table (Table 3) gives the values of $4(\Gamma_2 - 1)/\Gamma_2$ for different values of $1 - \beta$.

TABLE 3

$1-\beta$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$4(\Gamma_2 - 1)/\Gamma_2$ . . .	1.6	1.304	1.177	1.108	1.065	1.039	1.022	1.010	1.004	1.000	1.000

It follows from Table 3 that if the energy sources are concentrated towards the center then as we approach the central regions the radiative gradient will become unstable. For the case of vanishing radiation pressure, the nature of the steady state that will be set up if the radiative gradient becomes unstable has been investigated in detail

[9] See, for instance, the author's monograph on *An Introduction to the Study of Stellar Structure* (Chicago, 1939), pp. 222–227.

[10] See reference 9, pp. 55–59.

by Cowling and Biermann. These authors come to the conclusion that if

$$\frac{\kappa\eta}{\overline{\kappa\eta}(r)} \geqslant 1.6, \tag{85}$$

then the adiabatic gradient obtains:

$$p \propto T^{\gamma/(\gamma-1)}. \tag{86}$$

Now a rigorous attack on the equations of equilibrium (75), (76), (77) and (78) [11] is possible only if the dependence of ϵ and κ on ρ and T are known. Until recently we had no information concerning the nature of the dependence of ϵ on ρ and T. At the present time we have some notions on this [12] but even now our knowledge is by no means entirely satisfactory. It is for this reason that the analysis of stellar structure has depended so largely on the study of stellar models. It would appear at first sight that the uncertainty in the law of energy generation is a serious matter. We can however show that *for stars in which the radiation pressure is negligible throughout the configuration*

$$L = \text{Constant}\,\frac{1}{\kappa_0}\frac{M^{5+s}}{R^{s}}(\mu\bar{\beta})^{7+s}, \tag{87}$$

if the rate of generation of energy ϵ and the coefficient of opacity κ follow the laws

$$\epsilon = \epsilon_0\rho^{\alpha}T^{\nu}, \qquad \kappa = \kappa_0\rho T^{-3-s}, \tag{88}$$

where α, ν and s are arbitrary. The constant in (87) *depends on the exponents α, ν* and *s*.[13]

To prove this let us first consider stars in radiative equilibrium. Then the equations of equilibrium can be written as

$$\frac{dP}{dr} = -\frac{GM(r)}{r^2}\rho, \tag{89}$$

$$\frac{dM(r)}{dr} = 4\pi r^2\rho, \tag{90}$$

$$P = \frac{k}{\bar{\beta}\mu H}\rho T, \tag{91}$$

$$\frac{dT}{dr} = -\frac{3}{4ac}\kappa_0\rho^2T^{-6-s}\frac{\int_0^r \rho^{\alpha+1}T^{\nu}r^2dr}{r^2}. \tag{92}$$

[11] Or (86) depending upon whether (84) is satisfied or not.
[12] See H. N. Russell's article in this volume.
[13] This theorem is due to B. Strömgren.

The system of equations (89)–(92) has to be solved with the boundary conditions

$$M(r) = M, \qquad \rho = 0, \qquad T = 0 \qquad \text{at} \qquad r = R, \tag{93}$$

and

$$M(r) = 0 \qquad \text{at } r = 0. \tag{94}$$

These provide four boundary conditions and since the system of equations (89)–(92) is equivalent to a single differential equation of the fourth order, it follows that there is just exactly one solution which will satisfy the boundary conditions (93) and (94). We shall now show that from such a solution we can construct another solution such that it will describe another configuration with a different M, R, and $\bar{\beta}\mu$; we shall see, in fact, that the transformations required to go over from one set of values M, R and μ to another set M_1, R_1 and μ_1 is the successive application of three elementary homologous transformations. To show this we proceed as follows:

Let the physical variables, after a general homologous transformation has been applied be denoted by attaching a suffix " 1." for a general homologous transformation we should have

$$\begin{aligned} r_1 &= y^{n_1} r, & (\bar{\beta}\mu)_1 &= y^{n_5}\bar{\beta}\mu \\ P_1 &= y^{n_2} P, & T_1 &= y^{n_6} T \\ M(r_1)_1 &= y^{n_3} M(r), & (\kappa_0\epsilon_0)_1 &= y^{n_7}(\kappa_0\epsilon_0) \\ \rho_1 &= y^{n_4}\rho, \end{aligned} \tag{95}$$

where $n_1, \cdots, n_7$ are, for the present, arbitrary constants and y is the transformation constant. The exponents $n_1, \cdots, n_7$ should satisfy certain conditions, namely, those which are necessary for the continued validity of equations (89)–(92) in the suffixed variables. Substituting (95) in (89) we find that we should have

$$y^{n_2-n_1} = y^{n_3+n_4-2n_1}, \tag{96}$$

or

$$n_2 - n_1 = n_3 + n_4 - 2n_1. \tag{97}$$

In the same way equations (90), (91) and (92) yield:

$$n_3 - n_1 = 2n_1 + n_4, \tag{98}$$

$$n_2 = n_4 + n_6 - n_5, \tag{99}$$

$$n_6 - n_1 = n_7 + (\alpha + 3)n_4 - (6 + s - \nu)n_6 + n_1. \tag{100}$$

We have thus four equations between the seven unknowns. Hence, we should be able to express any four of the n's in terms of the other three. We shall choose n_1, n_3 and n_6 as the independent quantities.

Solving for n_2, n_4, n_6 and n_7 in terms of n_1, n_3 and n_5 we find that

$$n_2 = -4n_1 + 2n_3, \tag{101}$$

$$n_4 = -3n_1 + n_3, \tag{102}$$

$$n_6 = -n_1 + n_3 + n_5, \tag{103}$$

$$n_7 = -(s - \nu - 3\alpha)n_1 + (4 + s - \nu - \alpha)n_3 + (7 + s - \nu)n_5. \tag{104}$$

If we choose $n_1 = 1$, $n_3 = 0$ and $n_5 = 0$ we have a homologous transformation in which a star of a given mass M and mean molecular mass $\mu\bar{\beta}$ is expanded or contracted. In the same way, the choice $n_1 = 0$, $n_3 = 1$ and $n_5 = 0$ corresponds to an alteration of M while R and $\mu\bar{\beta}$ are kept unchanged. Finally, the choice $n_1 = 0$, $n_3 = 0$ and $n_5 = 1$ corresponds to an alteration of $\mu\bar{\beta}$ while M and R are kept unchanged.

We have now to consider how the luminosity is changed by a homologous transformation. Since

$$L = 4\pi \int_0^R r^2 \rho \epsilon dr, \tag{105}$$

we have according to our law (88) for ϵ

$$\kappa_0 L = 4\pi \kappa_0 \epsilon_0 \int_0^R r^2 \rho^{\alpha+1} T^\nu dr. \tag{106}$$

Hence, by a general homologous transformation $\kappa_0 L$ alters to $(\kappa_0 L)_1$ where

$$(\kappa_0 L)_1 = y^{n_7 + 3n_1 + (\alpha+1)n_4 + \nu n_6}(\kappa_0 L), \tag{107}$$

or by (101), (102), (103) and (104)

$$(\kappa_0 L)_1 = y^{-sn_1 + (5+s)n_3 + (7+s)n_5}(\kappa_0 L). \tag{108}$$

In other words

$$L = \text{constant}\, \frac{M^{5+s}}{\kappa_0 R^s} (\mu\bar{\beta})^{7+s}. \tag{109}$$

Another relation of importance is that which is equivalent to (104):

$$\kappa_0 \epsilon_0 = \text{constant}\, R^{(3\alpha+\nu-s)} M^{(4+s-\nu-\alpha)} (\mu\bar{\beta})^{7+s-\nu}. \tag{110}$$

It is clear that the constants in (109) and (110) can depend only on the exponents α, ν and s. We have thus proved the invariance of the luminosity formula for stars in radiative equilibrium. If, however, the law of energy generation is such that it leads to a sufficiently strong concentration of the energy sources towards the center then we should reach a stage when

$$\frac{\kappa\eta}{\overline{\kappa\eta}(r)} > 4\frac{\Gamma_2 - 1}{\Gamma_2}. \tag{111}$$

In other words going inward toward the interior of a star the radiative gradient will become unstable at some definite point $r = r_i$ (say). For stars with negligible radiation pressure we have

$$\frac{\kappa\eta}{\overline{\kappa\eta}(r_i)} = 1.6 \qquad (r = r_i). \tag{112}$$

For $r < r_i$, $\kappa\eta/\overline{\kappa\eta}(r) > 1.6$. Now the right hand side of (112) is a pure number, while the quantity on the left hand side is homology invariant. Hence the fraction $q = r_i/R$ of the radius at which the instability of the radiative gradient sets in, is the same for all stars with vanishing radiation pressure. The fraction q depends only on the exponents α, ν and s which occur in the laws for ϵ and κ. Further the material interior to r_i will be in convective equilibrium and we should have

$$\frac{p}{p_i} = \left(\frac{\rho}{\rho_i}\right)^{\gamma}, \tag{113}$$

where p_i and ρ_i refer to the pressure and the density at the interface, *i.e.*, at $r_i = qR$. Equation (113) is clearly homology invariant. We have thus proved the invariance of the form of the mass-luminosity-radius relation quite generally.

The next thing that we shall have to examine is the range of variation in the constant of proportionality in the relation (109) for the possible range of stellar models. In these discussions we shall restrict ourselves to the case $s = 1/2$, *i.e.*, assume for the law of opacity the form

$$\kappa = \kappa_0 \rho T^{-3.5}, \tag{114}$$

where κ_0 will depend upon the chemical composition.

We shall consider two models: (*a*) the model ϵ = constant and (*b*) the point source model $\epsilon = 0$, $r \neq 0$. The model ϵ = constant corresponds to a uniform distribution of the energy sources while the point source model corresponds to the complete concentration of the energy sources at the center of the star. We may expect that in the actual stars an energy source distribution is realized which is intermediate to these two limiting cases.

Consider first the model ϵ = constant. Remembering that we are restricting ourselves to stars with negligible radiation pressure we can rewrite (61) in the form:

$$\frac{k}{\mu H}\frac{3}{a}\frac{d(\rho T)}{d(T^4)} = \frac{4\pi c G M}{L\kappa\eta}. \tag{115}$$

For the model under consideration $\eta = 1$. Using (88) as our law of opacity we have

$$\frac{k}{\mu H}\frac{3}{a}\frac{d(\rho T)}{d(T^4)} = \frac{4\pi cGM}{\kappa_0 L}\frac{T^{3+s}}{\rho}. \tag{116}$$

Let

$$\rho = \frac{\mu H}{k}\frac{a}{3}T^3 y. \tag{117}$$

Then we have

$$\frac{d(yT^4)}{d(T^4)} = \frac{4\pi cGM}{\kappa_0 L}\frac{k}{\mu H}\frac{3}{a}\frac{T^s}{y}, \tag{118}$$

or after some minor transformations:

$$\frac{1}{4}T\frac{dy}{dT} = \frac{4\pi cGM}{\kappa_0 L}\frac{k}{\mu H}\frac{3}{a}\frac{T^s}{y} - y. \tag{119}$$

Introduce the new variable x defined by

$$x = \frac{4\pi cGM}{\kappa_0 L}\frac{k}{\mu H}\frac{3}{a}T^s. \tag{120}$$

(119) now takes the form

$$\frac{1}{4}sxy\frac{dy}{dx} = x - y^2. \tag{121}$$

The general solution of (121) is easily seen to be

$$y^2 = \frac{8}{8+s}x + Bx^{-8/s}, \tag{122}$$

where B is a constant of integration. From (122) we see that if $s > 0$ the second term in (122) becomes rapidly small compared to the first term as we descend into the deeper layers of the star. Hence for layers not immediately near the boundary

$$y^2 \backsimeq \frac{8}{8+s}x. \tag{123}$$

From (117), (120) and (123) we now have

$$\rho = \text{constant } T^{3+\frac{1}{2}s}; \tag{124}$$

or again

$$P = \frac{k}{\mu H}\rho T = \text{constant } T^{4+\frac{1}{2}s}. \tag{125}$$

From (124) and (125) we see that

$$P = \text{constant } \rho^{(1+(1/(3+\frac{1}{2}s)))}. \tag{126}$$

In other words, the configurations are polytropes of index $n = 3 + \frac{1}{2}s$. For the physically interesting case $s = 1/2$ so that the stars on this model are polytropes of index $n = 3.25$. The march of density and temperature in such configurations is given in Table 4 (see Figs. 2 and 3).

TABLE 4

DENSITY AND TEMPERATURE DISTRIBUTIONS FOR THE MODEL ϵ = CONSTANT

ξ	ρ/ρ_c	T/T_c
0	1.000	1.000
0.4	0.918	0.974
0.8	0.719	0.903
1.2	0.495	0.805
1.6	0.311	0.698
2.0	0.184	0.594
2.4	0.105	0.500
2.8	0.0588	0.418
3.2	0.0326	0.349
3.6	0.0179	0.290
4.0	0.00970	0.240
4.4	0.00518	0.198
4.8	0.00271	0.162
5.2	0.00137	0.131
5.6	6.57×10^{-4}	0.105
6.0	2.92×10^{-4}	0.0818
6.4	1.16×10^{-4}	0.0615
6.8	3.78×10^{-5}	0.0436
7.2	8.63×10^{-6}	0.0277
7.6	8.19×10^{-7}	0.0134
8.0	2.96×10^{-11}	0.0006
8.0189	0	0

Further from the integration appropriate for the polytrope $n = 3.25$ we find that [14]

$$\left.\begin{aligned} \rho_c &= 88.15\,\bar{\rho}, \\ T_c &= 0.968\,\frac{\mu H}{k}\,\frac{GM}{R}, \\ P_c &= 20.37\,\frac{GM^2}{R^4}. \end{aligned}\right\} \tag{127}$$

Further by a simple transformation of the luminosity formula (66) we obtain [15]

$$L = 1.43 \times 10^{25}\,\frac{1}{\kappa_0}\,\frac{M^{5.5}}{R^{0.5}}\,\mu^{7.5}, \tag{128}$$

where L, M and R are expressed in the corresponding solar units.

[14] The integration for the polytrope $n = 3.25$ is given in *Ap. J.*, **89**, 116, 1939.
[15] For the details of the derivation see the author's monograph (Ref. 9, pp. 322–327).

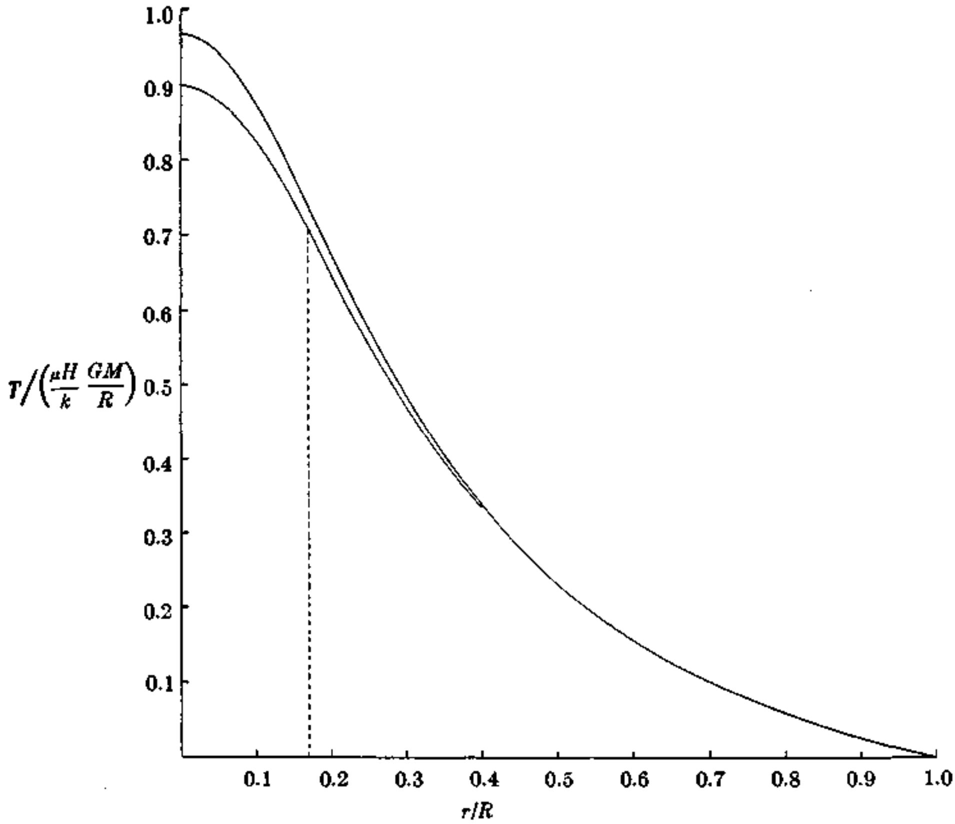


FIG. 2. The temperature distributions in the models (*a*) ϵ = constant and (*b*) the point source model.

Let us next consider the point source model. For this model it is clear that the central regions must be in convective equilibrium and therefore consist of polytropic cores of index $n = 1.5$ surrounded by "point-source envelopes," *i.e.*, regions governed by the equations

$$\frac{k}{\mu H}\frac{d}{dr}(\rho T) = -\frac{GM(r)}{r^2}\rho, \tag{129}$$

$$\frac{dp_r}{dr} = -\frac{\kappa_0 L}{4\pi c r^2}\frac{\rho^2}{T^{3.5}}, \tag{130}$$

$$\frac{dM(r)}{dr} = 4\pi r^2 \rho. \tag{131}$$

The integration has to be effected numerically. To start the integration we assume that the polytropic core extends to a fraction q of the

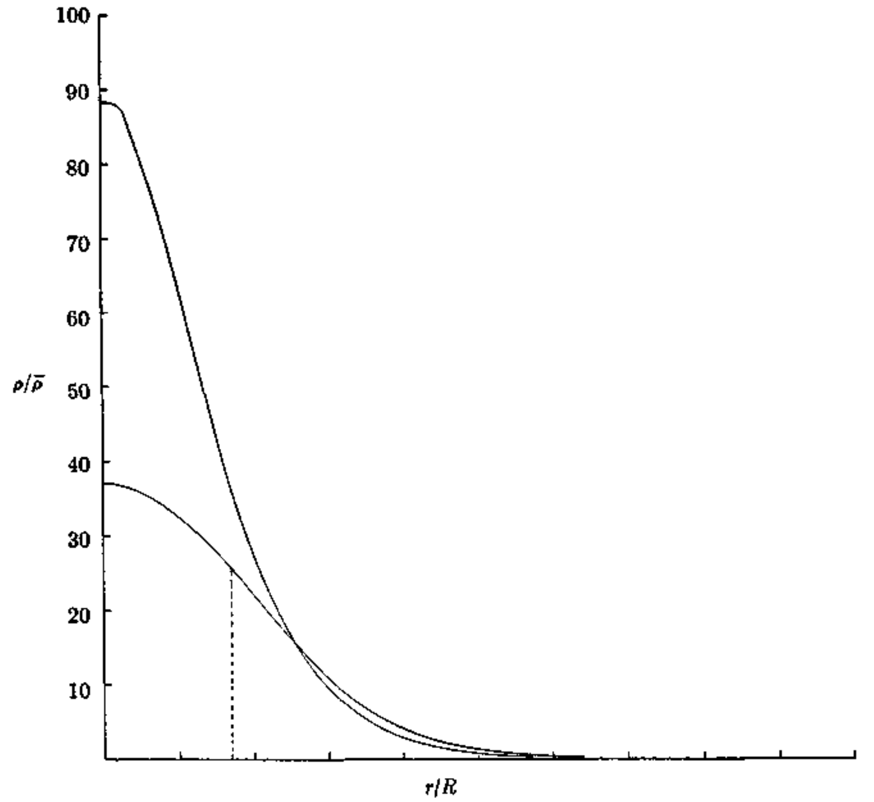


FIG. 3. The density distributions in the models (*a*) ϵ = constant and (*b*) the point source model.

radius of the star. From this point onward the integration is continued by means of equations (129)–(131). For an arbitrarily assigned fraction q, the density ρ and the temperature T will not tend to zero simultaneously. We can, however, adjust q by trial and error until ρ and T do tend to zero simultaneously. In this way we can construct a configuration of assigned mass M, radius R and mean molecular mass μ. From the homology argument it follows that the relative distributions of density, temperature, etc., will be the same for all stars with negligible radiation pressure built on this model. It is found [16] that the convective core extends to a fraction 0.17 of the radius of the star and encloses 14.5 per cent of the mass. The march of ρ and T for this model is shown in Table 5 (see Figs. 2 and 3). Further, it is found that

[16] The integration was first effected by Cowling. For greater details see the author's monograph (Ref. 9), pp. 351–355.

TABLE 5

DENSITY AND TEMPERATURE DISTRIBUTIONS FOR THE POINT-SOURCE MODEL

ξ	ρ/ρ_c	T/T_c
0	1.000	1.000
0.4	0.961	0.974
0.8	0.852	0.898
1.2	0.693	0.784
1.6	0.508	0.661
2.0	0.330	0.551
2.4	0.197	0.455
2.8	0.110	0.372
3.2	5.90×10^{-2}	0.303
3.6	3.05×10^{-2}	0.245
4.0	1.51×10^{-2}	0.196
4.4	7.16×10^{-3}	0.156
4.8	3.20×10^{-3}	0.121
5.2	1.30×10^{-3}	0.0919
5.6	4.60×10^{-4}	0.0666
6.0	1.26×10^{-4}	0.0448
6.4	2.06×10^{-5}	0.0256
6.8	6.22×10^{-7}	0.00873
7.0	5.4×10^{-10}	0.00100
7.027	0	0

$$\left.\begin{aligned} \rho_c &= 37.0\,\bar{\rho}, \\ T_c &= 0.900\,\frac{\mu H}{k}\,\frac{GM}{R}, \\ P_c &= 7.95\,\frac{GM^2}{R^4}. \end{aligned}\right\} \tag{132}$$

The luminosity formula is found to be

$$L = 5.43 \times 10^{24}\,\frac{1}{\kappa_0}\,\frac{M^{5.5}}{R^{0.5}}\,\mu^{7.5}, \tag{133}$$

where L, M and R are expressed in the corresponding solar units.

We now see that the mass-luminosity-radius relations for the two models are of the same form (in agreement with our general discussion). In addition, we notice that the constants of proportionality differ only by a factor 2.6 for the extreme range in the possible distributions of the energy sources.

For a comparison with these two models we may note that on the standard model in which stars are polytropes of index $n = 3$ we have

$$\left.\begin{aligned} \rho_c &= 54.2\,\bar{\rho}, \\ T_c &= 0.854\,\frac{\mu H}{k}\,\frac{GM}{R}, \\ P_c &= 11.05\,\frac{GM^2}{R^4}. \end{aligned}\right\} \tag{134}$$

We see that this model is " intermediate " to the two other models considered.

A fundamental result of importance which has come out of the present discussion is that there exists a relation of the type

$$L = \text{constant}\,\frac{1}{\kappa_0}\frac{M^{5.5}}{R^{0.5}}\mu^{7.5}, \tag{135}$$

in which the uncertainty in the constant of proportionality due to possible range of stellar models is less than the other uncertainties inherent in the problem.

As we have already pointed out κ_0 will depend on the chemical composition. The discussion of the stellar opacity is beyond the scope of the present report [17] except to mention that it is found that we can write

$$\kappa_0 = \frac{3.89 \times 10^{25}\,(1 - X_0{}^2)}{t}, \tag{136}$$

where X_0 is the hydrogen content by weight and t is a factor which is slowly varying function of ρ and T. For individual stars t can be replaced by a certain appropriate average value of $t = \bar{t}$. Thus for the sun it is found that $\bar{t} = 5$ while for Capella it is practically unity.

It is now obvious that an application of the relation (135) to the observational material on the masses, luminosities and the radii of the stars enables the determination of the mean molecular weight of individual stars. This problem has been investigated in great detail by Strömgren who finds a *systematic* variation of X_0 in the plane of the Hertzsprung-Russell diagram (more clearly however in a mass-radius diagram). As a result of his investigation Strömgren finds that within the limits of uncertainty of the observational material the stars can be satisfactorily arranged as a two parametric family, the two parameters being the mass M and the hydrogen content X_0. The interpretation of the Hertzsprung-Russell diagram which Strömgren arrives at is the following:

The main series up to spectral class A is the locus of stars of hydrogen content varying between 25 to 45 per cent—*i.e.*, about a mean of 35 per cent—and masses running up to $2.5\odot$. Stars of small mass and low hydrogen content are relatively rare, they occur as subgiants of spectral classes G to K. The gap between M giants and the corresponding dwarfs (on the main series) arises from the circumstance that not even stars of low hydrogen content " scatter " in this region. The massive stars ($M > 5\odot$) occurring in the region of the B-stars which are rich in hydrogen (X_0 sometimes going up to 95 per cent)

[17] For details see the author's monograph, pp. 261-272.

form the continuation of the main series, the continuation arising from the circumstance that massive stars with " medium " hydrogen content ($0.4 < X_0 < 0.8$) which are on the main series occur in a very small region of the H.R diagram. (We shall obtain evidence in Section III for the breakdown of the model underlying these computations for the very massive stars. Further, along the main series the breakdown probably sets in at about $M = 10\odot$. The investigations of the hydrogen contents of the B-stars are therefore somewhat inconclusive.) The giant branch is characterized by stars having about the same hydrogen content as (or somewhat less than) the main series stars. The giant branch is limited on the side of low luminosity, since stars of low luminosity are relatively rare. On the side of high luminosity it is limited again, because for X_0 a little greater than 0.35 the characteristic bend of the curves of constant X_0 (see Fig. 4) disappears and also because the stars of large mass with hydrogen content greater than about 40 per cent scatter over a large area in the H.R diagram, which must, therefore, be sparsely populated. The gap in the giant branch in the region of the spectral class F is probably due to a real scarcity of stars with masses between 2.5 and $4.5\odot$. The supergiants then are interpreted as massive stars with medium hydrogen content (see Fig. 4).

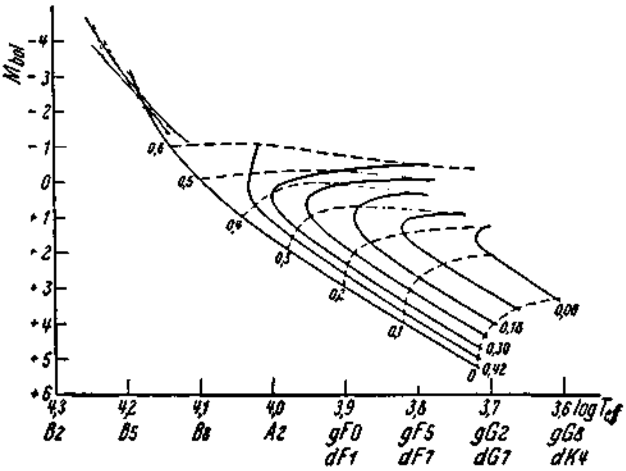


FIG. 4. Curves of constant X_0 (the full line curves) and the curves of constant M (the dotted curves) in the plane of the Hertzsprung-Russell diagram.

An important application of the method described has been recently made by Kuiper. By determining μ and X_0 for a few stars in the Hyades cluster for which he had derived the L, M and R values, he was able to show that the stars in this cluster are relatively poor in hydrogen as compared to the normal main series stars (*i.e.*, sun, α Centauri, etc.).

So far we have considered only the relation (109). But from the homology argument we established another relation, namely (110). Now, κ_0 and ϵ_0 will depend upon the chemical composition, *i.e.*, on μ and X_0. Hence if we consider stars with a given X_0 then we can eliminate R between (109) and (110) and obtain a *pure* mass-luminosity relation. Hence for stars with constant X_0 we have:

$$L = \text{constant}\ M^{5+s+\frac{(4+s-\nu-\alpha)s}{3\alpha+\nu-s}}. \tag{137}$$

If we assume $s = 1/2$ and $\alpha = 2$, (137) reduces to

$$L = \text{constant}\ M^{\frac{63+10\nu}{11+2\nu}}. \tag{138}$$

Hence by selecting stars of a given X_0 we can determine ν by a comparison of (138) with the result of observations. This in turn will give some indications of the type of nuclear reactions that are responsible as the source of stellar energy.[18]

III. The Method of Stellar Envelopes

In this method the equilibrium of the stellar envelopes is studied. By a stellar envelope we shall mean the outer parts of a star which though containing only a small fraction (for definiteness, we shall assume this fraction to be 10 per cent) of the total mass M nevertheless occupy a good fraction of the radius R. A study of stellar envelopes has a twofold application to astrophysical theories: first, it extends the study of the conventional stellar atmospheres into the far interior and, secondly, it has also a very definite bearing on the studies of the deep interiors which are the main concern in this report. Thus the central condensation of a star, defined as the fraction ξ^* of the radius R which encloses the inner 90 per cent of the mass M, must give some indication of the concentration of mass toward the center of the star under consideration. It is clear that $(1 - \xi^*)$ is a measure of the *extent* of the stellar envelope.

In writing down the equations of equilibrium of the stellar envelope we introduce two simplifications, (a) that there are no sources of energy in the stellar envelope and (b) that the mass contained in the envelope can be neglected in comparison with the mass of the star as a whole. Indeed, these two assumptions can be regarded as defining the stellar

[18] G. Gamow, *Ap. J.*, 89, 130, 1939.

envelope. The equations of equilibrium then are

$$\frac{d}{dr}\left(\frac{k}{\mu H}\rho T + \frac{1}{3}aT^4\right) = -\frac{GM}{r^2}\rho, \tag{139}$$

$$\frac{dp_r}{dr} = -\frac{\kappa_0 L}{4\pi c r^2}\frac{\rho^2}{T^{3.5}}, \tag{140}$$

and

$$\frac{dM(r)}{dr} = 4\pi r^2 \rho. \tag{141}$$

The equations (139) and (140) can be solved explicitly [19] to give the distribution of density and temperature in the stellar envelope. Finally, equation (141) enables us to determine how far inwards we have to go to cover the first 10 per cent of the mass. The equations which determine the central condensation are found to be [20]

$$\alpha = 6.25 \times 10^{-3}\left[\frac{L^2 R\mu(1 - X_0^2)^2}{M^3}\right]^{1/4}, \tag{142}$$

$$f(\alpha; w^*) = 0.0618\,\frac{(1 - X_0^2)^{0.5}}{\mu^{3.75}}\left(\frac{LR^{0.5}}{M^{5.5}}\right)^{0.5}, \tag{143}$$

and

$$\xi^* = \frac{1}{(w^* + \alpha)^3\left(w^* + \frac{19}{51}\alpha\right) + 1 - \frac{19}{51}\alpha^4}. \tag{144}$$

In the above equations L, M and R are expressed in solar units. $f(\alpha; w)$ is a function defined by means of a definite integral. Tables of this function have been provided.[21]

The method of evaluating ξ^* for a star of given L, M and R and an assumed value of μ proceeds as follows:

We first determine α. Then by interpolation in the tables of the function $f(\alpha; w)$ we find the value w^* such that $f(\alpha; w^*)$ has the value given by the right hand side of (143). (144) then determines ξ^*.

Figs. 5, 6 and 7 illustrate the dependence of ξ^* on X_0 for different stars.

Without going into too much detail it is clear that the evaluation of ξ^* for the normal stars (sun, Capella, ζ Herculis) confirms the conclusions drawn by Strömgren on the basis of the standard model. To consider an example: compare the sun and ζ Herculis. Both are stars of small mass and hence of negligible radiation pressure. According to our

[19] For the solutions see the author's monograph, Chapter VIII.
[20] See reference 19.
[21] Tables of the function $f(\alpha; w)$ will be found in the author's monograph, p. 361.

discussion of the homologous transformations in Section II, it is clear that these two stars must be homologous—and hence must be characterized by the same value of ξ^*. Fig. 5 shows that if ζ Herculis and the sun should have the same value, then the former must be poorer in hydrogen than the sun. This confirms the conclusion of Strömgren who has derived for the sun and ζ Herculis the values $X_0 = 0.37$ and 0.11 respectively.

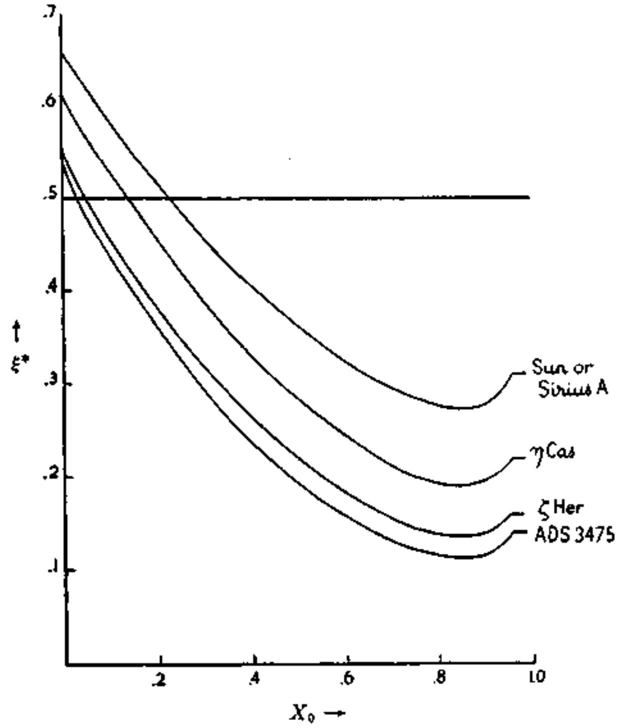


FIG. 5. (ξ^*, X_0) curves for the normal stars.

Considering now Fig. 6 we see that if we go along the sequence of stars, the sun; ζ Aurigæ B star ($M = 8.1\odot$); μ_1 Scorpii ($M = 12\odot$); V Puppis ($M = 18.6\odot$); VV Cephei, B star ($M = 31\odot$) and the Trumpler stars ($M \sim 100\odot$) the (ξ^*, X_0) curves change continuously. This strongly suggests the breakdown of the standard model for stars on the main series sets in at about $M = 10\odot$, becoming more and more pronounced on passing toward the larger masses. This breakdown is most clearly shown by the Trumpler stars where no adjustment of the mean molecular mass can make them homologous to the normal stars.

Turning next to Fig. 7 we notice that if we go along the sequence of stars, the sun ($M = \odot, R = R_\odot$); ζ Herculis ($M = .98\odot$, $R = 1.9R_\odot$);

Capella ($M = 4.2\odot$, $R = 15.8R_\odot$); ζ Aurigæ, K-star ($M = 14.8\odot$, $R = 200R_\odot$); ϵ Aurigæ, I-star ($M = 24.6\odot$, $R = 2140R_\odot$); *VV* Cephei, M-star ($M = 49\odot$, $R = 2130R_\odot$) we infer again the possibility of a breakdown of the standard model also in the region of the massive supergiants (stars of high luminosity and large radius). The

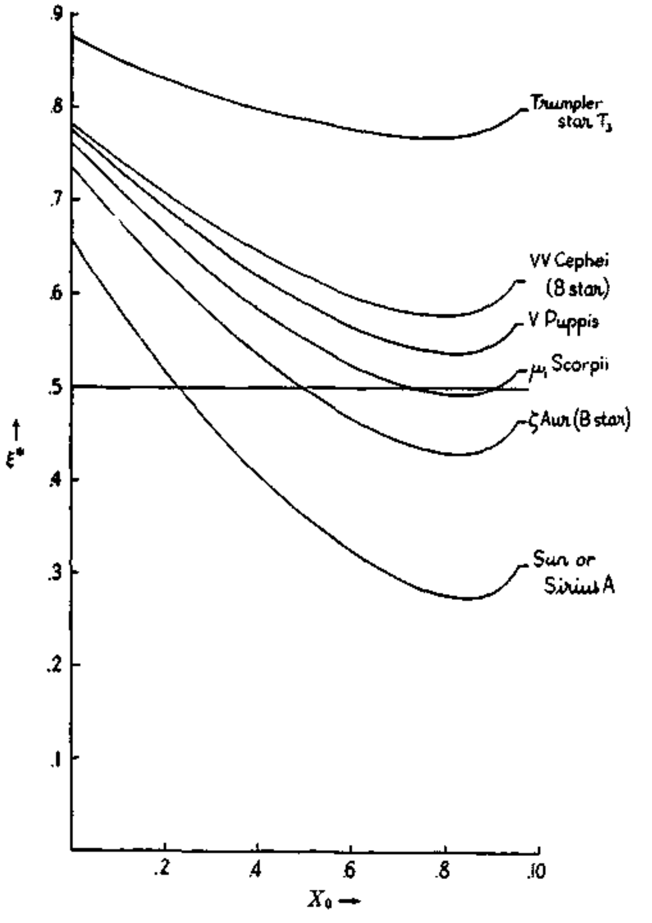


FIG. 6. (ξ^*, X_0) curves for the stars on the main series.

breakdown is now, however, in the sense of becoming more centrally condensed; this differs from the behavior of the massive stars which form an extension of the main series; the latter are certainly more homogeneous than the normal stars. Among the supergiants the possibility of finding stars with ξ^* as small as 0.05 (*e.g.*, *VV* Cephei, *M*-component) cannot be excluded.

The main results which emerge from the discussion of the central condensations of stars can be summarized as follows:

(*a*) The general way in which the theory of stellar envelopes supports the essential conclusions reached in Section II concerning the structures and the hydrogen contents of the normal stars.

(*b*) The increasing homogeneity of the massive stars on the main series, the breakdown of the standard model setting in probably at values of the mass of about 10 ☉.

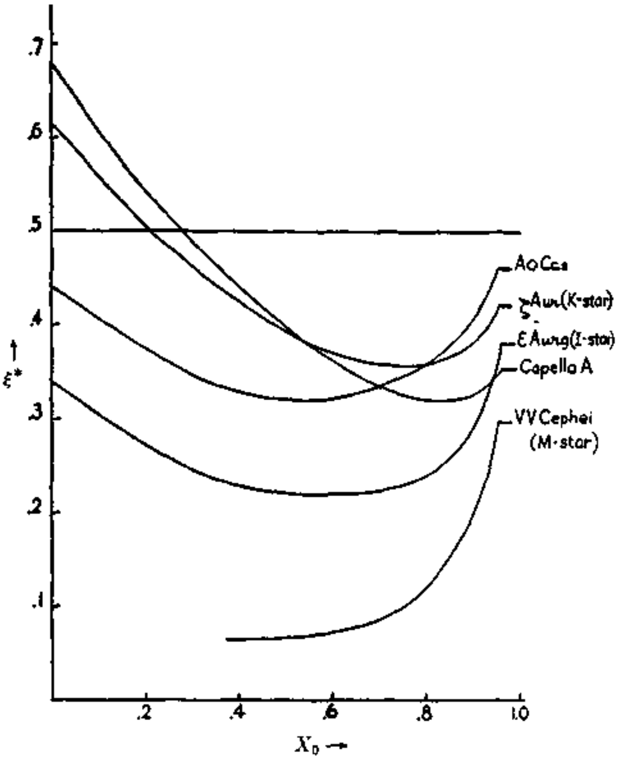


FIG. 7. (ξ^*, X_0) curves for the giants and supergiants.

(*c*) The centrally condensed nature of the massive supergiants.

We may therefore infer from the examples discussed that a certain systematic variation of the stellar model in the (M, R) plane exists.

To avoid misunderstanding, it should be pointed out that the discussion of the mass-luminosity-radius relation is satisfactory only for stars with negligible radiation pressure and consequently can be expected to be valid only for stars of normal masses ($M < 5$ ☉). On the other hand, in the theory of stellar envelopes the effect of the radiation pressure is taken into account " exactly " and it is for this reason that we are able to draw unambiguous conclusions concerning the structures of stars to which the standard theory cannot be applied. There is thus no contradiction involved in our discovering the " break-down " nature of the massive stars by the method of stellar envelopes.

IV. The Theory of White Dwarfs

So far we have restricted ourselves to a consideration of gaseous stars. However, as is well known, R. H. Fowler showed for the first time that the electrons assembling in the interior of the white dwarfs must be highly degenerate in the sense of the Fermi-Dirac statistics.

The theory of degeneracy in the form required for application in the theory of white dwarfs can be derived in an entirely elementary way. Now, a given number N of electrons can be confined in a given volume V by one of two methods, either by means of potential walls such that electrons inside the " potential hole " cannot escape, or by means of imposing certain periodicity conditions. We shall not consider these restrictions but the essential result of such discussions is that we can label the possible energy states for an electron inside a given volume V by means of quantum numbers in somewhat the same manner as the quantum states of an electron in an atom. If we assume that volume V is large, then it follows from the general theory that the number of quantum states with momenta between p and $p + dp$ is given by

$$V\frac{8\pi p^2 dp}{h^3}. \qquad (145)$$

The meaning of (145) is simply that the accessible six-dimensional phase space can be divided into " cells " of volume h^3 and that in each cell there are two possible states. Now, the Pauli principle states that no two electrons can occupy the same quantum state. This implies that if $N(p)dp$ denotes the number of electrons with momenta between p and $p + dp$ then

$$N(p)dp \leqslant V\frac{8\pi p^2 dp}{h^3}. \qquad (146)$$

A completely degenerate electron gas is one in which the lowest quantum states are all occupied. In other words, we should have

$$N(p) = V\frac{8\pi p^2}{h^3}. \qquad (147)$$

It is clear that if there is only a finite number N of electrons in the specified volume, then all the electrons must have momenta less than a certain maximum value p_0 such that

$$N = V\int_0^{p_0}\frac{8\pi p^2}{h^3}dp = \frac{8\pi V}{3h^3}p_0^3. \tag{148}$$

This " threshold value " p_0 of p is related to the electron concentration n by

$$n = \frac{N}{V} = \frac{8\pi}{3h^3}p_0^3. \tag{149}$$

To calculate the pressure, we recall that by definition the pressure P exerted by a gas is simply the mean rate of transfer of momentum across an ideal surface of unit area in the gas. From this definition it follows quite generally that

$$PV = \frac{1}{3}\int_0^{\infty} N(p)pv_p dp, \tag{150}$$

where v_p is the velocity associated with the momentum p. According to (147) we have for the case under consideration

$$P = \frac{8\pi}{3h^3}\int_0^{p_0} p^3\frac{\partial E}{\partial p}dp, \tag{151}$$

where E is the kinetic energy of the electron having a momentum p. According to the special theory of relativity

$$E = mc^2\left\{\left(1 + \frac{p^2}{m^2c^2}\right)^{1/2} - 1\right\}, \tag{152}$$

which gives

$$\frac{\partial E}{\partial p} = \frac{1}{m}\left(1 + \frac{p^2}{m^2c^2}\right)^{-1/2}p. \tag{153}$$

Substituting (153) in (151) we have

$$P = \frac{8\pi}{3mh^3}\int_0^{p_0}\frac{p^4 dp}{\left(1 + \frac{p^2}{m^2c^2}\right)^{1/2}}. \tag{154}$$

The integral occurring in (154) can be evaluated and we find that we can express P as

$$P = \frac{\pi m^4c^5}{3h^3}f(x), \tag{155}$$

where

$$f(x) = x(2x^2 - 3)(x^2 + 1)^{1/2} + 3\sinh^{-1}x, \tag{156}$$

and

$$x = p_0/mc. \tag{157}$$

We can now write (149) in the form

$$n = \frac{8\pi m^3 c^3}{3h^3} x^3. \tag{158}$$

Equations (155), (156) and (158) represent parametrically the equation of state of a completely degenerate electron gas. (158) can be written alternatively as

$$\rho = n\mu_e H = Bx^3, \tag{159}$$

where

$$B = \frac{8\pi m^3 c^3 \mu_e H}{3h^3} = 9.82 \times 10^5 \mu_e. \tag{160}$$

Similarly we can write (155) as

$$P = Af(x), \tag{161}$$

where

$$A = \frac{\pi m^4 c^5}{3h^3} = 6.01 \times 10^{22}. \tag{162}$$

Completely degenerate stellar configurations are then those which are in hydrostatic equilibrium and in which P and ρ are related according to (159) and (161). We should therefore introduce equations (159) and (161) in the equation of equilibrium,

$$\frac{1}{r^2}\frac{d}{dr}\left(\frac{r^2}{\rho}\frac{dP}{dr}\right) = -4\pi G\rho. \tag{163}$$

By the transformations

$$r = \left(\frac{2A}{\pi G}\right)^{1/2}\frac{1}{By_0}\eta; \qquad y = y_0\phi, \tag{164}$$

where

$$y_0^2 = x_0^2 + 1, \tag{165}$$

equation (163) reduces to

$$\frac{1}{\eta^2}\frac{d}{d\eta}\left(\eta^2\frac{d\phi}{d\eta}\right) = -\left(\phi^2 - \frac{1}{y_0^2}\right)^{3/2}. \tag{166}$$

Equation (166) has to be solved with the boundary conditions

$$\phi = 1; \qquad \frac{d\phi}{d\eta} = 0 \qquad \text{at} \qquad \eta = 0. \tag{167}$$

For each specified value of y_0 we have one such solution. The boundary is defined at the point where the density vanishes, and this by (159) and (165) means that if η_1 specified the boundary

$$\phi(\eta_1) = 1/y_0. \tag{168}$$

The integrations for the function ϕ have been carried out for ten different values of y_0 and the physical characteristics of the resulting configuration are shown in Table 6. The mass radius relation is shown in Fig. 8.

TABLE 6*

THE PHYSICAL CHARACTERISTICS OF COMPLETELY DEGENERATE CONFIGURATIONS

$1/y_0^2$	$M/\odot$	ρ_0 in Grams per Cubic Centimeter	ρ_{mean} in Grams per Cubic Centimeter	Radius in Centimeters
0	5.75	∞	∞	0
0.01	5.51	9.85×10^{8}	3.70×10^{7}	4.13×10^{8}
0.02	5.32	3.37×10^{8}	1.57×10^{7}	5.44×10^{8}
0.05	4.87	8.13×10^{7}	5.08×10^{6}	7.69×10^{8}
0.1	4.33	2.65×10^{7}	2.10×10^{6}	9.92×10^{8}
0.2	3.54	7.85×10^{6}	$7.9\ \times10^{5}$	1.29×10^{9}
0.3	2.95	3.50×10^{6}	4.04×10^{5}	1.51×10^{9}
0.4	2.45	1.80×10^{6}	2.29×10^{5}	1.72×10^{9}
0.5	2.02	9.82×10^{5}	1.34×10^{5}	1.93×10^{9}
0.6	1.62	5.34×10^{5}	$7.7\ \times10^{4}$	2.15×10^{9}
0.8	0.88	1.23×10^{5}	1.92×10^{4}	2.79×10^{9}
1.0	0	0	0	∞

* The values given in this table differ slightly from the published values (S. Chandrasekhar, *M. N.*, **95**, 208, 1935, Table III). The difference is due to the change in the accepted values of the fundamental physical constants.

The calculations are for $\mu_e = 1$. For other values of μ_e, M should be multiplied by μ_e^{-2}, R by μ_e^{-1} and ρ_0 by μ_e.

The most important characteristic of these configurations is that they possess a natural limit, *i.e.* as

$$y_0 \rightarrow \infty, \qquad \phi \rightarrow \theta_3 \tag{169}$$

(where θ_3 is the Lane-Emden function of index 3), and the mass tends to a finite limit M_3. Numerically it is found that

$$M_3 = 5.75\, \mu_e^{-2}\, \odot. \tag{170}$$

A glance at Table 6 shows that the mean density, the mass and the radius of these degenerate configurations are all of the right order of magnitude to provide the basis for the theoretical discussion of the white dwarfs. However, a really satisfactory test of the theory will consist in providing an observational basis for the existence of a mass such that as we approach it the mean density increases several times, even for a slight increase in mass. At the present time there is just one case which seems to support this aspect of the theoretical prediction. The case in question is Kuiper's white dwarf (AC 70° 8247) which is from several points of view a very remarkable star. According to Kuiper, the most probable values of L and R are

$$\operatorname{Log} L = -1.76; \qquad \operatorname{Log} R = -2.38, \tag{171}$$

L and R being expressed in solar units. If we assume that $\mu = 1.5$, the mass-radius relation leads to a mass of 2.5 ☉. On the other hand if we neglect relativity effects then the unrelativistic mass-radius relation (the

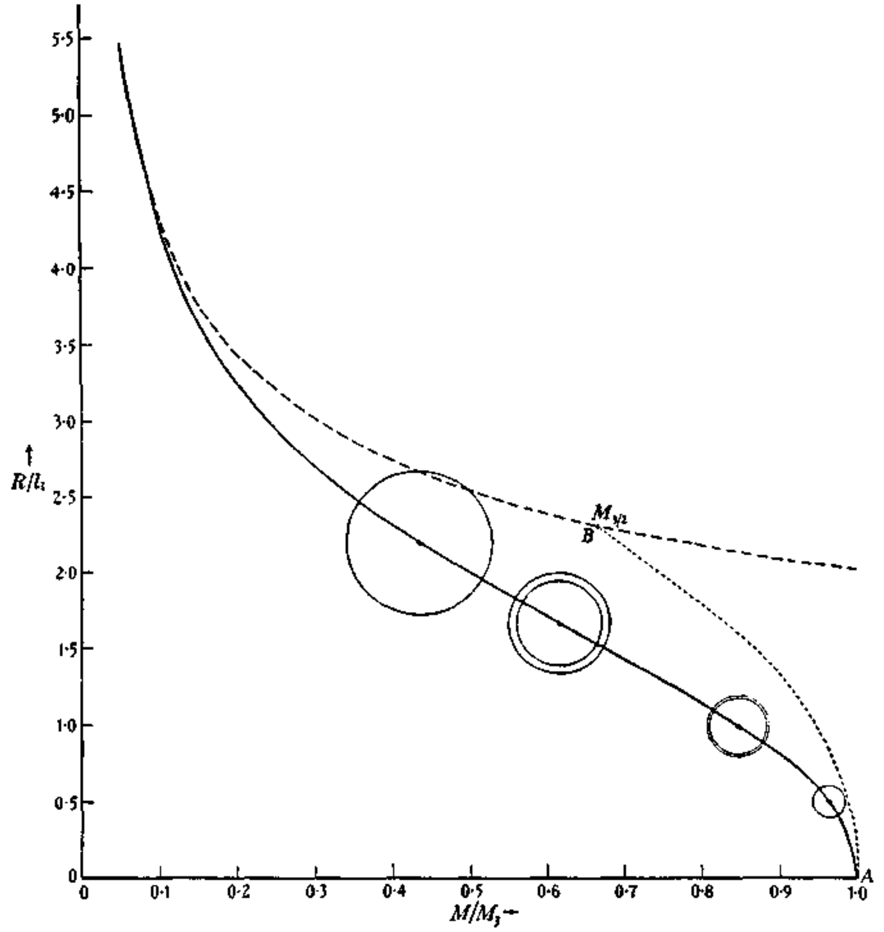


Fig. 8. The mass-radius relation for the white dwarfs.

dotted curve in Fig. 6) leads to a minimum mass of 28 ☉. The "probable" value predicted on the unrelativistic theory will be of the order of 100 solar masses. We conclude then that the relativistic effects are confirmed by observations.[22]

[22] There is a possibility that Wolf 219, another white dwarf discovered by Kuiper, may be comparable to AC 70° 8247). If confirmed this star would be even more extraordinary than AC 70° 8247) since it is of lower luminosity.

The Cosmological Constants

Prof. P. A. M. Dirac's recent letter in Nature[1] encourages me to direct attention to certain 'coincidences' which I had noticed some years ago, but which I have been hesitating to publish from the conviction that purely 'dimensional arguments' will not lead one very far.

If we consider the natural constants h (Planck's constant), c (velocity of light), H (mass of the proton), G (the constant of gravitation), we can form the following combination M_α which is of the dimension of mass:

$$M_\alpha = \left(\frac{hc}{G}\right)^{\alpha} \frac{1}{H^{2\alpha-1}}, \tag{1}$$

where α is an arbitrary numerical constant. Now a particular case of the above occurs in the theory of stellar interiors, namely, when $\alpha = 3/2$. Then

$$M_{3/2} = \left(\frac{hc}{G}\right)^{3/2} \frac{1}{H^2} \doteqdot 5 \cdot 76 \times 10^{34} \text{ gm.}, \tag{2}$$

which is about thirty times the mass of the sun. Now, the apparent success of steady state considerations in 'explaining' the observed order of stellar masses can be traced to the circumstance that the above combination (2) of the natural constants gives a mass of the correct order. It may be noticed that apart from numerical constants, (2) is the same as the upper limit to the mass of completely degenerate (degenerate in the sense of the Fermi-Dirac statistics) configurations[2]. The occurrence of (2) in stellar structure equations need not cause any surprise, since one can easily convince oneself by considering two homologous stellar configurations that if a formula for mass exists, it must contain the mean molecular weight μH with an inverse power 2, and this would, according to (1), fix the value of the exponent α as 3/2.

It is of interest to see what (1) leads to for other values of α. If $\alpha = 2$, then

$$M_2 = \left(\frac{hc}{G}\right)^{2} \frac{1}{H^3} \doteqdot 9 \cdot 5 \times 10^{20} \text{ mass of sun} \tag{3}$$

If we divide M_2 by H, then we get for the corresponding 'number of protons or/and neutrons',

$$N = \left(\frac{hc}{G}\right)^{2} \frac{1}{H^4} \doteqdot 1 \cdot 1 \times 10^{78}, \tag{4}$$

which is of the right order as the 'number of particles in the universe'. We may notice that if $G \sim t^{-1}$ (t is Milne's cosmological time), then $N \sim t^2$, which agrees with Dirac's speculation.

It may be further pointed out that if $\alpha = 1\frac{3}{4}$, then

$$M_{1\frac{3}{4}} = 1 \cdot 7 \times 10^{11} \text{ mass of sun}, \tag{5}$$

which is of the same order as the mass of our Milky Way system. If we 'identify' $M_{1\frac{3}{4}}$ as representing the mass of a galaxy (external or otherwise), then we should have, according to Dirac's ideas, that the 'number of particles in the galaxy' should vary as $t^{1 \cdot 75}$. Similarly, the number of particles in a star should vary as $t^{1 \cdot 5}$.

S. Chandrasekhar.

Yerkes Observatory,
Wisconsin.

[1] Nature, 139, 323 (Feb. 20, 1937).
[2] Chandrasekhar, S., *Mon. Not. Roy. Ast. Soc.*, 91, 456 (1931).

II. Radiative Transfer, the Polarization of the Sunlit Sky, and the Negative Ion of Hydrogen

Chandra used to say that the years 1943–48 were some of the happiest and most satisfying in his scientific life. His researches on radiative transfer ending up with the monograph *Radiative Transfer* (Oxford University Press, 1950) produced a series of papers in rapid succession. As the subject developed, he also developed, attaining for the first time a degree of self-assurance. He did not have to seek problems as problems arose one by one, each more complex and difficult than the previous one, and they were solved. The subject evolved on its own, at its own momentum, and attained a beauty which, according to his own admission, was not to be found in any of his other work.

For the above reasons, it is extremely difficult to pick one or two papers out of a series of twenty-four published in *The Astrophysical Journal*. However, I found in Paper 1 an excellent summary of the development of the whole subject of radiative transfer, beginning with Chandra's own ideas and methods and how they blended into the work influenced by Ambartsumian's ideas. It is based on the talk Chandra gave at a symposium at the Erevan Observatory in Armenia in October 1981. The symposium was organized to mark the fortieth anniversary of Ambartsumian's and Chandra's pioneering contributions to the field of radiative transfer. It was a kind of sentimental journey for Chandra, since he had not visited Russia since his visit in 1934 soon after his graduation. He had met Ambartsumian then, and the latter was partly responsible for encouraging Chandra to go ahead with his work on white dwarfs.

Likewise, Papers 2 and 3 are reviews of mostly his own work. Paper 2 describes the phenomenon of scattering of light in moving atmospheres, a subject that was of considerable interest in astrophysics, dealing with a variety of objects such as novae, Wolf–Rayet stars, planetary nebulae, the solar prominences, and the corona. In this paper Chandra establishes the equation of transfer, its approximate forms, and the associated boundary value problem and its solution. Paper 3 is based on the twentieth Josiah Willard Gibbs lecture delivered at Swarthmore, Pennsylvania on December 26, 1946. Chandra begins by saying that the advances in the various branches of theoretical physics have often resulted in the creation of new mathematical disciplines. He goes on to describe how advances in astrophysical studies relating to the transfer of radiation brought forth some new and characteristic developments in the theory of integrodifferential and functional equations.

The explanation of the polarization of unpolarized light by the earth's atmosphere was a problem that had remained unsolved since the classic work of Lord Rayleigh in 1871. In explaining the blue color of the sky based on Maxwell's equations, Rayleigh had made the approximation of a single scattering of the radiation and predicted nonvanishing polarization in all directions, except directly towards or away from the sun. It was known, however, that there existed two, sometimes three, neutral points of zero polarization on the sun's meridian circle, called the Babinet, Brewster and Arago points. It was not until 1946 that the problem found completely satisfactory solution in the hands of Chandra. In a series of papers, he formulated the full scattering problem with polarization and found solutions for the sunlit sky exhibiting precisely the character of the observations, in particular the above-described neutral points. I have chosen Paper 4, published with Elbert in 1954, to present this seminal work since it contains not only an excellent review of the problem and its solution, but also complete polarization sky maps computed from theory and comparison with observations.

The negative ion of hydrogen is the subject of the next three papers. H.A. Bethe in 1929 and independently E.A. Hylleraas in 1930 had established that a neutral hydrogen atom, together with an electron, could form a stable configuration. It was a delicate quantum-mechanical calculation involving variational methods and it had led to its possible existence in the sun's atmosphere, where there is an abundance of neutral hydrogen atoms as well as a supply of electrons due to the ionization of other elements. Theory anticipated that under such circumstances, there should be bound negative ions of hydrogen and they should have an effect on the absorption spectrum of the sun. Rupert Wildt in 1938 had indeed produced strong evidence for the presence of the negative ion of hydrogen in sufficient quantities to be the principal source of continuous absorption in the solar atmosphere and in the atmospheres of certain types of stars.

After the fundamental work of Bether and Hylleraas, several others tried to determine the electron affinity to the hydrogen atom (binding energy) with greater precision by trying out wave functions with more parameters than the original calculations of Bethe and Hylleraas. However, these efforts led to ambiguous results. It was Chandra who realized that the form of the wave function used by the later authors hinged on an analogy with the wave function that was immensely successful in explaining the helium atom with two bound electrons. He argued correctly that the physical situation of two electrons in the negative hydrogen ion was totally different and it required a more general parameterization of the trial wave function. With such a parameterization, he was able to obtain a more precise and stable value for the electron affinity. Chandra went on to calculate the continuous absorption coefficient of the negative hydrogen ion. This required a careful evaluation of the transition matrix element between the ground state of the hydrogen ion and the state with a neutral hydrogen atom and an outgoing free electron. He was able to obtain a precise value for this absorption coefficient and the consequent cross-sections for the radiative processes leading to its ionization. These calculations played an extremely important role in understanding the continuous spectrum of the sun and the stars.

II. Radiative Transfer, the Polarization of the Sunlit Sky, and the Negative Ion of Hydrogen

Radiative Transfer—A Personal Account

I had the pleasure of meeting Academician Viktor Amazaspovitch Ambartsumian in the company of Lev Davidovich Landau in the summer of 1934 in Leningrad; and the visit to the Hermitage with them is still very vivid in my memory. This is now my first opportunity, since that time, to visit the Soviet Union and to renew my personal acquaintance with Academician Ambartsumian. I feel greatly privileged to visit this Observatory, conceived and founded by his foresight and his efforts, and especially on this occasion when we are assembled to celebrate his innovative introduction of principles of invariance in the study of radiative transfer and in the fluctuations in brightness of the Milky Way.

Because of the exigencies of the Second World War, I first became aware of Academician Ambartsumian's first paper on principles of invariance[1] only in the summer of 1945; and I became aware of his second paper[2] very much later. At that time, I had already been involved for some two years in solving problems in radiative transfer by a different technique; and the impact of Ambartsumian's papers was immediate and profound. However, because of the particular circumstances in which I came to know of these innovative papers, their influence in midstream, so to say, in altering the course of my work was perhaps somewhat different from their influence on those of you who were more directly and immediately inspired by them. I hope, then, that you will forgive me if, on this occasion, I allow myself the license of giving an account of the evolution of my own investigations on radiative transfer during the years 1943–48 and how they were redirected by Ambartsumian's ideas. I am afraid that I must confine myself exclusively to those years, since, after the publication of my book on *Radiative Transfer* in 1950, my interests have strayed very far away and only rarely have I returned—I must confess with some nostalgia—to my interests of those youthful years.

Paper presented at the Symposium devoted to the 40th Anniversary of the Principle of Invariance: Introduction to the Radiation Transfer Theory, Byurakan, Armenia, USSR, 26–30 October 1981.

1. *C.R. (Doklady) Acad. URSS* 38 (1943): 257.
2. *J. Phys. Acad. Sci. USSR* 8 (1944): 65.

I. Preliminaries

I shall begin by formulating the problems I was concerned with when I became seriously interested in the theory of radiative transfer in the fall of 1943.

The problems relate to the transfer of radiation in atmospheres, stratified in parallel planes, which are either *semi-infinite* (normal to the plane of stratification) or *finite* (confined between two of the planes of stratification). The atmospheres are further characterized by a scattering coefficient, σ, so that a pencil of radiation of intensity I, incident on an element of mass dm, scatters the amount of radiation

$$\sigma\, dm\, I p(\cos\Theta)\,\frac{d\omega}{4\pi} \tag{1}$$

into an element of solid angle $d\omega$, in a direction inclined at an angle Θ to the direction of incidence, and $p(\cos\Theta)$ is the *phase function* for single scattering. If scattering is the only process by which matter and radiation interact with each other, then we must have

$$\int p(\cos\Theta)\,\frac{d\omega}{4\pi} = 1. \tag{2}$$

This is the *conservative case.* If, however, a certain fraction, $1 - \varpi_0$, of the incident radiant energy is consumptively absorbed, then we should have, instead,

$$\int p(\cos\Theta)\,\frac{d\omega}{4\pi} = \varpi_0 \quad (<1). \tag{3}$$

This is the *nonconservative case;* and ϖ_0 is the *albedo* for single scattering. In general we may suppose that the phase function can be expanded in a series of Legendre polynomials, $P_l(\cos\Theta)$, in the form

$$p(\cos\Theta) = \sum_{l=0}^{N} \varpi_l\, P_l(\cos\Theta), \tag{4}$$

where the ϖ_l's are constants and $\varpi_0 = 1$ in the conservative case.

In writing the equation of transfer, we shall introduce the normal *optical thickness*

$$\tau = \int_0^z \rho\sigma\, dz, \tag{5}$$

where ρ denotes the density and z is the normal distance from the boundary (at $\tau = 0$) measured *inward.* Also, we shall measure the polar angle θ with respect to the *outward* normal and let φ be the azimuthal angle referred to a suitably chosen x-axis in the plane of stratification. The equation of transfer then takes the form

$$\mu \frac{dI(\tau, \mu, \varphi)}{d\tau} = I(\tau, \mu, \varphi) - \frac{1}{4\pi} \int_{-1}^{+1} \int_{0}^{2\pi} p(\mu, \varphi; \mu', \varphi')\, I(\tau, \mu', \varphi')\, d\mu' d\varphi', \quad (6)$$

where $\mu = \cos\theta$.

In the conservative case ($\varpi_0 = 1$) equation (6) allows the *flux integral*,

$$\pi F = \int_{-1}^{+1} \int_{0}^{2\pi} \mu\, I(\tau, \mu, \varphi)\, d\mu\, d\varphi = \text{constant}. \quad (7)$$

There is a further integral which conservative problems admit: this is the *K-integral*,

$$\begin{aligned} K(\tau) &= \frac{1}{4\pi} \int_{-1}^{+1} \int_{0}^{2\pi} I(\tau, \mu, \varphi) \mu^2\, d\mu\, d\varphi \\ &= \tfrac{1}{4} F[(1 - \tfrac{1}{3}\varpi_1)\, \tau + Q], \end{aligned} \quad (8)$$

where Q is a constant.

In the foregoing framework there are two problems one generally considers: (1) the problem of a semi-infinite atmosphere with a constant net flux, πF, in the conservative case; and (2) the problem of diffuse reflection and transmission when a parallel beam of radiation of net flux πF per unit area normal to itself is incident in some specified direction, $(-\mu_0, \varphi_0)$.

In the first problem, the radiation field, at any point in the atmosphere, is clearly axisymmetric about the z-axis, and the intensity I, now a function only of τ and μ, satisfies the integrodifferential equation

$$\mu \frac{dI(\tau, \mu)}{d\tau} = I(\tau, \mu) - \frac{1}{2} \int_{-1}^{+1} p^{(0)}(\mu; \mu')\, I(\tau, \mu')\, d\mu', \quad (9)$$

where

$$p^{(0)}(\mu; \mu') = \frac{1}{2\pi} \int_{0}^{2\pi} p(\mu, \varphi; \mu', \varphi')\, d\varphi'. \quad (10)$$

And we require to solve equation (9) together with the boundary conditions

$$I(0, -\mu) = 0 \quad (0 < \mu \leq 1) \quad \text{and} \quad I(\tau, \mu)\, e^{-\tau} \to 0 \quad \text{as } \tau \to \infty. \quad (11)$$

In the problem of diffuse reflection and transmission, one considers, in general, an atmosphere of optical thickness τ_1 and asks for the intensity $I(0, +\mu, \varphi)$ $(0 < \mu \leq 1)$ *diffusely reflected* from the surface $\tau = 0$, and the intensity $I(\tau, -\mu, \varphi)$ $(0 < \mu \leq 1)$ *diffusely transmitted* below the surface $\tau = \tau_1$. The resulting laws of diffuse reflection and transmission are expressed in terms of a *scattering* and a *transmission* function,

$$S(\tau_1; \mu, \varphi; \mu_0, \varphi_0) \quad \text{and} \quad T(\tau_1; \mu, \varphi; \mu_0, \varphi_0), \quad (12)$$

defined in terms of the reflected and transmitted intensities by

$$I(0, +\mu, \varphi) = \frac{F}{4\mu} S(\tau_1; \mu, \varphi; \mu_0, \varphi_0)$$

and (13)

$$I(\tau_1, -\mu, \varphi) = \frac{F}{4\mu} T(\tau_1; \mu, \varphi; \mu_0, \varphi_0).$$

It is to be specifically noted that the reflected and the transmitted intensities refer only to the radiation that has suffered one or more scattering processes: $I(\tau_1; -\mu, \varphi)$ does not, for example, include the directly transmitted intensity, $\frac{1}{4}F\, e^{-\tau_1/\mu_0}$, in the direction $(-\mu_0, \varphi_0)$, which has not suffered any scattering process. More generally, in the treatment of the problem of diffuse reflection and transmission, we distinguish between the reduced incident radiation, $\pi F e^{-\tau/\mu_0}$, which penetrates to the depth τ without having suffered any scattering process, and the diffuse radiation field which has arisen from one or more scattering processes suffered by the incident beam. With this distinction, the equation of transfer takes the form

$$\mu \frac{dI(\tau, \mu, \varphi)}{d\tau} = I(\tau, \mu, \varphi) - \frac{1}{4\pi} \int_{-1}^{+1}\int_0^{2\pi} p(\mu, \varphi; \mu', \varphi')\, I(\tau, \mu', \varphi')\, d\mu' d\varphi' - \tfrac{1}{4}F\, e^{-\tau/\mu_0}\, p(\mu, \varphi; -\mu_0, \varphi_0); \tag{14}$$

and we require to solve equation (14) together with the boundary conditions

$$I(0, -\mu, \varphi) = 0 \quad (0 < \mu \leq 1) \quad \text{at} \quad \tau = 0$$

and (15)

$$I(\tau_1, +\mu, \varphi) = 0 \quad (0 < \mu \leq 1) \quad \text{at} \quad \tau = \tau_1.$$

II. The Solutions for Some Typical Problems by the Method of Discrete Ordinates

At the time I became interested in problems of radiative transfer, the two problems formulated in § I had been considered only for the case of conservative isotropic scattering when

$$p(\cos \Theta) = 1, \tag{16}$$

and the appropriate equations of transfer are

$$\mu \frac{dI(\tau, \mu)}{d\tau} = I(\tau, \mu) - \frac{1}{2} \int_{-1}^{+1} I(\tau, \mu')\, d\mu', \tag{17}$$

and

$$\mu \frac{dI(\tau, \mu)}{d\tau} = I(\tau, \mu) - \frac{1}{2} \int_{-1}^{+1} I(\tau, \mu')\, d\mu' - \tfrac{1}{4} F e^{-\tau/\mu_0}. \tag{18}$$

Also, in the context of equation (18) only the problem of diffuse reflection by a semi-infinite atmosphere had been considered. And through the combined researches of E. A. Milne, E. Hopf, and M. Bronstein, the solutions for the angular distributions of the emergent radiations, for the two problems, had been expressed in the forms

$$I(0, \mu) = \frac{\sqrt{3}}{4} F\, H(\mu), \tag{19}$$

and

$$I(0, \mu) = \tfrac{1}{4} F \frac{\mu_0}{\mu + \mu_0} H(\mu)\, H(\mu_0), \tag{20}$$

where $H(0) = 1$ and an integral representation for $H(\mu)$ had been given.

Since only the conservative case of isotropic scattering had been considered, I asked myself whether some systematic method could not be developed for solving the same problems for more general laws of scattering. I was particularly interested in solving the problem for *Rayleigh's phase function*,

$$p(\cos\Theta) = \tfrac{3}{4}(1 + \cos^2\Theta), \tag{21}$$

and the phase function,

$$p(\cos\Theta) = \varpi_0(1 + x\cos\Theta) \quad (0 \le |x| \le 1), \tag{22}$$

for the problem of diffuse reflection. A paper by G. C. Wick in which equation (17) had been treated by a novel method was most opportune for me. Wick's method consisted in replacing the integral over μ', which occurs in equation (17), by a sum using Gauss's quadrature formula. Thus,

$$\int_{-1}^{+1} I(\tau, \mu')\, d\mu' \quad \text{is replaced by} \quad \sum_{j=-n}^{+n} a_j\, I(\tau, \mu_j), \tag{23}$$

where $\pm\mu_j$ $(j = 1, \ldots, n)$ are the zeros of the Legendre polynomial $P_{2n}(\mu)$ and $a_{+j} = a_{-j}$ $(j = 1, \ldots, n)$ are the "Gaussian weights" which satisfy the conditions

$$\sum_{j=-n}^{+n} a_j \mu_j^{\,l} = \frac{2\delta_{l,\text{even}}}{l+1} \quad (l \le 4n - 1), \tag{24}$$

where

$$\left.\begin{aligned} \delta_{l,\text{even}} &= 1 \quad \text{if } l \text{ is even} \\ &= 0 \quad \text{if } l \text{ is odd} \end{aligned}\right\}. \tag{25}$$

By this device, the equations of transfer are replaced by systems of linear equations with constant coefficients for the intensities, $I_{\pm j}(\tau) = I(\tau, \pm\mu_j)$, at the points of the Gaussian division; and the solution of these equations presents no difficulty of principle. The method could clearly be applied to the problems in which I was interested. But what was totally unexpected was that in the general "*n*th approximation" (in which the quadrature formula with $2n$ divisions of the interval $-1 \leq \mu \leq 1$ is used) the solutions for the emergent radiations, in all cases considered, could be obtained in closed forms and expressed in terms of *H-functions* defined in the manner

$$H(\mu) = \frac{1}{\mu_1 \dots \mu_n} \frac{\prod_{i=1}^{n} (\mu + \mu_i)}{\prod_{\alpha=1}^{n} (1 + k_\alpha \mu)}, \tag{26}$$

where the μ_i's are the points of the Gaussian division in the positive half of the interval $(-1, +1)$, and the k_α's are the distinct positive (or zero) roots of a *characteristic equation,*

$$1 = 2 \sum_{j=1}^{n} \frac{a_j \Psi(\mu_j)}{1 - k^2 \mu^2}, \tag{27}$$

and $\Psi(\mu)$ is an even polynomial in μ satisfying the condition

$$\int_0^1 \Psi(\mu) = \sum_{j=0}^{n} a_j \Psi(\mu_j) = 1 \tag{28}$$

—the equality sign always obtaining for the conservative case.

Different physical problems lead to different characteristic equations and therefore to different H-functions. However, as the H-functions differ from one another only through the characteristic equations which determine the characteristic roots, k_α, the function $\Psi(\mu)$ is called the *characteristic function* which defines $H(\mu)$.

We tabulate below the solutions to the problems that were considered in the first instance. (A convenient reference to the historically minded reader is the author's Gibbs Lecture to the American Mathematical Society (given on 26 December 1946).[3]

A. *Isotropic Scattering*

Problem with constant flux: the law of darkening:

$$I(\mu) - \frac{\sqrt{3}}{4} F\, H(\mu). \tag{29}$$

Law of diffuse reflection:

$$I(\mu; \mu_0) = \tfrac{1}{4} F \frac{\mu_0}{\mu + \mu_0} H(\mu)\, H(\mu_0). \tag{30}$$

3. *Bull. Am. Math. Soc.* 53 (1947): 641–711 (paper 24 in this volume).

The characteristic function in terms of which $H(\mu)$ is defined is

$$\Psi(\mu) = \tfrac{1}{2}. \tag{31}$$

B. *Scattering in Accordance with Rayleigh's Phase Function* $\tfrac{3}{4}(1 + \cos^2 \Theta)$
Problem with constant net flux: the law of darkening:

$$I(\mu) = \tfrac{3}{4}qF\, H^{(0)}(\mu). \tag{32}$$

Law of diffuse reflection:

$$\begin{aligned} I(\mu, \varphi; \mu_0, \varphi_0) = \tfrac{3}{32} F \{ & H^{(0)}(\mu)\, H^{(0)}(\mu_0)[3 - (3 - 8q^2)^{1/2}(\mu + \mu_0) + \mu\mu_0] \\ & - 4\mu\mu_0 (1 - \mu^2)^{1/2}(1 - \mu_0^2)^{1/2}\, H^{(1)}(\mu)\, H^{(1)}(\mu_0) \cos(\varphi - \varphi_0) \\ & + (1 - \mu^2)(1 - \mu_0^2)\, H^{(2)}(\mu)\, H^{(2)}(\mu_0) \cos 2(\varphi - \varphi_0)\} \frac{\mu_0}{\mu + \mu_0}. \end{aligned} \tag{33}$$

The characteristic functions in terms of which $H^{(0)}(\mu)$, $H^{(1)}(\mu)$, and $H^{(2)}(\mu)$ are defined are, respectively,

$$\Psi^{(0)}(\mu) = \tfrac{3}{16}(3 - \mu^2), \qquad \Psi^{(1)}(\mu) = \tfrac{3}{8}\mu^2 (1 - \mu^2),$$
$$\text{and} \quad \Psi^{(2)}(\mu) = \tfrac{3}{32}(1 - \mu^2)^2. \tag{34}$$

And, finally,

$$q = \frac{2\sqrt{3}}{H^{(0)}(+\sqrt{3}) - H^{(0)}(-\sqrt{3})}. \tag{35}$$

C. *Scattering in Accordance with the Phase Function* $\varpi_0(1 + x \cos \Theta)$
Law of diffuse reflection:

$$\begin{aligned} I(\mu, \varphi; \mu_0, \varphi_0) = \tfrac{1}{4} \varpi_0 F\{ & H^{(0)}(\mu)\, H^{(0)}(\mu_0)\, [1 - c(\mu + \mu_0) - x(1 - \varpi_0)\mu\mu_0] \\ & + x (1 - \mu^2)^{1/2} (1 - \mu_0^2)^{1/2}\, H^{(1)}(\mu)\, H^{(1)}(\mu_0) \cos(\varphi - \varphi_0)\} \frac{\mu_0}{\mu + \mu_0}. \end{aligned} \tag{36}$$

The characteristic functions in terms of which $H^{(0)}(\mu)$ and $H^{(1)}(\mu)$ are defined are, respectively,

$$\Psi^{(0)}(\mu) = \tfrac{1}{2}\varpi_0 [1 + x(1 - \varpi_0)\mu^2] \quad \text{and} \quad \Psi^{(1)}(\mu) = \tfrac{1}{4} x\varpi_0(1 - \mu^2). \tag{37}$$

The constant c depends in a complicated manner on the characteristic roots defining $H^{(0)}(\mu)$.[4]

Law of diffuse reflection for the case of isotropic scattering with albedo ϖ_0:

$$I(\mu) = \tfrac{1}{4} \varpi_0 F \frac{\mu_0}{\mu + \mu_0} H^{(0)}(\mu)\, H^{(0)}(\mu_0), \tag{38}$$

4. Cf. S. Chandrasekhar, *Ap. J.* 103 (1946): 165, eq. (108). See p. 148 in this volume.

where $H_0(\mu)$ is defined in terms of the characteristic function

$$\Psi(\mu) = \tfrac{1}{2}\varpi_0. \tag{39}$$

It was at this stage that I became aware of Ambartsumian's first paper on the principle of invariance for the problem of diffuse reflection by a semi-infinite atmosphere. An immediate inference that could be drawn from a comparison of Ambartsumian's solution for the isotropic case with the solution (38) was that in the limit of infinite approximation, i.e., in an *exact theory*, the H-functions, which appear in the solutions, derived by the method of discrete ordinates, must be redefined as solutions of the nonlinear integral equation

$$H(\mu) = 1 + \mu\, H(\mu) \int_0^1 \frac{\Psi(\mu')\, H(\mu')}{\mu + \mu'}\, d\mu'. \tag{40}$$

Once this became apparent, it was not difficult to prove that $H(\mu)$ *defined by equations* (26) *and* (27) *is indeed the unique solution of the finite-difference form of equation* (40), namely,

$$H(\mu) = 1 + \mu H(\mu) \sum_{j=1}^{n} \frac{a_j \Psi(\mu_j)\, H(\mu_j)}{\mu + \mu_j}. \tag{41}$$

At this point, it was abundantly clear to me that Ambartsumian's paper had opened an entirely original and novel approach to the problems of radiative transfer. But I had to put the matter aside without a closer examination for some months, since I was, at that time, engrossed in another aspect of the theory of radiative transfer. I consider this aspect, first, in §§ III and IV; and I return to the principles of invariance in § V.

III. The Equations of Transfer Incorporating the Polarization of the Radiation Field—The Problem with a Constant Net Flux

Early in 1946, I became interested in the question of the polarization of the continuous radiation of early-type stars. The reason for my interest was that, in the atmospheres of these stars, hydrogen will be almost completely ionized and the transfer of radiation will be controlled by the scattering by the free electrons. And according to J. J. Thomson's laws governing this process, an incident beam of unpolarized radiation will become partially polarized on scattering. On this account, one may expect the continuous radiation of early-type stars to exhibit partial polarization. To investigate this matter, one requires, of course, to formulate the equation of transfer correctly allowing for the state of polarization of the prevalent radiation field. But no such formulation existed in 1946.

In the context of the problem in plane-parallel atmospheres with no incident radiation, the formulation of the relevant equations of transfer is much simplified, since the axial symmetry of the radiation field, at each point, with respect to the outward normal, implies that the plane of polarization, at each point, coincides with the principal meridian. Therefore, the radiation field can be fully specified by the intensities I_l and I_r, in the two states of polarization, parallel and perpendicular, respectively, to the principal meridian. For Thomson scattering by free electrons, the equations of transfer governing the intensities I_l and I_r are

$$\mu \frac{dI_l(\tau, \mu)}{d\tau} = I_l(\tau, \mu) - \frac{3}{8}\left\{\int_{-1}^{+1} I_l(\tau, \mu')[2(1 - \mu'^2) + \mu^2 (3\mu'^2 - 2)]d\mu' + \mu^2 \int_{-1}^{+1} I_r(\tau, \mu')d\mu'\right\} \tag{42}$$

and

$$\mu \frac{dI_r(\tau, \mu')}{d\tau} = I_r(\tau, \mu) - \frac{3}{8}\left\{\int_{-1}^{+1} I_l(\tau, \mu')\mu'^2 d\mu' + \int_{-1}^{+1} I_r(\tau, \mu')d\mu'\right\}. \tag{43}$$

And solutions of these equations are required which satisfy the boundary conditions

$$I_l(0, \mu) = I_r(0, \mu) = 0 \quad (0 < \mu \leq 1),$$

and (44)

$$I_l(\tau, \mu)\, e^{-\tau} \to 0 \quad \text{and} \quad I_r(\tau, \mu)\, e^{-\tau} \to 0 \quad \text{for } \tau \to \infty.$$

Again, it was found that by the method of discrete ordinates the solutions of equations (42) and (43) for the emergent radiation in the two states of polarization can be found in the closed forms

$$I_l(0, \mu) = \tfrac{3}{8}F(1 - c^2)^{1/2}H_l(\mu)$$

and (45)

$$I_r(0, \mu) = \tfrac{3}{8}F \frac{1}{\sqrt{2}} H_r(\mu)\,(\mu + c),$$

where $H_l(\mu)$ and $H_r(\mu)$ are defined in terms of the characteristic functions

$$\Psi_l = \tfrac{3}{4}(1 - \mu^2) \quad \text{and} \quad \Psi_r(\mu) = \tfrac{3}{8}(1 - \mu^2), \tag{46}$$

and

$$c = \frac{H_l(+1)\, H_r(-1) + H_l(-1)\, H_r(+1)}{H_l(+1)\, H_r(-1) - H_l(-1)\, H_r(+1)}. \tag{47}$$

From the solutions for $I_l(0, \mu)$ and $I_r(0, \mu)$ given by equations (45)–(47), it was concluded that the radiation emerging tangentially at the "limb" ($\theta = \pi/2$

and $\mu = 0$) must be partially polarized to the extent of 11.4%. (The exact solution obtained a few months later gave 11.7%.) Figure 2 in paper 7 (p. 151) illustrates the predicted difference in the laws of darkening in the two states of polarization.

When the effect illustrated in this figure was found in 1946, it seemed to me that it would be worthwhile to look for it during the eclipse of binary stars, one component of which is expected to show the predicted polarization. And I suggested this observation to several leading photoelectric astronomers of the time. Dr. W. A. Hiltner (who was then a colleague of mine at the Yerkes Observatory) undertook, together with Dr. J. Hall, to look for the predicted effect. The first binary star they observed did indeed show a polarization during the last phases of eclipse; but contrary to prediction, the degree of polarization did not change with phase and it persisted even after the eclipse. Hiltner and Hall had discovered interstellar polarization instead!

By a curious coincidence, at the time of this writing (26 December 1982) I have in front of me a preprint of a paper by Dr. A. Kemp of the University of Oregon and his collaborators at Oregon and at the University of New Mexico which concludes with the statement: "We assert that the long-sought Chandrasekhar limb polarization in eclipsing binaries has now been discovered in the star system Algol."

IV. The General Vector Equation of Transfer in Terms of Stokes Parameters

As I stated at the conclusion of my paper written in February 1946 (in which the solution [45] is derived), "the successful solution of a specific problem in the theory of radiative transfer, distinguishing the different states of polarization, justifies the hope that it will be possible to solve other astrophysical problems in which polarization plays a significant part."[5] I was particularly anxious to solve the problem of diffuse reflection by a semi-infinite atmosphere on Rayleigh's laws of scattering. Indeed, a note added on 6 May 1946 to the same paper states: "The problem of diffuse reflection from a semi-infinite plane-parallel atmosphere, allowing for the partial polarization of the diffuse radiation, has now been solved." (The paper giving the solution[6] was communicated on 13 May 1946.) But before the solution was obtained, it was necessary to formulate the equations of radiative transfer in which the plane of polarization is also considered as one of the variables and not as something which is known from symmetry considerations (as in the problem considered in § III). The formulation was possible only after discovering a long-forgotten paper by Stokes published in 1852. Perhaps I may be allowed to go into some detail as to how I came to

5. *Ap. J.* 103 (1946): 351 (paper 7 in this volume).
6. *Ap. J.* 104 (1946): 110 (paper 8 in this volume).

discover Stokes's paper. The following is an abridged version of an account that I wrote more than thirty years ago, unpublished hitherto.

The fundamental question that confronted me was the following. We have an element of gas immersed in an anisotropic radiation field; and the anisotropy refers not only to the dependence of the intensity of the radiation on direction but also to the dependence of the polarization characteristics on direction. Clearly the first question concerns how one is to characterize a pencil of partially polarized radiation. All the standard and not so standard books on optics available until the early fifties contain only the obvious statements to the effect that in order to characterize partially polarized light we should specify the intensity, the plane of polarization, and the degree of polarization; and if the light is elliptically polarized, then we should in addition specify the ellipticity of the ellipse characterizing the polarized part of the radiation. In other words, the parameters of the problem are an intensity, a direction, a ratio, and a geometrical property of an ellipse. For the particular problem we have on hand, all these quantities of diverse dimensions should depend in turn on the level of the atmosphere at which we are and on the direction as well; and all these factors must be incorporated in a basic equation of transfer. As I said, none of the extant books of that time were helpful; and I also sought in vain the advice of several physicists (including G. Herzberg, G. Placzek, J. Wheeler, and G. Breit). Nevertheless, it did not seem to me that the basic question could have been overlooked by the great masters of the nineteenth century. And I recall how one afternoon I took down from the library shelves the collected papers of Rayleigh, Kelvin, and Stokes. I started with Stokes's papers; and as I was glancing down the table of contents of volume 3 of his *Collected Papers*, I came across the title "On the Composition and Resolution of Streams of Polarized Light from Different Sources." I was at once certain that this paper by Stokes ought to contain the answer to my question. This certain belief was confirmed when in the opening paragraph of the paper I read, "But when two polarized streams from different sources mix together, the mixture possesses properties intermediate between those of the original streams. . . ." And Stokes continues with characteristic modesty, "The properties of such mixtures form but an uninviting subject of investigation; and accordingly, though to a certain extent they are obvious, and must have forced themselves upon the attention of all who have paid any special attention to the physical theory of light, they do not seem hitherto to have been studied in detail." That was written in 1852; but no book on optics that had been written for one hundred years after Stokes (with one exception,[7] as I came to know later) had included an account of the topic, and no physicist whom I had consulted seemed even to be aware of the problem.

The meaning of the Stokes parameters for a partially plane-polarized beam

7. J. Walker, *The Analytical Theory of Light* (Cambridge: Cambridge University Press, 1904), §§ 20–22, pp. 28–32.

is particularly simple. If l and r refer to arbitrarily chosen directions at right angles to one another in the plane transverse to the direction of propagation of the beam, the intensity $I(\psi)$ in a direction at an angle ψ to the direction of l must go through two complete cycles as ψ goes through 360°. It must therefore be expressible in the form

$$I(\psi) = \tfrac{1}{2}(I + Q \cos 2\psi + U \sin 2\psi). \tag{48}$$

The coefficients I, Q, and U in this representation are the Stokes parameters. If one is dealing with an elliptically polarized beam, then we ask for the intensity in the direction ψ when we introduce a retardation ε in one component relative to the other; then the corresponding expression is

$$I(\psi) = \tfrac{1}{2}(I + Q \cos 2\psi + U \sin 2\psi \cos \varepsilon - V \sin 2\psi \sin \varepsilon). \tag{49}$$

In terms of the foregoing parameters the plane of polarization (referred to the direction $\psi = 0$) and the ellipticity ($= \tan \beta =$ ratio of axes of the ellipse) are given by

$$\tan 2\chi = \frac{U}{Q} \quad \text{and} \quad \sin 2\beta = \frac{V}{(Q^2 + U^2 + V^2)^{1/2}}. \tag{50}$$

The additive property of the Stokes parameters which makes them ideal for treating transfer problems is evident: *if two independent streams of polarized light are mixed, then the Stokes parameters characterizing the mixture are the sum of the Stokes parameters of the individual streams.*

In terms of the Stokes parameters, a law of scattering is specified by a matrix, $\boldsymbol{R}$ (cos Θ), since an elementary act of scattering results in a linear transformation of the parameters of the incident beam. Consequently, by considering the intensity as a vector $\boldsymbol{I}$, with components I_l, I_r, U, and V (where l and r refer to directions parallel and perpendicular, respectively, to the meridian through the point under consideration and the plane containing the directions of the beam and of the normal to the plane of stratification of the atmosphere) and by replacing the "phase function" commonly introduced to describe the angular distribution of the scattered radiation by a *phase matrix*, $\boldsymbol{P}$, we can formulate the basic equation of transfer without any difficulty of principle. The equation of transfer is

$$\mu \frac{d\boldsymbol{I}(\tau, \mu, \varphi)}{d\tau} = \boldsymbol{I}(\tau, \mu, \varphi) - \frac{1}{4\pi} \int_{-1}^{+1} \int_0^{2\pi} \boldsymbol{P}(\mu, \varphi; \mu', \varphi')\, \boldsymbol{I}(\tau, \mu', \varphi') d\mu' d\varphi'$$
$$- \tfrac{1}{4} e^{-\tau/\mu_0} \boldsymbol{P}(\mu, \varphi; -\mu_0, \varphi_0)\, \boldsymbol{F}, \tag{51}$$

where

$$\boldsymbol{F} = (F_l, F_r, F_U, F_V) \tag{52}$$

is the Stokes vector representing the parallel beam of radiation incident on the atmosphere in the direction $(-\mu_0, \varphi_0)$: πF_l, πF_r, πF_U, and πF_V denote the net fluxes per unit area normal to the beam in the four Stokes parameters; and the phase matrix $\boldsymbol{P}$ is obtained from $\boldsymbol{R}(\cos\Theta)$ by applying to it a linear transformation $\boldsymbol{L}(-i_1)$ on the *right*, to transform $\boldsymbol{I}(\tau,\mu',\varphi')$ to the orientation of the axes to which $\boldsymbol{R}(\cos\Theta)$ is generally specified, *and* a linear transformation $\boldsymbol{L}(\pi-i_2)$ on the *left* to transform the resulting scattered intensity to the orientation of the axes chosen at (μ,φ).[8]

For the particular case of Rayleigh scattering and an incident beam of partially plane-polarized light, the phase matrix is given by[9]

$$\boldsymbol{P}(\mu,\varphi;\mu',\varphi') = \boldsymbol{Q}[\boldsymbol{P}^{(0)}(\mu,\mu') + (1-\mu^2)^{1/2}(1-\mu'^2)^{1/2}\boldsymbol{P}^{(1)}(\mu,\varphi;\mu',\varphi') + \boldsymbol{P}^{(2)}(\mu,\varphi;\mu',\varphi')], \tag{53}$$

where

$$\boldsymbol{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \tag{54}$$

$$\boldsymbol{P}^{(0)}(\mu,\mu') = \frac{3}{4}\begin{bmatrix} 2(1-\mu^2)(1-\mu'^2)+\mu^2\mu'^2 & \mu^2 & 0 \\ \mu'^2 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{55}$$

$$\boldsymbol{P}^{(1)}(\mu,\varphi;\mu',\varphi') = \frac{3}{4}\begin{bmatrix} 4\mu\mu'\cos(\varphi'-\varphi) & 0 & 2\mu\sin(\varphi'-\varphi) \\ 0 & 0 & 0 \\ -2\mu'\sin(\varphi'-\varphi) & 0 & \cos(\varphi'-\varphi) \end{bmatrix}, \tag{56}$$

and

$$\boldsymbol{P}^{(2)}(\mu,\varphi;\mu',\varphi') = \frac{3}{4}\begin{bmatrix} \mu^2\mu'^2\cos 2(\varphi'-\varphi) & -\mu^2\cos 2(\varphi'-\varphi) & \mu^2\mu'\sin 2(\varphi'-\varphi) \\ -\mu'^2\cos 2(\varphi'-\varphi) & \cos 2(\varphi'-\varphi) & -\mu'\sin 2(\varphi'-\varphi) \\ -\mu\mu'^2\sin 2(\varphi'-\varphi) & \mu\sin 2(\varphi'-\varphi) & \mu\mu'\cos 2(\varphi'-\varphi) \end{bmatrix}. \tag{57}$$

The solution of equation (51) for the problem of diffuse reflection and transmission must, of course, satisfy the standard boundary conditions:

$$\boldsymbol{I}(0,-\mu,\varphi) = 0 \quad (0<\mu\le 1,\ \ 0\le\varphi\le 2\pi)$$

and

$$\boldsymbol{I}(\tau_1,+\mu,\varphi) = 0 \quad (0<\mu\le 1,\ \ 0\le\varphi\le 2\pi). \tag{58}$$

And the laws of diffuse reflection and transmission are expressed in terms of a *scattering matrix*, $\boldsymbol{S}(\tau_1;\mu,\varphi;\mu_0,\varphi_0)$, and a *transmission matrix*, $\boldsymbol{T}(\tau_1;\mu,\varphi;\mu_0,\varphi_0)$, such that the reflected and the transmitted intensities are given by

8. For a detailed derivation of eq. (51) see S. Chandrasekhar, *Radiative Transfer* (Oxford: Clarendon Press, 1950), §§ 16 and 17.

9. Cf. S. Chandrasekhar, *Ap. J.* 104 (1946): 110 (see eqs. [49]–[51]).

$$I(0, +\mu, \varphi; \mu_0, \varphi_0) = \frac{1}{4\mu} QS(\tau_1; \mu, \varphi; \mu_0, \varphi_0)F$$

and (59)

$$I(\tau_1; -\mu, \varphi; \mu_0, \varphi_0) = \frac{1}{4\mu} QT(\tau_1; \mu, \varphi; \mu_0, \varphi_0)F.$$

The solution for the problem of diffuse reflection by a semi-infinite atmosphere was found in closed form in the early summer of 1946; and the solution for the scattering matrix that was obtained is given in equation (60) on page 525, where $H_l(\mu)$ and $H_r(\mu)$ are the same H-functions introduced in § III and $H^{(1)}(\mu)$ and $H^{(2)}(\mu)$ are two additional H-functions defined in terms of the characteristic functions,

$$\Psi^{(1)}(\mu) = \tfrac{3}{8}(1 - \mu^2)(1 + 2\mu^2) \quad \text{and} \quad \Psi^{(2)}(\mu) = \tfrac{3}{16}(1 + \mu^2)^2, \tag{61}$$

and the constant c has the same value given in (47).

V. The Impact of Ambartsumian's Principles of Invariance

By the fall of 1946, when I had completed the solution of the problem of diffuse reflection by a semi-infinite Rayleigh-scattering atmosphere, it became imperative that I relate the solutions I had obtained for the various problems to the coupled systems of nonlinear integral equations that the application of Ambartsumian's principles of invariance would provide. Studying then his two papers carefully—his second paper devoted to the diffuse reflection and transmission by atmospheres of finite optical thicknesses had also become available meantime—I felt the need to supplement his ideas in several directions. But first let me state the principles of invariance as Ambartsumian stated them. They are the following:

I. *The law of diffuse reflection by a semi-infinite plane-parallel atmosphere must be invariant to the addition (or subtraction) of layers of arbitrary thicknesses to (or from) the atmosphere.*

II. *The laws of diffuse reflection and transmission by a plane-parallel atmosphere of a finite optical thickness, τ_1, must be invariant to the addition (or removal) of layers of arbitrary optical thicknesses to (or from) the atmosphere at the top (at $\tau = 0$) and the simultaneous removal (or addition) of layers of equal optical thicknesses from (or to) the atmosphere at the bottom (at $\tau = \tau_1$).*

Tracing the implications of these principles, when the layers added or subtracted are of infinitesimal thicknesses, Ambartsumian derived nonlinear integral equations for the scattering and the transmission functions defined in equation (13). These equations, when applied to specific examples of phase functions, lead to coupled systems of nonlinear equations of orders two, four, or eight even for the simplest cases. On the other hand, since the method of discrete ordinates had provided, for the cases considered, solutions in closed

$$
\left(\frac{1}{\mu} + \frac{1}{\mu_0}\right)\boldsymbol{S}(\mu, \varphi; \mu_0, \varphi_0) =
$$

$$
\begin{bmatrix}
\begin{array}{l} 2H_l(\mu)H_l(\mu_0)[1 + \mu\mu_0 - c(\mu + \mu_0)] \\ -4\mu\mu_0(1 - \mu^2)^{1/2}(1 - \mu_0^2)^{1/2}H^{(1)}(\mu)H^{(1)}(\mu_0)\cos(\varphi - \varphi_0) \\ +\ \mu^2\mu_0^2H^{(2)}(\mu)H^{(2)}(\mu_0)\cos 2(\varphi - \varphi_0) \end{array} &
\begin{array}{l} [2(1 - c^2)]^{1/2}H_r(\mu_0)H_l(\mu)(\mu + \mu_0) \\ \\ -\mu^2H^{(2)}(\mu)H^{(2)}(\mu_0)\cos 2(\varphi - \varphi_0) \end{array} &
\begin{array}{l} \\ -2\mu(1 - \mu^2)^{1/2}(1 - \mu_0^2)^{1/2}H^{(1)}(\mu)H^{(1)}(\mu_0)\sin(\varphi - \varphi_0) \\ +\ \mu^2\mu_0H^{(2)}(\mu)H^{(2)}(\mu_0)\sin 2(\varphi - \varphi_0) \end{array} \\[2ex]
\begin{array}{l} [2(1 - c^2)]^{1/2}H_r(\mu)H_l(\mu_0)(\mu + \mu_0) \\ -\mu_0^2H^{(2)}(\mu)H^{(2)}(\mu_0)\cos 2(\varphi - \varphi_0) \end{array} &
\begin{array}{l} H_r(\mu)H_r(\mu_0)[1 + \mu\mu_0 + c(\mu + \mu_0)] \\ +\ H^{(2)}(\mu)H^{(2)}(\mu_0)\cos 2(\varphi - \varphi_0) \end{array} &
\begin{array}{l} \\ -\mu_0H^{(2)}(\mu)H^{(2)}(\mu_0)\sin 2(\varphi - \varphi_0) \end{array} \\[2ex]
\begin{array}{l} -2\mu_0(1 - \mu_0^2)^{1/2}(1 - \mu^2)^{1/2}H^{(1)}(\mu)H^{(1)}(\mu_0)\sin(\varphi - \varphi_0) \\ +\ \mu\mu_0^2H^{(2)}(\mu)H^{(2)}(\mu_0)\sin 2(\varphi - \varphi_0) \end{array} &
\begin{array}{l} \\ -\mu H^{(2)}(\mu)H^{(2)}(\mu_0)\sin 2(\varphi - \varphi_0) \end{array} &
\begin{array}{l} (1 - \mu^2)^{1/2}(1 - \mu_0^2)^{1/2}H^{(1)}(\mu)H^{(1)}(\mu_0)\cos(\varphi - \varphi_0) \\ -\mu\mu_0H^{(2)}(\mu)H^{(2)}(\mu_0)\cos 2(\varphi - \varphi_0) \end{array}
\end{bmatrix}
\tag{60}
$$

forms, it was clear that with the knowledge of the *forms* of the solutions, one should be able to reduce Ambartsumian's coupled systems of equations to one or more H-functions (for problems of reflection by semi-infinite atmospheres) or X- and Y-functions (for problems of reflection and transmission by atmospheres of finite optical thicknesses, as we shall see presently) defined with respect to suitable characteristic functions. Besides, Ambartsumian's principles required generalization to the case when the intensity is represented by a Stokes vector and scattering is governed by a phase matrix instead of a phase function. And finally, it appeared that Ambartsumian's principles could be formulated not so much in the sense of expressing invariance as of grasping the essential mathematical content of the geometrical scaffolding of the physical description that is sought.

Considering, then, the incidence of a parallel beam of radiation, of a net flux πF per unit area normal to itself, on a plane-parallel atmosphere of normal optical thickness τ_1, in a direction $(-\mu_0, \varphi_0)$ and distinguishing the outward, $(0 < \mu \leq 1)$, and the inward, $(-1 \leq \mu < 0)$, directed radiations,

$$I(\tau, +\mu, \varphi) \quad \text{and} \quad I(\tau, -\mu, \varphi) \quad (0 < \mu \leq 1), \tag{62}$$

prevailing at depth τ, we can give mathematical expression to the following four manifest consequences of the *definitions* of the scattering and the transmission functions.

I. *The intensity $I(\tau_1 + \mu, \varphi)$ in the outward direction at any level τ results from the reflection of the reduced incident flux $\pi F e^{-\tau/\mu_0}$ and the diffuse radiation $I(\tau, -\mu', \varphi')$, $(0 < \mu' \leq 1)$, incident on the surface τ, by the atmosphere of optical thickness $(\tau_1 - \tau)$ below τ.*

The mathematical expression of this principle is clearly (see fig. 1)

$$I(\tau, +\mu, \varphi) = \frac{F}{4\mu} e^{-\tau/\mu_0} S(\tau_1 - \tau; \mu, \varphi; \mu_0, \varphi_0)$$
$$+ \frac{1}{4\pi\mu} \int_0^1 \int_0^{2\pi} S(\tau_1 - \tau; \mu, \varphi; \mu', \varphi')\, I(\tau, -\mu', \varphi') d\mu' d\varphi'. \tag{63}$$

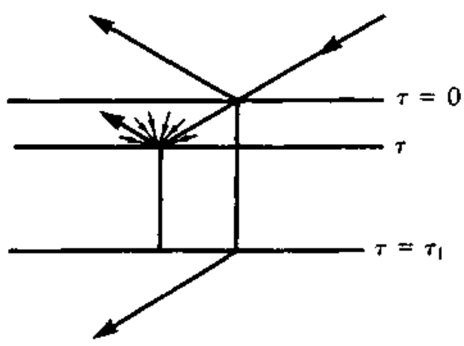


FIG. 1

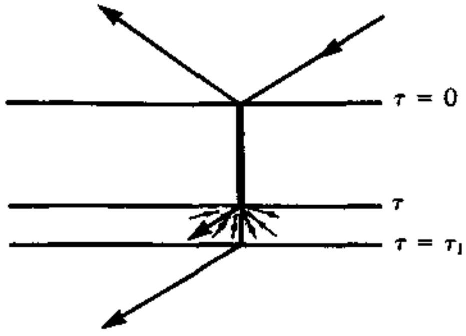


FIG. 2

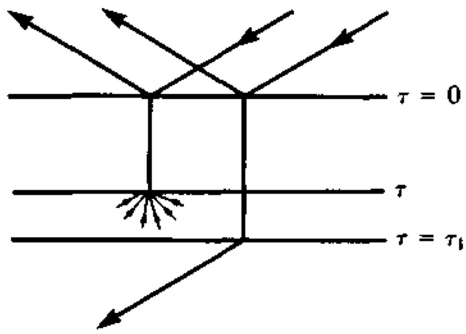


FIG. 3

II. *The intensity $I(\tau, -\mu, \varphi)$ in the inward direction at any level τ results from the transmission of the incident flux by the atmosphere of optical thickness τ, above the surface τ, and the reflection by this same surface of the diffuse radiation $I(\tau, +\mu', \varphi')$, $(0 < \mu' \leq 1)$, incident on it from below.*

The mathematical expression of this principle is (see fig. 2)

$$I(\tau, -\mu, \varphi) = \frac{F}{4\mu} T(\tau; \mu, \varphi; \mu_0, \varphi_0) + \frac{1}{4\pi\mu} \int_0^1 \int_0^{2\pi} S(\tau; \mu, \varphi; \mu', \varphi')\, I(\tau, +\mu', \varphi') d\mu' d\varphi'. \quad (64)$$

III. *The diffuse reflection of the incident light by the entire atmosphere is equivalent to the reflection by the part of the atmosphere of optical thickness τ, above the level τ, and the transmission by this same atmosphere of the diffuse radiation $I(\tau, +\mu', \varphi')$, $(0 < \mu' \leq 1)$, incident on the surface τ from below.*

The mathematical expression of this principle is (see fig. 3)

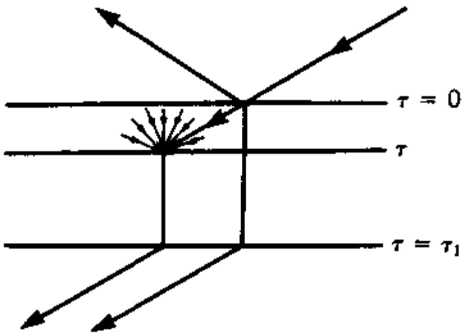


Fig. 4

$$\frac{F}{4\mu} S(\tau_1; \mu, \varphi; \mu_0, \varphi_0) = \frac{F}{4\mu} S(\tau; \mu, \varphi; \mu_0, \varphi_0) + e^{-\tau/\mu} I(\tau, +\mu, \varphi)$$
$$+ \frac{1}{4\pi\mu} \int_0^1 \int_0^{2\pi} T(\tau; \mu, \varphi; \mu', \varphi') \, I(\tau, +\mu', \varphi') d\mu' d\varphi'. \quad (65)$$

IV. *The diffuse transmission of the incident light by the entire atmosphere is equivalent to the transmission of the reduced incident flux $\pi F e^{-\tau/\mu_0}$ and the diffuse radiation $I(\tau, -\mu', \varphi')$, $(0 < \mu' \leq 1)$, incident on the surface τ by the atmosphere of optical thickness $(\tau_1 - \tau)$ below τ.*

The mathematical expression of this principle is (see fig. 4)

$$\frac{F}{4\mu} T(\tau_1; \mu, \varphi; \mu', \varphi') = \frac{F}{4\mu} e^{-\tau/\mu_0} T(\tau_1 - \tau; \mu, \varphi; \mu', \varphi') + e^{-(\tau_1-\tau)/\mu} I(\tau, -\mu, \varphi)$$
$$+ \frac{1}{4\pi\mu} \int_0^1 \int_0^{2\pi} T(\tau_1 - \tau; \mu, \varphi; \mu', \varphi') \, I(\tau, -\mu', \varphi') d\mu' d\varphi'. \quad (66)$$

Similar expressions can be written down which relate the angular distribution of the emergent radiation, in the problem with the constant net flux, to the scattering and the transmission functions.

The basic integral equations which govern the scattering and the transmission functions can be obtained by differentiating equations (63)–(66) with respect to τ and passing to the limit $\tau = 0$ of expressions (63) and (66), and to the limit $\tau = \tau_1$ of expressions (64) and (65). This procedure leads to a set of four integrodifferential equations for the scattering and the transmission functions. Two linear combinations of these equations are equivalent to the integral equations derived by Ambartsumian from his principles of invariance; and we are left with two additional relations.

For the case of isotropic scattering with an albedo ϖ_0, one finds that with the definitions

$$X(\mu) = 1 + \frac{1}{2}\int_0^1 S(\tau_1; \mu, \mu')\,\frac{d\mu'}{\mu'} \tag{67}$$

and

$$Y(\mu) = e^{-\tau_1/\mu} + \frac{1}{2}\int_0^1 T(\tau_1; \mu, \mu')\,\frac{d\mu'}{\mu'}, \tag{68}$$

we obtain four equations,

$$\left(\frac{1}{\mu_0} + \frac{1}{\mu}\right) S(\tau_1; \mu, \mu_0) = \varpi_0[X(\mu)X(\mu_0) - Y(\mu)Y(\mu_0)], \tag{69}$$

$$\left(\frac{1}{\mu_0} - \frac{1}{\mu}\right) T(\tau_1; \mu, \mu_0) = \varpi_0[Y(\mu)X(\mu_0) - X(\mu)Y(\mu_0)], \tag{70}$$

$$\frac{\partial S(\tau_1; \mu, \mu_0)}{\partial \tau_1} = \varpi_0 Y(\mu)Y(\mu_0), \tag{71}$$

and

$$\left(\frac{1}{\mu_0} - \frac{1}{\mu}\right)\frac{\partial T(\tau_1; \mu, \mu_0)}{\partial \tau_1} = \varpi_0\left[\frac{1}{\mu_0}X(\mu)Y(\mu_0) - \frac{1}{\mu}Y(\mu)X(\mu_0)\right]. \tag{72}$$

Inserting expressions (69) and (70) for S and T in equations (67) and (68), we obtain the equations

$$X(\mu) = 1 + \tfrac{1}{2}\varpi_0\mu\int_0^1 \frac{d\mu'}{\mu + \mu'}\,[X(\mu)X(\mu') - Y(\mu)Y(\mu')] \tag{73}$$

and

$$Y(\mu) = e^{-\tau_1/\mu} + \tfrac{1}{2}\varpi_0\mu\int_0^1 \frac{d\mu'}{\mu - \mu'}\,[Y(\mu)X(\mu') - X(\mu)Y(\mu')], \tag{74}$$

and from equations (71) and (72) we obtain

$$\frac{\partial X(\mu, \tau_1)}{\partial \tau_1} = \tfrac{1}{2}\varpi_0 Y(\mu, \tau_1)\int_0^1 \frac{d\mu'}{\mu'}Y(\mu', \tau_1) \tag{75}$$

and

$$\frac{\partial Y(\mu, \tau_1)}{\partial \tau_1} + \frac{Y(\mu, \tau_1)}{\mu} = \tfrac{1}{2}\varpi_0 X(\mu, \tau_1)\int_0^1 \frac{d\mu'}{\mu'}\,Y(\mu', \tau_1). \tag{76}$$

From a comparison of equations (73) and (74) with equation (40), it would appear that the pair of coupled equations

$$X(\mu) = 1 + \mu\int_0^1 \frac{\Psi(\mu')}{\mu + \mu'}\,[X(\mu)X(\mu') - Y(\mu)Y(\mu')]d\mu' \tag{77}$$

and

$$Y(\mu) = e^{-\tau_1/\mu} + \mu \int_0^1 \frac{\Psi(\mu')}{\mu - \mu'} [Y(\mu)X(\mu') - X(\mu)Y(\mu')]d\mu', \tag{78}$$

where $\Psi(\mu)$ is a characteristic function, will play the same role in the theory of radiative transfer in atmospheres of finite optical thicknesses as equation (40) does in the theory of transfer in semi-infinite atmospheres. We may note parenthetically that in the context of equations (77) and (78), equations (75) and (76) are replaced by

$$\frac{\partial X(\mu, \tau_1)}{\partial \tau_1} = Y(\mu, \tau_1) \int_0^1 \frac{d\mu'}{\mu'} \Psi(\mu')Y(\mu', \tau_1) \tag{79}$$

and

$$\frac{\partial Y(\mu, \tau_1)}{\partial \tau_1} + \frac{Y(\mu, \tau_1)}{\mu} = X(\mu, \tau_1) \int_0^1 \frac{d\mu'}{\mu'} \Psi(\mu')Y(\mu', \tau_1). \tag{80}$$

These equations have played important roles in subsequent developments.

We shall consider in § VI below the manner of application of the coupled systems of equations, to which the principles of invariance lead, for laws of scattering more general than isotropic. But it is important to point out that even in the case of isotropic scattering the equations governing the functions $X(\mu)$ and $Y(\mu)$ do not, by themselves, suffice to specify the scattering and the transmission functions in the conservative case when $\varpi_0 = 1$. For, quite generally, when

$$\int_0^1 \Psi(\mu)d\mu = \tfrac{1}{2}, \tag{81}$$

it can be readily verified that if $X(\mu)$ and $Y(\mu)$ are solutions of equations (77) and (78), then so are

$$X(\mu) + Q\mu[X(\mu) + Y(\mu)]$$

and (82)

$$Y(\mu) - Q\mu\,[X(\mu) + Y(\mu)],$$

where Q is an arbitrary constant. On this account, in the conservative case (81), one defines *standard solutions* by the requirements

$$x_0 = \int_0^1 X(\mu)\Psi(\mu)d\mu = 1$$

and (83)[10]

$$y_0 = \int_0^1 Y(\mu)\Psi(\mu)d\mu = 0.$$

10. One can impose these conditions in the case (81), since, in general, we have identity,

$$x_0 = 1 - \left[1 - 2\int_0^1 \Psi(\mu)d\mu + y_0^2\right]^{1/2}.$$

The resulting ambiguity in the solution for the scattering and the transmisson functions can be removed by the consideration of the K-integral. In this manner, one finds that with the X- and Y-functions defined as the standard solutions, the required X- and Y-functions are given by equation (82) with

$$Q = -\frac{\alpha_1 - \beta_1}{(\alpha_1 + \beta_1)\tau_1 + 2(\alpha_2 + \beta_2)}, \tag{84}$$

where α_n and β_n denote the moments of order n of $X(\mu)$ and $Y(\mu)$:

$$\alpha_n = \int_0^1 X(\mu)\mu^n d\mu \quad \text{and} \quad \beta_n = \int_0^1 Y(\mu)\mu^n d\mu. \tag{85}$$

VI. Solutions of the Coupled Systems of Integral Equations which Follow from the Principles of Invariance

Considering first the problem of diffuse reflection by semi-infinite atmospheres for the scattering function, $S(\mu, \varphi; \mu_0, \varphi_0)$, we find that, in the isotropic case, the principle of invariance directly leads to the solution (as Ambartsumian first showed),

$$\left(\frac{1}{\mu_0} + \frac{1}{\mu}\right) S(\mu, \mu_0) = \varpi_0 H(\mu) H(\mu_0), \tag{86}$$

where $H(\mu)$ is defined in terms of the characteristic function, $\frac{1}{2}\varpi_0$. This solution confirms that $H(\mu)$, defined as the solution of the difference equation (41), becomes, in the limit of infinite approximation, the solution of the integral equation (40).

When we proceed to a consideration of problems with even slightly more general phase functions, the reduction of the integral equations, which follows from the principle of invariance, is not as straightforward. Thus, considering the case of Rayleigh's phase function (21), we find that the azimuth-dependent terms in the emergent intensity are, indeed, directly expressed in terms of the H-functions, $H^{(1)}(\mu)$ and $H^{(2)}(\mu)$, defined in equations (33) and (34). But the azimuth-independent term in the scattering function is expressed in the manner

$$\left(\frac{1}{\mu} + \frac{1}{\mu_0}\right) S^{(0)}(\mu, \mu_0) = \tfrac{3}{8}\psi(\mu)\psi(\mu_0) + \tfrac{3}{8}\phi(\mu)\phi(\mu_0), \tag{87}$$

where

$$\psi(\mu) = 3 - \mu^2 + \frac{3}{16}\int_0^1 (3 - \mu'^2) S^{(0)}(\mu, \mu') \frac{d\mu'}{\mu'} \tag{88}$$

and

$$\phi(\mu) = \mu^2 + \frac{3}{16}\int_0^1 \mu'^2 S^{(0)}(\mu, \mu') \frac{d\mu'}{\mu'}. \tag{89}$$

The substitution of equation (87) in equations (88) and (89) provides the pair of coupled nonlinear integral equations

$$\psi(\mu) = 3 - \mu^2 + \tfrac{1}{16}\mu\psi(\mu)\int_0^1 \frac{3-\mu'^2}{\mu+\mu'}\psi(\mu')d\mu' + \tfrac{1}{2}\mu\phi(\mu)\int_0^1 \frac{3-\mu'^2}{\mu+\mu'}\phi(\mu')d\mu' \tag{90}$$

and

$$\phi(\mu) = \mu^2 + \tfrac{1}{16}\,\mu\psi(\mu)\int_0^1 \frac{\mu'^2}{\mu+\mu'}\psi(\mu')d\mu' + \tfrac{1}{2}\,\mu\phi(\mu)\int_0^1 \frac{\mu'^2}{\mu+\mu'}\phi(\mu')d\mu'. \tag{91}$$

I should not have known (in 1946) how to reduce these equations had it not been for the fact that the solution (33) (derived by the method of discrete ordinates) suggests that we seek solutions of the form

$$\psi(\mu) = (3 - c\mu)\,H^{(0)}(\mu) \quad\text{and}\quad \phi(\mu) = q\mu\,H^{(0)}(\mu), \tag{92}$$

where $H^{(0)}(\mu)$ is defined in terms of the characteristic function (cf. eq. [34])

$$\Psi^{(0)}(\mu) = \tfrac{3}{16}(3 - \mu^2), \tag{93}$$

and q and c are constants unspecified in the first instance. The substitution of solutions of the forms (92) in equations (90) and (91) determines the constants q and c. We find

$$q = \frac{2}{3\alpha_1} \quad\text{and}\quad c = \frac{\alpha_2}{\alpha_1}, \tag{94}$$

where α_1 and α_2 are the first and the second moments of $H^{(0)}(\mu)$.

Similarly, for scattering in accordance with the phase function (22), we find that the azimuth-independent term in the scattering function is expressed in the form

$$\left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)S^{(0)}(\mu,\mu_0) = \psi(\mu)\psi(\mu_0) - x\phi(\mu)\phi(\mu_0), \tag{95}$$

where

$$\psi(\mu) = 1 + \tfrac{1}{2}\varpi_0\int_0^1 S^{(0)}(\mu,\mu')\frac{d\mu'}{\mu'} \tag{96}$$

and

$$\phi(\mu) = \mu - \tfrac{1}{2}\varpi_0\int_0^1 S^{(0)}(\mu,\mu')\,d\mu'. \tag{97}$$

The substitution of equation (95) in equations (96) and (97) now leads to the pair of equations

$$\psi(\mu) = 1 + \tfrac{1}{2}\varpi_0\mu\psi(\mu)\int_0^1 \frac{d\mu'}{\mu+\mu'}\,\psi(\mu') - \tfrac{1}{2}x\varpi_0\mu\phi(\mu)\int_0^1 \frac{d\mu'}{\mu+\mu'}\,\phi(\mu') \quad (98)$$

and

$$\phi(\mu) = \mu - \tfrac{1}{2}\varpi_0\mu\psi(\mu)\int_0^1 \frac{d\mu'}{\mu+\mu'}\,\mu'\psi(\mu') + \tfrac{1}{2}x\varpi_0\mu\phi(\mu)\int_0^1 \frac{d\mu'}{\mu+\mu'}\,\mu'\phi(\mu'). \quad (99)$$

Again from the knowledge of solution (36), obtained by the method of discrete ordinates, we find that the required solution of equations (98) and (99) is given by

$$\psi(\mu) = (1 - c\mu)H^{(0)}(\mu) \quad \text{and} \quad \phi(\mu) = q\mu H^{(0)}(\mu), \quad (100)$$

where $H^{(0)}(\mu)$ is defined in terms of the characteristic function $\Psi^{(0)}(\mu)$ (cf. eq. [37]):

$$\Psi^{(0)}(\mu) = \tfrac{1}{2}\varpi_0[1 + x(1 - \varpi_0)\mu^2], \quad (101)$$

and c and q are constants given by

$$c = x\varpi_0(1 - \varpi_0)\frac{\alpha_1}{2 - \varpi_0\alpha_0} \quad \text{and} \quad q = \frac{2(1 - \varpi_0)}{2 - \varpi_0\alpha_0}, \quad (102)$$

where α_n denotes the moment of order n of $H^{(0)}(\mu)$.

Finally, considering the case of diffuse reflection by a Rayleigh-scattering atmosphere, we find that the principle of invariance (now generalized to allow for scattering in accordance with a phase matrix) requires that the azimuth-independent part of the scattering matrix is of the form

$$\left(\frac{1}{\mu'} + \frac{1}{\mu}\right)S^{(0)}(\mu, \mu') = \begin{bmatrix} \psi(\mu) & \phi(\mu)\sqrt{2} \\ \chi(\mu) & \zeta(\mu)\sqrt{2} \end{bmatrix}\begin{bmatrix} \psi(\mu') & \chi(\mu') \\ \phi(\mu')\sqrt{2} & \zeta(\mu')\sqrt{2} \end{bmatrix}, \quad (103)$$

where

$$\psi(\mu) = \mu^2 + \frac{3}{8}\int_0^1 \frac{d\mu'}{\mu'}[\mu'^2 S_{ll}^{(0)}(\mu, \mu') + S_{lr}^{(0)}(\mu, \mu')], \quad (104)$$

$$\phi(\mu) = 1 - \mu^2 + \frac{3}{8}\int_0^1 \frac{d\mu'}{\mu'}(1 - \mu'^2)S_{ll}^{(0)}(\mu, \mu'), \quad (105)$$

$$\chi(\mu) = 1 + \frac{3}{8}\int_0^1 \frac{d\mu'}{\mu'}[\mu'^2 S_{rl}^{(0)}(\mu, \mu') + S_{rr}^{(0)}(\mu, \mu')], \quad (106)$$

and

$$\zeta(\mu) = \frac{3}{8}\int_0^1 \frac{d\mu'}{\mu'}(1 - \mu'^2)S_{rl}^{(0)}(\mu, \mu'). \quad (107)$$

And we are led to the following coupled system of equations of order 4:

$$\psi(\mu) = \mu^2 + \tfrac{3}{8}\mu\psi(\mu) \int_0^1 \frac{d\mu'}{\mu + \mu'} [\mu'^2 \psi(\mu') + \chi(\mu')]$$
$$+ \tfrac{3}{4}\mu\phi(\mu) \int_0^1 \frac{d\mu'}{\mu + \mu'} [\mu'^2\phi(\mu') + \zeta(\mu')], \tag{108}$$

$$\phi(\mu) = 1 - \mu^2 + \tfrac{3}{8}\mu\psi(\mu) \int_0^1 \frac{d\mu'}{\mu + \mu'} (1 - \mu'^2)\psi(\mu')$$
$$+ \tfrac{3}{4}\mu\phi(\mu) \int_0^1 \frac{d\mu'}{\mu + \mu'} (1 - \mu'^2)\phi(\mu'), \tag{109}$$

$$\chi(\mu) = 1 + \tfrac{3}{8}\mu\chi(\mu) \int_0^1 \frac{d\mu'}{\mu + \mu'} [\mu'^2 \psi(\mu') + \chi(\mu')]$$
$$+ \tfrac{3}{4}\mu\zeta(\mu) \int_0^1 \frac{d\mu'}{\mu + \mu'} [\mu'^2\phi(\mu') + \zeta(\mu')], \tag{110}$$

and

$$\zeta(\mu) = \tfrac{3}{8}\mu\chi(\mu) \int_0^1 \frac{d\mu'}{\mu + \mu'} (1 - \mu'^2)\, \psi(\mu')$$
$$+ \tfrac{3}{4}\mu\zeta(\mu) \int_0^1 \frac{d\mu'}{\mu + \mu'} (1 - \mu'^2)\phi(\mu'). \tag{111}$$

From the solution for the scattering matrix given in equation (60), we now deduce that the solution of the foregoing equations is given by

$$\psi(\mu) = q\mu\, H_l(\mu), \quad \phi(\mu) = (1 - c\mu)\, H_l(\mu),$$
$$\zeta(\mu) = \tfrac{1}{2}q\mu\, H_r(\mu), \quad \text{and} \quad \chi(\mu) = (1 + c\mu)\, H_r(\mu), \tag{112}$$

where $H_l(\mu)$ and $H_r(\mu)$ are defined in terms of the same characteristic functions, Ψ_l and Ψ_r, given in equations (46), and q and c are constants given by

$$q = \frac{8(A_1 + 2\alpha_1) - 6(A_0\alpha_1 + \alpha_0 A_1)}{3(A_1{}^2 + 2\alpha_1{}^2)} \tag{113}$$

and

$$c = \frac{8(A_1 - \alpha_1) + 3(2\alpha_1\alpha_0 - A_1A_0)}{3(A_1{}^2 + 2\alpha_1{}^2)}, \tag{114}$$

and α_n and A_n are the moments of order n of $H_l(\mu)$ and $H_r(\mu)$.

Turning next to the problem of diffuse reflection and transmission by atmospheres of finite optical thicknesses, we find that they are considerably more complex than the same problems in the context of semi-infinite atmospheres, at both levels: at the level of eliminating the constants and of obtaining solutions in closed forms in the method of discrete ordinates *and* at the level of reducing the coupled systems of equations, derived from the principles of invariance, to a few X- and Y-equations. The complexity of the analysis does not permit a review of even the most salient features. But I shall briefly dwell on an aspect of the problem of the elimination of the constants, in the method of discrete ordinates, which does not appear to have attracted any attention in the subsequent literature: it is the central role played by a problem which, in some sense, is a generalization of the problem underlying Lagrange's interpolation formula. The particular problem[11] is the determination of two polynomials, $C_0(\mu)$ and $C_1(\mu)$, both of degree n, which are related in the manner

$$C_0(1/k_\alpha) = e^{k_\alpha \tau_1} \frac{P(-1/k_\alpha)}{P(+1/k_\alpha)} C_1(-1/k_\alpha) \qquad (\alpha = \pm 1, \ldots, \pm n), \tag{115}$$

where $\pm k_\alpha (\alpha = 1, \ldots, n)$ are the distinct roots of the characteristic equation (27) and

$$P(\mu) = \prod_{i=1}^{n} (\mu - \mu_i). \tag{116}$$

(We are ignoring for the present the conservative case when the characteristic equation allows a pair of zero roots.) The solution of the problem (none too straightforward to obtain) is

$$C_0(\mu) = \sum_{2^{n-1}\text{ terms}}^{l=n,n-2,\ldots} \varepsilon_l^{(0)} \frac{\prod_{i=1}^{l}\prod_{m=1}^{n-l}(k_{r_i} + k_{s_m})}{\prod_{i=1}^{l}\prod_{m=1}^{n-l}(k_{r_i} - k_{s_m})} \prod_{i=1}^{l}(1 + k_{r_i}\mu) \prod_{m=1}^{n-l} \frac{1}{\lambda_{s_m}}(1 - k_{s_m}\mu) \tag{117}$$

and

$$C_1(\mu) = (-1)^{n-1} \sum_{2^{n-1}\text{ terms}}^{l=n-1,n-3,\ldots} \varepsilon_l^{(1)} \frac{\prod_{i=1}^{l}\prod_{m=1}^{n-l}(k_{r_i} + k_{s_m})}{\prod_{i=1}^{l}\prod_{m=1}^{n-l}(k_{r_i} - k_{s_m})}$$

$$\times \prod_{i=1}^{l}(1 + k_{r_i}\mu) \prod_{m=1}^{n-l} \frac{1}{\lambda_{s_m}}(1 - k_{s_m}\mu), \tag{118}$$

where $r_1, \ldots, r_l$ and $s_1, \ldots, s_{n-l}$ are selections of l and $n - l$ distinct integers from the set $(1, 2, \ldots, n)$,

11. The abstract mathematical problem that is posed is the determination of two polynomials, $F(x)$ and $G(x)$, both of degree n, which are related in the manner

$$F(x_j) = +\lambda_j F(-x_j) \qquad \text{and} \qquad G(x_j) = -\lambda_j G(-x_j),$$

where x_j $(j = 1, \ldots, n)$ are n distinct values of the argument and λ_j $(j = 1, \ldots, n)$ are n assigned numbers, all different from one another.

$$\left.\begin{aligned}
\varepsilon_l^{(0)} &= +1 \quad \text{for integers of the form } n - 4m \\
&= -1 \quad \text{for integers of the form } n - 4m - 2 \\
&= 0 \quad\;\; \text{otherwise,} \\
\varepsilon_l^{(1)} &= +1 \quad \text{for integers of the form } n - 4m - 1 \\
&= -1 \quad \text{for integers of the form } n - 4m - 3 \\
&= 0 \quad\;\; \text{otherwise,} \\
\lambda_\alpha &= e^{k_\alpha \tau_1} \frac{P(-1/k_\alpha)}{P(+1/k_\alpha)} \qquad (\alpha = \pm 1, \ldots, \pm n).
\end{aligned}\right\} \tag{119}$$

And the functions in terms of which the solutions to the various problems are obtained in closed forms are

$$X(\mu) = \frac{(-1)^n}{\mu_1 \ldots \mu_n} \frac{1}{[C_0^2(0) - C_1^2(0)]^{1/2}} \frac{1}{W(\mu)} [P(-\mu)C_0(-\mu) - e^{-\tau_1/\mu}P(\mu)C_1(\mu)]$$

and (120)

$$Y(\mu) = \frac{(-1)^n}{\mu_1 \ldots \mu_n} \frac{1}{[C_0^2(0) - C_1^2(0)]^{1/2}} \frac{1}{W(\mu)} [e^{-\tau_1/\mu}P(\mu)C_0(\mu) - P(-\mu)C_1(-\mu)],$$

where

$$W(\mu) = \sum_{\alpha=1}^{n} (1 - k_\alpha^2 \mu^2). \tag{121}$$

In the limit of infinite approximation, these rational functions are replaced, in the exact theory, by the X- and Y-functions defined by the integral equations (77) and (78).

In the conservative case, the characteristic equation allows a pair of zero roots. We then restrict relations (115) to the nonvanishing roots and consider polynomials, $C_0(\mu)$ and $C_1(\mu)$, of degree $(n - 1)$. The corresponding expressions for $X(\mu)$ and $Y(\mu)$, defined in the same manner as in equations (120), become, in the exact theory, the *standard solutions* defined in § V. The rigorous justification of these "empirical" facts remains to be established.

The facts then are that for the different laws of scattering considered in §§ II and IV, the solutions for the scattering and the transmission functions (or matrices) can be obtained in closed forms. With the forms of the solutions thus known, and with the correspondence enunciated between the rational representations (120) and the exact X- and Y-functions, we can complete the solution of the coupled systems of equations, derived from the principles of invariance, in the manner we have described earlier in the context of semi-infinite atmospheres. I shall leave the matter with this bare statement.

VII. The Polarization of the Sunlit Sky

I now turn to a problem in the theory of radiative transfer which bears on the polarization of the sunlit sky. I shall begin by tracing briefly the history of this problem.

In 1871, Lord Rayleigh at the age of twenty-nine and in his eighth published paper (his six volumes of *Scientific Papers* list four hundred and forty-six papers), entitled "On the Light from the Sky, Its Polarization and Color," accounted for the principal features of the phenomena to which his paper was addressed in terms of the law of scattering that has since come to be known under his name. In particular, he interpreted the blue color of the sky as a result of the inverse fourth power of the wavelength which appears in his law of scattering; and he interpreted the almost complete polarization of the sky in a direction at right angles to the sun in terms of the complete polarization predicted by his law of scattering in that direction. But of course at no point of the sky is the light completely polarized. Lord Rayleigh refers to this fact and interprets it correctly by attributing the incomplete polarization to multiple scattering. But a precise formulation and solution to Rayleigh's problem had to wait for nearly three-quarters of a century. In the meantime, however, Babinet, Brewster, and Arago had made interesting discoveries concerning the directions in which the sunlight is unpolarized. For angles of incidence not exceeding 70°, the *neutral points* (that is, the points of zero polarization) occur at about 10°–20° above and below the sun; these are the points of Babinet and Brewster. And when the sun is low, then near the horizon opposite the sun and about 20° above the antisolar point, a neutral point occurs: this is the Arago point. In fact, the setting of the Brewster point in the western sky coincides with the rising of the Arago point in the eastern sky. These facts should be contrasted with what should be expected on Rayleigh's laws of scattering, namely, that the polarization should be zero in the forward and in the backward directions.

The neutral points were observed and discussed a great deal during the nineteenth century, culminating in the earlier years of this century in the monumental work of the Swiss meteorologist C. Dorno. Dorno not only observed the neutral points on the principal meridian, but he also investigated in detail the continuation of these neutral points over the entire hemisphere along the so-called *neutral lines.* These neutral lines separate the regions of negative polarization from the regions of positive polarization and show a remarkable dependence on the direction of the sun, as can be seen in figure 4 of Paper 11 in this volume (p. 211). To explain these curves, one needs an exact solution of the underlying problem in radiative transfer.

It is clear that a rigorous theory allowing for all orders of scattering cannot be developed without deriving the equations governing the transfer of radiation in an atmosphere in which each element scatters in accordance with Rayleigh's laws. The need for a development along these lines was clearly stated by L. V. King in 1913.[12] But he also stated:

> The complete solution of the problem from this aspect would require us to split up the incident radiation into two components one of which is polarized in the principal plane and the other at right angles to it: the effect of self-illumination would lead to simultaneous integral equations in three variables the solution of which would be much too complicated to be useful.

With this comment, King proceeded to a highly approximative treatment of the problem; and so did others who followed him.[13]

With the equation of transfer now properly formulated in vector form for the Stokes parameters, the problem we are required to solve is no more than the extension of the solution (60), obtained in the context of semi-infinite atmospheres, to atmospheres of finite optical thicknesses. The required solution was obtained by the procedure described toward the end of § VI. Thus, the solutions for the scattering and the transmission matrices were first obtained in closed forms by the method of discrete ordinates.[14] With the forms of the solutions thus determined, the exact solutions were then obtained with the aid of the systems of equations derived from the proper vector generalizations of the principles of invariance. Without further ado, I shall simply transcribe the solution that was obtained in the fall of 1947.[15]

Writing the scattering and the transmission matrices in the forms

$$\begin{aligned} \boldsymbol{S}(\mu, \varphi; \mu_0, \varphi_0) = \boldsymbol{Q}[\tfrac{3}{4}\boldsymbol{S}^{(0)}(\mu; \mu_0) &+ (1 - \mu^2)^{1/2} (1 - \mu_0^2)^{1/2} \boldsymbol{S}^{(1)}(\mu, \varphi; \mu_0, \varphi_0) \\ &+ \boldsymbol{S}^{(2)}(\mu, \varphi; \mu_0, \varphi_0)] \end{aligned} \tag{122}$$

and

$$\begin{aligned} \boldsymbol{T}(\mu, \varphi; \mu_0, \varphi_0) = \boldsymbol{Q}[\tfrac{3}{4}\boldsymbol{T}^{(0)}(\mu; \mu_0) &+ (1 - \mu^2)^{1/2} (1 - \mu_0^2)^{1/2} \boldsymbol{T}^{(1)}(\mu, \varphi; \mu_0, \varphi_0) \\ &+ \boldsymbol{T}^{(2)}(\mu, \varphi; \mu_0, \varphi_0)], \end{aligned} \tag{123}$$

we find, directly, that the dependence of the azimuth-dependent terms ($\boldsymbol{S}^{(1)}$, $\boldsymbol{T}^{(1)}$) and ($\boldsymbol{S}^{(2)}$, $\boldsymbol{T}^{(2)}$) on ($\varphi - \varphi_0$) is essentially the same as that of $\boldsymbol{P}^{(1)}$ and $\boldsymbol{P}^{(2)}$ in expression (53) for the phase matrix. Indeed, we find that

12. *Phil. Trans. Roy. Soc. London, A* 212 (1913): 375.
13. E.g., A. Hammad and S. Chapman, *Phil. Mag.*, ser. 7, 28 (1939): 99.
14. *Ap. J.* 106 (1947): 184–216.
15. *Ap. J.* 107 (1948): 199–214.

$$\left(\frac{1}{\mu_0}+\frac{1}{\mu}\right)\boldsymbol{S}^{(i)} = [X^{(i)}(\mu)X^{(i)}(\mu_0) - Y^{(i)}(\mu)Y^{(i)}(\mu_0)]\, \boldsymbol{P}^{(i)}(\mu, \varphi; -\mu_0, \varphi_0)$$

$$(i = 1, 2), \quad (124)$$

and

$$\left(\frac{1}{\mu_0}-\frac{1}{\mu}\right)\boldsymbol{T}^{(i)} = [Y^{(i)}(\mu)X^{(i)}(\mu_0) - X^{(i)}(\mu)Y^{(i)}(\mu_0)]\, \boldsymbol{P}^{(i)}(-\mu, \varphi; -\mu_0, \varphi_0)$$

$$(i = 1, 2), \quad (125)$$

where $(X^{(1)}, Y^{(1)})$ and $(X^{(2)}, Y^{(2)})$ are a pair of X- and Y-functions belonging to the characteristic functions $\Psi^{(1)}$ and $\Psi^{(2)}$ defined in equations (61).

In contrast to the azimuth-dependent terms, the azimuth-independent terms, $\boldsymbol{S}^{(0)}$ and $\boldsymbol{T}^{(0)}$, have a very complicated structure. They have the forms

$$\left(\frac{1}{\mu_0}+\frac{1}{\mu}\right)\boldsymbol{S}^{(0)}(\mu;\mu_0) = \begin{bmatrix} \psi(\mu) & \phi(\mu)\sqrt{2} & 0 \\ \chi(\mu) & \zeta(\mu)\sqrt{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \psi(\mu_0) & \chi(\mu_0) & 0 \\ \phi(\mu_0)\sqrt{2} & \zeta(\mu_0)\sqrt{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$
$$- \begin{bmatrix} \xi(\mu) & \eta(\mu)\sqrt{2} & 0 \\ \sigma(\mu) & \theta(\mu)\sqrt{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \xi(\mu_0) & \sigma(\mu_0) & 0 \\ \eta(\mu_0)\sqrt{2} & \theta(\mu_0)\sqrt{2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (126)$$

and

$$\left(\frac{1}{\mu_0}-\frac{1}{\mu}\right)\boldsymbol{T}^{(0)}(\mu;\mu_0) = \begin{bmatrix} \xi(\mu) & \eta(\mu)\sqrt{2} & 0 \\ \sigma(\mu) & \theta(\mu)\sqrt{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \psi(\mu_0) & \chi(\mu_0) & 0 \\ \phi(\mu_0)\sqrt{2} & \zeta(\mu_0)\sqrt{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$
$$- \begin{bmatrix} \psi(\mu) & \phi(\mu)\sqrt{2} & 0 \\ \chi(\mu) & \zeta(\mu)\sqrt{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \xi(\mu_0) & \sigma(\mu_0) & 0 \\ \eta(\mu_0)\sqrt{2} & \theta(\mu_0)\sqrt{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (127)$$

where ψ, ϕ, χ, etc., are eight functions which satisfy a coupled system of non-linear integral equations. By the procedure described earlier, we find that these eight functions are expressible in terms of a pair of X- and Y-functions, (X_l, Y_l) and (X_r, Y_r) belonging to the characteristic functions Ψ_l and Ψ_r defined in equations (46). However, since Ψ_l belongs to the conservative class, we choose for (X_l, Y_l) the standard solutions. In terms of (X_r, Y_r) and the standard solutions (X_l, Y_l), we find

$$\psi(\mu) = \mu[\nu_1 Y_l(\mu) - \nu_2 X_l(\mu)],$$
$$\xi(\mu) = \mu[\nu_2 Y_l(\mu) - \nu_1 X_l(\mu)],$$
$$\phi(\mu) = (1 + \nu_4\mu)X_l(\mu) - \nu_3\mu Y_l(\mu),$$
$$\eta(\mu) = (1 - \nu_4\mu)Y_l(\mu) + \nu_3\mu X_l(\mu),$$
$$\chi(\mu) = (1 - u_4\mu)X_r(\mu) + u_3\mu Y_r(\mu) + Q(u_4 - u_3)\mu^2[X_r(\mu) - Y_r(\mu)],$$

$$\sigma(\mu) = (1 + u_4\mu)Y_r(\mu) - u_3\mu X_r(\mu) - Q(u_4 - u_3)\mu^2[X_r(\mu) - Y_r(\mu)],$$
$$\zeta(\mu) = \tfrac{1}{2}\mu[\nu_1 Y_r(\mu) - \nu_2 X_r(\mu)] + \tfrac{1}{2}Q(\nu_2 - \nu_1)\mu^2[X_r(\mu) - Y_r(\mu)],$$
$$\theta(\mu) = \tfrac{1}{2}\mu[\nu_2 Y_r(\mu) - \nu_1 X_r(\mu)] - \tfrac{1}{2}Q(\nu_2 - \nu_1)\mu^2[X_r(\mu) - Y_r(\mu)], \tag{128}$$

where the constants $\nu_1, \nu_2, \nu_3, \nu_4, u_3, u_4$, and Q are determined by the following formulae:

$$\nu_2 + \nu_1 = 2\Delta_1(\kappa_1\delta_1 - \kappa_2\delta_2); \quad \nu_2 - \nu_1 = 2\Delta_2(\kappa_1\delta_1 - \kappa_2\delta_2),$$
$$\nu_4 + \nu_3 = \Delta_1(d_1\kappa_1 - d_0\kappa_2); \quad \nu_4 - \nu_3 = \Delta_2[c_1\delta_1 - c_0\delta_2 - 2Q(d_0\delta_1 - d_1\delta_2)],$$
$$u_4 + u_3 = \Delta_1(c_1\delta_1 - c_0\delta_2); \quad u_4 - u_3 = \Delta_2(d_1\kappa_1 - d_0\kappa_2),$$
$$\Delta_1 = (d_0\delta_1 - d_1\delta_2)^{-1}; \quad \Delta_2 = [c_0\kappa_1 - c_1\kappa_2 - 2Q(d_1\kappa_1 - d_0\kappa_2)]^{-1},$$
$$Q = (c_0 - c_2)[(d_0 - d_2)\tau_1 + 2(d_1 - d_3)]^{-1},$$
$$c_0 = A_0 + B_0 - \tfrac{8}{3}; \quad d_0 = A_0 - B_0 - \tfrac{8}{3},$$
$$c_n = A_n + B_n; \quad d_n = A_n - B_n; \quad \kappa_n = \alpha_n + \beta_n; \quad \delta_n = \alpha_n - \beta_n$$
$$(n = 1, 2, 3, \ldots), \tag{129}$$

and α_n, β_n, A_n, and B_n are the moments of order n of X_l, Y_l, X_r, and Y_r, respectively.

In the theory of the illumination of the sky we are interested in the transmitted light in the case of incident natural light. In this latter case $F_l = F_r = \tfrac{1}{2}F$ (where πF denotes the net flux of the incident natural light) and $F_U = 0$. The equations governing the intensity and polarization of the sky, as witnessed by an observer at $\tau = \tau_1$ in these circumstances, readily follow from the solutions already given; thus by setting $\boldsymbol{F} = \tfrac{1}{2}(F, F, 0)$ and combining equations (123), (125), and (127) appropriately, we find

$$I_l(\tau_1; -\mu, \varphi; \mu_0, \varphi_0) = \tfrac{3}{32}\{[\psi(\mu_0) + \chi(\mu_0)]\,\xi(\mu) + 2[\phi(\mu_0) + \zeta(\mu_0)]\eta(\mu)$$
$$- [\xi(\mu_0) + \sigma(\mu_0)]\psi(\mu) - 2[\theta(\mu_0) + \eta(\mu_0)]\phi(\mu)$$
$$+ 4\mu\mu_0(1 - \mu^2)^{1/2}(1 - \mu_0^2)^{1/2}[X^{(1)}(\mu_0)Y^{(1)}(\mu) - Y^{(1)}(\mu_0)X^{(1)}(\mu)]\cos(\varphi_0 - \varphi)$$
$$- \mu^2(1 - \mu_0^2)[X^{(2)}(\mu_0)Y^{(2)}(\mu) - Y^{(2)}(\mu_0)X^{(2)}(\mu)]\cos 2(\varphi_0 - \varphi)\}\,\frac{F\mu_0}{\mu - \mu_0}, \tag{130}$$

$$I_r(\tau_1; -\mu, \varphi; \mu_0, \varphi_0) = \tfrac{3}{32}\{[\psi(\mu_0) + \chi(\mu_0)]\,\sigma(\mu) + 2[\phi(\mu_0) + \zeta(\mu_0)]\theta(\mu)$$
$$- [\xi(\mu_0) + \sigma(\mu_0)]\chi() - 2[\theta(\mu_0) + \eta(\mu_0)]\zeta(\mu)$$
$$+ (1 - \mu_0^2)[X^{(2)}(\mu_0)Y^{(2)}(\mu) - Y^{(2)}(\mu_0)X^{(2)}(\mu)]\cos 2(\varphi_0 - \varphi)\}\,\frac{F\mu_0}{\mu - \mu_0}, \tag{131}$$

and

$$U(\tau_1; -\mu, \varphi; \mu_0, \varphi_0) =$$
$$\tfrac{3}{16}\{2(1 - \mu^2)^{1/2}(1 - \mu_0^2)^{1/2}\mu_0[X^{(1)}(\mu_0)Y^{(1)}(\mu) - Y^{(1)}(\mu_0)X^{(1)}(\mu)]\sin(\varphi_0 - \varphi)$$

$$- \mu(1 - \mu_0^2)[X^{(2)}(\mu_0)Y^{(2)}(\mu) - Y^{(2)}(\mu_0)X^{(2)}(\mu)] \sin 2(\varphi_0 - \varphi)\} \frac{F\mu_0}{\mu - \mu_0}. \quad (132)$$

Equations (130)–(132) can be used to determine the distribution of the polarization over the sky for various zenith distances of the sun. Figure 5 in paper 11 (p. 212) illustrates the predicted neutral lines for an atmosphere of optical thickness 0.15. It is seen that the theoretically delineated neutral lines show a remarkable correspondence with Dorno's observations.

And so, at long last, the problem of the polarization of the sunlit sky reached its port.

May I, in concluding this personal account, express my deep-felt gratitude to Director Professor L. V. Mirzoyan, Professor M. A. Mnatsakanian, and the other organizers of the symposium for this opportunity to recall the happiest years of my scientific life? I could not have chosen a finer occasion than the present, since the incisive and original ideas of Academician Ambartsumian provided so much sustenance and inspiration to my efforts of those years.

Postscript

I greatly regret that in my account I have not acknowledged the important contributions to the theory of radiative transfer by the "Leningrad School," and most especially those of Professors V. V. Sobolev and V. V. Ivanov. Fortunately, an adequate account of these contributions has been given by Professor I. N. Minin (*Astrophysica* [Academy of Sciences of the Armenian Republic of USSR] 17 [1980]: 585–618).

I must also refer to the contributions of Professor T. W. Mullikin, who has, in particular, given an alternative (and a more rigorous) derivation of the solution, quoted in § VII, to the problem of diffuse reflection and transmission by a Rayleigh-scattering atmosphere; see in particular the papers by T. W. Mullikin, "Radiative Transfer in Finite Homogeneous Atmospheres with Anisotropic Scattering. I. Linear Singular Equations" (*Ap. J.* 139 [1964]: 379–95), and "The Complete Rayleigh-scattered Field within a Homogeneous Plane-parallel Atmosphere" (*Ap. J.* 145 [1966]: 886–931).

As I stated at the outset, my account has been confined exclusively to the years 1943–47, when I was engrossed in this subject: after writing my book on *Radiative Transfer* (Oxford: Clarendon Press, 1950) during the spring and summer of 1948, I have only on rare occasions returned to the subject. For an account which brings the subject to the present (with full references to the literature), the reader should consult the two magnificent volumes of H. C. van de Hulst (*Multiple Scattering: Tables, Formulas, and Applications,* vols. 1 and 2 [New York: Academic Press, 1980]).

The Formation of Absorption Lines in a Moving Atmosphere

S. Chandrasekhar
Yerkes Observatory of the University of Chicago, Williams Bay, Wisconsin

1. INTRODUCTION

THE phenomenon of the scattering of light in a moving atmosphere has considerable interest for astrophysics. It occurs in Novae, Wolf-Rayet stars, planetary nebulae, the solar prominences, and the Corona. And more recently Struve's studies[1] of the spectra of stars like 48 Librae and 17 Leporis have emphasized its importance for stellar spectroscopy in general. But on consideration one soon realizes the unusual difficulties which must confront a rigorous theoretical analysis of these problems. For, in atmospheres in which large scale motions are present, on account of Doppler effect, the radiation scattered in different directions will have different frequencies, and, as a result of this, the radiation field in the different frequencies will interact with each other in a manner which is not always easy to visualize. However, in the astrophysical contexts, two circumstances simplify the problem. First, the velocities which are involved are small compared to the velocity of light, c, and second, the only effects of consequence are those which arise from the sensitive dependence of the scattering coefficient $\sigma(\nu)$ on the frequency ν. This last circumstance in particular allows us to ignore all effects such as aberration etc., and concentrate only on the effects arising from the change of frequency on scattering. The equation of transfer appropriate to these conditions has been written down by W. H. McCrea and K. K. Mitra.[2] But these writers did not succeed in solving any specific problem. However, we shall show how with certain approximations explicit solutions can be found which illustrate the effects which may be expected in the contours of absorption lines formed in an atmosphere in which differential motions exist. On the mathematical side, the novelty of the problem arises from the very unusual type of boundary value problem in hyperbolic equations which it presents.

2. THE EQUATION OF TRANSFER AND ITS APPROXIMATE FORMS

We shall consider an atmosphere stratified in parallel planes and in which all the properties are assumed to be constant over the planes z = constant (see Fig. 1).

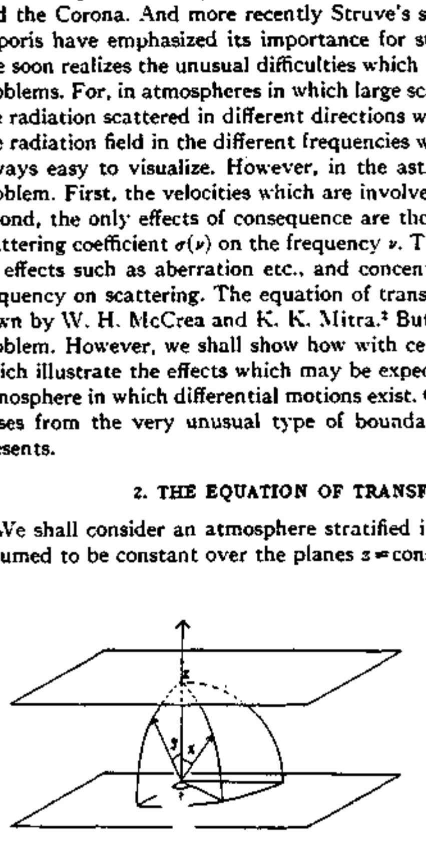


Fig. 1.

Let $\rho(z)$ be the density of the scattering material at height z and $w(z)$ the velocity of the material at the same height assumed parallel to the z direction. Further, let $\sigma(\nu)$ denote the mass scattering coefficient for the frequency ν as judged by an observer at rest with respect to the material. Since our principal interest is in the formation of absorption lines, we shall suppose that $\sigma(\nu)$ differs appreciably from zero only in a small range of ν. However, it is in the essence of the astrophysical problem that the "half-width" of $\sigma(\nu)$ is of the same order as the Doppler shifts in the frequency caused by the differential motions in the atmosphere. Indeed, it is this last

[1] O. Struve, Astrophys. J. 98, 98 (1943). Also W. Hiltner, Astrophys. J. 99, 103 (1944); P. W. Merrill and R. Sanford, Astrophys. J. 100, 14 (1944).
[2] W. H. McCrea and K. K. Mitra, Zeits. f. Astrophys. 11, 359 (1936).

circumstance which makes the change of frequency on scattering the only optical effect of the motion $w(z)$ which has any importance.

Consider then a pencil of radiation inclined at an angle ϑ to the positive normal and having a frequency ν as judged by a stationary observer. This radiation will appear to an observer at rest with respect to the material at z as having a frequency

$$\nu\left(1-\frac{w}{c}\cos\vartheta\right). \tag{1}$$

It will accordingly be scattered as such in all directions with a scattering coefficient

$$\sigma\left(\nu\left[1-\frac{w}{c}\cos\vartheta\right]\right). \tag{2}$$

We may, therefore, write the equation of transfer for the specific intensity $I(\nu, z, \vartheta)$ in the form

$$\cos\vartheta\frac{\partial I(\nu, z, \vartheta)}{\rho\partial z}=-\sigma\left(\nu\left[1-\frac{w}{c}\cos\vartheta\right]\right)I(\nu, z, \vartheta)+\mathfrak{J}(\nu, z, \vartheta), \tag{3}$$

where $\mathfrak{J}(\nu, z, \vartheta)$ denotes the emission per unit time and per unit solid angle in the frequency ν and in the direction ϑ. It is seen that this emission is given by

$$\mathfrak{J}(\nu, z, \vartheta)=\sigma\left(\nu\left[1-\frac{w}{c}\cos\vartheta\right]\right)\int_0^{2\pi}\int_0^{\pi} I\left(\nu\left[1-\frac{w}{c}\cos\vartheta+\frac{w}{c}\cos\chi\right], z, \chi\right)\sin\chi d\chi\frac{d\varphi}{4\pi}, \tag{4}$$

or in view of the symmetry about the z direction

$$\mathfrak{J}(\nu, z, \vartheta)=\tfrac{1}{2}\sigma\left(\nu\left[1-\frac{w}{c}\cos\vartheta\right]\right)\int_0^{\pi} I\left(\nu\left[1-\frac{w}{c}\cos\vartheta+\frac{w}{c}\cos\chi\right], z, \chi\right)\sin\chi d\chi. \tag{5}$$

To verify the foregoing expression for $\mathfrak{J}(\nu, z, \vartheta)$, we observe that the emission in the direction ϑ arises from the scattering into this direction of radiation from other directions. And, considering the contribution to $\mathfrak{J}$ from the scattering of the radiation in the direction specified by the polar angles χ and φ (see Fig. 1) into the direction $(\vartheta, 0)$, it is evident that the radiation must have the frequency

$$\nu\left(1-\frac{w}{c}\cos\vartheta+\frac{w}{c}\cos\chi\right), \tag{6}$$

as judged by a stationary observer; for, radiation of this frequency in the χ direction will appear to an observer at rest with respect to the material at z as having a frequency

$$\nu\left(1-\frac{w}{c}\cos\vartheta+\frac{w}{c}\cos\chi\right)\left(1-\frac{w}{c}\cos\chi\right)\simeq\nu\left(1-\frac{w}{c}\cos\vartheta\right), \tag{7}$$

which will accordingly be scattered uniformly in all directions with a scattering coefficient

$$\sigma\left(\nu\left[1-\frac{w}{c}\cos\vartheta\right]\right); \tag{8}$$

the radiation scattered into the ϑ-direction will have the same frequency (7) with respect to the material; to a stationary observer, it will appear as having a frequency ν. And summing over the contributions from all directions (χ, φ) we obtain (4).

Combining Eqs. (3) and (5) we have

$$\mu\frac{\partial I(\nu, z, \mu)}{\rho\partial z}=-\sigma\left(\nu\left[1-\frac{w}{c}\mu\right]\right)\left\{I(\nu, z, \mu)-\tfrac{1}{2}\int_{-1}^{+1} I\left(\nu\left[1-\frac{w}{c}\mu+\frac{w}{c}\mu'\right], z, \mu'\right)d\mu'\right\}, \tag{9}$$

where we have written μ and μ' for $\cos\vartheta$ and $\cos\chi$, respectively.

In solving Eq. (9) we shall adopt the method of approximation which has recently been developed in connection with the various problems of radiative transfer in the theory of stellar atmospheres.[3] The essence of this method is to replace the integrals which appear in the equation of transfer by sums according to Gauss's formula for numerical quadratures. Thus, considering Eq. (9) we replace it in the nth approximation by the system of $2n$ equations

$$\mu_i\frac{\partial I_i(\nu,z)}{\rho\partial z}=-\sigma\left(\nu\left[1-\frac{w}{c}\mu_i\right]\right)\left\{I_i(\nu,z)-\tfrac{1}{2}\sum a_jI_j\left(\nu\left[1-\frac{w}{c}\mu_i+\frac{w}{c}\mu_j\right],z\right)\right\},\quad (i=\pm1,\cdots,\pm n)\qquad(10)$$

where the μ_i's, $(i=\pm1,\cdots,\pm n)$, are the zeros of the Legendre polynomial $P_{2n}(\mu)$, and the a_i's are the appropriate weights. Further, in Eq. (10) we have written $I_i(\nu,z)$ for $I(\nu,z,\mu_i)$.

At this stage one further simplification of Eq. (9) is possible. In evaluating the Doppler shifts, we need not distinguish between

$$\nu\left(1-\frac{w}{c}\mu\right)\quad\text{and}\quad\nu-\nu_0\frac{w}{c}\mu,$$

where ν_0 denotes the frequency of the center of the line. We may, therefore, replace Eq. (10) by the simpler one

$$\mu_i\frac{\partial I_i(\nu,z)}{\rho\partial z}=-\sigma\left(\nu-\nu_0\frac{w}{c}\mu_i\right)\left\{I_i(\nu,z)-\tfrac{1}{2}\sum a_jI_j\left(\nu-\nu_0\frac{w}{c}\mu_i+\nu_0\frac{w}{c}\mu_j,z\right)\right\},\quad(i=\pm1,\cdots,\pm n).\qquad(11)$$

The form of Eq. (11) suggests that instead of considering the intensities I_i, $(i=\pm1,\cdots,\pm n)$, for some fixed frequency ν, we consider them for the frequencies

$$\nu_i=\nu+\nu_0\frac{w}{c}\mu_i\quad(i=\pm1,\cdots,\pm n),\qquad(12)$$

which are functions of z. In Eq. (12), ν is a "fixed" frequency. If we now let

$$I_i(\nu_i,z)=\psi_i(\nu,z)\quad(i=\pm1,\cdots,\pm n),\qquad(13)$$

we have

$$\frac{\partial\psi_i}{\partial z}=\left[\frac{\partial I_i(\nu,z)}{\partial z}\right]_{\nu=\nu_i}+\frac{\partial I_i(\nu_i,z)}{\partial\nu_i}\frac{\partial\nu_i}{\partial z},\qquad(14)$$

or, according to Eqs. (12) and (13)

$$\frac{\partial\psi_i}{\partial z}=\left[\frac{\partial I_i(\nu,z)}{\partial z}\right]_{\nu=\nu_i}+\mu_i\frac{\nu_0}{c}\frac{dw}{dz}\frac{\partial\psi_i}{\partial\nu}.\qquad(15)$$

Substituting for the first term on the right-hand side of the foregoing equation from Eq. (11), we obtain

$$\mu_i\frac{\partial\psi_i}{\partial z}-\mu_i^2\frac{\nu_0}{c}\frac{dw}{dz}\frac{\partial\psi_i}{\partial\nu}=-\rho\sigma(\nu)(\psi_i-\tfrac{1}{2}\sum a_j\psi_j)\quad(i=\pm1,\cdots,\pm n),\qquad(16)$$

which is clearly the most convenient form in which to study the equation of transfer for a moving atmosphere.

In our subsequent work we shall restrict ourselves to the first approximation. In this approximation

$$\mu_1=-\mu_{-1}=1/\sqrt{3}\quad\text{and}\quad a_1=a_{-1}=1,\qquad(17)$$

and Eqs. (11) and (16) lead to the two pairs of equations

[3] S. Chandrasekhar, Astrophys. J. 100, 76, 117 (1944) and 101, 95, 328, 348 (1945).

$$\mu_1 \frac{\partial I_{+1}(\nu, z)}{\rho \partial z} = -\tfrac{1}{2}\sigma\left(\nu - \nu_0 \frac{w}{c}\mu_1\right)\left[I_{+1}(\nu, z) - I_{-1}\left(\nu - 2\nu_0 \frac{w}{c}\mu_1, z\right)\right] \tag{18}$$

$$\mu_1 \frac{\partial I_{-1}(\nu, z)}{\rho \partial z} = +\tfrac{1}{2}\sigma\left(\nu + \nu_0 \frac{w}{c}\mu_1\right)\left[I_{-1}(\nu, z) - I_{+1}\left(\nu + 2\nu_0 \frac{w}{c}\mu_1, z\right)\right], \tag{19}$$

and

$$\mu_1 \frac{\partial \psi_{+1}}{\partial z} - \mu_1^2 \frac{\nu_0}{c} \frac{dw}{dz} \frac{\partial \psi_{+1}}{\partial \nu} = -\tfrac{1}{2}\rho\sigma(\nu)(\psi_{+1} - \psi_{-1}), \tag{20}$$

$$\mu_1 \frac{\partial \psi_{-1}}{\partial z} + \mu_1^2 \frac{\nu_0}{c} \frac{dw}{dz} \frac{\partial \psi_{-1}}{\partial \nu} = -\tfrac{1}{2}\rho\sigma(\nu)(\psi_{+1} - \psi_{-1}), \tag{21}$$

where it may be recalled that

$$\psi_{+1}(\nu, z) = I_{+1}\left(\nu + \mu_1 \nu_0 \frac{w}{c}, z\right), \quad \psi_{-1}(\nu, z) = I_{-1}\left(\nu - \mu_1 \nu_0 \frac{w}{c}, z\right). \tag{22}$$

3. SCHUSTER'S PROBLEM FOR A MOVING ATMOSPHERE

A classical problem first formulated by Schuster[4] provides the simplest model in terms of which the formation of absorption lines in a stellar atmosphere can be analyzed. In this model we consider a plane-stratified scattering atmosphere lying above a plane surface which radiates in a known manner and absorbs all radiation falling on it. The problem is to determine the radiation field in the atmosphere and in particular to relate the distribution in intensity of the emergent radiation with that radiated by the surface below. The appropriateness of this model for a first analysis of stellar absorption lines consists in the suitable idealization which it provides of the notions of a *photospheric surface* and the *reversing layers*. Consequently, when considering moving atmospheres it would seem proper that we retain the essentials of the Schuster model and generalize it only to the extent of admitting large scale motions. More particularly, we shall suppose that the photospheric surface is at $z=0$, and that it radiates uniformly in all outward directions $(0 \leqslant \vartheta < \pi/2)$ and in all frequencies. In other words, we suppose that

$$I(\nu, z, \vartheta) = \text{constant at } z=0 \text{ for } 0 \leqslant \vartheta < \pi/2 \text{ and for all frequencies.} \tag{23}$$

The state of motions in the atmosphere will be specified by the function $w(z)$ giving the velocity (assumed parallel to the z direction) at height z.

Finally, if $z=z_1$ defines the outer boundary of the atmosphere, we must require that here

$$I(\nu, z, \vartheta) \equiv 0, \quad \pi/2 < \vartheta < \pi \text{ at } z = z_1, \tag{24}$$

in accordance with the assumed non-existence of any radiation from the outside being incident on the atmosphere.

Schuster's problem for a moving atmosphere consists then in solving the equation of transfer (9), or the equivalent systems of equations in the various approximations, together with the boundary conditions (23) and (24). In the first approximation, the equivalent boundary conditions are that

$$I_{+1}(\nu, z) = \text{constant independent of } \nu \text{ at } z=0, \tag{25}$$

and

$$I_{-1}(\nu, z) \equiv 0 \text{ at } z = z_1. \tag{26}$$

4. THE REDUCTION TO A BOUNDARY VALUE PROBLEM FOR THE CASE $\sigma(\nu)$ = CONSTANT FOR $\nu_0 - \Delta\nu \leqslant \nu \leqslant \nu_0 + \Delta\nu$ AND ZERO OUTSIDE THIS INTERVAL AND FOR A LINEAR INCREASE OF w WITH THE OPTICAL DEPTH

In this paper we shall consider the solution to Schuster's problem formulated in the preceding

[4] A. Schuster, Astrophys. J. 21, 1 (1905).

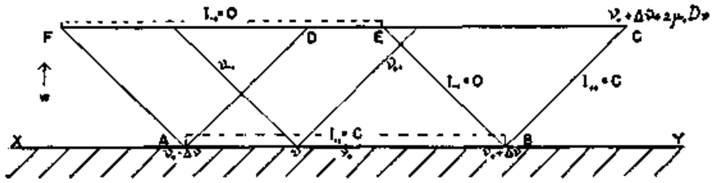


Fig. 2.

section for the case

$$\left.\begin{aligned}\sigma(\nu) &= \text{constant} = \sigma_0 \text{ for } \nu_0-\Delta\nu \leqslant \nu \leqslant \nu_0+\Delta\nu, \\ &= 0 \text{ otherwise,}\end{aligned}\right\} \tag{27}$$

and

$$\frac{1}{\rho}\frac{dw}{dz} = \text{constant}. \tag{28}$$

Further, we shall restrict ourselves to the first approximation.

When $\sigma(\nu)$ has the form (27) some care is required in the formulation of the boundary conditions. For, according to Eqs. (18) and (19)

$$\partial I_{+1}/\partial z \neq 0 \text{ only if } \nu_0-\Delta\nu+\mu_1\nu_0\frac{w}{c} \leqslant \nu \leqslant \nu_0+\Delta\nu+\mu_1\nu_0\frac{w}{c}, \tag{29}$$

and

$$\partial I_{-1}/\partial z \neq 0 \text{ only if } \nu_0-\Delta\nu-\mu_1\nu_0\frac{w}{c} \leqslant \nu \leqslant \nu_0+\Delta\nu-\mu_1\nu_0\frac{w}{c}. \tag{30}$$

Accordingly, in the (ν, w) plane the lines

$$\nu = \nu_0-\Delta\nu+\mu_1\nu_0\frac{w}{c} \quad \text{and} \quad \nu = \nu_0+\Delta\nu+\mu_1\nu_0\frac{w}{c}, \tag{31}$$

demark the regions in which I_{+1} is different from a constant from the regions in which it is a constant for varying z. The situation is further clarified in Fig. 2 where AD and BC represent the lines (31). Similarly, the lines (AF and BE in Fig. 2)

$$\nu = \nu_0-\Delta\nu-\mu_1\nu_0\frac{w}{c} \quad \text{and} \quad \nu = \nu_0+\Delta\nu-\mu_1\nu_0\frac{w}{c}, \tag{32}$$

demark the regions in which I_{-1} is different from a constant from the regions in which it is a constant for varying z.

Now, since the outward intensity I_{+1} is a constant independent of ν on the photospheric surface (represented by the line $XABY$ in Fig. 2), it is clear that, we must, in accordance with our foregoing remarks, require that

$$I_{+1}(\nu, z) = \text{constant along } AB \text{ and } BC. \tag{33}$$

Similarly, the non-existence of any radiation incident on the atmosphere from the outside requires that

$$I_{-1}(\nu, z) = 0 \text{ along } BE \text{ and } EF. \tag{34}$$

When we pass to the intensities ψ_{+1} and ψ_{-1} defined as in Eq. (22), the boundary conditions (33) and (34) are equivalent to (cf. Fig. 3):

$$\left.\begin{aligned}\psi_{+1} &= C = \text{constant on } AB\text{:} \quad z=0 \text{ and } \nu_0-\Delta\nu \leqslant \nu \leqslant \nu_0+\Delta\nu, \\ &= \text{the same constant on } BC\text{:} \quad \nu=\nu_0+\Delta\nu \text{ and } 0 \leqslant z \leqslant z_1,\end{aligned}\right\} \tag{35}$$

and

$$\left.\begin{aligned}\psi_{-1} &= 0 \text{ on } CD\text{:} \quad z=z_1 \text{ and } \nu_0-\Delta\nu \leqslant \nu \leqslant \nu_0+\Delta\nu, \\ &= 0 \text{ on } BC\text{:} \quad \nu=\nu_0+\Delta\nu \text{ and } 0 \leqslant z \leqslant z_1.\end{aligned}\right\} \tag{36}$$

We now transform Eqs. (20) and (21) to forms which are more convenient for their solution:

Let t denote the optical depth of the atmosphere measured from the boundary inward in terms of σ_0. Then

$$\rho\sigma_0 dz = -dt. \tag{37}$$

In transforming Eqs. (20) and (21) it is, however, more convenient to use instead of the optical depth t the variable

$$x = \frac{1}{2\mu_1}t = \frac{\sqrt{3}}{2}t. \tag{38}$$

In terms of x Eqs. (20) and (21) are

$$\frac{\partial\psi_{+1}}{\partial x} - \mu_1\frac{\nu_0}{c}\frac{dw}{dx}\frac{\partial\psi_{+1}}{\partial\nu} = \psi_{+1} - \psi_{-1}, \tag{39}$$

and

$$\frac{\partial\psi_{-1}}{\partial x} + \mu_1\frac{\nu_0}{c}\frac{dw}{dx}\frac{\partial\psi_{-1}}{\partial\nu} = \psi_{+1} - \psi_{-1}. \tag{40}$$

Now the assumption (28) concerning the variation of w clearly implies that the velocity is a linear function of x. And as it entails no loss of generality, we shall suppose that $w=0$ at the base of the atmosphere. Further, let

$$w = w_1 \text{ at } t=0 \text{ and } x=0. \tag{41}$$

Under these circumstances we can write

$$w = w_1(1 - x/x_1) = w_1(1 - t/t_1), \tag{42}$$

where t_1 denotes the optical thickness in σ_0 of the entire atmosphere lying above the radiating surface. According to Eqs. (41) and (42) w_1 denotes the difference in velocity between the top and the bottom of the atmosphere. This velocity can be expressed in terms of a *Doppler width* $D\nu$ according to

$$D\nu = \tfrac{1}{2}\nu_0 w_1/c. \tag{43}$$

With these definitions

$$\mu_1\frac{\nu_0}{c}\frac{dw}{dx} = -\mu_1\frac{\nu_0}{c}\frac{w_1}{x_1} = -\frac{2}{\sqrt{3}}\frac{D\nu}{x_1} = -\frac{4}{3}\frac{D\nu}{t_1}, \tag{44}$$

and Eqs. (39) and (40) can be rewritten as

$$\frac{\partial\psi_{+1}}{\partial x} + 2\mu_1\frac{D\nu}{x_1}\frac{\partial\psi_{+1}}{\partial\nu} = \psi_{+1} - \psi_{-1}, \tag{45}$$

and

$$\frac{\partial\psi_{-1}}{\partial x} - 2\mu_1\frac{D\nu}{x_1}\frac{\partial\psi_{-1}}{\partial\nu} = \psi_{+1} - \psi_{-1}. \tag{46}$$

We now introduce the variable y defined by

$$(\nu_0 + \Delta\nu) - \nu = 2\mu_1\frac{D\nu}{x_1}y. \tag{47}$$

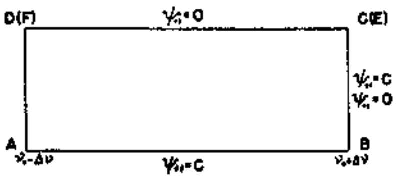


FIG. 3.

In other words, y measures the frequency shifts from the *violet edge* of $\sigma(\nu)$ in units of

$$2\mu_1 D\nu/x_1 \quad \text{(unit of frequency)}. \tag{48}$$

Equations (45) and (46) simplify to the forms

$$\partial\psi_{+1}/\partial x - \partial\psi_{+1}/\partial y = \psi_{+1} - \psi_{-1}, \tag{49}$$

and

$$\partial\psi_{-1}/\partial x + \partial\psi_{-1}/\partial y = \psi_{+1} - \psi_{-1}. \tag{50}$$

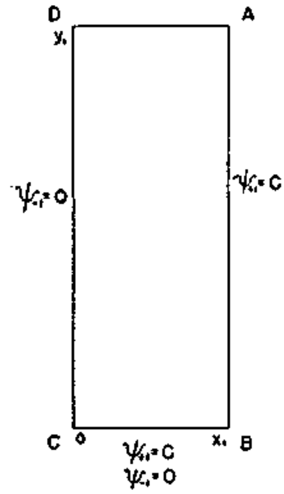


FIG. 4.

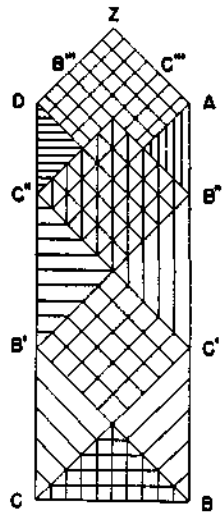


FIG. 5.

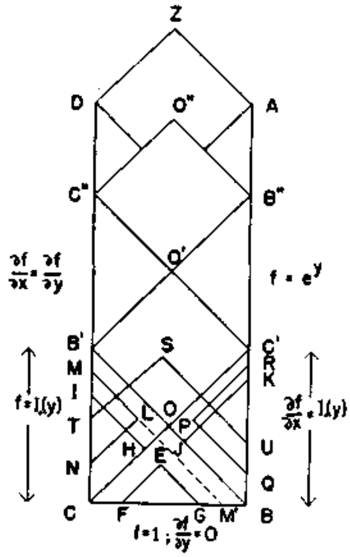


FIG. 6.

The range of the variables x and y in which the solution has to be sought is (Cf. Eq. [47])

$$0\leqslant x\leqslant x_1 \quad\text{and}\quad 0\leqslant y\leqslant y_1=\frac{1}{\mu_1}\frac{\Delta\nu}{D\nu}x_1. \tag{51}$$

And the boundary conditions with respect to which Eqs. (49) and (50) have to be solved in the ranges (51) are (see Fig. 4):

$$\left.\begin{aligned}\psi_{+1}&=C \text{ on } BA:\quad x=x_1 \text{ and } 0\leqslant y\leqslant y_1,\\ &=C \text{ on } CB:\quad y=0 \text{ and } 0\leqslant x\leqslant x_1.\end{aligned}\right\} \tag{52}$$

and

$$\left.\begin{aligned}\psi_{-1}&=0 \text{ on } CD:\quad x=0 \text{ and } 0\leqslant y\leqslant y_1,\\ &=0 \text{ on } CB:\quad y=0 \text{ and } 0\leqslant x\leqslant x_1.\end{aligned}\right\} \tag{53}$$

Since the Eqs. (49) and (50) are linear and homogeneous, there is no loss of generality if we set

$$C=1. \tag{54}$$

We shall assume this normalization in our further work. Finally, we may note that in terms of the variables x and y Eq. (12) allowing the passage from the ψ's to the I's becomes

$$y_{\pm1}=y\mp(x_1-x). \tag{55}$$

It is convenient to introduce one further transformation of the variables. Let

$$\psi_{+1}=e^{-y}f \quad\text{and}\quad \psi_{-1}=e^{-y}g. \tag{56}$$

Equations (49) and (50) reduce to

$$\partial f/\partial x-\partial f/\partial y=-g, \tag{57}$$

and

$$\partial g/\partial x+\partial g/\partial y=+f. \tag{58}$$

Eliminating g between Eqs. (57) and (58), we obtain

$$\partial^2 f/\partial x^2 - \partial^2 f/\partial y^2 + f = 0. \tag{59}$$

We require to solve this hyperbolic equation with the boundary conditions (cf. Eqs. (52) and (53) and Fig. 5)

$$\left.\begin{aligned} f = e^y \text{ on } AB:\quad & x = x_1 \text{ and } 0 \leqslant y \leqslant y_1, \\ f = 1 \text{ and } \partial f/\partial x = \partial f/\partial y = 0 \text{ on } BC:\quad & y = 0 \text{ and } 0 \leqslant x \leqslant x_1, \\ \partial f/\partial x = \partial f/\partial y \text{ on } CD:\quad & x = 0 \text{ and } 0 \leqslant y \leqslant y_1, \end{aligned}\right\} \tag{60}$$

since $\psi_{-1} = 0$ implies that $g = 0$ and according to Eq. (57) this, in turn, implies that $\partial f/\partial x = \partial f/\partial y$.

Since, in the problem of the formation of absorption lines, our principal interest is on the ratio of the emergent intensity $I_{+1}(\nu, t)$ at $t = 0$ to the constant outward intensity on the radiating surface, we are most interested in the value of f on CD and DA. We may recall in this connection that $\Delta\nu$, $D\nu$ and x_1 are to be regarded as the parameters of the problem. Given these, y_1 is determined according to the relation (cf. Eq. (51))

$$y_1 = \sqrt{3}x_1 \frac{\Delta\nu}{D\nu} = \frac{3}{2} \text{ optical depth } \frac{\text{line width}}{\text{Doppler width}}. \tag{61}$$

Further, according to Eqs. (48) and (61), y measures the frequency as it enters in ψ_{+1} and ψ_{-1}, from the violet edge of $\sigma(\nu)$ and in the units

$$\frac{2}{y_1}\Delta\nu = \frac{1}{y_1} \text{ line width}. \tag{62}$$

Accordingly, the line contour will have a width

$$\frac{2}{y_1}(x_1 + y_1)\Delta\nu = \left(1 + \frac{x_1}{y_1}\right) \text{ line width}. \tag{63}$$

This is in agreement with what can be inferred directly from Fig. 2. As can be seen from this figure, the contour (on our present first approximation) must extend from

$$\nu_0 - \Delta\nu \text{ to } \nu_0 + \Delta\nu + \mu_1 \nu_0 \frac{w_1}{c}, \tag{64}$$

and must, therefore, have the width

$$2\Delta\nu + \mu_1 \nu_0 \frac{w_1}{c}, \tag{65}$$

or, according to Eqs. (43) and (61):

$$2\Delta\nu + 2\mu_1 D\nu = 2\Delta\nu\left(1 + \mu_1 \frac{D\nu}{\Delta\nu}\right) = 2\Delta\nu\left(1 + \frac{x_1}{y_1}\right). \tag{66}$$

It is evident that the line contour will itself be given by

$$r = e^{-y} f, \quad x = 0, \quad 0 \leqslant y \leqslant y_1, \tag{67}$$

and

$$r = e^{-y_1} f, \quad y = y_1, \quad 0 \leqslant x \leqslant x_1. \tag{68}$$

Equation (67) refers to the part of the contour which extends from

$$\nu = \nu_0 - \Delta\nu + \mu_1 \nu_0 \frac{w_1}{c} \text{ to } \nu_0 + \Delta\nu + \mu_1 \nu_0 \frac{w_1}{c} \tag{69}$$

while Eq. (68) refers to the part

$$\nu_0 - \Delta\nu + \mu_1 \nu_0 \frac{w_1}{c} \geqslant \nu \geqslant \nu_0 - \Delta\nu. \tag{70}$$

Finally, it may be noted that according to Eq. (55) the scale of frequency is the same for both the x and the y axis.

5. THE SOLUTION OF THE BOUNDARY VALUE PROBLEM

In the preceding section we have seen how the determination of the radiation field in a scattering atmosphere in which differential motions are present can be reduced to a boundary value problem in partial differential equations of the hyperbolic type. Under the conditions (27) and (28), the hyperbolic equation is one with constant coefficients and is of the simplest kind; indeed it is of the same form as the well-known equation of telegraphy.[5] But where our problem differs from the standard ones is in the boundary conditions. And it is the nature of our boundary conditions which prevents a direct application of the methods of Cauchy or Riemann.[6] For in these latter methods, only those situations in which the function and its derivatives are assigned along curves which do not intersect any characteristic more than once are contemplated. Our boundary conditions (60) are not as simple, the "supporting curve" $DCBA$ in fact intersecting every characteristic through a point inside the fundamental rectangle twice. Moreover, the function and its derivatives are assigned only on a part of the contour namely CB, while on the rest of the contour either the function alone, or a relation between its derivatives is specified. We shall, however, show how the boundary conditions (60) just suffice to determine f uniquely in the region $ZDCBA$. The method of solution we are going to describe is an adaptation of Riemann's method and is based on Green's theorem.

Now Green's theorem as applied to Eq. (59) is that the integral

$$\oint Pdy - Qdx \tag{71}$$

where

$$P = v\frac{\partial f}{\partial x} - f\frac{\partial v}{\partial x} \quad \text{and} \quad Q = f\frac{\partial v}{\partial y} - v\frac{\partial f}{\partial y}, \tag{72}$$

around a closed contour vanishes if f and v are any two functions which satisfy Eq. (59) on and inside the contour.

As in Riemann's method we shall apply Green's theorem to contours which in parts are the characteristics $x-\xi = \pm(y-\eta)$ passing through some selected point (ξ, η) and choose for v a solution which is constant along the characteristics through (ξ, η). For Eq. (59) such a "Riemann function" $v(x, y; \xi, \eta)$ is known and depending on the quadrant in which the contour lies is

$$v(x, y; \xi, \eta) = I_0([(y-\eta)^2 - (x-\xi)^2]^{\frac{1}{2}}), \tag{73}$$

or

$$v(x, y; \xi, \eta) = J_0([(x-\xi)^2 - (y-\eta)^2]^{\frac{1}{2}}), \tag{74}$$

where J_0 and I_0 are the Bessel functions of order zero for real and imaginary arguments, respectively.

With the choice of the Riemann function for v, it is readily verified that

$$\int_{x-\xi=\pm(y-\eta)} Pdy - Qdx = \pm f \tag{75}$$

if the integral on the right-hand side is a line integral along the characteristic $x-\xi = \pm(y-\eta)$.

In solving Eq. (59) consistent with the boundary conditions (60), we shall find it necessary to treat the various regions distinguished in Fig. 5 separately.

---

[5] Cf. A. G. Webster, *Partial differential equations of mathematical physics* (1933), Section 46, p. 173.

[6] For a general exposition of these classical methods see Webster, reference 5, pp. 160–188 and 239–255; or P. Frank and R. von Mises, *Die Differential und Integralgleichungen der Mechanik und Physik* (Rosenberg, New York, 1943), Vol. I, pp. 779–817.

(a) The Solution in the Region OCB

Let the characteristic $x=y$ through C intersect AB at C' and the characteristic $x_1-x=y$ through B intersect CD at B'. Further, let CC' and BB' intersect at O. (See Fig. 6.)

Now, since the function and its derivatives are specified along CB, the solution inside the region OCB (including the sides OC and OB) can be found directly by Riemann's method. Thus applying Green's theorem to a contour such as $EFGE$ where EF and EG are the characteristics through $E=(\xi, \eta)$, and using Eq. (75) to evaluate the integrals along the characteristics we readily find that

$$f(\xi, \eta)=1-\tfrac{1}{2}\int_{\xi-\eta}^{\xi+\eta}\left(f\frac{\partial v}{\partial y}-v\frac{\partial f}{\partial y}\right)_{y=0}dx, \tag{76}$$

or, remembering that along CB

$$f=1 \quad \text{and} \quad \frac{\partial f}{\partial x}=\frac{\partial f}{\partial y}=0, \tag{77}$$

we have

$$f(\xi, \eta)=1-\tfrac{1}{2}\int_{\xi-\eta}^{\xi+\eta}\left(\frac{\partial v}{\partial y}\right)_{y=0}dx. \tag{78}$$

Now the Riemann function appropriate to our present contour is (73). Accordingly,

$$\begin{aligned}\left(\frac{\partial v}{\partial y}\right)_{y=0}&=\left[I_1([(y-\eta)^2-(x-\xi)^2]^{1/2})\frac{y-\eta}{[(y-\eta)^2-(x-\xi)^2]^{1/2}}\right]_{y=0}\\&=-I_1([\eta^2-(x-\xi)^2]^{1/2})\frac{\eta}{[\eta^2-(x-\xi)^2]^{1/2}}.\end{aligned} \tag{79}$$

Hence,

$$f(\xi, \eta)=1+\tfrac{1}{2}\eta\int_{\xi-\eta}^{\xi+\eta}I_1([\eta^2-(x-\xi)^2]^{1/2})\,\frac{dx}{[\eta^2-(x-\xi)^2]^{1/2}}. \tag{80}$$

Equation (80) determines f in the region OCB.

To evaluate the integral on the right-hand side of Eq. (80) we let

$$x-\xi=\eta\cos\vartheta, \tag{81}$$

and obtain

$$f(\xi, \eta)=1+\tfrac{1}{2}\eta\int_0^\pi I_1(\eta\sin\vartheta)d\vartheta. \tag{82}$$

Replacing I_1 in the foregoing equation by its equivalent series expansion and integrating term by term we find

$$\begin{aligned}f(\xi, \eta)&=1+\tfrac{1}{2}\eta\sum_{m=0}^{\infty}\int_0^\pi\frac{(\frac{1}{2}\eta\sin\vartheta)^{2m+1}}{m!\Gamma(m+2)}d\vartheta,\\&=1+\eta\sum_{m=0}^{\infty}(\tfrac{1}{2}\eta)^{2m+1}\frac{1}{m!\Gamma(m+2)}\int_0^{\pi/2}\sin^{2m+1}\vartheta d\vartheta,\\&=1+\eta\sum_{m=0}^{\infty}(\tfrac{1}{2}\eta)^{2m+1}\frac{1}{m!\Gamma(m+2)}\frac{2^{2m}(m!)^2}{(2m+1)!}\\&=1+\sum_{m=0}^{\infty}\frac{\eta^{2m+2}}{(2m+2)!}\\&=\sum_{m=0}^{\infty}\frac{\eta^{2m}}{(2m)!}.\end{aligned} \tag{83}$$

Thus, $$f(\xi, \eta) = \cosh \eta \tag{84}$$

inside and on the triangular contour OCB.

(b) The Integral Equation which Ensures the Continuity of the Solution Along OC

We have seen how the boundary conditions along CB determine the solution in the region OCB and on the sides OC and OB. We shall now show how this knowledge of the function along OC and OB together with the boundary conditions on CB' and BC' enables us to continue the solution into the region $O'B'COBC'O'$ (including the sides $B'O'$ and $O'C'$).

Thus, applying Green's theorem to contours such as $ICHI$ and $KJBK$ we shall obtain integral equations relating the values which the function takes along CO and OB with the values which the function and its derivatives take along CB' and BC'. And, as we shall see presently, these integral equations suffice to determine f along CB' and $\partial f/\partial x$ along BC' uniquely and secure at the same time the continuity of the solutions along OC and OB.

Considering first the condition which ensures the continuity of the solution along CO apply Green's theorem to a contour such as $ICHI$ where $H = (\eta, \eta)$ is a point on CO and HI is the characteristic $\eta - x = y - \eta$ through H. Using Eq. (75) to evaluate the integrals along the characteristics HI and CH and remembering that f takes the values 1 and $\cosh \eta$ at C and H, respectively, we find that

$$2 \cosh \eta - 1 = f(2\eta, 0) + \int_0^{2\eta} \left(v\frac{\partial f}{\partial x} - f\frac{\partial v}{\partial x} \right)_{x=0} dy, \tag{85}$$

or, since $\partial f/\partial x = \partial f/\partial y$ along CB', we have

$$2 \cosh \eta - 1 = f(0, 2\eta) + \int_0^{2\eta} \left(v\frac{\partial f}{\partial y} \right)_{x=0} dy - \int_0^{2\eta} \left(f\frac{\partial v}{\partial x} \right)_{x=0} dy. \tag{86}$$

Integrating by parts the first of the two integrals on the right-hand side of Eq. (86) we obtain

$$2 \cosh \eta = 2f(0, 2\eta) - \int_0^{2\eta} f(0, y) \left(\frac{\partial v}{\partial x} + \frac{\partial v}{\partial y} \right)_{x=0} dy. \tag{87}$$

The Riemann function appropriate to our present contour is (74) with $\xi = \eta$. With this choice of v we find after some minor reductions that

$$\cosh \eta = f(0, 2\eta) - \tfrac{1}{2} \int_0^{2\eta} f(0, y) J_1([\eta^2 - (y-\eta)^2]^{\frac{1}{2}}) \frac{y\,dy}{[\eta^2 - (y-\eta)^2]^{\frac{1}{2}}}, \tag{88}$$

which is seen to be an integral equation for f along CB'. It is seen that Eq. (88) is equivalent to an integral equation of Voltèrra's type. For, by differentiating the equation

$$\sinh \eta = \tfrac{1}{2} \int_0^{2\eta} f(0, y) J_0([\eta^2 - (y-\eta)^2]^{\frac{1}{2}}) dy \tag{89}$$

with respect to η we may readily verify that we recover Eq. (88).

(c) The Solution of the Integral Eq. (89)

To solve Eq. (89) we apply a Laplace transformation to this equation. Thus multiplying both sides of Eq. (89) by $e^{-s\eta}$ and integrating over η from 0 to ∞ we obtain

$$\frac{2}{s^2 - 1} = \int_0^{\infty} d\eta\, e^{-s\eta} \int_0^{2\eta} dy\, f(0, y) J_0([\eta^2 - (y-\eta)^2]^{\frac{1}{2}}), \tag{90}$$

or inverting the order of the integration on the right-hand side we have

$$\frac{2}{s^2-1}=\int_0^\infty dy f(0,y)\int_{y/2}^\infty d\eta e^{-s\eta}J_0([2\eta y-y^2]^{\frac{1}{2}}). \tag{91}$$

Introducing the variable

$$t=(2\eta y-y^2)^{\frac{1}{2}} \tag{92}$$

instead of η we find

$$\frac{2}{s^2-1}=\int_0^\infty \frac{dy}{y}e^{-sy/2}f(0,y)\int_0^\infty dt\, t\exp[-st^2/2y]J_0(t). \tag{93}$$

In Eq. (93) the integral over t is equivalent to the so-called Weber's first exponential integral in the theory of Bessel functions[7] and its value is given by

$$\int_0^\infty \exp[-st^2/2y]J_0(t)t\,dt=\frac{y}{s}e^{-y/2s}. \tag{94}$$

Using this result Eq. (93) reduces to

$$\frac{2s}{s^2-1}=\int_0^\infty f(0,y)\exp[-y(s+s^{-1})/2]dy. \tag{95}$$

If we now let

$$s+s^{-1}=2u, \tag{96}$$

Eq. (95) becomes

$$\frac{1}{(u^2-1)^{\frac{1}{2}}}=\int_0^\infty f(0,y)e^{-yu}dy. \tag{97}$$

In other words, we have shown that the Laplace transform of $f(0,y)$ is $(u^2-1)^{\frac{1}{2}}$. But it is known that the Laplace transform of $I_0(y)$ is exactly this. Hence

$$f(0,y)=I_0(y)\quad (0<y\leqslant x_1). \tag{98}$$

Thus, the requirement of the continuity of the solution along CO has determined f along CB'. Its derivatives along CB' are also deducible. We have

$$\left(\frac{\partial f}{dx}\right)_{x=0}=\left(\frac{\partial f}{\partial y}\right)_{x=0}=I_1(y)\quad (0<y\leqslant x_1). \tag{99}$$

(d) The Integral Equation Ensuring the Continuity of the Solution Along OB and Its Solution

Along BA we know f and its derivative with respect to y. But we do not know $\partial f/\partial x$ along this line. However, as the solution along OB is known the requirement that the solution be continuous on this line will determine $\partial f/\partial x$ along BC'. Thus, applying Green's theorem to a contour such as $JKBJ$ where $J=(x_1-\eta,\eta)$ is a point on OB and JK the characteristic $x-x_1+\eta=y-\eta$ through J we find in the usual manner that

$$2\cosh\eta=1+e^{2\eta}-\int_0^{2\eta}\left(v\frac{\partial f}{\partial x}-e^y\frac{\partial v}{\partial x}\right)_{x=x_1}dy. \tag{100}$$

The Riemann function appropriate to our present contour is

$$v=J_0([(x-x_1+\eta)^2-(y-\eta)^2]^{\frac{1}{2}}). \tag{101}$$

With v given by Eq. (101), Eq. (100) becomes

$$2\cosh\eta=1+e^{2\eta}-\int_0^{2\eta}e^yJ_1([\eta^2-(y-\eta)^2]^{\frac{1}{2}})\frac{\eta dy}{[\eta^2-(y-\eta)^2]^{\frac{1}{2}}}-\int_0^{2\eta}J_0([\eta^2-(y-\eta)^2]^{\frac{1}{2}})\left(\frac{\partial f}{\partial x}\right)_{x=x_1}dy. \tag{102}$$

[7] Cf. G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, England, 1944), p. 393.

Putting
$$y-\eta=\eta\cos\vartheta \tag{103}$$
in the first of the two integrals on the right-hand side of Eq. (102), it can be expressed in the form
$$2\cosh\eta=1+e^{2\eta}-e^{\eta}G(\eta)-\int_0^{2\eta}J_0([\eta^2-(y-\eta)^2]^{\frac{1}{2}})\left(\frac{\partial f}{\partial x}\right)_{x=x_1}dy, \tag{104}$$
where
$$G(\eta)=\eta\int_0^{\pi}e^{\eta\cos\vartheta}J_1(\eta\sin\vartheta)d\vartheta. \tag{105}$$

To evaluate $G(\eta)$ we replace $e^{\eta\cos\vartheta}$ and $J_1(\eta\sin\vartheta)$ by their respective series expansions and integrate term by term. In this manner we find that
$$\begin{aligned}
G(\eta)&=\eta\int_0^{\pi}\sum_{n=0}^{\infty}\frac{(\eta\cos\vartheta)^n}{n!}\sum_{m=0}^{\infty}(-1)^m\frac{(\frac{1}{2}\eta\sin\vartheta)^{2m+1}}{m!\Gamma(m+2)}d\vartheta,\\
&=\sum_{n=0}^{\infty}\frac{\eta^{2n+1}}{(2n)!}\Gamma(n+\tfrac{1}{2})\sum_{m=0}^{\infty}(-1)^m\frac{(\frac{1}{2}\eta)^{2m+1}}{\Gamma(m+2)\Gamma(m+n+\frac{3}{2})},\\
&=-\sum_{n=0}^{\infty}\frac{(2\eta)^{n+\frac{1}{2}}\Gamma(n+\frac{1}{2})}{(2n)!}\sum_{m=1}^{\infty}(-1)^m\frac{(\frac{1}{2}\eta)^{2m+n-\frac{1}{2}}}{\Gamma(m+1)\Gamma(m+n+\frac{1}{2})},\\
&=\sum_{n=0}^{\infty}\frac{(2\eta)^{n+\frac{1}{2}}\Gamma(n+\frac{1}{2})}{(2n)!}\left[\frac{(\frac{1}{2}\eta)^{n-\frac{1}{2}}}{\Gamma(n+\frac{1}{2})}-J_{n-\frac{1}{2}}(\eta)\right],\\
&=2\sum_{n=0}^{\infty}\frac{\eta^{2n}}{(2n)!}-\sum_{n=0}^{\infty}\frac{(2\eta)^{n+\frac{1}{2}}}{(2n)!}\Gamma(n+\tfrac{1}{2})J_{n-\frac{1}{2}}(\eta),\\
&=2\cosh\eta-(2\pi\eta)^{\frac{1}{2}}\sum_{n=0}^{\infty}\left(\frac{\eta}{2}\right)^n\frac{1}{n!}J_{n-\frac{1}{2}}(\eta).
\end{aligned} \tag{106}$$
But according to a formula of Lommel[8]
$$\sum_{n=0}^{\infty}(\tfrac{1}{2}\eta)^n\frac{1}{n!}J_{n-\frac{1}{2}}(\eta)=\left(\frac{2}{\pi\eta}\right)^{\frac{1}{2}}. \tag{107}$$
Hence
$$G(\eta)=2(\cosh\eta-1). \tag{108}$$
Substituting this result in Eq. (104) we find
$$\sinh\eta=\tfrac{1}{2}\int_0^{2\eta}J_0([\eta^2-(y-\eta)^2]^{\frac{1}{2}})\left(\frac{\partial f}{\partial x}\right)_{x=x_1}dy, \tag{109}$$
which is a Volterra integral equation for $(\partial f/\partial x)_{x=x_1}$.

It is seen that Eq. (109) is of the same form as Eq. (89). Accordingly
$$(\partial f/\partial x)_{x=x_1}=I_0(y)\quad(0\leqslant y\leqslant x_1). \tag{110}$$

(e) The Solution in the Region $O'B'COBC'O'$

With the determination of f along CB' and of $\partial f/\partial x$ along BC' our knowledge of the function and its derivatives along $B'CBC'$ is complete, and in the region $O'B'COBC'O'$ the solution becomes determinate. Thus, as in Riemann's method, applying Green's theorem to contours such as $LMNL$, $PQRP$, and $STCBUS$ we find that we can express f in the regions $OB'C$, OBC', and $O'B'OC'$ as follows:[9]

[8] See G. N. Watson, *Theory of Bessel Functions* (Cambridge University Press, England, 1944), p. 141, Eq. (7).
[9] For a point such as L, Green's theorem can be applied to either of the two contours $LMNL$ and $LNCM'L$. But clearly no ambiguity is implied as the continuity of the solution along OC (and OB) has already been ensured.

$$f(\xi,\eta)=I_0(\xi+\eta)-\tfrac{1}{2}\xi\int_0^{\pi} I_0(\eta+\xi\cos\vartheta)J_1(\xi\sin\vartheta)(1+\cos\vartheta)d\vartheta, \quad [(\xi,\eta) \text{ in } OB'C]. \tag{111}$$

$$f(\xi,\eta)=e^{\eta}-\tfrac{1}{2}(x_1-\xi)\int_0^{\pi} I_0(\eta+[x_1-\xi]\cos\vartheta)J_0([x_1-\xi]\sin\vartheta)\sin\vartheta d\vartheta, \quad [(\xi,\eta) \text{ in } OBC']; \tag{112}$$

and

$$\begin{aligned} f(\xi,\eta)=&I_0([\eta^2-\xi^2]^{\frac{1}{2}})+e^{\eta-x_1+\xi}+\eta\int_{\cos^{-1}(x_1-\xi)/\eta}^{\cos^{-1}\xi/\eta} I_1(\eta\sin\vartheta)d\vartheta \\ &-\xi\int_0^{\cosh^{-1}\eta/\xi} I_0(\eta-\xi\cosh\vartheta)I_1(\xi\sinh\vartheta)(\cosh\vartheta-1)d\vartheta \\ &+(x_1-\xi)\int_0^{\cosh^{-1}\eta/(x_1-\xi)} I_0([x_1-\xi]\sinh\vartheta)I_0(\eta-[x_1-\xi]\cosh\vartheta)\sinh\vartheta d\vartheta \\ &+(x_1-\xi)\int_0^{\cosh^{-1}\eta/(x_1-\xi)} e^{\eta-(x_1-\xi)\cosh\vartheta} I_1([x_1-\xi]\sinh\vartheta)d\vartheta, \quad [(\xi,\eta) \text{ in } O'B'OC']. \end{aligned} \tag{113}$$

In particular we may note that Eq. (113) will enable us to determine the solution along the sides $B'O'$ and $O'C'$.

(f) Further Continuation of the Solution

In the preceding paragraphs we have seen how the knowledge of the function along COB, together with the boundary conditions on CB' and BC' enables us to determine f in the region $O'B'COBC'$ including the sides $B'O'$ and $O'C'$. It is now apparent that in the same way we can utilize our present knowledge of the function along $B'O'C'$ to extend the solution further into the region $O'B'C''O''B''C'$. And this process can be continued until the solution inside the entire rectangle $DCBA$ is completed. However, in this paper we shall not consider these further extensions but content ourselves with the solution which has been completed in the first square $B'CBC'$. According to Eq. (61) this will suffice to determine the radiation field in all cases in which the ratio $D\nu:\Delta\nu$ exceeds $\sqrt{3}$.

6. THE CONTOURS OF THE ABSORPTION LINES FORMED. NUMERICAL ILLUSTRATIONS

In the preceding section we have seen how the boundary value problem formulated in Section 4 can, in principle, be solved. In terms of the solution thus found, we can specify the radiation field in an atmosphere with differential motions and under the conditions prescribed in Sections 3 and 4. While the determination of the radiation field in the entire atmosphere is necessary to answer all questions relating to the formation of the absorption lines (see Section 7 below), greatest interest is, however, attached to the contour of the resulting line. In the first approximation in which we have studied the problem, this is given in terms of the emergent value of the outward intensity $I_{+1}(\nu)$. More specifically the form of the line is given by Eqs. (67) and (68) where f is the solution of the boundary value problem. We shall now consider in some detail the predicted nature of these contours.

Now along CB' the solution is given by (cf. Eq. (98))

$$f=I_0(y) \quad (x=0,\ 0\leqslant y\leqslant x_1). \tag{114}$$

According to Eq. (68) we may, therefore, write down a formula for the *residual intensity* r which will be valid for a part of the line contour. Thus

$$r=e^{-y}I_0(y), \tag{115}$$

will describe the line in the frequency interval

$$\nu_0+\Delta\nu+2\mu_1 D\nu \geqslant \nu \geqslant \nu_0-\Delta\nu+2\mu_1 D\nu, \tag{116}$$

or

$$\nu_0+\Delta\nu+2\mu_1 D\nu \geqslant \nu \geqslant \nu_0+\Delta\nu, \tag{116'}$$

TABLE I. The function $r = e^{-y}I_0(y)$.

y	$e^{-y}I_0(y)$	y	$e^{-y}I_0(y)$	y	$e^{-y}I_0(y)$	y	$e^{-y}I_0(y)$
0	1.0000	0.7	0.5593	1.40	0.3831	3.0	0.2430
0.1	0.9071	0.8	0.5241	1.50	0.3674	3.5	0.2228
0.2	0.8269	0.9	0.4932	1.75	0.3346	4.0	0.2070
0.3	0.7576	1.0	0.4658	2.00	0.3085	4.5	0.1942
0.4	0.6974	1.1	0.4414	2.25	0.2874	5.0	0.1835
0.5	0.6450	1.2	0.4198	2.50	0.2700	6.0	0.1667
0.6	0.5993	1.3	0.4004	2.75	0.2555	8.0	0.1434
						10.0	0.1278

TABLE II. $f(\xi, \eta)$.

	η					
ξ	1.0	2.0	2.5	3.0	4.0	5.0
0	1.2661	2.2796	3.2898	4.8808	11.302	27.240
0.5	1.4762	2.9697	4.4134	6.6792	15.847	39.601
1.0	1.5431	3.4374	5.2449	8.0836	19.628	51.362
1.5	1.5431	3.6897	5.7814	9.0767	23.354	62.659
2.0	1.5431	3.7622	6.0569	9.6891	26.931	73.746
2.5	1.5431	3.7622	6.1323	10.6881	30.579	84.930
3.0	1.5431	3.7622	6.7927	11.8944	34.506	96.502
3.5	1.5431	4.3857	7.7767	13.4600	38.859	108.657
4.0	1.5431	5.2378	9.0554	15.4183	43.693	121.460
4.5	2.0969	6.2657	10.5584	17.6724	48.998	134.802
5.0	2.7183	7.3891	12.1825	20.0855	54.598	148.413

TABLE III. Line contours of absorption lines formed in a moving atmosphere ($x_1 = 5$; $y_1 = 1, 2, 2.5, 3, 4, 5$).

$\frac{\nu-\nu_0+\Delta\nu}{2\Delta\nu}$	r	$\frac{\nu-\nu_0+\Delta\nu}{2\Delta\nu}$	r	$\frac{\nu-\nu_0+\Delta\nu}{2\Delta\nu}$	r	$\frac{\nu-\nu_0+\Delta\nu}{2\Delta\nu}$	r	$\frac{\nu-\nu_0+\Delta\nu}{2\Delta\nu}$	r	$\frac{\nu-\nu_0+\Delta\nu}{2\Delta\nu}$	r
0.0	1.0000	0.0	1.0000	0.0	1.0000	0.0	1.0000	0.0	1.0000	0.0	1.0000
0.5	0.7714	0.25	0.8480	0.2	0.8667	$0.1\dot{6}$	0.8799	0.125	0.8974	0.1	0.9083
1.0	0.5677	0.50	0.7089	0.4	0.7433	$0.3\dot{3}$	0.7676	0.250	0.8003	0.2	0.8184
1.5	0.5677	0.75	0.5935	0.6	0.6384	0.50	0.6701	0.375	0.7117	0.3	0.7321
2.0	0.5677	1.00	0.5092	0.8	0.5576	$0.6\dot{6}$	0.5922	0.500	0.6320	0.4	0.6502
2.5	0.5677	1.25	0.5092	1.0	0.5034	$0.8\dot{3}$	0.5321	0.625	0.5601	0.5	0.5723
3.0	0.5677	1.50	0.5092	1.2	0.4972	1.00	0.4824	0.750	0.4933	0.6	0.4969
3.5	0.5677	1.75	0.4993	1.4	0.4746	$1.1\dot{6}$	0.4519	0.875	0.4277	0.7	0.4222
4.0	0.5677	2.00	0.4652	1.6	0.4305	$1.3\dot{3}$	0.4025	1.000	0.3595	0.8	0.3461
4.5	0.5431	2.25	0.4019	1.8	0.3623	1.50	0.3325	1.125	0.2902	0.9	0.2668
5.0	0.4658	2.50	0.3085	2.0	0.2700	$1.6\dot{6}$	0.2430	1.250	0.2070	1.0	0.1835
5.5	0.6450	2.75	0.3674	2.2	0.3085	$1.8\dot{3}$	0.2700	1.375	0.2228	1.1	0.1942
6.0	1.0000	3.00	0.4658	2.4	0.3674	2.00	0.3085	1.500	0.2430	1.2	0.2070
		3.25	0.6450	2.6	0.4658	$2.1\dot{6}$	0.3674	1.625	0.2700	1.3	0.2228
		3.50	1.0000	2.8	0.6450	$2.3\dot{3}$	0.4658	1.750	0.3085	1.4	0.2430
				3.0	1.0000	2.50	0.6450	1.875	0.3674	1.5	0.2700
						$2.6\dot{6}$	1.0000	2.000	0.4658	1.6	0.3085
								2.125	0.6450	1.7	0.3674
								2.250	1.0000	1.8	0.4658
										1.9	0.6450
										2.0	1.0000

depending on whether

$$\mu_1 D\nu \gtrless \Delta\nu \quad \text{or} \quad \mu D\nu \lessgtr \Delta\nu. \tag{117}$$

It should be noted in this connection that in our present context y measures the frequency shifts from the violet edge $\nu_0+\Delta\nu+2\mu_1 D\nu$ of the *line contour* in the unit $2\Delta\nu/y_1$.

For convenience we have provided a brief table of the function on the right-hand side of Eq. (115) (see Table I).

Again, according to Eqs. (68) and (84), when

$$D\nu > (2/\mu_1)\Delta\nu = 2\sqrt{3}\Delta\nu, \tag{118}$$

in the frequency interval

$$\nu_0 - 3\Delta\nu + 2\mu_1 D\nu \geqslant \nu \geqslant \nu_0 + \Delta\nu, \tag{119}$$

the contour is flat, the residual intensity having the constant value

$$r = e^{-y_1}\cosh y_1 \quad (y_1 \leqslant x \leqslant x_1 - y_1). \tag{120}$$

This flat portion occupies a fraction

$$(x_1 - 2y_1)/(x_1 + y_1) \tag{121}$$

of the entire contour. As $D\nu/\Delta\nu \to \infty$ and $y_1 \to 0$, the fraction (121) tends to unity: the line accordingly becomes very shallow and very broad. More specifically, as $y_1 \to 0$

$$1 - r \to y_1 \quad (y_1 \to 0), \tag{122}$$

and

$$\text{the width of the line contour} \to 2\Delta\nu x_1/y_1 \quad (y_1 \to 0). \tag{123}$$

The equivalent width, therefore, tends to the limiting value

$$\text{Equivalent width} \to 2\Delta\nu x_1 = \sqrt{3}\Delta\nu t_1 = \sqrt{3}\Delta\nu N\sigma_0 m^{-1} \quad (y_1 \to 0), \tag{124}$$

where N denotes the number of scattering atoms in a column of unit cross section in the atmosphere and m the mass of the atom.

Returning to the general case, it is seen that the specification of the line contour over its entire range

$$\nu_0 - \Delta\nu \leqslant \nu \leqslant \nu_0 + \Delta\nu + 2\mu_1 D\nu, \tag{125}$$

requires a knowledge of the function f along the line $y = y_1$ and for $0 \leqslant x \leqslant x_1$. However, since the solution for f has been found in an explicit form only in the first square, $B'CBC'$, complete contours can be given only for those cases in which $D\nu \geqslant \sqrt{3}\Delta\nu$. And even then, the part of the solution not included in the triangle OBC and the sides CB' and BC' can be found only after several numerical quadratures. For, in these regions the solution is given by the formulae (111)–(113), and it does not seem that the various integrals occurring in these formulae can be evaluated explicitly.

As illustrating the solution found in Section 5 we have considered in detail the case

$$x_1 = 5 \tag{126}$$

and determined the line contours for the following ratios of the Doppler width to the line width:

$$\mu_1 D\nu/\Delta\nu = 5,\ 2.5,\ 2,\ 1\tfrac{2}{3},\ 1.25, \text{ and } 1. \tag{127}$$

According to Eq. (61) the specification of the contours for these ratios of $\mu_1 D\nu : \Delta\nu$, requires the evaluation of f along the lines

$$y = 1,\ 2,\ 2.5,\ 3,\ 4, \text{ and } 5, \tag{128}$$

for $0 \leqslant x \leqslant 5$. The values of f for several points along these lines and intercepted in the region $O'B'COBC'$ were determined according to Eqs. (111)–(113). The various integrals occurring in these equations were evaluated numerically.[16] The results of these calculations are included in Table II. In Table III, the values of f given in Table II are converted into residual intensities according to

[16] The carrying out of the numerical quadratures were immensely facilitated by the *British Association Mathematical Tables*, VI: *Bessel functions of order zero and unity* (Cambridge University Press, England, 1937). I should record here my indebtedness to Mrs. Frances Herman Breen for assistance with these calculations.

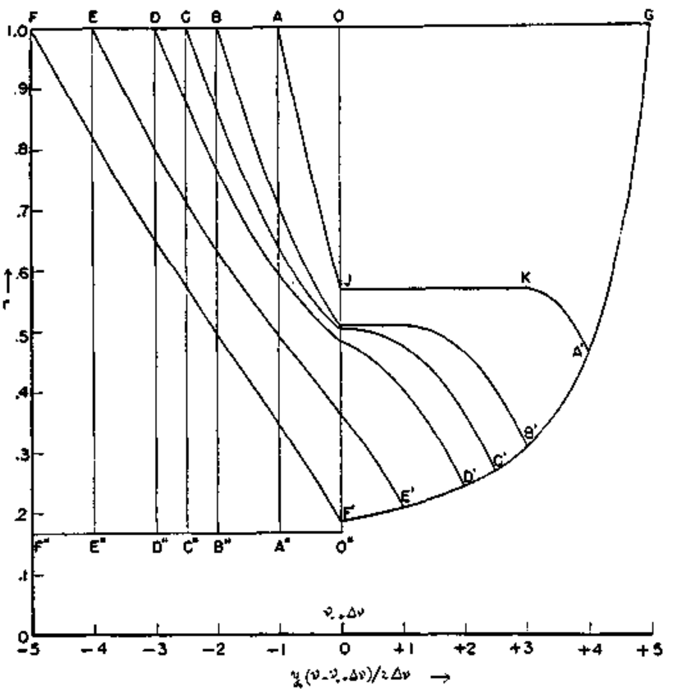


Fig. 7.

Eq. (68) and are tabulated together with the values of r for the remaining parts of the contours given by Eq. (115). The arguments in Table III are the frequency shifts measured from the red end $\nu_0-\Delta\nu$ of the contour in the unit $2\Delta\nu$.

The residual intensities tabulated in Table III are further illustrated as line contours in Fig. 7. In this figure the various contours are plotted on different frequency scales, the width $2\Delta\nu$ of $\sigma(\nu)$ always extending from the red end of the contour to O. Thus the contour $BB'G$ corresponds to a case in which the line formed under the same conditions in a static atmosphere would extend from B to O.

From Fig. 7 it is apparent that in all cases in which $D\nu > 2\sqrt{3}\Delta\nu$ the contour consists of four distinct parts, namely,

$$\left.\begin{array}{lr} \nu_0-\Delta\nu \leqslant \nu \leqslant \nu_0+\Delta\nu, & \text{(i)} \\ \nu_0+\Delta\nu \leqslant \nu \leqslant \nu_0-3\Delta\nu+2\mu_1 D\nu, & \text{(ii)} \\ \nu_0-3\Delta\nu+2\mu_1 D\nu \leqslant \nu \leqslant \nu_0-\Delta\nu+2\mu_1 D\nu, & \text{(iii)} \\ \nu_0-\Delta\nu+2\mu_1 D\nu \leqslant \nu \leqslant \nu_0+\Delta\nu+2\mu_1 D\nu. & \text{(iv)} \end{array}\right\} \tag{129}$$

In each of these parts r is given by a different analytical expression. It decreases from 1 in the first interval, remains constant in the second, and decreases some more in the third attaining its minimum at $\nu=\nu_0-\Delta\nu+2\mu_1 D\nu$. In the last interval it increases again to 1. It is in this fourth interval that the line contour is described by Eq. (115). In Fig. 7 we have indicated these four parts on the contour $AA'G$. The parts are respectively, AJ, JK, KA', and $A'G$. The reason for the existence of these four parts can be understood from a reference to Fig. 8. In this figure, which is similar to Fig. 2, the regions in which I_{+1} and I_{-1}, respectively, are different from constants (for varying z) are marked. We have further

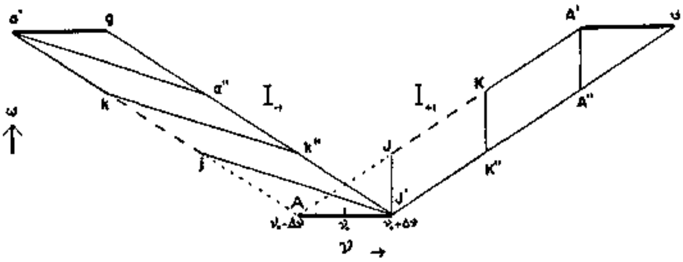


FIG. 8.

indicated the frequency intervals in which the different parts of the contour arise. (The lettering in Figs. 7 and 8 correspond). Now, according to our discussion in Sections 3 and 4, the outward intensity I_{+1} for a frequency ν interacts with the inward intensity I_{-1} for a frequency $\nu-2\mu_1\nu_0(w/c)$. Hence I_{+1} for the frequencies in the intervals AJ, JK, KA', and $A'G$ in their transfer through the atmosphere have interacted with I_{-1} for the frequencies in the regions AjJ', $jkk''J'$, $ka'a''k''$, and $a'ga''$, respectively. The reason for the existence of the four distinct parts in the line contour now becomes apparent. Moreover, this discussion makes it clear why it is that the problem increases in complexity as $\mu_1 D\nu/\Delta\nu$ decreases below unity.

Finally, it is of interest to compare the contours we have obtained with those which would be expected in an atmosphere in which no gradient of velocity exists. To discuss this case we have to go back to Eqs. (39) and (40). Setting $dw/dx=0$ in these equations and solving them with the boundary conditions appropriate to Schuster's problem we readily find that

$$\left.\begin{aligned} r&=\frac{1}{1+x_1} \quad (\nu_0-\Delta\nu\leqslant\nu\leqslant\nu_0+\Delta\nu)\\ &=0 \qquad\qquad \text{otherwise.}\end{aligned}\right\} \tag{130}$$

The contours are therefore rectangular. For $x_1=5$, $r=\frac{1}{6}$. These rectangular contours which will be obtained in the limit $D\nu=0$ are also shown in Fig. 7. Thus the contour $BB'G$ should be compared with $BB''O''O$; and similarly for the others.

7. REMARKS ON FUTURE WORK

The successful solution of a specific problem in the theory of moving atmospheres which we have presented in the preceding sections justifies the hope that it will be possible to solve problems more general and less idealized than the one considered in this paper. Indeed, there are several problems in the theory of moving atmospheres which come already within the scope of the methods developed in this paper. For example, there is the problem of the variation of line contours with the angle of emergence from the atmosphere. The solution to this problem will depend on the radiation field in the entire atmosphere. For, the intensity $I(\nu, z_1, \mu)$ of the radiation of frequency ν (as judged by an observer at rest with respect to the radiating surface at $z=0$) emergent in a direction with a direction cosine μ with respect to the positive normal can be expressed as an integral in the form (cf., Eq. (9))

$$I(\nu, z_1, \mu)=\int_0^{z_1} J(\nu, z, \mu)\exp\left\{-\int_z^{z_1}\rho\sigma(\nu-[\nu_0/c]w\mu)/\mu\right\}\frac{dz}{\mu}, \tag{131}$$

where

$$J(\nu, z, \mu)=\tfrac{1}{2}\int_{-1}^{+1} I\left(\nu-\nu_0\frac{w}{c}\mu+\nu_0\frac{w}{c}\mu', z, \mu'\right)d\mu'. \tag{132}$$

In the first approximation we can express the integral on the right-hand side of Eq. (132) as a Gauss sum with two terms. Thus,

$$J(\nu, z, \mu) = \tfrac{1}{2}\left\{ I_{+1}(\nu - \nu_0\frac{w}{c}\mu + \nu_0\frac{w}{c}\mu_1) + I_{-1}\left(\nu - \nu_0\frac{w}{c}\mu - \nu_0\frac{w}{c}\mu_1, z\right)\right\}. \tag{133}$$

The source function J can, therefore, be expressed in terms of the solutions for $I_{+1}(\nu, z)$ and $I_{-1}(\nu, z)$ which we have found in Section 5 and there will be no formal difficulty in solving for $I(\nu, z_1, \mu)$ according to Eq. (131).

Again, the determination of $I(\nu, z_1, \mu)$ by the procedure we have outlined above will be of particular importance for deriving contours comparable to those *observed* in cases in which the photospheric surface is itself moving with a velocity w_0. It will be recalled in this connection that our discussion of the equation of transfer involves no assumption concerning w_0 since everything was referred to an observer at rest with respect to the surface at $z=0$ and $t=t_1$. However, the line contour as seen by an observer outside the star will not be given by $F(\nu, z_1)$, as allowance will have to be made for the fact that the photospheric surface from which the radiation is emerging at an angle ϑ has a motion $w_0 \cos\vartheta$ towards the observer. Accordingly, the contour as judged by an external observer at a great distance from the star will be determined by

$$\mathfrak{F}(\nu) = 2\int_0^1 I\left(\nu + \nu_0\frac{w_0}{c}\mu,\ z_1, \mu\right)\mu d\mu, \tag{134}$$

where $I(\nu, z_1, \mu)$ has the same meaning as in Eq. (131).

Another problem which can be solved by the methods of the present paper is the radiative equilibrium of a planetary nebula. It is known that large differential motions are present in planetary nebulae and a problem of considerable interest relates to the question of the radiation pressure in the Lyman α-radiation. It can be shown that with the same assumptions (27) and (28) concerning $\sigma(\nu)$ and the variation of w through the atmosphere, the problem can be reduced to a boundary value problem very similar to the one considered in Sections 4 and 5 and the solution can also be found by similar methods.

And finally there is the general problem of line formation in moving atmospheres in which $\sigma(\nu)$ is allowed to be more general than the rectangular form considered in this paper. It can be shown that under these more general conditions the problem can still be reduced to a boundary value problem in hyperbolic equations. If the velocity be assumed to vary linearly with the optical depth, the equation we have to consider differs from (59) only in the occurrence of a factor depending on y in front of f. It does not seem impossible that with suitable simplifications, progress toward the solution of these more difficult problems can be made.

THE TRANSFER OF RADIATION IN STELLAR ATMOSPHERES

S. CHANDRASEKHAR

1. **Introductory remarks.** The advances in the various branches of theoretical physics have often resulted in the creation of new mathematical disciplines: disciplines which, in their way, are as characteristic of the subjects as the physical phenomena they are devoted to. That several of the earlier Gibbs lectures should illustrate this point is not surprising since the example provided by the scientific writings of Gibbs is indeed the most striking in this connection. With your permission, I would like to illustrate the same thing this evening, on a more modest level, by considering the recent advances in theoretical astrophysics. More specifically, I should like to show how astrophysical studies relating to the transfer of radiation in stellar atmospheres have led to some characteristic developments in the theory of integro-differential and functional equations.

2. **The physical problem.** In a general way, the preoccupation of the astrophysicist with the transport of radiant energy in an atmosphere which absorbs, emits and scatters radiation is not difficult to understand. For, it is in the characteristics of the radiation emergent from a star—in its variation over the stellar disc, and in its distribution with wavelength through the spectrum—that he has to seek for information concerning the constitution and structure of the stellar atmosphere.

3. **Definitions. Intensity.** Now, the basic equation which governs the transfer of radiation in any medium is the so-called *equation of transfer*. This is the equation which governs the variation of radiant intensity in terms of the local properties of the medium. But, first, I should explain that for most purposes it is sufficient to characterize the radiation field by the specific intensity I; in terms of this quantity the amount of radiant energy crossing an element of area $d\sigma$, in a direction inclined at an angle ϑ to the normal, and confined to an element of solid angle $d\omega$, and in time dt can be expressed in the form

$$I \cos \vartheta d\sigma d\omega dt. \tag{1}$$

Accordingly, to characterize a radiation field completely, the specific

The twentieth Josiah Willard Gibbs lecture delivered at Swarthmore, Pennsylvania, December 26, 1946, under the auspices of the American Mathematical Society; received by the editors January 13, 1947.

intensity will have to be defined at every point and for every direction through a point.

4. The equation of transfer. To establish the equation which the intensity must satisfy in a stationary radiation field (that is, in a radiation field which is constant with time) we consider the equilibrium with respect to the transport of energy of a small cylindrical element of cross section $d\sigma$ and height ds in the medium. From the definition of intensity, it follows that the difference in the radiant energy crossing the two faces normally, in a time dt and confined to an element of solid angle $d\omega$, is

$$\frac{dI}{ds}\, ds\, d\sigma\, d\omega\, dt. \tag{2}$$

Here I denotes the specific intensity in the direction of s. This difference (2) must arise from the excess of emission over absorption of radiant energy in time dt and the element of solid angle $d\omega$ considered. Now the absorption of radiation by an element of mass is expressed in terms of the *mass absorption coefficient* denoted by κ, such that, of the energy $I d\sigma d\omega dt$ normally incident on the face $d\sigma$, the amount absorbed in time dt is

$$\kappa \rho ds \times I d\sigma d\omega dt, \tag{3}$$

where ρ denotes the density of the material. Similarly, the emission coefficient j is defined in such a way that the element of mass $\rho d\sigma ds$ emits in directions confined to $d\omega$ and in time dt an amount of radiant energy given by

$$j \rho d\sigma ds d\omega dt. \tag{4}$$

It should be noted that, according to our definition, the emission coefficient j can very well depend on the direction at each point, differing in this respect from the absorption coefficient which is a function of position only.

From our earlier remarks and the expressions (3) and (4) for absorption and emission, we have the equation of transfer

$$\frac{dI}{\rho ds} = -\kappa I + j, \tag{5}$$

expressing the conservation of radiant energy.

5. The equation of transfer for plane-parallel atmospheres. The source function. In our further discussion we shall restrict ourselves to transfer problems in a semi-infinite plane-parallel atmosphere (Fig. 1). Let z measure distances normal to the plane of stratification and ϑ and φ the corresponding polar angles. The specific intensity I in such an atmosphere will, in general, depend on all three variables, z, ϑ

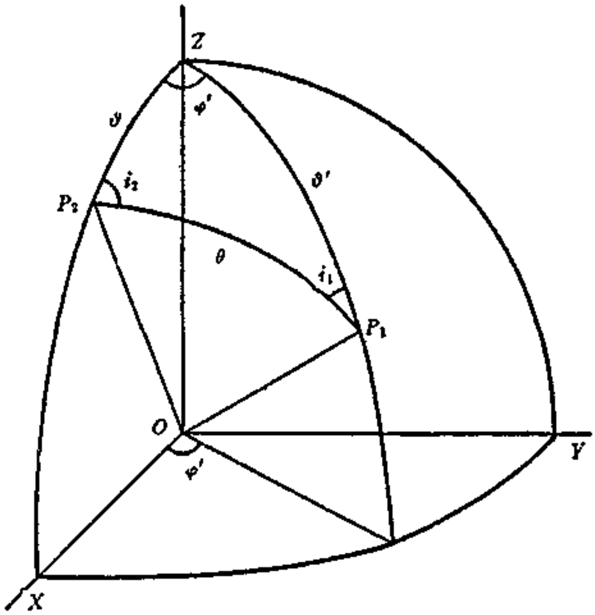


FIG. 1.

and φ. For such an atmosphere the equation of transfer can be written in the form

$$\cos\vartheta\frac{dI}{\rho dz} = -\kappa I + j. \tag{6}$$

It is convenient at this stage to introduce the *optical thickness*

$$\tau = \int_z^\infty \kappa\rho dz, \tag{7}$$

measured from the boundary *inward* as our variable instead of z. In terms of τ, the equation of transfer becomes

$$\cos\vartheta\frac{dI}{d\tau} = I(\tau, \vartheta, \varphi) - \mathfrak{J}(\tau, \vartheta, \varphi), \tag{8}$$

where we have written

$$\mathfrak{J} = \frac{j}{\kappa}. \tag{9}$$

This is the *source function.*

6. **The source function for a scattering atmosphere. The phase function.** Different physical problems lead to different *functional* dependences of the source function on I. To illustrate the nature of such functional dependences we shall consider the case of a *scattering atmosphere.* In such an atmosphere, absorption arises simply on account of the fact that when a pencil of radiation is incident on an element of gas, a determinate part of the radiant energy is scattered in other directions. In the same way, the radiation in any given direction can be reinforced by the scattering of radiation from other directions into the particular direction considered. It is now evident that, to formulate quantitatively the concept of scattering, we must specify the angular distribution of the scattered radiation when a pencil of radiation is incident on an element of gas. This angular distribution is generally given by a *phase function* $p(\cos\Theta)$ in such a way that

$$p(\cos\Theta)\frac{d\omega}{4\pi} \tag{10}$$

governs the probability that radiation will be scattered in a direction inclined at an angle Θ with the direction of incidence. If, as we have assumed, scattering is the only process by which radiation and matter interact with each other, then it is evident that

$$\int p(\cos\Theta)\frac{d\omega}{4\pi} = 1 \tag{11}$$

when the integration is extended over the complete sphere. The corresponding source function is

$$\mathfrak{J}(\tau,\vartheta,\varphi) = \frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi} I(\tau,\vartheta',\varphi')p(\cos\vartheta\cos\vartheta' + \sin\vartheta\sin\vartheta'\cos[\varphi'-\varphi])\sin\vartheta' d\vartheta' d\varphi'. \tag{12}$$

When the scattering is *isotropic*

$$p(\cos\Theta) = 1 \tag{13}$$

and the source function reduces to

$$\mathfrak{J}(\tau) = \frac{1}{4\pi}\int_0^\pi \int_0^{2\pi} I(\tau, \vartheta', \varphi') \sin \vartheta' d\vartheta' d\varphi'. \tag{14}$$

Next to this isotropic case, greatest interest is attached to ***Rayleigh's phase function***

$$p(\cos \Theta) = \frac{3}{4}(1 + \cos^2 \Theta). \tag{15}$$

In general, we may, however, suppose that the phase function can be expanded as a series in Legendre polynomials of the form

$$p(\cos \Theta) = \sum \varpi_l P_l(\cos \Theta) \tag{16}$$

where $\varpi_0 = 1$ and ϖ_l ($l \neq 0$) are some constants. (The condition $\varpi_0 = 1$ follows from equation (11).)

7. **Phase-functions with absorption.** Our foregoing remarks strictly apply to the case of perfect scattering; that is, to cases in which the interaction between matter and radiation results only in the transformation of radiation in one direction into radiation in other directions. However, cases are known in which, in the process of scattering, a certain amount of the radiation is consumptively transformed into other forms of energy. In such cases, we can still describe the scattered radiation by a phase function $p(\cos \Theta)$; but the "normalizing condition" (11) will no longer be satisfied and we shall have instead

$$\int p(\cos \Theta) \frac{d\omega}{4\pi} = \lambda \qquad (\lambda \leqq 1). \tag{17}$$

The quantity λ defined in this way is called the *albedo*. Formally, this does not introduce anything very essential, as the source function will continue to be given by equation (12).

The case

$$p(\cos \Theta) = \lambda(1 + x \cos \Theta) \qquad (-1 < x < 1) \tag{18}$$

is of particular interest for the analysis of planetary illumination.

8. **The problem of a semi-infinite plane-parallel atmosphere with a constant net flux.** Returning to the case of a perfectly scattering atmosphere, we observe that the equation of transfer

$$\cos \vartheta \frac{dI(\tau, \vartheta, \varphi)}{d\tau} = I(\tau, \vartheta, \varphi) - \frac{1}{4\pi}\int_0^\pi \int_0^{2\pi} I(\tau, \vartheta', \varphi') p(\cos \Theta) d\omega', \tag{19}$$

where

$$\cos \Theta = \cos \vartheta \cos \vartheta' + \sin \vartheta \sin \vartheta' \cos (\varphi' - \varphi), \tag{20}$$

admits the first integral

$$\frac{d}{d\tau}\int_0^{\pi}\int_0^{2\pi} I(\tau, \vartheta, \varphi)\cos\vartheta\sin\vartheta d\vartheta d\varphi = 0, \tag{21}$$

or

$$\pi F = \int_0^{\pi}\int_0^{2\pi} I(\tau, \vartheta, \varphi)\cos\vartheta\sin\vartheta d\vartheta d\varphi = \text{constant}. \tag{22}$$

The integral representing F is proportional to the net flux of radiant energy crossing unit area in the plane of stratification; and this is a constant for a perfectly scattering atmosphere. Because of this constancy of the net flux, a type of transfer problem which arises in connection with such perfectly scattering atmsopheres is that of a plane-parallel semi-infinite atmosphere with no incident radiation, and a constant net flux πF through the atmosphere normal to the plane of stratification. It is evident that for these problems, solutions of the appropriate equations of transfer must be sought which depend only on τ and ϑ, that is, solutions which exhibit rotational symmetry about the z-axis. Further, in these problems, the greatest interest is attached to the angular distribution of the *emergent radiation* $I(0, \mu)$ at $\tau = 0$ and for $\mu > 0$.

Finally, we may note that for the cases of isotropic scattering and scattering in accordance with Rayleigh's phase function, the equations of transfer have explicitly the forms

$$\boxed{\mu\frac{dI(\tau, \mu)}{d\tau} = I(\tau, \mu) - \frac{1}{2}\int_{-1}^{+1} I(\tau, \mu')d\mu',} \tag{I}$$

and

$$\boxed{\begin{aligned}\mu\frac{dI(\tau, \mu)}{d\tau} = I(\tau, \mu) - \frac{3}{16}\Big[(3 - \mu^2)\int_{-1}^{+1} I(\tau, \mu')d\mu' \\ + (3\mu^2 - 1)\int_{-1}^{+1} I(\tau, \mu')\mu'^2 d\mu'\Big].\end{aligned}} \tag{II}$$

In the foregoing equations we have written μ for $\cos\vartheta$.

We require solutions of the integro-differential equations (I) and (II) which satisfy the boundary condition

$$I(0, \mu) = 0 \qquad \text{for } -1 < \mu < 0. \tag{23}$$

9. **The problem of diffuse reflection.** Another type of transfer problem which arises quite generally (that is, also in cases when λ is different from unity) is that relating to the phenomenon of *diffuse reflection.* In these contexts we are principally interested in the angular distribution $I(\vartheta, \varphi; \vartheta_0, \varphi_0)$ of the radiation diffusely reflected when a parallel beam of radiation of net flux πF per unit area normal to itself is incident on the atmosphere in some specified direction $(\pi-\vartheta_0, \varphi_0)$.

One general remark relating to problems of diffuse reflection may be made here. It is that, at any level, we may distinguish between the reduced incident radiation $\pi F e^{-\tau \sec \vartheta_0}$ which penetrates to the depth τ without suffering any scattering or absorbing mechanism, and the *diffuse radiation field* which arises in consequence of multiple scattering. We shall characterize this diffuse radiation field by $I(\tau, \mu, \varphi)$. Making this distinction between these two fields of radiation, we may write the equation of transfer for problems of diffuse reflection in the form

$$\begin{aligned} \cos\vartheta \frac{dI(\tau, \vartheta, \varphi)}{d\tau} &= I(\tau, \vartheta, \varphi) - \frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi} I(\tau, \vartheta', \varphi') \\ &\times p(\cos\vartheta\cos\vartheta' + \sin\vartheta\sin\vartheta'\cos[\varphi'-\varphi])\sin\vartheta' d\vartheta' d\varphi' \\ &- \frac{1}{4}Fe^{-\tau\sec\vartheta_0}p(-\cos\vartheta\cos\vartheta_0 + \sin\vartheta\sin\vartheta_0\cos[\varphi_0-\varphi]). \end{aligned} \tag{24}$$

For the cases of isotropic scattering, and scattering in accordance with the phase functions (15) and (18), we have the equations

$$\text{III} \qquad \boxed{\mu \frac{dI(\tau, \mu)}{d\tau} = I(\tau, \mu) - \frac{1}{2}\int_{-1}^{+1} I(\tau, \mu')d\mu' - \frac{1}{4}Fe^{-\tau/\mu_0},}$$

$$\text{IV} \qquad \boxed{\begin{aligned} \mu \frac{dI(\tau, \vartheta, \varphi)}{d\tau} &= I(\tau, \vartheta, \varphi) \\ &- \frac{\lambda}{4\pi}\int_0^{\pi}\int_0^{2\pi} I(\tau, \vartheta', \varphi')[1 + x(\cos\vartheta\cos\vartheta' \\ &\qquad + \sin\vartheta\sin\vartheta'\cos[\varphi'-\varphi])]\sin\vartheta' d\vartheta' d\varphi' \\ &- \frac{1}{4}\lambda F e^{-\tau\sec\vartheta_0}[1 + x(-\cos\vartheta\cos\vartheta_0 + \sin\vartheta\sin\vartheta_0\cos\varphi)] \end{aligned}}$$

and

$$
\text{V} \qquad
\begin{aligned}
\cos\vartheta \frac{dI(\tau,\vartheta,\varphi)}{d\tau} &= I(\tau,\vartheta,\varphi) \\
&- \frac{3}{16\pi}\int_0^{\pi}\int_0^{2\pi} I(\tau,\vartheta',\varphi')\Big[1+\cos^2\vartheta\cos^2\vartheta' \\
&+ \frac{1}{2}\sin^2\vartheta\sin^2\vartheta' + 2\cos\vartheta\sin\vartheta\cos\vartheta'\sin\vartheta'\cos(\varphi'-\varphi) \\
&+ \frac{1}{2}\sin^2\vartheta\sin^2\vartheta'\cos 2(\varphi'-\varphi)\Big]\sin\vartheta' d\vartheta' d\varphi' \\
&- \frac{3}{16}Fe^{-\tau\sec\vartheta_0}\Big[1+\cos^2\vartheta\cos^2\vartheta_0 \\
&+ \frac{1}{2}\sin^2\vartheta\sin^2\vartheta_0 - 2\cos\vartheta\sin\vartheta\cos\vartheta_0\sin\vartheta_0\cos\varphi \\
&+ \frac{1}{2}\sin^2\vartheta\sin^2\vartheta_0\cos 2\varphi\Big].
\end{aligned}
$$

And again we require solutions of these equations which satisfy the boundary condition (23).

It appears from the foregoing equations that for problems of diffuse reflection which arise in practice, we can expand $I(\tau, \mu, \varphi)$ in a finite Fourier series of the form

$$(25) \qquad I(\tau,\mu,\varphi) = \sum_{m=0}^{N} I^{(m)}(\tau,\mu)\cos m\varphi.$$

And finally, it should be remarked that in the problem of diffuse reflection by a semi-infinite atmosphere, greatest interest is attached to the *scattering function* $S(\mu, \varphi; \mu_0, \varphi_0)$ which is related to the reflected intensity $I(0, \mu, \varphi)$ according to the equation

$$(26) \qquad I(0,\mu,\varphi) = \frac{1}{4\mu}FS(\mu,\varphi;\mu_0,\varphi_0).$$

Defined in this manner $S(\mu, \varphi; \mu_0, \varphi_0)$ *must be symmetrical in the pair of variables* (μ, φ) *and* (μ_0, φ_0). This is required by a *principle of reciprocity* due to Helmholtz.

10. The equations of transfer incorporating the polarization of the scattered radiation in accordance with Rayleigh's law. The axially symmetric case. The transfer problems which we have described in the preceding sections do not exhaust all the physical possibilities that may arise. For example, an important factor which we have not included in our discussion so far concerns the *state of polarization* of the radiation field. For many problems, this must be taken into account before we may be said to have an adequate physical description of the phenomenon under consideration, since, on scattering, light in general gets polarized. On Rayleigh's classical laws for instance, when an initially unpolarized beam is scattered in a direction inclined at an angle Θ with the direction of incidence, it becomes partially plane polarized with a ratio of intensities $1 : \cos^2 \Theta$ in directions perpendicular, respectively, parallel to the *plane of scattering*. (This is the plane which contains the directions of the incident and the scattered light.) It is, therefore, apparent that the diffuse radiation field in a scattering atmosphere must be partially polarized. The question immediately arises as to how best we can characterize the radiation field under these conditions in order that the relevant equations of transfer may be most conveniently formulated. This is a fundamental question: on its answer will depend the solution of a variety of problems, including the important one of the illumination and the polarization of the sunlit sky—a problem which was in fact first considered in a general way by Rayleigh in 1871. It is, therefore, somewhat surprising to find that even the basic equations of the problem should not have been written down before the problem arose in an astrophysical context earlier this year. Perhaps I may briefly explain the particular astrophysical problem which gave the final impetus for formulating and solving transfer problems in which the polarization characteristics of the radiation field are incorporated and allowance for the polarization of the scattered radiation is made:

It is now believed that in the atmospheres of stars with surface temperatures exceeding 15,000°K, the transfer of radiation must be predominantly controlled by the scattering by free electrons. And, according to J. J. Thomson's laws for this process, scattering by free electrons must result in the partial polarization of the scattered radiation. Thomson's laws agree, in fact, with Rayleigh's as regards the angular distribution and the state of polarization of the scattered radiation.

Now, if our belief in the important role played by the Thomson scattering in the atmospheres of hot stars is correct, we should expect that the radiation emergent from the atmospheres of such stars should be partially polarized. The question of the *degree of polarization* to be expected under such conditions therefore becomes one of exceptional astrophysical interest. But, before we can answer this question, we must formulate the relevant equations of transfer, distinguishing the different states of polarization, and solve them! One circumstance, however, simplifies this particular problem. It is, that in a plane-parallel atmosphere with no incident radiation, the radiation field at any point must exhibit rotational symmetry about the z-axis; accordingly, the plane of polarization must be along the principal meridian (or, at right angles to it). In other words, under the conditions of our present problem, the specific intensities I_l and I_r, referring respectively to the two states of polarization in which the electric vector vibrates in the meridian plane and at right angles to it, are sufficient to characterize the radiation field completely. And the relevant equations of transfer appropriate for this problem are found to be:

$$
\text{VI}\qquad \boxed{\begin{aligned}
\mu\frac{dI_l}{d\tau} &= I_l - \frac{3}{8}\left\{\int_{-1}^{+1} I_l(\tau,\mu')\left[2(1-\mu'^2)+\mu^2(3\mu'^2-2)\right]d\mu' \right.\\
&\qquad\left. + \mu^2\int_{-1}^{+1} I_r(\tau,\mu')d\mu'\right\},\\
\mu\frac{dI_r}{d\tau} &= I_r - \frac{3}{8}\left\{\int_{-1}^{+1} I_l(\tau,\mu')\mu'^2 d\mu' + \int_{-1}^{+1} I_r(\tau,\mu')d\mu'\right\}.
\end{aligned}}
$$

In other words, we now have a pair of simultaneous integro-differential equations to solve with the boundary conditions

$$
I_l(0,\mu) = I_r(0,\mu) = 0 \qquad (-1 < \mu < 0). \tag{27}
$$

11. **The parametric representation of partially plane-polarized light. The vector form of the equation of transfer.** Returning to the general problem of a partially polarized radiation field, we at once realize that the intensities I_l and I_r are not sufficient to characterize the radiation field completely. A third quantity is needed which will determine the plane of polarization. But the inclination of the plane of polarization itself to the meridian plane (for example) would be a most unsuitable parameter to choose: for, a proper parametric representation of partially polarized light should not only provide a complete set of variables to describe the radiation field, but also enable a convenient formulation of the rule of composition of a mixture of *independent* polarized streams in a given direction into a single partially polarized stream. On consideration it appears that the most convenient third variable to choose is a quantity U also of the dimensions of intensity and in terms of which the inclination, χ, of the plane of polarization to the direction to which I_l refers is given by

(28) $$\chi = \frac{1}{2}\tan^{-1}\frac{U}{I_l - I_r}.$$

The rule of composition of a number of independent plane-polarized streams can be formulated as follows: If we have a number of independent plane-polarized streams of intensities $I^{(n)}$ (say) in a given direction, the resulting mixture will be partially plane-polarized and will be described by the parameters

(29) $$I_l = \sum I^{(n)}\cos^2\chi_n; \qquad I_r = \sum I^{(n)}\sin^2\chi_n,$$

and

(29′) $$U = \sum I^{(n)}\sin 2\chi_n$$

where χ_n denotes the inclination of the plane of polarization of the stream $I^{(n)}$ to the direction (in the transverse plane of the electric and the magnetic vectors) to which I_l refers.

With the rule of composition, as we have just stated, the equations of transfer for I_l, I_r and U can be formulated in accordance, for example, with Rayleigh's laws. For the problem of diffuse reflection of a partially plane-polarized parallel beam of radiation (characterized by the parameters F_l, F_r and $U^{(0)}$) incident on a plane-parallel atmosphere in the direction $(-\mu_0, \varphi_0)$, the three equations of transfer for the intensities I_l, I_r and U of the diffuse radiation field can be combined into a single vector equation of the form

VII $$\boxed{\begin{aligned}\mu\frac{dI}{d\tau} &= I(\tau, \mu, \varphi)\\ &\quad - \frac{3}{16\pi}Q\int_{-1}^{+1}\int_0^{2\pi} J(\mu, \varphi; -\mu', \varphi')I(\tau, \mu', \varphi')d\mu' d\varphi'\\ &\quad - \frac{3}{16}QJ(\mu, \varphi; \mu_0, \varphi_0)Fe^{-\tau/\mu_0},\end{aligned}}$$

where

(30) $$I = (I_l, I_r, U), \qquad F = (F_l, F_r, U^{(0)}),$$

(31) $$Q = \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 2\end{pmatrix},$$

and

$$J(\mu,\varphi;\mu_0,\varphi_0)$$

$$(32)\quad = \begin{pmatrix} \begin{matrix} 2(1-\mu^2)(1-\mu_0^2)+\mu^2\mu_0^2 \\ -4\mu\mu_0(1-\mu^2)^{1/2}(1-\mu_0^2)^{1/2}\cos(\varphi-\varphi_0) \\ +\mu^2\mu_0^2\cos 2(\varphi-\varphi_0) \end{matrix} & \begin{matrix} \mu^2 \\ \\ -\mu^2\cos 2(\varphi-\varphi_0) \end{matrix} & \begin{matrix} -2\mu(1-\mu^2)^{1/2}(1-\mu_0^2)^{1/2}\sin(\varphi-\varphi_0) \\ +\mu^2\mu_0\sin 2(\varphi-\varphi_0) \end{matrix} \\ \\ \mu_0^2-\mu_0^2\cos 2(\varphi-\varphi_0) & 1+\cos 2(\varphi-\varphi_0) & -\mu_0\sin 2(\varphi-\varphi_0) \\ \\ \begin{matrix} -2\mu_0(1-\mu^2)^{1/2}(1-\mu_0^2)^{1/2}\sin(\varphi-\varphi_0) \\ +\mu\mu_0^2\sin 2(\varphi-\varphi_0) \end{matrix} & -\mu\sin 2(\varphi-\varphi_0) & \begin{matrix} (1-\mu^2)^{1/2}(1-\mu_0^2)^{1/2}\cos(\varphi-\varphi_0) \\ -\mu\mu_0\cos 2(\varphi-\varphi_0) \end{matrix} \end{pmatrix}.$$

The matrix QJ therefore plays the same role for this problem as the phase function $p(\cos\Theta)$ does in the simpler problems in which polarization is not allowed for.

We require to solve (VII) which satisfies the boundary condition

$$(33)\qquad I(0,\mu,\varphi)\equiv 0 \qquad (-1<\mu<0).$$

Since the equation of transfer is linear in F, it is clear that the intensity $I(\mu,\varphi;\mu_0,\varphi_0)$ diffusely reflected by the atmosphere in the outward direction (μ,φ) must be expressible in the form

$$(34)\qquad I(\mu,\varphi;\mu_0,\varphi_0)=\frac{3}{16\mu}QS(\mu,\varphi;\mu_0,\varphi_0)F.$$

And our principal interest in this problem is in the determination of this *scattering matrix* S.

12. The Schwarzschild-Milne integral equation. In the preceding sections we have formulated several typical problems in the theory of radiative transfer and we have seen how all of them lead to integro-differential equations of varying degrees of complexity. The question now arises as to whether there are any general methods for solving such equations. Until recently, the only example of such integro-differential equations which had been considered to any extent is the simplest we have written down, namely equation (I). And here the method used was one of reducing the integro-differential equation to an integral equation. This reduction to an integral equation was achieved in the following way:

Writing equation (I) in the form

$$(35)\qquad \mu\frac{dI}{d\tau}=I-J(\tau),$$

where

$$(36)\qquad J(\tau)=\frac{1}{2}\int_{-1}^{+1}I(\tau,\mu)d\mu,$$

we observe that its formal solution is

$$
(37)\qquad
\begin{aligned}
I(\tau, \mu) &= \int_\tau^\infty e^{-(t-\tau)/\mu} J(t) \frac{dt}{\mu} \qquad (0 < \mu \leqq 1) \\
&= -\int_0^\tau e^{-(t-\tau)/\mu} J(t) \frac{dt}{\mu} \qquad (-1 \leqq \mu < 0).
\end{aligned}
$$

Substituting this solution for $I(\tau, \mu)$ back into the equation defining $J(\tau)$ (equation (36)) we find, after some elementary reductions, that

$$
(38)\qquad J(\tau) = \frac{1}{2}\int_0^\infty E_1(|t - \tau|) J(t) dt,
$$

where

$$
(39)\qquad E_1(x) = \int_x^\infty e^{-y} \frac{dy}{y},
$$

is the "first exponential integral." This is the *Schwarzschild-Milne integral equation* for $J(\tau)$.

13. The method of Hopf and Wiener. The integral equation (38) for $J(\tau)$ is seen to be homogeneous with a symmetric kernel. However, the kernel has a logarithmic singularity at $t = \tau$. Integral equations of this general structure have therefore a definite mathematical interest and have, in fact, been considered by a number of mathematicians including E. Hopf, J. Brönstein and N. Wiener. But this method based on the reduction of the equation of transfer to a linear integral equation for the source function has not so far been extended successfully to the solution of the other more interesting equations of transfer we have formulated. I shall therefore pass on to a different method of attack on these equations which has been developed during the past three years and which allows the solution of practically all the different types of transfer problems which arise in practice, by a systematic method of approximation. Also, for the particular problems enumerated, *exact solutions* for the angular distribution of the emergent (or the diffusely reflected) radiation can be found explicitly by passing to the limit of "infinite" approximation.

14. The method of replacing the integro-differential equations by systems of linear equations. The fundamental idea in this new method of solving an integro-differential equation is to replace it in a certain approximation by a system of ordinary linear equations. This reduction to an equivalent linear system is made by approximating the integrals over μ which occur in the equation by a weighted average of the values which the various integrands take at a suitably selected set of points in the interval $(-1, +1)$. It is evident that in making the necessary division of the interval $(-1, +1)$, we must be guided by reasons of "economy" in the sense that with a given number of divisions we must try to achieve the maximum accuracy in the evaluation of the integrals. The problem which we encounter here is therefore the same as that considered by Gauss in 1814 in deriving his formula for numerical quadratures. In this formula of Gauss, the interval $(-1, +1)$ is divided according to the zeros of a Legendre polynomial, $P_m(\mu)$ (say), and the integral of a function $f(\mu)$ over the interval $(-1, +1)$ is expressed as a sum in the form

$$\int_{-1}^{+1} f(\mu)d\mu \simeq \sum a_j f(\mu_j) \tag{40}$$

where the *weights* a_j are given by

$$a_j = \frac{1}{P_m'(\mu_j)} \int_{-1}^{+1} \frac{P_m(\mu)}{\mu - \mu_j} d\mu. \tag{41}$$

The reason Gauss's formula is superior to other formulae for quadratures in the interval $(-1, +1)$ is that for a given m it evalues the integrals *exactly* for *all* polynomial of degree less than, or equal to, $2m-1$; in other words, it is almost "twice as accurate" as we should expect a formula which uses only m values of the function in the interval to be.

For certain formal reasons, we shall not enter into here, in our further work we shall use divisions of the interval $(-1, +1)$ only according to the zeros of the even Legendre polynomials. Further, when the division according to the zeros of $P_{2n}(\mu)$ is selected, we shall say that we are working in the nth *approximation*. It is, therefore, apparent that when an equation of transfer is reduced to an equivalent linear system in the nth approximation, we are effectively using a polynomial representation of order $4n-1$ (in μ) for the various intensities which occur in our problem.

Finally, it may be observed that, when working in the nth approximation, we may use the following relations between the points of the Gaussian division, μ_j, and the Gaussian weights a_j:

$$a_j = a_{-j}; \qquad \mu_j = -\mu_{-j} \qquad (j = \pm 1, \cdots, \pm n), \tag{42}$$

and

$$\sum_{j=1}^{n} a_j \mu_j^{2m} = \frac{1}{2m+1} \qquad (m = 1, \cdots, 2n-1). \tag{43}$$

The identity (43) arises from the fact that in the nth approximation Gauss's formula evaluates exactly integrals of all polynomials of degree less than or equal to $4n-1$.

15. **The solution of equation (I) by the new method.** We shall illustrate the method of solving linear integro-differential equations by reducing them to equivalent linear systems in the various approximations by considering equation I.

In our nth approximation we replace equation (I) by the system of $2n$ equations

$$\mu_i \frac{dI_i}{d\tau} = I_i - \frac{1}{2} \sum a_j I_j \qquad (i = \pm 1, \cdots, \pm n), \tag{44}$$

where, for the sake of brevity, we have written $I_i(\tau)$ for $I(\tau, \mu_i)$.

In solving the system of equations represented by equation (44), we shall first obtain the different linearly independent solutions and later, by combining them, obtain the general solution.

As our first trial, we seek a solution of the form

$$I_i = g_i e^{-k\tau} \qquad (i = \pm 1, \cdots, \pm n), \tag{45}$$

where the g_i's and k are constants, for the present unspecified. Introducing equation (45) in equation (44), we obtain the relation

$$g_i(1 + \mu_i k) = \frac{1}{2} \sum a_j g_j. \tag{46}$$

Hence,

$$g_i = \frac{\text{constant}}{1 + \mu_i k} \qquad (i = \pm 1, \cdots, \pm n), \tag{47}$$

where the "constant" is independent of i. Substituting the foregoing form for g_i in equation (46), we obtain the *characteristic equation* for k:

$$1 = \frac{1}{2} \sum \frac{a_j}{1 + \mu_j k}. \tag{48}$$

Remembering that $a_{-j} = a_j$ and $\mu_{-j} = -\mu_j$, we can rewrite the characteristic equation in the form

$$1 = \sum_{j=1}^{n} \frac{a_j}{1 - \mu_j^2 k^2}. \tag{49}$$

It is, therefore, seen that k^2 must satisfy an algebraic equation of order n. However, since

$$\sum_{j=1}^{n} a_j = 1, \tag{50}$$

$k^2=0$ is a root of equation (49). Equation (48) accordingly admits only $2n-2$ distinct nonzero roots which must occur in pairs as

$$\pm k_\alpha \qquad (\alpha = 1, \cdots, n-1). \tag{51}$$

And corresponding to these $2n-2$ roots, we have $2n-2$ independent solutions of equation (44). The solution is completed by observing that

$$I_i = b(\tau + \mu_i + Q) \qquad (i = \pm 1, \cdots, \pm n), \tag{52}$$

where b and Q are two arbitrary constants, also satisfies equation (44). The general solution of equation (44) can therefore be written in the form

$$I_i = b\left\{\sum_{\alpha=1}^{n-1} \frac{L_\alpha e^{-k_\alpha\tau}}{1+\mu_i k_\alpha} + \sum_{\alpha=1}^{n-1} \frac{L_{-\alpha} e^{+k_\alpha\tau}}{1-\mu_i k_\alpha} + \tau + \mu_i + Q\right\} \qquad (i = \pm 1, \cdots, \pm n), \tag{53}$$

where b, $L_{\pm\alpha}$ $(\alpha=1, \cdots, n-1)$ and Q are the $2n$ arbitrary constants of integration.

For the astrophysical problem on hand, the boundary conditions are that none of the I_i's increase exponentially as $\tau\to\infty$ and that, further, there is no radiation incident on $\tau=0$. The first of these conditions requires that in the general solution (53) we omit all terms in $\exp(+k_\alpha\tau)$, thus leaving

$$I_i = b\left\{\sum_{\alpha=1}^{n-1} \frac{L_\alpha e^{-k_\alpha\tau}}{1+\mu_i k_\alpha} + \tau + \mu_i + Q\right\} \qquad (i = \pm 1, \cdots, \pm n). \tag{54}$$

Next, the absence of any radiation in the directions $-1<\mu<0$ at $\tau=0$ implies that in our approximation we must require

$$I_{-i} = 0 \qquad \text{at } \tau = 0 \text{ for } i = 1, \cdots, n. \tag{55}$$

Hence, according to equation (54),

$$\sum_{\alpha=1}^{n-1} \frac{L_\alpha}{1-\mu_i k_\alpha} - \mu_i + Q = 0 \qquad (i = 1, \cdots, n). \tag{56}$$

These are the n equations which determine the n constants of integration L_α $(\alpha=1, \cdots, n-1)$ and Q. The constant b is left arbitrary, and this, as we shall presently see, is related to the assigned constant net flux of radiation through the atmosphere.

16. **Some elementary identities.** The further discussion of the solution obtained in §15 requires certain relations which we shall now establish.

Let

$$D_m(x) = \sum_i \frac{a_i\mu_i^m}{1+\mu_i x} = (-1)^m \sum_i \frac{a_i\mu_i^m}{1-\mu_i x} \qquad (m = 0, \cdots, 4n). \tag{57}$$

There is a simple recursion formula which $D_m(x)$ defined in this manner satisfies. We have:

$$D_m(x) = \frac{1}{x}\sum_i a_i\mu_i^{m-1}\left(1 - \frac{1}{1+\mu_i x}\right), \tag{58}$$

or, using equations (42) and (43),

$$D_m(x) = \frac{1}{x}\left[\frac{2}{m}\epsilon_{m,\mathrm{odd}} - D_{m-1}(x)\right], \tag{59}$$

where

$$\epsilon_{m,\mathrm{odd}} = \begin{cases} 1 & \text{if } m \text{ is odd}, \\ 0 & \text{if } m \text{ is even}. \end{cases} \tag{60}$$

For odd, respectively, even values of m, equation (59) takes the form

$$D_{2j-1}(x) = \frac{1}{x}\left[\frac{2}{2j-1} - D_{2j-2}(x)\right], \tag{61}$$

and

$$D_{2j}(x) = -\frac{1}{x}D_{2j-1}(x). \tag{62}$$

Combining these relations, we have

$$D_{2j-1}(x) = \frac{1}{x}\left[\frac{2}{2j-1} + \frac{1}{x}D_{2j-3}(x)\right] = -xD_{2j}(x). \tag{63}$$

From this formula, we readily deduce that

$$\begin{aligned} D_{2j-1}(x) = {} & \frac{2}{(2j-1)x} + \frac{2}{(2j-3)x^3} + \cdots \\ & + \frac{2}{3x^{2j-3}} + \frac{2}{3x^{2j-1}}\,[2 - D_0(x)], \end{aligned} \tag{64}$$

and

$$D_{2j}(x) = -\frac{2}{(2j-1)x^2} - \frac{2}{(2j-3)x^4} - \cdots \tag{65}$$
$$-\frac{2}{3x^{2j-2}} - \frac{2}{3x^{2j}}[2 - D_0(x)].$$

If we now let x be a root k of the characteristic equation (48),

$$D_0(k) = 2, \tag{66}$$

and we find from equations (64) and (65) that

$$D_1(k) = D_2(k) = 0, \tag{67}$$

$$D_{2j}(k) = -\frac{2}{(2j-1)k^2} - \frac{2}{(2j-3)k^4} - \cdots - \frac{2}{3k^{2j-2}} \qquad (j = 2, \cdots, n), \tag{68}$$

$$D_{2j-1}(k) = \frac{2}{(2j-1)k} + \frac{2}{(2j-3)k^3} + \cdots + \frac{2}{3k^{2j-3}} \qquad (j = 2, \cdots, n). \tag{69}$$

17. A relation between the characteristic roots and the zeros of the Legendre polynomial. In terms of the $D_{2j}(k)$'s introduced in §16, we can express the characteristic equation for k in a form which does not explicitly involve the Gaussian weights and divisions: Let p_{2j} be the coefficient of μ^{2j} in the polynomial representation of the Legendre polynomial $P_{2n}(\mu)$, so that

$$P_{2n}(\mu) = \sum_{j=0}^{n} p_{2j}\mu^{2j}. \tag{70}$$

Now consider

$$\sum_{j=0}^{n} p_{2j}D_{2j}(k) = \sum_i \frac{a_i}{1 + \mu_i k}\left(\sum_{j=0}^{n} p_{2j}\mu_i^{2j}\right). \tag{71}$$

Since the μ_i's are the zeros of $P_{2n}(\mu)$

$$\sum_{j=0}^{n} p_{2j}\mu_i^{2j} = 0 \tag{72}$$

and the characteristic equation can be expressed in the form

$$\sum_{j=0}^{n} p_{2j}D_{2j}(k) = 0, \tag{73}$$

where the D_{2j}'s are defined as in equation (68). Substituting in particular for D_{2n} and D_0, we have

$$\frac{2}{3}\frac{p_{2n}}{k^{2n-2}} + \cdots + 2p_0 = 0. \tag{74}$$

From this equation it follows that

$$\frac{1}{(k_1 \cdots k_{n-1})^2} = (-1)^{n-1}\frac{3p_0}{p_{2n}} = 3\mu_1^2 \cdots \mu_n^2, \tag{75}$$

or

$$k_1 \cdots k_{n-1}\mu_1 \cdots \mu_n = 1/3^{1/2}. \tag{76}$$

18. **The law of darkening.** Returning to the solution (54) of the transfer problem considered in §15, we evaluate the net flux πF according to the formula

$$F = 2\int_{-1}^{+1} I\mu d\mu. \tag{77}$$

In our present scheme of approximation, we can write

$$F = 2\sum a_i I_i \mu_i. \tag{78}$$

Evaluating this sum with the I_i's given by equation (54), we obtain

$$F = 2b\left\{\sum_{\alpha=1}^{n-1} L_\alpha e^{-k_\alpha \tau} D_1(k_\alpha) + \sum_i a_i\mu_i^2\right\}. \tag{79}$$

Using equations (43) and (67), we have

$$F = 4b/3. \tag{80}$$

We can now rewrite the solution (54) in the form

$$I_i = \frac{3}{4}F\left\{\sum_{\alpha=1}^{n-1}\frac{L_\alpha e^{-k_\alpha \tau}}{1+\mu_i k_\alpha} + \tau + \mu_i + Q\right\} \quad (i = \pm 1, \cdots, \pm n). \tag{81}$$

In terms of this solution, the source function for the problem under consideration can be obtained in the following manner. We have

$$\begin{aligned} J &= \frac{1}{2}\int_{-1}^{+1} I d\mu \simeq \frac{1}{2}\sum a_j I_j \\ &= \frac{3}{8}F\left\{\sum_{\alpha=1}^{n-1} L_\alpha e^{-k_\alpha \tau} D_0(k_\alpha) + 2(\tau + Q)\right\}, \end{aligned} \tag{82}$$

or, using equation (66),

$$J = \frac{3}{4} F \left\{ \sum_{\alpha=1}^{n-1} L_\alpha e^{-k_\alpha \tau} + \tau + Q \right\}. \tag{83}$$

The angular distribution $I(0, \mu)$ of the emergent radiation can be found from the source function (83) in accordance with the formula (cf. equation (37))

$$I(0, \mu) = \int_0^\infty J(\tau) e^{-\tau/\mu} \frac{d\tau}{\mu}. \tag{84}$$

We find

$$I(0, \mu) = \frac{3}{4} F \left\{ \sum_{\alpha=1}^{n-1} \frac{L_\alpha}{1 + \mu k_\alpha} + \mu + Q \right\}. \tag{85}$$

It is to be particularly noted that the foregoing expression for $I(0, \mu)$ is in agreement with the solution (81) for $\tau = 0$ and at the points of the Gaussian division $\mu = \mu_i$.

Comparing equation (85) with equation (56) which determines the constants L_α and Q, we observe that the angular distribution of the emergent radiation defined for the interval $0 \leq \mu \leq 1$ is determined in terms of a function which has zeros assigned in the complementary interval $-1 < \mu < 0$. Thus, letting

$$S(\mu) = \sum_{\alpha=1}^{n-1} \frac{L_\alpha}{1 - \mu k_\alpha} - \mu + Q, \tag{86}$$

the boundary conditions require that

$$S(\mu_i) = 0 \qquad (i = 1, \cdots, n), \tag{87}$$

while the angular distribution of the emergent radiation is given by

$$I(0, \mu) = \frac{3}{4} F S(-\mu). \tag{88}$$

19. The elimination of the constants and the expression of the solution in closed form. We shall now show how an explicit formula for $S(\mu)$ can be found without solving explicitly for the constants L_α and Q:

Consider the function

$$\prod_{\alpha=1}^{n-1} (1 - \mu k_\alpha) S(\mu). \tag{89}$$

This is a polynomial of degree n in μ which vanishes for $\mu=\mu_i$, $i=1, \cdots, n$. Consequently there must exist a proportionality of the form

$$\prod_{\alpha=1}^{n-1}(1-k_\alpha\mu)S(\mu) \propto \prod_{i=1}^{n-1}(\mu-\mu_i). \tag{90}$$

The constant of proportionality can be found from a comparison of the coefficients of the highest powers of μ on either side. In this manner we find that

$$S(\mu) = (-1)^n k_1 \cdots k_{n-1} \frac{\prod_{i=1}^{n}(\mu-\mu_i)}{\prod_{\alpha=1}^{n-1}(1-k_\alpha\mu)}. \tag{91}$$

This is the required formula.

According to equation (91)

$$S(-\mu) = k_1 \cdots k_{n-1} \frac{\prod_{i=1}^{n}(\mu+\mu_i)}{\prod_{\alpha=1}^{n-1}(1+k_\alpha\mu)}, \tag{92}$$

or, using the result (76), we can write

$$S(-\mu) = \frac{1}{3^{1/2}} H(\mu), \tag{93}$$

where

$$H(\mu) = \frac{1}{\mu_1 \cdots \mu_n} \frac{\prod_{i=1}^{n}(\mu+\mu_i)}{\prod_{\alpha=1}^{n-1}(1+k_\alpha\mu)}. \tag{94}$$

In terms of the function $H(\mu)$ defined in this manner, the angular distribution of the emergent radiation can be expressed in the form

$$I(0, \mu) = \frac{3^{1/2}}{4} F H(\mu). \tag{95}$$

20. The solution of equation (III) by the new method. A particular integral. As a further illustration of our method of solving equations of transfer, we shall next consider equation III appropriate for the problem of diffuse reflection by an isotropically scattering atmosphere. The equivalent system of linear equations in the nth approximation is

$$\mu_i \frac{dI_i}{d\tau} = I_i - \frac{1}{2}\sum a_j I_j - \frac{1}{4} F e^{-\tau/\mu_0} \qquad (i = \pm 1, \cdots, \pm n). \tag{96}$$

It is seen that the homogeneous system associated with equation (96) is the same as equation (44). Accordingly, the complimentary function for the solution of equation (96) is the same as the general solution (53) of the homogeneous system. To complete the solution we require only a particular integral. This can be found in the following manner: Setting

$$I_i = \frac{1}{4} F h_i e^{-\tau/\mu_0} \qquad (i = \pm 1, \cdots, \pm n) \tag{97}$$

in equation (96) (the h_i's are constants unspecified for the present) we verify that we must have

$$h_i(1 + \mu_i/\mu_0) = \frac{1}{2} \sum a_j h_j + 1. \tag{98}$$

Equation (98) implies that the constants h_i must be expressible in the form

$$h_i = \frac{\gamma}{1 + \mu_i/\mu_0} \qquad (i = \pm 1, \cdots, \pm n), \tag{99}$$

where the constant γ has to be determined from the condition (cf. equation (98))

$$\gamma = \frac{1}{2}\gamma \sum \frac{a_{i'}}{1 + \mu_i/\mu_0} + 1. \tag{100}$$

In other words

$$I_i = \frac{1}{4} F \frac{\gamma e^{-\tau/\mu_0}}{1 + \mu_i/\mu_0} \qquad (i = \pm 1, \cdots, \pm n), \tag{101}$$

with

$$\gamma = \frac{1}{1 - \sum_{j=1}^{n} \dfrac{a_j}{1 - \mu_j^2/\mu_0^2}}, \tag{102}$$

represents the required particular integral of equation (96).

The constant γ defined as in equation (102) can be expressed in terms of the H-function introduced in §19 (equation (94)). For this purpose consider the function

$$T(x) = 1 - \sum_{j=1}^{n} \frac{a_j}{1 - \mu_j^2 x}. \tag{103}$$

This clearly vanishes for

$$x = 0 \quad \text{and} \quad x = k_\alpha^2 \qquad (\alpha = 1, \cdots, n-1). \tag{104}$$

Accordingly

$$T(x) \prod_{j=1}^{n} (1 - \mu_j^2 x) \tag{105}$$

cannot differ from

$$x \prod_{\alpha=1}^{n-1} (x - k_\alpha^2) \tag{106}$$

by more than a constant factor since (105) represents a polynomial of degree n in x. The constant of proportionality can be determined by comparing the coefficients of the highest powers. In this manner we find that

$$T(x) = (-1)^n \mu_1^2 \cdots \mu_n^2 \frac{x \prod_{\alpha=1}^{n-1} (x - k_\alpha^2)}{\prod_{j=1}^{n} (1 - \mu_j^2 x)}. \tag{107}$$

But

$$\gamma = \frac{1}{T(\mu_0^{-2})}. \tag{108}$$

Hence

$$\gamma = (-1)^n \frac{1}{\mu_1^2 \cdots \mu_n^2} \frac{\prod_{i=1}^{n} (\mu_0^2 - \mu_i^2)}{\prod_{\alpha=1}^{n-1} (1 - k_\alpha^2 \mu_0^2)} \tag{109}$$

or according to our definition of $H(\mu)$

$$\gamma = H(\mu_0)H(-\mu_0). \tag{110}$$

21. **The law of diffuse reflection for the case of isotropic scattering.** Now, adding to the particular integral (101) the general solution of the homogeneous system compatible with our present requirement of boundedness of the solution for $\tau \to \infty$, we have

$$I_i = \frac{1}{4} F \left\{ \sum_{\alpha=1}^{n-1} \frac{L_\alpha e^{-k_\alpha \tau}}{1 + \mu_i k_\alpha} + Q + \frac{H(\mu_0)H(-\mu_0)}{1 + \mu_i/\mu_0} e^{-\tau/\mu_0} \right\} \tag{111}$$

$$(i = 1, \cdots, n),$$

where the constants L_α $(\alpha = 1, \cdots, n-1)$ and Q are to be determined by the boundary conditions at $\tau = 0$.

At $\tau = 0$ we have no diffuse radiation directed inward. The conditions which determine the constants of integration are therefore

$$\sum_{\alpha=1}^{n-1} \frac{L_\alpha}{1 - \mu_i k_\alpha} + Q + \frac{H(\mu_0)H(-\mu_0)}{1 - \mu_i/\mu_0} = 0 \quad (i = \pm 1, \cdots, \pm n). \tag{112}$$

The angular distribution of the diffusely reflected radiation can be found in terms of the source function

$$\mathfrak{I}(\tau) = \frac{1}{2} \sum a_i I_i + \frac{1}{4} F e^{-\tau/\mu_0}, \tag{113}$$

according to the formula

$$I(0, \mu) = \int_0^\infty \mathfrak{I}(\tau) e^{-\tau/\mu} \frac{d\tau}{\mu}. \tag{114}$$

We find

$$I(0, \mu; \mu_0) = \frac{1}{4} F \left\{ \sum_{\alpha=1}^{n-1} \frac{L_\alpha}{1 + \mu k_\alpha} + Q + \frac{H(\mu_0)H(-\mu_0)}{1 + \mu/\mu_0} \right\}. \tag{115}$$

As in §18, we again observe that the angular distribution of the diffusely reflected radiation defined in the interval $0 \leq \mu \leq 1$ is described in terms of a function which has zeros assigned in the interval $0 > \mu > -1$. Thus, letting

$$S(\mu) = \sum_{\alpha=1}^{n-1} \frac{L_\alpha}{1 - \mu k_\alpha} + Q + \frac{H(\mu_0)H(-\mu_0)}{1 - \mu/\mu_0}, \tag{116}$$

we have

$$S(\mu_i) = 0 \qquad (i = 1, \cdots, n), \tag{117}$$

and

$$I(0, \mu; \mu_0) = \frac{1}{4} FS(-\mu). \tag{118}$$

And, again as in the earlier problem, we can find an explicit formula for $S(\mu)$ without having to solve explicitly for the constants L_α and Q.

Thus, considering the function

$$(1 - \mu/\mu_0) \prod_{\alpha=1}^{n-1} (1 - k_\alpha \mu) S(\mu) \tag{119}$$

we observe that it is a polynomial of degree n in μ which vanishes for $\mu = \mu_i$, $i = 1, \cdots, n$. There must, therefore, exist a relation of the form

$$S(\mu) = (-1)^n \frac{X}{\mu_1 \cdots \mu_n} \frac{\prod_{i=1}^{n} (\mu - \mu_i)}{\prod_{\alpha=1}^{n-1} (1 - k_\alpha \mu)} \frac{1}{(1 - \mu/\mu_0)}, \tag{120}$$

where X is a constant. In terms of $H(\mu)$ we can rewrite the foregoing equation as

$$S(\mu) = X \frac{H(-\mu)}{1 - \mu/\mu_0}. \tag{121}$$

The constant X appearing in equation (121) can be determined from the relation (cf. equation (116))

$$\operatorname*{limit}_{\mu \to \mu_0} (1 - \mu/\mu_0) S(\mu) = H(\mu_0) H(-\mu_0). \tag{122}$$

According to equation (121) the left-hand side of equation (122) is $XH(-\mu_0)$. Hence

$$X = H(\mu_0) \tag{123}$$

and

$$S(\mu) = H(\mu_0) H(-\mu) \frac{\mu_0}{\mu_0 - \mu}. \tag{124}$$

The expression (118) for the angular distribution of the reflected radiation therefore becomes

$$I(0, \mu; \mu_0) = \frac{1}{4} FH(\mu) H(\mu_0) \frac{\mu_0}{\mu + \mu_0}. \tag{125}$$

22. **Tabulation of the solutions of equations (I)–(VII).** The method of solution of transfer problems which we have illustrated in the preceding sections by considering the two simplest problems is, on examination, found to be sufficiently general for adaptation to the solution of the more difficult problems presented by the other equations of transfer formulated in §§8–11. While the details of the solution of these other equations (particularly equations VI and VII) are considerably more elaborate and complex, the analysis nevertheless shows similarities with the simple problems we have considered in the broad features. Thus, it is found that, in all cases, the angular distribution of the emergent (equivalently, reflected) radiation is described by a function for the argument in the range (0, 1) while the boundary conditions specify the zeros of the same function in the interval (0, −1). This remarkable *reciprocity*, which exists in all the problems, enables the *elimination of the constants* and allows the reduction of the solutions to *closed forms* in the general nth approximation. And finally, these solutions, apart from certain constants, involve only H-functions of the form

$$H(\mu) = \frac{1}{\mu_1 \cdots \mu_n} \frac{\prod_{i=1}^{n} (\mu + \mu_i)}{\prod_{\alpha=1}^{n} (1 + k_\alpha \mu)}, \tag{126}$$

where the μ_i's are the zeros of the Legendre polynomial $P_{2n}(\mu)$ and the k_α's are the positive (or zero) roots of a characteristic equation of the form

$$1 = 2\sum_{j=1}^{n} \frac{a_j \Psi(\mu_j)}{1 - k^2 \mu_j^2}, \tag{127}$$

where $\Psi(\mu)$ is an even polynomial in μ satisfying the condition

$$\int_0^1 \Psi(\mu) d\mu \leq \frac{1}{2} \cdot \tag{128}$$

(This last condition is necessary for $H(\mu)$ to be real.)

The different physical problems naturally lead to different characteristic equations and therefore to different H-functions. However, as the H-functions differ from one another only through the characteristic equations which define the roots k_α, we may properly call $\Psi(\mu)$ the *characteristic function* in terms of which H is defined.

We tabulate below the solutions for the various transfer problems (equations I–VII) obtained in the nth approximation of our method of solutions.

Solutions for the emergent and the diffusely reflected radiation for various transfer problems

A. *Isotropic scattering*

Problem with constant net flux: the law of darkening:

$$I(\mu) = \frac{3^{1/2}}{4} F H(\mu).$$

Law of diffuse reflection:

$$I(\mu; \mu_0) = \frac{1}{4} F H(\mu) H(\mu_0) \frac{\mu_0}{\mu + \mu_0}.$$

The characteristic function in terms of which $H(\mu)$ is defined is

$$\Psi(\mu) = 1/2.$$

B. *Scattering in accordance with Rayleigh's phase function* $3(1+\cos^2\Theta)/4$

Problem with constant net flux: the law of darkening:

$$I(\mu) = \frac{3}{4} q F H^{(0)}(\mu).$$

Law of diffuse reflection:

$$\begin{aligned} I(\mu, \varphi; \mu_0, \varphi_0) = \frac{3}{32} F \{ & H^{(0)}(\mu) H^{(0)}(\mu_0) [3 - (3 - 8q^2)^{1/2}(\mu + \mu_0) + \mu\mu_0] \\ & - 4\mu\mu_0 (1 - \mu^2)^{1/2} (1 - \mu_0^2)^{1/2} H^{(1)}(\mu) H^{(1)}(\mu_0) \cos(\varphi - \varphi_0) \\ & + (1 - \mu^2)(1 - \mu_0^2) H^{(2)}(\mu) H^{(2)}(\mu_0) \cos 2(\varphi - \varphi_0) \} \frac{\mu_0}{\mu + \mu_0}. \end{aligned}$$

The characteristic functions in terms of which $H^{(0)}(\mu)$, $H^{(1)}(\mu)$ and $H^{(2)}(\mu)$ are defined are respectively

$$\Psi^{(0)}(\mu) = \frac{3}{16}(3 - \mu^2),$$

$$\Psi^{(1)}(\mu) = \frac{3}{8}\mu^2(1 - \mu^2),$$

$$\Psi^{(2)}(\mu) = \frac{3}{32}(1 - \mu^2)^2.$$

And finally,

$$q = \frac{2(3^{1/2})}{H^{(0)}(+3^{1/2}) - H^{(0)}(-3^{1/2})}.$$

C. *Scattering in accordance with the phase function* $\lambda(1+x\cos\Theta)$

Law of diffuse reflection:

$$I(\mu, \varphi; \mu_0, \varphi_0) = \frac{1}{4}\lambda F\{H^{(0)}(\mu)H^{(0)}(\mu_0)[1 - c(\mu+\mu_0) - x(1-\lambda)\mu\mu_0]$$
$$+ x(1-\mu^2)^{1/2}(1-\mu_0^2)^{1/2}H^{(1)}(\mu)H^{(1)}(\mu_0)\cos(\varphi-\varphi_0)\}\frac{\mu_0}{\mu+\mu_0}.$$

The characteristic functions in terms of which $H^{(0)}(\mu)$ and $H^{(1)}(\mu)$ are defined are respectively

$$\Psi^{(0)}(\mu) = \frac{1}{2}\lambda[1 + x(1-\lambda)\mu^2],$$

$$\Psi^{(1)}(\mu) = \frac{1}{4}x\lambda(1-\mu^2).$$

The constant c depends in a somewhat complicated manner on the characteristic roots defining $H^{(0)}(\mu)$ (cf. S. Chandrasekhar, Astrophysical Journal vol. 103 (1946) p. 165, equation [108]).

D. *Scattering in accordancé with Rayleigh's law and allowing for the polarization of the scattered radiation*

Problem with constant net flux: the law of darkening in the two states of polarization:

$$I_l(\mu) = \frac{3}{8}F(1-c^2)^{1/2}H_l(\mu),$$

$$I_r(\mu) = \frac{3}{8}F\frac{1}{2^{1/2}}H_r(\mu)(\mu+c).$$

Problem of diffuse reflection: the scattering matrix:

$$I(\mu, \varphi; \mu_0, \varphi_0) = \frac{3}{16\mu}QS(\mu, \varphi; \mu_0, \varphi_0)F$$

where

$$\left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)S(\mu,\varphi;\mu_0,\varphi_0)$$

$$=\left[\begin{array}{lll}
\begin{array}{l}2H_l(\mu)H_l(\mu_0)[1+\mu\mu_0-c(\mu+\mu_0)]\\ -4\mu\mu_0(1-\mu^2)^{1/2}(1-\mu_0^2)^{1/2}H^{(1)}(\mu)H^{(1)}(\mu_0)\cos(\varphi-\varphi_0)\\ +\mu^2\mu_0^2H^{(2)}(\mu)H^{(2)}(\mu_0)\cos 2(\varphi-\varphi_0)\end{array} &
\begin{array}{l}(2(1-c^2))^{1/2}H_r(\mu_0)H_l(\mu)(\mu+\mu_0)\\ \\ -\mu^2H^{(2)}(\mu)H^{(2)}(\mu_0)\cos 2(\varphi-\varphi_0)\end{array} &
\begin{array}{l} \\ -2\mu(1-\mu^2)^{1/2}(1-\mu_0^2)^{1/2}H^{(1)}(\mu)H^{(1)}(\mu_0)\sin(\varphi-\varphi_0)\\ +\mu^2\mu_0H^{(2)}(\mu)H^{(2)}(\mu_0)\sin 2(\varphi-\varphi_0)\end{array} \\
\\
\begin{array}{l}(2(1-c^2))^{1/2}H_r(\mu)H_l(\mu_0)(\mu+\mu_0)\\ -\mu_0^2H^{(2)}(\mu)H^{(2)}(\mu_0)\cos 2(\varphi-\varphi_0)\end{array} &
\begin{array}{l}H_r(\mu)H_r(\mu_0)[1+\mu\mu_0+c(\mu+\mu_0)]\\ +H^{(2)}(\mu)H^{(2)}(\mu_0)\cos 2(\varphi-\varphi_0)\end{array} &
\begin{array}{l} \\ -\mu_0H^{(2)}(\mu)H^{(2)}(\mu_0)\sin 2(\varphi-\varphi_0)\end{array} \\
\\
\begin{array}{l}-2\mu_0(1-\mu_0^2)^{1/2}(1-\mu^2)^{1/2}H^{(1)}(\mu)H^{(1)}(\mu_0)\sin(\varphi-\varphi_0)\\ +\mu\mu_0^2H^{(2)}(\mu)H^{(2)}(\mu_0)\sin 2(\varphi-\varphi_0)\end{array} &
\begin{array}{l} \\ -\mu H^{(2)}(\mu)H^{(2)}(\mu_0)\sin 2(\varphi-\varphi_0)\end{array} &
\begin{array}{l}(1-\mu^2)^{1/2}(1-\mu_0^2)^{1/2}H^{(1)}(\mu)H^{(1)}(\mu_0)\cos(\varphi-\varphi_0)\\ -\mu\mu_0H^{(2)}(\mu)H^{(2)}(\mu_0)\cos 2(\varphi-\varphi_0)\end{array}
\end{array}\right].$$

The characteristic functions in terms of which $H_l(\mu)$, $H_r(\mu)$, $H^{(1)}(\mu)$ and $H^{(2)}(\mu)$ are defined are respectively

$$\Psi_l(\mu) = \frac{3}{4}(1-\mu^2);$$

$$\Psi^{(1)}(\mu) = \frac{3}{8}(1-\mu^2)(1+2\mu^2);$$

$$\Psi_r(\mu) = \frac{3}{8}(1-\mu^2),$$

$$\Psi^{(2)}(\mu) = \frac{3}{16}(1+\mu^2)^2.$$

And finally, the constant c is given by

$$c = \frac{H_l(+1)H_r(-1) + H_l(-1)H_r(+1)}{H_l(+1)H_r(-1) - H_l(-1)H_r(+1)}.$$

23. **Remarks on the tabulated solutions. The possibility of passage to the "infinite approximation."** An examination of the solutions of the various transfer problems given in the preceding section discloses remarkable relationships between the *laws of darkening* of the emergent radiation from a semi-infinite plane-parallel atmosphere with constant net flux (and no incident radiation) and the *laws of diffuse reflection* by the same atmosphere. The relationship is naturally the simplest for the case of an isotropically scattering atmosphere and can indeed be established directly from the equations of transfer. However, in the other cases, the relationship is of a more complex nature and has to be sought between the darkening function for the problem with a constant net flux and the *azimuth independent* term in the law of diffuse reflection. It is seen that both these functions involve the same H-functions and the same constants.

While the relationship between the two problems has been established only in a particular scheme of approximation, it is apparent that the *relationship itself* must be an *exact one* since it is present in every approximation and must consequently be also present in the *limit of infinite approximation* when the solutions will become the exact ones of the problem. A further result of this train of thought is the realization that if we can solve the problem of diffuse reflection exactly, we shall, at the same time, have also solved exactly the axially symmetric problem with a constant net flux and no incident radiation.

Having thus been led to conceive the passage to the limit of infinite approximation, we naturally ask ourselves: *Can we in fact perform this limiting process and thus obtain the exact solutions for the various problems?* The answer to this question must clearly depend on our ability to pass to the limit of the H-functions as we have defined them, as $n \to \infty$. We shall now indicate how this limiting process can be achieved in practice.

24. The equation satisfied by $H(\mu)$. For the purposes of passing to the limit of infinite approximations of the solutions of the various transfer problems, we shall first establish the following basic theorem relating to the H-functions.

THEOREM 1. *Let $\Psi(\mu)$ be an even polynomial of degree $2m$ in μ such that*

$$\int_0^1 \Psi(\mu)d\mu \leqq \frac{1}{2}.$$

Let μ_j ($j = \pm 1, \cdots, \pm n$) denote the division of the interval $(-1, +1)$ according to the zeros of the Legendre polynomial of order $2n$ ($>m$); further, let a_j ($=a_{-j}$) denote the corresponding Gaussian weights. Finally, let k_α ($\alpha = 1, \cdots, n$) denote the distinct positive (or zero) roots of the characteristic equation

$$1 = \sum_j \frac{a_j\Psi(\mu_j)}{1 + k\mu_j} = 2\sum_{j=1}^{n} \frac{a_j\Psi(\mu_j)}{1 - k^2\mu_j^2}.$$

Then the function

$$H(\mu) = \frac{1}{\mu_1 \cdots \mu_n} \frac{\prod_{i=1}^{n} (\mu + \mu_i)}{\prod_{\alpha=1}^{n} (1 + k_\alpha\mu)}$$

satisfies identically the equation

$$H(\mu) = 1 + \mu H(\mu) \sum_{j=1}^{n} \frac{a_j H(\mu_j)\Psi(\mu_j)}{\mu + \mu_j}.$$

PROOF. We shall first consider the case

$$\int_0^1 \Psi(\mu)d\mu < \frac{1}{2}. \tag{129}$$

In this case the characteristic equation admits n distinct nonvanishing positive roots and we consider the function

$$S(\mu) = \sum_{\alpha=1}^{n} \frac{L_\alpha}{1 - k_\alpha \mu} + 1, \tag{130}$$

where L_α $(\alpha = 1, \cdots, n)$ are certain constants to be determined from the equations

$$\sum_{\alpha=1}^{n} \frac{L_\alpha}{1 - k_\alpha \mu_i} + 1 = 0 \qquad (i = 1, \cdots, n), \tag{131}$$

or equivalently

$$S(\mu_i) = 0 \qquad (i = 1, \cdots, n). \tag{132}$$

By considerations of the type we are now familiar with, it can be shown that

$$S(\mu) = k_1 \cdots k_n \mu_1 \cdots \mu_n \frac{(-1)^n}{\mu_1 \cdots \mu_n} \frac{\prod_i (\mu - \mu_i)}{\prod_\alpha (1 - k_\alpha \mu)}. \tag{133}$$

In other words,

$$S(\mu) = k_1 \cdots k_n \mu_1 \cdots \mu_n H(-\mu). \tag{134}$$

Since $H(0) = 1$, we can rewrite equation (134) alternatively in the form

$$S(\mu) = S(0) H(-\mu), \tag{135}$$

where

$$S(0) = \sum_{\alpha=1}^{n} L_\alpha + 1 = k_1 \cdots k_n \mu_1 \cdots \mu_n. \tag{136}$$

Now, since $\Psi(\mu)$ is even in μ, we can rewrite the equation which a characteristic root satisfies in either of the forms

$$1 = \sum_j \frac{a_j \Psi(\mu_j)}{1 + k\mu_j} \tag{137}$$

or

$$1 = \sum_j \frac{a_j \Psi(\mu_j)}{1 - k\mu_j}. \tag{138}$$

Let k_α denote a particular characteristic root. Then, on account of equations (137) and (138), which are satisfied by any of the characteristic roots, we can clearly write

$$
\begin{aligned}
S(0) &= \sum_{\beta=1}^{n} L_\beta + 1 \\
(139) \qquad &= \sum_{\beta=1}^{n} \frac{L_\beta}{k_\alpha + k_\beta}\left[\sum_j a_j\Psi(\mu_j)\left\{\frac{k_\alpha}{1+k_\alpha\mu_j} + \frac{k_\beta}{1-k_\beta\mu_j}\right\}\right] \\
&\quad + \sum_j \frac{a_j\Psi(\mu_j)}{1+k_\alpha\mu_j}.
\end{aligned}
$$

Simplifying the quantity in brackets in the foregoing equation, we have

$$
(140) \qquad S(0) = \sum_{\beta=1}^{n} L_\beta\left[\sum_j \frac{a_j\Psi(\mu_j)}{(1+k_\alpha\mu_j)(1-k_\beta\mu_j)}\right] + \sum_j \frac{a_j\Psi(\mu_j)}{1+k_\alpha\mu_j}
$$

or, inverting the order of the summation,

$$
(141) \qquad S(0) = \sum \frac{a_j\Psi(\mu_j)}{1+k_\alpha\mu_j}\left[\sum_{\beta=1}^{n} \frac{L_\beta}{1-k_\beta\mu_j} + 1\right].
$$

But the quantity in brackets in equation (141) is $S(\mu_j)$. Hence

$$
(142) \qquad S(0) = \sum_j \frac{a_j S(\mu_j)\Psi(\mu_j)}{1+k_\alpha\mu_j}.
$$

In equation (142) (as in equations (137)–(141)) the summation is, of course, extended over all values of j, positive and negative. However, since $S(+\mu_j) = 0$ (equation (132)), in equation (142) only the terms with negative j make a nonzero contribution. We can, therefore, write

$$
(143) \qquad S(0) = \sum_{j=1}^{n} \frac{a_j S(-\mu_j)\Psi(\mu_j)}{1-k_\alpha\mu_j} \qquad (\alpha = 1, \cdots, n)
$$

or, in view of equation (135),

$$
(144) \qquad 1 = \sum_{j=1}^{n} \frac{a_j H(\mu_j)\Psi(\mu_j)}{1-k_\alpha\mu_j} \qquad (\alpha = 1, \cdots, n).
$$

Now, consider the function

$$
(145) \qquad 1 - \mu\sum_{j=1}^{n} \frac{a_j H(\mu_j)\Psi(\mu_j)}{\mu+\mu_j}.
$$

According to equation (144), this vanishes for $\mu=-1/k_\alpha$ $(\alpha=1, \cdots, n)$; for

$$(146) \qquad 1+\frac{1}{k_\alpha}\sum_{j=1}^{n}\frac{a_jH(\mu_j)\Psi(\mu_j)}{(-1/k_\alpha)+\mu_j}=1-\sum_{j=1}^{n}\frac{a_jH(\mu_j)\Psi(\mu_j)}{1-k_\alpha\mu_j}=0.$$

Hence

$$(147) \qquad \prod_{j=1}^{n}(\mu+\mu_j)-\mu\sum_{j=1}^{n}a_jH(\mu_j)\Psi(\mu_j)\prod_{i\neq j}(\mu+\mu_i)$$

also vanishes for

$$(148) \qquad \mu=-1/k_\alpha \qquad (\alpha=1,\cdots,n).$$

But the expression (147) is a polynomial of degree n in μ. It cannot, therefore, differ from

$$(149) \qquad \prod_{\alpha=1}^{n}(1+k_\alpha\mu)$$

by more than a constant factor; and the constant of proportionality is seen to be

$$(150) \qquad \mu_1\cdots\mu_n$$

from a comparison of the two functions at $\mu=0$. It therefore follows that

$$(151) \qquad 1-\mu\sum_{j=1}^{n}\frac{a_jH(\mu_j)\Psi(\mu_j)}{\mu+\mu_j}\equiv\mu_1\cdots\mu_n\frac{\prod_{\alpha=1}^{n}(1+k_\alpha\mu)}{\prod_{j=1}^{n}(\mu+\mu_j)}=\frac{1}{H(\mu)}.$$

Hence

$$(152) \qquad H(\mu)\equiv 1+\mu H(\mu)\sum_{j=1}^{n}\frac{a_jH(\mu_j)\Psi(\mu_j)}{\mu+\mu_j}.$$

This proves the theorem for the case (129).

Turning now to the case

$$(153) \qquad \int_0^1\Psi(\mu)d\mu=\frac{1}{2},$$

we observe that, in this case, $k=0$ is a root of the characteristic equation, and we are left with only $(n-1)$ positive roots. We, therefore, consider in this case the function

$$S(\mu) = \sum_{\alpha=1}^{n-1} \frac{L_\alpha}{1 - k_\alpha \mu} + L_0 - \tag{154}$$

in place of equation (130). However, the constants L_0 and L_α ($\alpha = 1, \cdots, n-1$) are again to be determined by the conditions

$$S(\mu_i) = 0 \qquad (i = 1, \cdots, n). \tag{155}$$

With this definition of $S(\mu)$, equation (135) continues to be valid and the rest of the proof follows on similar lines. The only essential point of departure that needs to be noted is that, at the stage of the proof corresponding to equation (140) and before inverting the order of the summation, we must add the extra term

$$-\sum_j \frac{a_j \Psi(\mu_j)\mu_j}{1 + k_\alpha \mu_j} \tag{156}$$

to the right-hand side of the equation. We can do this without altering anything, since the quantity we thus add is zero; for

$$\begin{aligned} \sum_j \frac{a_j \Psi(\mu_j)\mu_j}{1 + k_\alpha \mu_j} &= \frac{1}{k_\alpha}\left[\sum_j a_j \Psi(\mu_j) \left\{1 - \frac{1}{1 + k_\alpha \mu_j}\right\}\right] \\ &= \frac{1}{k_\alpha}\left[1 - \sum_j \frac{a_j \Psi(\mu_j)}{1 + k_\alpha \mu_j}\right] = 0. \end{aligned} \tag{157}$$

(Note that we are permitted to set $\sum a_j \Psi(\mu_j) = 1$, since the Gauss sum in the nth approximation evaluates the integrals exactly for all polynomials of degree less than or equal to $4n-1$; and we have assumed $2n > m$.)

This completes the proof of the theorem.

25. The limit of $H(\mu)$ as $n \to \infty$. The basic functional equation. The theorem proved in §24 suggests how the limit of the H-function as $n \to \infty$ can be obtained. We shall state this in the form of a theorem.

THEOREM 2. *The solution of the functional equation*

$$H(\mu) = 1 + \mu H(\mu) \int_0^1 \frac{H(\mu')\Psi(\mu')}{\mu + \mu'} d\mu',$$

where $\Psi(\mu)$ is an even polynomial satisfying the condition

$$\int_0^1 \Psi(\mu) d\mu \leq \frac{1}{2},$$

is the limit function

$$\lim_{n\to\infty} \frac{1}{\mu_1 \cdots \mu_n} \frac{\prod_{i=1}^{n} (\mu + \mu_i)}{\prod_{\alpha=1}^{n} (1 + k_\alpha \mu)}$$

where the μ_i's and the k_α's have the same meanings as in Theorem 1.

SKETCH OF PROOF. The theorem arises in the following way: It is known that the integral of a bounded function over the interval (0, 1) can be approximated by a Gauss sum with any desired degree of accuracy by choosing a division of the interval according to the zeros of a Legendre polynomial of a sufficiently high degree. The integral which occurs on the right-hand side of the functional equation for $H(\mu)$ can, therefore, be replaced by the Gauss sum

$$\sum_{j=1}^{n} \frac{a_j H(\mu_j) \Psi(\mu_j)}{\mu + \mu_j} \tag{158}$$

to any desired accuracy by choosing a sufficiently large n. But by Theorem 1, for a finite n, no matter how large, the unique solution of the equation

$$H(\mu) = 1 + \mu H(\mu) \sum_{j=1}^{n} \frac{a_j H(\mu_j) \Psi(\mu_j)}{\mu + \mu_j} \tag{159}$$

is

$$\frac{1}{\mu_1 \cdots \mu_n} \frac{\prod_{i=1}^{n} (\mu + \mu_i)}{\prod_{\alpha=1}^{n} (1 + k_\alpha \mu)}. \tag{160}$$

If we now let $n \to \infty$, equation (159) becomes

$$H(\mu) = 1 + \mu H(\mu) \int_0^1 \frac{H(\mu')\Psi(\mu')}{\mu + \mu'} d\mu'. \tag{161}$$

The solution of this functional equation is, therefore, seen to be the limit of the function (160) as $n \to \infty$.

It is realized that the sketch of the proof of Theorem 2 we have just outlined does not meet the full demands of a rigorous mathematical demonstration. It is, moreover, probable that precisely the questions of uniqueness and existence which we have ignored will cause the principal difficulties in the constructions of a rigorous mathematical proof. However, as the equation arises in a physical context, and the physical situations are such as to leave no room for ambiguity, it is hardly to be doubted that the theorem is true. Indeed, as we shall presently show quite rigorously by an entirely different line of argument, the exact solutions for the various transfer problems are of the forms tabulated in §22 with the H-functions redefined in terms of functional equations of the form (161) instead of in terms of the Gaussian division and characteristic roots. Nevertheless, it may be of interest to pursue further the purely mathematical questions raised by Theorem 2.

26. A practical method of determining the exact H-functions as solutions of the functional equations they satisfy. Assuming for the present the indications of the preceding sections that the exact solutions for the various transfer problems are of the *forms* found in our method of solution and that the H-functions which occur in them have to be redefined as solutions of functional equations of the form (161), we may observe that Theorem 1 of §24 suggests a simple practical method for determining the exact H-functions numerically. For, starting with an approximate solution for $H(\mu)$ (in the third approximation, for example) we can determine the exact H-functions by a process of iteration using for this purpose the functional equation which it satisfies. In this manner the exact H-functions which occur in the solutions of the various transfer problems involving isotropic scattering with an albedo $\lambda \leq 1$, Rayleigh phase function and Rayleigh scattering (including the state of polarization of the scattered radiation) have all been numerically evaluated. We have therefore now available exact numerical solutions for the cases A, B, C (for $x=0$) and D tabulated in §22.

27. The constants in the solution. It is to be noted that the determination of the limit to which the H-functions which occur in the solutions for the emergent (or diffusely reflected) radiation in the nth approximation tend, as $n \to \infty$, still leaves open the question of the exact limiting values of the constants which occur in these solutions. It does not seem that any direct or simple limiting process can be applied to the formulae which define them in the nth approximation. Attention may be particularly drawn in this connection to the fact that the formulae which define these constants (in the nth approximation) often involve the values of H-functions (now defined as rational functions in terms of the Gaussian division and the characteristic roots) *outside* the interval (0, 1) (sometimes even for complex values of the argument!), whereas it would seem that in the limit of infinite approximation $H(\mu)$ has a meaning only in the interval (0, 1). (See, for example, the definitions of the constants c and q under the headings B and D in the tabulation of §22.) However, it appears that these constants can be determined *indirectly* by appealing to certain other identical relations which the problems must satisfy.

Thus, considering the transfer problems involving the Rayleigh phase function, we have

$$I(\mu) = \frac{3}{4} qFH^{(0)}(\mu), \tag{162}$$

where $H^{(0)}(\mu)$ is defined as the solution of the functional equation

$$H^{(0)}(\mu) = 1 + \frac{3}{16}\mu H^{(0)}(\mu)\int_0^1 \frac{(3-\mu'^2)H(\mu')}{\mu+\mu'}d\mu'. \tag{163}$$

Now, as equation (162) gives the angular distribution of the emergent radiation for the axially symmetric problem with constant net flux, it follows that the outward flux of the emergent radiation must also equal πF. In other words, we must have

$$F = 2\int_0^1 I(\mu)\mu d\mu = \frac{3}{2} qF\alpha_1, \tag{164}$$

where α_1 denotes the first moment of $H^{(0)}(\mu)$. Hence,

$$q = \frac{2}{3\alpha_1}. \tag{165}$$

Thus, once the solution of equation (163) has been determined (by iteration based on an approximate $H(\mu)$ as suggested in §26) the constant q can be determined directly in terms of its first moment. Since q is the only constant which occurs in the solutions, they become determinate in this fashion.

Similarly, in the transfer problems in which the polarization of the scattered radiation in accordance with Rayleigh's law has to be properly allowed for, the solutions again involve a constant c. It does not seem possible to pass directly to the limit of infinite approximation in the formula defining this constant in the nth approximation. But we can determine it by appealing to the flux condition in the problem with the constant net flux. Thus, with the solutions $I_l(\mu)$ and $I_r(\mu)$ for the emergent radiations in the two states of polarization as given in §22 (with the functions $H_l(\mu)$ and $H_r(\mu)$ now defined properly in terms of functional equations of the form (161)), we must have

$$(166)\qquad \begin{aligned} F &= 2\int_0^1 [I_l(\mu) + I_r(\mu)]\mu d\mu \\ &= \frac{3}{4} F\left[(1-c^2)^{1/2}\alpha_1 + \frac{1}{2^{1/2}}(A_2 + cA_1)\right] \end{aligned}$$

where α_1 denotes the first moment of $H_l(\mu)$ and A_1 and A_2 are the first and the second moments, respectively, of $H_r(\mu)$. Hence, we can determine c from the equation

$$(167)\qquad (2(1-c^2))^{1/2}\alpha_1 + A_2 + cA_1 = \frac{4}{3} 2^{1/2}.$$

28. **The functional equation for the problem of diffuse reflection.** The discussion in the preceding sections has shown how we can obtain in *practice* the exact solutions for the angular distribution of the emergent and the reflected radiations from a semi-infinite plane-parallel atmosphere for a wide variety of scattering laws. From a strictly mathematical point of view, the limiting process by which the passage to the limit of infinite approximation was achieved may not have been as rigorously justified as one might have wished. We shall, therefore, now show how the exact solutions obtained in the manner of the preceding sections (by redefining, for example, the H-functions which occur in the solutions in the general nth approximation, in terms of functional equations of the form (161)) can be justified by following an entirely different line of argument. The basic idea in the development we are now going to describe is due to the Armenian astrophysicist, V. A. Ambarzumian.

In the problem of diffuse reflection, we are interested in the solution of the relevant equations of transfer principally, only to the extent that we want to establish the law of diffuse reflection as specified by the $\sigma(\mu, \varphi; \mu_0, \varphi_0)$ which gives the intensity reflected in the direction (μ, φ) when a parallel beam of radiation of unit flux normal to itself is incident on the atmosphere in the direction $(-\mu_0, \varphi_0)$. Now, Ambarzumian starts with the almost trivial observation that the intensity $\sigma(\mu, \varphi; \mu_0, \varphi_0)$ *must be invariant to the addition (or removal) of layers of arbitrary optical thickness to (from) the atmosphere* and shows (and this is really the point of the observation) how this invariance can be used to derive a *functional equation* for the *scattering function* $\sigma(\mu, \varphi; \mu_0, \varphi_0)\mu$. Ambarzumian has explicitly derived the form of this functional equation for the law of diffuse reflection from an atmosphere scattering radiation in accordance with a general phase function, that is, the functional equation associated with the equation of transfer (24). However, when one proceeds to solve the resulting functional equation, one is soon led to simultaneous systems of nonlinear, nonhomogeneous functional equations of such a highly complex nature that one might almost despair of solving them! But, on examination, it soon appears that a knowledge of the *forms* of the solution obtained by the method described in the earlier parts of this lecture enables us, in all cases, to reduce the Ambarzumian type of functional equations to equations of the following standard form

$$H(\mu) = 1 + \mu H(\mu) \int_0^1 \frac{H(\mu')\Psi(\mu')}{\mu + \mu'} d\mu', \tag{168}$$

and helps us, moreover, to confirm the results obtained by our method of passage to the limit of infinite approximation.

29. **The functional equation for the scattering matrix.** In this section we shall derive, following Ambarzumian's general ideas, the functional equation for the scattering matrix S introduced in §11 (equation (34)). This problem is, therefore, more advanced than the ones concered by Ambarzumian; but it serves to illustrate the power of his idea.

To obtain the functional equation for S, we first rewrite the equation of transfer VII in the form

$$\mu \frac{dI}{d\tau} = I(\tau, \mu, \varphi) - B(\tau, \mu, \varphi), \tag{169}$$

where

$$\begin{aligned} B(\tau, \mu, \varphi) = {} & \frac{3}{16\pi} Q \int_{-1}^{+1} \int_0^{2\pi} J(\mu, \varphi; -\mu', \varphi') I(\tau, \mu', \varphi') d\mu' d\varphi' \\ & + \frac{3}{16} Q J(\mu, \varphi; \mu_0, \varphi_0) F e^{-\tau/\mu_0}, \end{aligned} \tag{170}$$

and express the intensity $I(\mu, \varphi; \mu_0, \varphi_0)$ reflected in the direction (μ, φ), where radiation with a net flux πF is incident in the direction $(-\mu_0, \varphi_0)$, in the form

$$I(\mu, \varphi; \mu_0, \varphi_0) = \frac{3}{16\mu} Q S(\mu, \varphi; \mu_0, \varphi_0) F. \tag{171}$$

Now, consider a level at a depth $d\tau$ below the boundary of the atmosphere at $\tau=0$. At this level the radiation field present can be decomposed into two parts: first, there is the reduced incident flux of amount

$$\pi F\left(1-\frac{d\tau}{\mu_0}\right) \tag{172}$$

and, second, there is a diffuse radiation field. The amount of this diffuse radiation field which is directed inward can be inferred from the equation of transfer: for, since at $\tau=0$ there is no inward intensity, at the level $d\tau$, we must have an inward intensity

$$I(d\tau,-\mu',\varphi')=B(0,-\mu',\varphi')\frac{d\tau}{\mu'} \tag{173}$$

in the direction $(-\mu',\varphi')$. Both of these radiation fields will be reflected by the atmosphere below $d\tau$ by the *same laws* as those by which the atmosphere below $\tau=0$ reflects. *This invariance is due to the fact that the removal of a layer of arbitrary thickness from a semi-infinite atmosphere cannot alter its reflecting power.* This is Ambarzumian's basic idea. Accordingly, the reflection of the radiations (172) and (173) by the atmosphere below $d\tau$ will contribute to an *outward* intensity, in the direction (μ,φ), the amount

$$\begin{aligned} I(d\tau,\mu,\varphi) &= \frac{3}{16\mu}\left(1-\frac{d\tau}{\mu_0}\right)QS(\mu,\varphi;\mu_0,\varphi_0)F \\ &\quad+\frac{3}{16\pi\mu}d\tau Q\int_0^1\int_0^{2\pi}S(\mu,\varphi;\mu',\varphi')B(0,-\mu',\varphi')\frac{d\mu}{\mu'}d\varphi'. \end{aligned} \tag{174}$$

On the other hand, from the equation of transfer, we conclude that

$$\begin{aligned} I(d\tau,\mu,\varphi) &= I(0,\mu,\varphi)+\frac{d\tau}{\mu}\left[I(0,\mu,\varphi)-B(0,\mu,\varphi)\right] \\ &= \frac{3}{16\mu}\left(1+\frac{d\tau}{\mu}\right)QS(\mu,\varphi;\mu_0,\varphi_0)F-\frac{d\tau}{\mu}B(0,\mu,\varphi). \end{aligned} \tag{175}$$

Combining equations (174) and (175) and passing to the limit $d\tau=0$, we have

$$\begin{aligned} \frac{3}{16}\left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)QS(\mu,\varphi;\mu_0,\varphi_0)F &= B(0,\mu,\varphi) \\ &\quad+\frac{3}{16\pi}Q\int_0^1\int_0^{2\pi}S(\mu,\varphi;\mu',\varphi')B(0,-\mu',\varphi')\frac{d\mu'}{\mu'}d\varphi'. \end{aligned} \tag{176}$$

But, according to equations (170) and (171)

$$(177)\quad \begin{aligned} B(0, \mu, \varphi) = \frac{3}{16} Q \Bigg[& J(\mu, \varphi; \mu_0, \varphi_0) \\ & + \frac{3}{16\pi}\int_0^1 \int_0^{2\pi} J(\mu, \varphi; -\mu'', \varphi'') Q S(\mu'', \varphi''; \mu_0, \varphi_0) \frac{d\mu''}{\mu''} d\varphi'' \Bigg] F. \end{aligned}$$

Substituting for $B(0, \mu, \varphi)$ from the foregoing equation in equation (176) and remembering that F can be an arbitrary vector, we find, after some minor reductions, that

$$(178)\quad \begin{aligned} \left(\frac{1}{\mu} + \frac{1}{\mu_0}\right) S(\mu, \varphi; \mu_0, \varphi_0) = {} & J(\mu, \varphi; \mu_0, \varphi_0) \\ & + \frac{3}{16\pi}\int_0^1 \int_0^{2\pi} J(\mu, \varphi; -\mu'', \varphi'') Q S(\mu'', \varphi''; \mu_0, \varphi_0) \frac{d\mu''}{\mu''} d\varphi'' \\ & + \frac{3}{16\pi}\int_0^1 \int_0^{2\pi} S(\mu, \varphi; \mu', \varphi') Q J(-\mu', \varphi'; \mu_0, \varphi_0) \frac{d\mu'}{\mu'} d\varphi' \\ & + \frac{9}{256\pi^2}\int_0^1 \int_0^{2\pi} \int_0^1 \int_0^{2\pi} S(\mu, \varphi; \mu', \varphi') Q J(-\mu', \varphi'; -\mu'', \varphi'') Q \\ & \qquad \times S(\mu'', \varphi''; \mu_0, \varphi_0) \frac{d\mu'}{\mu'} d\varphi' \frac{d\mu''}{\mu''} d\varphi'' \end{aligned}$$

This is the required functional equation for the scattering matrix.

Now, the matrix J (see equation (32)) has the property

$$(179)\qquad J_{ik}(\mu, \varphi; \mu_0, \varphi_0) = J_{ki}(\mu_0, \varphi; \mu, \varphi_0).$$

From this property of J for transposition, it follows from equation (178) that S has also the same property:

$$(180)\qquad S_{ik}(\mu, \varphi; \mu_0, \varphi_0) = S_{ki}(\mu_0, \varphi; \mu, \varphi_0).$$

It can be verified that equation (180) is equivalent to *Helmholtz' principle of reciprocity* for the problem under consideration.

30. **The reduction of the functional equation for S.** We shall now indicate how the functional equation for S derived in the preceding section can be solved.

From the form of the equation for S and the manner of its relation to J, it is evident that in a Fourier analysis of the elements of S in $(\varphi - \varphi_0)$ we must have the same nonvanishing components as in the corresponding elements of J. We may accordingly assume without loss of generality that S has the form

$$
(181)\quad S(\mu,\varphi;\mu_0,\varphi) = \left\{
\begin{array}{lll}
S_{11}^{(0)}(\mu,\mu_0) & S_{12}^{(0)}(\mu,\mu_0) & \\
-4\mu\mu_0(1-\mu^2)^{1/2}(1-\mu_0^2)^{1/2}S_{11}^{(1)}(\mu,\mu_0)\cos(\varphi-\varphi_0) & & -2\mu(1-\mu^2)^{1/2}(1-\mu_0^2)^{1/2}S_{13}^{(1)}(\mu,\mu_0)\sin(\varphi-\varphi_0) \\
+\mu^2\mu_0^2S_{11}^{(2)}(\mu,\mu_0)\cos 2(\varphi-\varphi_0) & -\mu^2S_{12}^{(2)}(\mu,\mu_0)\cos 2(\varphi-\varphi_0) & +\mu^2\mu_0S_{13}^{(2)}(\mu,\mu_0)\sin 2(\varphi-\varphi_0) \\
 & & \\
S_{21}^{(0)}(\mu,\mu_0) & S_{22}^{(0)}(\mu,\mu_0) & \\
-\mu_0^2S_{21}^{(2)}(\mu,\mu_0)\cos 2(\varphi-\varphi_0) & +S_{22}^{(2)}(\mu,\mu_0)\cos 2(\varphi-\varphi_0) & -\mu_0S_{23}^{(2)}(\mu,\mu_0)\sin 2(\varphi-\varphi_0) \\
 & & \\
-2\mu_0(1-\mu_0^2)^{1/2}(1-\mu^2)^{1/2}S_{31}^{(1)}(\mu,\mu_0)\sin(\varphi-\varphi_0) & & (1-\mu^2)^{1/2}(1-\mu_0^2)^{1/2}S_{33}^{(1)}(\mu,\mu_0)\cos(\varphi-\varphi_0) \\
+\mu_0^2\mu S_{31}^{(2)}(\mu,\mu_0)\sin 2(\varphi-\varphi_0) & -\mu S_{32}^{(2)}(\mu,\mu_0)\sin 2(\varphi-\varphi_0) & -\mu\mu_0S_{33}^{(2)}(\mu,\mu_0)\cos 2(\varphi-\varphi_0)
\end{array}
\right\}
$$

where, as the notation indicates, $S_{11}^{(0)}$, and so on, are all functions of μ and μ_0 only. From the property (180) of S for transposition, we now conclude that

$$S_{jk}^{(i)}(\mu, \mu_0) = S_{kj}^{(i)}(\mu_0, \mu). \tag{182}$$

If we now substitute the form (181) for S in equation (178) and equate the different Fourier components of the various elements, we shall clearly obtain three systems of functional equations governing the functions of the different orders, distinguished by their superscripts. Of these systems, the first, involving the zero-order functions $S_{11}^{(0)}$, $S_{12}^{(0)}$, $S_{21}^{(0)}$ and $S_{22}^{(0)}$, is the most important and, at the same time, the most difficult. We shall accordingly consider this system briefly.

First, we may write down the equations which are found for $S_{11}^{(0)}$, $S_{12}^{(0)}$, $S_{21}^{(0)}$, and $S_{22}^{(0)}$. The equations are

$$\begin{aligned}\left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)&S_{11}^{(0)}(\mu,\mu_0)\\ &= \left\{\mu^2+\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}\left[\mu'^2S_{11}^{(0)}(\mu,\mu')+S_{12}^{(0)}(\mu,\mu')\right]\right\}\\ &\quad\times\left\{\mu_0^2+\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}\left[\mu'^2S_{11}^{(0)}(\mu',\mu_0)+S_{21}^{(0)}(\mu',\mu_0)\right]\right\}\\ &\quad+2\left\{1-\mu^2+\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}(1-\mu'^2)S_{11}^{(0)}(\mu,\mu')\right\}\\ &\quad\times\left\{1-\mu_0^2+\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}(1-\mu'^2)S_{11}^{(0)}(\mu',\mu_0)\right\};\end{aligned}\tag{183}$$

$$\begin{aligned}\left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)&S_{12}^{(0)}(\mu,\mu_0)\\ &= \left\{\mu^2+\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}\left[\mu'^2S_{11}^{(0)}(\mu,\mu')+S_{12}^{(0)}(\mu,\mu')\right]\right\}\\ &\quad\times\left\{1+\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}\left[\mu'^2S_{12}^{(0)}(\mu',\mu_0)+S_{22}^{(0)}(\mu',\mu_0)\right]\right\}\\ &\quad+2\left\{1-\mu^2+\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}(1-\mu'^2)S_{11}^{(0)}(\mu,\mu')\right\}\\ &\quad\times\left\{\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}(1-\mu'^2)S_{12}^{(0)}(\mu',\mu_0)\right\};\end{aligned}\tag{184}$$

$$(185)\quad \begin{aligned}\left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)&S_{21}^{(0)}(\mu,\mu_0)\\ &=\left\{\mu_0^2+\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}\left[\mu'^2S_{11}^{(0)}(\mu',\mu_0)+S_{21}^{(0)}(\mu',\mu_0)\right]\right\}\\ &\times\left\{1+\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}\left[\mu'^2S_{21}^{(0)}(\mu,\mu')+S_{22}^{(0)}(\mu,\mu')\right]\right\}\\ &+2\left\{1-\mu_0^2+\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}(1-\mu'^2)S_{11}^{(0)}(\mu',\mu_0)\right\}\\ &\times\left\{\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}(1-\mu'^2)S_{21}^{(0)}(\mu,\mu')\right\};\end{aligned}$$

$$(186)\quad \begin{aligned}\left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)&S_{22}^{(0)}(\mu,\mu_0)\\ &=\left\{1+\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}\left[\mu'^2S_{12}^{(0)}(\mu',\mu_0)+S_{22}^{(0)}(\mu',\mu_0)\right]\right\}\\ &\times\left\{1+\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}\left[\mu'^2S_{21}^{(0)}(\mu,\mu')+S_{22}^{(0)}(\mu,\mu')\right]\right\}\\ &+\frac{9}{32}\int_0^1\frac{d\mu'}{\mu'}(1-\mu'^2)S_{21}^{(0)}(\mu,\mu')\\ &\times\int_0^1\frac{d\mu'}{\mu'}(1-\mu'^2)S_{12}^{(0)}(\mu',\mu_0).\end{aligned}$$

An inspection of these equations shows what we have already seen from the functional equation for S, that among these functions of zero-order the relation (182) must hold in particular. In other words

$$(187)\quad \begin{gathered}S_{11}^{(0)}(\mu,\mu_0)=S_{11}^{(0)}(\mu_0,\mu);\qquad S_{12}^{(0)}(\mu,\mu_0)=S_{21}^{(0)}(\mu_0,\mu);\\ S_{22}^{(0)}(\mu,\mu_0)=S_{22}^{(0)}(\mu_0,\mu).\end{gathered}$$

In view of these relations, it follows from equations (183)–(186) that we can express the functions $S_{11}^{(0)}$, $S_{12}^{(0)}$, $S_{21}^{(0)}$ and $S_{22}^{(0)}$ in the forms

$$(188)\quad \left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)S_{11}^{(0)}(\mu,\mu_0)=\psi(\mu)\psi(\mu_0)+2\phi(\mu)\phi(\mu_0),$$

$$(189)\quad \left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)S_{12}^{(0)}(\mu,\mu_0)=\psi(\mu)\chi(\mu_0)+2\phi(\mu)\zeta(\mu_0),$$

$$(190)\qquad \left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)S_{21}^{(0)}(\mu,\mu_0)=\psi(\mu_0)\chi(\mu)+2\phi(\mu_0)\zeta(\mu),$$

and

$$(191)\qquad \left(\frac{1}{\mu}+\frac{1}{\mu_0}\right)S_{22}^{(0)}(\mu,\mu_0)=\chi(\mu)\chi(\mu_0)+2\zeta(\mu)\zeta(\mu_0),$$

where

$$(192)\qquad \psi(\mu)=\mu^2+\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}\left[\mu'^2S_{11}^{(0)}(\mu,\mu')+S_{12}^{(0)}(\mu,\mu')\right],$$

$$(193)\qquad \phi(\mu)=1-\mu^2+\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}(1-\mu'^2)S_{11}^{(0)}(\mu,\mu'),$$

$$(194)\qquad \chi(\mu)=1+\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}\left[\mu'^2S_{21}^{(0)}(\mu,\mu')+S_{22}^{(0)}(\mu,\mu')\right],$$

and

$$(195)\qquad \zeta(\mu)=\frac{3}{8}\int_0^1\frac{d\mu'}{\mu'}(1-\mu'^2)S_{21}^{(0)}(\mu,\mu').$$

Substituting for $S_{11}^{(0)}$, and so on, from equations (188)–(191) back into equations (192)–(195) we obtain the functional equations for the problem in their normal forms. We have

$$(196)\qquad \begin{aligned}\psi(\mu)=\mu^2&+\frac{3}{8}\mu\psi(\mu)\int_0^1\frac{d\mu'}{\mu+\mu'}\left[\mu'^2\psi(\mu')+\chi(\mu')\right]\\&+\frac{3}{4}\mu\phi(\mu)\int_0^1\frac{d\mu'}{\mu+\mu'}\left[\mu'^2\phi(\mu')+\zeta(\mu')\right],\end{aligned}$$

$$(197)\qquad \begin{aligned}\phi(\mu)=1-\mu^2&+\frac{3}{8}\mu\psi(\mu)\int_0^1\frac{d\mu'}{\mu+\mu'}(1-\mu'^2)\psi(\mu')\\&+\frac{3}{4}\mu\phi(\mu)\int_0^1\frac{d\mu'}{\mu+\mu'}(1-\mu'^2)\phi(\mu'),\end{aligned}$$

$$(198)\qquad \begin{aligned}\chi(\mu)=1&+\frac{3}{8}\mu\chi(\mu)\int_0^1\frac{d\mu'}{\mu+\mu'}\left[\mu'^2\psi(\mu')+\chi(\mu')\right]\\&+\frac{3}{4}\mu\zeta(\mu)\int_0^1\frac{d\mu'}{\mu+\mu'}\left[\mu'^2\phi(\mu')+\zeta(\mu')\right],\end{aligned}$$

$$(199)\qquad \begin{aligned}\zeta(\mu) = \frac{3}{8}\mu\chi(\mu)\int_0^1 \frac{d\mu'}{\mu+\mu'}(1-\mu'^2)\psi(\mu') \\ + \frac{3}{4}\mu\zeta(\mu)\int_0^1 \frac{d\mu'}{\mu+\mu'}(1-\mu'^2)\phi(\mu').\end{aligned}$$

31. **The solution of the functional equations (196)–(199).** Equations (196)–(199) represent a nonlinear, nonhomogeneous system of four simultaneous functional equations which one might despair of even attempting to solve. However, with the guidance provided by the form of the solution obtained in our general nth approximation, it is possible to reduce the solution of equations (196)–(199) to two simple functional equations, each of form (168). Thus, it can be verified by direct substitution that the solution of the system of equations (196)–(199) is given by

$$(200)\qquad \psi(\mu) = (2(1-c^2))^{1/2}\mu H_l(\mu),$$

$$(201)\qquad \phi(\mu) = H_l(\mu)(1-c\mu),$$

$$(202)\qquad \chi(\mu) = H_r(\mu)(1+c\mu),$$

and

$$(203)\qquad \zeta(\mu) = \frac{(1-c^2)^{1/2}}{2^{1/2}}\mu H_r(\mu),$$

where $H_l(\mu)$ and $H_r(\mu)$ are defined in terms of the functional equations

$$(204)\qquad H_l(\mu) = 1 + \frac{3}{4}\mu H_l(\mu)\int_0^1 \frac{H_l(\mu')}{\mu+\mu'}(1-\mu'^2)d\mu'$$

and

$$(205)\qquad H_r(\mu) = 1 + \frac{3}{8}\mu H_r(\mu)\int_0^1 \frac{H_r(\mu')}{\mu+\mu'}(1-\mu'^2)d\mu',$$

and c is a constant related in a determinate way with the moments of $H_l(\mu)$ and $H_r(\mu)$. We find

$$(206)\qquad c = \frac{8(A_1-\alpha_1)+3(2\alpha_1\alpha_0 - A_1A_0)}{3(A_1^2+2\alpha_1^2)},$$

where α_0, A_0 and α_1, A_1 are the moments of order zero and one of $H_l(\mu)$ and $H_r(\mu)$, respectively.

With $\psi(\mu)$, $\phi(\mu)$, and so on, defined as in equations (200)–(203) we can verify that the solutions for $S_{\text{II}}^{(0)}$, and so on, are in entire agreement with our earlier results obtained by passing to the limit of infinite approximation.

32. **The completion of the solution for S.** The discussion of the other two systems for the functions of order one and two turns out to be very simple, as it appears that all the functions $S^{(1)}$ are equal to each other and similarly all the functions $S^{(2)}$ are equal to each other. Therefore, writing

$$S_{i,j}^{(1)}(\mu, \mu_0) = S^{(1)}(\mu, \mu_0), \tag{207}$$

and

$$S_{i,j}^{(2)}(\mu, \mu_0) = S^{(2)}(\mu, \mu_0), \tag{208}$$

it is found that the equations governing $S^{(1)}$ and $S^{(2)}$ are

$$\begin{aligned} \left(\frac{1}{\mu}+\frac{1}{\mu_0}\right) & S^{(1)}(\mu, \mu_0) \\ &= \left\{1+\frac{3}{8}\int_0^1 \frac{d\mu'}{\mu'}(1-\mu'^2)(1+2\mu'^2)S^{(1)}(\mu', \mu_0)\right\} \\ &\quad \times \left\{1+\frac{3}{8}\int_0^1 \frac{d\mu'}{\mu'}(1-\mu'^2)(1+2\mu'^2)S^{(1)}(\mu, \mu')\right\}, \end{aligned} \tag{209}$$

and

$$\begin{aligned} \left(\frac{1}{\mu}+\frac{1}{\mu_0}\right) S^{(2)}(\mu, \mu_0) &= \left\{1+\frac{3}{16}\int_0^1 \frac{d\mu'}{\mu'}(1+\mu'^2)^2 S^{(2)}(\mu', \mu_0)\right\} \\ &\quad \times \left\{1+\frac{3}{16}\int_0^1 \frac{d\mu'}{\mu'}(1+\mu'^2)^2 S^{(2)}(\mu, \mu')\right\}. \end{aligned} \tag{210}$$

From these equations it follows that the functions $S^{(1)}(\mu, \mu_0)$ and $S^{(2)}(\mu, \mu_0)$ are symmetrical in the variables μ and μ_0 and that they are expressible in the forms

$$\left(\frac{1}{\mu}+\frac{1}{\mu_0}\right) S^{(1)}(\mu, \mu_0) = H^{(1)}(\mu)H^{(1)}(\mu_0), \tag{211}$$

and

$$\left(\frac{1}{\mu}+\frac{1}{\mu_0}\right) S^{(2)}(\mu, \mu_0) = H^{(2)}(\mu)H^{(2)}(\mu_0) \tag{212}$$

where $H^{(1)}(\mu)$ and $H^{(2)}(\mu)$ are solutions of the functional equations

$$H^{(1)}(\mu) = 1 + \frac{3}{8}\mu H^{(1)}(\mu)\int_0^1 \frac{H^{(1)}(\mu')}{\mu+\mu'}(1-\mu'^2)(1+2\mu'^2)d\mu', \tag{213}$$

and

$$H^{(2)}(\mu) = 1 + \frac{3}{16}\mu H^{(2)}(\mu)\int_0^1 \frac{H^{(2)}(\mu')}{\mu+\mu'}(1+\mu'^2)^2 d\mu'. \tag{214}$$

With this, the solution of the functional equation for S is completed and it will be observed that the solution for S which we have now obtained is of exactly the same form as that given in §22 (under D) with the only difference that the H-functions which appear in the solution are now defined in terms of the exact functional equations which they satisfy; further, the constant c is shown to be related in a definite way with the moments of $H_l(\mu)$ and $H_r(\mu)$.

33. **A class of functional equations and their solution.** The solution of the functional equations for the problem of diffuse reflection for laws of scattering, other than the one we have described, can be carried out in an analogous manner. It is not our intention to go into the details of the solution of these other cases here, but it may be of some mathematical interest to see the type of functional equations which these problems lead us to consider.

The problem of diffuse reflection in accordance with Rayleigh's phase function leads to the following simultaneous pair of functional equations:

$$\begin{aligned}\psi(\mu) = 3 - \mu^2 &+ \frac{1}{16}\mu\psi(\mu)\int_0^1 \frac{\psi(\mu')}{\mu+\mu'}(3-\mu'^2)d\mu' \\ &+ \frac{1}{2}\mu\phi(\mu)\int_0^1 \frac{\phi(\mu')}{\mu+\mu'}(3-\mu'^2)d\mu'\end{aligned} \tag{215}$$

and

$$\begin{aligned}\phi(\mu) = \mu^2 &+ \frac{1}{16}\mu\psi(\mu)\int_0^1 \frac{\psi(\mu')}{\mu+\mu'}\mu'^2 d\mu' \\ &+ \frac{1}{2}\mu\phi(\mu)\int_0^1 \frac{\phi(\mu')}{\mu+\mu'}\mu'^2 d\mu'.\end{aligned} \tag{216}$$

Again, guided by the form of the solution obtained by the direct solution of the equation of transfer (see the tabulation in §22, under B), we are led to surmise that the solutions of equations (215) and

(216) must be of the forms

$$\psi(\mu) = H^{(0)}(\mu)(3 - c\mu) \tag{217}$$

and

$$\phi(\mu) = q\mu H^{(0)}(\mu), \tag{218}$$

where $H^{(0)}(\mu)$ satisfies the functional equation

$$H^{(0)}(\mu) = 1 + \frac{3}{16}\mu H^{(0)}(\mu)\int_0^1 \frac{H^{(0)}(\mu')}{\mu+\mu'}(3 - \mu'^2)d\mu', \tag{219}$$

and q and c are two constants related in the manner

$$8q^2 = 3 - c^2. \tag{220}$$

Direct substitution confirms that the solutions of equations (215) and (216) are indeed of the forms surmised and shows further that, in agreement with equation (165),

$$q = \frac{2}{3\alpha_1} \tag{221}$$

where α_1 is the first moment of $H^{(0)}(\mu)$.

Similarly, the problem of diffuse reflection in accordance with the phase function $\lambda(1+x\cos\Theta)$ leads to the following pair of functional equations:

$$\begin{aligned}\psi(\mu) = 1 &+ \frac{1}{2}\lambda\mu\psi(\mu)\int_0^1 \frac{\psi(\mu')}{\mu+\mu'}d\mu' \\ &- \frac{1}{2}x\lambda\mu\phi(\mu)\int_0^1 \frac{\phi(\mu')}{\mu+\mu'}d\mu',\end{aligned} \tag{222}$$

and

$$\begin{aligned}\phi(\mu) = \mu &- \frac{1}{2}\lambda\mu\psi(\mu)\int_0^1 \frac{\psi(\mu')}{\mu+\mu'}\mu' d\mu' \\ &+ \frac{1}{2}x\lambda\mu\phi(\mu)\int_0^1 \frac{\phi(\mu')}{\mu+\mu'}\mu' d\mu'.\end{aligned} \tag{223}$$

And, it is found by direct verification that the solution of these equations can be expressed in the form

$$\psi(\mu) = H^{(0)}(\mu)(1 - c\mu), \tag{224}$$

and

$$\phi(\mu) = q\mu H^{(0)}(\mu), \tag{225}$$

where $H^{(0)}(\mu)$ now satisfies the functional equation

$$H^{(0)}(\mu) = 1 + \frac{1}{2}\lambda\mu H^{(0)}(\mu)\int_0^1 \frac{H^{(0)}(\mu')}{\mu+\mu'}[1 + x(1-\lambda)\mu'^2]d\mu', \tag{226}$$

and q and c are two constants related in the manner

$$xq^2 = c^2 + x(1-\lambda), \tag{227}$$

and given explicitly by the formulae

$$q = \frac{2(1-\lambda)}{2-\lambda\alpha_0} \quad \text{and} \quad c = x\lambda(1-\lambda)\frac{\alpha_1}{2-\lambda\alpha_0}, \tag{228}$$

α_0 and α_1 being the moments of order zero and one of $H^{(0)}(\mu)$.

34. **Some general remarks.** In some ways it is remarkable that systems of functional equations as complex in appearance as equations (196)–(199), (215)–(216) and (222)–(223) are should be capable of being reduced to single functional equations of the form (168). There must clearly be something in the structure of these equations which makes this reduction possible. But, as to what it precisely is, is at present shrouded in mystery!

35. **Some integral properties of the functions $H(\mu)$.** The discussion in the preceding sections has disclosed the important role which functional equations of the form

$$H(\mu) = 1 + \mu H(\mu)\int_0^1 \frac{H(\mu')}{\mu+\mu'}\Psi(\mu')d\mu' \tag{229}$$

play in the theory of radiative transfer. The investigation of the properties of these equations is, therefore, a matter of considerable interest. Of course, from the practical standpoint of solving such equations numerically, the most important property is that derived from Theorem 1 (§24), namely, that when we replace the integral on the right-hand side by a Gauss sum, the solution can be explicitly written down as a rational function involving the points of the Gaussian division and the roots of the associated characteristic equation

$$1 = 2\sum_{j=1}^{n} \frac{a_j\Psi(\mu_j)}{1-k^2\mu_j^2}. \tag{230}$$

However, in addition to this property, there are a number of integral theorems (of an essentially elementary kind) which can be proved for functions satisfying equations of the form (229). We shall give two examples.

THEOREM 3. $\int_0^1 H(\mu)\Psi(\mu)d\mu = 1 - [1 - 2\int_0^1 \Psi(\mu)d\mu]^{1/2}$.

PROOF. Multiplying the equation satisfied by $H(\mu)$ by $\Psi(\mu)$ and integrating over the range of μ, we have

$$
\begin{aligned}
(231)\qquad & \int_0^1 H(\mu)\Psi(\mu)d\mu \\
& = \int_0^1 \Psi(\mu)d\mu + \int_0^1 \int_0^1 \frac{\mu}{\mu + \mu'} H(\mu)\Psi(\mu)H(\mu')\Psi(\mu')d\mu d\mu'.
\end{aligned}
$$

Interchanging μ and μ' in the double integral in equation (231) and taking the average of the two equations, we obtain

$$
\begin{aligned}
(232)\qquad & \int_0^1 H(\mu)\Psi(\mu)d\mu \\
& = \int_0^1 \Psi(\mu)d\mu + \frac{1}{2}\int_0^1 \int_0^1 H(\mu)\Psi(\mu)H(\mu')\Psi(\mu')d\mu d\mu'
\end{aligned}
$$

or, alternatively,

$$
(233)\qquad \frac{1}{2}\left[\int_0^1 H(\mu)\Psi(\mu)d\mu\right]^2 - \int_0^1 H(\mu)\Psi(\mu)d\mu + \int_0^1 \Psi(\mu)d\mu = 0.
$$

Solving this equation for the integral in question, we have

$$
(234)\qquad \int_0^1 H(\mu)\psi(\mu)d\mu = 1 \pm \left[1 - 2\int_0^1 \Psi(\mu)d\mu\right]^{1/2}.
$$

The ambiguity in the sign in equation (234) can be removed by the consideration that the integral on the left-hand side must uniformly converge to zero when $\Psi(\mu)$ tends to zero uniformly in the interval (0, 1). This requires us to choose the negative sign in equation (234) and the result stated follows.

COROLLARY. *A necessary and sufficient condition that $H(\mu)$ be real is*

$$
\int_0^1 \Psi(\mu)d\mu \leq \frac{1}{2}\cdot
$$

This is, of course, an immediate consequence of the theorem.

The physical meaning of the limitation on $\Psi(\mu)$ implied by this corollary is interesting: it is really equivalent to the condition that, on each scattering, more radiation should not be emitted than was incident; further, the equality sign is admissible only in the case of perfect scattering in the sense of equation (11).

THEOREM 4. $[1-2\int_0^1\Psi(\mu)d\mu]^{1/2}\int_0^1 H(\mu)\Psi(\mu)\mu^2 d\mu+[\int_0^1 H(\mu)\Psi(\mu)\mu d\mu]^2/2 =\int_0^1\Psi(\mu)\mu^2 d\mu$.

PROOF. To prove this theorem, we multiply the equation defining $H(\mu)$ by $\Psi(\mu)\mu^2$ and integrate over the range of μ. We find

$$(235)\quad \begin{aligned} &\int_0^1 H(\mu)\Psi(\mu)\mu^2 d\mu \\ &= \int_0^1 \Psi(\mu)\mu^2 d\mu + \int_0^1\int_0^1 \frac{H(\mu)H(\mu')\Psi(\mu)\Psi(\mu')}{\mu+\mu'}\mu^2 d\mu d\mu' \\ &= \int_0^1 \Psi(\mu)\mu^2 d\mu \\ &\quad + \frac{1}{2}\int_0^1\int_0^1 \frac{H(\mu)H(\mu')\Psi(\mu)\Psi(\mu')}{\mu+\mu'}(\mu^3+\mu'^3)d\mu d\mu' \\ &= \int_0^1 \Psi(\mu)\mu^2 d\mu \\ &\quad + \frac{1}{2}\int_0^1\int_0^1 H(\mu)H(\mu')\Psi(\mu)\Psi(\mu')(\mu^2-\mu\mu'+\mu'^2)d\mu d\mu' \\ &= \int_0^1 \Psi(\mu)\mu^2 d\mu + \left[\int_0^1 H(\mu)\Psi(\mu)\mu^2 d\mu\right]\left[\int_0^1 H(\mu)\Psi(\mu)d\mu\right] \\ &\quad - \frac{1}{2}\left[\int_0^1 H(\mu)\Psi(\mu)\mu d\mu\right]^2. \end{aligned}$$

Using Theorem 3 we obtain, after some minor reductions,

$$(236)\quad \begin{aligned} &\left[1-2\int_0^1 \Psi(\mu)d\mu\right]^{1/2}\int_0^1 H(\mu)\Psi(\mu)\mu^2 d\mu \\ &\qquad + \frac{1}{2}\left[\int_0^1 H(\mu)\Psi(\mu)\mu d\mu\right]^2 = \int_0^1 \Psi(\mu)\mu^2 d\mu, \end{aligned}$$

which is the required result.

COROLLARY. *For the case of a perfectly scattering atmosphere when*

$$\int_0^1 \Psi(\mu)d\mu = 1/2,$$

we have the further integral

$$\int_0^1 H(\mu)\Psi(\mu)\mu d\mu = \left[2\int_0^1 \Psi(\mu)\mu^2 d\mu \right]^{1/2}.$$

The corollary we have just stated generalizes a classical result of Hopf and Bronstein for the case of an isotropically scattering atmosphere. For, in this latter case

$$\Psi(\mu) = \text{constant} = 1/2, \tag{237}$$

and, according to Theorem 3 and the corollary of Theorem 4, we have

$$\int_0^1 H(\mu)d\mu = 2, \tag{238}$$

$$\int_0^1 H(\mu)\mu d\mu = \frac{2}{3^{1/2}}. \tag{239}$$

Hence

$$\frac{J(0)}{F} = \frac{\int_0^1 H(\mu)d\mu}{4\int_0^1 H(\mu)\mu d\mu} = \frac{3^{1/2}}{4}. \tag{240}$$

This is the Hopf-Bronstein relation. It, therefore, follows that for all cases of perfect scattering we have an integral of the Hopf-Bronstein type which is essentially that given by the corollary of Theorem 4.

36. **The equation of transfer in spherical atmospheres, and its reduction for the case $\kappa\rho\alpha r^{-n}$.** So far we have restricted ourselves to transfer problems in plane parallel atmospheres. We shall now briefly indicate how the methods we have described can be extended to treat transfer problems in spherically symmetric atmospheres. In such cases, the intensity in the radiation field will be a function of the distance r from the center of symmetry and the angle ϑ measured from the positive direction of the radius vector; and the equation of transfer will take the form

$$\mu\frac{\partial I}{\partial r} + \frac{1-\mu^2}{r}\frac{\partial I}{\partial \mu} = -\kappa\rho(I - \mathfrak{J}), \tag{241}$$

where $\mu = \cos\vartheta$ and $\mathfrak{J}$ denotes, as usual, the source function.

In outlining the manner in which equations of transfer of the form (241) can be solved, we shall restrict ourselves to an isotropically scattering atmosphere. In this case equation (241) becomes

$$\mu\frac{\partial I}{\partial r}+\frac{1-\mu^2}{r}\frac{\partial I}{\partial\mu}=-\kappa\rho I+\frac{1}{2}\kappa\rho\int_{-1}^{+1}I(r,\mu')d\mu'. \tag{242}$$

According to the ideas developed in the earlier parts of this lecture, we shall replace the integral which occurs on the right-hand side of equation (242) by a sum according to Gauss's formula for numerical quadratures, and reduce the integrodifferential equation to the system

$$\mu_i\frac{dI_i}{dr}+\frac{1-\mu_i^2}{r}\left(\frac{\partial I}{\partial\mu}\right)_{\mu=\mu_i}=-\kappa\rho I_i+\frac{1}{2}\kappa\rho\sum a_jI_j \qquad (i=\pm 1,\cdots,\pm n) \tag{243}$$

of $2n$ ordinary linear equations in the nth approximation. It is at once seen that our present system of equations differs, in an essential way, from those which we have considered so far: equation (243) now involves $(\partial I/\partial\mu)_{\mu=\mu_i}$ and, before we can proceed any further, we must know the values which we are to assign to $\partial I/\partial\mu$ at the points of the Gaussian division in our present scheme of approximation. At first sight it might be supposed that the assignment of values to $\partial I/\partial\mu$ at $\mu=\mu_i$, $i=\pm 1,\cdots,\pm n$, is largely arbitrary, particularly when n is small. However, on consideration, it appears that this assignment can be done in a satisfactory manner in only one way and, indeed, according to the following device:

Define the polynomials $Q_m(\mu)$ according to the formula

$$P_m(\mu)=-\frac{dQ_m}{d\mu} \qquad (m=1,\cdots,2n), \tag{244}$$

and adjust the constant of integration in Q_m by requiring that

$$Q_m=0 \qquad \text{for } |\mu|=1. \tag{245}$$

This can always be accomplished, since, when m is odd, Q_m is even and when m is even the indefinite integral of $P_m(\mu)$ already contains $(1-\mu^2)$ as a factor. The first few polynomials $Q_m(\mu)$ are given below:

m	$P_m(\mu)$	$Q_m(\mu)$	$\mathfrak{Q}_m(\mu)$
1	μ	$(1-\mu^2)/2$	$1/2$
2	$(3\mu^2-1)/2$	$\mu(1-\mu^2)/2$	$\mu/2$
3	$(5\mu^3-3\mu)/2$	$1/8(5\mu^2-1)(1-\mu^2)$	$1/8(5\mu^2-1)$
4	$1/8(35\mu^4-30\mu^2+3)$	$1/8\mu(7\mu^2-3)(1-\mu^2)$	$1/8\mu(7\mu^2-3)$
5	$1/8(63\mu^5-70\mu^3+15\mu)$	$1/16(21\mu^4-14\mu^2+1)(1-\mu^2)$	$1/16(21\mu^4-14\mu^2+1)$

Now, by an integration by parts, we arrive at the identity

$$\text{(246)} \qquad \int_{-1}^{+1} Q_m(\mu)\frac{\partial I}{\partial \mu}d\mu = -\int_{-1}^{+1} I\frac{dQ_m}{d\mu}d\mu = \int_{-1}^{+1} IP_m(\mu)d\mu.$$

Expressing the first and the last integrals in equation (246) as sums according to Gauss's formula, we have, in the nth approximation,

$$\text{(247)} \qquad \sum a_i Q_m(\mu_i)\left(\frac{\partial I}{\partial \mu}\right)_{\mu=\mu_i} = \sum a_i I_i P_m(\mu_i) \qquad (m = 1, \cdots, 2n).$$

Equation (247) provides us with exactly the right number of equations to express $(\partial I/\partial \mu)_{\mu=\mu_i}$ $(i = \pm 1, \cdots, \pm n)$ as linear combinations of I_i. Essentially what equation (247) allows is to determine in a "best possible way" the derivatives of a function in terms of its values at the points of the Gaussian division. This problem has apparently not been considered before.

Returning to equation (243) we now observe that this equation, together with equation (246), provides the required reduction of the equation of transfer (242) to an equivalent system of linear equations.

For purposes of practical solution it appears most convenient to combine equations (243) and (247) in the following manner.

Since we have arranged $Q_m(\mu)$ to be divisible by $(1-\mu^2)$, we can clearly write

$$\text{(248)} \qquad Q_m(\mu) = \mathfrak{Q}_m(\mu)(1-\mu^2).$$

The first few of the polynomials $\mathfrak{Q}_m(\mu)$ are listed in the preceding tabulation.

Now, multiply equation (243) by $a_i\mathfrak{Q}_m(\mu_i)$ and sum over all i's. We obtain

$$\text{(249)} \qquad \begin{aligned} &\frac{d}{dr}\sum a_i\mu_i\mathfrak{Q}_m(\mu_i)I_i + \frac{1}{r}\sum a_i(1-\mu_i^2)\mathfrak{Q}_m(\mu_i)\left(\frac{\partial I}{\partial \mu}\right)_{\mu=\mu_i} \\ &\qquad = -\kappa\rho\sum a_i\mathfrak{Q}_m(\mu_i)I_i + \frac{1}{2}\kappa\rho\left(\sum a_i I_i\right)\left[\sum a_i\mathfrak{Q}_m(\mu_i)\right] \end{aligned}$$

$$(m = 1, \cdots, 2n).$$

But, according to equations (247) and (248)

$$\text{(250)} \qquad \begin{aligned} \sum a_i(1-\mu_i^2)\mathfrak{Q}_m(\mu_i)\left(\frac{\partial I}{\partial \mu}\right)_{\mu=\mu_i} &= \sum a_i Q_m(\mu_i)\left(\frac{\partial I}{\partial \mu}\right)_{\mu=\mu_i} \\ &= \sum a_i P_m(\mu_i)I_i. \end{aligned}$$

Equation (249), therefore, reduces to

$$
(251)\quad
\begin{aligned}
&\frac{d}{dr}\sum a_i\mu_i\mathfrak{Q}_m(\mu_i)I_i + \frac{1}{r}\sum a_iP_m(\mu_i)I_i \\
&\qquad = -\kappa\rho\sum a_i\mathfrak{Q}_m(\mu_i)I_i + \frac{1}{2}\kappa\rho\left(\sum a_iI_i\right)\left[\sum a_i\mathfrak{Q}_m(\mu_i)\right]
\end{aligned}
$$

$$(m = 1, \cdots, 2n).$$

This is the required system of linear equations in the nth approximation.

Equation (251) for the case $m=1$ admits of immediate integration. For, when $m=1$

$$(252)\qquad P_1(\mu) = \mu \quad\text{and}\quad \mathfrak{Q}_1(\mu) = 1/2,$$

and equation (251) yields

$$(253)\qquad \frac{1}{2}\frac{d}{dr}\sum a_i\mu_iI_i + \frac{1}{r}\sum a_i\mu_iI_i = 0$$

or

$$(254)\qquad \sum a_i\mu_iI_i = \frac{1}{2}\frac{F_0^1}{r^2},$$

where F_0 is a constant. This is the equivalent, in our approximation, of the flux integral

$$(255)\qquad F = 2\int_{-1}^{+1} I\mu d\mu = \frac{F_0}{r^2},$$

which the equation of transfer (242) admits directly.

Again, since $\mathfrak{Q}_m(\mu)$ is odd when m is even, equation (251) reduces for even values of m to the form

$$(256)\qquad \frac{d}{dr}\sum a_i\mu_i\mathfrak{Q}_m(\mu_i)I_i + \frac{1}{r}\sum a_iP_m(\mu_i)I_i = -\kappa\rho\sum a_i\mathfrak{Q}_m(\mu_i)I_i$$

$$(m = 2, 4, \cdots, 2n).$$

For $m=2n$, the foregoing equation further simplifies to

$$(257)\qquad \frac{d}{dr}\sum a_i\mu_i\mathfrak{Q}_{2n}(\mu_i)I_i = -\kappa\rho\sum a_i\mathfrak{Q}_{2n}(\mu_i)I_i.$$

Finally, we may note the explicit forms of the equations in the second approximation. They are

$$\sum a_i\mu_i I_i = \frac{1}{2}\frac{F_0}{r^2},$$

$$\frac{d}{dr}\sum a_i\mu_i^2 I_i + \frac{1}{r}\sum a_i(3\mu_i^2 - 1)I_i = -\kappa\rho\sum a_i\mu_i I_i,$$

$$(258)\quad \frac{d}{dr}\sum a_i\mu_i(5\mu_i^2 - 1)I_i + \frac{4}{r}\sum a_i\mu_i(5\mu_i^2 - 3)I_i$$

$$= -\frac{5}{3}\kappa\rho\sum a_i(3\mu_i^2 - 1)I_i,$$

$$\frac{d}{dr}\sum a_i\mu_i^2(7\mu_i^2 - 3)I_i = -\kappa\rho\sum a_i\mu_i(7\mu_i^2 - 3)I_i.$$

From the point of view of astrophysical applications, greatest interest is attached to the case when $\kappa\rho$ varies as some inverse of power. And when

$$(259)\qquad \kappa\rho = \frac{\text{constant}}{r^n} \qquad (n > 1),$$

the equations of the second approximation (258) can be solved explicitly and the various physical quantities expressed as integrals over the Bessel functions, I_ν and K_ν, of purely imaginary argument, and of order

$$(260)\qquad \nu = \frac{n+5}{2(n-1)}.$$

It does not, however, seem that the passage to the limit of infinite approximation can be achieved as simply as in the case of transfer problems in plane parallel atmospheres.

37. **The equation of transfer in a differentially moving atmosphere. The influence of Doppler effect.** Finally, we shall turn to a class of transfer problems which is of an altogether different character from the ones we have considered so far. The problems in question arise in connection with the study of the transfer of radiation in atmospheres in which the different parts are in relative motion. To be specific, consider a plane-parallel atmosphere in which the material at height z has a velocity $w(z_1)$ with respect to a stationary observer. The novelty of the situation arises on account of Doppler effect which makes the radiation scattered in different directions have different frequencies

as judged by a stationary observer. Consequently, the radiation field in the different frequencies will interact with each other in a manner which is not always easy to visualize. However, in the astrophysical contexts, two circumstances simplify the problem. First, the velocities which are involved are small compared to the velocity of light, c, and second, the only effects of consequence are those which arise from the sensitive dependence of the scattering coefficient $\sigma(\nu)$ on frequency. This last circumstance, in particular, allows us to ignore all effects such as aberration and so on, and concentrate only on the effects arising from the change of frequency on scattering. Under these conditions, the equation of transfer can be shown to take the form

$$\begin{aligned}\mu\frac{\partial I(\nu, z, \mu)}{\rho\partial z} = &-\sigma\left(\nu - \nu_0\frac{w}{c}\mu\right)\Big\{I(\nu, z, \mu)\\ &-\frac{1}{2}\int_{-1}^{+1} I\left(\nu - \nu_0\frac{w}{c}\mu + \nu_0\frac{w}{c}\mu', z, \mu'\right)d\mu'\Big\},\end{aligned}\tag{261}$$

where ν_0 denotes the frequency of the "center of the line."

With suitable simplifying assumptions concerning $\sigma(\nu)$ and $w(z)$, the discussion of the equation of transfer (261), in the first approximation in our scheme of replacing integrals by Gauss sums, leads to a variety of novel types of boundary value problems in hyperbolic equations. It may be of some interest to specify the nature of these boundary value problems and indicate the methods which have been developed for their solution.

38. **A new class of boundary value problems in hyperbolic equations.** As related to the equation

$$\frac{\partial^2 f}{\partial x^2} - \frac{\partial^2 f}{\partial y^2} + f = 0\tag{262}$$

the boundary value problems which are of most frequent occurrence are of the following general type:

The value of f and its derivatives are assigned for

$$y = 0 \quad\text{and}\quad 0 \leqq x \leqq l_1,\tag{263}$$

that is, along AB in Fig. 2. Along AD ($x=0$ and $0\leqq y\leqq l_2$) and BC ($x=l_1$, $0\leqq y\leqq l_2$), we are further given that

$$\left(\frac{\partial f}{\partial x}\right)_{x=0} = \left(\frac{\partial f}{\partial y}\right)_{x=0} + \phi(y) \qquad (0 \leqq y \leqq l_2),\tag{264}$$

$$f(l_1, y) = \psi(y) \qquad (0 \leqq y \leqq l_2)\tag{265}$$

where $\phi(y)$ and $\psi(y)$ are two known functions. The problem is to solve equation (1) in the rectangular strip $ABCD$ satisfying the stated boundary conditions. For the particular boundary value problems which occur in the astrophysical contexts, the following "systematic method" of solution has been found convenient.

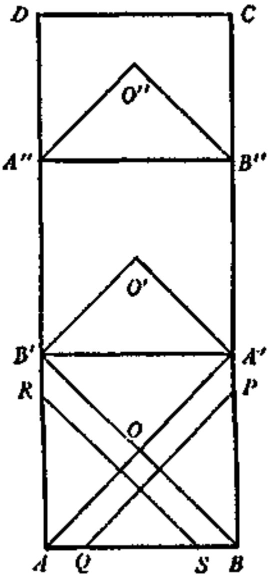


FIG. 2.

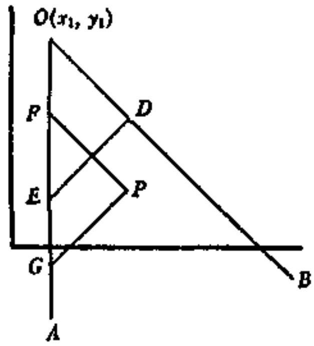


FIG. 3.

Let the characteristics $x=y$ and $l_1-x=y$ through A and B intersect BC and AD at A' and B', respectively. Further, let AA' and BB' intersect at O. First, since the function and its derivatives are known along AB, we can find the solution inside and on the sides of the triangle OAB directly by Riemann's well known method. Next we use the requirement that the solution be unambiguously defined along AO and OB together with the boundary conditions specified along AD and BC to determine f along AB' and $\partial f/\partial x$ along BA' as solutions of certain integral equations of Volterra's type. With the knowledge of the function and its derivatives thus completed along the part $B'ABA'$ of the "supporting curve" $ABCD$, the solution inside the entire region $O'B'ABA'$ becomes determinate by Riemann's method. In particular, the function and its derivatives along $B'A'$ can be found and the continuation of the solution in the second square $B'A'B''A''$ follows along similar lines.

While the method outlined above shows how solutions satisfying the given boundary conditions can be found in principle, it suffers from the disadvantage that the method of solution depends on solving a succession of Volterra integral equations; and, unless the boundary conditions specified along $DABC$ are specially simple, we should not expect to go very far in the explicit carrying out of the solution by this method. It would, therefore, be useful if an alternative method of solution can be devised which will eliminate the need of solving integral equations and reduce the practical problem to one involving (at most!) only quadratures. It is remarkable that this can actually be accomplished by constructing suitable Green's functions and applying Green's theorem to contours, such as RAS and PQB.

39. **The Green's functions $C(x, y; x_1, y_1)$ and $\Gamma(x, y; x_1, y_1)$.** It is found that Green's functions which are appropriate for the solution of the boundary value problems of the type formulated in §38 are $C(x, y; x_1, y_1)$ and $\Gamma(x, y; x_1, y_1)$, defined as follows:

$C(x, y; x_1, y_1)$ is a solution of the hyperbolic equation (262) which satisfies the boundary conditions

$$C(x_1, y; x_1, y_1) = 1 \qquad (y \leq y_1) \tag{266}$$

and

$$C(x, x_1 + y_1 - x; x_1, y_1) = 1 \qquad (x \geq x_1). \tag{267}$$

In other words, if O represents the point (x_1, y_1), OB the characteristic $x+y=x_1+y_1$ through O and OA the line through O parallel to the y-axis, then the boundary conditions require that C take the value 1 along both OA and OB (see Fig. 3). A solution satisfying these boundary conditions can be found explicitly. It is given by

$$\begin{aligned} C(x, y; x_1, y_1) = \cos (x - x_1) + \frac{1}{2}(x - x_1) \int_0^{\pi} J_0([x - x_1] \sin \vartheta) \\ \times Ii_1(y_1 - y - [x - x_1] \cos \vartheta) \sin \vartheta d\vartheta, \end{aligned} \tag{268}$$

where J_0 denotes the Bessel function of order zero and $Ii_1(z)$ the "Bessel integral"

$$Ii_1(z) = \int_0^z \frac{I_1(t)}{t} dt. \tag{269}$$

($I_1(t)$ denotes the Bessel function of order 1 for a purely imaginary argument.)

The second function, $\Gamma(x, y; x_1, y_1)$, is defined in terms of

$C(x, y; x_1, y_1)$ according to the formula

$$\Gamma(x, y; x_1, y_1) = \frac{\partial C}{\partial x} - \frac{\partial C}{\partial y}. \tag{270}$$

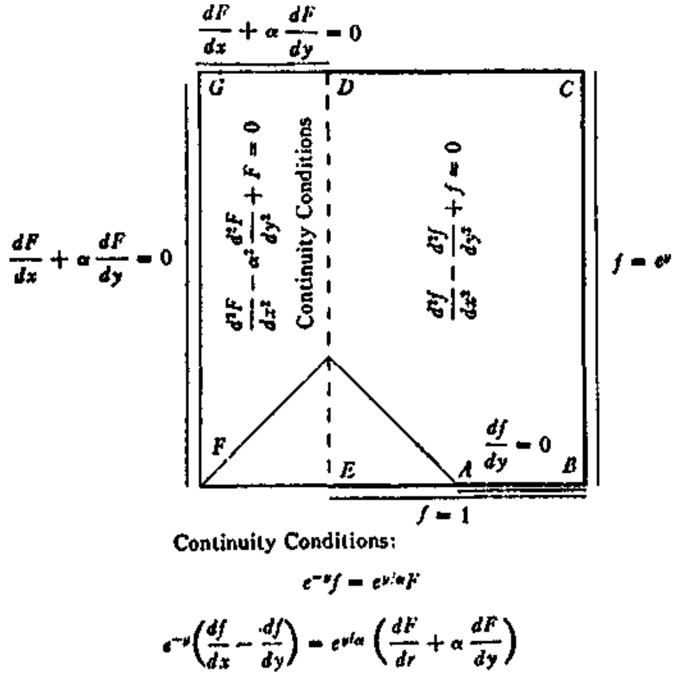


FIG. 4.

40. **Some further boundary problems.** The boundary value problem formulated in §38 does not exhaust the type of problems which occur in theory of radiative transfer in moving atmospheres. However, no progress has so far been made in the solution of these other problems. It may, therefore, be of particular interest to describe here the nature of these more complex boundary value problems.

A typical problem is to solve (see Fig. 4)

$$\frac{\partial^2 F}{\partial x^2} - \alpha^2 \frac{\partial^2 F}{\partial y^2} + F = 0 \qquad (\alpha = \text{constant}), \tag{271}$$

for F in the rectangular strip $FEDG$ $(0 \leq x \leq x_1,\ 0 \leq y \leq y_1)$, and

$$\frac{\partial^2 f}{\partial x^2} - \frac{\partial^2 f}{\partial y^2} + f = 0, \tag{272}$$

for f in the rectangular strip $EBCD$ ($x_2 \leqq x \leqq x_1$, $0 \leqq y \leqq y_1$) satisfying the following boundary conditions:

Along FG ($x=0$, $0 \leqq y \leqq y_1$) and GD ($y=y_1$, $0 \leqq x \leqq x_2$) relations of the form

$$\left(\frac{\partial F}{\partial x} + \alpha \frac{\partial F}{\partial y}\right)_{x=0} = \phi(y) \qquad (0 \leqq y \leqq y_1) \tag{273}$$

and

$$\left(\frac{\partial F}{\partial x} + \alpha \frac{\partial F}{\partial y}\right)_{y=y_1} = \psi(x) \qquad (0 \leqq x \leqq x_2) \tag{274}$$

are specified where $\phi(y)$ and $\psi(x)$ are two known functions. Along EB ($y=0$, $x_2 \leqq x \leqq x_1$) and BC ($x=x_1$, $0 \leqq y \leqq y_1$) f is given, while along the part AB ($x^* = x_2(1+\alpha) \leqq x \leqq x_1$, $y=0$) of the x-axis, the derivative $\partial f/\partial y$ is also given. And, finally, along ED ($x=x_2$, $0 \leqq y \leqq y_1$) certain "continuity conditions" of the type

$$f(x_2, y) = Q(y)F(x_2, y) \qquad (0 \leqq y \leqq y_1) \tag{275}$$

and

$$\left(\frac{\partial f}{\partial x} - \frac{\partial f}{\partial y}\right)_{x=x_2} = Q(y)\left(\frac{\partial F}{\partial x} + \alpha \frac{\partial F}{\partial y}\right)_{x=x_2} \quad (0 \leqq y \leqq y_1) \tag{276}$$

are specified, where $Q(y)$ is a known function.

It would be of considerable interest to know how such boundary value problems can be solved. (In Fig. 4 the particular boundary conditions which occur in a specific problem are indicated.)

41. **Concluding remarks.** In concluding, I may recall what I said at the beginning, namely, that the advance of a branch of theoretical physics often leads to the creation of a new mathematical discipline. I think it may be conversely said, with almost equal truth, that the creation of a new mathematical discipline is often the sign that the particular branch of theoretical physics has reached maturity. I hope I have given you the impression that theoretical astrophysics has now come of age.

Since the lecture was given in December, the mathematical theory of radiative transfer has advanced along several directions and in this Addendum we shall briefly summarize the results of these newer investigations. The particular sections of the lecture to which these advances refer are indicated by the numbering of the paragraphs which follow; however, §42 breaks new ground not covered by the lecture.

11a. Elliptically polarized radiation field. In §11 we outlined how the equations of transfer for a partially plane-polarized radiation field can be formulated and gave the explicit form of these equations for the particular case of Rayleigh scattering. It is not difficult to extend this discussion to include the case of a general elliptic polarization of the radiation field. In this latter case, we must consider, in addition to the intensities I_l and I_r in two directions at right angles to each other in the plane of the electric and the magnetic vectors, the two further quantities

$$(277) \quad U = (I_l - I_r) \tan 2\chi \text{ and } V = (I_l - I_r) \tan 2\beta \sec 2\chi$$

where χ denotes the inclination of the plane of polarization to the direction to which l refers and $-\pi/2 \leq \beta \leq +\pi/2$ is an angle the tangent of which is equal to the ratio of axes of the ellipse characterizing the state of polarization. (The sign of β depends on whether the polarization is right-handed (+) or left-handed (−).) And the rule of composition (due to Stokes) is that a mixture of several *independent* streams of polarized light is characterized by values of the parameters I_l, I_r, U and V which are the sums of the respective parameters of the individual streams. With this rule of composition, the equations of transfer for I_l, I_r, U and V can be formulated in terms of the basic laws of single scattering.

For the case of Rayleigh scattering, it is found that the equations of transfer for I_l, I_r and U are of exactly the same forms as when, only, plane polarization is contemplated; that is, in the case of Rayleigh scattering, the equations are *reducible*. Moreover, it is found that for Rayleigh scattering, V is simply scattered in accordance with a phase function $3/2 \cos \Theta$; and the exact solution of the equation for V therefore presents no difficulty.

23a. The functional equation relating the law of darkening and the scattering function for semi-infinite plane-parallel atmospheres. In §23 we have remarked on the remarkable relationships between the angular distributions of the emergent radiation in the problem with a constant net flux and the law of diffuse reflection. It has since been possible to trace the origin of these relationships: they arise simply in consequence of the invariance of the emergent radiation $I(0, \mu)$ (in the problem with a constant net flux) to the addition (or removal) of layers of arbitrary thickness to (or from) the atmosphere. The mathematical expression of this invariance is that the outward radiation $I(\tau, +\mu)$ $(0<\mu<1)$ at any level τ differs from the emergent radiation $I(0, \mu)$ only on account of the fact that at τ there is an inward directed radiation field specified by $I(\tau, -\mu')$ $(0<\mu'<1)$ which will be reflected by the atmosphere below τ by the law of diffuse reflection of a semi-infinite atmosphere. In other words, we must have

$$I(\tau, +\mu) = I(0, \mu) + \frac{1}{2\mu}\int_0^1 S^{(0)}(\mu, \mu')I(\tau, -\mu')d\mu', \tag{278}$$

where

$$S^{(0)}(\mu, \mu') = \frac{1}{2\pi}\int_0^1 S(\mu, \phi; \mu', \phi')d\phi' \tag{279}$$

is the azimuth independent term in the scattering function $S(\mu, \phi; \mu', \phi')$ (cf. equation (26)).

Differentiating equation (278) with respect to τ and passing to the limit $\tau=0$, we obtain

$$\left[\frac{dI(\tau, +\mu)}{d\tau}\right]_{\tau=0} = \frac{1}{2\mu}\int_0^1 S^{(0)}(\mu, \mu')\left[\frac{dI(\tau, -\mu')}{d\tau}\right]_{\tau=0} d\mu'. \tag{280}$$

On the other hand, from the equation of transfer

$$\mu\frac{dI}{d\tau} = I(\tau, \mu) - B(\tau, \mu) \tag{281}$$

where $B(\tau, \mu)$ is the source function appropriate for the problem (see equations I, II and VI) we conclude that

$$\left[\frac{dI(\tau, +\mu)}{d\tau}\right]_{\tau=0} = \frac{1}{\mu}\left[I(0, \mu) - B(0, \mu)\right], \tag{282}$$

and

$$\left[\frac{dI(\tau, -\mu')}{d\tau}\right]_{\tau=0} = \frac{1}{\mu'}B(0, -\mu'). \tag{283}$$

Now combining equations (280), (282) and (283), we obtain

$$I(0, \mu) = B(0, \mu) + \frac{1}{2}\int_0^1 S^{(0)}(\mu, \mu')B(0, -\mu')\frac{d\mu'}{\mu'}, \tag{284}$$

which is a functional equation relating $I(0, \mu)$ and $S^{(0)}(\mu, \mu')$; it can be shown that it is precisely in consequence of this equation that the relationship between $I(0, \mu)$ and $S^{(0)}(\mu, \mu')$ noticed in §23 arises.

35a. **Representation of $H(\mu)$ as a complex integral.** According to equations (102) and (110)

$$1-\sum_{j=1}^{n}\frac{a_j}{1-\mu_j^2/z^2}=\frac{1}{H(z)H(-z)} \tag{285}$$

where $H(z)$ is defined as usual in terms of the roots of the characteristic equation (see equation (49))

$$1=\sum_{j=1}^{n}\frac{a_j}{1-\mu_j^2k^2}. \tag{286}$$

The arguments leading to equation (285) (§20) are seen to be sufficiently general to establish the identity

$$\begin{aligned}1-2\sum_{j=1}^{n}\frac{a_j\Psi(\mu_j)}{1-\mu_j^2/z^2}&=\frac{1}{H(z)H(-z)}\\&=\frac{\prod_{\alpha}(1-k_\alpha^2z^2)}{\prod_{j}(1-z^2/\mu_j^2)},\end{aligned} \tag{287}$$

where $H(z)$ is now defined as in equation (126) in terms of the roots of the characteristic equation

$$1=2\sum_{j=1}^{n}\frac{a_j\Psi(\mu_j)}{1-k^2\mu_j^2}, \tag{288}$$

and $\Psi(\mu)$ has the same meaning as in equation (127).

Now let

$$G(x)=\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\log T(z)\,\frac{xdz}{z^2-x^2}, \tag{289}$$

where

$$T(z)=1-2z^2\sum_{j=1}^{n}\frac{a_j\Psi(\mu_j)}{z^2-\mu_j^2}=\frac{\prod_{\alpha}(1-k_\alpha^2z^2)}{\prod_{j}(1-z^2/\mu_j^2)}. \tag{290}$$

It is seen that defined in this manner $G(x)$ is regular for $R(x)>0$.

By evaluating the residue at the pole on the right

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\log\left(1+\frac{z}{a}\right)\frac{xdz}{z^2-x^2}=-\frac{1}{2}\log\left(1+\frac{x}{a}\right), \tag{291}$$

if $R(x)>0$ and $R(a)>0$. Similarly, by evaluating the residue at the pole on the left, we have

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\log\left(1-\frac{z}{a}\right)\frac{xdz}{z^2-x^2}=-\frac{1}{2}\log\left(1+\frac{x}{a}\right). \tag{292}$$

Hence,

$$\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\log\left(1-\frac{z^2}{a^2}\right)\frac{xdz}{z^2-x^2}=-\log\left(1+\frac{x}{a}\right). \tag{293}$$

Accordingly

$$\begin{aligned} G(x) &= \frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\left\{\sum_{\alpha=1}^{n}\log\left(1-k_\alpha^2z^2\right)-\sum_{j=1}^{n}\log\left(1-\frac{z^2}{\mu_j^2}\right)\right\}\frac{xdz}{z^2-x^2} \\ &= -\sum_{\alpha=1}^{n}\log\left(1+xk_\alpha\right)+\sum_{j=1}^{n}\log\left(1+\frac{z}{\mu_j}\right) \\ &= \log H(z). \end{aligned} \tag{294}$$

We have thus shown that

$$\log H(x)=\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\log T(z)\,\frac{xdz}{z^2-x^2}\,. \tag{295}$$

From the representation (295) of the H-function as a complex integral, it would appear that the solution of the *functional equation*

$$H(\mu)=1+\mu H(\mu)\int_0^1\frac{H(\mu')}{\mu+\mu'}\Psi(\mu')d\mu' \tag{296}$$

has the representation (cf. Theorem 2, §25)

$$\log H(\mu)=\frac{1}{2\pi i}\int_{-i\infty}^{i\infty}\log T(z)\,\frac{\mu dz}{z^2-\mu^2}\,, \tag{297}$$

where, now (see equation (290)),

$$T(z)=1-2z^2\int_0^1\frac{\Psi(\mu)d\mu}{z^2-\mu^2}\,. \tag{298}$$

Our arguments do not of course establish rigorously the representation (297) of the solution of equation (296). However, Professor E. C. Titchmarsh, with whom I have corresponded, has kindly communicated to me a rigorous demonstration of the representation (297) by one of his colleagues, Mr. M. M. Crum.

42. The theory of radiative transfer in atmospheres of finite optical thicknesses. In the lecture attention was directed almost exclusively to transfer problems in semi-infinite plane-parallel atmospheres. The extension of this theory to the study of the transfer of radiation in plane-parallel atmospheres of finite optical thicknesses raises problems of a higher order of difficulties; these difficulties arise principally from the circumstance that boundary conditions have to be *explicitly* satisfied on both sides of the atmosphere. Thus, if we consider an atmosphere of optical thickness τ_1 ($< \infty$), we are interested, for example, in solutions of the equations of transfer I–VII which satisfy the boundary conditions

$$I(0, -\mu) = 0 \qquad (0 < \mu < 1), \tag{299}$$

and

$$I(\tau_1, +\mu) = 0 \qquad (0 < \mu < 1). \tag{300}$$

However, in analogy with the theory of semi-infinite atmospheres, we may expect that in the case of finite atmospheres, also, the angular distributions of the emergent radiations can be expressed in terms of functions (like $H(\mu)$) which will be explicitly known in any finite approximation and which, in the limit of infinite approximation, will become solutions of functional equations of a standard form. It now appears that this reduction can in fact be achieved and that the pair of functional equations

$$X(\mu) = 1 + \mu \int_0^1 \frac{X(\mu)X(\mu') - Y(\mu)Y(\mu')}{\mu + \mu'} \Psi(\mu')d\mu', \tag{301}$$

and

$$Y(\mu) = e^{-\tau_1/\mu} + \mu \int_0^1 \frac{Y(\mu)X(\mu') - X(\mu)Y(\mu')}{\mu - \mu'} \Psi(\mu')d\mu', \tag{302}$$

where $\Psi(\mu)$ is an even polynomial in μ satisfying the condition

$$\int_0^1 \Psi(\mu)d\mu < \frac{1}{2}, \tag{303}$$

plays the same basic role in the theory of atmospheres of finite optical thicknesses as the functional equation

$$H(\mu) = 1 + \mu H(\mu) \int_0^1 \frac{H(\mu')}{\mu + \mu'} \Psi(\mu')d\mu' \tag{304}$$

played in the theory of semi-infinite atmospheres. And just as

$$H(\mu) = \frac{1}{\mu_1 \cdots \mu_n} \frac{\prod_i (\mu + \mu_i)}{\prod_\alpha (1 + k_\alpha \mu)}, \tag{305}$$

where the k_α's are the positive roots of the characteristic equation

$$1 = 2 \sum_{j=1}^{n} \frac{a_j \Psi(\mu_j)}{1 - k^2 \mu_j^2}, \tag{306}$$

provides a rational approximation to the solution of equation (304), so also the functions $X(\mu)$ and $Y(\mu)$ defined in the manner of the following equations provides an approximation to the solution of equations (301) and (302):

$$X(\mu) = \frac{(-1)^n}{\mu_1 \cdots \mu_n} \frac{1}{[C_0^2(0) - C_1^2(0)]^{1/2}} \frac{1}{W(\mu)} \cdot [P(-\mu) C_0(-\mu) - e^{-\tau_1/\mu} P(\mu) C_1(\mu)] \tag{307}$$

$$Y(\mu) = \frac{(-1)^n}{\mu_1 \cdots \mu_n} \frac{1}{[C_0^2(0) - C_1^2(0)]^{1/2}} \frac{1}{W(\mu)} \cdot [e^{-\tau_1/\mu} P(\mu) C_0(\mu) - P(-\mu) C_1(-\mu)] \tag{308}$$

where

$$P(\mu) = \prod_{i=1}^{n} (\mu - \mu_i), \qquad W(\mu) = \prod_{\alpha=1}^{n} (1 - k_\alpha^2 \mu^2), \tag{309}$$

$$C_0(\mu) = \sum_{2^{n-1} \text{ terms}} \epsilon_{2l}^{(0)} \frac{\prod_{i=1}^{2l} \prod_{m=1}^{n-2l} (k_{r_i} + k_{s_m})}{\prod_{i=1}^{2l} \prod_{m=1}^{n-2l} (k_{r_i} - k_{s_m})} \times \prod_{i=1}^{2l} \xi_{r_i}(1 + k_{r_i}\mu) \prod_{m=1}^{n-2l} \eta_{s_m}(1 - k_{s_m}\mu), \tag{310}$$

$$C_1(\mu) = \sum_{2^{n-1} \text{ terms}} \epsilon_{2l+1}^{(1)} \frac{\prod_{i=1}^{2l+1} \prod_{m=1}^{n-2l-1} (k_{r_i} + k_{s_m})}{\prod_{i=1}^{2l+1} \prod_{m=1}^{n-2l-1} (k_{r_i} - k_{s_m})} \times \prod_{i=1}^{2l+1} \xi_{r_i}(1 + k_{r_i}\mu) \prod_{m=1}^{n-2l-1} \eta_{r_i}(1 - k_{s_m}\mu). \tag{311}$$

In equations (310) and (311), $(r_1, \cdots, r_j)$ and $(s_1, \cdots, s_{n-j})$ are j, respectively, $n-j$ *distinct* integers from the set $(1, 2, \cdots, n)$,

$$(312) \qquad \epsilon_{2l}^{(0)} = \begin{cases} +1 \text{ for even integers of the form } 4l, \\ -1 \text{ for even integers of the form } 4l+2, \end{cases}$$

$$(313) \qquad \epsilon_{2l+1}^{(1)} = \begin{cases} +1 \text{ for odd integers of the form } 4l+1, \\ -1 \text{ for odd integers of the form } 4l+3, \end{cases}$$

and

$$(314) \qquad \xi_\alpha = e^{k_\alpha \tau_1/2} P(-1/k_\alpha) \quad \text{and} \quad \eta_\alpha = e^{-k_\alpha \tau_1/2} P(+1/k_\alpha) \qquad (\alpha = 1, \cdots, n).$$

Finally, it should be noted that the definitions of $C_0(\mu)$ and $C_1(\mu)$ according to equations (307) and (308) are valid only in even orders of approximation; in odd orders the role of C_0 and C_1 should be interchanged.

Moreover, there exist also functional equations for the scattering and the transmission functions for the problem of diffuse reflection and transmission by atmospheres of finite optical thicknesses. These equations arise from general invariances of the type considered in §§28, 29 and 23a and lead to a whole new class of systems of functional equations which can all be reduced to the solution of pairs of functional equations of the form (301) and (302). It is therefore apparent that the study of the transfer of radiation in atmospheres of finite optical thicknesses will lead to the development of a mathematical theory at least as extensive as the one described in the lecture in the context of semi-infinite atmospheres.

References

The following papers and monographs substantially form the basis of the report given in the lecture. The list is not intended to be a complete bibliography of the subject of the transfer of radiation in stellar atmospheres.

V. A. Ambarzumian
C. R. (Doklady) Acad. Sci. URSS. vol. 38 (1943) p. 257.
Russian Astronomical Journal vol. 19 (1942).
Journal of Physics of the Academy of Sciences of the U.S.S.R. vol. 8 (1944) p. 65.

S. Chandrasekhar
Astrophysical Journal vol. 99 (1944) p. 180; vol. 100 (1944) pp. 76, 117, 355; vol. 101 (1945) pp. 95, 328, 348; vol. 102 (1945) p. 402; vol. 103 (1946) pp. 165, 351; vol. 104 (1946) pp. 110, 191; vol. 105 (1947) pp. 151, 164, 424, 435, 441, 461. (Further papers in this series are in course of publication in the same journal.)

Reviews of Modern Physics vol. 17 (1945) p. 138.
Proc. Cambridge Philos. Soc. vol. 42 (1946) p. 250.

E. Hopf
Mathematical problems of radiative equilibrium, Cambridge Mathematical Tracts, No. 31, Cambridge, England, 1934.

G. Münch
Astrophysical Journal vol. 104 (1946) p. 87.

Lord Rayleigh
Scientific papers of Lord Rayleigh, vol. 1, Cambridge, England, 1899, pp. 87, 104, 518.

Sir George Stokes
Trans. Cambridge Philos. Soc. vol. 9 (1852) p. 399.

M. Tuberg
Astrophysical Journal vol. 103 (1946) p. 145.

G. C. Wick
Zeitschrift für Physik vol. 120 (1943) p. 702.

Yerkes Observatory, University of Chicago

THE ILLUMINATION AND POLARIZATION OF THE SUNLIT SKY ON RAYLEIGH SCATTERING

S. Chandrasekhar and Donna D. Elbert

CONTENTS

1. INTRODUCTION

Since 1871 when Lord Rayleigh first accounted for the principal features of the brightness and polarization of the sunlit sky in terms of the laws of scattering now associated with his name, it has been generally recognized that a problem of fundamental importance both for meteorological optics and for theories of planetary illumination is the following:

A parallel beam of radiation in a given state of polarization is incident on a plane-parallel atmosphere of optical thickness τ_1 in some specified direction. Each element of the atmosphere scatters radiation in accordance with Rayleigh's laws. It is required to find the distribution of intensity and polarization of the light diffusely transmitted by the atmosphere below $\tau = \tau_1$ and of the light diffusely reflected by the atmosphere above $\tau = 0$.

In the theory of planetary illumination one is principally interested in the reflected light while in the theory of sky illumination one is similarly interested in the transmitted light. In this paper we shall be concerned only with the latter.

It is clear that an exact treatment of the foregoing problem in the theory of diffuse reflection and transmission will require the formulation and solution of the appropriate equations of radiative transfer. This was accomplished six years ago[1] and the theory is described and briefly illustrated in the book *Radiative Transfer* (Oxford, 1950) by one of us. A general account of the theory, together with a comparison of its predictions with observations particularly those relating to the polarization of the sunlit sky, was published in 1951 in a brief article.[2]

In this paper we shall present the calculations which we have made (at intervals) during the past five years with the object of giving the theory a concrete form. At one time it was our hope to present our calculations based on the exact mathematical solution of the problem with detailed comparisons not only with the available observational data but also with the calculations of the earlier investigators based on approximations of various kinds. But pressure of time and circumstance have forced us to abandon this plan: this paper will be restricted to giving the results of our calculations with only such comparisons with observations as seemed to us of particular interest.

[1] S. Chandrasekhar, On the radiative equilibrium of a stellar atmosphere. XXII. (V. Rayleigh scattering), *Astrophys. Jour.* **107**: 199, 1947.

[2] S. Chandrasekhar and Donna Elbert, Polarization of the sunlit sky, *Nature* **167**: 51, 1951.

2. THE SOLUTION OF THE FUNDAMENTAL PROBLEM

As we have stated, the problem in the theory of diffuse reflection and transmission formulated in § 1 has been exactly solved; the solution is given in *Radiative Transfer* (§§ 69–73). We shall not describe in any detail how the solution was obtained. But a few explanatory remarks on the parameters in terms of which the solution was obtained and on the structure of the solution itself may be useful in the present connection.

Since on Rayleigh's laws light gets partially plane-polarized whenever it is scattered, it is clear that in formulating the equations of radiative transfer we must allow for the partial plane-polarization of the radiation field. Now to describe a radiation field which is partially plane-polarized we need three parameters to specify the intensity, the degree of polarization, and the plane of polarization. It would scarcely be expected that one could include such diverse quantities as an intensity, a ratio, and an angle in any satisfactory way in formulating the basic equations of the problem. It appears that for these latter purposes the most convenient representation of polarized light is a set of parameters first introduced by Stokes[3] in 1852. The meaning of these parameters for a partially plane-polarized beam is simple: Let l and r refer to two arbitrarily chosen directions at right-angles to one another in the plane transverse to the direction of propagation of the beam. The intensity $I(\psi)$ in a direction making an angle ψ (measured clock-wise) to the direction of l can be expressed in the form

$$I(\psi) = I_l \cos^2\psi + I_r \sin^2\psi + \tfrac{1}{2} U \sin 2\psi. \quad (1)$$

The coefficients I_l, I_r and U in this representation are the Stokes parameters. In terms of these parameters the angle χ which the plane of polarization makes with the direction l and the degree of polarization, δ, are given by

$$\tan 2\chi = U/(I_l - I_r) \quad (2)$$

[3] G. G. Stokes, On the composition and resolution of streams of polarized light from different sources, *Trans. Camb. Phil. Soc.* **9**: 399, 1852.

and

$$\delta = (I_l - I_r)\sec 2\chi/(I_l + I_r). \tag{3}$$

The additive property of the Stokes parameters which makes them so convenient for treating problems of radiative transfer is evident from the representation (1): If two independent streams of polarized light are mixed, then the Stokes parameter characterizing the mixture is the sum of the Stokes parameters of the individual streams.

In terms of the Stokes parameters a law of scattering is specified by a matrix, since an elementary act of scattering results in a linear transformation of the parameters. Consequently, by considering the intensity as a vector $\boldsymbol{I}$ with the components I_l, I_r and U (where l and r from now on refer to directions parallel and perpendicular, respectively, to the meridian through the point under consideration and in the plane containing the directions of the beam and of the normal to the plane of stratification of the atmosphere) and by replacing the "phase function" commonly introduced to describe the angular distribution of the scattered radiation by a *phase matrix*, $\boldsymbol{P}$, we can formulate the basic equation of transfer without any difficulty of principle. In this manner we find that the equation we have to solve is[4]

$$\mu\frac{d\boldsymbol{I}(\tau,\mu,\varphi)}{d\tau} = \boldsymbol{I}(\tau,\mu,\varphi) - \frac{1}{4\pi}\int_{-1}^{+1}\int_0^{2\pi}\boldsymbol{P}(\mu,\varphi;\mu',\varphi')\boldsymbol{I}(\tau,\mu',\varphi')\,d\mu'\,d\varphi' - \tfrac{1}{4}e^{-\tau/\mu_0}\boldsymbol{P}(\mu,\varphi;-\mu_0,\varphi_0)\boldsymbol{F}. \tag{4}$$

where

$$\boldsymbol{F} = (F_l, F_r, F_U) \tag{5}$$

is the (Stokes) vector which represents the parallel beam of radiation incident on the atmosphere in the direction $(-\mu_0, \varphi_0)$: πF_l, πF_r and πF_U denote the net fluxes per unit area normal to the beam in the three Stokes parameters. Further, in equation (4) μ denotes the cosine of the angle to the outward normal and the azimuthal angle. And, finally, for the case of Rayleigh scattering the phase matrix $\boldsymbol{P}(\mu,\varphi;\mu',\varphi')$ has the explicit form:

$$\boldsymbol{P}(\mu,\varphi;\mu',\varphi') = Q[\boldsymbol{P}^{(0)}(\mu,\mu') + (1-\mu^2)^{\frac{1}{2}}(1-\mu'^2)^{\frac{1}{2}}\boldsymbol{P}^{(1)}(\mu,\varphi;\mu',\varphi') + \boldsymbol{P}^{(2)}(\mu,\varphi;\mu',\varphi')], \tag{6}$$

where

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \tag{7}$$

$$\boldsymbol{P}^{(0)}(\mu,\mu') = \frac{3}{4}\begin{pmatrix} 2(1-\mu^2)(1-\mu'^2)+\mu^2\mu'^2 & \mu^2 & 0 \\ \mu'^2 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{8}$$

$$\boldsymbol{P}^{(1)}(\mu,\varphi;\mu',\varphi') = \frac{3}{4}\begin{pmatrix} 4\mu\mu'\cos(\varphi'-\varphi) & 0 & 2\mu\sin(\varphi'-\varphi) \\ 0 & 0 & 0 \\ -2\mu'\sin(\varphi'-\varphi) & 0 & \cos(\varphi'-\varphi) \end{pmatrix} \tag{9}$$

and

$$\boldsymbol{P}^{(2)}(\mu,\varphi;\mu',\varphi') = \frac{3}{4}\begin{pmatrix} \mu^2\mu'^2\cos 2(\varphi'-\varphi) & -\mu^2\cos 2(\varphi'-\varphi) & \mu^2\mu'\sin 2(\varphi'-\varphi) \\ -\mu'^2\cos 2(\varphi'-\varphi) & \cos 2(\varphi'-\varphi) & -\mu'\sin 2(\varphi'-\varphi) \\ -\mu\mu'^2\sin 2(\varphi'-\varphi) & \mu\sin 2(\varphi'-\varphi) & \mu\mu'\cos 2(\varphi'-\varphi) \end{pmatrix}. \tag{10}$$

The solution of equation (4) appropriate to the problem on hand must satisfy the boundary conditions

$$\left.\begin{aligned} &\boldsymbol{I}(0,-\mu,\varphi) \equiv 0 \quad (0<\mu\leqslant 1,\ 0\leqslant\varphi\leqslant 2\pi) \\ &\text{and} \\ &\boldsymbol{I}(\tau_1,+\mu,\varphi) \equiv 0 \quad (0<\mu\leqslant 1,\ 0\leqslant\varphi\leqslant 2\pi), \end{aligned}\right\} \tag{11}$$

since there is no diffuse radiation in any inward direction at $\tau = 0$ and in any outward direction at $\tau = \tau_1$. And the solution to the problem of diffuse reflection and transmission will be completed when we specify the angular distribution and the state of polarization of the diffuse light which emerges from $\tau = 0$ and $\tau = \tau_1$.

The laws of diffuse reflection and transmission by a plane-parallel atmosphere are generally expressed (*cf. Radiative Transfer*, 44) in terms of a *scattering matrix*, $\boldsymbol{S}(\mu,\varphi;\mu_0,\varphi_0)$ and a *transmission matrix*, $\boldsymbol{T}(\mu,\varphi;\mu_0,\varphi_0)$ such that the reflected and the transmitted intensities are given by

$$\boldsymbol{I}(0;\mu,\varphi;\mu_0,\varphi_0) = \frac{1}{4\mu}\boldsymbol{S}(\mu,\varphi;\mu_0,\varphi_0)\boldsymbol{F}$$

and

$$\boldsymbol{I}(\tau_1;-\mu,\varphi;\mu_0,\varphi_0) = \frac{1}{4\mu}\boldsymbol{T}(\mu,\varphi;\mu_0,\varphi_0)\boldsymbol{F}. \tag{12}$$

Our problem, then, is to specify $\boldsymbol{S}$ and $\boldsymbol{T}$ for an atmosphere scattering radiation in accordance with Rayleigh's laws.

[4] The derivation of the equations which follow will be found in *Radiative Transfer* § 16: 35–45.

The elements of S and T are clearly functions of the four variables μ, φ, μ_0 and φ_0 in addition, of course, to the optical thickness τ_1 which may, however, be treated as a parameter. If the variables μ, φ, μ_0 and φ_0 were not separable the problem of tabulating S and T may indeed be considered as impracticable. But the essential feature of the solution for S and T (which we shall presently write down) which makes the problem a practicable one is that S and T involve only four pairs of functions $X_l(\mu)$, $Y_l(\mu)$; $X_r(\mu)$, $Y_r(\mu)$; $X^{(1)}(\mu)$, $Y^{(1)}(\mu)$; and $X^{(2)}(\mu)$, $Y^{(2)}(\mu)$ all of the single variable μ. Further, these four pairs of functions belong to a general class (the X- and Y-functions) which satisfy a simultaneous pair of integral equations of the form[5]

$$X(\mu) = 1 + \mu \int_0^1 \frac{\Psi(\mu')}{\mu + \mu'} \times [X(\mu)X(\mu') - Y(\mu)Y(\mu')]d\mu' \quad (13)$$

and

$$Y(\mu) = e^{-\tau_1/\mu} + \mu \int_0^1 \frac{\Psi(\mu')}{\mu - \mu'} \times [Y(\mu)X(\mu') - X(\mu)Y(\mu')]d\mu', \quad (14)$$

where the *characteristic function* $\Psi(\mu)$ is in problems of radiative transfer, an even polynomial in μ satisfying the condition

$$\int_0^1 \Psi(\mu)\,d\mu \leqslant \tfrac{1}{2}. \quad (15)$$

The case when equality occurs in (15) (the so-called *conservative case*) is special: The solutions of equations (13) and (14) are, then, no longer unique; they form instead a one-parameter family. In conservative cases one therefore defines what are called *standard solutions* which have the property:

$$\int_0^1 X(\mu)\Psi(\mu)\,d\mu = 1$$

and $\quad (16)$

$$\int_0^1 Y(\mu)\Psi(\mu)\,d\mu = 0.$$

As we have already stated, the solutions for S and T (for an atmosphere scattering radiation in accordance with Rayleigh's laws) involve only four pairs of X- and Y-functions: X_l, Y_l; X_r, Y_r; $X^{(1)}$, $Y^{(1)}$; $X^{(2)}$, $Y^{(2)}$; and the characteristic functions in terms of which these are defined are:

$$\Psi_l(\mu) = \tfrac{3}{4}(1-\mu^2);$$
$$\Psi_r(\mu) = \tfrac{3}{8}(1-\mu^2),$$
$$\Psi^{(1)}(\mu) = \tfrac{3}{8}(1-\mu^2)(1+2\mu^2)$$

and

$$\Psi^{(2)}(\mu) = \tfrac{3}{16}(1+\mu^2)^2, \quad (17)$$

respectively. The function $\Psi_l(\mu)$ belongs to the conservative class; accordingly, in this case we define $X_l(\mu)$, $Y_l(\mu)$ as the standard solutions having the property

$$\tfrac{3}{4}\int_0^1 X_l(\mu)(1-\mu^2)\,d\mu = 1$$

and $\quad (18)$

$$\int_0^1 Y_l(\mu)(1-\mu^2)\,d\mu = 0.$$

After these explanatory remarks we shall now write down the solutions for S and T given in *Radiative Transfer:*

The scattering and the transmission matrices allow a decomposition into azimuth independent and azimuth dependent terms in the same manner as the phase matrix [equation (6)] and have the forms:

$$S(\mu,\varphi;\mu_0,\varphi_0) = Q[\tfrac{3}{4}S^{(0)}(\mu;\mu_0) + (1-\mu^2)^{\frac{1}{2}}(1-\mu_0^2)^{\frac{1}{2}}S^{(1)}(\mu,\varphi;\mu_0,\varphi_0) + S^{(2)}(\mu,\varphi;\mu_0,\varphi_0)] \quad (19)$$

and

$$T(\mu,\varphi;\mu_0,\varphi_0) = Q[\tfrac{3}{4}T^{(0)}(\mu;\mu_0) + (1-\mu^2)^{\frac{1}{2}}(1-\mu_0^2)^{\frac{1}{2}}T^{(1)}(\mu,\varphi;\mu_0,\varphi_0) + T^{(2)}(\mu,\varphi;\mu_0,\varphi_0)]. \quad (20)$$

The dependence of the azimuth dependent terms $(S^{(1)}, T^{(1)})$ and $(S^{(2)}, T^{(2)})$ on $\varphi_0 - \varphi$ are essentially the same as $P^{(1)}$ and $P^{(2)}$; indeed, we have

$$\left(\frac{1}{\mu_0}+\frac{1}{\mu}\right)S^{(i)} = [X^{(i)}(\mu)X^{(i)}(\mu_0) - Y^{(i)}(\mu)Y^{(i)}(\mu_0)]P^{(i)}(\mu,\varphi;-\mu_0,\varphi_0)$$

and

$$\left(\frac{1}{\mu_0}-\frac{1}{\mu}\right)T^{(i)} = [Y^{(i)}(\mu)X^{(i)}(\mu_0) - X^{(i)}(\mu)Y^{(i)}(\mu_0)]P^{(i)}(-\mu,\varphi;-\mu_0,\varphi_0) \quad (i=1,2). \quad (21)$$

In contrast, the solutions for the azimuth independent terms $S^{(0)}$ and $T^{(0)}$ are very complicated. They are given by

$$\left(\frac{1}{\mu_0}+\frac{1}{\mu}\right)S^{(0)}(\mu;\mu_0) = \begin{pmatrix}\psi(\mu) & 2^{\frac{1}{2}}\phi(\mu) & 0\\ \chi(\mu) & 2^{\frac{1}{2}}\zeta(\mu) & 0\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\psi(\mu_0) & \chi(\mu_0) & 0\\ 2^{\frac{1}{2}}\phi(\mu_0) & 2^{\frac{1}{2}}\zeta(\mu_0) & 0\\ 0 & 0 & 0\end{pmatrix} - \begin{pmatrix}\xi(\mu) & 2^{\frac{1}{2}}\eta(\mu) & 0\\ \sigma(\mu) & 2^{\frac{1}{2}}\theta(\mu) & 0\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}\xi(\mu_0) & \sigma(\mu_0) & 0\\ 2^{\frac{1}{2}}\eta(\mu_0) & 2^{\frac{1}{2}}\theta(\mu_0) & 0\\ 0 & 0 & 0\end{pmatrix} \quad (22)$$

[5] For the theory of the X- and Y-functions see *Radiative Transfer*, chap. VIII.

and

$$\left(\frac{1}{\mu_0}-\frac{1}{\mu}\right) T^{(0)}(\mu;\mu_0) = \begin{pmatrix} \xi(\mu) & 2^{\frac12}\eta(\mu) & 0 \\ \sigma(\mu) & 2^{\frac12}\theta(\mu) & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \psi(\mu_0) & \chi(\mu_0) & 0 \\ 2^{\frac12}\phi(\mu_0) & 2^{\frac12}\zeta(\mu_0) & 0 \\ 0 & 0 & 0 \end{pmatrix}$$
$$- \begin{pmatrix} \psi(\mu) & 2^{\frac12}\phi(\mu) & 0 \\ \chi(\mu) & 2^{\frac12}\zeta(\mu) & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \xi(\mu_0) & \sigma(\mu_0) & 0 \\ 2^{\frac12}\eta(\mu_0) & 2^{\frac12}\theta(\mu_0) & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (23)$$

where ψ, ϕ, χ, etc., are eight functions expressible in terms of the two pairs of X- and Y-functions, X_l, Y_l and X_r, Y_r in the forms

$$\begin{aligned}
\psi(\mu) &= \mu[\nu_1 Y_l(\mu) - \nu_2 X_l(\mu)], \\
\xi(\mu) &= \mu[\nu_2 Y_l(\mu) - \nu_1 X_l(\mu)], \\
\phi(\mu) &= (1+\nu_4\mu) X_l(\mu) - \nu_3\mu Y_l(\mu), \\
\eta(\mu) &= (1-\nu_4\mu) Y_l(\mu) + \nu_3\mu X_l(\mu), \\
\chi(\mu) &= (1-u_4\mu) X_r(\mu) + u_3\mu Y_r(\mu) + Q(u_4-u_3)\mu^2[X_r(\mu) - Y_r(\mu)], \\
\sigma(\mu) &= (1+u_4\mu) Y_r(\mu) - u_3\mu X_r(\mu) - Q(u_4-u_3)\mu^2[X_r(\mu) - Y_r(\mu)], \\
\zeta(\mu) &= \tfrac12\mu[\nu_1 Y_r(\mu) - \nu_2 X_r(\mu)] + \tfrac12 Q(\nu_2-\nu_1)\mu^2[X_r(\mu) - Y_r(\mu)], \\
\theta(\mu) &= \tfrac12\mu[\nu_2 Y_r(\mu) - \nu_1 X_r(\mu)] - \tfrac12 Q(\nu_2-\nu_1)\mu^2[X_r(\mu) - Y_r(\mu)],
\end{aligned} \quad (24)$$

where the constants ν_1, ν_2, ν_3, ν_4, u_3, u_4 and Q are to be determined by the following formulae:

$$\begin{gathered}
\nu_2+\nu_1 = 2\Delta_1(\kappa_1\delta_1 - \kappa_2\delta_2); \quad \nu_2-\nu_1 = 2\Delta_2(\kappa_1\delta_1 - \kappa_2\delta_2), \\
\nu_4+\nu_3 = \Delta_1(d_1\kappa_1 - d_0\kappa_2); \quad \nu_4-\nu_3 = \Delta_2[c_1\delta_1 - c_0\delta_2 - 2Q(d_0\delta_1 - d_1\delta_2)], \\
u_4+u_3 = \Delta_1(c_1\delta_1 - c_0\delta_2); \quad u_4-u_3 = \Delta_2(d_1\kappa_1 - d_0\kappa_2), \\
\Delta_1 = (d_0\delta_1 - d_1\delta_2)^{-1}; \quad \Delta_2 = [c_0\kappa_1 - c_1\kappa_2 - 2Q(d_1\kappa_1 - d_0\kappa_2)]^{-1}, \\
Q = (c_0 - c_2)[(d_0 - d_2)\tau_1 + 2(d_1 - d_3)]^{-1}, \\
c_0 = A_0 + B_0 - \frac{8}{3}; \quad d_0 = A_0 - B_0 - \frac{8}{3}, \\
c_n = A_n + B_n; \quad d_n = A_n - B_n; \quad \kappa_n = \alpha_n + \beta_n; \quad \delta_n = \alpha_n - \beta_n \qquad (n = 1, 2, 3, \cdots),
\end{gathered} \quad (25)$$

α_n, β_n, A_n and B_n are the moments of order n of X_l, Y_l, X_r and Y_r, respectively. (It may be recalled here that X_l and Y_l are the standard solutions for the case.)

In the theory of the illumination of the sky we are interested in the transmitted light in the case of incident natural light. In this latter case $F_l = F_r = \frac12 F$ (where πF denotes the net flux of the incident natural light) and $F_U = 0$. The equations governing the intensity and polarization of the sky as witnessed by an observer at $\tau = \tau_1$ in these circumstances readily follow from the solutions already given; thus by setting $F = \frac12(F_l, F_r, 0)$ in equation (12) and combining equations (20), (21), and (23) appropriately, we find:

$$\begin{aligned}
I_l(\tau_1; -\mu, \varphi; \mu_0, \varphi_0) = \tfrac{3}{32}[&\{\psi(\mu_0) + \chi(\mu_0)\}\xi(\mu) + 2\{\phi(\mu_0) + \zeta(\mu_0)\}\eta(\mu) - \{\xi(\mu_0) + \sigma(\mu_0)\}\psi(\mu) \\
&- 2\{\theta(\mu_0) + \eta(\mu_0)\}\phi(\mu) + 4\mu\mu_0(1-\mu^2)^{\frac12}(1-\mu_0^2)^{\frac12}\{X^{(1)}(\mu_0)Y^{(1)}(\mu) - Y^{(1)}(\mu_0)X^{(1)}(\mu)\}\cos(\varphi_0 - \varphi) \\
&- \mu^2(1-\mu_0^2)\{X^{(2)}(\mu_0)Y^{(2)}(\mu) - Y^{(2)}(\mu_0)X^{(2)}(\mu)\}\cos 2(\varphi_0 - \varphi)]\frac{F\mu_0}{\mu - \mu_0}, \\
I_r(\tau_1; -\mu, \varphi; \mu_0, \varphi_0) = \tfrac{3}{32}[&\{\psi(\mu_0) + \chi(\mu_0)\}\sigma(\mu) + 2\{\phi(\mu_0) + \zeta(\mu_0)\}\theta(\mu) - \{\xi(\mu_0) + \sigma(\mu_0)\}\chi(\mu) \\
&- 2\{\theta(\mu_0) + \eta(\mu_0)\}\zeta(\mu) + (1-\mu_0^2)\{X^{(2)}(\mu_0)Y^{(2)}(\mu) - Y^{(2)}(\mu_0)X^{(2)}(\mu)\}\cos 2(\varphi_0 - \varphi)]\frac{F\mu_0}{\mu - \mu_0}
\end{aligned}$$

and

$$\begin{aligned}
U(\tau_1; -\mu, \varphi; \mu_0, \varphi_0) = \tfrac{3}{16}[&2(1-\mu^2)^{\frac12}(1-\mu_0^2)^{\frac12}\mu_0\{X^{(1)}(\mu_0)Y^{(1)}(\mu) - Y^{(1)}(\mu_0)X^{(1)}(\mu)\}\sin(\varphi_0 - \varphi) \\
&- \mu(1-\mu_0^2)\{X^{(2)}(\mu_0)Y^{(2)}(\mu) - Y^{(2)}(\mu_0)X^{(2)}(\mu)\}\sin 2(\varphi_0 - \varphi)]\frac{F\mu_0}{\mu - \mu_0}.
\end{aligned} \quad (26)$$

3. THE EFFECT OF REFLECTION BY THE GROUND

Before we can apply the solution for S and T given in § 2 to the problem of the illumination of the sky, we must consider the effect of the ground at $\tau = \tau_1$. The solution for S and T given in § 2 was derived on the assumption that at $\tau = \tau_1$ there is no diffuse radiation in the outward direction [*cf.* equation (11)]. The presence of the ground will alter this. However, if the law of reflection by the ground is specified then it is not a difficult matter to relate the solution of the problem when there is a ground to the solution of the problem when there is no ground. This reduction is particularly simple if the ground reflects according to Lambert's law with a certain albedo λ_0; that is, if the light reflected by the ground is unpolarized and uniform in the outward hemisphere independently of the state of polarization and the angular distribution of the incident light, and if, further, the outward flux of the reflected light is always a certain fixed fraction, λ_0, of the inward flux of the radiation incident on the surface. Under these latter circumstances it can be shown (*Radiative Transfer*, § 73) that the effect of the ground is to increase the diffuse intensities emergent at $\tau = 0$ and directed inward at $\tau = \tau_1$ by amounts $I^*(0; \mu, \varphi)$ and $I^*(\tau_1; -\mu, \varphi)$ given by

$$I^*(0; \mu, \varphi) = \frac{\lambda_0\mu_0}{2(1-\lambda_0\bar{s})}\,\mathbf{\Gamma}(\mu; \mu_0)F \qquad (27)$$

and

$$I^*(\tau_1; -\mu, \varphi) = \frac{\lambda_0\mu_0}{2(1-\lambda_0\bar{s})}\,\mathbf{A}(\mu; \mu_0)F, \qquad (28)$$

where $\bar{s}$ is a constant (to be defined presently) and

$$\mathbf{\Gamma}(\mu; \mu_0) = \begin{bmatrix} \gamma_l(\mu)\gamma_l(\mu_0) & \gamma_l(\mu)\gamma_r(\mu_0) & 0 \\ \gamma_r(\mu)\gamma_l(\mu_0) & \gamma_r(\mu)\gamma_r(\mu_0) & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad (29)$$

and

$$\mathbf{A}(\mu; \mu_0) = \begin{bmatrix} \{1-\gamma_l(\mu)\}\gamma_l(\mu_0) & \{1-\gamma_l(\mu)\}\gamma_r(\mu_0) & 0 \\ \{1-\gamma_r(\mu)\}\gamma_l(\mu_0) & \{1-\gamma_r(\mu)\}\gamma_r(\mu_0) & 0 \\ 0 & 0 & 0 \end{bmatrix}. \qquad (30)$$

In equations (29) and (30) $\gamma_l(\mu)$ and $\gamma_r(\mu)$ are two functions which are related to X_l, Y_l, X_r and Y_r by

$$\gamma_l(\mu) = \tfrac{3}{8}Q(\nu_2 - \nu_1)(d_0 - d_2)[X_l(\mu) + Y_l(\mu)], \qquad (31)$$

and

$$\begin{aligned} \gamma_r(\mu) = \tfrac{3}{8}Q(d_0 - d_2) & \\ \times [(u_4 - u_3)\{X_r(\mu) + Y_r(\mu)\} & \\ - u_5\mu\{X_r(\mu) + Y_r(\mu)\}], & \qquad (32) \end{aligned}$$

where

$$u_5 = \Delta_2(c_0\kappa_1 - c_1\kappa_2), \qquad (33)$$

and the remaining constants have the same meanings as in equations (25). Finally, the constant $\bar{s}$ in equations (27) and (28) is given by

$$\begin{aligned} \bar{s} = 1 - \tfrac{3}{8}Q(d_0 - d_2) & \\ \times [(\nu_2 - \nu_1)\kappa_1 + (u_4 - u_3)c_1 - u_5 d_2]. & \qquad (34) \end{aligned}$$

Again, when the incident light is natural the corrections which have to be made to the intensities given by equations (26) to allow for a ground surface at $\tau = \tau_1$ which reflects according to Lambert's law with an albedo λ_0 are given by

$$\begin{aligned} I_l^*(\tau_1; -\mu, \varphi; \mu_0, \varphi_0) & \\ = \frac{\lambda_0}{4(1-\lambda_0\bar{s})}\{\gamma_l(\mu_0) + \gamma_r(\mu_0)\} & \\ \times \{1 - \gamma_l(\mu)\}\mu_0 F, & \\ I_r^*(\tau_1; -\mu, \varphi; \mu_0, \varphi_0) & \\ = \frac{\lambda_0}{4(1-\lambda_0\bar{s})}\{\gamma_l(\mu_0) + \gamma_r(\mu_0)\} & \\ \times \{1 - \gamma_r(\mu)\}\mu_0 F, & \\ U^*(\tau_1; -\mu, \varphi; \mu_0, \varphi_0) = 0. & \qquad (35) \end{aligned}$$

4. DESCRIPTION OF THE TABLES*

The solution of the fundamental problem in the theory of the illumination of the sky given in the two preceding sections was obtained some six years ago. A detailed examination of its predictions had to await the tabulation of the basic eight functions X_l, Y_l, X_r, Y_r, $X^{(1)}$, $Y^{(1)}$, $X^{(2)}$, and $Y^{(2)}$ of the variable μ for various values of τ_1. This tabulation has now been completed for $\tau_1 = 0.05$, 0.10, 0.15, 0.20, 0.25, 0.50, and 1.00 with the cooperation of the Watson Scientific Computing Laboratory (New York).[6] The solutions were obtained by a direct process of iteration applied to the governing integral equations. The

*The tables are omitted from this volume. See, instead, Kinsell L. Coulson, Jitendra V. Dave, and Zdenek Sekera, *Tables Related to Radiation Emerging from a Planetary Atmosphere with Rayleigh Scattering* (Berkeley: University of California Press, 1960).

[6] While these calculations were in progress, Dr. Z. Sekera initiated a similar program at the Department of Meteorology of the University of California at Los Angeles in cooperation with the Institute for Numerical Analysis of the National Bureau of Standards at Los Angeles. Their Report No. 3 (prepared by the Air Material Command, Air Force Cambridge Research Center) provides some calculations for $\tau_1 = 0.15$, 0.25, and 1.0. Their calculations, while they are much less extensive than ours, do provide a valuable check.

iterations were started with the solutions in the corrected second approximation described in *Radiative Transfer*, Chapter VIII [§ 60, see particularly equations (117), (118), and (120)]. The corrected second approximations were computed by one of us (D. D. E.) at the Yerkes Observatory. The iterations were carried out at the Watson Scientific Laboratory with IBM pluggable sequence relay calculators by Miss Ann Franklin to whom and to Dr. Wallace Eckert we are very greatly indebted. The functions obtained after the iterations showed, however, a certain "raggedness" between $\mu = 0.9$ and 1.0. The solutions have therefore been "smoothed" by plotting the deviations of the iterated solutions from the corrected second approximations. Table 1 presents these smoothed solutions. While the solutions have been tabulated to five decimals the last place is definitely not reliable. But it is expected that if the tabulated solutions are rounded to one less place the solution may be trusted to two or three units in the surviving place. The solutions for $\tau_1 \leqslant 0.20$ are very probably more accurate than this while for $\tau_1 = 0.5$ and 1.0 they may be less accurate. Nevertheless, the solutions are given to five places since the functions as tabulated do have smooth differences and as such they can be used for further iterations to improve their accuracy if the need for it should arise.

The moments α_n, β_n, A_n and B_n of order n of X_l, Y_l, X_r and Y_r, respectively, which are needed in the evaluation of the various terms of S and T are given in table 2. The theory of the X- and Y-functions leads to a number of identical relations which must exist between their moments; these relations among the moments listed in table 2 have been verified within the accuracy of the tabulated values. Table 2 also includes the values of the constants ν_1, ν_2, ν_3, ν_4, u_3, u_4 and Q which occur in the definitions of the functions ψ, ϕ, etc. [*equations* (24) and (25)]; the value of the constant $\bar{s}$ [equation (34)] which occurs in the expressions for the ground corrections [equations (27) and (28)] is also listed in this table.

According to equations (22) and (23) the calculation of S and T in a given case can be most easily carried out in terms of the auxiliary functions ψ, ϕ, χ, ζ, ξ, η, σ and θ. These functions computed with the aid of tables 1 and 2 are given in table 3. Similarly, table 4 gives the functions $\gamma_l(\mu)$ and $\gamma_r(\mu)$ which are needed to allow for reflection by the ground [*cf.* equations (31) and (32)].

With the basic functions tabulated we can calculate the theoretical illumination and polarization of the sky on Rayleigh scattering for plane-parallel atmospheres. To illustrate the use of tables 1–4 we have made some model calculations to which the remaining tables are devoted.

The most extensive calculations were made for $\tau_1 = 0.15$; this is approximately the value of the optical thickness at λ4500A. For $\tau_1 = 0.15$, angles of incidence (or, zenith distances) $\theta_0 = 90°$, 85.4°, 76.1°, 58.7°, 50.1°, 43.9°, 36.9°, 19.95°, and 0° (corresponding to $\mu_0 = 0$, 0.08, 0.24, 0.52, 0.64, 0.72, 0.80, 0.94, and 1.00) and azimuthal differences $\varphi_0 - \varphi$ (denoted, simply, by φ in the tables) $= 0°(10°)90°$ were considered. The results of the calculations are summarized in table 5; it gives the total intensity, $I_l + I_r$, in units of F, the inclination, χ, of the plane of polarization with the meridian through the direction of observation and the degree of polarization, δ. Less extensive calculations were made for $\tau_1 = 0.10$ and 0.20. For these two values of the optical thicknesses the total intensity $(I_l + I_r)$ and the degree of polarization, δ, were found only in the principal meridian $(\varphi_0 - \varphi = 0°)$ containing the sun and in the plane at right-angles $(\varphi_0 - \varphi = 90°)$ for zenith distances $\theta_0 = 90°$, 80.8°, 60°, 30.7°, and 0° (corresponding to $\mu_0 = 0$, 0.16, 0.50, 0.86, and 1.00). The results of the calculations are given in tables 6 and 7.

In the calculations presented in tables 5–7 no allowance has been made for the reflection by the ground surface. This is taken into account in tables 8–10 in accordance with equations (35). Again, the most extensive calculations were made for $\tau_1 = 0.15$. For angles of incidence corresponding to $\mu_0 = 0$, 0.24, 0.52, 0.64, 0.72, 0.80, and 1.00 the effect of a ground reflecting according to Lambert's law for two values of the albedo, λ_0, on the intensity and the degree of polarization in the principal meridian $(\varphi_0 - \varphi = 0°)$ was determined. For $\mu_0 = 0.64$ the effect on the intensity for $\varphi_0 - \varphi = 0°(10°)90°$ and $\lambda_0 = 0.10$ was determined; and for the same angle of incidence the effect on the intensity for $\varphi_0 - \varphi = 90°$ was also found for $\lambda_0 = 0.25$. The results of all these calculations are given in table 8. For $\tau_1 = 0.10$ and 0.20 the effect of ground reflection on the intensity and polarization in the principal meridian is illustrated for $\lambda_0 = 0.10$ and 0.20 and for angles of incidence corresponding to $\mu_0 = 0$, 0.16, 0.50, 0.86, and 1.00. The results of these calculations are given in tables 9 and 10.

Table 11 gives the positions of the neutral points. The results given in this table will be described and discussed in the following section.

Finally the supplementary table 12 gives the functions ψ, ϕ, χ, ζ, ξ, η, σ, θ, $X^{(1)}$, $Y^{(1)}$, $X^{(2)}$, $Y^{(2)}$, γ_l, and γ_r for $\tau_1 = 0.01(0.01)0.20(0.05)0.50(0.10)1.0$ and for representative values of μ. At the head of the table for each value τ_1, the value of $\bar{s}$ (needed for the evaluation of the ground correction) is also given. The values of the functions listed in this table are *not* based on the solutions of the basic X- and Y-functions derived from the integral equations they satisfy; they are based, instead, on the corrected second approximations (*cf. Radiative Transfer*, § 60) for the relevant X- and Y-functions. However, for $\tau_1 \leqslant 0.25$ the table should suffice to calculate S and T to well within a fraction of a per cent; for the larger values of τ_1 accuracy within a few per cent may be expected. But one could, if one wished, obtain from these tables values of considerably higher precision by differencing the values of the functions for $\tau_1 = 0.05$, 0.10, 0.15, 0.20, 0.25, 0.50, and 1.0 given in tables 3 and 12 and interpolating among these differences to estimate the corrections (to the values given by the corrected second approximation) for any other intermediate value of τ_1. With these supplementary tables, then, the theory developed and described in *Radiative Transfer* has been, finally, brought to a point where it is capable of giving numerical values for any of the desired quantities under most conditions in which they are likely to be of interest.

5. THE POLARIZATION OF THE SUNLIT SKY: THE THEORY OF THE NEUTRAL POINTS AND LINES

As we have already stated in the introductory section we shall not attempt in this paper any detailed comparison between the calculations presented here and those of earlier investigators based on approximations of various kinds. But an exception might be made with regard to the quantitative explanation which the present theory affords for the phenomena associated with the *neutral points* of Arago, Babinet, and Brewster. The phenomena in question are these:

The neutral points are the points of zero polarization; from symmetry we should, of course, expect them to occur in the principal meridian. And for a long time it has been known that there are in general two such neutral points. For angles of incidence not exceeding 70° these neutral points occur between 0° and 20° above and below the sun; these are the neutral points of Babinet and Brewster, respectively. But when the sun is low, the neutral point occurs about 20° above the anti-solar point in the opposite sky: this is the Arago point. These facts concerning the

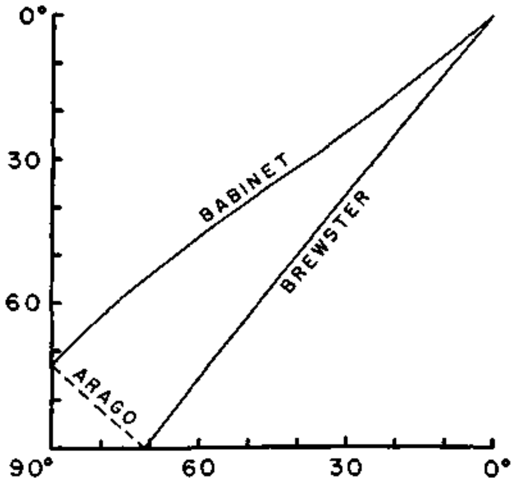


Fig. 1. Calculated positions of the neutral points for various angles of incidence for an atmosphere of optical thickness $\tau_1 = 0.10$. The abscissa gives the zenith distance of the sun and the ordinate gives the corresponding positions of the neutral points. The Arago point occurs on the side of the horizon opposite the sun; to emphasize this its position on the sky is indicated by the dashed curve.

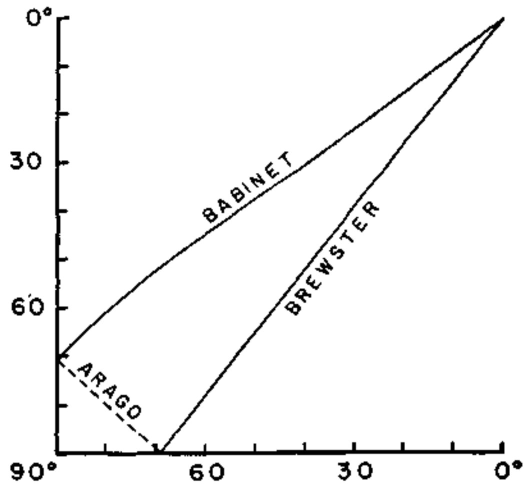


Fig. 2. Calculated positions of the neutral points for various angles of incidence for an atmosphere of optical thickness $\tau_1 = 0.15$. The abscissa and the ordinate have otherwise the same meanings as in fig. 1.

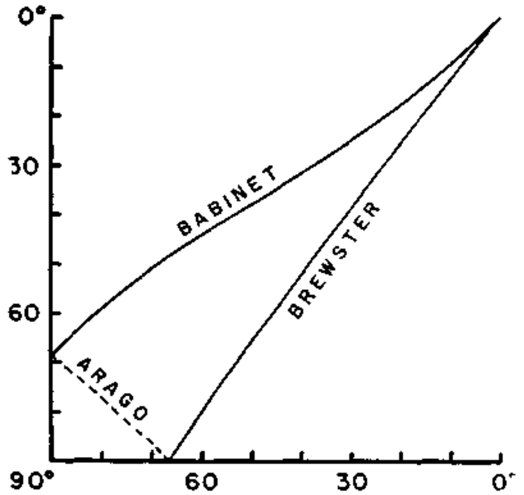


Fig. 3. Calculated positions of the neutral points for various angles of incidence for an atmosphere of optical thickness $\tau_1 = 0.20$. The abscissa and the ordinate have otherwise the same meanings as in fig. 1.

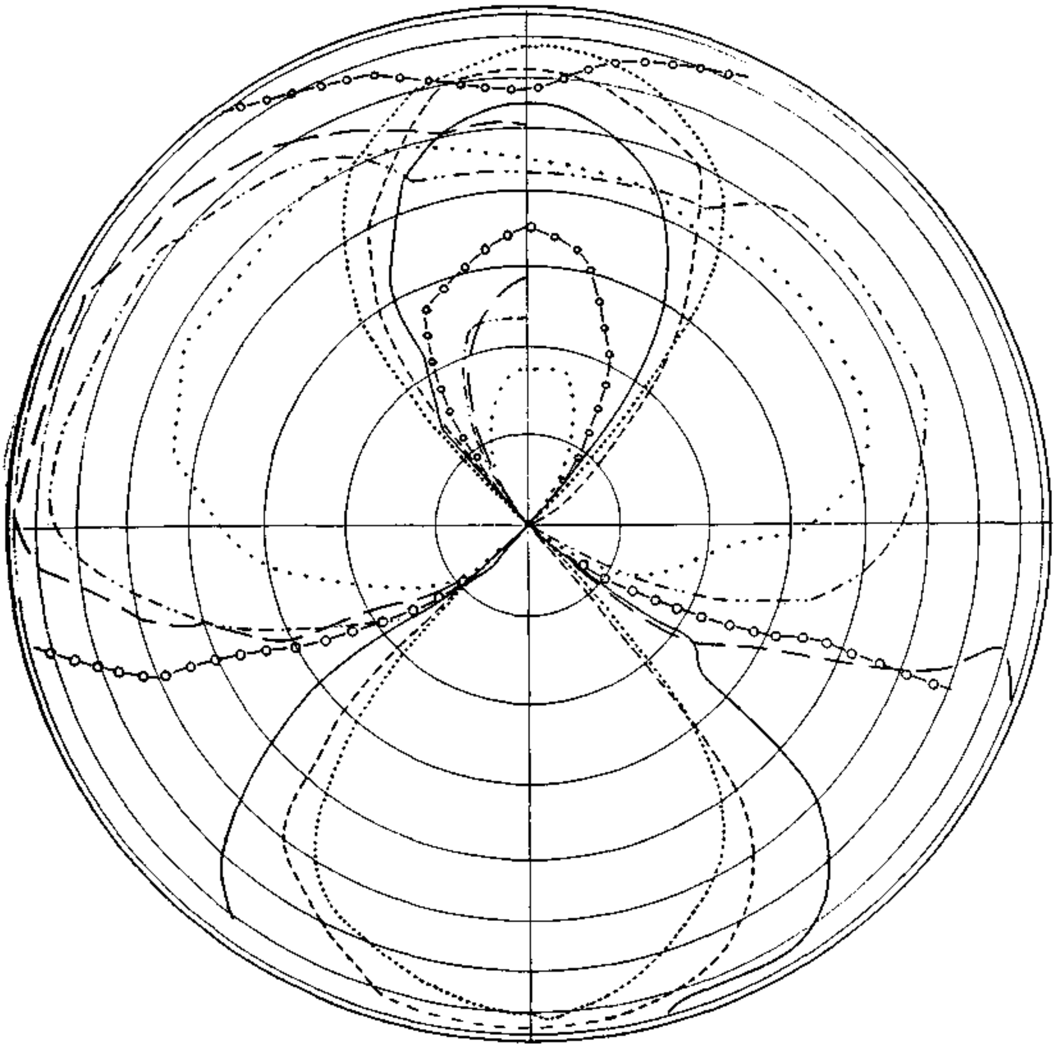


FIG. 4. Dorno's observations of the neutral lines on May 17, 1917, at Davos. The various curves were determined during the following times when the zenith distance of the sun varied by the amounts given:

··············	$7^p\,19$– $7^p\,28$	$\theta_0 = 87°\,10'$– $0°$
- - - - - - - - -	$6^p\,13$– $6^p\,33$	$\theta_0 = 78°\,14'$–$81°\,30'$
————	$5^p\,14$– $5^p\,34$	$\theta_0 = 68°\,20'$–$71°\,44'$
-o-o-o-o-o-	$3^p\,15$– $3^p\,42$	$\theta_0 = 48°\,12'$–$52°\,40'$
— — — —	$9^a\;\;9$– $9^a\,34$	$\theta_0 = 44°\,21'$–$40°\,33'$
—··—··—	$9^a\,44$–$10^a\,11$	$\theta_0 = 39°\;\;5'$–$35°\,26'$
· · · · · · · · ·	$12^p\,56$– $1^p\,14$	$\theta_0 = 29°\,50'$–$31°\,26'$

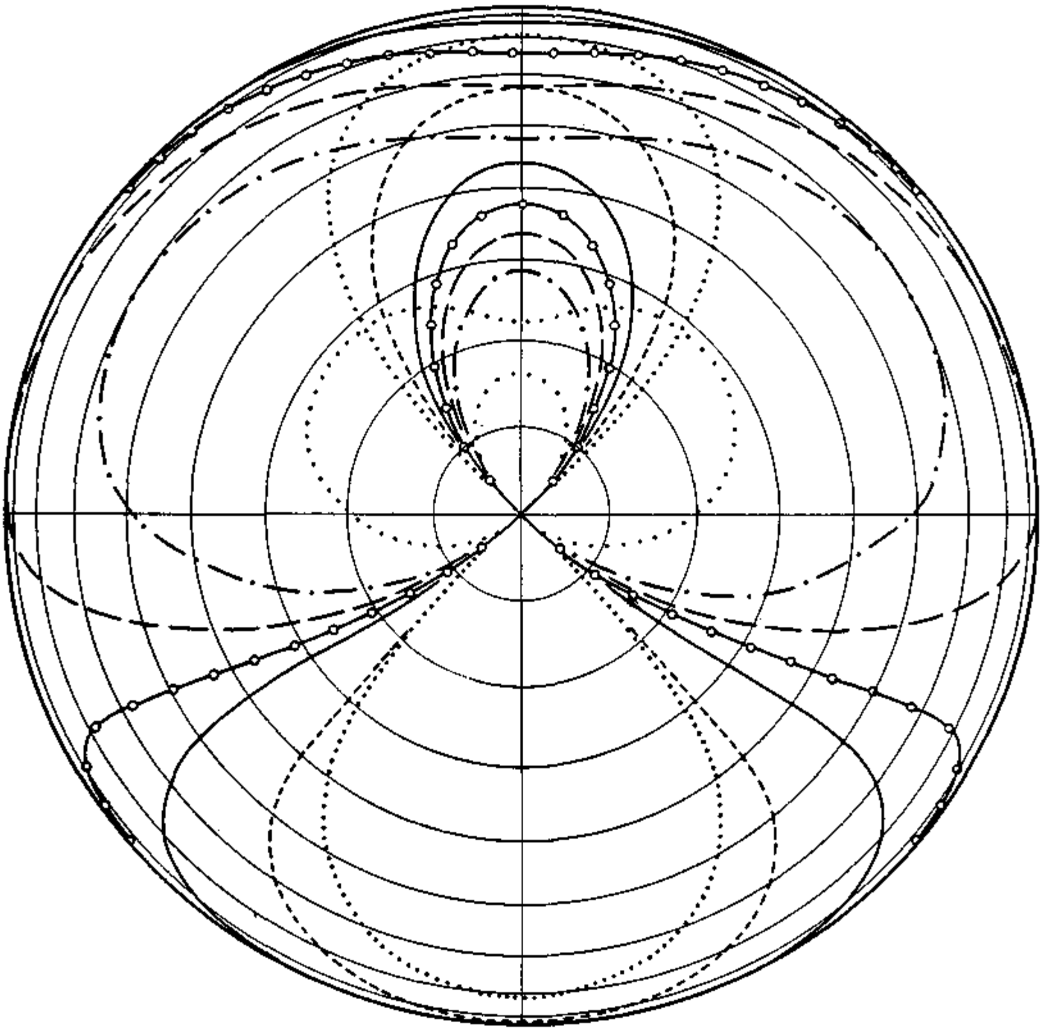


FIG. 5. The neutral lines as predicted by the theory. The various curves refer to the following zenith distances of the sun:

·········· = 90°;	———— = 58.7°;	— — — — = 43.9°;
·+·+·+·+· = 76.1°;	–o–o–o–o– = 50.2°;	—·—·— = 36.9°;
	········ = 19.9°.	

The curves in this figure which roughly correspond to Dorno's observations are marked similarly.

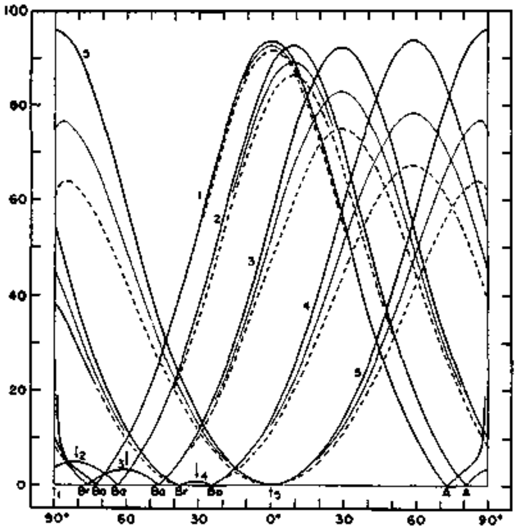


FIG. 6. Variation of the degree of polarization in the principal meridian for various angles of incidence for an atmosphere of optical thickness $\tau_1 = 0.10$. The abscissa gives the zenith distance and the ordinate gives the degree of polarization in per cent. The curves marked 1, 2, 3, 4, and 5 represent the variation for the angles of incidence $\theta_0 = 90°, 80.8°, 60.0°, 30.7°$ and $0°$, respectively. The thick solid curves are obtained before any ground corrections have been applied. The dashed curves are obtained if we allow for a ground reflecting according to Lambert's law with an albedo $\lambda_0 = 0.20$. The thin intermediate curves are obtained if $\lambda_0 = 0.10$. The positions of the neutral points are also indicated.

points of zero polarization should be contrasted with what should be expected on the laws of single scattering, namely that the polarization should tend to zero as we approach the direction towards the sun.

During the nineteenth century the existence of these neutral points and their behavior with the direction of the sun were regarded as among the most remarkable phenomena in meteorological optics. As such they were studied with great care and attention and by none more than Carl Dorno whose monumental work on the subject[7] contains a wealth of information painstakingly gathered. Dorno not only observed the neutral points on the principal meridian, but he also investigated in detail the continuation of these neutral points over the entire hemisphere along what he called the neutral lines. These lines separate the regions of positive from the regions of negative polarization;[8] they show a remarkable dependence on the direction of the sun. A diagram representing Dorno's principal results is reproduced in figure 4. It will be seen from this figure that when the sun is nearly on the horizon the neutral line connects the Babinet and the Arago points by a closed symmetrical curve of the shape of a lemniscate: this is the so-called lemniscate of Busch. As the sun rises the lemniscate becomes more and more asymmetrical and when the angle of incidence exceeds about 70° the lemniscate opens out and a part of the locus appears on the horizon below the sun and passes through the Brewster point which has now risen. The neutral line consists of two such separated curves until the angle of incidence becomes about 45° when the opposite ends join together to form a closed re-entrant curve. For still smaller angles of incidence the neutral line collapses towards the center and finally reduces to a point when the sun is at the zenith.

We shall now see how this entire range of phenomena associated with the neutral points and lines are faithfully reproduced by our calculations. In

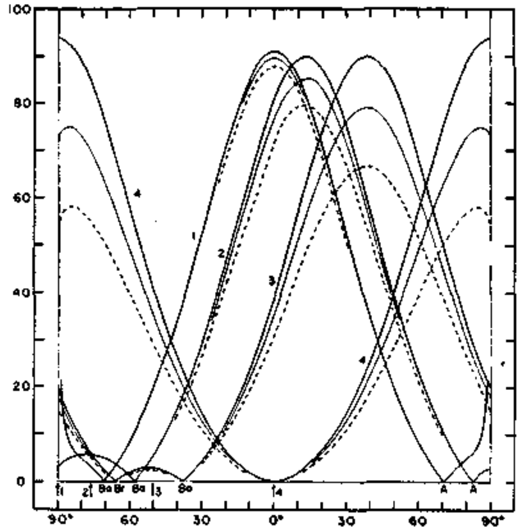


FIG. 7. Variation of the degree of polarization in the principal meridian for various angles of incidence for an atmosphere of optical thickness $\tau_1 = 0.15$. The abscissa gives the zenith distance and the ordinate gives the degree of polarization in per cent. The curves marked 1, 2, 3 and 4 represent the variation for the angles of incidence $\theta_0 = 90°, 76.1°, 50.2°$ and $0°$, respectively. The thick solid curves are obtained before any ground corrections have been applied. The dashed curves are obtained if we allow for a ground reflecting according to Lambert's law with an albedo $\lambda_0 = 20.5$. The thin intermediate curves are obtained if $\lambda_0 = 0.10$. The positions of the neutral points are also indicated.

[7] C. Dorno, Himmelshelligkeit, Himmelspolarisation und Sonnenintensität in Davos 1911 bis 1918, *Veröffentl. Preuss. Met. Inst.*, No. 303, Berlin, 1919.

[8] The polarization is assumed positive if I_l is less than I_r and negative if the reverse is true. On this convention the polarization is negative on the principal meridian between the neutral points of Babinet and Brewster; these points, therefore, separate the regions of positive from the regions of negative polarization on this meridian. The neutral lines of Dorno do the same for the entire hemisphere.

table 11 we have collected all the information contained in tables 5–10 regarding the points at which the polarization changes sign for various zenith distances of the sun; they are further illustrated in figures 1, 2, 3, and 5. Considering first figures 1–3 (which refer to the principal meridian), we observe that the calculations predict the occurrence of the neutral points as observed. In particular it will be noticed that the Brewster point sets when the angle of incidence is about 70°: its dependence on the values of the optical thickness in the range of interest is not pronounced. Also as the Brewster point sets the Arago point rises in the opposite sky. And as the sun sinks lower, the Arago point continues to rise until, when the sun sets, the Babinet and the Arago points are both at an equal

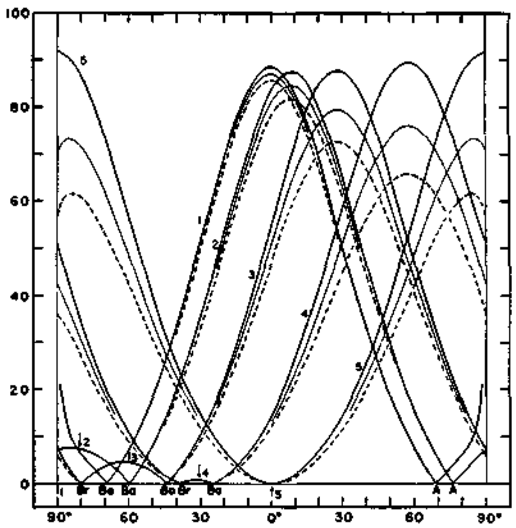


FIG. 8. Variation of the degree of polarization in the principal meridian for various angles of incidence for an atmosphere of optical thickness $\tau_1 = 0.20$. The abscissa gives the zenith distance and the ordinate gives the degree of polarization in per cent. The curves marked 1, 2, 3, 4, and 5 represent the variation for the angles of incidence $\theta_0 = 90°$, 80.8°, 60.0°, 30.7°, and 0°, respectively. The thick solid curves are obtained before any ground corrections have been applied. The dashed curves are obtained if we allow for a ground reflecting according to Lambert's law with an albedo $\lambda_0 = 0.20$. The thin intermediate curves are obtained if $\lambda_0 = 0.10$. The positions of the neutral points are also indicated.

elevation of about 20° (it varies from 17° to 21° for τ_1 in the range $0.10 \leqslant \tau_1 \leqslant 0.20$) from the horizon.

With the calculations for the different values of $\varphi_0 - \varphi$ for $\tau_1 = 0.15$ given in table 5, we can draw an entire system of calculated neutral lines. This has been done in figure 5. Comparing it with the results of Dorno's observations (fig. 4) we observe how well the two sets of curves match.

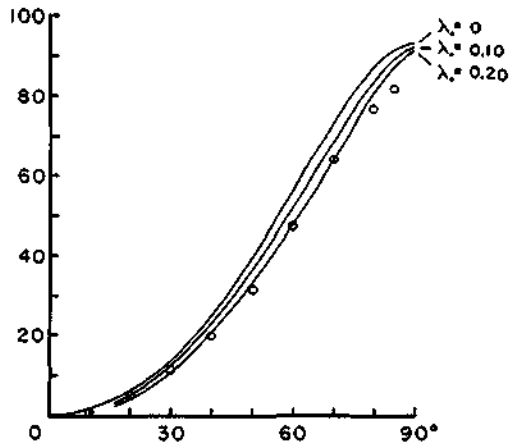


FIG. 9. Variation of the degree of polarization at the zenith for various angles of incidence and for an atmosphere of optical thickness $\tau_1 = 0.10$. The curve $\lambda_0 = 0$ allows for no ground reflection. The other two curves allow for reflection by a ground with the albedos indicated. The circles represent the observations of Tousey and Hulburt for values of τ_1 and λ_0 estimated at 0.10 and 0.20, respectively.

Turning next to the calculated degrees of polarization, we have illustrated its variation on the principal meridian for various angles of incidence and for the three values of the optical thickness for which calculations have been made in figures 6, 7, and 8. It will be noticed from these figures that the effect of a ground surface with an albedo even as high as 0.25 does not

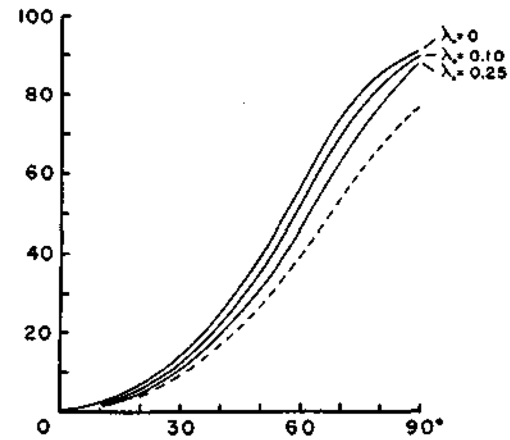


FIG. 10. Variation of the degree of polarization at the zenith for various angles of incidence and for an atmosphere of optical thickness $\tau_1 = 0.15$. The curve $\lambda_0 = 0$ allows for no ground reflection. The other two curves allow for reflection by a ground with the albedos indicated. The dashed curve represents the observations of Tichanowsky.

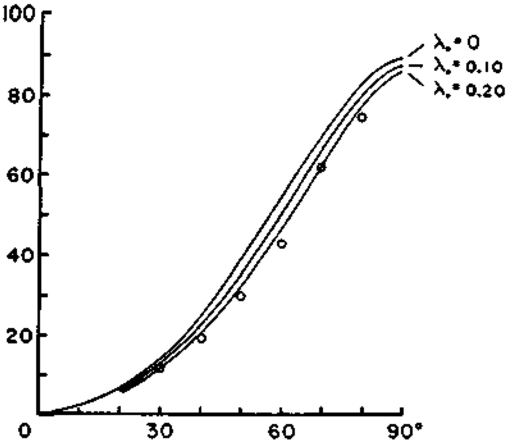


FIG. 11. Variation of the degree of polarization at the zenith for various angles of incidence and for an atmosphere of optical thickness $\tau_1 = 0.20$. The curve $\lambda_0 = 0$ allows for no ground reflection. The other two curves allow for reflection by a ground with the albedos indicated. The circles represent the observations of Richardson and Hulburt for a value of τ_1 estimated at 0.3.

make any essential difference to the predicted positions of the neutral points. This independence of the neutral points (and lines) on ground reflection is not difficult to understand. As is well known, the laws of Rayleigh scattering give the maximum polarization for the scattered light; all other laws give much less polarization. A ground reflecting according to Lambert's law can, therefore, hardly compete with Rayleigh scattering for producing polarization. The effect of reflection by the ground is therefore essentially one of adding a component of natural light to the polarized light already present. This last statement is, of course, not strictly true. The difference between γ_l and γ_r is precisely a measure of the polarization of the ground contribution to the sky brightness; but as is apparent from the tables the difference between γ_l and γ_r is generally very small. For these same reasons we should not expect that the direction of maximum polarization will be influenced by ground reflection. This is, indeed, the case: the direction of maximum polarization always occurs in a direction which is very nearly at right-angles to the direction of the sun. In contrast, ground reflection has a very pronounced effect on the *degree* of maximum polarization: In the absence of ground corrections, the maximum polarization varies between 94 and 89 per cent depending on τ_1 but is very nearly independent of the altitude of the sun (though there is a slight increase for higher altitudes). However, when the effect of ground reflection is taken into account the maximum polarization shows a very marked decrease with altitude. Observations do show such a behavior and we conclude that this is probably due to the effect of ground reflection.

Finally, in figures 9, 10, and 11 we have compared the degree of polarization at the zenith for various angles of incidence with the observations of Tichanowsky,[9] Tousey and Hulburt,[10] and Richardson and Hulburt.[11] It will be seen that the agreement between the theory and the observations is as good as one might expect.

In concluding this paper we should again like to record our thanks to Dr. Wallace Eckert and Miss Ann Franklin of the Watson Scientific Computing Laboratory for their generous co-operation in obtaining the solutions for the basic X- and Y-functions.

[9] J. J. Tichanowsky, Resultate der Messungen der Himmelspolarisation in verschiedenen Spektrumabschnitten, *Meteor. Ztschr.* 43: 288, 1926.

[10] R. Tousey and E. O. Hulburt, Brightness and polarization of the daylight sky at various altitudes above sea level, *Jour. Opt. Soc. Amer.* 37: 78, 1947.

[11] R. A. Richardson and E. O. Hulburt, Sky-brightness measurements near Bocaiuva, Brazil, *Jour. Geophys. Research* 54: 215, 1949.

SOME REMARKS ON THE NEGATIVE HYDROGEN ION AND ITS ABSORPTION COEFFICIENT

S. Chandrasekhar
Yerkes Observatory
Received June 28, 1944

ABSTRACT

Some remarks on the quantum theory of the negative hydrogen ion are made, and attention is drawn to certain facts which make the evaluation of its continuous absorption coefficient a problem of extreme difficulty.

This paper will consist of a few disconnected remarks on the quantum theory of the negative hydrogen ion.

1. *The wave function for the ground state of* H^-.—Since the discovery of the stability of the negative ion of hydrogen by Bethe[1] and Hylleraas[2] and the recognition of its astrophysical importance by Wildt,[3] attempts[4] have been made to determine the electron affinity of hydrogen with as high a precision as possible. In these latter attempts the energy of the ground state is determined by applications of the Ritz principle, using forms for the wave functions suggested by Hylleraas' successful treatment of the ground state of the helium atom. Thus Williamson's six-parameter wave function is of exactly the same form as the "best" wave function of Hylleraas for helium. Similarly, Henrich's eleven-parameter wave function includes terms beyond those used by Williamson. While there can hardly be any doubt that Henrich's value for the electron affinity of 0.747 electron volts can be in error by more than a fraction of 1 per cent, the relatively weak convergence of the entire process (cf., e.g., Table 7 in Henrich's paper) leaves one with a suspicion that the formal analogy between the atomic configurations of H^- and He has perhaps been taken too literally. From one point of view it would seem that the structures of these two atoms must be very different indeed; for, while helium is a stable closed structure, the negative hydrogen ion is an open structure which exists principally on account of incomplete screening and polarization (see below). This suggests that it might be possible to obtain better representations of the true wave function by seeking forms which will explicitly take into account this difference. That such attempts may not prove unsuccessful is suggested by the following preliminary considerations.

As is well known, the success of Hylleraas' investigations on helium is due principally to the circumstance that a wave function of the form

$$\psi = e^{-\alpha(r_1+r_2)}\ , \tag{1}$$

which ascribes a hydrogen-like wave function to each of the electrons in a suitably screened Coulomb field, already provides a good first approximation. More particularly the wave function of the form (1), which gives the lowest energy, is

$$\psi = e^{-(Z-[5/16])(r_1+r_2)}\ . \tag{2}$$

(In the foregoing equation r_1 and r_2 are measured in units of the Bohr radius. Similarly, in the rest of the paper we shall systematically use Hartree's atomic units.)

[1] *Zs. f. Phys.*, **57**, 815, 1929.

[2] *Zs. f. Phys.*, **60**, 624, 1930.

[3] *Ap. J.*, **89**, 295, 1939.

[4] R. E. Williamson, *Ap. J.*, **96**, 438, 1942; and L. R. Henrich, *Ap. J.*, **99**, 59, 1943.

When $Z = 1$, the wave function (2) predicts an energy $E = -0.473$, which actually makes H^- an unstable structure and is in error by fully 12 per cent. In other words, the first approximation, which is so satisfactory for *He*, fails completely for H^-. That this should happen is not surprising in view of our earlier remarks concerning the difference between the two atoms. On the other hand, it would appear that in contrast to *He* a natural first approximation for H^- is to ignore the screening of one of the electrons and adjust the screening constant for the second electron only. In other words, the starting-point for H^- should rather be a wave function of the form

$$\psi = e^{-r_1 - br_2} + e^{-r_2 - br_1} , \tag{3}$$

where b is the screening constant for the second electron. More generally, we may write

$$\psi = e^{-ar_1 - br_2} + e^{-ar_2 - br_1} , \tag{4}$$

where a and b are constants to be appropriately chosen. The Ritz principle applied to a wave function of the form (4) showed that the lowest value of energy is attained when

$$a = 1.03925 \quad \text{and} \quad b = 0.28309 . \tag{5}$$

The corresponding value for the energy is

$$E_1 = -0.51330 , \tag{6}$$

which predicts the stability of H^-. Moreover, in confirmation of our expectations it is seen that, while the inner electron is practically unscreened, the outer one is screened considerably and to the extent of 72 per cent. In view of this, it appears that a good second approximation may be provided by considering a wave function of the form

$$\psi = (e^{-ar_1 - br_2} + e^{-ar_2 - br_1})(1 + cr_{12}), \tag{7}$$

where a, b, and c are constants to be so chosen as to lead to a minimum value for the energy. It is found that with

$$a = 1.07478 , \quad b = 0.47758 , \quad \text{and} \quad c = 0.31214 \tag{8}$$

we minimize the energy integral and obtain for it the value

$$E_2 = -0.52592 . \tag{9}$$

This value for the energy, while inferior to those predicted by Williamson (0.5265) and Henrich (0.5276), is substantially better than the value 0.5253 given by the three-parameter wave function of Bethe and Hylleraas.

An interesting feature of the wave function (7) with the constants as given by equation (8) is that the inclusion of the term r_{12} reduces the screening of the outer electron from 0.72 to 0.52. This relatively large reduction in the screening is due to the strong polarizability of the hydrogen atom. Indeed, according to equations (6) and (9) we may say that the electron affinity of hydrogen is due about equally to the incomplete screening of the nucleus and to the polarization of the hydrogenic core.

The foregoing discussion suggests that it might be profitable to improve the wave function (7) by including further terms. This would be particularly useful for estimating the inherent uncertainty in the absorption cross-sections derived from different wave functions, all of which predict (within limits) the same value for energy. The practical importance for carrying out such a discussion will be apparent from our remarks in the following section.

2. *The absorption cross-sections for* H^-.—The calculations of the absorption cross-sections which have been carried out so far (Massey and Bates; Williamson; Henrich)

are based on two approximations. The first consists in the use of the wave function for describing the bound state the ones derived from the minimal calculations and the second, in the use of a plane wave representation of the ejected outgoing electron. The validity or otherwise of these approximations will depend upon whether the principal contributions to the matrix element,

$$\mu = \int \Psi_d (r_1 + r_2) \Psi_c d\tau \ , \tag{10}$$

come from those regions of the configuration space in which the two approximations may be expected to be satisfactory. In equation (10) Ψ_d denotes the normalized wave function for the ground state of H^-, and Ψ_c the wave function of the continuous state normalized to correspond to an outgoing electron of unit density.

It appears that the use of the plane-wave representation for the free electron will not introduce any very serious error, since, as has been pointed out on an earlier occasion,[5] parts of the configuration space which are only relatively far from the hydrogenic core are relevant for the absorption process. But if this be admitted, the question immediately arises as to whether the wave function for the ground state derived from the Ritz principle can be trusted to these distances. It appears that the matter can be decided in the following manner.

First, we may observe that it might prove to be an adequate approximation to use for the continuous wave function that of an electron moving in the Hartree field of a hydrogen atom. In other words, it might be sufficient to use for Ψ_c the expression

$$\Psi_c = \frac{1}{\sqrt{2\pi}} \{ e^{-r_2} \phi(r_1) + e^{-r_1} \phi(r_2) \} \ , \tag{11}$$

where $\phi(r)$ satisfies the wave equation

$$\nabla^2 \phi + \left[k^2 + 2 \left(1 + \frac{1}{r} \right) e^{-2r} \right] \phi = 0 \tag{12}$$

and tends asymptotically at infinity to a plane wave of unit amplitude along some chosen direction. If this direction in which the ejected electron moves at infinity be chosen as the polar axis of a spherical system of co-ordinates, it is readily shown that the appropriate solution for ϕ can be expressed in the form

$$\phi = \sum_{l=0}^{\infty} \frac{1}{kr} (2l+1) P_l (\cos \vartheta) \chi_l (r) , \tag{13}$$

where the radial function χ_l is a solution of the equation

$$\frac{d^2 \chi_l}{dr^2} + \left\{ k^2 - \frac{l(l+1)}{r^2} + 2 \left(1 + \frac{1}{r} \right) e^{-2r} \right\} \chi_l = 0 \ , \tag{14}$$

which tends to a pure sinusoidal wave of unit amplitude at infinity. Thus, on our present approximation Ψ_c can be written in the form

$$\left.\begin{aligned} \Psi_c = \frac{1}{\sqrt{2\pi}} \Big\{ & e^{-r_2} \sum_{l=0}^{\infty} \frac{1}{kr_1} (2l+1) P_l (\cos \vartheta_1) \chi_l (r_1; k) \\ & + e^{-r_1} \sum_{l=0}^{\infty} \frac{1}{kr_2} (2l+1) P_l (\cos \vartheta_2) \chi_l (r_2; k) \Big\} . \end{aligned}\right\} \tag{15}$$

[5] S. Chandrasekhar and M. K. Krogdahl, *Ap. J.*, 98, 205, 1943.

Using the foregoing form for Ψ_c, the standard formula for the absorption cross-section for a process in which a photoelectron with k atomic units of momentum is ejected can be reduced to the form

$$\kappa = 9.266 \times 10^{-19} \frac{k^2 + 0.05512}{k} \left| \int_0^\infty W(r) \chi_1(r) dr \right|^2 \text{cm}^2, \quad (16)$$

where $W(r)$ is a certain weight function which can be derived from and depends only on the wave function for the bound state. It is seen that, according to equation (16), the absorption cross-section depends only on the single radial function χ_1. This is to be expected, since the ground state, being an s-state, transitions can take place only to a p-state. It may be noted here that on the plane-wave representation of the free electron the appropriate form for χ_1 is

$$\chi_1(\text{plane wave}) = \frac{\sin kr}{kr} - \cos kr. \quad (17)$$

The function $W(r)$ corresponding to Henrich's eleven-parameter wave function has been computed and is tabulated in Table 1. The run of the function is further illustrated in Figure 1.

TABLE 1

THE WEIGHT FUNCTION $W(r)$

r	$W(r)$	r	$W(r)$	r	$W(r)$	r	$W(r)$
0	0	4.0	1.597	11.0	0.833	19.0	0.131
0.5	0.210	4.5	1.623	12.0	.703	20.0	.096
1.0	0.553	5.0	1.620	13.0	.585	21.0	.069
1.5	0.861	6.0	1.548	14.0	.478	22.0	.049
2.0	1.108	7.0	1.422	15.0	.383	23.0	.034
2.5	1.298	8.0	1.273	16.0	.301	24.0	.024
3.0	1.439	9.0	1.120	17.0	.233	25.0	0.016
3.5	1.538	10.0	0.972	18.0	0.177	∞	0

An examination of the values given in Table 1 discloses the somewhat disquieting fact that substantial contributions to the integral

$$\int_0^\infty W(r) \chi_1(r) dr \quad (18)$$

arise from values of r up to 25, while as much as 30–40 per cent of the entire value comes from $r \geqslant 10$. This result has two consequences. The first is that the use of the p-spherical wave (17) instead of the solution derived from (cf. eq. [14])

$$\frac{d^2\chi_1}{dr^2} + \left\{ k^2 - \frac{2}{r^2} + 2\left(1 + \frac{1}{r}\right) e^{-2r} \right\} \chi_1 = 0, \quad (19)$$

will not lead to any serious error; for the solution of equation (19), which tends to a sine wave of unit amplitude at infinity, has the behavior

$$\chi_1 \to \frac{\sin(kr + \delta)}{kr} - \cos(kr + \delta) \qquad (r \to \infty), \quad (20)$$

and the "phase shift" δ may be taken as a measure of the distortion of the p-spherical wave by the hydrogen atom at the origin. Integrations of equation (19) for various values of k^2 have been carried out numerically, and the resulting phase shifts for some of them are given in Table 2. It is seen that the phase shifts are indeed quite small for values of k^2, which are of astrophysical interest.

The second consequence of the run of the function $W(r)$ is not so satisfactory; for an examination of the energy integral minimized in the Ritz principle reveals that over 95 per cent of the contribution to the integral arises from regions of the configuration space which correspond to $r < 10$. Accordingly, it would appear that the choice of the wave

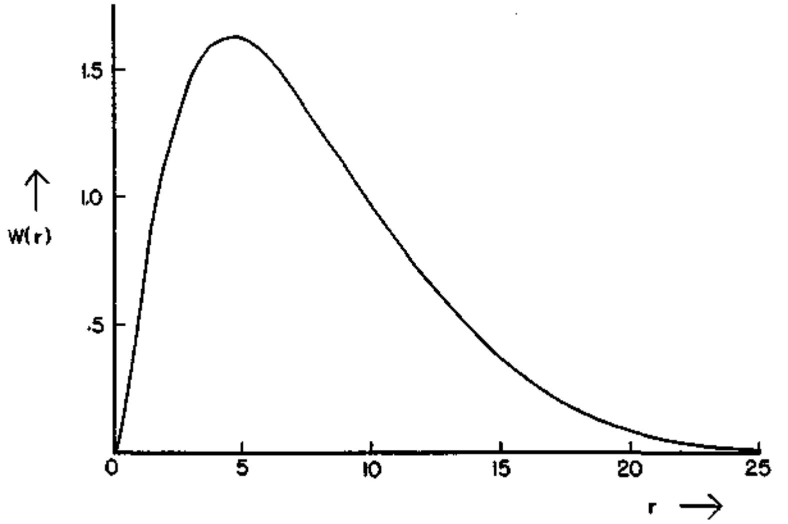


FIG. 1

TABLE 2

PHASE SHIFTS δ FOR THE p-SPHERICAL WAVES IN THE HARTREE FIELD OF A HYDROGEN ATOM

k^2	δ	k^2	δ	k^2	δ	k^2	δ
1.50.......	0.1486	0.80......	0.09244	0.25......	0.02605	0.100.....	0.007689
1.00.......	0.1115	0.50......	0.05838	0.125.....	0.01046	0.035.....	0.001709

function in accordance only with the Ritz principle cannot be expected to lead to values of $W(r)$ which are necessarily trustworthy for $r > 10$. Under these circumstances the best hope for improving the current wave functions would consist in first determining the true asymptotic forms of the wave function for large distances and later choosing functions which would lead not only to the best value for the energy but also to the correct asymptotic forms. However, such calculations are likely to be extremely laborious.

I am greatly indebted to Miss Frances Herman for valuable assistance in the numerical parts of the present investigation.

ON THE CONTINUOUS ABSORPTION COEFFICIENT OF THE NEGATIVE HYDROGEN ION

S. CHANDRASEKHAR
Yerkes Observatory
Received June 25, 1945

ABSTRACT

In this paper it is shown that the continuous absorption coefficient of the negative hydrogen ion is most reliably determined by a formula for the absorption cross-section which involves the matrix element of the momentum operator. A new absorption curve for H^- has been determined which places the maximum at λ 8500 A; at this wave length the atomic absorption coefficient has the value 4.37×10^{-17} cm².

1. *Introduction.*—In earlier discussions[1] by the writer attention has been drawn to the fact that the continuous absorption coefficient of the negative hydrogen ion, evaluated in terms of the matrix element

$$\mu = \int \Psi_d^* (r_1 + r_2) \Psi_c d\tau \tag{1}$$

(where Ψ_d denotes the wave function of the ground state of the ion and Ψ_c the wave function belonging to a continuous state normalized to correspond to an outgoing electron of unit density), depends very much on Ψ_d in regions of the configuration space which are relatively far from the hydrogenic core. This has the consequence that the absorption cross-sections are not trustworthily determined if wave functions derived by applications of the Ritz principle are used in the calculation of the matrix elements according to equation (1). This is evident, for example, from Figure 1, in which we have plotted the absorption coefficients as determined by Williamson[2] and Henrich,[3] using wave functions of the forms

$$\Psi_d = \mathfrak{N}\, e^{-\alpha s/2} (1 + \beta u + \gamma t^2 + \delta s + \epsilon s^2 + \zeta u^2) \tag{2}$$

and

$$\left.\begin{aligned}\Psi_d = \mathfrak{N}\, e^{-\alpha s/2} (1 + \beta u + \gamma t^2 + \delta s + \epsilon s^2 + \zeta u^2 + \chi_6 t^4 + \chi_7 t^6 \\ + \chi_8 t^4 u^2 + \chi_9 t^2 u^2 + \chi_{10} t^2 u^4)\,,\end{aligned}\right\} \tag{3}$$

respectively. (In eqs. [2] and [3] $\mathfrak{N}$ is the normalizing factor; and α, β, γ, etc., are constants determined by the Ritz condition of minimum energy,

$$s = r_2 + r_1\,, \quad t = r_2 - r_1\,, \quad \text{and} \quad u = r_{12}\,, \tag{4}$$

where r_1, r_2, and r_{12} are the distances of the two electrons from the nucleus and from each other, respectively.) The wide divergence between the two curves in Figure 1 is too large to be explained in terms of only the improvement in energy effected by the wave function (3): it must arise principally from the fact that in the evaluation of the matrix elements according to equation (1) parts of the wave function are used which do not contribute appreciably to the energy integral and are therefore poorly determined. Indeed, this sen-

[1] *Ap. J.*, **100**, 176, 1944; also *Rev. Mod. Phys.*, **16**, 301, 1944.

[2] *Ap. J.*, **96**, 438, 1942.

[3] *Ap. J.*, **99**, 59, 1944.

sitiveness of the derived absorption coefficients to wave functions effecting only relatively slight improvements in the energy makes it difficult to assess the reliability of the computed absorption coefficients. However, in this paper we shall show how these difficulties can be avoided by using a somewhat different formula for the absorption cross-section.

2. *Alternative formulae for evaluating the absorption coefficient.*—It is well known that in the classical theory the radiative characteristics of an oscillating dipole can be expressed in terms of either its dipole moment, its momentum, or its acceleration. There are, of course, analogous formulations in the quantum theory, the matrix element

$$(a \mid z_j \mid b) = \int \psi_a^* z_j \psi_b d\tau \tag{5}$$

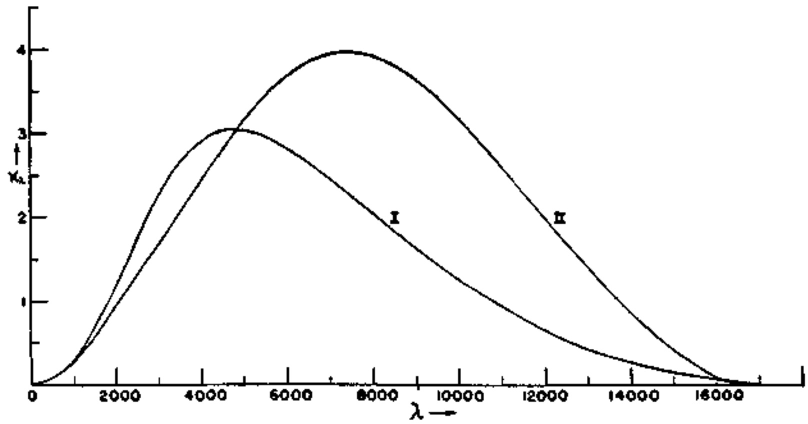


FIG. 1.—A comparison of the continuous absorption coefficient of H^- computed according to formula (I) and with wave functions of forms (2) (curve *I*) and (3) (curve *II*). The ordinates denote the absorption coefficients in units of 10^{-17} cm^2; the abscissae, the wave length in angstroms.

for the co-ordinate z_j of the jth electron in an atom being simply related to the corresponding matrix element of the momentum operator or the acceleration. Thus, we have the relations

$$(a \mid z_j \mid b) = \frac{1}{(E_a - E_b)} \int \frac{\partial \psi_a^*}{\partial z_j} \psi_b d\tau = -\frac{1}{(E_a - E_b)} \int \psi_a^* \frac{\partial \psi_b}{\partial z_j} d\tau \tag{6}$$

and

$$(a \mid z_j \mid b) = \frac{1}{(E_a - E_b)^2} \int \psi_a^* \frac{\partial V}{\partial z_j} \psi_b d\tau \tag{7}$$

if all the quantities are measured in Hartree's atomic units and where E_a and E_b denote the energies of the states indicated by the letters a and b and where V denotes the potential energy arising from Coulomb interactions between the particles. More particularly for an atom (or ion) with two electrons, we have

$$\mu_z = \int \Psi_d^* (z_1 + z_2) \Psi_c d\tau , \tag{8}$$

$$\mu_z = -\frac{1}{(E_d - E_c)} \int \Psi_d^* \left(\frac{\partial}{\partial z_1} + \frac{\partial}{\partial z_2} \right) \Psi_c d\tau , \tag{9}$$

and

$$\mu_z = \frac{1}{(E_d - E_c)^2} \int \Psi_d^* \left(\frac{z_1}{r_1^3} + \frac{z_2}{r_2^3} \right) \Psi_c d\tau . \tag{10}$$

While the foregoing formulae are entirely equivalent to each other if Ψ_d abd Ψ_c are exact solutions of the wave equation, they are of different merits for the evaluation of μ_z if approximate wave functions are used. Thus, it is evident that formula (8) uses parts of the configuration space, which are more distant than relevant, for example, in the evaluation of the energy; similarly, formula (10) uses the wave functions in regions much nearer the origin. It would appear that formula (9) is the most suitable one for the evaluation of μ_z, particularly when wave functions derived by applications of the Ritz principle are used. The calculations which we shall present in the following sections confirm this anticipation; but before we proceed to such calculations, it is useful to have the explicit formulae for the absorption cross-sections on the basis of equations (8), (9), and (10).

In ordinary (c.g.s.) units the standard formula for the atomic absorption coefficient κ_ν for radiation of frequency ν, in which an electron with a velocity v is ejected, is

$$\kappa_\nu = \frac{32\pi^4 m^2 e^2}{3h^3 c} \nu v \left| \int \Psi_d^* (z_1 + z_2) \Psi_c d\tau \right|^2 , \tag{11}$$

where m, e, h, and c have their usual meanings. (In writing eq. [11] it has been assumed that the electron is ejected in the z-direction; see eq. [15] below.) By inserting the numerical values for the various atomic constants equation (11) can be expressed in the form

$$\kappa_\nu = 8.561 \times 10^{-19} (\nu_{at} k \,|\mu_z|^2) \text{ cm}^2 , \tag{12}$$

where k denotes the momentum of the ejected electron and ν_{at} the frequency of the radiation absorbed, both measured in atomic units, and where, moreover, the matrix element μ_z has also to be evaluated in atomic units.

If I denotes the electron affinity (also expressed in atomic units)

$$4\pi \nu_{at} = k^2 + 2I , \tag{13}$$

and depending on which of the formulae (8), (9), and (10) we use for evaluating κ_ν, we have

$$\kappa_\nu = 6.812 \times 10^{-20} k (k^2 + 2I) \left| \int \Psi_d^* (z_1 + z_2) \Psi_c d\tau \right|^2 . \tag{I}$$

$$\kappa_\nu = 2.725 \times 10^{-19} \frac{k}{(k^2 + 2I)} \left| \int \Psi_d^* \left(\frac{\partial}{\partial z_1} + \frac{\partial}{\partial z_2} \right) \Psi_c d\tau \right|^2 , \tag{II}$$

and

$$\kappa_\nu = 1.090 \times 10^{-18} \frac{k}{(k^2 + 2I)^3} \left| \int \Psi_d^* \left(\frac{z_1}{r_1^3} + \frac{z_2}{r_2^3} \right) \Psi_c d\tau \right|^2 . \tag{III}$$

Finally, we may note that if λ denotes the wave length of the radiation measured in angstroms, then

$$\lambda = \frac{911.3}{k^2 + 2I} \text{ A} . \tag{14}$$

3. *The continuous absorption coefficient of* H$^-$ *evaluated according to formula (III)*.—As we have already indicated, in the customary evaluations of κ_ν according to formula (I) the relatively more distant parts of the wave function are used. It is evident that we shall be going to the opposite extreme in using the wave function principally only near the origin if we evaluate κ_ν according to formula (III). For this reason it is of interest to consider first the absorption coefficient as determined by this formula.

In evaluating κ_ν according to formula (III), we shall use for Ψ_d a wave function of form (3) and for Ψ_c a plane wave representation of the outgoing electron:

$$\Psi_c = \frac{1}{\sqrt{2\pi}} \left(e^{-r_1 + ikz_2} + e^{-r_2 + ikz_1} \right) . \tag{15}$$

(In § 5 we refer to an improvement in Ψ_c which can be incorporated without much difficulty at this stage.) For Ψ_d and Ψ_c of forms (3) and (15) the evaluation of the matrix element

$$\int \Psi_d^* \left(\frac{z_1}{r_1^3} + \frac{z_2}{r_2^3} \right) \Psi_c d\tau \tag{16}$$

is straightforward, though it is somewhat involved. We find

$$\left.\begin{aligned}\int \Psi_d^* \left(\frac{z_1}{r_1^3}+\frac{z_2}{r_2^3}\right)\Psi_c d\tau = -(2048\pi^3)^{1/2}\frac{\mathfrak{N}}{(1+a)^3}\frac{i}{k^2}\Bigg[\sum_{j=-2}^{6} l_j \mathcal{L}_j^{(a)} \\ +\sum_{j=-2}^{3}\lambda_j \mathcal{L}_j^{(1+2a)} - \tfrac{1}{30}\beta(1+a)^4\left\{\sum_{j=-1}^{3} a_j S_j^{(1+2a)} + \sum_{j=-1}^{3} b_j C_j^{(1+2a)}\right\}\Bigg],\end{aligned}\right\} \quad (17)$$

where we have used the following abbreviations:

$$\left.\begin{aligned}\mathcal{L}_j^{(p)} &= \int_0^\infty e^{-py}\left(k\cos ky - \frac{\sin ky}{y}\right)y^j dy && (j=-2,\,-1,\ldots\ldots), \\ &= (j-1)!\rho^j\{j\rho k\cos[(j+1)\xi] - \sin j\xi\} && (j \geqslant 1), \\ &= \rho k \cos\xi - \xi && (j=0), \\ &= p\xi - k && (j=-1), \\ &= \frac{1}{2}\left(pk - \frac{\xi}{\rho^2}\right) && (j=-2),\end{aligned}\right\} \quad (18)$$

$$\left.\begin{aligned}S_j^{(p)} &= \int_0^\infty e^{-py}y^j \sin ky\, dy && (j=-1,\,0,\ldots\ldots), \\ &= j!\rho^{j+1}\sin[(j+1)\xi] && (j=0,\,1,\ldots\ldots), \\ &= \xi && (j=-1),\end{aligned}\right\} \quad (19)$$

and

$$\left.\begin{aligned}C_j^{(p)} &= \int_0^\infty e^{-py}y^j\cos ky\, dy = j!\rho^{j+1}\cos[(j+1)\xi] && (j=0,\,1,\ldots\ldots), \\ &= \int_0^\infty e^{-py}(e^{ay}-\cos ky)\frac{dy}{y} = \log\left(\frac{p}{p-a}\,|\sec\xi|\right) && (j=-1),\end{aligned}\right\} \quad (20)$$

where

$$\rho = \frac{1}{(k^2+p^2)^{1/2}} \quad \text{and} \quad \xi = \tan^{-1}\frac{k}{p}; \quad (21)$$

and

$$\left.\begin{aligned}l_{-2} &= 4q^2\beta, \\ l_{-1} &= 1+q(\tfrac{1}{3}\beta+\delta)+12q^2(\gamma+\epsilon+\zeta)+360q^4(\chi_6+\chi_9)+20{,}160q^6(\chi_7+\chi_8+\chi_{10}), \\ l_0 &= (\delta+\beta)-6q(\gamma-\epsilon)-120q^3(2\chi_6+\chi_9)-5040q^5(3\chi_7+2\chi_8+\chi_{10}), \\ l_1 &= -\frac{\beta}{6q}+(\gamma+\epsilon+\zeta)+24q^2(3\chi_6+\chi_9)+120q^4(45\chi_7+21\chi_8+13\chi_{10}), \\ l_2 &= -\frac{\zeta}{3q}-4q(3\chi_6+2\chi_9)-40q^3(30\chi_7+13\chi_8+12\chi_{10}) \\ l_3 &= (\chi_6+\tfrac{7}{3}\chi_9)+4q^2(45\chi_7+29\chi_8+21\chi_{10}), \\ l_4 &= -\frac{\chi_9}{3q}-2q(9\chi_7+12\chi_8+7\chi_{10}), \\ l_5 &= \chi_7+\tfrac{11}{3}(\chi_8+\chi_{10}), \\ l_6 &= -\frac{1}{3q}(\chi_8+2\chi_{10}).\end{aligned}\right\} \quad (22)$$

$$\left.\begin{aligned} \lambda_{-2} &= -4\beta q^2\,; & \lambda_1 &= \frac{\beta}{15q}\,, \\ \lambda_{-1} &= -\tfrac{6}{5}\beta q\,; & \lambda_2 &= -\frac{\beta}{30q^2}\,, \\ \lambda_0 &= -\tfrac{1}{5}\beta\,; & \lambda_3 &= \frac{\beta}{30q^3}\,, \end{aligned}\right\} \tag{23}$$

where

$$q = \frac{1}{1+a}\,; \tag{24}$$

and

$$\left.\begin{aligned} a_{-1} &= 6\eta^4\left[4\eta k^2(5a^4-10a^2k^2+k^4)+(a^4-6a^2k^2+k^4)\right], \\ a_0 &= 6\eta^3 a\left[16\eta k^2(a^2-k^2)+(a^2-3k^2)\right], \\ a_1 &= 3\eta^2\left[4\eta k^2(3a^2-k^2)+(a^2-k^2)\right], \\ a_2 &= \eta a\,(8\eta k^2+1)\,, \\ a_3 &= \eta k^2\,, \end{aligned}\right\} \tag{25}$$

$$\left.\begin{aligned} b_{-1} &= +24\eta^4 a k\left[\eta(a^4-10a^2k^2+5k^4)-(a^2-k^2)\right], \\ b_0 &= -6\eta^3 k\left[4\eta(a^4-6a^2k^2+k^4)-(3a^2-k^2)\right], \\ b_1 &= -6\eta^2 a k\left[2\eta(a^2-3k^2)-1\right], \\ b_2 &= -\eta k\left[4\eta(a^2-k^2)-1\right], \\ b_3 &= -\eta a k\,, \end{aligned}\right\} \tag{26}$$

where

$$\eta = (a^2+k^2)^{-1}\,. \tag{27}$$

Putting $\chi_6 = \chi_7 = \ldots. = \chi_{10} = 0$ in the foregoing equations, we shall obtain the formulae which can be used with a wave function of form (2).

By using for the constants of wave functions (2) and (3) the values determined by Williamson and Henrich, the atomic absorption coefficient κ_ν has been computed according to the foregoing formulae for various wave lengths. The results of the calculations are given in Table 1 and are further illustrated in Figure 2. It is seen that, in contrast to what happened when formula (I) was used (cf. Fig. 1), wave function (2) now predicts systematically *larger* values for κ_ν than does wave function (3). The divergence between the two curves must now be attributed to the overweighting of the wave function near the origin, where it is again poorly determined by the Ritz method.

4. *The continuous absorption coefficient of* H^- *evaluated according to formula (II).*—Finally, returning to formula (II), which would appear to have the best chances for determining κ_ν most reliably, the calculations were again carried through for wave functions Ψ_d of forms (2) and (3) and for Ψ_c of form (15). Before we give the results of the calculations, we may note that for Ψ_d of form (3) and for Ψ_c of form (15)

$$\left.\begin{aligned} \int\Psi_d^*\left(\frac{\partial}{\partial z_1}+\frac{\partial}{\partial z_2}\right)\Psi_c d\tau = -(2048\pi^3)^{1/2}\frac{\mathfrak{N}}{(1+a)^3}\frac{i}{k^2}\Bigg[\sum_{j=-1}^{6} l_j\mathcal{L}_j^{(a)} \\ +\sum_{j=-1}^{+1}\lambda_j\mathcal{L}_j^{(1+2a)}+k^2\left\{\sum_{j=0}^{7} s_j S_j^{(a)}+\sum_{j=0}^{+1}\sigma_j S_j^{(1+2a)}\right\}\Bigg], \end{aligned}\right\} \tag{28}$$

TABLE 1

THE CONTINUOUS ABSORPTION COEFFICIENT OF H^- COMPUTED ACCORDING TO FORMULA III AND WITH WAVE FUNCTIONS OF FORMS (2) AND (3)

λ (A)	$\kappa_\lambda \times 10^{17}$ cm^2		λ (A)	$\kappa_\lambda \times 10^{17}$ cm^2	
	With Wave Function (3)	With Wave Function (2)		With Wave Function (3)	With Wave Function (2)
1000	0.225	0.241	7000	5.173	6.732
2000	0.955	1.010	7500	5.225	7.070
2500	1.459	1.538	8000	5.204	7.333
3000	2.010	2.125	8500	5.106	7.496
3500	2.580	2.752	9000	4.946	7.567
4000	3.139	3.400	9500	4.724	7.536
4500	3.657	4.046	10000	4.453	7.411
5000	4.118	4.676	12000	3.031	5.952
5500	4.505	5.271	14000	1.407	3.355
6000	4.812	5.820	16000	0.149	0.401
6500	5.036	6.310			

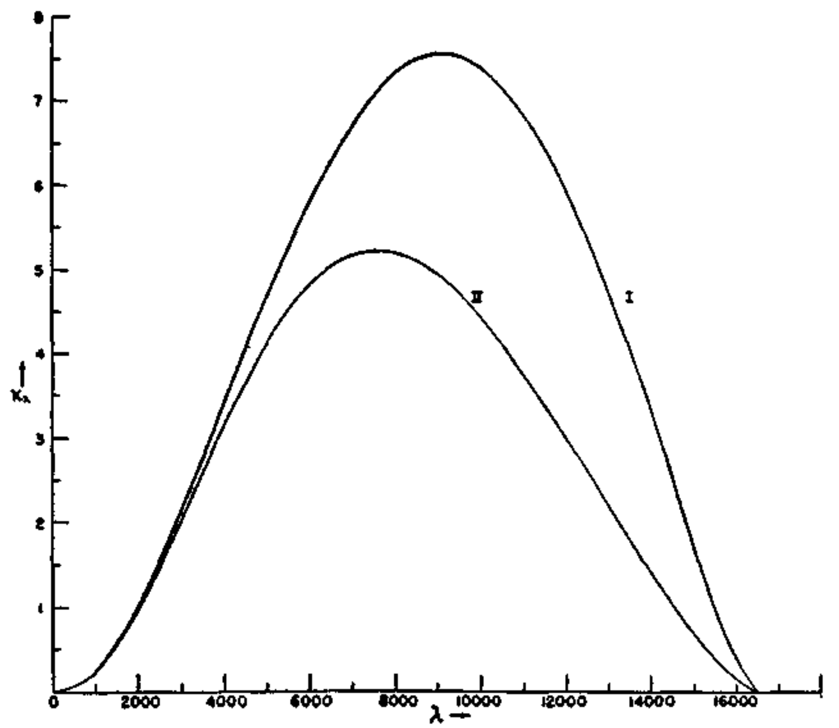


FIG. 2.—A comparison of the continuous absorption coefficient of H^- computed according to formula (III) and with wave functions of forms (2) (curve *I*) and (3) (curve *II*). The ordinates denote the absorption coefficients in units of 10^{-17} cm^2; the abscissae, the wave length in angstroms.

where

$$\left.\begin{aligned} &l_{-1} = 4\beta q^3\,; \quad l_0 = 0\,; \quad l_1 = -\beta q\,. \\ &l_2 = -2q\zeta - 40q^3\chi_9 - 1680q^5(\chi_8 + 2\chi_{10})\,. \\ &l_3 = 16q^2\chi_9 + 960q^4(\chi_8 + \chi_{10})\,. \\ &l_4 = -2q\chi_9 - 80q^3(3\chi_8 + 2\chi_{10})\,. \\ &l_5 = 32q^2(\chi_8 + \chi_{10})\,; \quad l_6 = -2q(\chi_8 + 2\chi_{10})\,, \end{aligned}\right\} \tag{29}$$

$$\lambda_{-1} = -4\beta q^3\,; \quad \lambda_0 = -4\beta q^2\,; \quad \lambda_1 = -\beta q\,, \tag{30}$$

$$\left.\begin{aligned} &s_0 = 4\beta q^2\,, \\ &s_1 = 1 + 3q\delta + 12q^2(\gamma + \epsilon + \zeta) + 360q^4(\chi_6 + \chi_9) + 20{,}160q^6(\chi_7 + \chi_8 + \chi_{10})\,, \\ &s_2 = (\delta + \beta) - 6q(\gamma - \epsilon) - 120q^2(2\chi_6 + \chi_9) - 5040q^5(3\chi_7 + 2\chi_8 + \chi_{10})\,, \\ &s_3 = (\gamma + \epsilon + \zeta) + 24q^2(3\chi_6 + \chi_9) + 120q^4(45\chi_7 + 21\chi_8 + 13\chi_{10})\,, \\ &s_4 = -6q(2\chi_6 + \chi_9) - 80q^3(15\chi_7 + 6\chi_8 + 5\chi_{10})\,, \\ &s_5 = (\chi_6 + \chi_9) + 4q^2(45\chi_7 + 21\chi_8 + 13\chi_{10})\,. \\ &s_6 = -6q(3\chi_7 + 2\chi_8 + \chi_{10})\,, \\ &s_7 = \chi_7 + \chi_8 + \chi_{10} \end{aligned}\right\} \tag{31}$$

and

$$\sigma_0 = -4\beta q^2\,; \quad \sigma_1 = -\beta q\,. \tag{32}$$

Further, in equation (28) the quantities $\mathcal{L}_j^{(p)}$, $S_j^{(p)}$, and q have the same meanings as in equations (18), (19), (21), and (24).

TABLE 2

THE CONTINUOUS ABSORPTION COEFFICIENT OF H^- COMPUTED ACCORDING TO FORMULA II AND WITH WAVE FUNCTIONS OF FORMS (2) AND (3)

λ (A)	$\kappa_\lambda \times 10^{17}$ cm^2		λ (A)	$\kappa_\lambda \times 10^{17}$ cm^2	
	With Wave Function (3)	With Wave Function (2)		With Wave Function (3)	With Wave Function (2)
1000	0.271	0.270	7000	4.174	4.113
2000	0.945	0.991	7500	4.296	4.080
2500	1.335	1.461	8000	4.363	3.993
3000	1.730	1.955	8500	4.372	3.858
3500	2.119	2.437	9000	4.324	3.682
4000	2.498	2.880	9500	4.221	3.471
4500	2.860	3.265	10000	4.065	3.233
5000	3.197	3.581	12000	2.995	2.108
5500	3.504	3.822	14000	1.502	0.954
6000	3.773	3.989	16000	0.167	0.097
6500	3.998	4.084			

The absorption cross-sections, as calculated according to formula (II), and the foregoing equations are given in Table 2 and further illustrated in Figure 3. It is seen that, as anticipated, the two curves now do not diverge more than can be reasonably attributed to the betterment of the wave function in consequence of the increased number of parameters used in the Ritz method.

5. *Concluding remarks.*—A comparison of Figures 1, 2, and 3 clearly illustrates the superiority of formula (II) for the purposes of evaluating the continuous absorption coefficient of the negative hydrogen ion. The general reliability of the absorption cross-

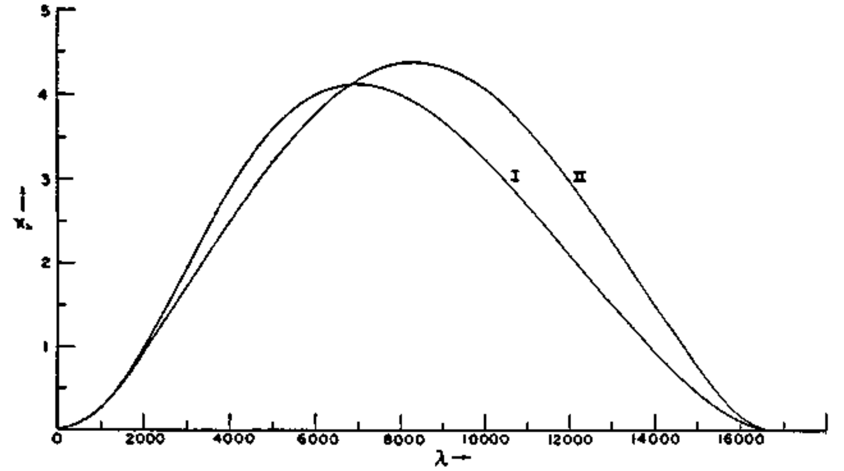


FIG. 3.—A comparison of the continuous absorption coefficient of H^- computed according to formula (II) and with wave functions of forms (2) (curve *I*) and (3) (curve *II*). The ordinates denote the absorption coefficients in units of 10^{-17} cm^2; the abscissae, the wave length in angstroms.

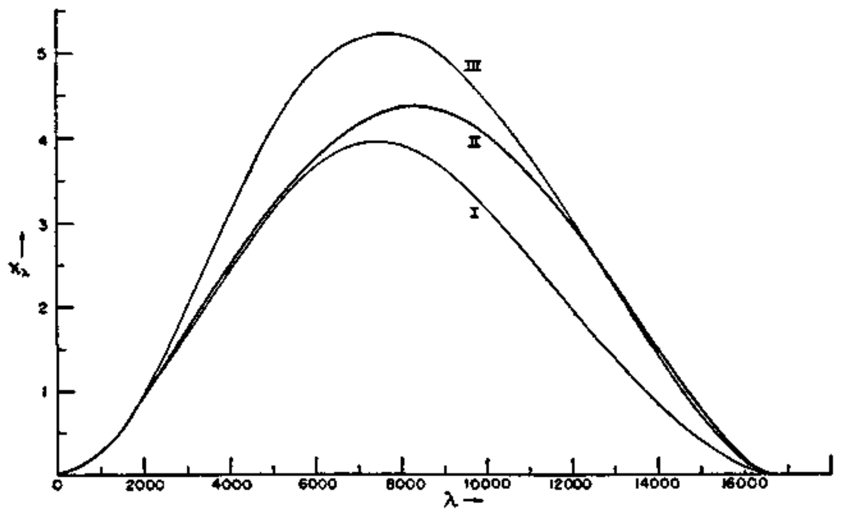


FIG. 4.—A comparison of the continuous absorption coefficient of H^- computed according to formulae (I) (curve *I*), (II) (curve *II*), and (III) (curve *III*) with a wave function of form (3). The ordinates denote the absorption coefficients in units of 10^{-17} cm^2; the abscissae, the wave length in angstroms.

sections derived on the basis of formula (II) and wave function (3) can be seen in another way. In Figure 4 we have plotted κ_ν as given by the three formulae and as obtained in each case with wave function (3). It is seen that, while the cross-sections given by formula (II) agree with those given by formula (I) in the visual and the violet part of the spectrum ($\lambda < 6000$ A), they agree with those given by formula (III) in the infrared ($\lambda > 12{,}000$ A). This is readily understood when it is remembered that on all the three formulae the absorption cross-sections in the infrared are relatively more dependent on the wave function at large distances than they are in the visual and the violet parts of the spectrum. Accordingly, it is to be expected that, as we approach the absorption limit of H^- at 16,550 A, formula (III) must give less unreliable values than it does at shorter wave lengths; formula (I), of course, ceases to be valid in the infrared. It is also clear that, as we go toward the violet, we have the converse situation.

Summarizing our conclusions so far, it may be said that in the framework of the approximation in which a plane-wave representation of the outgoing electron is used, formula (II), together with wave function (3), gives sufficiently reliable values for the absorption coefficient over the entire range of the spectrum. Attention may be particularly drawn to the fact that the maximum of the absorption-curve is now placed at λ 8500 A, where $\kappa_\lambda = 4.37 \times 10^{-17}\,\mathrm{cm}^2$.

The question still remains as to the improvements which can be effected in the choice of Ψ_c. As shown in an earlier paper,[4] it may be sufficient to use for Ψ_c the wave functions in the Hartree field of a hydrogen atom. On this approximation we should use (*op. cit.*, eq. [15])

$$\left.\begin{aligned}\Psi_c = \frac{1}{\sqrt{2\pi}}\Bigg\{ e^{-r_1} \sum_{l=0}^{\infty} \frac{i^l}{k\,r_2} (2l+1)\,P_l(\cos\vartheta_2)\,\chi_l(r_2;\,k) \\ + e^{-r_2} \sum_{l=0}^{\infty} \frac{i^l}{k\,r_1} (2l+1)\,P_l(\cos\vartheta_1)\,\chi_l(r_1;\,k)\Bigg\},\end{aligned}\right\} \tag{33}$$

where χ_l is the solution of the equation

$$\frac{d^2\chi_l}{d\,r^2} + \left\{ k^2 - \frac{l\,(l+1)}{r^2} + 2\left(1+\frac{1}{r}\right) e^{-2r} \right\} \chi_l = 0\,, \tag{34}$$

which tends to a pure sinusoidal wave of unit amplitude at infinity. We shall return to these further improvements in a later paper.

It is a pleasure to acknowledge my indebtedness to Professor E. P. Wigner for many helpful discussions and much valuable advice. My thanks are also due to Mrs. Frances Herman Breen for assistance with the numerical work.

[4] *Ap. J.*, **100**, 176, 1944.

THE CONTINUOUS SPECTRUM OF THE SUN AND THE STARS

S. Chandrasekhar and Guido Münch[1]
Yerkes Observatory
Received September 5, 1946

Using the recent determination of the continuous absorption coefficient of H^- by Chandrasekhar and Breen, we have shown that the dependence of the continuous absorption coefficient with wave length in the range 4000–24,000 A, which can be inferred from the intensity distribution in the continuous spectrum of the sun, can be quantitatively accounted for as due to H^-; and, further, that the color temperatures measured in the wave-length intervals 4100–6500 A (Greenwich) and 4000–4600 A (Barbier and Chalonge) for stars of the main sequence and of spectral types A0–G0 can also be interpreted in terms of the continuous absorption of H^- and neutral hydrogen atoms.

The problem of the discontinuities at the head of the Balmer and the Paschen series is also briefly considered on the revised physical theory of the continuous absorption coefficient.

1. *Introduction.*—The two principal problems in the theory of the continuous spectrum of the stars are, first, to identify the source of the continuous absorption in the solar atmosphere which will account for the intensity distribution in the continuous spectrum of the sun and the law of darkening in the different wave lengths and, second, to account for the observed relations between the color and the effective temperatures of the stars. In this paper we shall show that the major aspects of these two problems find their natural solution in terms of the continuous absorption coefficient of the negative hydrogen ion as recently determined by Chandrasekhar and Breen.[2] More particularly, we shall show that the dependence of the continuous absorption coefficient with wave length in the range 4000–24,000 A, which can be deduced from the solar data, can be quantitatively accounted for as due to H^-; and, further, that the color temperatures measured in the wave-length intervals 4100–6500 A (Greenwich[3]) and 4000–4600 A (Barbier and Chalonge[4]) for stars of the main sequence and of spectral types A0–G0 can also be interpreted in terms of the continuous absorption of H^- and neutral hydrogen.

In addition to the two problems we have mentioned, we shall also consider some related questions concerning the discontinuities at the head of the Balmer and the Paschen series.

2. *The mean absorption coefficients of* H$^-$ *and* H.—As is well known, the character of the emergent continuous radiation from a stellar atmosphere is determined in terms of the temperature distribution in the atmosphere; and, as has recently been shown,[5] the temperature distribution in a nongray atmosphere will be given approximately by a formula of the standard type

$$T^4 = \tfrac{3}{4} T_e^4 (\tau + q[\tau]), \tag{1}$$

where T_e denotes the effective temperature and $q(\tau)$ a certain monotonic increasing function of the optical depth τ, provided that the mean absorption coefficient κ, in terms of which τ is measured, is defined as a straight average of the monochromatic absorption coefficient weighted according to the net flux $F_\nu^{(1)}$ of radiation of frequency ν in a gray

[1] Fellow of the John Simon Guggenheim Memorial Foundation at the Yerkes Observatory.

[2] *Ap. J.*, **104**, 430, 1946.

[3] Sir Frank Dyson, *Observations of Color-Temperatures of Stars, 1926–1932*, London, 1932; also *M.N.*, **100**, 189, 1940.

[4] *Ann. d'ap.*, **4**, 30, Table 4, 1941.

[5] S. Chandrasekhar, *Ap. J.*, **101**, 328, 1945. This paper will be referred to as "Radiative Equilibrium VII."

atmosphere and if, further, $\kappa_\nu/\bar{\kappa}$ is independent of depth. On this approximation, then, the emergent intensity in a given frequency and in a given direction will depend not only on the continuous absorption coefficient at the frequency under consideration but also on the mean absorption coefficient κ over all frequencies.

As the discussion in this paper will establish for the stellar atmospheres considered, the contributions to κ in the visible and the infrared regions of the spectrum are essentially from only two sources: H^- and the neutral hydrogen atoms. The cross-sections for the absorption by H^- for various temperatures and wave lengths have been tabulated by Chandrasekhar and Breen in Table 7 of their paper, while those for hydrogen can be found from the formulae of Kramers and Gaunt, standardized, for example, by B. Strömgren.[6] However, the evaluation of the mean absorption coefficient for wave lengths shorter than 4000 A is made uncertain on two accounts: First, there is the absorption by the metals and the excessive crowding of the absorption lines toward the violet, which is particularly serious for spectral types later than F0; and, second, there is the absorption in the Lyman continuum. On both these accounts the true values of $\bar{\kappa}$ will be larger than those determined by ignoring them. But the exact amount by which they will be larger will be difficult to predict without a detailed theory of "blanketing,"[7] on the one hand, and without going into a more exact theory[8] of radiative transfer than represented by the approximations leading to equation (1), on the other. However, since in this paper our primary object is to establish only the adequacy of H^- as the source of absorption in the solar atmosphere over the entire visible and infrared regions of the spectrum and the corresponding role of H^- and H for stellar atmospheres with spectral types A2–G0, it appeared best to ignore the refinements indicated and simply determine $\bar{\kappa}$ by weighting κ_ν due to H^- and H (without the Lyman absorption) at the conditions prevailing at $\tau = 0.6$ by the flux $F_\nu^{(1)}$ at this level.[9] For only in this way can we use the solution to the transfer problem in the form of equation (1) in a consistent manner. It should, however, be remembered that the effects we have ignored may easily increase $\bar{\kappa}$, determined in terms of H^- and H (without the absorption in the Lyman continuum) by factors of the order of 1.5 and probably not exceeding 2.[10]

Turning our attention, next, to the evaluation of $\bar{\kappa}$, we may first observe that, since our present method of averaging is a straight one, the contributions to κ_ν from different sources are simply additive. We may, accordingly, consider the mean absorption coefficient of H^- and H separately.

Now the absorption coefficient of H^-, including both the bound-free and the free-free transitions, is most conveniently expressed as per neutral hydrogen atom and per unit electron pressure in the unit cm^4/dyne. The monochromatic coefficients κ_ν', after allowing for the stimulated emission factor $(1 - e^{-h\nu/kT})$, are tabulated in Chandrasekhar and Breen's paper for various values of $\theta(=5040/T)$. If we now denote by $a(H^-)$ the average

[6] "Tables of Model Stellar Atmospheres," *Publ. mind. Meddel. Kobenhavns Obs.*, No. 138, 1944.

[7] Cf. G. Münch, *Ap. J.*, **104**, 87, 1946.

[8] Such as, e.g., the (2, 2) approximation given in "Radiative Equilibrium VII," § 6.

[9] The choice of $\tau = 0.6$ for the "representative point" was made after some preliminary trials (cf. G. Münch, *Ap. J.*, **102**, 385, 1945, esp. Table 3), though it is evident on general grounds that a level such as $\tau = 0.6$, where the local temperature is approximately the same as the effective temperature, would be the correct one in the scheme of approximations leading to eq. (1).

[10] In all earlier evaluations of $\bar{\kappa}$ the absorption in the Lyman continuum did not, indeed, play any role. This was due to the manner in which $\bar{\kappa}$ was defined in those investigations as the Rosseland mean. But in "Radiative Equilibrium VII" it has been shown that there is no justification for taking the Rosseland means as they have been hitherto. Since the method of averaging, by which we have now replaced the Rosseland mean, is a straight one, it is no longer permissible simply to ignore the absorption in the Lyman continuum. At the same time, it is not possible to take it into account satisfactorily in the (2, 1) approximation leading to eq. (1). We should have to go at least to the (2, 2) approximation of "Radiative Equilibrium VII."

value of the coefficients κ'_ν, weighted according to the flux $F_\nu^{(1)}$ in a gray atmosphere at $\tau = 0.6$, where the temperature is approximately the effective temperature T_e, then the contribution $\bar{\kappa}(H^-)$ to the mass absorption coefficient $\bar{\kappa}$ by H^- is given by

$$\bar{\kappa}\,(H^-) = \frac{(1 - x_H)\, p_e}{m_H}\, a\,(H^-)\,, \tag{2}$$

where m_H is the mass of the hydrogen atom, p_e the electron pressure, and x_H the degree of ionization of hydrogen under the physical conditions represented by T_e and p_e.[11] The values of $a(H^-)$ found by graphical integration in accordance with the formula

$$a\,(H^-) = \frac{1}{F}\int_0^\infty \kappa'_\nu F^{(1)}\,(0.6)\,d\nu \tag{3}$$

for various values of $\theta = \theta_e$ are given in Table 1.

TABLE 1

THE MEAN ABSORPTION COEFFICIENTS $a(H^-)$ AND $a(H)$

θ	$a(H^-)$	$a(H)$	θ	$a(H^-)$	$a(H)$
0.5	0.563×10^{-26}	1.65×10^{-22}	0.9	6.08×10^{-26}	5.52×10^{-27}
0.6	1.145×10^{-26}	1.22×10^{-23}	1.0	9.32×10^{-26}	3.65×10^{-28}
0.7	2.25×10^{-26}	1.13×10^{-24}	1.2	2.00×10^{-25}	....
0.8	3.88×10^{-26}	8.00×10^{-26}	1.4	3.89×10^{-25}	....

Similarly, the contribution to $\bar{\kappa}$ by hydrogen can also be expressed in the form

$$\bar{\kappa}\,(H) = \frac{1 - x_H}{m_H}\, a\,(H)\,, \tag{4}$$

where

$$a\,(H) = \int_0^\infty \frac{fD}{a^3}\,(1 - e^{-a})\,\frac{F_a^{(1)}\,(0.6)}{F}\,da\,, \tag{5}$$

where $a = h\nu/kT_e$ and f and D are certain functions of temperature and frequency, respectively, which have been tabulated by Strömgren.[12] For reasons which we have already explained, we do not include the Lyman absorption in evaluating $a(H)$. The values of $a(H)$ for various temperatures are also listed in Table 1.

In terms of $a(H^-)$ and $a(H)$ given in Table 1, we can determine the combined mass absorption coefficient $\bar{\kappa}$ according to

$$\bar{\kappa} = \frac{1 - x_H}{m_H}\,[a\,(H^-)\,p_e + a\,(H)]\;. \tag{6}$$

Values of $\bar{\kappa}(H^-)$, $\bar{\kappa}(H)$, and $\bar{\kappa}$ determined in accordance with the foregoing equations for various temperatures and electron pressures are given in Table 2.

3. *The continuous absorption in the solar atmosphere*—As we have already stated in the introduction, one of the principal problems in the interpretation of the solar spectrum is the identification of the source of absorption which will predict the same dependence

[11] It will be noted that, in writing the mass absorption coefficient in the form (2), we have assumed the preponderant abundance of hydrogen in the stellar atmosphere.

[12] See the reference quoted in n. 6.

of the absorption coefficient with wave length in the range 4000–24,000 A, which can be inferred from the intensity distribution in the continuous spectrum and the law of darkening in the different wave lengths. While the amount and variation of the continuous absorption in the spectral region λλ 4000–24,000 A can be deduced in a variety of ways,[13] it appears that, for the purposes of the identification of the physical source of absorption, it is most direct to adopt the following procedure:

TABLE 2

THE MEAN MASS ABSORPTION COEFFICIENTS $\bar{\kappa}(H^-)$, $\bar{\kappa}(H)$, AND $\bar{\kappa}$ FOR VARIOUS TEMPERATURES AND ELECTRON PRESSURES

		$p_e=1$	$p_e=10$	$p_e=10^2$	$p_e=10^3$	$p_e=10^4$
$\theta_e=0.5$	$\bar{\kappa}(H^-)$	5.85×10^{-6}	3.70×10^{-4}	4.98×10^{-2}	2.12	32.1
	$\bar{\kappa}(H)$	1.71×10^{-1}	1.08	14.5	61.8	93.5
	$\bar{\kappa}$	1.71×10^{-1}	1.08	14.6	63.9	126
$\theta_e=0.6$	$\bar{\kappa}(H^-)$	3.93×10^{-4}	2.60×10^{-2}	5.82×10^{-1}	6.68	68.4
	$\bar{\kappa}(H)$	4.18×10^{-1}	2.76	6.19	7.10	7.27
	$\bar{\kappa}$	4.18×10^{-1}	2.79	6.77	13.8	75.7
$\theta_e=0.7$	$\bar{\kappa}(H^-)$	9.08×10^{-3}	1.28×10^{-1}	1.34	13.4	134
	$\bar{\kappa}(H)$	4.54×10^{-1}	6.41×10^{-1}	0.67	0.7	1
	$\bar{\kappa}$	4.63×10^{-1}	7.69×10^{-1}	2.01	14.1	135
$\theta_e=0.8$	$\bar{\kappa}(H^-)$	2.26×10^{-2}	2.32×10^{-1}	2.32	23.2	232
	$\bar{\kappa}(H)$	4.66×10^{-2}	0.48×10^{-1}	0.05	....	....
	$\bar{\kappa}$	6.92×10^{-2}	2.80×10^{-1}	2.37	23.2	232
$\theta_e=0.9$	$\bar{\kappa}(H^-)$	3.63×10^{-2}	3.63×10^{-1}	3.63	36.3	363
	$\bar{\kappa}(H)$	0.33×10^{-2}	0.03×10^{-1}	....	....	....
	$\bar{\kappa}$	3.96×10^{-2}	3.66×10^{-1}	3.63	36.3	363
$\theta_e=1.0$	$\kappa(H^-)$	5.57×10^{-2}	5.57×10^{-1}	5.57	55.7	557
	$\bar{\kappa}(H)$	0.02×10^{-2}	....	....	....	....
	$\bar{\kappa}$	5.59×10^{-2}	5.57×10^{-1}	5.57	55.7	557
$\theta_e=1.2$	$\bar{\kappa}(H^-)$	1.20×10^{-1}	1.20	12.0	120	1200
$\theta_e=1.4$	$\bar{\kappa}(H^-)$	2.33×10^{-1}	2.33	23.3	233	2330

We compare the observed intensity distribution in the emergent solar flux F_λ (obs.) with the flux $F_\lambda^{(1)}(0)$ to be expected in a gray atmosphere.[14] It is evident that the *departures*,

$$\Delta \log F_\lambda = \log F_\lambda \text{ (obs.)} - \log F_\lambda^{(1)}(0) , \tag{7}$$

must be related more or less directly with the dependence of the continuous absorption coefficient κ_ν' with wave length. Indeed, in the approximations leading to the temperature distribution (1) this relation must be one-one, since, with the adopted definition of $\bar{\kappa}$, the temperature distributions in the gray and the nongray atmospheres agree.[15] This suggests that, with the known value of κ_ν' due to H^-, we compare the predicted de-

[13] G. Mülders, *Zs. f. Ap.*, **11**, 132, 1935; G. Münch, *Ap. J.*, **102**, 385, 1945; D. Chalonge and V. Kourganoff, *Ann. d'ap.* (in press).

[14] The values of $F_\lambda^{(1)}(0)$ can be readily derived from the entries along the line $\tau = 0$ in Table 2 of "Radiative Equilibrium VII."

[15] Cf. the remarks in italics on p. 343 in "Radiative Equilibrium VII."

partures from $F_\lambda^{(1)}(0)$ with those observed. The only uncertainty in these predictions will be of the nature of a "zero-point" correction, since a value of $\bar{\kappa}$ different from the one adopted will lead to an approximately constant additive correction to $\Delta \log F_\lambda$.[16]

In order, then, to make the comparison suggested in the preceding paragraph, we need to determine $\Delta \log F_\lambda$ in terms of κ'_λ due to H^- and an adopted $\bar{\kappa}$. Assuming in the first instance that the contribution to $\bar{\kappa}$ is only H^-, we find that

$$a(H^-) = 5.62 \times 10^{-26}\ \text{cm}^4/\text{dyne} \tag{8}$$

for an adopted value of

$$\theta_e = 0.8822\ . \tag{9}$$

The value of κ'_λ for $\theta = 0.8822$ can be found by simple interpolation in Table 7 of Chandrasekhar and Breen's paper. The ratios $\kappa'_\lambda/a(H^-)$ derived in this manner are

TABLE 3

THE PREDICTED DEPARTURES [LOG F_λ (obs.)−LOG $F_\lambda^{(1)}(0)$] FROM GRAYNESS OF THE SOLAR ATMOSPHERE DUE TO THE ABSORPTION BY H^-

λA	$\dfrac{\kappa'_\lambda}{\bar{\kappa}(H^-)}$	$\dfrac{\kappa'_\lambda}{1.42\bar{\kappa}(H^-)}$	LOG F_λ (THEO.)		LOG $F_\lambda^{(1)}(0)$	Δ LOG F_λ	
			$\bar{\kappa}=\bar{\kappa}(H^-)$	$\bar{\kappa}=1.42\bar{\kappa}(H^-)$		$\bar{\kappa}=\bar{\kappa}(H^-)$	$=\bar{\kappa}1.42\bar{\kappa}(H^-)$
4000	0.686	0.483	14.489	14.532	14.358	+0.131	+0.174
4500	0.783	.551	14.468	....	14.390	+ .078	....
5000	0.881	.620	14.431	14.534	14.399	+ .032	+ .135
6000	1.029	.725	14.341	14.423	14.350	− .009	+ .073
7000	1.132	.797	14.259	14.329	14.285	− .026	+ .044
8000	1.188	.837	14.164	14.224	14.194	− .030	+ .030
9000	1.183	.833	14.070	14.128	14.098	− .028	+ .030
10,000	1.125	.792	13.982	14.034	13.999	− .017	+ .035
11,000	1.028	.724	13.897	13.946	13.901	− .004	+ .045
12,000	0.911	.642	13.817	13.862	13.805	+ .012	+ .057
13,000	0.788	.555	13.742	13.784	13.712	+ .030	+ .072
14,000	0.651	.458	13.674	13.713	13.624	+ .050	+ .089
15,000	0.523	.368	13.608	13.643	13.537	+ .071	+ .106
16,000	0.481	.339	13.532	13.566	13.454	+ .078	+ .112
17,000	0.486	.342	13.448	13.480	13.374	+ .074	+ .106
18,000	0.516	.363	13.362	13.395	13.297	+ .065	+ .098
19,000	0.562	.396	13.279	13.313	13.223	+ .056	+ .090
20,000	0.618	.435	13.197	13.230	13.152	+ .045	+ .078
21,000	0.679	.478	13.118	13.149	13.083	+ .035	+ .066
22,000	0.749	.527	13.042	13.073	13.016	+ .026	+ .057
23,000	0.820	0.577	12.971	13.003	13.952	+0.019	+0.051

given in Table 3 for various values of λ. With these values of $\kappa'_\lambda/a(H^-)$, the theoretical determination of $\Delta \log F_\lambda$ is straightforward with the help of the nomogram of Burkhardt's table,[17] which one of us has recently published.[18] The results of the determination are given in Table 3. In Figure 1 we have further compared the computed departures $\Delta \log F_\lambda$ with those observed.[19] It is seen that the predicted variation of the

[16] This is seen most directly in an approximation in which we expand the source function $B_\lambda(T)$ as a Taylor series about a suitable point and determine the emergent flux in terms of it (see, e.g., A. Unsöld, *Physik der Sternatmosphären*, p. 109. eq. [31.18], Berlin, 1938).

[17] *Zs.f.Ap.*, 13, 56, 1936.

[18] G. Münch, *Ap. J.*, 102, 385, Fig. 2, 1945.

[19] For λ > 9000 A the observed departures were obtained from a reduction of the solar data by M. Minnaert, *B.A.N.*, 2, No. 51, 75, 1924; see also Unsöld, *op. cit.*, p. 32. For λ < 9000 A the reduction of G. Mülders (dissertation, Utrecht, 1934) was used.

departures runs remarkably parallel with the observed departures over the wavelength range 4000–20,000 A. (The observational data do not seem specially reliable for $\lambda > 20,000$ A.) However, the absolute values of the predicted departures are systematically less than the predicted departures by approximately a constant amount, indicating a zero-point correction in the sense that the adopted value of $\bar{\kappa}$ as due to H^- alone is somewhat too small. The calculations were accordingly repeated for other slightly larger values of $\bar{\kappa}$, and it was found that with $\bar{\kappa} = 1.42\ \kappa(H^-)$, the predicted and the observed departures agree entirely within the limits of the observational uncertainties over the whole region of the spectrum in which H^- contributes to the absorption. The

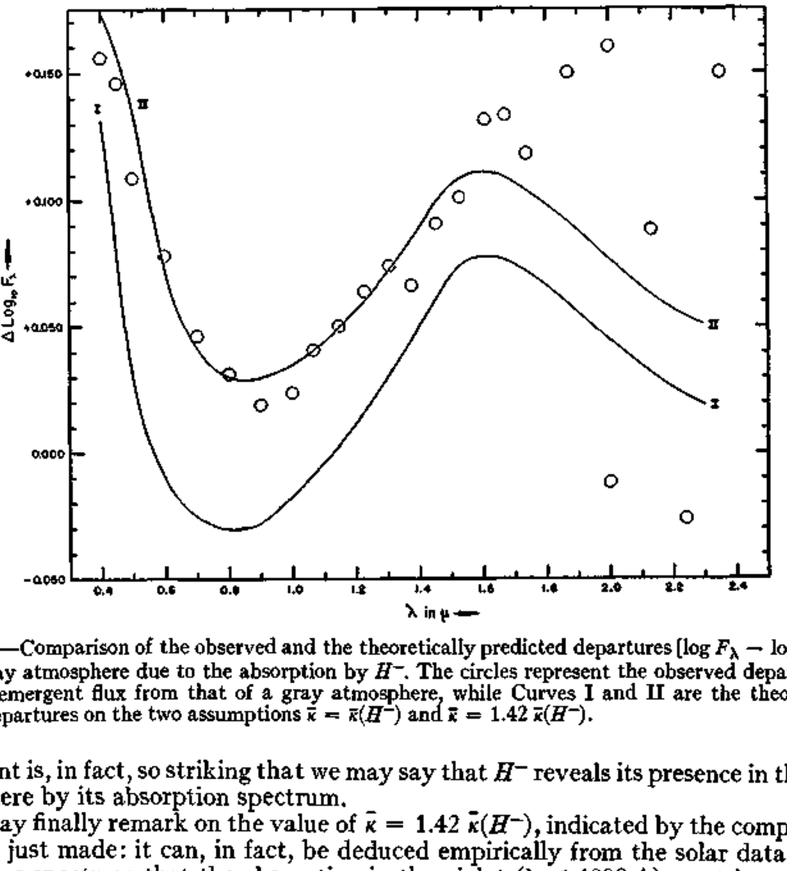


FIG. 1.—Comparison of the observed and the theoretically predicted departures $[\log F_\lambda - \log F_\lambda^1(0)]$ from a gray atmosphere due to the absorption by H^-. The circles represent the observed departures of the solar emergent flux from that of a gray atmosphere, while Curves I and II are the theoretically derived departures on the two assumptions $\bar{\kappa} = \bar{\kappa}(H^-)$ and $\bar{\kappa} = 1.42\ \bar{\kappa}(H^-)$.

agreement is, in fact, so striking that we may say that H^- reveals its presence in the solar atmosphere by its absorption spectrum.

We may finally remark on the value of $\bar{\kappa} = 1.42\ \bar{\kappa}(H^-)$, indicated by the comparisons we have just made: it can, in fact, be deduced empirically from the solar data on the continuous spectrum that the absorption in the violet ($\lambda < 4000$ A) must increase the value of κ derived from the visible and the infrared regions of the spectrum by a factor of the order of 1.5.[20]

4. *The predicted color-effective temperature relations: comparison with observations.*— All earlier attempts[21] to predict the color temperatures in the region $\lambda\lambda$ 4000–6500 A for stars of spectral types A0–G0 in agreement with the observations have failed. This failure in the past has been due to the following circumstance: The observed relation between the color and the effective temperatures and, in particular, the fact that $T_c > T_e$

[20] Cf. G. Münch, *Ap. J.*, **102**, 385, 1945, esp. the remarks preceding eq. (19) on p. 394.

[21] R. Wildt, *Ap. J.*, **93**, 47, 1941, and *Observatory*, **64**, 195, 1942; R. E. Williamson, *Ap. J.*, **97**, 51, 1943.

implies that the continuous absorption coefficient is an increasing function of λ in the spectral region observed. But the physical theory on which the calculations were made placed the maximum of the absorption-curve in the region of λ 4500 A; this was incompatible with the observations and, moreover, predicted color temperatures less than the effective temperatures, contrary to all evidence. Indeed, on the strength of this discrepancy, it was concluded that H^- as a source of absorption was inadequate even in the region $\lambda\lambda$ 4500–6500 A, and the existence of an unknown source operative in this region was further inferred. However, later evaluations[22] of the bound-free transitions of H^- showed the unreliability of earlier determinations and placed the maximum of the absorption-curve in the neighborhood of λ 8500 A. The addition of the free-free transitions pushes this maximum only still further to the red. It is therefore evident that on the revised physical theory we should be able to remove the major discrepancies of the subject. We shall now show how complete the resolution of these past difficulties is.

From the point of view of establishing the adequacy of the physical theory in the region $\lambda\lambda$ 4000–6500 A, it is most instructive to consider the theoretical predictions for color temperatures which can be directly compared with the color determinations at Greenwich,[3] for the Greenwich measures are based on the mean gradients in the wave-length interval 4100–6500 A, and it is in the prediction of these colors that the earlier calculations were most discordant.[23]

Now, from the Planck formula in the form

$$i_\lambda = \frac{2hc^2}{\lambda^5}\frac{1}{e^{c_2/\lambda T}-1}, \tag{10}$$

it readily follows that

$$\frac{1}{i_\lambda}\frac{d i_\lambda}{d(1/\lambda)} = 5\lambda - \frac{c_2}{T}\left(1 - e^{-c_2/\lambda T}\right)^{-1}. \tag{11}$$

Defining the gradient

$$\phi = \frac{c_2}{T}\left(1 - e^{-c_2/\lambda T}\right)^{-1} \tag{12}$$

in the usual manner, we can write

$$\frac{1}{M}\frac{d\log_{10} i_\lambda}{d(1/\lambda)} = 5\lambda - \phi \qquad \left(\frac{1}{M} = 2.303\right). \tag{13}$$

If F_{λ_1} and F_{λ_2} are the emergent fluxes at two wave lengths λ_1 and λ_2 and if ϕ is the mean gradient in this wave-length interval, then we can write, in accordance with equation (13)

$$\phi = 5\lambda_m - \frac{1}{M}\frac{\log_{10}(F_{\lambda_1}/F_{\lambda_2})}{\lambda_1^{-1} - \lambda_2^{-1}}, \tag{14}$$

where λ_m denotes an appropriate mean wave length for the interval to which the gradient ϕ refers. Equation (14) can be re-written in the following form:

$$\phi = 5\lambda_m - \frac{1}{M}\frac{\Delta\log_{10} F}{\Delta(1/\lambda)}. \tag{15}$$

According to equation (15), the theoretical determination of color temperatures will proceed by determining, first, the gradient ϕ from the values of F_λ at the end-points of the wave-length interval and then determining the temperature which will give this gradient.

[22] S. Chandrasekhar, *Ap. J.*, **102**, 223, 395, 1945

[23] Cf. Fig. 6 in Williamson's paper (*op. cit.*).

For the Greenwich measures $\lambda_m = 0.55\ \mu$, and equation (15) becomes

$$\phi\ (\text{Greenwich}) = 2.75 - 2.56 \log_{10}\left(\frac{F_{4100}}{F_{6500}}\right), \tag{16}$$

provided that, in determining the gradients, wave lengths are measured in microns.[24]

In Table 4 we have listed the values of $\kappa'_\lambda/\bar{\kappa}$ for the wave lengths 4100 A and 6500 A for various values of θ_e and p_e. In terms of these values the determination of the fluxes at the two wave lengths 4100 A and 6500 A is straightforward with the help of Burkhardt's table. The gradient ϕ then follows according to equation (16) and, from that, the color temperature. The reciprocal color temperatures $\theta_c = 5040/T_c$ derived in this manner are given in Table 5. The resulting color–effective temperature relations are illustrated

TABLE 4

$\kappa'_\lambda/\bar{\kappa}$ IN MODEL STELLAR ATMOSPHERES

$5040/T_e$	λA	$p_e=10$	$p_e=10^2$	$p_e=10^3$	$p_e=10^4$
	λ 3647	....	3.45 0.161	3.22 0.181	2.68 0.332
	λ 4000	0.210	0.212	0.238	0.397
$\theta=0.5$............	λ 4600	0.316	0.318	0.346	0.513
	λ 6500	0.826	0.834	0.848	0.991
	λ 8203	....	1.51 0.490	1.51 0.526	1.55 0.788
	λ 3647	....	4.04 0.178	2.60 0.427	1.09 0.685
	λ 4000	0.148	0.205	0.481	0.770
$\theta=0.6$............	λ 4600	0.221	0.278	0.585	0.901
	λ 6500	0.579	0.637	0.935	1.25
	λ 8203	....	1.27 0.424	1.36 0.885	1.45 1.36
	λ 3647	6.06 0.176	2.82 0.461	0.961 0.625	0.684 0.645
	λ 4000	0.208	0.521	0.695	0.725
$\theta=0.7$............	λ 4600	0.276	0.617	0.810	0.848
	λ 6500	0.546	0.901	1.11	1.14
	λ 8203	0.889 0.356	1.105 0.893	1.22 1.19	1.24
	λ 3647	1.54 0.518	0.720 0.601	0.625 0.614	.... 0.611
	λ 4000	0.582	0.671	0.685	0.685
$\theta=0.8$............	λ 4600	0.681	0.781	0.800	0.800
	λ 6500	0.961	1.06	1.09	1.09
	λ 8203	1.11 1.01	1.178 1.164	1.19	1.19
	λ 3647	0.708 0.631	0.642 0.636	0.636	0.636
	λ 4000	0.690	0.690	0.690	0.690
$\theta=0.9$............	λ 4600	0.800	0.800	0.800	0.800
	λ 6500	1.09	1.09	1.09	1.09
	λ 8203	1.16	1.16	1.16	1.16 ....

[24] The constant c_2 in eq. (12) then has the value 14,320.

in Figure 2. For comparison we have also plotted in this figure the Greenwich determinations for stars on the main sequence and of spectral types A0–G0 (reduced, however, to the Morgan, Keenan, and Kellman system of spectral classification). The color temperature of the sun for this wave-length interval is also plotted in Figure 1. It is seen from Figure 1 that the agreement between the observed and the theoretical color temperatures is entirely satisfactory, particularly when it is remembered that the earlier calculations failed even to predict the correct sign for $\theta_c - \theta_e$. It will, however, be noted that the observed values of θ_c for spectral types later than F0 are somewhat larger than the pre-

TABLE 5

THEORETICAL RECIPROCAL COLOR TEMPERATURES AND THE PREDICTED DISCONTINUITIES AT THE HEAD OF THE BALMER AND THE PASCHEN SERIES*

p_e		θ_e = 0.5	0.6	0.7	0.8	0.9	1.0
10	$\theta_c(G)$	....	....	....	0.68	0.80	0.88
	$\theta_c(B \text{ and } C)$	....	....	....	.61	.69	.75
	D_B	....	....	....	.31	.015	....
10^2	$\theta_c(G)$	....	0.45	0.60	.73	.80	.88
	$\theta_c(B \text{ and } C)$	....	....	.47	.63	.69	.75
	D_B	....	....	[.50]	.07	....	....
	D_P	....	....	.030	.001	....	....
10^3	$\theta_c(G)$	[0.31]	.52	.65	.74	.80	.88
	$\theta_c(B \text{ and } C)$	....	.42	.56	.64	.69	.75
	D_B	[.58]	[.34]	.13	....	....	....
	D_P	.114	.051	.003	....	....	....
10^4	$\theta_c(G)$	[.39]	.59	.66	.74	.80	.88
	$\theta_c(B \text{ and } C)$	[.34]	.50	.58	.64	.69	.75
	D_B	[.40]	.12	.018	....	....	....
	D_P	[.071]	.006	....	....	....	....
Pure H^-	$\theta_c(G)$	.52	.59	.66	.74	.80	.88
	$\theta_c(B \text{ and } C)$	0.41	0.50	0.58	0.64	0.69	0.75

* $\theta_c(G)$ and $\theta_c(B \text{ and } C)$ are the reciprocal color temperatures, $5040/T_c$, appropriate for the wave-length intervals 4100–6500 A and 4000–4600 A, respectively; D_B and D_P, representing the logarithm of the ratio of the fluxes at the two sides of the series limits, are the expected Balmer and Paschen discontinuities, respectively.

dicted values, though the agreement is as good as can be expected in the case of the sun. The reason for this must undoubtedly be the crowding of the absorption lines toward the violet in the later spectral types and the consequent depression of the continuous spectrum in this region. The correctness of this explanation is apparent when it is noted that in the case of the sun, in which allowance has been made for this effect of the lines on the continuum, the discordance is not present.

Comparisons similar to those we have just made also can be made with the measurements of Barbier and Chalonge[4] on the color temperatures based on the observed gradients in the wave-length interval 4000–4600 A. The formula giving the theoretical gradient for this wave-length interval takes the form

$$\phi\,(B \text{ and } C) = 2.175 - 7.06 \log_{10}\left(\frac{F_{4000}}{F_{4600}}\right). \qquad (17)$$

The values of $\kappa_\lambda'/\bar{\kappa}$ at λ 4000 A and λ 4600 A are given in Table 4, and the reciprocal color temperatures derived from these in Table 5. The results are further illustrated in Figure 3, where the theoretical relations for various electron pressures are compared with the measures of Barbier and Chalonge (reduced also to the Morgan, Keenan, and Kellman system of spectral classification). It is seen that the general agreement is again good, though there are now somewhat larger differences between the computed and the observed color temperatures for spectral types later than F0 than were encountered in the comparison with the Greenwich colors. This must again be due to the crowding of the absorption lines toward the violet in the later spectral types and the further fact that

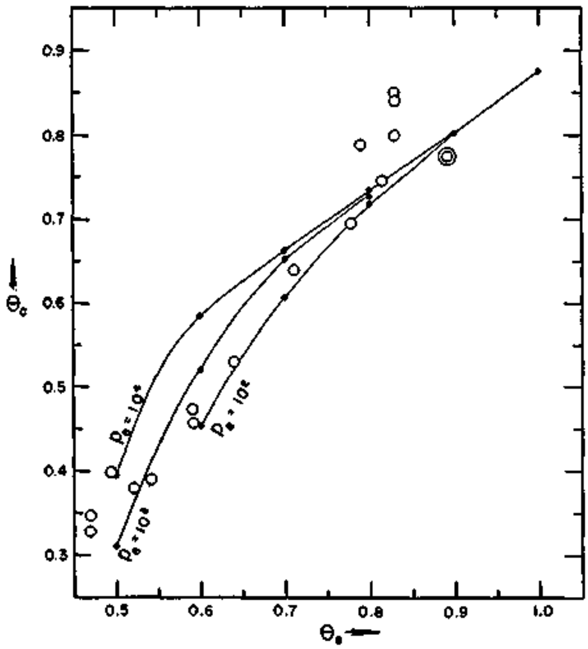


FIG. 2.—Comparison of the observed and the theoretical color effective-temperature relations for the wave-length interval 4100–6500 A. The ordinates denote the reciprocal color temperatures and the abscissae denote the reciprocal effective temperatures ($\theta = 5040/T$). The circles represent the Greenwich color determinations reduced to the Morgan, Keenan, and Kellman system of spectral classification. The double circle represents the sun.

the base line for the Barbier and Chalonge colors is much shorter than that for the Greenwich colors.

5. *The discontinuities at the head of the Balmer and the Paschen series.*—With the physical theory of the continuous absorption coefficient now available, we can also predict the extent of the discontinuities which we may expect at the head of the Balmer and the Paschen series of hydrogen. For this purpose the values of $\kappa_\lambda'/\bar{\kappa}$ on the two sides of the series limits are also given in Table 4. From these values it is a simple matter to estimate the discontinuities which will exist at the head of the Balmer and the Paschen series, and they are given in Table 5. The results for the Balmer discontinuities are further illustrated in Figure 4, in which the discontinuities measured by Barbier and Chalonge[4] for various stars are also plotted. The progressive increase of the electron pressure as we go from the later to the earlier spectral types is particularly apparent

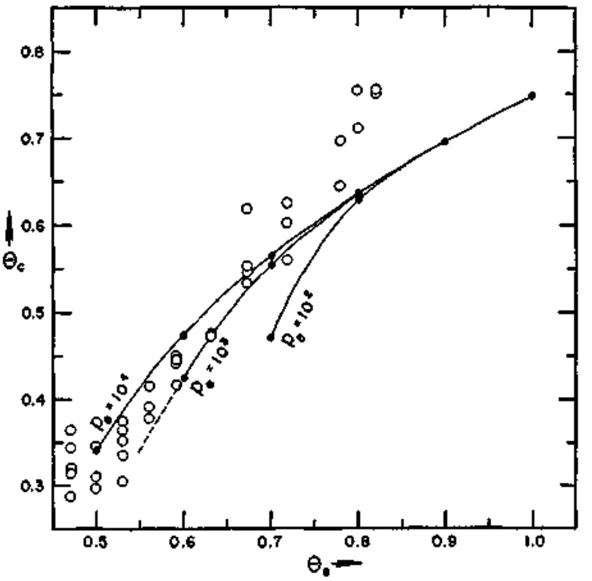


FIG. 3.—Comparison of the observed and the theoretical color effective-temperature relations for the wave-length interval 4000–4600 A. The ordinates denote the reciprocal color temperatures, and the abscissae denote the reciprocal effective temperatures $\theta = 5040/T$). The circles represent the color determinations of Barbier and Chalonge for the wave-length interval 4000–4600 A, reduced to the Morgan, Keenan, and Kellman system of spectral classification.

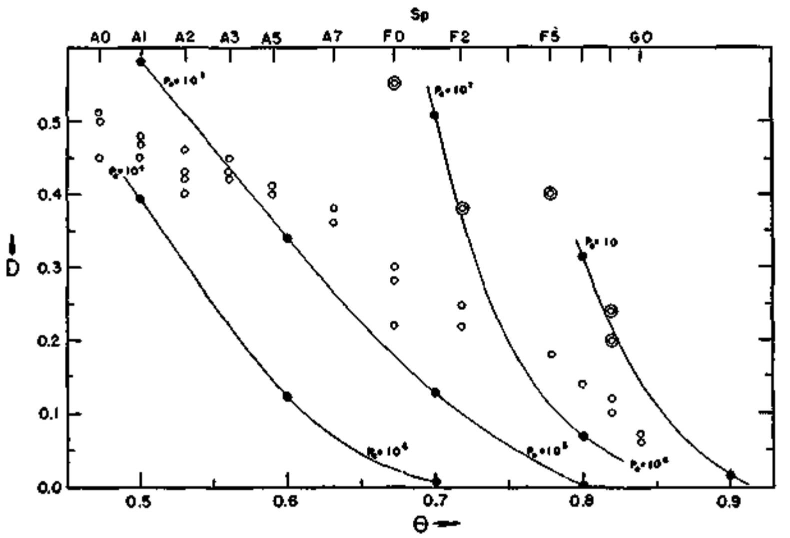


FIG. 4.—The predicted discontinuities (D) at the head of the Balmer series for various effective temperatures and electron pressures. The circles represent the discontinuities as measured by Barbier and Chalonge. (The double circles represent the observations for supergiants.)

from Figure 4; this progression is, moreover, in agreement with what is indicated by the color determinations (cf. Figs. 2 and 3). In Table 6 we give the electron pressures for the main-sequence stars of various spectral types estimated in this manner.

6. *Concluding remarks.*—While our discussion in the preceding sections has established the unique role which H^- plays in determining the character of the continuous spectrum of the sun and the stars, it should not be concluded that the various other

TABLE 6

ELECTRON PRESSURES FOR STARS ON THE MAIN SEQUENCE

Type	log p_e	Type	log p_e	Type	log p_e
A1	3.7	F0	2.4	F8	1.4
A2	3.3	F2	2.2	G0	1.2
A3	3.0	F4	2.0	G2	1.0
A5	2.8	F5	1.8		
A7	2.6	F6	1.6		

astrophysical elements of the theory are equally well established. Indeed, the theory of model stellar atmospheres as developed by Strömgren in recent years must not only be revised on the basis of the new absorption coefficients of H^- but also be advanced still further before we can be said to have a completely satisfactory account of all the classical problems of the theory of stellar atmospheres. But our discussion in this paper does give us the confidence that the continuous absorption by H^- discovered by Wildt must provide the key to the solution of many of these problems.

This page is intentionally left blank

III. Stochastic and Statistical Problems in Astronomy

Paper 1 in this part is one of the most celebrated, most widely quoted papers in scientific literature. Chandra had prepared this article for himself, with no intention of publishing it. He was in Princeton at the time, working with John von Neumann. When Chandra showed the article to Neumann, the latter insisted that it be published and took it upon himself to send it to *Reviews of Modern Physics*. The paper reviews certain fundamental probability methods that find applications in a wide variety of problems and in fields as different as colloidal chemistry and stellar dynamics.

The subject matter of Paper 2 is stellar encounters and the influence of fluctuating stellar distribution on the motion of stars. In estimating the influence, it was generally assumed that the effect was a cumulative result of a large number of separate two-body encounters. Chandra provides physical arguments to the effect that this is not a good approximation. He introduces a new approach to treating the fluctuating field based on statistical methods. This forms the starting point of a series of papers along with John von Neumann's in which a statistical analysis of the gravitational field arising from a random distribution of stars is presented. Paper 3 provides a condensed but excellent account of the new methods Chandra had developed in the field of stellar dynamics.

Paper 4 marks the beginning of still another series of papers involving statistical methods. In an introductory note on this series, Chandra discusses the influence of the principle of invariance formulated by Ambartsumian (Author's Note, p. xiii, in *Selected Papers,* Vol. 3; University of Chicago Press, 1989). On the assumption of a uniform distribution of interstellar clouds and stars, satisfying a condition on the transparency factor, Ambartsumian had formulated the invariance principle and derived a nonlinear integral equation governing the required probability distribution in the limit of infinite optical depth. Chandra, along with Munch, extended the principle in the case of a more general and a more realistic system of finite extension, and obtained the moments of the probability distribution in terms of the reciprocal of the Vandermondie matrix.

III. Stochastic and Statistical Problems in Astronomy

Stochastic Problems in Physics and Astronomy

S. Chandrasekhar
Yerkes Observatory, The University of Chicago, Williams Bay, Wisconsin

Contents

APPENDIXES

INTRODUCTION

IN this review we shall consider certain fundamental probability methods which are finding applications increasingly in a wide variety of problems and in fields as different as colloid chemistry and stellar dynamics. However, a common characteristic of all these problems is that interest is focused on a property which is the result of superposition of a large number of variables, the values which these variables take being governed by certain probability laws. We may cite as illustrations two examples:

(i) The first example is provided by the *problem of random flights*. In this problem, a particle undergoes a sequence of displacements $\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_i, \cdots$, the magnitude and direction of each displacement being independent of all the preceding ones. But the probability that the displacement $\mathbf{r}_i$ lies between $\mathbf{r}_i$ and $\mathbf{r}_i+d\mathbf{r}_i$ is governed by a distribution function $\tau_i(\mathbf{r}_i)$ assigned *a priori*. We ask: What is the probability $W(\mathbf{R})d\mathbf{R}$ that after N displacements the coordinates of the particle lie in the interval $\mathbf{R}(=[x, y, z])$ and $\mathbf{R}+d\mathbf{R}$. It is seen that in this problem the position $\mathbf{R}$ of the particle is the resultant of N vectors, $\mathbf{r}_i$, $(i=1, \cdots, N)$ the position and direction of each vector being governed by the probability distributions $\tau_i(\mathbf{r}_i)$. As we shall see the solution to this problem provides us with one of the principal weapons of the theory.[1]

(ii) We shall take our second illustration from stellar dynamics. The gravitational force acting on a star (per unit mass) is given by

$$\mathbf{F}=G\Sigma M_i\mathbf{r}_i/|\mathbf{r}_i|^3 \tag{1}$$

where M_i denotes the mass of a typical "field" star and $\mathbf{r}_i$ its position vector relative to the star under consideration and G the constant of gravitation. Further in Eq. (1) the summation is extended over all the neighboring stars. We now suppose that the distribution of stars in the neighborhood of a given one is subject to fluctuations and that stars of different masses occur in the stellar system according to some well defined empirically established law. However, the fluctuations in density are assumed to be subject to the restriction of a constant average density of n stars per unit volume. We ask: What is the probability that $\mathbf{F}$ lies between $\mathbf{F}$ and $\mathbf{F}+d\mathbf{F}$? Again, the force acting on a star is the resultant of the forces due to all the neighboring stars while the spatial distribution of these stars and their masses are subject to well-defined laws of fluctuations.

From the foregoing two examples it is clear that one of the principal problems under the circumstances envisaged is the specification of the distribution function $W(\mathbf{\Phi})$ of a quantity $\mathbf{\Phi}$ (in general a vector in hyper-space) which is the resultant of a large number of other quantities having assigned distributions over a range of values. A second fundamental problem in the theories we shall consider concerns questions relating to *probability after-effects*[2]—a notion first introduced by Smoluchowski. We may broadly describe the nature of these questions in the following terms: A certain quantity $\mathbf{\Phi}$ is characterized by a stationary distribution $W(\mathbf{\Phi})$. We first make an observation of $\mathbf{\Phi}$ at a certain instant of time $t=0$ (say) and again repeat our

[1] For historical remarks on this problem of random flights see the Bibliographical Notes at the end of the article.

[2] This is the translation of the German word "Wahrscheinlichkeitsnachwirkung" coined by M. von Smoluchowski.

observation at a later time t. We ask: What can we say about the possible values of Φ which we may expect to observe at time t when we already know that Φ had a particular value at $t=0$? It is clear that if the second observation were made after a sufficiently long interval of time, we should not, in general, expect any correlation with the fact that Φ had a particular value at a very much earlier epoch. On the other hand as $t\to 0$ the values which we would expect to observe on the second occasion will be strongly dependent on what we observed on the earlier occasion.

An example considered by Smoluchowski in colloid statistics illustrates the nature of the problem presented in theories of probability after-effects: Suppose we observe by means of an ultramicroscope a small well-defined element of volume of a colloidal solution and count the number of particles in the element at definite intervals of time τ, 2τ, 3τ, etc., and record them consecutively. We shall further suppose that the interval τ between successive observations is not large. Then the number which is observed on any particular occasion will be correlated in a definite manner with what was observed on the immediately preceding occasion. This correlation will depend on a variety of physical factors including the viscosity of the medium: thus it is clear from general considerations that the more viscous the surrounding medium the greater will be the correlation in the numbers counted on successive occasions. We shall discuss this problem following Smoluchowski in some detail in Chapter III but pass on now to the consideration of another example typical of this theory.

We have already indicated that a fundamental problem in stellar dynamics is the specification of the distribution function $W(\boldsymbol{F})$ governing the probability of occurrence of a force $\boldsymbol{F}$ per unit mass acting on a star. Suppose that $\boldsymbol{F}$ has a definite value at a given instant of time. We can ask: How long a time should elapse on the average before the force acting on the star can be expected to have no appreciable correlation with the fact of its having had a particular value at the earlier epoch? In other words, what is the *mean life* of the state of fluctuation characterized by $\boldsymbol{F}$? In a general way it is clear that this mean life will depend on the state of stellar motions in the neighborhood of the star under consideration in contrast to the probability distribution $W(\boldsymbol{F})$ which depends only on the average number of stars per unit volume. The two examples we have cited are typical of the problems which are properly in the province of the theory dealing with probability after-effects.

A physical problem, the complete elucidation of which requires both the types of theories outlined in the preceding paragraphs, is provided by Brownian motion. We shall accordingly consider certain phases of this theory also.

CHAPTER I

THE PROBLEM OF RANDOM FLIGHTS

The problem of random flights which in its most general form we have already formulated in the introduction provides an illustrative example in reference to which we may develop several of the principal methods of the theories we wish to describe. Accordingly, in this chapter, in addition to providing the general solution of the problem, we shall also discuss it from several different points of view.

1. The Simplest One-Dimensional Problem: The Problem of Random Walk

The principal features of the solution of the problem of random flights in its most general form are disclosed and more clearly understood by considering first the following simplest version of the problem in one dimension:

A particle suffers displacements along a straight line in the form of a series of *steps* of equal length, each step being taken, either in the forward, or in backward direction with equal probability $\frac{1}{2}$. After taking N such steps the particle *could* be at any of the points[3]

$$-N,\ -N+1,\ \cdots,\ -1,\ 0,\ +1,\ \cdots,\ N-1 \text{ and } N.$$

We ask: What is the probability $W(m, N)$ that the particle arrives at the point m after suffering N displacements?

We first remark that in accordance with the conditions of the problem each individual step is equally likely to be taken either in the back-

[3] These can be regarded as the coordinates along a straight line if the unit of length be chosen to be equal to the length of a single step.

ward or in the forward direction quite independently of the direction of all the preceding ones. Hence, all possible sequences of steps each taken in a definite direction have the same probability. In other words, the probability of any given sequence of N steps is $(\frac{1}{2})^N$. The required probability $W(m, N)$ is therefore $(\frac{1}{2})^N$ times the number of distinct sequences of steps which will lead to the point m after N steps. But in order to arrive at m among the N steps, *some* $(N+m)/2$ steps should have been taken in the positive direction and the remaining $(N-m)/2$ steps in the negative direction. (Notice that m can be even or odd only according as N is even or odd.) The number of such distinct sequences is clearly

$$N!/[\tfrac{1}{2}(N+m)]![\tfrac{1}{2}(N-m)]!. \tag{2}$$

Hence

$$W(m, N)=\frac{N!}{[\frac{1}{2}(N+m)]![\frac{1}{2}(N-m)]!}\left(\frac{1}{2}\right)^N. \tag{3}$$

In terms of the binomial coefficients C_r^n's we can rewrite Eq. (3) in the form

$$W(m, N)=C^N_{(N+m)/2}\left(\frac{1}{2}\right)^N, \tag{4}$$

in other words we have a *Bernoullian distribution*. Accordingly, the expectation and the mean square deviation of $(N+m)/2$ are (see Appendix I)

$$\left.\begin{aligned}\tfrac{1}{2}\langle N+m\rangle_{\text{Av}}&=\tfrac{1}{2}N,\\ \langle[\tfrac{1}{2}(N+m)-\tfrac{1}{2}N]^2\rangle_{\text{Av}}&=\tfrac{1}{4}N.\end{aligned}\right\} \tag{5}$$

Hence,

$$\langle m\rangle_{\text{Av}}=0;\quad \langle m^2\rangle_{\text{Av}}=N. \tag{6}$$

The root mean square displacement is therefore $\sqrt{N}$.

We return to formula (3): The case of greatest interest arises when N is large and $m\ll N$. We can then simplify our formula for $W(m, N)$ by using Stirling's formula

$$\log n!=(n+\tfrac{1}{2})\log n-n+\tfrac{1}{2}\log 2\pi+O(n^{-1})(n\to\infty). \tag{7}$$

Accordingly when $N\to\infty$ and $m\ll N$ we have

$$\begin{aligned}\log W(m, N)\simeq&(N+\tfrac{1}{2})\log N\\&-\tfrac{1}{2}(N+m+1)\log\left[\frac{N}{2}\left(1+\frac{m}{N}\right)\right]\\&-\tfrac{1}{2}(N-m+1)\log\left[\frac{N}{2}\left(1-\frac{m}{N}\right)\right]\\&-\tfrac{1}{2}\log 2\pi-N\log 2.\end{aligned} \tag{8}$$

But since $m\ll N$ we can use the series expansion

$$\log\left(1\pm\frac{m}{N}\right)=\pm\frac{m}{N}-\frac{m^2}{2N^2}+O(m^3/N^3). \tag{9}$$

Equation (8) now becomes

$$\begin{aligned}\log W(m, N)\simeq&(N+\tfrac{1}{2})\log N-\tfrac{1}{2}\log 2\pi-N\log 2\\&-\tfrac{1}{2}(N+m+1)\left(\log N-\log 2+\frac{m}{N}-\frac{m^2}{2N^2}\right)\\&-\tfrac{1}{2}(N-m+1)\left(\log N-\log 2-\frac{m}{N}-\frac{m^2}{2N^2}\right).\end{aligned} \tag{10}$$

Simplifying the right-hand side of this equation we obtain

$$\log W(m, N)\simeq-\tfrac{1}{2}\log N+\log 2-\tfrac{1}{2}\log 2\pi-m^2/2N. \tag{11}$$

In other words, for large N we have the asymptotic formula

$$W(m, N)=(2/\pi N)^{\frac{1}{2}}\exp(-m^2/2N). \tag{12}$$

A numerical comparison of the two formulae (3) and (12) is made in Table I for $N=10$. We see that even for $N=10$ the asymptotic formula gives sufficient accuracy.

Now, when N is large it is convenient to introduce instead of m the net displacement x from the starting point as the variable:

$$x=ml \tag{13}$$

where l is the length of a step. Further, if we consider intervals Δx along the straight line which are large compared with the length of a

TABLE I. The problem of random walk: the distribution $W(m, N)$ for $N=10$.

m	From (3)	From (12)
0	0.24609	0.252
2	0.20508	0.207
4	0.11715	0.113
6	0.04374	0.042
8	0.00977	0.010
10	0.00098	0.002

step we can ask the probability $W(x)\Delta x$ that the particle is likely to be in the interval $x, x+\Delta x$ after N displacements. We clearly have

$$W(x, N)\Delta x = W(m, N)(\Delta x/2l), \quad (14)$$

since m can take only even or odd values depending on whether N is even or odd. Combining Eqs. (12), (13), and (14) we obtain

$$W(x, N) = \frac{1}{(2\pi Nl^2)^{\frac{1}{2}}} \exp(-x^2/2Nl^2). \quad (15)$$

Suppose now that the particle suffers n displacements per unit time. Then the probability $W(x, t)\Delta x$ that the particle will find itself between x and $x+\Delta x$ after a time t is given by

$$W(x, t)\Delta x = \frac{1}{2(\pi Dt)^{\frac{1}{2}}} \exp(-x^2/4Dt)\Delta x, \quad (16)$$

where we have written

$$D = \tfrac{1}{2}nl^2. \quad (17)$$

We shall see in §4 that the solution to the general problem of random flights has precisely this form.

2. Random Walk with Reflecting and Absorbing Barriers

In this section we shall continue the discussion of the problem of random walk in one dimension but with certain restrictions on the motion of the particle introduced by the presence of reflecting or absorbing walls. We shall first consider the influence of a reflecting barrier.

(a) *A Reflecting Barrier at* $m = m_1$

Without loss of generality we can suppose that $m_1 > 0$. Then, the interposition of the reflecting barrier at m_1 has simply the effect that whenever the particle arrives at m_1 it has a probability unity of retracing its step to $m_1 - 1$ when it takes the next step. We now ask the probability $W(m, N; m_1)$ that the particle will arrive at $m(\leqslant m_1)$ after N steps.

For the discussion of this problem it is convenient to trace the course of the particle in an (m, N)-plane as in Fig. 1. In this diagram, the displacement of a particle by a step means that the representative point moves upward by

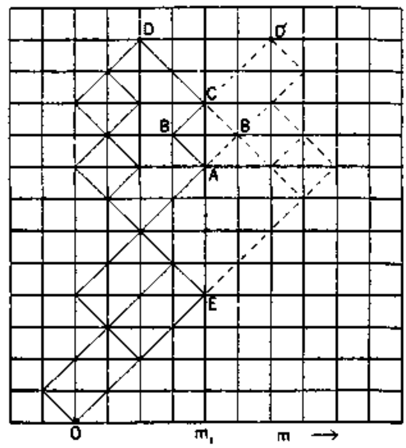


FIG. 1.

one unit while at the same time it suffers a lateral displacement also by one unit either in the positive or in the negative direction.

In the absence of a reflecting wall at $m = m_1$ the probability that the particle arrives at m after N steps is of course given by Eq. (3). But the presence of the reflecting wall requires $W(m, N)$ according to (3) to be modified to take account of the fact that a path reaching m after n reflections must be counted 2^n times since at each reflection it has a probability unity of retracing its step. It is now seen that we can take account of the relevant factors by adding to $W(m, N)$ the probability $W(2m_1 - m, N)$ of arriving at the "*image*" point $(2m_1 - m)$ after N steps (also in the absence of the reflecting wall), i.e.,

$$W(m, N; m_1) = W(m, N) + W(2m_1 - m, N). \quad (18)$$

We can verify the truth of this assertion in the following manner: Consider first a path like OED which has suffered just one reflection at m_1. By reflecting this path about the vertical line through m_1 we obtain a trajectory leading to the image point $(2m_1 - m)$ and conversely, for every trajectory leading to the image point, having crossed the line through m_1 once, there is exactly one which leads to m after a single reflection. Thus, instead of counting twice each trajectory reflected once, we can add a uniquely defined trajectory leading to $(2m_1 - m)$. Consider next a trajectory like $OABCD$ which leads to m after two reflections. A trajectory like this should be counted four times. But there are two trajectories ($OAB'CD$ and $OABCD'$) leading to the image point and a third ($OAB'CD$) which we should exclude on account of the barrier. These three additional trajectories together with $OABCD$ give exactly four trajectories leading either to m or its image $2m_1-m$ in the absence of the reflecting barrier. In this manner the arguments can be extended to prove the general validity of (18).

If we pass to the limit of large N Eq. (18) becomes [cf. Eq. (12)]

$$W(m,N;m_1)=\left(\frac{2}{\pi N}\right)^{\frac{1}{2}}\{\exp(-m^2/2N)+\exp[-(2m_1-m)^2/2N]\}. \quad (19)$$

Again, if as in §1 we use the net displacement $x=ml$ as the variable and consider the probability $W(x,t;x_1)\Delta x$ that the particle is between x and $x+\Delta x$, ($\Delta x\gg l$) after a time t (during which time it has taken nt steps) in the presence of a reflecting barrier at $x_1=m_1l$, we have

$$W(x,t;x_1)=\frac{1}{2(\pi Dt)^{\frac{1}{2}}}\{\exp(-x^2/4Dt)+\exp[-(2x_1-x)^2/4Dt]\}. \quad (20)$$

We may note here for future reference that according to Eq. (20)

$$(\partial W/\partial x)_{x=x_1}\equiv 0. \quad (21)$$

(*b*) *Absorbing Wall at* $m=m_1$

We shall now consider the case when there is a perfectly absorbing barrier at $m=m_1$. The interposition of the perfect absorber at m_1 means that whenever the particle arrives at m_1 it at once becomes incapable of suffering further displacements.[4] There are two questions which we should like to answer under these circumstances. The first is the analog of the problems we have considered so far, namely the probability that the particle arrives at $m(\leqslant m_1)$ after taking N steps. The second question which is characteristic of the present problem concerns the average rate at which the particle will deposit itself on the absorbing screen.

Considering first the probability $W(m,N;m_1)$, it is clear that in counting the number of distinct sequences of steps which lead to m we should be careful to exclude all sequences which include even a single arrival to m_1. In other words, if we first count *all* possible sequences which lead to m in the absence of the absorbing screen we should then exclude a certain number of "*forbidden*" sequences. It is evident, on the other hand, that every such forbidden sequence uniquely defines another sequence leading to the image $(2m_1-m)$ of m on the line $m=m_1$ in the (m,N)-plane (see Fig 1) and conversely. For, by reflecting about the line $m=m_1$ the part of a forbidden trajectory *above* its last point of contact with the line $m=m_1$ before arriving at m we are led to a trajectory leading to the image point, and conversely for every trajectory leading to $2m_1-m$ we necessarily obtain by reflection a forbidden trajectory leading to m (since any trajectory leading to $2m_1-m$ must necessarily cross the line $m=m_1$). Hence,

$$W(m,N;m_1)=W(m,N)-W(2m_1-m,N). \quad (22)$$

For large N we have

$$W(m,N;m_1)=(2/\pi N)^{\frac{1}{2}}\{\exp(-m^2/2N)-\exp[-(2m_1-m)^2/2N]\}. \quad (23)$$

Similarly, analogous to Eq. (21) we now have

$$W(x,t;x_1)=\frac{1}{2(\pi Dt)^{\frac{1}{2}}}\{\exp(-x^2/4Dt)-\exp[-(2x_1-x)^2/4Dt]\}. \quad (24)$$

We may further note that according to this equation

$$W(x_1,t;x_1)\equiv 0. \quad (25)$$

Turning next to our second question concerning the probable rate at which the particle deposits itself on the absorbing screen, we may first formulate the problem more specifically. What we wish to know is simply the probability $a(m_1,N)$ that after taking N steps the particle will arrive at m_1 *without ever* having touched or crossed the line $m=m_1$ at any earlier step.

First of all it is clear that N should have to be even or odd depending on whether m_1 is even

[4] This problem has important applications to other physical problems.

or odd. We shall suppose that this is the case. Suppose now that there is no absorbing screen. Then the arrival of the particle at m_1 after N steps implies that its position after $(N-1)$ steps must have been either (m_1-1) or (m_1+1). (See Fig. 2.) But every trajectory which arrives at (m_1, N) from $(m_1+1, N-1)$ is a forbidden one in the presence of the absorbing screen since such a trajectory must necessarily have crossed the line $m=m_1$. It does *not* however follow that *all* trajectories arriving at (m_1, N) from $(m_1-1, N-1)$ are permitted ones: For, a certain number of these trajectories will have touched or crossed the line $m=m_1$ earlier than its last step. The number of such trajectories arriving at $(m_1-1, N-1)$ but having an earlier contact with, or a crossing of, the line $m=m_1$ is equal to those arriving at $(m_1+1, N-1)$. The argument is simply that by reflection about the line $m=m_1$ we can uniquely derive from a trajectory leading to $(m_1+1, N-1)$ another leading to $(m_1-1, N-1)$ which has a forbidden character, and conversely. Thus, the number of permitted ways of arriving at m_1 for the first time after N steps is equal to *all* the possible ways of arriving at m_1 after N steps in the absence of the absorbing wall *minus* twice the number of ways of arriving at $(m_1+1, N-1)$ again in the absence of the absorbing screen: i.e.,

$$\frac{N!}{[\frac{1}{2}(N-m_1)]![\frac{1}{2}(N+m_1)]!}-2\frac{(N-1)!}{[\frac{1}{2}(N+m_1)]![\frac{1}{2}(N-m_1-2)]!}$$
$$=\frac{N!}{[\frac{1}{2}(N-m_1)]![\frac{1}{2}(N+m_1)]!}\left(1-\frac{N-m_1}{N}\right). \quad (26)$$
$$=\frac{m_1}{N}\frac{N!}{[\frac{1}{2}(N-m_1)]![\frac{1}{2}(N+m_1)]!}.$$

The required probability $a(m_1, N)$ is therefore given by

$$a(m_1, N)=\frac{m_1}{N}W(m_1, N). \quad (27)$$

For the limiting case of large N we have

$$a(m_1, N)=\frac{m_1}{N}\left(\frac{2}{\pi N}\right)^{\frac{1}{2}}\exp\left(-m_1^2/2N\right). \quad (28)$$

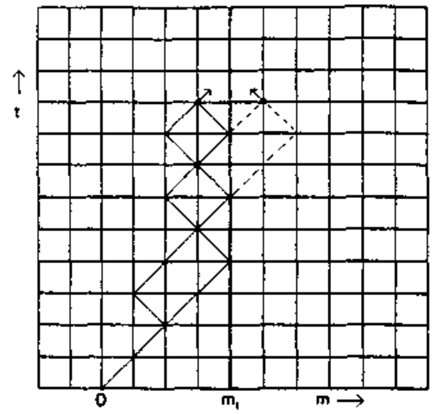


FIG. 2.

If we further write

$$x_1=m_1l;\quad N=nt;\quad D=\tfrac{1}{2}nl^2, \quad (29)$$

where l is the length of each step and n the number of displacements (assumed constant) which the particle suffers in unit time, then

$$a(x_1, t)=\frac{x_1}{nt}\frac{1}{(\pi Dt)^{\frac{1}{2}}}\exp\left(-x_1^2/4Dt\right). \quad (30)$$

Finally, if we ask the probability $q(x_1, t)\Delta t$ that the particle arrives at x_1 during t and $t+\Delta t$ for the first time, then

$$q(x_1, t)\Delta t=\tfrac{1}{2}a(x_1, t)n\Delta t, \quad (31)$$

since (30) is the number which arrive at x_1 in the time taken to traverse two steps. Thus,

$$q(x_1, t)=\frac{x_1}{t}\frac{1}{2(\pi Dt)^{\frac{1}{2}}}\exp\left(-x_1^2/4Dt\right). \quad (32)$$

We can interpret Eq. (32) as giving the fraction of a large number of particles initially at $x=0$ and which are deposited on the absorbing screen per unit time, at time t.

We readily verify that $q(x_1, t)$ as defined by Eq. (32) satisfies the relation

$$q(x_1, t)=-D(\partial W/\partial x)_{x=x_1}, \quad (33)$$

with W defined as in Eq. (24). This equation has an important physical interpretation to which we shall draw attention in §5.

3. The General Problem of Random Flights: Markoff's Method

In the general problem of random flights, the position $\boldsymbol{R}$ of the particle after N displacements is given by

$$\boldsymbol{R}=\sum_{i=1}^{N}\boldsymbol{r}_i, \tag{34}$$

where the $\boldsymbol{r}_i$'s ($i=1,\cdots,N$) denote the different displacements. Further, the probability that the ith displacement lies between $\boldsymbol{r}_i$ and $\boldsymbol{r}_i+d\boldsymbol{r}_i$ is given by

$$\tau_i(x_i,y_i,z_i)dx_idy_idz_i=\tau_id\boldsymbol{r}_i \quad (i=1,\cdots,N). \tag{35}$$

We require the probability $W_N(\boldsymbol{R})d\boldsymbol{R}$ that the position of the particle after N displacements lies in the interval $\boldsymbol{R}$, $\boldsymbol{R}+d\boldsymbol{R}$. In this general form the problem can be solved by using a method originally devised by A. A. Markoff. Now, Markoff's method is of such extreme generality that it actually enables us to solve the first of the two fundamental problems outlined in the introductory section. We shall accordingly describe Markoff's method in a form in which it can readily be applied to other problems besides that of random flights.

Let

$$\boldsymbol{\phi}_j=(\phi_j^1,\phi_j^2,\cdots,\phi_j^n) \quad (j=1,\cdots,N) \tag{36}$$

be N, n-dimensional vectors, the components of each of these vectors being functions of s coordinates:

$$\phi_j^k=\phi_j^k(q_j^1,\cdots,q_j^s) \quad (k=1,\cdots,n;\ j=1,\cdots,N). \tag{37}$$

The probability that the q_j^i's occur in the range

$$q_j^1,\ q_j^1+dq_j^1;\ q_j^2,\ q_j^2+dq_j^2;\ \cdots;\ q_j^s,\ q_j^s+dq_j^s, \quad (j=1,\cdots,N) \tag{38}$$

is given by

$$\tau_j(q_j^1,\cdots,q_j^s)dq_j^1\cdots dq_j^s=\tau_j(\boldsymbol{q}_j)d\boldsymbol{q}_j. \tag{39}$$

Further, let

$$(\Phi^1,\Phi^2,\cdots,\Phi^n)=\boldsymbol{\Phi}=\sum_{j=1}^{N}\boldsymbol{\phi}_j. \tag{40}$$

The problem is: What is the probability that

$$\boldsymbol{\Phi}_0-\tfrac{1}{2}d\boldsymbol{\Phi}_0\leqslant\boldsymbol{\Phi}\leqslant\boldsymbol{\Phi}_0+\tfrac{1}{2}d\boldsymbol{\Phi}_0 \tag{41}$$

where $\boldsymbol{\Phi}_0$ is some preassigned value for $\boldsymbol{\Phi}$.

If we denote the required probability by

$$W_N(\boldsymbol{\Phi}_0)d\Phi_0^1\cdots d\Phi_0^n=W(\boldsymbol{\Phi}_0)d\boldsymbol{\Phi}_0, \tag{42}$$

we clearly have

$$W_N(\boldsymbol{\Phi}_0)d\boldsymbol{\Phi}_0=\int\cdots\int\prod_{j=1}^{N}\{\tau_j(\boldsymbol{q}_j)d\boldsymbol{q}_j\}, \tag{43}$$

where the integration is effected over only those parts of the Ns-dimensional configuration space $(q_1^1,\cdots,q_N^s)$ in which the inequalities (41) are satisfied.

We shall now introduce a factor $\Delta(\boldsymbol{q}_1,\cdots,\boldsymbol{q}_N)$ having the following properties:

$$\left.\begin{aligned}\Delta(\boldsymbol{q}_1,\cdots,\boldsymbol{q}_N)&=1 \quad \text{whenever} \quad \boldsymbol{\Phi}_0-\tfrac{1}{2}d\boldsymbol{\Phi}_0\leqslant\boldsymbol{\Phi}\leqslant\boldsymbol{\Phi}_0+\tfrac{1}{2}d\boldsymbol{\Phi}_0,\\ &=0 \quad \text{otherwise.}\end{aligned}\right\} \tag{44}$$

Then,

$$W_N(\boldsymbol{\Phi}_0)d\boldsymbol{\Phi}_0=\int\cdots\int\Delta(\boldsymbol{q}_1,\cdots,\boldsymbol{q}_N)\prod_{j=1}^{N}\{\tau_j(\boldsymbol{q}_j)d\boldsymbol{q}_j\} \tag{45}$$

where, in contrast to (43), the integration is now extended over *all* the accessible regions of the configuration space. The introduction of the factor Δ under the integral sign in Eq. (45) in this manner appears at first sight as a very formal device to extend the range of integration over the entire configuration space. But the essence of Markoff's method is that an explicit expression for this factor can be given.

Consider the integrals

$$\delta_k = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\sin \alpha_k \rho_k}{\rho_k} \exp (i\rho_k \gamma_k) d\rho_k \quad (k = 1, \cdots, n). \tag{46}$$

The integral defining δ_k is the well-known discontinuous integral of Dirichlet and has the property

$$\left.\begin{aligned} \delta_k &= 1 \quad \text{whenever} \quad -\alpha_k < \gamma_k < \alpha_k, \\ &= 0 \quad \text{otherwise.} \end{aligned}\right\} \tag{47}$$

Now, let

$$\alpha_k = \tfrac{1}{2} d\Phi_0^k; \quad \gamma_k = \sum_{j=1}^{N} \phi_j^k - \Phi_0^k \quad (k = 1, \cdots, n). \tag{48}$$

According to Eq. (47)

$$\left.\begin{aligned} \delta_k &= 1 \quad \text{whenever} \quad \Phi_0^k - \tfrac{1}{2} d\Phi_0^k < \sum_{j=1}^{N} \phi_j^k < \Phi_0^k + \tfrac{1}{2} d\Phi_0^k, \\ &= 0 \quad \text{otherwise.} \end{aligned}\right\} \tag{49}$$

Consequently

$$\Delta = \prod_{k=1}^{n} \delta_k \tag{50}$$

has the required properties (44).

Substituting for Δ from Eqs. (46) and (50) in Eq. (45), we obtain

$$\left.\begin{aligned} W_N(\Phi_0) d\Phi_0 &= \frac{1}{\pi^n} \int \underset{(\varrho)}{\cdots} \int \int \underset{(q)}{\cdots} \int \left\{ \prod_{j=1}^{N} \tau_j(q_j) dq_j \right\} \left\{ \prod_{k=1}^{n} \frac{\sin (\tfrac{1}{2} d\Phi_0^k \rho_k)}{\rho_k} \right\} \\ &\qquad \times \exp \left\{ i \left[\sum_{k=1}^{n} \sum_{j=1}^{N} \phi_j^k \rho_k - \sum_{k=1}^{n} \Phi_0^k \rho_k \right] \right\} d\rho_1 \cdots d\rho_n \\ &= \frac{d\Phi_0}{2^n \pi^n} \int \cdots \int \exp (-i\varrho \cdot \Phi_0) A_N(\varrho) d\varrho \end{aligned}\right\} \tag{51}$$

where we have written

$$A_N(\varrho) = \prod_{j=1}^{N} \int \cdots \int dq_j^1 \cdots dq_j^s \exp (i\varrho \cdot \phi_j) \tau_j (q_j^1, \cdots, q_j^s). \tag{52}$$

The case of greatest interest is when all the functions τ_j (of the respective q_j's) are identical. Equation (52) then becomes

$$A_N(\varrho) = \left[\int \exp (i\varrho \cdot \phi) \tau(q) dq \right]^N. \tag{53}$$

According to Eq. (51), $A_N(\varrho)$ is the n-dimensional Fourier-transform of the probability function $W(\Phi_0)$. And Markoff's procedure illustrates a very general principle that it is the Fourier transform of the probability function, rather than the function itself, that has a more direct relation to the physical situations.

For $N\to\infty$, $A_N(\rho)$ generally tends to the form [see §4 Eq. (91)]

$$\operatorname*{Limit}_{N\to\infty} A_N(\rho)=\exp[-C(\rho)]. \tag{54}$$

4. The Solution to the General Problem of Random Flights

We shall now apply Markoff's method to the problem of random flights. According to Eqs. (34), (51), and (52), the probability $W_N(R)dR$ that the position R of the particle will be found in the interval $(R, R+dR)$ after N displacements is given by

$$W_N(R)=\frac{1}{8\pi^3}\int_{-\infty}^{+\infty}\exp(-i\rho\cdot R)A_N(\rho)d\rho, \tag{55}$$

where

$$A_N(\rho)=\prod_{j=1}^{N}\int_{-\infty}^{+\infty}\tau_j(r_j)\exp(i\rho\cdot r_j)dr_j. \tag{56}$$

In Eq. (55), $\tau_j(r_j)$ governs the probability of occurrence of a displacement r_j on the jth occasion. The explicit form which $W_N(R)$ takes will naturally depend on the assumptions made concerning the $\tau_j(r_j)$'s. We shall now consider several cases of interest.

(a) A Gaussian Distribution of the Displacements r_j

A case of special interest arises when

$$\tau_j=\frac{1}{(2\pi l_j^2/3)^{\frac{3}{2}}}\exp(-3|r_j|^2/2l_j^2), \tag{57}$$

where l_j^2 denotes the mean square displacement to be expected on the jth occasion. While l_j^2 may differ from one displacement to another we assume that *all* the displacements occur in random directions.

For τ_j of the form (57), our expression for $A_N(\rho)$ becomes

$$\begin{aligned}A_N(\rho)&=\prod_{j=1}^{N}\frac{1}{(2\pi l_j^2/3)^{\frac{3}{2}}}\iiint_{-\infty}^{+\infty}\exp[i(\rho_1x_j+\rho_2y_j+\rho_3z_j)-3(x_j^2+y_j^2+z_j^2)/2l_j^2]dx_jdy_jdz_j\\&=\prod_{j=1}^{N}\exp[-(\rho_1^2+\rho_2^2+\rho_3^2)l_j^2/6]=\exp[-(|\rho|^2\sum_{j=1}^{N}l_j^2)/6].\end{aligned} \tag{58}$$

Let $\langle l^2\rangle_{Av}$ stand for

$$\langle l^2\rangle_{Av}=\frac{1}{N}\sum_{j=1}^{N}l_j^2. \tag{59}$$

Equation (58) becomes

$$A_N(\rho)=\exp[-N\langle l^2\rangle_{Av}|\rho|^2/6]. \tag{60}$$

Substituting this expression for $A_N(\rho)$ in Eq. (55), we obtain

$$W_N(R)=\frac{1}{8\pi^3}\iiint_{-\infty}^{+\infty}\exp[-i(\rho_1X+\rho_2Y+\rho_3Z)-N\langle l^2\rangle_{Av}(\rho_1^2+\rho_2^2+\rho_3^2)/6]d\rho_1d\rho_2d\rho_3. \tag{61}$$

The integrations in (61) are readily performed and we find

$$W_N(R)=\frac{1}{(2\pi N\langle l^2\rangle_{Av}/3)^{\frac{3}{2}}}\exp[-3|R|^2/2N\langle l^2\rangle_{Av}]. \tag{62}$$

This is an *exact* solution valid for any value of N. That an exact solution can be found for a Gaussian distribution of the different displacements is simply a consequence of the "*addition theorem*" which these functions satisfy.

(b) Each Displacement of a Constant Length But in Random Directions

Let the displacement on the jth occasion be of length l_j in a random direction. Under these circumstances, we can define the distribution functions τ_j by

$$\tau_j=\frac{1}{4\pi l_j^2}\delta(|r_j|^2-l_j^2),\quad (j=1,\cdots,N) \tag{63}$$

where δ stands for Dirac's δ function.

Accordingly, our expression for $A_N(\varrho)$ becomes

$$A_N(\varrho)=\prod_{j=1}^{N}\frac{1}{4\pi l_j^2}\int_{-\infty}^{+\infty}\exp(i\varrho\cdot r_j)\delta(r_j^2-l_j^2)dr_j, \tag{64}$$

or, using polar coordinates with the z axis in the direction of ϱ

$$A_N(\varrho)=\prod_{j=1}^{N}\frac{1}{4\pi l_j^2}\int_0^{\infty}\int_0^{\pi}\int_0^{2\pi}\exp\left[i|\varrho|r_j\cos\vartheta\right]\delta(r_j^2-l_j^2)r_j^2\sin\vartheta dr_jd\vartheta d\omega. \tag{65}$$

The integrations over the polar and the azimuthal angles ϑ and ω are readily effected:

$$\left.\begin{aligned}A_N(\varrho)&=\prod_{j=1}^{N}\frac{1}{2l_j^2}\int_0^{\infty}\int_0^{\pi}\exp(i|\varrho|r_j\cos\vartheta)r_j^2\delta(r_j^2-l_j^2)\sin\vartheta d\vartheta dr_j\\&=\prod_{j=1}^{N}\frac{1}{l_j^2|\varrho|}\int_0^{\infty}\sin(|\varrho|r_j)r_j\delta(r_j^2-l_j^2)dr_j\\&=\prod_{j=1}^{N}\frac{\sin(|\varrho|l_j)}{|\varrho|l_j}.\end{aligned}\right\} \tag{66}$$

Thus,

$$W_N(R)=\frac{1}{8\pi^3}\int_{-\infty}^{+\infty}\exp(-i\varrho\cdot R)\prod_{j=1}^{N}\frac{\sin(|\varrho|l_j)}{|\varrho|l_j}d\varrho. \tag{67}$$

Again, choosing polar coordinates but with the z axis pointing this time in the direction of $\boldsymbol{R}$, we have

$$W_N(R)=\frac{1}{8\pi^3}\int_0^{\infty}\int_{-1}^{+1}\int_0^{2\pi}\exp(-i|\varrho||R|t)\left\{\prod_{j=1}^{N}\frac{\sin(|\varrho|l_j)}{|\varrho|l_j}\right\}|\varrho|^2d\omega dt d|\varrho|. \tag{68}$$

The integrations over ω and t are readily performed and we obtain

$$W_N(R)=\frac{1}{2\pi^2|R|}\int_0^{\infty}\sin(|\varrho||R|)\left\{\prod_{j=1}^{N}\frac{\sin(|\varrho|l_j)}{|\varrho|l_j}\right\}|\varrho|d|\varrho| \tag{69}$$

which represents the formal solution to the problem. In this form, the solution for the problem of random flights is due to Rayleigh.[5]

[5] Lord Rayleigh, *Collected Papers*, Vol. 6, p. 604. We may, however, draw attention to the fact that our formulation of the general problem of random flights is wider in its scope than Rayleigh's. Rayleigh's formulation of the problem corresponds to our special case (63).

The case of greatest interest arises when all the l_j's are equal. We shall assume that this is the case in the rest of our discussion:

$$l_j = l = \text{constant} \quad (j=1, \cdots, N). \tag{70}$$

Equation (69) becomes

$$W_N(R) = \frac{1}{2\pi^2|R|}\int_0^\infty \sin(|\rho||R|)\left\{\frac{\sin(|\rho|l)}{|\rho|l}\right\}^N |\rho|d|\rho|. \tag{71}$$

(*i*) *N finite.*—We shall illustrate (following Rayleigh) the method of evaluating the integral on the right-hand side of Eq. (71) for finite values of N by considering the cases $N=3$ and 4.

When $N=3$, Eq. (71) becomes

$$W_3(R) = \frac{1}{2\pi^2|R|l^3}\int_0^\infty \sin(|\rho||R|)\sin^3(|\rho|l)\frac{d|\rho|}{|\rho|^2}. \tag{72}$$

But

$$\sin(|\rho||R|)\sin^3(|\rho|l) = \tfrac{1}{8}\{3\cos[(|R|-l)|\rho|] - 3\cos[(|R|+l)|\rho|] - \cos[(|R|-3l)|\rho|] + \cos[(|R|+3l)|\rho|]\}. \tag{73}$$

Further

$$\left.\begin{aligned}\int_0^\infty \{\cos[(|R|-l)|\rho|] - \cos[(|R|+l)|\rho|]\}\frac{d|\rho|}{|\rho|^2} &= 2\int_0^\infty\left\{\sin^2\frac{(|R|+l)|\rho|}{2} - \sin^2\frac{(|R|-l)|\rho|}{2}\right\}\frac{d|\rho|}{|\rho|^2}\\ &= \tfrac{1}{2}\pi(|R|+l-||R|-l|).\end{aligned}\right\} \tag{74}$$

We have a similar formula for the integral involving the other pair of cosines in Eq. (73). Combining these results we obtain

$$W_3(R) = \frac{1}{32\pi|R|l^3}\{2|R| - 3||R|-l| + ||R|-3l|\}, \tag{75}$$

or, more explicitly

$$\left.\begin{aligned}W_3(R) &= \frac{1}{8\pi l^3} && (0<|R|<l),\\ &= \frac{1}{16\pi l^3|R|}(3l-|R|) && (l<|R|<3l),\\ &= 0 && (3l<|R|<\infty).\end{aligned}\right\} \tag{76}$$

We shall consider next the case $N=4$. According to Eq. (71) we have

$$W_4(R) = \frac{1}{2\pi^2|R|l^4}\int_0^\infty \frac{d|\rho|}{|\rho|^3}\sin(|\rho||R|)\sin^4(|\rho|l). \tag{77}$$

From this equation we derive

$$
\left.\begin{aligned}
-\frac{d^2}{d|R|^2}[|R|W_4(R)]&=\frac{1}{2\pi^2l^4}\int_0^\infty \frac{d|\rho|}{|\rho|}\sin(|\rho||R|)\sin^4(|\rho|l)\\
&=\frac{1}{32\pi^2l^4}\int_0^\infty \frac{d|\rho|}{|\rho|}\{\sin[(|R|+4l)|\rho|]+\sin[(|R|-4l)|\rho|]\\
&\qquad-4\sin[(|R|+2l)|\rho|]-4\sin[(|R|-2l)|\rho|]+6\sin(|R||\rho|)\}\\
&=\frac{1}{64\pi l^4}(1\pm1-4\mp4+6)=\frac{1}{64\pi l^4}(3\pm1\mp4),
\end{aligned}\right\}\tag{78}
$$

where the two alternatives in the last two steps of Eq. (78) depend, respectively, on the signs of $(|R|-4l)$ and $(|R|-2l)$. Thus

$$
\left.\begin{aligned}
64\pi l^4\frac{d^2}{d|R|^2}[|R|W_4(R)]&=-6 \quad (0<|R|<2l),\\
&=+2 \quad (2l<|R|<4l),\\
&=0 \quad (4l<|R|<\infty).
\end{aligned}\right\}\tag{79}
$$

We can integrate the foregoing equation working backwards from large values of $|R|$ where all derivatives must vanish. We find

$$
\left.\begin{aligned}
64\pi l^4\frac{d}{d|R|}[|R|W_4(R)]&=2(|R|-4l) \quad (2l<|R|<4l),\\
&=-6|R|+8l \quad (0<|R|<2l),
\end{aligned}\right\}\tag{80}
$$

where we have used the continuity of the quantity on the left-hand side of this equation at $|R|=2l$. Integrating Eq. (80) once again we similarly obtain

$$
\left.\begin{aligned}
64\pi l^4|R|W_4(R)&=|R|^2-8l|R|+16l^2\\
&=(4l-|R|)^2
\end{aligned}\right\}(2l<|R|<4l), \tag{81}
$$

and

$$
64\pi l^4|R|W_4(R)=-3|R|^2+8l|R| \,(2l>|R|>0). \tag{82}
$$

Thus, finally

$$
\left.\begin{aligned}
W_4(R)&=\frac{1}{64\pi l^4|R|}(8l|R|-3|R|^2) \quad (0<|R|<2l),\\
&=\frac{1}{64\pi l^4|R|}(4l-|R|)^2 \quad (2l<|R|<4l),\\
&=0 \quad (4l<|R|<\infty).
\end{aligned}\right\}\tag{83}
$$

In like manner it is possible, in principle, to evaluate the integral for $W_N(R)$ for any finite value of N. But the calculations become very tedious. We may however note the following solution obtained by Rayleigh for the case $N=6$.

$$
\left.\begin{aligned}
W_4(R) &= \frac{1}{2^9\pi^2|R|l^6}(16l^3|R|-4l|R|^3+(5/6)|R|^4) && (0<R<2l)\\
&= \frac{1}{2^8\pi|R|l^6}(-20l^4+56l^3|R|-30l^2|R|^2+6l|R|^3-(5/12)|R|^4) && (2l<|R|<4l)\\
&= \frac{1}{2^8\pi|R|l^6}(108l^4-72l^3|R|+18l^2|R|^2-2l|R|^3+(1/12)|R|^4) && (4l<|R|<6l)\\
&= 0 && (6l<|R|<\infty).
\end{aligned}\right\} \quad (84)
$$

(ii) $N\gg 1$.—By far the most interesting case is when N is very large. Under these circumstances

$$
\left.\begin{aligned}
\operatorname*{Limit}_{N\to\infty}\left(\frac{\sin(|\varrho|l)}{|\varrho|l}\right)^N &= \operatorname*{Limit}_{N\to\infty}(1-\tfrac{1}{6}|\varrho|^2l^2+\cdots)^N,\\
&= \exp(-N|\varrho|^2l^2/6).
\end{aligned}\right\} \quad (85)
$$

Accordingly, from Eq. (69) we conclude that for large values of N

$$
W(R)=\frac{1}{2\pi^2|R|}\int_0^\infty \exp(-Nl^2|\varrho|^2/6)|\varrho|\sin(|R||\varrho|)d|\varrho|, \quad (86)
$$

where we have written $W(R)$ for $W_N(R)$, $N\to\infty$. Evaluating the integral on the right-hand side of Eq. (86), we find

$$
W(R)=\frac{1}{(2\pi Nl^2/3)^{\frac{3}{2}}}\exp(-3|R|^2/2Nl^2). \quad (87)
$$

We notice the formal similarity of Eqs. (62) and (87). However, on our present assumptions, Eq. (87) is valid only for large values of N.

(c) A Spherical Distribution of the Displacements. $N\gg 1$

We shall assume that

$$
\tau_j(r_j)=\tau(|r_j|^2) \quad (j=1,\cdots,N). \quad (88)
$$

Then

$$
A_N(\varrho)=\left[\int_{-\infty}^{+\infty}\exp(i\varrho\cdot r)\tau(r^2)dr\right]^N. \quad (89)
$$

By using polar coordinates, the integral inside the square brackets in Eq. (89) becomes

$$
\int_{-\infty}^{+\infty}\exp(i\varrho\cdot r)\tau(r^2)dr=\int_0^\infty\int_{-1}^{+1}\int_0^{2\pi}\exp(i|\varrho|rt)r^2\tau(r^2)d\omega dt dr=4\pi\int_0^\infty\frac{\sin(|\varrho|r)}{|\varrho|r}r^2\tau(r^2)dr. \quad (90)
$$

Hence

$$
\left.\begin{aligned}
\operatorname*{Limit}_{N\to\infty}A_N(\varrho) &= \operatorname*{Limit}_{N\to\infty}\left[4\pi\int_0^\infty\frac{\sin(|\varrho|r)}{|\varrho|r}r^2\tau(r^2)dr\right]^N,\\
&= \operatorname*{Limit}_{N\to\infty}\left[4\pi\int_0^\infty(1-\tfrac{1}{6}|\varrho|^2r^2+\cdots)r^2\tau(r^2)dr\right]^N,\\
&= \exp(-N|\varrho|^2\langle r^2\rangle_{Av}/6)
\end{aligned}\right\} \quad (91)
$$

where $\langle r^2\rangle_{hh}$ is the mean square displacement to be expected on any occasion. Substituting the foregoing result in Eq. (55) we obtain

$$W(R)=\frac{1}{8\pi^3}\int_{-\infty}^{+\infty}\exp\left(-i\boldsymbol{\varrho}\cdot\boldsymbol{R}-N|\boldsymbol{\varrho}|^2\langle r^2\rangle_{hh}/6\right)d\boldsymbol{\varrho}, \tag{92}$$

or, [cf. Eq. (62)]

$$W(R)=\frac{1}{(2\pi N\langle r^2\rangle_{hh}/3)^{\frac{3}{2}}}\exp\left(-3|\boldsymbol{R}|^2/2N\langle r^2\rangle_{hh}\right). \tag{93}$$

It is seen that Eq. (93) includes the result obtained earlier in Section (b) under case (ii) [Eq. (87)] as a special case.

(d) *The Solution to the General Problem of Random Flights for* $N\gg 1$

We shall now obtain the general expression for $W_N(\boldsymbol{R})$ for large values of N with no special assumptions concerning the distribution of the different displacements except that all the τ_j's represent the same function. Accordingly, we have to examine quite generally the behavior for $N\to\infty$ of $A_N(\boldsymbol{\varrho})$ defined by [cf. Eq. (53)]

$$A_N(\boldsymbol{\varrho})=\left[\int_{-\infty}^{+\infty}\exp\left(i\boldsymbol{\varrho}\cdot\boldsymbol{r}\right)\tau(\boldsymbol{r})d\boldsymbol{r}\right]^N. \tag{94}$$

Let ρ_1, ρ_2, ρ_3 denote the components of $\boldsymbol{\varrho}$ in some fixed system of coordinates. Then

$$\left.\begin{aligned}
A_N(\boldsymbol{\varrho})&=\left[\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\exp\left[i(\rho_1x+\rho_2y+\rho_3z)\right]\tau(x,y,z)dxdydz\right]^N,\\
&=\left[\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\{1+i(\rho_1x+\rho_2y+\rho_3z)-\tfrac{1}{2}(\rho_1^2x^2+\rho_2^2y^2+\rho_3^2z^2+2\rho_1\rho_2xy\right.\\
&\qquad\left.+2\rho_2\rho_3yz+2\rho_3\rho_1zx)+\cdots\}\tau(x,y,z)dxdydz\right]^N,\\
&=[1+i(\rho_1\langle x\rangle+\rho_2\langle y\rangle+\rho_3\langle z\rangle)-\tfrac{1}{2}(\rho_1^2\langle x^2\rangle+\rho_2^2\langle y^2\rangle+\rho_3^2\langle z^2\rangle+2\rho_1\rho_2\langle xy\rangle\\
&\qquad+2\rho_2\rho_3\langle yz\rangle+2\rho_3\rho_1\langle zx\rangle)+\cdots]^N
\end{aligned}\right\} \tag{95}$$

where $\langle x\rangle,\cdots,\langle zx\rangle$ denote the various first and second moments of the function $\tau(x, y, z)$. Hence for $N\to\infty$ we have

$$A_N(\boldsymbol{\varrho})=\exp\left[iN(\rho_1\langle x\rangle+\rho_2\langle y\rangle+\rho_3\langle z\rangle)-\tfrac{1}{2}NQ(\boldsymbol{\varrho})\right] \tag{96}$$

where $Q(\boldsymbol{\varrho})$ stands for the homogeneous quadratic form

$$Q(\boldsymbol{\varrho})=\langle x^2\rangle\rho_1^2+\langle y^2\rangle\rho_2^2+\langle z^2\rangle\rho_3^2+2\langle xy\rangle\rho_1\rho_2+2\langle yz\rangle\rho_2\rho_3+2\langle zx\rangle\rho_3\rho_1. \tag{97}$$

Substituting for $A_N(\boldsymbol{\varrho})$ from Eq. (96) in Eq. (55) we obtain for the probability distribution for large values of N the expression:

$$W(R)=\frac{1}{8\pi^3}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\exp\left[-\tfrac{1}{2}NQ(\boldsymbol{\varrho})-i\{\rho_1(X-N\langle x\rangle)+\rho_2(Y-N\langle y\rangle)+\rho_3(Z-N\langle z\rangle)\}\right]d\rho_1d\rho_2d\rho_3. \tag{98}$$

To evaluate this integral we first rotate our coordinate system to bring the quadratic form $Q(\boldsymbol{\varrho})$ to its diagonal form,

$$Q(\boldsymbol{\varrho})=\langle\xi^2\rangle\rho_\xi^2+\langle\eta^2\rangle\rho_\eta^2+\langle\zeta^2\rangle\rho_\zeta^2. \tag{99}$$

In Eq. (99) $\langle\xi^2\rangle$, $\langle\eta^2\rangle$ and $\langle\zeta^2\rangle$ are the eigenvalues of the symmetric matrix formed by the second moments:

$$\begin{vmatrix} \langle x^2\rangle & \langle xy\rangle & \langle xz\rangle \\ \langle yx\rangle & \langle y^2\rangle & \langle yz\rangle \\ \langle zx\rangle & \langle zy\rangle & \langle z^2\rangle \end{vmatrix} \tag{100}$$

Further, the three eigenvectors of the matrix (100) form an orthogonal system which we have denoted by (ξ, η, ζ). Let

$$R=(\Xi,\ \mathrm{H},\ \mathrm{Z}) \tag{101}$$

in this system of coordinates. Equation (98) now reduces to

$$W(R)=\frac{1}{8\pi^3}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\exp\left[-\tfrac{1}{2}N(\langle\xi^2\rangle\rho_\xi^2+\langle\eta^2\rangle\rho_\eta^2+\langle\zeta^2\rangle\rho_\zeta^2)\right.$$
$$\left.-i\{\rho_\xi(\Xi-N\langle\xi\rangle)+\rho_\eta(\mathrm{H}-N\langle\eta\rangle)+\rho_\zeta(\mathrm{Z}-N\langle\zeta\rangle)\}\right]d\rho_\xi d\rho_\eta d\rho_\zeta. \tag{102}$$

The integrations over ρ_ξ, ρ_η and ρ_ζ are now readily performed, and we find

$$W(R)=\frac{1}{(8\pi^3N^3\langle\xi^2\rangle\langle\eta^2\rangle\langle\zeta^2\rangle)^{\frac{1}{2}}}\exp\left[-\frac{(\Xi-N\langle\xi\rangle)^2}{2N\langle\xi^2\rangle}-\frac{(\mathrm{H}-N\langle\eta\rangle)^2}{2N\langle\eta^2\rangle}-\frac{(\mathrm{Z}-N\langle\zeta\rangle)^2}{2N\langle\zeta^2\rangle}\right]. \tag{103}$$

According to Eq. (103), the probability distribution $W(R)$ of the position R of the particle after suffering a large number of displacements (governed by a basic distribution function $\tau[x, y, z]$) is an *ellipsoidal distribution* centered at $(N\langle\xi\rangle, N\langle\eta\rangle, N\langle\zeta\rangle)$—in other words the particle suffers an average systematic net displacement of amount $(N\langle\xi\rangle, N\langle\eta\rangle, N\langle\zeta\rangle)$ and superposed on this a general random distribution.

The principal axes of this ellipsoidal distribution are along the principal directions of the moment-ellipsoid defined by (100) and the mean square net displacements about $(N\langle\xi\rangle, N\langle\eta\rangle, N\langle\zeta\rangle)$ along the three principal directions are

$$\langle(\Xi-N\langle\xi\rangle)^2\rangle_{Av}=N\langle\xi^2\rangle;\quad \langle(\mathrm{H}-N\langle\eta\rangle)^2\rangle_{Av}=N\langle\eta^2\rangle;\quad \langle(\mathrm{Z}-N\langle\zeta\rangle)^2\rangle_{Av}=N\langle\zeta^2\rangle. \tag{104}$$

5. The Passage to a Differential Equation: The Reduction of the Problem of Random Flights for Large N to a Boundary Value Problem

In the preceding sections we have obtained the solution to the problem of random flights under various conditions. Though in each case the problem was first formulated and solved for a finite number of displacements, the greatest interest is attached to the limiting form of the solutions for large values of N. And, for large values of N the solutions invariably take very simple forms. Thus, according to Eq. (93) a particle starting from the origin and suffering n displacements per unit time, each displacement r being governed by a probability distribution $\tau(|r|^2)$, will find itself in the element of volume defined by R and $R+dR$ after a time t with the probability

$$W(R)dR=\frac{1}{(2\pi n\langle r^2\rangle_{Av}t/3)^{\frac{3}{2}}}\exp\left(-3|R|^2/2n\langle r^2\rangle_{Av}t\right)dR. \tag{105}$$

In the foregoing equation $\langle r^2\rangle_{Av}$ denotes the mean square displacement that is to be expected on any given occasion. If we put

$$D=n\langle r^2\rangle_{Av}/6 \tag{106}$$

Eq. (105) takes the form [cf. Eq. (16)]

$$W(R)dR=\frac{1}{(4\pi Dt)^{\frac{3}{2}}}\exp\left(-|R|^2/4Dt\right)dR. \tag{107}$$

In view of the simplicity of this and the other solutions, the question now arises whether we cannot obtain the asymptotic distributions directly, without passing to the limit of large N, in each case, individually. This problem is of particular importance when restrictions on the motion of the particle in the form of reflecting and absorbing barriers are introduced. Our discussion in §2 of the simple problem of random walk in one dimension with such restrictions already indicates how very complicated the method of enumeration must become under even somewhat more general conditions than those contemplated in §2. The fact, however, that for the solutions obtained in §2, W vanishes on an absorbing wall [Eq. (25)] while grad W vanishes on a reflecting wall [Eq. (21)] suggests that the solutions perhaps correspond to solving a partial differential equation with appropriate boundary conditions. We shall now show how this passage to a differential equation and a boundary value problem is to be achieved.

First, we shall introduce a somewhat different language from that we have used so far in discussing the problem of random flights. Up to the present we have spoken of a *single* particle suffering displacements according to a given probability law, and asking for the probability of finding this particle in some given element of volume at a later time. It is clear that we can instead imagine a very large number of particles starting under the same initial conditions and undergoing the displacements without any mutual interference, and ask the *fraction* of the original number which will be found in a given element of volume at a later time. On this picture, the interpretation of the quantity on the right-hand side of Eq. (106) is that it represents the fraction of a large number of particles which will be found between $\boldsymbol{R}$ and $\boldsymbol{R}+d\boldsymbol{R}$ at time t if all the particles started from $\boldsymbol{R}=0$ at $t=0$. However, the two methods of interpretation are fully equivalent and we shall adopt the language of whichever of the two happens to be more convenient.

We pass on to considerations which lead to a differential equation for $W(\boldsymbol{R}, t)$:

Let Δt denote an interval of time long enough for a particle to suffer a large number of displacements but *still* short enough for the net mean square increment $\langle|\Delta\boldsymbol{R}|^2\rangle_{\text{Av}}$ in $\boldsymbol{R}$ to be small. Under these circumstances, the probability that a particle suffers a net displacement $\Delta\boldsymbol{R}$ in time Δt is given by

$$\psi(\Delta\boldsymbol{R};\Delta t)=\frac{1}{(4\pi D\Delta t)^{\frac{3}{2}}}\exp\left(-|\Delta\boldsymbol{R}|^2/4D\Delta t\right) \tag{108}$$

and is independent of $\boldsymbol{R}$. With Δt chosen in this manner, we seek to derive the probability distribution $W(\boldsymbol{R}, t+\Delta t)$ at time $t+\Delta t$ from the distribution $W(\boldsymbol{R}, t)$ at the earlier time t. In view of (108) and its independence of $\boldsymbol{R}$ we have the integral equation

$$W(\boldsymbol{R}, t+\Delta t)=\int_{-\infty}^{+\infty} W(\boldsymbol{R}-\Delta\boldsymbol{R}, t)\psi(\Delta\boldsymbol{R};\Delta t)d(\Delta\boldsymbol{R}). \tag{109}$$

Since $\langle|\Delta\boldsymbol{R}|^2\rangle_{\text{Av}}$ is assumed to be small we can expand $W(\boldsymbol{R}-\Delta\boldsymbol{R}, t)$ under the integral sign in (109) in a Taylor series and integrate term by term. We find

$$\left.\begin{aligned}
W(\boldsymbol{R}, t+\Delta t)=&\frac{1}{(4\pi D\Delta t)^{\frac{3}{2}}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\exp\left(-|\Delta\boldsymbol{R}|^2/4D\Delta t\right)\Bigg\{W(\boldsymbol{R}, t)-\Delta X\frac{\partial W}{\partial X}-\Delta Y\frac{\partial W}{\partial Y}\\
&-\Delta Z\frac{\partial W}{\partial Z}+\frac{1}{2}\left[(\Delta X)^2\frac{\partial^2 W}{\partial X^2}+(\Delta Y)^2\frac{\partial^2 W}{\partial Y^2}+(\Delta Z)^2\frac{\partial^2 W}{\partial Z^2}+2\Delta X\Delta Y\frac{\partial^2 W}{\partial X\partial Y}\right.\\
&\left.+2\Delta Y\Delta Z\frac{\partial^2 W}{\partial Y\partial Z}+2\Delta Z\Delta X\frac{\partial^2 W}{\partial Z\partial X}\right]+\cdots\Bigg\}d(\Delta X)d(\Delta Y)d(\Delta Z)\\
=&W(\boldsymbol{R}, t)+D\Delta t\left(\frac{\partial^2 W}{\partial X^2}+\frac{\partial^2 W}{\partial Y^2}+\frac{\partial^2 W}{\partial Z^2}\right)+O([\Delta t]^2).
\end{aligned}\right\} \tag{110}$$

Accordingly,

$$\frac{\partial W}{\partial t}\Delta t+O([\Delta t]^2)=D\left(\frac{\partial^2 W}{\partial X^2}+\frac{\partial^2 W}{\partial Y^2}+\frac{\partial^2 W}{\partial Z^2}\right)\Delta t+O([\Delta t]^2). \tag{111}$$

Passing now to the limit of $\Delta t=0$ we obtain

$$\frac{\partial W}{\partial t}=D\left(\frac{\partial^2 W}{\partial X^2}+\frac{\partial^2 W}{\partial Y^2}+\frac{\partial^2 W}{\partial Z^2}\right) \tag{112}$$

which is the required differential equation. And, it is seen that $W(R, t)$ defined according to Eq. (107) is indeed the fundamental solution of this differential equation.

Equation (112) is the standard form of the *equation of diffusion* or of heat conduction. This analogy that exists between our differential Eq. (112) to the equation of diffusion provides a new interpretation of the problem of random flights in terms of a *diffusion coefficient* D.

It is well known that in the *macroscopic* theory of diffusion if $W(R, t)$ denotes the concentration of the diffusing substance at R and at time t, then the amount crossing an area $\Delta\sigma$ in time Δt is given by

$$-D(\mathbf{1}_{\Delta\sigma}\cdot\text{grad } W)\Delta\sigma\Delta t, \tag{113}$$

where $\mathbf{1}_{\Delta\sigma}$ is a unit vector normal to the element of area $\Delta\sigma$. The diffusion equation is an elementary consequence of this fact. Consequently, we may describe the motion of a large number of particles describing random flights without mutual interference as a process of diffusion with the diffusion coefficient

$$D=n\langle r^2\rangle_{Av}/6. \tag{114}$$

With this visualization of the problem, the boundary conditions

$$W=0 \text{ on an element of surface which is a perfect absorber} \tag{115}$$

and

$$\text{grad } W=0 \text{ normal to an element surface which is a perfect reflector} \tag{116}$$

become intelligible. Further, according to Eq. (113), the rate at which particles appear on an absorbing screen per unit area, and per unit time, is given by

$$-D(\mathbf{1}\cdot\text{grad } W)_{W=0} \tag{117}$$

where $\mathbf{1}$ is a unit vector normal to the absorbing surface. This is in agreement with Eq. (33).

We shall now derive the differential equation for the problem of random flights in its general form considered in §4, subsection (d). This problem differs from the one we have just considered in that the probability distribution $\tau(r)$ governing the individual displacements $\mathbf{r}$ is now a function with no special symmetry properties. Accordingly, the first moments of τ cannot be assumed to vanish; further, the second moments define a general symmetric tensor of the second rank. Under these circumstances, the probability of finding the particle between R and $R+dR$ after it has suffered a large number of displacements is given by [cf. Eq. (103)]

$$W(R)dR=\frac{1}{(8\pi^3N^3\langle x^2\rangle\langle y^2\rangle\langle z^2\rangle)^{\frac{1}{2}}}\exp\left[-\frac{(X-N\langle x\rangle)^2}{2N\langle x^2\rangle}-\frac{(Y-N\langle y\rangle)^2}{2N\langle y^2\rangle}-\frac{(Z-N\langle z\rangle)^2}{2N\langle z^2\rangle}\right]dR. \tag{118}$$

In writing the probability distribution $W(R)$ in this form we have supposed that the coordinate system has been so chosen that the X, Y, and Z directions are along the principal axes of the moment ellipsoid.

Assuming that, on the average, the particle suffers n displacements per unit time we can rewrite our expression for $W(R)$ more conveniently in the form

$$W(\boldsymbol{R}) = \frac{1}{8(\pi t)^{3/2}(D_1 D_2 D_3)^{1/2}} \exp\left[-\frac{(X+\beta_1 t)^2}{4D_1 t} - \frac{(Y+\beta_2 t)^2}{4D_2 t} - \frac{(Z+\beta_3 t)^2}{4D_3 t} \right] \tag{119}$$

where we have written

$$\left.\begin{aligned} &D_1 = \tfrac{1}{2} n \langle x^2 \rangle; \quad D_2 = \tfrac{1}{2} n \langle y^2 \rangle; \quad D_3 = \tfrac{1}{2} n \langle z^2 \rangle, \\ &\beta_1 = -n \langle x \rangle; \quad \beta_2 = -n \langle y \rangle; \quad \beta_3 = -n \langle z \rangle. \end{aligned}\right\} \tag{120}$$

To make the passage to a differential equation, we consider, as before, an interval Δt which is long enough for the particle to suffer a large number of individual displacements but short enough for the mean square increment $\langle |\Delta \boldsymbol{R}|^2 \rangle_{Av}$ to be small. The probability that the particle suffers an increment $\Delta \boldsymbol{R}$ in the interval Δt is therefore governed by the distribution function

$$\psi(\Delta \boldsymbol{R}; \Delta t) = \frac{1}{8(\pi \Delta t)^{3/2}(D_1 D_2 D_3)^{1/2}} \exp\left[-\frac{(\Delta X+\beta_1 \Delta t)^2}{4D_1 \Delta t} - \frac{(\Delta Y+\beta_2 \Delta t)^2}{4D_2 \Delta t} - \frac{(\Delta Z+\beta_3 \Delta t)^2}{4D_3 \Delta t} \right]. \tag{121}$$

Hence, analogous to Eqs. (109) and (110) we now have

$$\left.\begin{aligned} W(\boldsymbol{R}, t+\Delta t) &= W(\boldsymbol{R}, t) + \frac{\partial W}{\partial t}\Delta t + O([\Delta t]^2) = \int_{-\infty}^{+\infty} W(\boldsymbol{R}-\Delta \boldsymbol{R}, t)\psi(\Delta \boldsymbol{R}; \Delta t) d(\Delta \boldsymbol{R}) \\ &= \frac{1}{8(\pi \Delta t)^{3/2}(D_1 D_2 D_3)^{1/2}} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp\left[-\frac{(\Delta X+\beta_1 \Delta t)^2}{4D_1 \Delta t} - \frac{(\Delta Y+\beta_2 \Delta t)^2}{4D_2 \Delta t} \right. \\ &\quad \left. - \frac{(\Delta Z+\beta_3 \Delta t)^2}{4D_3 \Delta t} \right] \left\{ W(\boldsymbol{R}, t) - \left(\Delta X \frac{\partial W}{\partial X} + \Delta Y \frac{\partial W}{\partial Y} + \Delta Z \frac{\partial W}{\partial Z} \right) \right. \\ &\quad + \frac{1}{2}\left(\Delta X^2 \frac{\partial^2 W}{\partial X^2} + \Delta Y^2 \frac{\partial^2 W}{\partial Y^2} + \Delta Z^2 \frac{\partial^2 W}{\partial Z^2} + 2\Delta X \Delta Y \frac{\partial^2 W}{\partial X \partial Y} + 2\Delta Y \Delta Z \frac{\partial^2 W}{\partial Y \partial Z} \right. \\ &\quad \left.\left. + 2\Delta Z \Delta X \frac{\partial^2 W}{\partial Z \partial X} \right) - \cdots \right\} d(\Delta X) d(\Delta Y) d(\Delta Z). \end{aligned}\right\} \tag{122}$$

Since for the distribution function (121)

$$\langle \Delta X \rangle_{Av} = -\beta_1 \Delta t; \quad \langle \Delta Y \rangle_{Av} = -\beta_2 \Delta t; \quad \langle \Delta Z \rangle_{Av} = -\beta_3 \Delta t, \tag{123}$$

and

$$\left.\begin{aligned} &\langle \Delta X^2 \rangle_{Av} = 2D_1 \Delta t + \beta_1^2 \Delta t^2; \quad \langle \Delta Y \Delta Z \rangle_{Av} = \beta_2 \beta_3 \Delta t^2, \\ &\langle \Delta Y^2 \rangle_{Av} = 2D_2 \Delta t + \beta_2^2 \Delta t^2; \quad \langle \Delta Z \Delta X \rangle_{Av} = \beta_3 \beta_1 \Delta t^2, \\ &\langle \Delta Z^2 \rangle_{Av} = 2D_3 \Delta t + \beta_3^2 \Delta t^2; \quad \langle \Delta X \Delta Y \rangle_{Av} = \beta_1 \beta_2 \Delta t^2, \end{aligned}\right\} \tag{124}$$

we conclude from Eq. (122) that

$$\frac{\partial W}{\partial t}\Delta t + O([\Delta t]^2) = \left(\beta_1 \frac{\partial W}{\partial X} + \beta_2 \frac{\partial W}{\partial Y} + \beta_3 \frac{\partial W}{\partial Z} \right)\Delta t + \left(D_1 \frac{\partial^2 W}{\partial X^2} + D_2 \frac{\partial^2 W}{\partial Y^2} + D_3 \frac{\partial^2 W}{\partial Z^2} \right)\Delta t + O([\Delta t]^2). \tag{125}$$

Passing now to the limit $\Delta t = 0$ we obtain

$$\frac{\partial W}{\partial t} = \beta_1 \frac{\partial W}{\partial X} + \beta_2 \frac{\partial W}{\partial Y} + \beta_3 \frac{\partial W}{\partial Z} + D_1 \frac{\partial^2 W}{\partial X^2} + D_2 \frac{\partial^2 W}{\partial Y^2} + D_3 \frac{\partial^2 W}{\partial Z^2}, \tag{126}$$

which is the required differential equation. According to this equation we can describe the phenomenon under discussion as a general process of diffusion in which the number of particles crossing elements of area normal to the X, Y, and Z direction per unit area and per unit time are given, respectively, by

$$-\beta_1 W - D_1\frac{\partial W}{\partial X};\quad -\beta_2 W - D_2\frac{\partial W}{\partial Y};\quad -\beta_3 W - D_3\frac{\partial W}{\partial Z}. \tag{127}$$

For the purposes of solving the differential Eq. (126) it is convenient to introduce a change in the independent variable. Let

$$W = U\exp\left[-\frac{\beta_1}{2D_1}(X-X_0)-\frac{\beta_2}{2D_2}(Y-Y_0)-\frac{\beta_3}{2D_3}(Z-Z_0)-\frac{\beta_1^2}{4D_1}t-\frac{\beta_2^2}{4D_2}t-\frac{\beta_3^2}{4D_3}t\right]. \tag{128}$$

We verify that Eq. (126) now reduces to

$$\frac{\partial U}{\partial t} = D_1\frac{\partial^2 U}{\partial X^2}+D_2\frac{\partial^2 U}{\partial Y^2}+D_3\frac{\partial^2 U}{\partial Z^2}. \tag{129}$$

The fundamental solution of this differential equation is

$$U = \frac{\text{Constant}}{(D_1D_2D_3t^3)^{\frac{1}{2}}}\exp\left[-\frac{(X-X_0)^2}{4D_1t}-\frac{(Y-Y_0)^2}{4D_2t}-\frac{(Z-Z_0)^2}{4D_3t}\right]. \tag{130}$$

Returning to the variable W, we have

$$W = \frac{\text{Constant}}{(D_1D_2D_3t^3)^{\frac{1}{2}}}\exp\left[-\frac{(X-X_0+\beta_1t)^2}{4D_1t}-\frac{(Y-Y_0+\beta_2t)^2}{4D_2t}-\frac{(Z-Z_0+\beta_3t)^2}{4D_3t}\right]. \tag{131}$$

In other words, the distribution (119) does indeed represent the fundamental solution of the differential Eq. (126).

CHAPTER II

THE THEORY OF THE BROWNIAN MOTION

1. Introductory Remarks. Langevin's Equation

In the studies on Brownian motion we are principally concerned with the perpetual irregular motions exhibited by small grains or particles of colloidal size immersed in a fluid. As is now well known, we witness in Brownian movement the phenomenon of molecular agitation on a reduced scale by particles very large on the molecular scale—so large in fact as to be readily visible in an ultramicroscope. The perpetual motions of the Brownian particles are maintained by fluctuations in the collisions with the molecules of the surrounding fluid. Under normal conditions, in a liquid, a Brownian particle will suffer about 10^{21} collisions per second and this is so frequent that we cannot really speak of separate collisions. Also, since each collision can be thought of as producing a kink in the path of the particle, it follows that we cannot hope to follow the path in any detail—indeed, to our senses the details of the path are impossibly fine.

The modern theory of the Brownian motion of a *free particle* (i.e., in the absence of an external field of force) generally starts with Langevin's equation

$$du/dt = -\beta u + A(t), \tag{132}$$

where u denotes the velocity of the particle. According to this equation, the influence of the surrounding medium on the motion of the particle can be split up into two parts: first, a systematic part $-\beta u$ representing a *dynamical friction* experienced by the particle and second, a fluctuating part $A(t)$ which is characteristic of the Brownian motion.

Regarding the frictional term $-\beta u$ it is assumed that this is governed by Stokes' law which states that the frictional force decelerating a spherical particle of radius a and mass m is given by $6\pi a\eta u/m$ where η denotes the coefficient of viscosity of the surrounding fluid. Hence

$$\beta = 6\pi a\eta/m. \tag{133}$$

As for the fluctuating part $A(t)$ the following principal assumptions are made:

(i) $A(t)$ is independent of u.
(ii) $A(t)$ varies extremely rapidly compared to the variations of u.

The second assumption implies that time intervals of duration Δt exist such that during Δt the variations in u that are to be expected are very small indeed while during the same interval $A(t)$ may undergo several fluctuations. Alternatively, we may say that though $u(t)$ and $u(t+\Delta t)$ are expected to differ by a negligible amount, no correlation between $A(t)$ and $A(t+\Delta t)$ exists. (The assumptions which are made here are quite analogous to those made in Chapter I, §5 in the passage to the differential equation for the problem of random flights; also see §§2 and 4 in this chapter.)

We shall show in the following sections how with the assumptions made in the foregoing paragraphs, we can derive from Langevin's equation all the physically significant relations concerning the motions of the Brownian particles. But we should draw attention even at this stage to the very drastic nature of assumptions implicit in the very writing of an equation of the form (132). For we have in reality supposed that we can divide the phenomenon into two parts, one in which the discontinuity of the events taking place is essential while in the other it is trivial and can be ignored. In view of the discontinuities in all matter and all events, this is a prima facie, an *ad-hoc* assumption. They are however made with reliance on physical intuition and the *aposteriori* justification by the success of the hypothesis. However, the correct procedure would be to treat the phenomenon in its entirety without appealing to the laws of continuous physics except insofar as they can be explicitly justified. As we shall see in Chapter IV a problem which occurs in stellar dynamics appears to provide a model in which the rigorous procedure can be explicitly followed.

2. The Theory of the Brownian Motion of a Free Particle

Our problem is to solve the stochastic differential equation (132) subject to the restrictions on $A(t)$ stated in the preceding section. But "solving" a stochastic differential equation like (132) is not the same thing as solving any ordinary differential equation. For one thing, Eq. (132) involves the function $A(t)$ which, as we shall presently see, has only statistically defined properties. Consequently, "solving" the Langevin Eq. (132) has to be understood rather in the sense of specifying a probability distribution $W(u, t; u_0)$ which governs the probability of occurrence of the velocity u at time t given that $u = u_0$ at $t = 0$. Of this function $W(u, t; u_0)$ we should clearly require that, as $t \to 0$,

$$W(u, t; u_0) \to \delta(u_x - u_{x,0})\delta(u_y - u_{y,0})\delta(u_z - u_{z,0}) \quad (t \to 0), \tag{134}$$

where the δ's are Dirac's δ functions. Further, the physical circumstances of the problem require that we demand of $W(u, t; u_0)$ that it tend to a Maxwellian distribution for the temperature T of the surrounding fluid, *independently* of u_0 as $t \to \infty$:

$$W(u, t; u_0) \to \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} \exp\left(-m|u|^2/2kT\right) \quad (t \to \infty). \tag{135}$$

This last demand on $W(u, t; u_0)$ conversely requires that $A(t)$ satisfy certain statistical requirements. For, according to the Langevin equation we have the formal solution

$$u - u_0 e^{-\beta t} = e^{-\beta t}\int_0^t e^{\beta \xi} A(\xi) d\xi. \tag{136}$$

Consequently, the statistical properties of

$$u-u_0e^{-\beta t} \tag{137}$$

must be the same as those of

$$e^{-\beta t}\int_0^t e^{\beta\xi}A(\xi)d\xi. \tag{138}$$

And, as $t\to\infty$ the quantity (137) tends to u; hence the distribution of

$$\underset{t\to\infty}{\text{Limit}}\left\{e^{-\beta t}\int_0^t e^{\beta\xi}A(\xi)d\xi\right\} \tag{139}$$

must be the Maxwellian distribution

$$(m/2\pi kT)^{\frac{3}{2}}\exp\left(-m|u|^2/2kT\right). \tag{140}$$

Now one of our principal assumptions concerning $A(t)$ is that it varies extremely rapidly compared to any of the other quantities which enter into our discussion. Further, the fluctuating acceleration experienced by the Brownian particles is statistical in character in the sense that Brownian particles having the same initial coordinates and/or velocities will suffer accelerations which will differ from particle to particle both in magnitude and in their dependence on time. However, on account of the rapidity of these fluctuations, we can always divide an interval of time which is long enough for any of the physical parameters like the position or the velocity of a Brownian particle to change appreciably, into a very large number of subintervals of duration Δt such that during each of these subintervals we can treat all functions of time except $A(t)$ which enter in our formulae as constants. Thus, the quantity on the right-hand side of Eq. (136) can be written as

$$e^{-\beta t}\sum_j e^{\beta j\Delta t}\int_{j\Delta t}^{(j+1)\Delta t}A(\xi)d\xi. \tag{141}$$

Let

$$B(\Delta t)=\int_t^{t+\Delta t}A(\xi)d\xi. \tag{142}$$

The physical meaning of $B(\Delta t)$ is that it represents the net acceleration which a Brownian particle may suffer on a given occasion during an interval of time Δt.

Equation (136) becomes

$$u-u_0e^{-\beta t}=\sum_j e^{\beta(j\Delta t-t)}B(\Delta t), \tag{143}$$

and we require that as $t\to\infty$ the quantity on the right-hand side tends to the Maxwellian distribution (140). We now assert that this requires *the probability of occurrence of different values for* $B(\Delta t)$ *be governed by the distribution function*

$$w(B[\Delta t])=\frac{1}{(4\pi q\Delta t)^{\frac{3}{2}}}\exp\left(-|B(\Delta t)|^2/4q\Delta t\right) \tag{144}$$

where

$$q=\beta kT/m. \tag{145}$$

To prove this assertion we have to show that the distribution function $W(u, t; u_0)$ derived on the basis of Eqs. (143) and (144) does in fact tend to the Maxwellian distribution (140) as $t\to\infty$. We shall presently show that this is the case but we may remark meantime on the formal similarity of Eq. (144) giving the probability distribution of the acceleration $B(\Delta t)$ suffered by a Brownian particle in time Δt and Eq. (108) giving the probability distribution of the increment ΔR in the position of a particle describing random flights in time Δt. It will be recalled that for the validity of Eq. (108) it is necessary that Δt be long enough for a large number of individual displacements to occur; analogously, our expression for $w(B[\Delta t])$ is valid only for times Δt large compared to the average period of a single fluctuation of $A(t)$. Now, the period of fluctuation of $A(t)$ is clearly of the order of the time between successive collisions between the Brownian particle and the molecules of the surrounding fluid; in a liquid this is generally of the order of 10^{-21} sec. Accordingly, the similarity of our expression for $w(B[\Delta t])$ with Eq. (108) in the theory of random flights, leads us to interpret the acceleration $B(\Delta t)$ suffered by a Brownian particle (in a time Δt large compared with the frequency of collisions with the surrounding molecules) as the result of superposition of the large number of random accelerations caused by collisions with the individual molecules. This is of course eminently reasonable; but the reason why q in Eq. (144) has to be precisely that given by Eq. (145) is due to our requirement that $W(u, t; u_0)$ tend to the Maxwellian distribution (140) as $t \to \infty$. We shall return to these questions again in §5.

We now proceed to prove our assertion concerning Eqs. (143), (144) and (145):

We first prove the following lemma:

Lemma I. Let

$$R = \int_0^t \psi(\xi) A(\xi) d\xi. \tag{146}$$

Then, the probability distribution of R is given by

$$W(R) = \frac{1}{\left[4\pi q \int_0^t \psi^2(\xi) d\xi\right]^{\frac{3}{2}}} \exp\left(-|R|^2 \Big/ 4q \int_0^t \psi^2(\xi) d\xi\right). \tag{147}$$

In order to prove this, we first divide the interval $(0, t)$ into a large number of subintervals of duration Δt. We can then write

$$R = \sum_j \psi(j\Delta t) \int_{j\Delta t}^{(j+1)\Delta t} A(\xi) d\xi. \tag{148}$$

Remembering our definition of $B(\Delta t)$ [Eq. (142)] we can express R in the form

$$R = \sum_j r_j, \tag{149}$$

where

$$r_j = \psi(j\Delta t) B(\Delta t) = \psi_j B(\Delta t). \tag{150}$$

According to Eq. (144) the probability distribution of r_j is governed by

$$\tau(r_j) = \frac{1}{(2\pi l_j^2/3)^{\frac{3}{2}}} \exp\left(-3|r_j|^2/2l_j^2\right), \tag{151}$$

where we have written

$$l_j^2 = 6q\psi_j^2 \Delta t. \tag{152}$$

A comparison of Eqs. (149) and (151) with Eqs. (34) and (57) shows that we have reduced our present problem to the special case in the theory of random flights considered in Chapter I, §4 case (a). Hence, [cf. Eqs. (59) and (62)]

$$W(R) = \frac{1}{(2\pi \sum l_j^2/3)^{\frac{3}{2}}} \exp\left(-3|R|^2/2\sum l_j^2\right). \tag{153}$$

But

$$\left.\begin{aligned} \sum l_j^2 &= 6q \sum_j \psi_j^2 \Delta t = 6q \sum_j \psi^2(j\Delta t)\Delta t, \\ &= 6q \int_0^t \psi^2(\xi) d\xi. \end{aligned}\right\} \tag{154}$$

We therefore have

$$W(R)=\frac{1}{\left[4\pi q\int_0^t \psi^2(\xi)d\xi\right]^{\frac{3}{2}}}\exp\left(-|R|^2\Big/4q\int_0^t \psi^2(\xi)d\xi\right), \tag{155}$$

which proves the lemma.

Returning to Eq. (136) we notice that we can express the right-hand side of this equation in the form

$$\int_0^t \psi(\xi)A(\xi)d\xi \tag{156}$$

if we define

$$\psi(\xi)=e^{\beta(\xi-t)}. \tag{157}$$

We can therefore apply lemma I and with the foregoing definition of $\psi(\xi)$, Eq. (155) governs the probability distribution of

$$u-u_0e^{-\beta t}. \tag{158}$$

Since, now,

$$\int_0^t \psi^2(\xi)d\xi=\int_0^t e^{2\beta(\xi-t)}d\xi=\frac{1}{2\beta}(1-e^{-2\beta t}), \tag{159}$$

and remembering that according to Eq. (145)

$$q/\beta=kT/m \tag{160}$$

we have proved that

$$W(u,t;u_0)=\left[\frac{m}{2\pi kT(1-e^{-2\beta t})}\right]^{\frac{3}{2}}\exp\left[-m|u-u_0e^{-\beta t}|^2/2kT(1-e^{-2\beta t})\right]. \tag{161}$$

We verify that according to this equation

$$W(u,t;u_0)\rightarrow\left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}}\exp\left(-m|u|^2/2kT\right)\quad(t\rightarrow\infty) \tag{162}$$

i.e., the Maxwellian distribution (140). This proves the assertion we made that with the statistical properties of $B(\Delta t)$ implied in Eqs. (144) and (145), Eq. (143) leads to a distribution $W(u,t;u_0)$ which tends to be Maxwellian independent of u_0 as $t\rightarrow\infty$.

We shall now show how with the assumptions already made concerning $B(\Delta t)$ we can further derive the distribution of the displacement r of a Brownian particle at time t given that the particle is at r_0 with a velocity u_0 at time $t=0$:

Since

$$r-r_0=\int_0^t u(t)dt, \tag{163}$$

we have according to Eq. (136)

$$r-r_0=\int_0^t d\eta\left\{u_0e^{-\beta\eta}+e^{-\beta\eta}\int_0^\eta d\xi e^{\beta\xi}A(\xi)\right\} \tag{164}$$

or

$$r-r_0-\beta^{-1}u_0(1-e^{-\beta t})=\int_0^t d\eta e^{-\beta\eta}\int_0^\eta d\xi e^{\beta\xi}A(\xi). \tag{165}$$

We can simplify the right-hand side of this equation by an integration by parts. We find

$$r-r_0-\beta^{-1}u_0(1-e^{-\beta t})=-\beta^{-1}e^{-\beta t}\int_0^t e^{\beta\xi}A(\xi)d\xi+\beta^{-1}\int_0^t A(\xi)d\xi. \tag{166}$$

Again, we can reduce this equation to the form

$$\boldsymbol{r}-\boldsymbol{r}_0-\beta^{-1}\boldsymbol{u}_0(1-e^{-\beta t})=\int_0^t \psi(\xi)\boldsymbol{A}(\xi)d\xi, \tag{167}$$

by defining

$$\psi(\xi)=\beta^{-1}(1-e^{\beta(\xi-t)}). \tag{168}$$

Thus lemma I can be applied and with the definition of $\psi(\xi)$ according to Eq. (168), Eq. (155) governs the probability distribution of

$$\boldsymbol{r}-\boldsymbol{r}_0-\beta^{-1}\boldsymbol{u}_0(1-e^{-\beta t}) \tag{169}$$

i.e., of $\boldsymbol{r}$ at time t for given $\boldsymbol{r}_0$ and $\boldsymbol{u}_0$. Since,

$$\left.\begin{aligned}\int_0^t \psi^2(\xi)d\xi&=\frac{1}{\beta^2}\int_0^t(1-e^{\beta(\xi-t)})^2d\xi,\\ &=\frac{1}{2\beta^3}(2\beta t-3+4e^{-\beta t}-e^{-2\beta t}),\end{aligned}\right\} \tag{170}$$

we have

$$W(\boldsymbol{r},t;\boldsymbol{r}_0,\boldsymbol{u}_0)=\left\{\frac{m\beta^2}{2\pi kT[2\beta t-3+4e^{-\beta t}-e^{-2\beta t}]}\right\}^{\frac{3}{2}}\exp-\left\{\frac{m\beta^2|\boldsymbol{r}-\boldsymbol{r}_0-\boldsymbol{u}_0(1-e^{-\beta t})/\beta|^2}{2kT[2\beta t-3+4e^{-\beta t}-e^{-2\beta t}]}\right\}. \tag{171}$$

For intervals of time long compared to β^{-1} the foregoing expression simplifies considerably. For, under these circumstances we can ignore the exponential and the constant terms as compared to $2\beta t$. Further, as we shall presently show, $\langle|\boldsymbol{r}-\boldsymbol{r}_0|^2\rangle_{Av}$ is of order t [cf. Eq. (174)]; hence we can also neglect $\boldsymbol{u}_0(1-e^{-\beta t})\beta^{-1}$ compared to $\boldsymbol{r}-\boldsymbol{r}_0$. Thus Eq. (171) reduces to

$$W(\boldsymbol{r},t;\boldsymbol{r}_0,\boldsymbol{u}_0)\simeq\frac{1}{(4\pi Dt)^{\frac{3}{2}}}\exp\left(-|\boldsymbol{r}-\boldsymbol{r}_0|^2/4Dt\right)\quad(t\gg\beta^{-1}) \tag{172}$$

where we have introduced the "diffusion coefficient" D defined by

$$D=kT/m\beta=kT/6\pi a\eta. \tag{173}$$

In Eq. (173) we have substituted for β according to Eq. (133).

From Eq. (172) we obtain for the mean square displacement along any given direction (say, the x direction) the formula

$$\langle(x-x_0)^2\rangle_{Av}=\tfrac{1}{3}\langle|\boldsymbol{r}-\boldsymbol{r}_0|^2\rangle_{Av}=2Dt=(kT/3\pi a\eta)t. \tag{174}$$

This is Einstein's result. Equation (174) has been verified by Perrin to lead to consistent and satisfactory values for the Boltzmann constant k by observation of $\langle(x-x_0^2)\rangle_{Av}/t$ over wide ranges of T, η and a.

The law of distribution of displacements (172) has been exhaustively tested by observation. Perrin gives the following sets of counts of the displacements of a grain of radius 2.1×10^{-5} cm at 30 sec. intervals. Out of a number N of such observations the number of observed values of x displacements between x_1 and x_2 should be

$$\frac{N}{\pi^{\frac{1}{2}}}\int_{x_1}^{x_2}\exp\left(-x^2/4Dt\right)\frac{dx}{(4Dt)^{\frac{1}{2}}}.$$

The agreement is satisfactory. See Table II.

Comparing Eq. (172) with the solution for the problem of random flights obtained in Eq. (107) we conclude that for times $t\gg\beta^{-1}$ we can regard the motion of a Brownian particle as one of random

TABLE II. Observations and calculations of the distribution of the displacements of a Brownian particle.

Range $x\times10^4$ cm	1st set Obs.	1st set Calc.	2nd set Obs.	2nd set Calc.	Total Obs.	Total Calc.
0 - 3.4	82	91	86	84	168	175
3.4- 6.8	66	70	65	63	131	132
6.8-10.2	46	39	31	36	77	75
10.2-17.0	27	23	23	21	50	44

flights. And therefore, according to the ideas of *I* §5, describe the motion of Brownian particles also as one of diffusion and governed by the diffusion equation. We shall return to this connection with the diffusion equation from a more general point of view in §4.

Returning to Eq. (171) we see that, quite generally, we have

$$\langle|\mathbf{r}-\mathbf{r}_0|^2\rangle_{Av}=\frac{|\mathbf{u}_0|^2}{\beta^2}(1-e^{-\beta t})^2+3\frac{kT}{m\beta^2}(2\beta t-3+4e^{-\beta t}-e^{-2\beta t}). \tag{175}$$

Averaging this equation over all values of $\mathbf{u}_0$ and remembering that $\langle|\mathbf{u}_0|^2\rangle_{Av}=3kT/m$ we obtain

$$\langle\langle|\mathbf{r}-\mathbf{r}_0|^2\rangle\rangle_{Av}=6\frac{kT}{m\beta^2}(\beta t-1+e^{-\beta t}). \tag{175'}$$

For $t\to\infty$, Eq. (175′) is in agreement with our result (174), while for $t\to 0$ we have instead

$$\langle\langle|\mathbf{r}-\mathbf{r}_0|^2\rangle\rangle_{Av}=3\frac{kT}{m}t^2=\langle|\mathbf{u}_0|^2\rangle_{Av}t^2. \tag{175''}$$

So far we have only inquired into the law of distributions of $\mathbf{u}$ and $\mathbf{r}$ separately. But we can also ask for the distribution $W(\mathbf{r},\mathbf{u},t;\mathbf{u}_0,\mathbf{r}_0)$ governing the probability of the simultaneous occurrence of the velocity $\mathbf{u}$ and the position $\mathbf{r}$ at time t, given that $\mathbf{u}=\mathbf{u}_0$ and $\mathbf{r}=\mathbf{r}_0$ at $t=0$. The solution to this problem can be obtained from the following lemma:

Lemma II. *Let*

$$\mathbf{R}=\int_0^t\psi(\xi)\mathbf{A}(\xi)d\xi, \tag{176}$$

and

$$\mathbf{S}=\int_0^t\phi(\xi)\mathbf{A}(\xi)d\xi. \tag{177}$$

Then, the bivariate probability distribution of $\mathbf{R}$ *and* $\mathbf{S}$ *is given by*

$$W(\mathbf{R},\mathbf{S})=\frac{1}{8\pi^3(FG-H^2)^{\frac{3}{2}}}\exp\left[-(G|\mathbf{R}|^2-2H\mathbf{R}\cdot\mathbf{S}+F|\mathbf{S}|^2)/2(FG-H^2)\right] \tag{178}$$

where

$$F=2q\int_0^t\psi^2(\xi)d\xi;\quad G=2q\int_0^t\phi^2(\xi)d\xi;\quad H=2q\int_0^t\phi(\xi)\psi(\xi)d\xi. \tag{179}$$

The lemma is proved by writing $\mathbf{R}$ and $\mathbf{S}$ in the forms [cf. Eqs. (149) and (150)]

$$\mathbf{R}=\textstyle\sum_j\psi(j\Delta t)\mathbf{B}(\Delta t);\quad \mathbf{S}=\sum_j\phi(j\Delta t)\mathbf{B}(\Delta t) \tag{180}$$

and remembering that the distribution of $\mathbf{B}$ is Gaussian according to Eq. (144). The problem then reduces to the one considered in Appendix II and the solution stated readily follows.

To obtain the distribution $W(\mathbf{r}, \mathbf{u}, t; \mathbf{u}_0, \mathbf{r}_0)$ we have only to set [cf. Eqs. (157), (158), (167) and (168)]

$$\left.\begin{aligned} \mathbf{R} &= \mathbf{r} - \mathbf{r}_0 - \beta^{-1}\mathbf{u}_0(1-e^{-\beta t}); \quad \psi(\xi) = \beta^{-1}(1-e^{\beta(\xi-t)}), \\ \mathbf{S} &= \mathbf{u} - \mathbf{u}_0 e^{-\beta t}; \qquad\qquad\qquad \phi(\xi) = e^{\beta(\xi-t)}, \end{aligned}\right\} \tag{181}$$

and [cf. Eqs. (159) and (170)]

$$F = q\beta^{-3}(2\beta t - 3 + 4e^{-\beta t} - e^{-2\beta t}); \quad G = q\beta^{-1}(1-e^{-2\beta t}), \tag{182}$$

and finally

$$H = 2q\beta^{-1}\int_0^t e^{\beta(\xi-t)}(1-e^{\beta(\xi-t)})d\xi = q\beta^{-2}(1-e^{-\beta t})^2. \tag{183}$$

3. The Theory of the Brownian Motion of a Particle in a Field of Force. The Harmonically Bound Particle

In the presence of an external field of force, the Langevin Eq. (132) is generalized to

$$d\mathbf{u}/dt = -\beta\mathbf{u} + \mathbf{A}(t) + \mathbf{K}(\mathbf{r}, t) \tag{184}$$

where $\mathbf{K}(\mathbf{r}, t)$ is the acceleration produced by the field. In writing this equation we are making the same general assumptions as are involved in writing the original Langevin equation (cf. the remarks at the end of §1).

In solving the stochastic equation (184) we attribute to $\mathbf{A}(t)$ or more particularly for

$$\mathbf{B}(\Delta t) = \int_t^{t+\Delta t} \mathbf{A}(\xi)d\xi \tag{185}$$

the statistical properties already assigned in the preceding section [Eq. (144)]. The method of solution is illustrated sufficiently by a one-dimensional harmonic oscillator describing Brownian motion. The appropriate stochastic equation is

$$du/dt = -\beta u + A(t) - \omega^2 x, \tag{186}$$

where ω denotes the circular frequency of the oscillator. We can write Eq. (184) alternatively in the form

$$d^2x/dt^2 + \beta dx/dt + \omega^2 x = A(t). \tag{187}$$

What we seek from this equation are, of course, the probability distributions $W(x, t; x_0, u_0)$, $W(u, t; x_0, u_0)$ and $W(x, u, t; x_0, u_0)$. To obtain these distributions we first write down the formal solution of Eq. (187) regarded as an ordinary differential equation. The method of solution most appropriate for our present purposes is that of the variation of the parameters. In this method, as applied to Eq. (187), we express the solution in terms of that of the homogeneous equation:

$$x = a_1 \exp(\mu_1 t) + a_2 \exp(\mu_2 t) \tag{188}$$

where μ_1 and μ_2 are the roots of

$$\mu^2 + \beta\mu + \omega^2 = 0; \tag{189}$$

i.e.,

$$\mu_1 = -\tfrac{1}{2}\beta + (\tfrac{1}{4}\beta^2 - \omega^2)^{\frac{1}{2}}; \quad \mu_2 = -\tfrac{1}{2}\beta - (\tfrac{1}{4}\beta^2 - \omega^2)^{\frac{1}{2}}. \tag{190}$$

We assume that the solution of Eq. (187) is of the form (188) where a_1 and a_2 are functions of time restricted however to satisfy the equation

$$\exp(\mu_1 t)(da_1/dt) + \exp(\mu_2 t)(da_2/dt) = 0. \tag{191}$$

From Eq. (187) we derive the further relation

$$\mu_1 \exp(\mu_1 t)(da_1/dt)+\mu_2 \exp(\mu_2 t)(da_2/dt)=A(t). \tag{192}$$

Solving Eqs. (191) and (192) we readily obtain the integrals

$$\left.\begin{aligned} a_1&=+\frac{1}{\mu_1-\mu_2}\int_0^t \exp(-\mu_1\xi)A(\xi)d\xi+a_{10},\\ a_2&=-\frac{1}{\mu_1-\mu_2}\int_0^t \exp(-\mu_2\xi)A(\xi)d\xi+a_{20},\end{aligned}\right\} \tag{193}$$

where a_{10} and a_{20} are constants. Accordingly, we have the solution

$$x=\frac{1}{\mu_1-\mu_2}\left\{\exp(\mu_1 t)\int_0^t \exp(-\mu_1\xi)A(\xi)d\xi-\exp(\mu_2 t)\int_0^t \exp(-\mu_2\xi)A(\xi)d\xi\right\}$$
$$+a_{10}\exp(\mu_1 t)+a_{20}\exp(\mu_2 t). \tag{194}$$

From the foregoing equation we obtain for the velocity u the formula

$$u=\frac{1}{\mu_1-\mu_2}\left\{\mu_1\exp(\mu_1 t)\int_0^t \exp(-\mu_1\xi)A(\xi)d\xi-\mu_2\exp(\mu_2 t)\int_0^t \exp(-\mu_2\xi)A(\xi)d\xi\right\}$$
$$+\mu_1 a_{10}\exp(\mu_1 t)+\mu_2 a_{20}\exp(\mu_2 t). \tag{195}$$

The constants a_{10} and a_{20} can now be determined from the conditions that $x=x_0$ and $u=u_0$ at $t=0$. We find

$$a_{10}=-\frac{x_0\mu_2-u_0}{\mu_1-\mu_2};\quad a_{20}=+\frac{x_0\mu_1-u_0}{\mu_1-\mu_2}. \tag{196}$$

Thus, we have the solutions

$$x+\frac{1}{\mu_1-\mu_2}[(x_0\mu_2-u_0)\exp(\mu_1 t)-(x_0\mu_1-u_0)\exp(\mu_2 t)]=\int_0^t A(\xi)\psi(\xi)d\xi, \tag{197}$$

and

$$u+\frac{1}{\mu_1-\mu_2}[\mu_1(x_0\mu_2-u_0)\exp(\mu_1 t)-\mu_2(x_0\mu_1-u_0)\exp(\mu_2 t)]=\int_0^t A(\xi)\phi(\xi)d\xi, \tag{198}$$

where we have written

$$\left.\begin{aligned}\psi(\xi)&=\frac{1}{\mu_1-\mu_2}[\exp[\mu_1(t-\xi)]-\exp[\mu_2(t-\xi)]],\\ \phi(\xi)&=\frac{1}{\mu_1-\mu_2}[\mu_1\exp[\mu_1(t-\xi)]-\mu_2\exp[\mu_2(t-\xi)]].\end{aligned}\right\} \tag{199}$$

It is now seen that the quantities on the right-hand sides of Eqs. (197) and (198) are of the forms considered in lemmas I and II in §2. Accordingly, we can at once write down the distribution functions $W(x, t; x_0, u_0)$, $W(u, t; x_0, u_0)$ and $W(x, u, t; x_0, u_0)$ in terms of the integrals

$$\int_0^t \psi^2(\xi)d\xi;\quad \int_0^t \phi^2(\xi)d\xi\quad\text{and}\quad\int_0^t \psi(\xi)\phi(\xi)d\xi. \tag{200}$$

With $\psi(\xi)$ and $\phi(\xi)$ defined as in Eqs. (199) we readily verify that

$$\int_0^t \psi^2(\xi)d\xi = \frac{1}{(\mu_1-\mu_2)^2}\left[\frac{1}{2\mu_1\mu_2}(\mu_2 \exp(2\mu_1 t)+\mu_1 \exp(2\mu_2 t)) - \frac{2}{\mu_1+\mu_2}(\exp[(\mu_1+\mu_2)t]-1) - \frac{\mu_1+\mu_2}{2\mu_1\mu_2}\right], \tag{201}$$

$$\int_0^t \phi^2(\xi)d\xi = \frac{1}{(\mu_1-\mu_2)^2}\left[\tfrac{1}{2}(\mu_1 \exp(2\mu_1 t)+\mu_2 \exp(2\mu_2 t)) - \frac{2\mu_1\mu_2}{\mu_1+\mu_2}(\exp[(\mu_1+\mu_2)t]-1) - \tfrac{1}{2}(\mu_1+\mu_2)\right], \tag{202}$$

and

$$\int_0^t \psi(\xi)\phi(\xi)d\xi = \frac{1}{2(\mu_1-\mu_2)^2}(\exp(\mu_1 t)-\exp(\mu_2 t))^2. \tag{203}$$

At this point it is convenient to introduce in the foregoing expressions the values of μ_1 and μ_2 explicitly according to Eq. (190): We find that the quantities on the left-hand sides of Eqs. (197) and (198) become, respectively,

$$x - x_0 e^{-\beta t/2}\cosh \tfrac{1}{2}\beta_1 t - \frac{x_0\beta+2u_0}{\beta_1}e^{-\beta t/2}\sinh \tfrac{1}{2}\beta_1 t, \tag{204}$$

and

$$u - u_0 e^{-\beta t/2}\cosh \tfrac{1}{2}\beta_1 t + \frac{2x_0\omega^2+\beta u_0}{\beta_1}e^{-\beta t/2}\sinh \tfrac{1}{2}\beta_1 t, \tag{205}$$

where we have introduced the quantity β_1 defined by

$$\beta_1 = (\beta^2 - 4\omega^2)^{\frac{1}{2}}. \tag{206}$$

Similarly, we find

$$\int_0^t \psi^2(\xi)d\xi = \frac{1}{2\omega^2\beta} - \frac{e^{-\beta t}}{2\omega^2\beta_1^2\beta}(2\beta^2\sinh^2 \tfrac{1}{2}\beta_1 t + \beta\beta_1 \sinh \beta_1 t + \beta_1^2), \tag{207}$$

$$\int_0^t \phi^2(\xi)d\xi = \frac{1}{2\beta} - \frac{e^{-\beta t}}{2\beta_1^2\beta}(2\beta^2\sinh^2 \tfrac{1}{2}\beta_1 t - \beta\beta_1 \sinh \beta_1 t + \beta_1^2), \tag{208}$$

and

$$\int_0^t \psi(\xi)\phi(\xi)d\xi = 2\beta_1^{-2}e^{-\beta t}\sinh^2 \tfrac{1}{2}\beta_1 t. \tag{209}$$

It is seen that all the foregoing expressions remain finite and real even when β_1 is zero or imaginary. Thus, while all the expressions remain valid as they stand in the "overdamped" case (β_1 real) the formulae appropriate for the periodic (β_1 imaginary) and the aperiodic (β_1 zero) cases can be readily written down by replacing

$$\cosh \tfrac{1}{2}\beta_1 t,\ \beta_1^{-1}\sinh \tfrac{1}{2}\beta_1 t \quad \text{and} \quad \beta_1^{-1}\sinh \beta_1 t, \tag{210}$$

respectively, by

$$\cos \omega_1 t,\quad \frac{1}{2\omega_1}\sin \omega_1 t \quad \text{and} \quad \frac{1}{2\omega_1}\sin 2\omega_1 t \text{ where } \omega_1 = (\omega^2 - \tfrac{1}{4}\beta^2)^{\frac{1}{2}} \tag{211}$$

in the periodic case, and by

$$1,\ \tfrac{1}{2}t \quad \text{and} \quad t \tag{212}$$

in the aperiodic case.

As we have already remarked, we can immediately write down the distribution functions for the quantities on the left-hand sides of the Eqs. (197) and (198) [i.e., the quantities (204) and (205)] according to lemmas I and II of §2 in terms of the integrals (207)–(209). Thus,

$$W(x, t; x_0, u_0) = \left[\frac{m}{4\pi\beta kT\int_0^t \psi^2(\xi)d\xi}\right]^{\frac{1}{2}} \exp - \frac{\left(x - x_0 e^{-\beta t/2}\left[\cosh \tfrac{1}{2}\beta_1 t + \dfrac{\beta}{\beta_1}\sinh \tfrac{1}{2}\beta_1 t\right] - \dfrac{2u_0}{\beta_1}e^{-\beta t/2}\sinh \tfrac{1}{2}\beta_1 t\right)^2}{\dfrac{2kT}{m\omega^2}\left\{1 - e^{-\beta t}\left(\dfrac{2\beta^2}{\beta_1^2}\sinh^2 \tfrac{1}{2}\beta_1 t + \dfrac{\beta}{\beta_1}\sinh \beta_1 t + 1\right)\right\}}. \tag{213}$$

We have similar expressions for $W(u, t; x_0, u_0)$ and $W(x, u, t; x_0, u_0)$.

The quantities of greatest interest are the moments $\langle x\rangle_{\text{Av}}$, $\langle u\rangle_{\text{Av}}$, $\langle x^2\rangle_{\text{Av}}$, $\langle u^2\rangle_{\text{Av}}$ and $\langle xu\rangle_{\text{Av}}$. We find

$$\left.\begin{aligned}
\langle x\rangle_{\text{Av}} &= x_0 e^{-\beta t/2}\left(\cosh \tfrac{1}{2}\beta_1 t + \frac{\beta}{\beta_1}\sinh \tfrac{1}{2}\beta_1 t\right) + \frac{2u_0}{\beta_1}e^{-\beta t/2}\sinh \tfrac{1}{2}\beta_1 t,\\
\langle u\rangle_{\text{Av}} &= u_0 e^{-\beta t/2}\left(\cosh \tfrac{1}{2}\beta_1 t - \frac{\beta}{\beta_1}\sinh \tfrac{1}{2}\beta_1 t\right) - \frac{2x_0\omega^2}{\beta_1}e^{-\beta t/2}\sinh \tfrac{1}{2}\beta_1 t,\\
\langle x^2\rangle_{\text{Av}} &= \langle x\rangle_{\text{Av}}^2 + \frac{kT}{m\omega^2}\left\{1 - e^{-\beta t}\left(2\frac{\beta^2}{\beta_1^2}\sinh^2 \tfrac{1}{2}\beta_1 t + \frac{\beta}{\beta_1}\sinh \beta_1 t + 1\right)\right\},\\
\langle u^2\rangle_{\text{Av}} &= \langle u\rangle_{\text{Av}}^2 + \frac{kT}{m}\left\{1 - e^{-\beta t}\left(2\frac{\beta^2}{\beta_1^2}\sinh^2 \tfrac{1}{2}\beta_1 t - \frac{\beta}{\beta_1}\sinh \beta_1 t + 1\right)\right\},\\
\langle xu\rangle_{\text{Av}} &= \langle x\rangle_{\text{Av}}\langle u\rangle_{\text{Av}} + \frac{4\beta kT}{\beta_1^2 m}e^{-\beta t}\sinh^2 \tfrac{1}{2}\beta_1 t.
\end{aligned}\right\} \tag{214}$$

The foregoing expressions are the average values of the various quantities at time t for assigned values of x and u (namely, x_0 and u_0) at time $t=0$. We see that

$$\left.\begin{aligned}
&\langle x\rangle_{\text{Av}}\to 0;\quad \langle u\rangle_{\text{Av}}\to 0;\quad \langle xu\rangle_{\text{Av}}\to 0,\\
&\langle x^2\rangle_{\text{Av}}\to kT/m\omega^2;\quad \langle u^2\rangle_{\text{Av}} = kT/m,
\end{aligned}\right\}\quad t\to\infty. \tag{215}$$

By averaging the various moments over all values of u_0 and remembering that

$$\langle u_0\rangle_{\text{Av}} = 0;\quad \langle u_0^2\rangle_{\text{Av}} = kT/m. \tag{216}$$

we obtain from Eqs. (214) that

$$\left.\begin{aligned}
\langle\langle x\rangle\rangle_{\text{Av}} &= x_0 e^{-\beta t/2}\left(\cosh \tfrac{1}{2}\beta_1 t + \frac{\beta}{\beta_1}\sinh \tfrac{1}{2}\beta_1 t\right),\\
\langle\langle u\rangle\rangle_{\text{Av}} &= -\frac{2x_0\omega^2}{\beta_1}e^{-\beta t/2}\sinh \tfrac{1}{2}\beta_1 t,\\
\langle\langle x^2\rangle\rangle_{\text{Av}} &= \frac{kT}{m\omega^2} + \left(x_0^2 - \frac{kT}{m\omega^2}\right)e^{-\beta t}\left(\cosh \tfrac{1}{2}\beta_1 t + \frac{\beta}{\beta_1}\sinh \tfrac{1}{2}\beta_1 t\right)^2,\\
\langle\langle u^2\rangle\rangle_{\text{Av}} &= \frac{kT}{m} + \frac{4\omega}{\beta_1^2}\left(x_0^2 - \frac{kT}{m\omega^2}\right)e^{-\beta t}\sinh^2 \tfrac{1}{2}\beta_1 t,\\
\langle\langle xu\rangle\rangle_{\text{Av}} &= \frac{2\omega^2}{\beta_1}\left(\frac{kT}{m\omega^2} - x_0^2\right)e^{-\beta t}\sinh \tfrac{1}{2}\beta_1 t\left(\cosh \tfrac{1}{2}\beta_1 t + \frac{\beta}{\beta_1}\sinh \tfrac{1}{2}\beta_1 t\right).
\end{aligned}\right\} \tag{217}$$

Equations (214) and (217) show how the equipartition values (215) are reached as $t\to\infty$.

4. The Fokker-Planck Equation. The Generalization of Liouville's Theorem

As we have already remarked on several occasions, in an analysis of the Brownian movement we regard as impracticable a detailed description of the motions of the individual particles. Instead, we emphasize the essential stochastic nature of the phenomenon and seek a description in terms of the probability distributions of position and/or velocity at a later time starting from given initial distributions. Thus, in our discussion of the Brownian movement of a free particle in §2 we obtain explicitly the distribution functions $W(\boldsymbol{u}, t; \boldsymbol{u}_0)$, $W(\boldsymbol{r}, t; \boldsymbol{u}_0, \boldsymbol{r}_0)$ and $W(\boldsymbol{r}, \boldsymbol{u}, t; \boldsymbol{r}_0, \boldsymbol{u}_0)$ for given initial values of $\boldsymbol{r}_0$ and $\boldsymbol{u}_0$; similarly, in §3 we determined the distributions $W(u, t; x_0, u_0)$, $W(x, t; x_0, u_0)$ and $W(x, u, t; x_0, u_0)$ for a harmonically bound particle describing Brownian motion. In deriving these distributions in §§2 and 3 we started with the Langevin equation [Eq. (132) in the field-free case, and Eq. (184) when an external field is present] and solved it in a manner appropriate to the problem. We shall now consider the question whether we cannot reduce the determination of these distribution functions to appropriate boundary value problems of suitably chosen partial differential equations. We have in mind a reduction similar to that achieved in Chapter I, §5 where we showed how, under certain circumstances, the solution to the problem of random flights can be obtained as solutions of boundary-value problems long familiar in the theory of diffusion or conduction of heat. That a similar reduction should be possible under our present circumstances is apparent when we recall that the interpretation of the problem of random flights as one in diffusion (or heat conduction) is possible only if there exist time intervals Δt long enough for the particle to suffer a large number of individual displacements but still short enough for the *net* mean square displacement $\langle |\Delta \boldsymbol{R}|^2 \rangle_{\mathrm{Av}}$ to be small and of $O(\Delta t)$. And, it is in the essence of Brownian motion that there exist time intervals Δt during which the physical parameters (like position and velocity of the Brownian particle) change by "infinitesimal" amounts while there occur a very large number of fluctuations characteristic of the motion and arising from the collisions with the molecules of the surrounding fluid.

It is clear that for the solutions of the most general problem we shall require the density function $W(\boldsymbol{r}, \boldsymbol{u}, t)$; in other words, we should really consider the problem in the six-dimensional phase space. Accordingly, we may state our principal objective by the remark that what we are seeking is essentially a generalization of Liouville's theorem of classical dynamics to include Brownian motion. But before we proceed to establish such a general theorem it will be instructive to consider the simplest problem of the Brownian motion of a free particle in the velocity space and obtain a differential equation for $W(\boldsymbol{u}, t)$; this leads us to the discussion of the Fokker-Planck equation in its most familiar form.

(i) *The Fokker-Planck Equation in Velocity Space to Describe the Brownian Motion of a Free Particle*

Let Δt denote an interval of time long compared to the periods of fluctuations of the acceleration $\boldsymbol{A}(t)$ occurring in the Langevin equation but short compared to intervals during which the velocity of a Brownian particle changes by appreciable amounts. Under these circumstances we should expect to derive the distribution function $W(\boldsymbol{u}, t+\Delta t)$ governing the probability of occurrence of $\boldsymbol{u}$ at time $t+\Delta t$ from the distribution $W(\boldsymbol{u}, t)$ at time t and a knowledge of the *transition probability* $\psi(\boldsymbol{u}; \Delta\boldsymbol{u})$ that $\boldsymbol{u}$ suffers an increment $\Delta\boldsymbol{u}$ in time Δt. More particularly, we expect the relation

$$W(\boldsymbol{u}, t+\Delta t) = \int W(\boldsymbol{u}-\Delta\boldsymbol{u}, t)\psi(\boldsymbol{u}-\Delta\boldsymbol{u}; \Delta\boldsymbol{u})d(\Delta\boldsymbol{u}), \tag{218}$$

to be valid. We may parenthetically remark that in expecting this integral equation between $W(\boldsymbol{u}, t+\Delta t)$ and $W(\boldsymbol{u}, t)$ to be true we are actually supposing that the course which a Brownian particle will take depends only on the instantaneous values of its physical parameters and is entirely independent of its whole previous history. In general probability theory, a stochastic process which has this characteristic, namely, that what happens at a given instant of time t depends *only* on the state of the system at time t is said to be a *Markoff* process. We may describe a Markoff process picturesquely by the statement that it represents "the gradual unfolding of a transition probability" in exactly the same sense as the development of a conservative dynamical system can be described as "the gradual unfolding of a contact transformation." That we should be able to idealize Brownian motion as a Markoff process appears very reasonable. But we should be careful not to conclude too hastily that every stochastic process is necessarily of the Markoff type. For, it can happen that the future course of a system is conditioned by its past history: i.e., what happens at a given instant of time t may depend on what has already happened during all time preceding t.

Returning to Eq. (218), for the case under discussion we have

$$\psi(\boldsymbol{u};\Delta\boldsymbol{u})=\frac{1}{(4\pi q\Delta t)^{\frac{3}{2}}}\exp\left(-|\Delta\boldsymbol{u}+\beta\boldsymbol{u}\Delta t|^2/4q\Delta t\right)\quad(q=\beta kT/m). \tag{219}$$

For, according to the Langevin equation [cf. Eq. (142)]

$$\Delta\boldsymbol{u}=-\beta\boldsymbol{u}\Delta t+\boldsymbol{B}(\Delta t) \tag{220}$$

where $\boldsymbol{B}(\Delta t)$ denotes the net acceleration arising from fluctuations which a Brownian particle suffers in time Δt; and, since the distribution of $\boldsymbol{B}(\Delta t)$ is given by Eq. (144), the transition probability (218) follows at once.

Expanding $W(\boldsymbol{u},t+\Delta t)$, $W(\boldsymbol{u}-\Delta\boldsymbol{u},t)$ and $\psi(\boldsymbol{u}-\Delta\boldsymbol{u};\Delta\boldsymbol{u})$ in Eq. (218) in the form of Taylor series, we obtain

$$W(\boldsymbol{u},t)+\frac{\partial W}{\partial t}\Delta t+O(\Delta t^2)$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left\{W(\boldsymbol{u},t)-\sum_i\frac{\partial W}{\partial u_i}\Delta u_i+\frac{1}{2}\sum_i\frac{\partial^2 W}{\partial u_i^2}\Delta u_i^2+\sum_{i<j}\frac{\partial^2 W}{\partial u_i\partial u_j}\Delta u_i\Delta u_j+\cdots\right\}$$

$$\times\left\{\psi(\boldsymbol{u};\Delta\boldsymbol{u})-\sum_i\frac{\partial\psi}{\partial u_i}\Delta u_i+\frac{1}{2}\sum_i\frac{\partial^2\psi}{\partial u_i^2}\Delta u_i^2+\sum_{i<j}\frac{\partial^2\psi}{\partial u_i\partial u_j}\Delta u_i\Delta u_j+\cdots\right\}d(\Delta u_1)d(\Delta u_2)d(\Delta u_3) \tag{221}$$

or, writing

$$\left.\begin{aligned}\langle\Delta u_i\rangle_{\text{Av}}&=\int_{-\infty}^{+\infty}\Delta u_i\psi(\boldsymbol{u};\Delta\boldsymbol{u})d(\Delta\boldsymbol{u}),\\ \langle\Delta u_i^2\rangle_{\text{Av}}&=\int_{-\infty}^{+\infty}\Delta u_i^2\psi(\boldsymbol{u};\Delta\boldsymbol{u})d(\Delta\boldsymbol{u}),\\ \langle\Delta u_i\Delta u_j\rangle_{\text{Av}}&=\int_{-\infty}^{+\infty}\Delta u_i\Delta u_j\psi(\boldsymbol{u};\Delta\boldsymbol{u})d(\Delta\boldsymbol{u}),\end{aligned}\right\} \tag{222}$$

we have

$$\begin{aligned}\frac{\partial W}{\partial t}\Delta t+O(\Delta t^2)=&-\sum_i\frac{\partial W}{\partial u_i}\langle\Delta u_i\rangle_{\text{Av}}+\frac{1}{2}\sum_i\frac{\partial^2 W}{\partial u_i^2}\langle\Delta u_i^2\rangle_{\text{Av}}+\sum_{i<j}\frac{\partial^2 W}{\partial u_i\partial u_j}\langle\Delta u_i\Delta u_j\rangle_{\text{Av}}-\sum_i W\frac{\partial}{\partial u_i}\langle\Delta u_i\rangle_{\text{Av}}\\ &+\sum_i\frac{\partial}{\partial u_i}\langle\Delta u_i^2\rangle_{\text{Av}}\frac{\partial W}{\partial u_i}+\sum_{i\neq j}\frac{\partial W}{\partial u_i}\frac{\partial}{\partial u_j}\langle\Delta u_i\Delta u_j\rangle_{\text{Av}}+\frac{1}{2}\sum_i\frac{\partial^2}{\partial u_i^2}\langle\Delta u_i^2\rangle_{\text{Av}}W\\ &+\sum_{i<j}W\frac{\partial^2}{\partial u_i\partial u_j}\langle\Delta u_i\Delta u_j\rangle_{\text{Av}}+O(\langle\Delta u_i\Delta u_j\Delta u_k\rangle_{\text{Av}}).\end{aligned} \tag{223}$$

where the remainder term involves the averages of the quantities

$$\Delta u_i^3, \quad \Delta u_i^2 \Delta u_j \quad \text{and} \quad \Delta u_i \Delta u_j \Delta u_k, \quad (i, j, k = 1, 2, 3).$$

Equation (223) can be written more conveniently as

$$\frac{\partial W}{\partial t}\Delta t + O(\Delta t^2) = -\sum_i \frac{\partial}{\partial u_i}(W\langle\Delta u_i\rangle_{Av}) + \frac{1}{2}\sum_i \frac{\partial^2}{\partial u_i^2}(W\langle\Delta u_i^2\rangle_{Av}) + \sum_{i<j}\frac{\partial^2}{\partial u_i \partial u_j}(W\langle\Delta u_i \Delta u_j\rangle_{Av}) + O(\langle\Delta u_i \Delta u_j \Delta u_k\rangle_{Av}), \tag{224}$$

which is the ***Fokker-Planck equation*** in its most general form.

For the transition probability (219),

$$\langle\Delta u_i\rangle_{Av} = -\beta u_i \Delta t; \quad \langle\Delta u_i \Delta u_j\rangle_{Av} = O(\Delta t^2); \quad \langle\Delta u_i^2\rangle_{Av} = 2q\Delta t + O(\Delta t^2). \tag{225}$$

Hence, Eq. (224) reduces in our case to

$$\frac{\partial W}{\partial t}\Delta t + O(\Delta t^2) = \{\beta \operatorname{div}_u (W\mathbf{u}) + q\nabla_u^2 W\}\Delta t + O(\Delta t^2), \tag{226}$$

and passing now to the limit $\Delta t = 0$ we have

$$\partial W/\partial t = \beta \operatorname{div}_u (W\mathbf{u}) + q\nabla_u^2 W. \tag{227}$$

We shall now show that the distribution function $W(\mathbf{u}, t; \mathbf{u}_0)$ obtained in §2, Eq. (161) is the fundamental solution of the Fokker-Planck Eq. (227) in the sense that this is the solution which tends to the δ function

$$\delta(u_1 - u_{1,0})\delta(u_2 - u_{2,0})\delta(u_3 - u_{3,0}) \tag{228}$$

as $t \to 0$. To prove this, we first note that but for the Laplacian term, Eq. (227) is a linear partial differential equation of the first order. Hence, it is natural to expect that the general solution of Eq. (227) will be intimately connected with that of the associated first-order equation

$$(\partial W/\partial t) - \beta \operatorname{div}_u (W\mathbf{u}) = 0. \tag{229}$$

The general solution of this first-order equation involves the three first integrals of the Lagrangian subsidiary system

$$d\mathbf{u}/dt = -\beta \mathbf{u}. \tag{230}$$

The required first integrals are therefore

$$\mathbf{u}e^{\beta t} = \mathbf{u}_0 = \text{constant}. \tag{231}$$

Accordingly, for solving Eq. (227) we introduce a new vector $\boldsymbol{\varrho}$ defined by

$$\boldsymbol{\varrho} = (\xi, \eta, \zeta) = \mathbf{u}e^{\beta t}. \tag{232}$$

Equation (227) now becomes

$$\partial W/\partial t = 3\beta W + qe^{2\beta t}\nabla_\varrho^2 W. \tag{233}$$

This equation can be further simplified by introducing the variable

$$\chi = We^{-3\beta t}. \tag{234}$$

We have

$$\frac{\partial \chi}{\partial t} = qe^{2\beta t}\left(\frac{\partial^2 \chi}{\partial \xi^2} + \frac{\partial^2 \chi}{\partial \eta^2} + \frac{\partial^2 \chi}{\partial \zeta^2}\right). \tag{235}$$

The solution of this equation can be readily written down by using the following lemma:

Lemma I. If $\phi(t)$ is an arbitrary function of time, the solution of the partial differential equation

$$\partial\chi/\partial t=\phi^2(t)\nabla_{\varrho}{}^2\chi \tag{236}$$

which has a source at $\varrho=\varrho_0$ at time $t=0$ is

$$\chi=\frac{1}{\left[4\pi\int_0^t\phi^2(t)dt\right]^{\frac{3}{2}}}\exp\left(-|\varrho-\varrho_0|^2\Big/4\int_0^t\phi^2(t)dt\right). \tag{237}$$

We shall omit the proof of this lemma as it is very elementary.

Applying this lemma to Eq. (235) we have the fundamental solution

$$\chi=\frac{1}{\left[4\pi q\int_0^t e^{2\beta t}dt\right]^{\frac{3}{2}}}\exp\left(-|ue^{\beta t}-u_0|^2\Big/4q\int_0^t e^{2\beta t}dt\right), \tag{238}$$

or, returning to the variable W according to Eq. (234) we have

$$W(u,t;u_0)=\frac{1}{[2\pi q(1-e^{-2\beta t})/\beta]^{\frac{3}{2}}}\exp\left[-\beta|u-u_0e^{-\beta t}|^2/2q(1-e^{-2\beta t})\right] \tag{239}$$

which agrees with our earlier result in §2, Eq. (161).

(ii) *The Generalization of Liouville's Theorem to Include Brownian Motion*

We shall now consider the general problem of a particle describing Brownian motion and under the influence of an external field of force.

Let Δt again denote an interval of time which is long compared to the periods of fluctuations of the acceleration $\boldsymbol{A}(t)$ occurring in the Langevin Eq. (184) but short compared to the intervals in which any of the physical parameters change appreciably. Then, the increments $\Delta\boldsymbol{r}$ and $\Delta\boldsymbol{u}$ in position and velocity which the particle suffers during Δt are

$$\Delta\boldsymbol{r}=\boldsymbol{u}\Delta t;\quad \Delta\boldsymbol{u}=-(\beta\boldsymbol{u}-\boldsymbol{K})\Delta t+\boldsymbol{B}(\Delta t), \tag{240}$$

where $\boldsymbol{K}$ denotes the acceleration per unit mass caused by the external field of force and $\boldsymbol{B}(\Delta t)$ the net acceleration arising from fluctuations which the particle suffers in time Δt. The distribution of $\boldsymbol{B}(\Delta t)$ is again given by Eq. (144).

Assuming as before that the Brownian movement can be idealized as a Markoff process the probability distribution $W(\boldsymbol{r},\boldsymbol{u},t+\Delta t)$ in *phase space* at time $t+\Delta t$ can be derived from the distribution $W(\boldsymbol{r},\boldsymbol{u},t)$ at the earlier time t by means of the integral equation

$$W(\boldsymbol{r},\boldsymbol{u},t+\Delta t)=\int\int W(\boldsymbol{r}-\Delta\boldsymbol{r},\boldsymbol{u}-\Delta\boldsymbol{u},t)\Psi(\boldsymbol{r}-\Delta\boldsymbol{r},\boldsymbol{u}-\Delta\boldsymbol{u};\Delta\boldsymbol{r},\Delta\boldsymbol{u})d(\Delta\boldsymbol{r})d(\Delta\boldsymbol{u}). \tag{241}$$

According to the Eqs. (240) we can write

$$\Psi(\boldsymbol{r},\boldsymbol{u};\Delta\boldsymbol{r},\Delta\boldsymbol{u})=\psi(\boldsymbol{r},\boldsymbol{u};\Delta\boldsymbol{u})\delta(\Delta x-u_1\Delta t)\delta(\Delta y-u_2\Delta t)\delta(\Delta z-u_3\Delta t), \tag{242}$$

where the δ's denote Dirac's δ functions and $\psi(\boldsymbol{r},\boldsymbol{u};\Delta\boldsymbol{u})$ the transition probability in the velocity space. With this form for the transition probability in the phase space the integration over $\Delta\boldsymbol{r}$ in

Eq. (241) is immediately performed and we get

$$W(r, u, t+\Delta t) = \int W(r - u\Delta t, u - \Delta u, t)\psi(r - u\Delta t, u - \Delta u; \Delta u)d(\Delta u). \tag{243}$$

Alternatively, we can write

$$W(r+u\Delta t, u, t+\Delta t) = \int W(r, u-\Delta u, \Delta t)\psi(r, u-\Delta u; \Delta u)d(\Delta u). \tag{244}$$

Expanding the various functions in the foregoing equation in the form of Taylor series and proceeding as in our derivation of the Fokker-Planck equation, we obtain [cf. Eq. (221)]

$$\left(\frac{\partial W}{\partial t} + u\cdot\mathrm{grad}_r\, W\right)\Delta t + O(\Delta t^2) = -\sum_i \frac{\partial}{\partial u_i}(W\langle\Delta u_i\rangle_{Av}) + \frac{1}{2}\sum_i \frac{\partial^2}{\partial u_i^2}(W\langle\Delta u_i^2\rangle_{Av})$$
$$+\sum_{i<j}\frac{\partial^2}{\partial u_i \partial u_j}(W\langle\Delta u_i \Delta u_j\rangle_{Av}) + O(\langle\Delta u_i \Delta u_j \Delta u_k\rangle_{Av}). \tag{245}$$

This is the complete analog in the phase space of the Fokker-Planck equation in the velocity space.

For the case (240), the transition probability $\psi(u; \Delta u)$ is given by [cf. Eq. (144)]

$$\psi(u; \Delta u) = \frac{1}{(4\pi q\Delta t)^{\frac{3}{2}}} \exp\left(-|\Delta u + (\beta u - K)\Delta t|^2/4q\Delta t\right). \tag{246}$$

And with this expression for the transition probability we clearly have

$$\langle\Delta u_i\rangle_{Av} = -(\beta u_i - K_i)\Delta t;\quad \langle\Delta u_i^2\rangle_{Av} = 2q\Delta t + O(\Delta t^2);\quad \langle\Delta u_i \Delta u_j\rangle_{Av} = O(\Delta t^2). \tag{247}$$

Accordingly Eq. (245) simplifies to

$$\left\{\frac{\partial W}{\partial t} + u\cdot\mathrm{grad}_r\, W\right\}\Delta t + O(\Delta t^2) = \left\{\sum_i \frac{\partial}{\partial u_i}[(\beta u_i - K_i)W] + q\sum_i \frac{\partial^2 W}{\partial u_i^2}\right\}\Delta t + O(\Delta t^2), \tag{248}$$

and now passing to the limit $\Delta t = 0$ and after some minor rearranging of the terms we finally obtain

$$\partial W/\partial t + u\cdot\mathrm{grad}_r\, W + K\cdot\mathrm{grad}_u\, W = \beta\, \mathrm{div}_u\,(Wu) + q\nabla_u^2 W. \tag{249}$$

The foregoing equation represents the complete generalization of the Fokker-Planck Eq. (227) to the phase space. At the same time Eq. (249) represents also the generalization of Liouville's theorem of classical dynamics to include Brownian motion; more particularly, on the right-hand side of Eq. (249) we have the terms arising from Brownian motion while on the left-hand side we have the usual Stokes operator D/Dt acting on W.

(iii) The Solution of Equation (249) *for the Field Free Case*

When no external field is present Eq. (249) becomes

$$\partial W/\partial t + u\cdot\mathrm{grad}_r\, W = 3\beta W + \beta u\cdot\mathrm{grad}_u\, W + q\nabla_u^2 W. \tag{250}$$

To solve this equation we again note that the equation

$$\partial W/\partial t + u\cdot\mathrm{grad}_r\, W = 3\beta W + \beta u\cdot\mathrm{grad}_u\, W \tag{251}$$

derived from (250) by ignoring the Laplacian term $q\nabla_u^2 W$ is a linear homogeneous first-order partial differential equation for $We^{-3\beta t}$. Accordingly, the general solution of Eq. (251) can be expressed in terms of any six independent integrals of the Lagrangian subsidiary system

$$du/dt=-\beta u;\quad dr/dt=u. \tag{252}$$

Two vector integrals of this system are

$$ue^{\beta t}=I_1;\quad r+u/\beta=I_2. \tag{253}$$

Accordingly, to solve Eq. (119) we introduce the new variables

$$\varrho=(\xi,\eta,\zeta)=ue^{\beta t};\quad P=(X,Y,Z)=r+u/\beta. \tag{254}$$

For this transformation of the variables we have

$$\left.\begin{aligned}\frac{\partial W}{\partial t}&=\frac{\partial}{\partial t}W(\varrho,P,t)+\beta\varrho\cdot\mathrm{grad}_\varrho W,\\ \mathrm{grad}_r W&=\mathrm{grad}_P W,\\ \mathrm{grad}_u W&=e^{\beta t}\,\mathrm{grad}_\varrho W+(1/\beta)\,\mathrm{grad}_P W,\end{aligned}\right\} \tag{255}$$

and finally

$$\nabla_u^2W=e^{2\beta t}\nabla_\varrho^2W+(2/\beta)e^{\beta t}\nabla_\varrho\cdot\nabla_PW+(1/\beta^2)\nabla_P^2W. \tag{256}$$

Substituting the foregoing equations in Eq. (250) we obtain

$$\partial W/\partial t=3\beta W+q\{e^{2\beta t}\nabla_\varrho^2W+(2/\beta)e^{\beta t}\nabla_\varrho\cdot\nabla_PW+(1/\beta^2)\nabla_P^2W\}. \tag{257}$$

Again, we introduce the variable

$$\chi=We^{-3\beta t}. \tag{258}$$

Equation (257) reduces to

$$\partial\chi/\partial t=q\{e^{2\beta t}\nabla_\varrho^2\chi+(2/\beta)e^{\beta t}\nabla_\varrho\cdot\nabla_P\chi+(1/\beta^2)\nabla_P^2\chi\}, \tag{259}$$

or, written out explicitly

$$\frac{\partial\chi}{\partial t}=q\left\{e^{2\beta t}\left(\frac{\partial^2\chi}{\partial\xi^2}+\frac{\partial^2\chi}{\partial\eta^2}+\frac{\partial^2\chi}{\partial\zeta^2}\right)+\frac{2}{\beta}e^{\beta t}\left(\frac{\partial^2\chi}{\partial\xi\partial X}+\frac{\partial^2\chi}{\partial\eta\partial Y}+\frac{\partial^2\chi}{\partial\zeta\partial Z}\right)+\frac{1}{\beta^2}\left(\frac{\partial^2\chi}{\partial X^2}+\frac{\partial^2\chi}{\partial Y^2}+\frac{\partial^2\chi}{\partial Z^2}\right)\right\}. \tag{260}$$

To solve this equation we first prove the following lemma:

Lemma II. Let $\phi(t)$ and $\psi(t)$ be two arbitrary functions of time. The solution of the differential equation

$$\frac{\partial\chi}{\partial t}=\phi^2(t)\frac{\partial^2\chi}{\partial\xi^2}+2\phi(t)\psi(t)\frac{\partial^2\chi}{\partial\xi\partial X}+\psi^2(t)\frac{\partial^2\chi}{\partial X^2} \tag{261}$$

which has a source at $\xi=X=0$ at $t=0$ is

$$\chi=\frac{1}{2\pi\Delta^{\frac{1}{2}}}\exp\left[-(a\xi^2+2h\xi X+bX^2)/2\Delta\right] \tag{262}$$

where

$$a=2\int_0^t\psi^2(t)dt;\quad h=-2\int_0^t\phi(t)\psi(t)dt;\quad b=2\int_0^t\phi^2(t)dt, \tag{263}$$

and

$$\Delta=ab-h^2. \tag{264}$$

To prove this lemma we substitute for χ according to Eq. (262) in the differential Eq. (261). After some minor reductions we find that we are left with

$$\frac{1}{\Delta}\frac{d\Delta}{dt}+\xi^2\frac{da_1}{dt}+2\xi X\frac{dh_1}{dt}+X^2\frac{db_1}{dt}+2\phi^2(a_1^2\xi^2+2a_1h_1\xi X+h_1^2X^2-a_1)$$
$$+4\phi\psi(a_1h_1\xi^2+h_1b_1X^2+\xi X[h_1^2+a_1b_1]-h_1)+2\psi^2(h_1^2\xi^2+2h_1b_1\xi X+b_1^2X^2-b_1)=0, \quad (265)$$

where we have written

$$a_1=a/\Delta; \quad h_1=h/\Delta; \quad b_1=b/\Delta. \quad (266)$$

Equating the coefficients of ξ^2, ξX and X^2 in (265) we obtain the set of equations

$$\left.\begin{aligned} da_1/dt&=-2(a_1\phi+h_1\psi)^2,\\ db_1/dt&=-2(h_1\phi+b_1\psi)^2,\\ dh_1/dt&=-2(a_1\phi+h_1\psi)(h_1\phi+b_1\psi),\end{aligned}\right\} \quad (267)$$

and

$$d\Delta/dt=2\Delta(a_1\phi^2+2h_1\phi\psi+b_1\psi^2). \quad (268)$$

It is readily verified that Eq. (268) is consistent with the Eq. (267) [see Eqs. (271) and (272) below].

Since [cf. Eqs. (266)]

$$da/dt=\Delta(da_1/dt)+a_1(d\Delta/dt), \quad (269)$$

we have according to Eqs. (267) and (268)

$$da/dt=-2\Delta(a_1\phi+h_1\psi)^2+2\Delta(a_1^2\phi^2+2a_1h_1\phi\psi+a_1b_1\psi^2)=2\Delta(a_1b_1-h_1^2)\psi^2, \quad (270)$$

or

$$da/dt=2\psi^2. \quad (271)$$

Similarly we prove that

$$db/dt=2\phi^2; \quad dh/dt=-2\phi\psi. \quad (272)$$

Hence,

$$a=2\int^t \psi^2 dt; \quad h=-2\int^t \phi\psi dt; \quad b=2\int^t \phi^2 dt. \quad (273)$$

The lemma now follows as an immediate consequence of the boundary conditions at $t=0$ stated.

In order to apply the foregoing lemma to Eq. (260) we first notice that the equation is separable in the pairs of variables (ξ, X), (η, Y) and (ζ, Z). Expressing therefore the solution in the form

$$\chi=\chi_1(\xi, X)\chi_2(\eta, Y)\chi_3(\zeta, Z), \quad (274)$$

we see that each of the functions χ_1, χ_2 and χ_3 satisfies an equation of the form (261) with

$$\phi(t)=q^{\frac{1}{2}}e^{\beta t}; \quad \psi(t)=q^{\frac{1}{2}}/\beta. \quad (275)$$

Hence, the solution of Eq. (260) with the boundary condition

$$\mathbf{r}=\mathbf{r}_0, \quad \mathbf{P}=\mathbf{P}_0 \quad \text{at} \quad t=0 \quad (276)$$

is

$$\chi=\frac{1}{8\pi^3\Delta^{\frac{3}{2}}}\exp\{-[a|\mathbf{r}-\mathbf{r}_0|^2+2h(\mathbf{r}-\mathbf{r}_0)\cdot(\mathbf{P}-\mathbf{P}_0)+b|\mathbf{P}-\mathbf{P}_0|^2]/2\Delta\} \quad (277)$$

where

$$\left.\begin{aligned} a&=2q\beta^{-2}\int_0^t dt=2q\beta^{-2}t,\\ b&=2q\int_0^t e^{2\beta t}dt=q\beta^{-1}(e^{2\beta t}-1),\\ h&=-2q\beta^{-1}\int_0^t e^{\beta t}dt=-2q\beta^{-2}(e^{\beta t}-1),\end{aligned}\right\} \quad (278)$$

and

$$\varrho - \varrho_0 = e^{\beta t}u - u_0; \quad P - P_0 = r + u/\beta - r_0 - u_0/\beta. \tag{279}$$

In Eq. (279) r_0 and u_0 denote the position and velocity of the Brownian particle at time $t=0$. Finally,

$$W = \frac{e^{3\beta t}}{8\pi^3\Delta^{\frac{3}{2}}} \exp \{ -[a|\varrho - \varrho_0|^2 + 2h(\varrho - \varrho_0)\cdot(P - P_0) + b|P - P_0|^2]/2\Delta \}. \tag{280}$$

We shall now verify that the foregoing solution for W obtained as the fundamental solution of Eq. (250) agrees with what we obtained in §2 through a discussion of the Langevin equation: With R and S as defined in Eqs. (181) we have

$$\varrho - \varrho_0 = e^{\beta t}S; \quad P - P_0 = R + (1/\beta)S. \tag{281}$$

Accordingly,

$$\left. \begin{aligned} a|\varrho - \varrho_0|^2 + 2h(\varrho - \varrho_0)\cdot(P - P_0) + b|P - P_0|^2 &= ae^{2\beta t}|S|^2 + 2he^{\beta t}(R\cdot S + (1/\beta)|S|^2) + b|R + (1/\beta)S|^2, \\ &= e^{2\beta t}(F|S|^2 - 2HR\cdot S + G|R|^2), \end{aligned} \right\} \tag{282}$$

where

$$F = a + 2h\beta^{-1}e^{-\beta t} + b\beta^{-2}e^{-2\beta t}; \quad G = be^{-2\beta t}; \quad H = -(he^{-\beta t} + b\beta^{-1}e^{-2\beta t}). \tag{283}$$

With a, b and h as given by Eqs. (278) we find that

$$F = q\beta^{-3}(2\beta t - 3 + 4e^{-\beta t} - e^{-2\beta t}); \quad G = q\beta^{-1}(1 - e^{-2\beta t}); \quad H = q\beta^{-2}(1 - e^{-\beta t})^2. \tag{284}$$

Further,

$$FG - H^2 = (ab - h^2)e^{-2\beta t} = \Delta e^{-2\beta t}. \tag{285}$$

Thus the solution (280) can be expressed alternatively in the form

$$W = \frac{1}{8\pi^3(FG - H^2)^{\frac{3}{2}}} \exp [-(F|S|^2 - 2HR\cdot S + G|R|^2)/2(FG - H^2)]. \tag{286}$$

Comparing Eqs. (284) and (286) with Eqs. (178), (182) and (183) we see that the verification is complete.

(iv) The Solution of Equation (249) for the Case of a Harmonically Bound Particle

The method of solution is sufficiently illustrated by considering the case of a one-dimensional oscillator describing Brownian motion. Equation (249) then reduces to

$$\frac{\partial W}{\partial t} + u\frac{\partial W}{\partial x} - \omega^2 x\frac{\partial W}{\partial u} = \beta u\frac{\partial W}{\partial u} + \beta W + q\frac{\partial^2 W}{\partial u^2}. \tag{287}$$

As in our discussion in the two preceding sections we introduce as variables two first integrals of the associated subsidiary system:

$$dx/dt = u; \quad du/dt = -\beta u - \omega^2 x. \tag{288}$$

Two independent first integrals of Eqs. (288) are readily seen to be

$$(x\mu_1 - u)\exp(-\mu_2 t) \quad \text{and} \quad (x\mu_2 - u)\exp(-\mu_1 t) \tag{289}$$

where μ_1 and μ_2 have the same meanings as in §3 [cf. Eqs. (189) and (190)]. Accordingly we set

$$\xi = (x\mu_1 - u)\exp(-\mu_2 t); \quad \eta = (x\mu_2 - u)\exp(-\mu_1 t). \tag{290}$$

In these variables Eq (287) becomes

$$\frac{\partial W}{\partial t}=\beta W+q\left(\exp\left(-2\mu_2 t\right)\frac{\partial^2 W}{\partial \xi^2}+2\exp\left(-(\mu_1+\mu_2)t\right)\frac{\partial^2 W}{\partial\xi\partial\eta}+\exp\left(-2\mu_1 t\right)\frac{\partial^2 W}{\partial\eta^2}\right). \tag{291}$$

Introducing the further transformation

$$W=\chi e^{\beta t}, \tag{292}$$

we finally obtain

$$\frac{\partial \chi}{\partial t}=q\left(\exp\left(-2\mu_2 t\right)\frac{\partial^2 \chi}{\partial \xi^2}+2\exp\left[-(\mu_1+\mu_2)t\right]\frac{\partial^2 \chi}{\partial\xi\partial\eta}+\exp\left(-2\mu_1 t\right)\frac{\partial^2 \chi}{\partial\eta^2}\right). \tag{293}$$

This equation is of the same form as Eq. (261) in lemma II. Hence the solution of this equation which tends to $\delta(\xi-\xi_0)\delta(\eta-\eta_0)$ as $t\to 0$ is

$$\chi=\frac{1}{2\pi\Delta^{\frac{1}{2}}}\exp\left\{-\left[a(\xi-\xi_0)^2+2h(\xi-\xi_0)(\eta-\eta_0)+b(\eta-\eta_0)^2\right]/2\Delta\right\}, \tag{294}$$

where

$$\left.\begin{aligned} a&=2q\int_0^t \exp\left(-2\mu_1 t\right)dt=\frac{q}{\mu_1}\left[1-\exp\left(-2\mu_1 t\right)\right],\\ b&=2q\int_0^t \exp\left(-2\mu_2 t\right)dt=\frac{q}{\mu_2}\left[1-\exp\left(-2\mu_2 t\right)\right],\\ h&=-2q\int_0^t \exp\left[-(\mu_1+\mu_2)t\right]dt=-\frac{2q}{\mu_1+\mu_2}\left\{1-\exp\left[-(\mu_1+\mu_2)t\right]\right\}.\end{aligned}\right\} \tag{295}$$

Further,

$$\xi_0=x_0\mu_1-u_0;\quad \eta_0=x_0\mu_2-u_0, \tag{296}$$

where x_0 and u_0 denote the position and velocity of the particle at time $t=0$. It is again verified that the solution

$$W=\frac{e^{\beta t}}{2\pi\Delta^{\frac{1}{2}}}\exp\left\{-\left[a(\xi-\xi_0)^2+2h(\xi-\xi_0)(\eta-\eta_0)+b(\eta-\eta_0)^2\right]/2\Delta\right\}, \tag{297}$$

obtained as the fundamental solution of Eq. (287) is in agreement with the distributions obtained in §3 through a discussion of the Langevin equation.

(v) *The General Case*

Our discussion in the two preceding sections suggests that in dealing with Eq. (249) quite generally we may introduce as new variables six independent first integrals of the equations of motion

$$d\mathbf{r}/dt=\mathbf{u};\quad d\mathbf{u}/dt=-\beta\mathbf{u}+\mathbf{K}. \tag{298}$$

These are the Lagrangian subsidiary equations of the linear first-order equation derived from (249) after ignoring the Laplacian term $q\nabla_u^2 W$. If $I_1,\cdots,I_6$ are six such integrals, we introduce

$$I_1(\mathbf{r},\mathbf{u},t),\cdots,I_6(\mathbf{r},\mathbf{u},t) \tag{299}$$

as the new independent variables. If we further set

$$W=\chi e^{3\beta t}, \tag{300}$$

Eq. (249) will transform to

$$\partial\chi/\partial t=q\left[\nabla_u^2\chi\right]_{I_1,\cdots I_6}, \tag{301}$$

where the Laplacian of χ on the right-hand side has to be expressed in terms of the new variables $I_1,\cdots,I_6$.

We shall thus be left with a general linear partial differential equation of the second order for χ; and we seek a solution of this equation of the form

$$\chi = \frac{1}{8\pi^3\Delta^{\frac{1}{2}}} e^{-Q/2\Delta}. \tag{302}$$

where Q stands for a general homogeneous quadratic form in the six variables $I_1, \cdots, I_6$ with coefficients which are functions of time only. Further, in Eq. (302) Δ is the determinant of the matrix formed by the coefficients of the quadratic form. In this manner we can expect to solve the general problem.

(*vi*) *The Differential Equation for the Displacement* ($t \gg \beta^{-1}$). *The Smoluchowski Equation*

We have seen that all the physically significant questions concerning the motion of a free Brownian particle can be answered by solving Eq. (250) with appropriate boundary conditions. However, if we are interested only in time intervals very large compared to the "*time of relaxation*" β^{-1} we can apply the method of the Fokker-Planck equation to configuration space ($\mathbf{r}$) independently of the velocity space. For, according to Eq. (172), we may say that for a free Brownian particle, the transition probability that r suffers an increment Δr in time $\Delta t \gg \beta^{-1}$ is given by

$$\psi(\Delta r) = \frac{1}{(4\pi D\Delta t)^{\frac{3}{2}}} \exp\left(-|\Delta r|^2/4D\Delta t\right), \tag{303}$$

where

$$D = q/\beta^2 = kT/m\beta. \tag{304}$$

Thus, again with the understanding that $\Delta t \gg \beta^{-1}$ we can write [cf. Eq. (218) and the remarks following it]

$$w(r, t+\Delta t) = \int w(r-\Delta r, t)\psi(\Delta r)d(\Delta r). \tag{305}$$

Applying now to this equation the procedure that was followed in the derivation of the Fokker-Planck equation in the velocity space we readily obtain the "diffusion equation"

$$\partial w/\partial t = D\nabla_r^2 w. \tag{306}$$

That we should be led to the diffusion equation is not surprising since Eq. (303) implies that for time intervals $\Delta t \gg \beta^{-1}$ the motion of the particle reduces to the elementary case of the problem of random flights (Chapter I, §4 case [c]) and the analysis of *I* §5 leading to Eq. (112) applies.

Equation (306) is valid for a free Brownian particle. To extend this result for the case when an external field is acting we start from Eq. (249) which is quite generally true in phase space. We first rewrite this equation in the form

$$\frac{\partial W}{\partial t} = \beta\left(\operatorname{div}_u - \frac{1}{\beta}\operatorname{div}_r\right)\left(Wu + \frac{q}{\beta}\operatorname{grad}_u W - \frac{K}{\beta}W + \frac{q}{\beta^2}\operatorname{grad}_r W\right) + \operatorname{div}_r\left(\frac{q}{\beta^2}\operatorname{grad}_r W - \frac{K}{\beta}W\right). \tag{307}$$

We now integrate this equation along the straight line

$$r + u/\beta = \text{constant} = r_0, \tag{308}$$

from $u = -\infty$ to $+\infty$. We obtain

$$\frac{\partial}{\partial t}\int_{r+u\beta^{-1}=r_0} W du = \int_{r+u\beta^{-1}=r_0} \operatorname{div}_r\left(\frac{q}{\beta^2}\operatorname{grad}_r W - \frac{K}{\beta}W\right)du. \tag{309}$$

We shall now suppose that $K(r)$ does not change appreciably over distances of the order of $(q/\beta^3)^{\frac{1}{2}}$. Then, starting from an arbitrary initial distribution $W(r, u, 0)$ at time $t=0$ we should expect that a Maxwellian distribution of the velocities will be established at all points after time intervals $\Delta t \gg \beta^{-1}$. Consequently, if we are not interested in time intervals of the order of β^{-1} we can write

$$W(r, u, t) \simeq \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} \exp\left(-m|u|^2/2kT\right) w(r, t). \tag{310}$$

With these assumptions Eq. (309) becomes

$$\frac{\partial w}{\partial t} \simeq \operatorname{div}_{r_0}\left\{\frac{q}{\beta^2}\operatorname{grad}_{r_0} w(r_0) - \frac{K(r_0)}{\beta} w(r_0)\right\} \tag{311}$$

The passage from Eqs. (309) to (311) is the result of our supposition that in the domain of u from which the dominant contribution to the integral on the right-hand side of Eq. (309) arises (namely, $|u| \lesssim (kT/m)^{\frac{1}{2}} = (q/\beta)^{\frac{1}{2}}$) the variation of r (which is of the order $|u|/\beta \simeq (q/\beta^3)^{\frac{1}{2}}$) is small compared to the distances in the configuration space in which K and w change appreciably. The required generalization of Eq. (306) is therefore

$$\frac{\partial w}{\partial t} = \operatorname{div}_r\left(\frac{q}{\beta^2}\operatorname{grad}_r w - \frac{K}{\beta} w\right). \tag{312}$$

Equation (312) is sometimes called Smoluchowski's equation.

An immediate consequence of Eq. (312) may be noted. According to this equation a *stationary diffusion current* j obeys the law

$$j = \beta^{-1} K w - q\beta^{-2} \operatorname{grad} w = \text{constant}. \tag{313}$$

If K can be derived from a potential $\mathfrak{B}$ so that

$$K = -\operatorname{grad} \mathfrak{B} \tag{314}$$

Eq. (313) can be rewritten in the form

$$j = -q\beta^{-2} \exp\left(-\beta\mathfrak{B}/q\right) \operatorname{grad}\left(w \exp\left(\beta\mathfrak{B}/q\right)\right), \tag{315}$$

where it may be noted $q/\beta = kT/m$. Integrating Eq. (315) between any two points A and B we obtain

$$j \cdot \int_A^B \beta \exp\left(\beta\mathfrak{B}/q\right) ds = \frac{kT}{m} w \exp\left(\beta\mathfrak{B}/q\right)\Big|_B^A, \tag{316}$$

an important equation, first derived by Kramers.

We may finally again draw attention to the fact that Eqs. (306) and (312) are valid only if we ignore effects which happen in time intervals of the order of β^{-1} and space intervals of the order of $(q/\beta^3)^{\frac{1}{2}}$; when such effects are of interest we should go back to Eqs. (249) or (250) which are rigorously valid in phase space.

(vii) *General Remarks*

So far we have only shown that the discussion based on Eq. (249) and its various special forms leads to results in agreement with those already derived on the basis of the Langevin equation. However, the special importance of the partial differential equations arises when further restrictions on the problem are imposed. For, these additional restrictions can also be expressed in the form of boundary conditions which the solutions will have to satisfy and the consequent reduction to a boundary value problem in partial differential equations provides a very direct method for obtaining the necessary solutions. The alternative analysis based on the Langevin equation would in general be too involved.

Further examples of the use of the partial differential equations obtained in this section will be found in Chapters III and IV.

5. General Remarks

A general characteristic of the stochastic processes of the type considered in the preceding sections is that the increment in the velocity, $\Delta \boldsymbol{u}$ which a particle suffers in a time Δt long compared to the periods of the elementary fluctuations can be expressed as the sum of two distinct terms: a term $\boldsymbol{K}\Delta t$ which represents the action of the external field of force, and a term $\delta\boldsymbol{u}(\Delta t)$ which denotes a fluctuating quantity with a definite law of distribution. Thus

$$\Delta \boldsymbol{u} = \boldsymbol{K}\Delta t + \delta\boldsymbol{u}(\Delta t); \tag{317}$$

the corresponding increment in the position, $\Delta \boldsymbol{r}$ is given by

$$\Delta \boldsymbol{r} = \boldsymbol{u}\Delta t, \tag{318}$$

where $\boldsymbol{u}$ is the instantaneous velocity of the particle.

When dealing with stochastic processes of the *strictly* Brownian motion type we further suppose that the term $\delta\boldsymbol{u}(\Delta t)$ in Eq. (317) can in turn be decomposed into two parts: a part $-\beta\boldsymbol{u}\Delta t$ representing the deceleration caused by the dynamical friction $-\beta\boldsymbol{u}$ and a fluctuating part $\boldsymbol{B}(\Delta t)$ which is really the vector sum of a very large number of very "minute" accelerations arising from collisions with individual molecules of the surrounding fluid:

$$\delta\boldsymbol{u}(\Delta t) = -\beta\boldsymbol{u}\Delta t + \boldsymbol{B}(\Delta t). \tag{319}$$

It is *this* particular decomposition of $\delta\boldsymbol{u}(\Delta t)$ that is peculiarly characteristic of stochastic processes of the Brownian type.

Concerning $\boldsymbol{B}(\Delta t)$ in Eq. (319) we have supposed in §§2, 3, and 4 that it is governed by the distribution function [cf. Eqs. (144) and (145)]

$$w(\boldsymbol{B}[\Delta t]) = \frac{1}{(4\pi q\Delta t)^{3/2}} \exp\left(-|\boldsymbol{B}(\Delta t)|^2/4q\Delta t\right), \tag{320}$$

where

$$q = \beta kT/m. \tag{321}$$

In this choice of the distribution function for $\boldsymbol{B}(\Delta t)$ we were guided by two considerations: *First*, that starting from any arbitrarily assigned distribution of the velocities we shall always be led to the Maxwellian distribution as $t \to \infty$ (or, alternatively that the Maxwellian distribution of the velocities is invariant to stochastic processes of the type considered); and *second* that during a time Δt in which the position and the velocity of the particle will change by an "infinitesimal" amount of order Δt the particle will in reality suffer an *exceedingly* large number of individual accelerations by collisions with the molecules of the surrounding fluid. This second consideration would suggest, from analogy with the simple case of the problem of random flights [Eq. (108)], a formula of the *form* (320). The particular value of q (321) then follows from the first requirement.

Combining Eqs. (319) and (320) we obtain for the *transition probability* $\psi(\boldsymbol{u};\delta\boldsymbol{u})$ for $\boldsymbol{u}$ to suffer an increment $\delta\boldsymbol{u}$ due to the Brownian forces only, the expression

$$\psi(\boldsymbol{u};\delta\boldsymbol{u}) = \frac{1}{(4\pi q\Delta t)^{3/2}} \exp\left(-|\beta\boldsymbol{u}\Delta t + \delta\boldsymbol{u}|^2/4q\Delta t\right). \tag{322}$$

We shall now briefly re-examine the problem of continuous stochastic processes more generally from the point of view of the invariance of the *Maxwell-Boltzmann* distribution

$$W=\text{constant}\exp\{-[m|u|^2+2m\mathfrak{V}(r)]/2kT\};\quad K=-\text{grad}\,\mathfrak{V} \tag{323}$$

to processes governed by Eqs. (317) and (318) *only* i.e., without making the further assumptions included in Eqs. (319)–(322).

Assuming, as we have done hitherto, that the stochastic process we are considering is of the Markoff type we can write the integral equation [cf. Eq. (241)]

$$W(r,u,t+\Delta t)=\int\int W(r-\Delta r,u-\Delta u,t)\Psi(r-\Delta r,u-\Delta u;\Delta r,\Delta u)d(\Delta r)d(\Delta u). \tag{324}$$

According to Eqs. (318) we expect that [cf. Eq. (242)]

$$\Psi(r,u;\Delta r,\Delta u)=\psi(r,u;\Delta u)\delta(\Delta x-u_1\Delta t)\delta(\Delta y-u_2\Delta t)\delta(\Delta z-u_3\Delta t). \tag{325}$$

Equation (324) becomes

$$W(r,u,t+\Delta t)=\int W(r-u\Delta t,u-K\Delta t-\delta u,t)\psi(r-u\Delta t,u-K\Delta t-\delta u;K\Delta t+\delta u)d(\delta u), \tag{326}$$

where we have further substituted for Δu according to Eq. (317). Equation (326) can be written alternatively as

$$W(r+u\Delta t,u+K\Delta t,t+\Delta t)=\int W(r,u-\delta u,t)\psi(r,u-\delta u;\delta u)d(\delta u). \tag{327}$$

Applying to this equation the same procedure as was adopted in the derivation of the Fokker-Planck and the generalized Liouville equations in §4, we readily find that [cf. Eq. (245)]

$$\left\{\frac{\partial W}{\partial t}+u\cdot\text{grad}_r\,W+K\cdot\text{grad}_u\,W\right\}\Delta t+O(\Delta t^2)=-\sum_i\frac{\partial}{\partial u_i}(W\langle\delta u_i\rangle_{\text{Av}})+\frac{1}{2}\sum_i\frac{\partial^2}{\partial u_i^2}(W\langle\delta u_i^2\rangle_{\text{Av}})$$
$$+\sum_{i<j}\frac{\partial^2}{\partial u_i\partial u_j}(W\langle\delta u_i\delta u_j\rangle_{\text{Av}})+O(\langle\delta u_i\delta u_j\delta u_k\rangle_{\text{Av}}) \tag{328}$$

where $\langle\delta u_i\rangle_{\text{Av}}$ etc., denote the various moments of the transition probability $\psi(r,u;\delta u)$.

We shall now suppose that

$$\langle\delta u_i\rangle_{\text{Av}}=\mu_i\Delta t+O(\Delta t^2);\quad\langle\delta u_i^2\rangle_{\text{Av}}=\mu_{ii}\Delta t+O(\Delta t^2);\quad\langle\delta u_i\delta u_j\rangle_{\text{Av}}=\mu_{ij}\Delta t+O(\Delta t^2), \tag{329}$$

and that all averages of quantities like $\delta u_i\delta u_j\delta u_k$ are of order higher than one in Δt. With this understanding we shall obtain from Eq. (328), on passing to the limit $\Delta t=0$ the result

$$\frac{\partial W}{\partial t}+u\cdot\text{grad}_r\,W+K\cdot\text{grad}_u\,W=-\sum_i\frac{\partial}{\partial u_i}(W\mu_i)+\frac{1}{2}\sum_i\frac{\partial^2}{\partial u_i^2}(W\mu_{ii})+\sum_{i<j}\frac{\partial^2}{\partial u_i\partial u_j}(W\mu_{ij}). \tag{330}$$

We now require that the Maxwell-Boltzmann distribution (323) satisfy Eq. (330) identically. On substituting this distribution in Eq. (330) we find that the left-hand side of this equation vanishes and we are left with

$$-\sum_i\frac{\partial}{\partial u_i}[\exp(-m|u|^2/2kT)\mu_i]+\frac{1}{2}\sum_i\frac{\partial^2}{\partial u_i^2}[\exp(-m|u|^2/2kT)\mu_{ii}]$$
$$+\sum_{i<j}\frac{\partial^2}{\partial u_i\partial u_j}[\exp(-m|u|^2/2kT)\mu_{ij}]=0. \tag{331}$$

Equation (331) is to be regarded as the general condition on the moments.

For the distribution (322)

$$\mu_i = -\beta u_i; \quad \mu_{ii} = 2q = 2\beta kT/m; \quad \mu_{ij} = 0. \tag{332}$$

Also, the third and higher moments of δu do not contain terms linear in Δt.

We readily verify that with the μ's given by (332) we satisfy Eq. (331). It is not, however, to be expected that (332) represents the most general solution for the μ's which will satisfy Eq. (331). It would clearly be a matter of considerable interest to investigate Eq. (331) (or the generalization of this equation to include terms involving μ_{ijk} etc.,) with a view to establishing the nature of the restrictions on the μ's implied by Eq. (331). Such an investigation might lead to the discovery of new classes of Markoff processes which will leave the Maxwell-Boltzmann distribution invariant but which will not be of the classical Brownian motion type. It is not proposed to undertake this investigation in this article. We may, however, draw special attention to the fact that according to Eqs. (331) and (332), β can very well depend on the spatial coordinates (though $q/\beta[=kT/m]$ must be a constant throughout the system). Thus, the generalized Liouville Eq. (249) and the Smoluchowski Eq. (312) are valid as they stand, also when $\beta = \beta(r)$.

CHAPTER III

PROBABILITY AFTER-EFFECTS: COLLOID STATISTICS; THE SECOND LAW OF THERMODYNAMICS. THE THEORY OF COAGULATION, SEDIMENTATION, AND THE ESCAPE OVER POTENTIAL BARRIERS

In this chapter we shall consider certain problems in the theory of Brownian motion which require the more explicit introduction than we had occasion hitherto, of the notion of *probability after-effects*. The fundamental ideas underlying this notion have already been described in the introductory section where we have also seen that colloid statistics (or, more generally, the phenomenon of density fluctuations in a medium of constant average density) provides a very direct illustration of the problem. The theory of this phenomenon which has been developed along very general lines by Smoluchowski has found beautiful confirmation in the experiments of Svedberg, Westgren, and others. This theory of Smoluchowski in addition to providing a striking application of the principles of Brownian motion has also important applications to the elucidation of the statistical nature of the second law of thermodynamics. In view, therefore, of the fundamental character of Smoluchowski's theory we shall give a somewhat detailed account of it in this chapter (§§1–3). (In the later sections of this chapter we consider further miscellaneous applications of the theory of Brownian motion which have bearings on problems considered in Chapter IV.)

1. The General Theory of Density Fluctuations for Intermittent Observations. The Mean Life and the Average Time of Recurrence of a State of Fluctuation*

Consider a geometrically well-defined small element of volume v in a solution containing Brownian particles under conditions of diffusion equilibrium. (More generally, we may also consider v as an element in a very much larger volume containing a large number of particles in equilibrium.) Suppose now that we observe the number of particles contained in v systematically at constant intervals of time τ apart. Then the frequency $W(n)$ with which different numbers of particles will be observed in v will follow the *Poisson distribution* (see Appendix III),

$$W(n) = e^{-\nu}\nu^n/n!, \tag{333}$$

where ν denotes the average number of particles that will be contained in v:

$$\langle n\rangle_{Av} = \sum_{n=0}^{\infty} nW(n) = e^{-\nu}\nu \sum_{n=1}^{\infty} \frac{\nu^{n-1}}{(n-1)!} = \nu. \tag{334}$$

In other words, the number of particles that will be observed in v is subject to *fluctuations* and the different *states of fluctuations* (which, in this case, can be labelled by n) occur with definite frequencies.

According to Eq. (333) the mean square deviation δ^2 from the average value ν is given by

$$\delta^2 = \langle (n-\nu)^2\rangle_{Av} = \langle n^2\rangle_{Av} - \nu^2. \tag{335}$$

*Professor M. S. Bartlett has pointed out (Nature 165, 727, 1950) that the treatment of the average time of recurrence in this section is not free of errors.

or, since

$$\left.\begin{aligned}\langle n^2\rangle_{Av} &= \sum_{n=0}^{\infty} n^2 \frac{e^{-\nu}\nu^n}{n!} \\ &= e^{-\nu}\left\{\nu\sum_{n=2}^{\infty}\frac{\nu^{n-2}}{(n-2)!}+\sum_{n=1}^{\infty}\frac{\nu^{n-1}}{(n-1)!}\right\} \\ &= \nu^2+\nu,\end{aligned}\right\} \tag{336}$$

we have

$$\delta^2 = \nu. \tag{337}$$

It is seen that the frequency with which the different states of fluctuation n occur is independent of all physical parameters describing the particle (e.g., radius and density) and the surrounding fluid (e.g., viscosity). The situation is, however, completely changed when we consider the *speed* with which the different states of fluctuations follow each other in time. More specifically, consider the number of particles n and m contained in v at an interval of time τ apart. We expect that the number m observed on the second occasion will be correlated with the number n observed on the first occasion. This correlation should be such, that as $t\rightarrow 0$ the result of the second observation can be predicted with certainty as n, while as $t\rightarrow\infty$ we shall observe on the second occasion numbers which will increasingly be distributed according to the Poisson distribution (333). For finite intervals of time τ we can therefore ask for the *transition probability* $W(n;m)$ that m particles will be counted in v after a time τ from the instant when there was observed to be n particles in it.

In solving the problem stated toward the end of the preceding paragraph we shall make, following Smoluchowski, the two assumptions: (1) that the motions of the individual particles are not mutually influenced and are independent of each other and (2) that all positions in the element of volume considered have equal *a priori* probability. Under these circumstances we can expect to define *a probability P that a particle somewhere inside v will have emerged from it during the time τ*. The exact value of this *probability after-effect factor P* will depend on the precise circumstances of the problem including the geometry of the volume v. In §2 we shall obtain the explicit formula for P when the motions of the individual particles are governed by the laws of Brownian motion [Eq. (380)]; and similarly in §3 we shall obtain the formula for P for the case when the particles describe linear trajectories [Eq. (413)]. Meantime, we shall continue the discussion of the speed of fluctuations on the assumption that the factor P as defined can be unambiguously evaluated depending, however, on circumstances.

It is clear that the required transition probability $W(n;m)$ can be written down in an entirely elementary way if we know the probabilities with which particles enter and leave the element of volume. More precisely, let $A_i^{(n)}$ denote the probability that starting from an initial situation in which there are n particles inside v *some* i particles will have emerged from it during τ; this probability of emergence of a certain number of particles will clearly depend on the initial number of particles inside v. Similarly, let E_i denote the probability that i particles will have entered the element of volume v during τ. Since one of our principal assumptions is that the motions of the particles are not mutually influenced, the probability of entrance of a certain number of particles cannot depend on the number already contained in it. We shall now obtain explicit expressions for these two probabilities in terms of P.

The expression for $A_i^{(n)}$ can be written down at once when we recall that this must be equal to the product of the probability P^i that some particular group of i particles leaves v during τ, the probability $(1-P)^{n-i}$ that the remaining $(n-i)$ particles do not leave v during τ, and the number of distinct ways C_i^n of selecting i particles from the initial group of n. Accordingly,

$$A_i^{(n)} = C_i^n P^i(1-P)^{n-i} = \frac{n!}{i!(n-i)!}P^i(1-P)^{n-i}, \tag{338}$$

which is a Bernoulli distribution.

To obtain the expression for E_i we first remark that this must equal the probability that i particles *emerge* from the element of volume v on an *arbitrary* occasion; since, under *equilibrium conditions* the *a priori* probabilities for the entrance and emergence of particles must be equal. Remembering further, that E_i is independent of the number of particles initially contained in v, we clearly have

$$E_i = \langle A_i^{(n)} \rangle_{Av} = \sum_{n=i}^{\infty} W(n) A_i^{(n)}, \tag{339}$$

where $W(n)$ is the probability that v initially contained n particles; $W(n)$ accordingly is given by (333). Combining Eqs. (333), (338), and (339) we therefore have

$$\left.\begin{aligned} E_i &= \sum_{n=i}^{\infty} \frac{e^{-\nu}\nu^n}{n!} \frac{n!}{i!(n-i)!} P^i(1-P)^{n-i}, \\ &= \frac{e^{-\nu}(\nu P)^i}{i!} \sum_{n=i}^{\infty} \frac{\nu^{n-i}(1-P)^{n-i}}{(n-i)!}, \\ &= \frac{e^{-\nu}(\nu P)^i}{i!} e^{\nu(1-P)}. \end{aligned}\right\} \tag{340}$$

Thus,

$$E_i = e^{-\nu P}(\nu P)^i / i!, \tag{341}$$

in other words, a Poisson distribution with variance νP.

Using the formulae (338) and (341) for $A_i^{(n)}$ and E_i we can at once write down the expression for the transition-probability $W(n; n+k)$ that there is an increase in the number of particles from n to $n+k$. We clearly have

$$W(n; n+k) = \sum_{i=0}^{n} A_i^{(n)} E_{i+k}. \tag{342}$$

Similarly, for the transition probability $W(n; n-k)$ that there is a decrease in the number of particles from n to $n-k$ we have

$$W(n; n-k) = \sum_{i=k}^{n} A_i^{(n)} E_{i-k}, \quad (k \leqslant n). \tag{343}$$

From Eqs. (338), (341), (342), and (343) we therefore obtain

$$W(n; n+k) = e^{-\nu P} \sum_{i=0}^{n} C_i^n P^i (1-P)^{n-i} \times (\nu P)^{i+k}/(i+k)!, \tag{344}$$

and

$$W(n; n-k) = e^{-\nu P} \sum_{i=k}^{n} C_i^n P^i (1-P)^{n-i} \times (\nu P)^{i-k}/(i-k)! \tag{345}$$

The foregoing expressions for the transition probabilities are due to Smoluchowski.

The formulae (344) and (345) in spite of their apparent complexity have in reality very simple structures. To see this we first introduce the Bernoulli and the Poisson distributions

$$w_1^{(n)}(x) = C_x^n (1-P)^x P^{n-x} \quad (0 \leqslant x \leqslant n), \tag{346}$$

and

$$w_2(y) = e^{-\nu P}(\nu P)^y / y! \quad (0 \leqslant y < \infty). \tag{347}$$

$w_1^{(n)}(x)$ is the probability that *some* x particles *remain* in v after a time τ when initially there were n particles in it; similarly, $w_2(y)$ is the probability that y particles enter v in time τ. In terms of the distributions (346) and (347) we can rewrite Eqs. (344) and (345) as

$$W(n; n+k) = \sum_{i=0}^{n} w_1^{(n)}(n-i) w_2(i+k), \tag{348}$$

and

$$W(n; n-k) = \sum_{i=k}^{n} w_1^{(n)}(n-i) w_2(i-k), \tag{349}$$

or, writing m for $n+k$, respectively $n-k$, we see that both Eqs. (348) and (349) can be included in the single formula

$$W(n, m) = \sum_{x+y=m} w_1^{(n)}(x) w_2(y). \tag{350}$$

In other words, *the distribution* $W(n, m)$ for a fixed value of n *is the "sum" of the two distributions* (346) *and* (347). And, therefore, the mean and the mean square deviation for the distribution of m according to (350) is the sum of the means and the mean square deviations of the component distributions (346) and (347) (see Appendix IV). Since [cf. Eqs. (334) and (335) and Appendix I Eqs. (621) and (624)]

$$\langle x \rangle_{Av} = n(1-P); \quad \langle (x - \langle x \rangle_{Av})^2 \rangle_{Av} = nP(1-P), \tag{351}$$

and

$$\langle y \rangle_{Av} = \nu P; \quad \langle (y - \langle y \rangle_{Av})^2 \rangle_{Av} = \nu P, \tag{352}$$

we conclude that

$$\langle m \rangle_{Av} = n(1-P) + \nu P, \tag{353}$$

and

$$\langle (m - \langle m \rangle_{Av})^2 \rangle_{Av} = nP(1-P) + \nu P. \tag{354}$$

Let

$$\Delta_n = m - n. \tag{355}$$

Then, according to Eqs. (354) and (355)

$$\langle \Delta_n \rangle_{Av} = \langle m \rangle_{Av} - n = (\nu - n)P, \tag{356}$$

and

$$\left.\begin{aligned} \langle \Delta_n^2 \rangle_{Av} &= \langle (m - \langle m \rangle_{Av} + \langle m \rangle_{Av} - n)^2 \rangle_{Av} \\ &= \langle (m - \langle m \rangle_{Av})^2 \rangle_{Av} + (\langle m \rangle_{Av} - n)^2 \\ &= nP(1-P) + \nu P + (\nu - n)^2 P^2, \end{aligned}\right\} \tag{357}$$

or

$$\langle \Delta_n^2 \rangle_{Av} = P^2[(\nu - n)^2 - n] + (n+\nu)P. \tag{358}$$

It is seen that according to Eq. (356) the number of particles inside v changes, on the average, in the direction of making n approach its mean value, namely ν. In other words, the density fluctuations studied here in terms of a "microscopic" analysis of the stochastic motions of the individual particles are in complete agreement with the macroscopic theory of diffusion.

The quantities $\langle\Delta_n\rangle_{Av}$ and $\langle\Delta_n^2\rangle_{Av}$ represent the mean and the mean square of the differences that are to be expected in the numbers observed on two occasions at an interval of time τ apart when on the first occasion n particles were observed. If now, we further average $\langle\Delta_n\rangle_{Av}$ and $\langle\Delta_n^2\rangle_{Av}$ over all values of n with the weight function $W(n)$ we shall obtain the mean and the mean square of the differences in the numbers of particles observed on consecutive occasions in a long sequence of observations made at constant intervals τ apart. Thus [cf. Eq. (334)]

$$\langle\Delta\rangle_{Av}=\langle\langle\Delta_n\rangle_{Av}\rangle_{Av}=\langle\nu-n\rangle_{Av}P=0, \qquad (359)$$

a result which is to be expected. On the other hand [cf. Eq. (337)]

$$\left.\begin{aligned}\langle\Delta^2\rangle_{Av}&=\langle\langle\Delta_n^2\rangle_{Av}\rangle_{Av}\\&=P^2[\langle(\nu-n)^2\rangle_{Av}-\langle n\rangle_{Av}]+\langle n+\nu\rangle_{Av}P\\&=P^2(\delta^2-\langle n\rangle_{Av})+(\langle n\rangle_{Av}+\nu)P,\end{aligned}\right\} \qquad (360)$$

or

$$\langle\Delta^2\rangle_{Av}=2\nu P. \qquad (361)$$

Equation (361) suggests a direct method for the experimental determination of the probability after-effect factor P from the simple evaluation of the mean square differences $\langle\Delta^2\rangle_{Av}$ from long sequences of observations of n (see §2 below). Further, according to Eq. (361)

$$\langle\Delta^2\rangle_{Av}=2\nu \quad \text{when} \quad P=1. \qquad (362)$$

This result is in agreement with what we should expect, since, when $P=1$ there will be no correlation between the numbers that will be observed on two occasions at an interval τ apart; $\langle\Delta^2\rangle_{Av}$ then simply becomes the mean square of the differences between two numbers each of which (without correlation) is governed by the same Poisson distribution; and, therefore [cf. Eqs. (333) and (336)],

$$\begin{aligned}\langle\Delta^2\rangle_{Av}&=\langle(n-m)^2\rangle_{Av}=\langle n^2\rangle_{Av}+\langle m^2\rangle_{Av}-2\langle n\rangle_{Av}\langle m\rangle_{Av}\\&=2(\nu^2+\nu)-2\nu^2=2\nu, \quad (P=1).\end{aligned} \qquad (363)$$

We shall now show how we can define the *mean life* and *the average time of recurrence* for a given state of fluctuation in terms of the transition probability $W(n;n)$:

$$W(n;n)=e^{-\nu P}\sum_{i=0}^{n} C_i^n P^i(1-P)^{n-i}(\nu P)^i/i!, \qquad (364)$$

which gives the probability that n will be observed on two consecutive occasions. Accordingly, the probability $\phi_n(k\tau)$ that the same number n will be observed on $(k-1)$ consecutive occasions (at constant intervals τ apart) and that on the kth occasion some number different from n will be observed is given by

$$\phi_n(k\tau)=W^{k-1}(n;n)[1-W(n;n)]. \qquad (365)$$

On the other hand, in terms of $\phi_n(k\tau)$ we can give a natural definition to the mean life to the state of fluctuation n by the equation

$$T_n=\sum_{k=1}^{\infty} k\tau\phi_n(k\tau). \qquad (366)$$

Combining Eqs. (365) and (366) we obtain

$$T_n=\tau[1-W(n;n)]\sum_{k=1}^{\infty} kW^{k-1}(n;n). \qquad (367)$$

The infinite series in Eq. (367) is readily evaluated and we find

$$T_n=\frac{\tau}{1-W(n;n)}. \qquad (368)$$

In an analogous manner we can define the time of recurrence of the state n by the equation

$$\Theta_n=\sum_{k=1}^{\infty} k\tau\psi_n(k\tau), \qquad (369)$$

where $\psi_n(k\tau)$ denotes the probability that starting from *an arbitrary state which is not n* we shall observe on $k-1$ successive occasions states which are not n and on the kth occasion observe the state n. If

$$W(Nn;Nn) \qquad (370)$$

denotes the probability that from an arbitrary state $\neq n$ we shall have a transition to a state which is also $\neq n$, then clearly

$$\psi_n(k\tau)=W^{k-1}(Nn;Nn)[1-W(Nn;Nn)]. \qquad (371)$$

Substituting the foregoing expression for $\psi_n(k\tau)$ in Eq. (369) we obtain [cf. Eqs. (365) and (368)]

$$\Theta_n = \frac{\tau}{1 - W(Nn;Nn)}. \tag{372}$$

We shall now obtain a formula for $W(Nn;Nn)$. First of all it is clear that

$$1 - W(Nn;Nn) = W(Nn;n), \tag{373}$$

where $W(Nn;n)$ is the probability that from an arbitrary state $\neq n$ we shall have a transition to the state n. Now, under equilibrium conditions, the number of transitions from states $\neq n$ to the state n must equal the number of transisitions from the state n to states $\neq n$; accordingly

$$[1 - W(n)]W(Nn;n) = W(n)[1 - W(n;n)], \tag{374}$$

where $W(n)$ is given by Eq. (333). Hence,

$$W(Nn;n) = W(n)\frac{1 - W(n;n)}{1 - W(n)}. \tag{375}$$

Combining Eqs. (372), (373), and (375) we obtain

$$\Theta_n = \frac{\tau}{1 - W(n;n)} \frac{1 - W(n)}{W(n)}. \tag{376}$$

Finally, we may note that between T_n and Θ_n we have the relation

$$\Theta_n = T_n \frac{1 - W(n)}{W(n)}. \tag{377}$$

In the next section we shall give a brief account of the experiments of Svedberg and Westgren on colloid statistics which have provided complete confirmation of Smoluchowski's theory of density fluctuations which we have developed in this section. Also, the formulae for T_n and Θ_n which we have derived have important applications to the elucidation of the second law of thermodynamics to which we shall return in §4.

2. Experimental Verification of Smoluchowski's Theory: Colloid Statistics

In the experiments of Svedberg, Westgren, and others on colloid statistics observations are made by means of an ultramicroscope on the numbers of particles in a well-defined element of volume in a colloidal solution. These observations, made systematically at constant intervals τ apart, are secured either by the use of intermittent illumination (Svedberg) or by counting on the ticks of a metronome (Westgren). The volumes in which the counts are made are defined either optically by illuminating only plane parallel layers several microns in thickness (Svedberg) or mechanically by having the solution under observation sealed between the objective of the microscope and a glass plate and observing with the help of a cardioid condenser (Westgren). The dimensions of the element of volume at right angles to the line of sight are defined directly by limiting the field of observation (see Figs. 3 and 4).

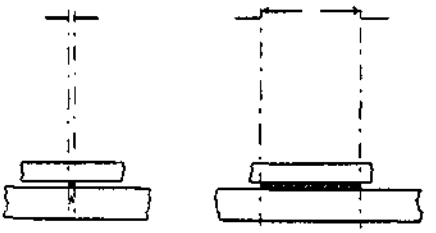


FIG. 3.

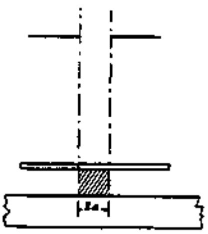


FIG. 4.

The colloidal particles describe Brownian motion and since the intervals of time we are normally interested in are never less than a few hundredths of a second we can suppose that the motions of the particles are governed by the diffusion equation [cf. Eqs. (133), (304), and (306)]

$$\partial w/\partial t = D\nabla^2 w; \qquad D = q/\beta^2 = kT/m\beta = kT/6\pi a\eta. \tag{378}$$

For, according to our discussion in Chapter II, §§2 and 4 the validity of the diffusion equation requires that we only ignore what happens in time intervals of order β^{-1} and for colloidal gold particles of radius $a=50\mu\mu$ this time of relaxation is of the order of 10^{-9}–10^{-10} second.

From Eq. (378) we conclude that the probability of occurrence of a particle at $\mathbf{r}_2$ at time t when it was at $\mathbf{r}_1$ at time $t=0$ is given by [cf. Eq. (172)]

$$\frac{1}{(4\pi Dt)^{\frac{3}{2}}}\exp\left(-|\mathbf{r}_2-\mathbf{r}_1|^2/4Dt\right). \qquad (379)$$

On this basis we can readily write down a general formula for the probability after-effect factor P introduced in §1. For, by definition, P denotes the probability that a particle somewhere inside the given element of volume v (with uniform probability) at time $t=0$ will find itself outside of it at time $t=\tau$. Accordingly

$$P=\frac{1}{(4\pi D\tau)^{\frac{3}{2}}v}\int\int\exp\left(-|\mathbf{r}_1-\mathbf{r}_2|^2/4D\tau\right)d\mathbf{r}_1d\mathbf{r}_2, \qquad (380)$$

where the integration over $\mathbf{r}_1$ is extended over all points in the interior of v while that over $\mathbf{r}_2$ is extended over all points exterior to v. Alternatively, we can also write

$$1-P=\frac{1}{(4\pi D\tau)^{\frac{3}{2}}v}\int_{\mathbf{r}_1\epsilon v}\int_{\mathbf{r}_2\epsilon v}\times\exp\left(-|\mathbf{r}_1-\mathbf{r}_2|^2/4D\tau\right)d\mathbf{r}_1d\mathbf{r}_2, \qquad (381)$$

where, now, the integrations over *both* $\mathbf{r}_1$ and $\mathbf{r}_2$ are extended over all points inside v (indicated by the symbols $\mathbf{r}_1\epsilon v$ and $\mathbf{r}_2\epsilon v$).

We thus see that for any geometrically well-defined element of volume in a colloidal solution we can always evaluate, in principle, the probability after-effect factor P in terms of the physical parameters of the problem, namely, the geometry of the volume v, the radius a of the colloidal particles, and the coefficient of viscosity η of the surrounding liquid. On the other hand, this factor P can also be determined empirically from a direct evaluation of the mean square of the differences in the numbers of particles observed on consecutive occasions in a long sequence of observations made at constant intervals τ apart and using the formula [Eq. (361)]

$$\langle\Delta^2\rangle_{\text{Av}}=2\nu P, \qquad (382)$$

where ν is the average of all the numbers observed. A comparison of the predictions of the theory with the data of colloid statistics therefore becomes possible. Once P has been determined [either theoretically according to Eq. (381) or empirically from Eq. (382)] we can predict the frequency of occurrence, $H(n, m)$, of the *pair* (n, m) in the observed sequence of numbers. For, clearly;

$$H(n, m)=W(n)W(n;m), \qquad (383)$$

where $W(n)$ is the frequency of occurrence of n according to Eq. (333) and $W(n;m)$ is the transition probability from the state n to the state m according to Smoluchowski's formulae (344) and (345). Again a comparison between the predictions of the theory with the results of observations becomes possible.

Comparisons of the kind indicated in the preceding paragraph were first made by Smoluchowski himself who used for this purpose the data provided by Svedberg's experiments. However, later experiments by Westgren carried out with the expressed intention of verifying Smoluchowski's theory provide a more stringent comparison between the predictions of the theory and the results of observations. We shall therefore limit ourselves to describing the results of Westgren's experiments only.

Westgren conducted two series of experiments with the arrangements shown in Figs. 3 and 4. In the first of the two arrangements (Fig. 3) the particles under observation are confined to a long rectangular parallelepiped (see the shaded portions in Fig. 3). Under the conditions of this arrangement it is clear that the variation in the number of particles observed is predominantly due to diffusion at right angles to the lengthwise edge. Consequently, the formula for P appropriate to this arrangement is [cf. Eq. (381)]

$$1-P=\frac{1}{h(4\pi D\tau)^{\frac{1}{2}}}\int_0^h\int_0^h\times\exp\left[-(x_1-x_2)^2/4D\tau\right]dx_1dx_2, \qquad (384)$$

where h denotes the width of the element of volume under observation (see Fig. 3). Introducing $2(D\tau)^{\frac{1}{2}}$ as the unit of length, Eq. (384) becomes

$$1-P=\frac{1}{\alpha\pi^{\frac{1}{2}}}\int_0^\alpha\int_0^\alpha \exp\left[-(\xi_1-\xi_2)^2\right]d\xi_1 d\xi_2, \quad (385)$$

where we have written

$$\alpha=h/2(D\tau)^{\frac{1}{2}}. \quad (386)$$

We readily verify that Eq. (385) is equivalent to

$$1-P=\frac{2}{\alpha\pi^{\frac{1}{2}}}\int_0^\alpha d\xi_1\int_0^{\xi_1} d\eta \exp(-\eta^2), \quad (387)$$

or, after an integration by parts we find

$$P=1-\frac{2}{\pi^{\frac{1}{2}}}\int_0^\alpha \exp(-\xi^2)d\xi + \frac{1}{\alpha\pi^{\frac{1}{2}}}\left[1-\exp(-\alpha^2)\right]. \quad (388)$$

For the second of Westgren's arrangements (Fig. 4) the element under observation is a cylindrical volume and the variations in the numbers observed are in this case due to the diffusion of particles in all directions at right angles to the line of sight. Accordingly we have

$$P=\frac{4}{\alpha^2\pi}\int_\alpha^\infty d\xi_1\xi_1\int_0^\alpha d\xi_2\xi_2\int_0^\pi \times\exp(-\xi_1^2-\xi_2^2+2\xi_1\xi_2\cos\vartheta)d\vartheta, \quad (389)$$

where

$$\alpha=r_0/2(D\tau)^{\frac{1}{2}}, \quad (390)$$

r_0 denoting the radius of the cylindrical element under observation. The integrals in (389) can be evaluated in terms of Bessel functions with imaginary arguments and we find

$$P=e^{-2\sqrt{\alpha}}\left[I_0(2\sqrt{\alpha})+I_1(2\sqrt{\alpha})\right]. \quad (391)^6$$

Westgren has made several series of counts with both of his experimental arrangements. We give below a sample extract from one of his sequences:

$$\begin{matrix} 2\ 1\ 1\ 1\ 1\ 1\ 0\ 2\ 2\ 1\ 1\ 1\ 2\ 3\ 2\ 3\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 1\ 2\ 2\ 3\ 3\ 4\ 5 \\ 3\ 4\ 2\ 2\ 1\ 2\ 1\ 3\ 2\ 0\ 2\ 2\ 1\ 0\ 2\ 2\ 2\ 1\ 2\ 3\ 2\ 2\ 2\ 3\ 2\ 2\ 2\ 2\ 2\ 2\ 1\ 3\ 3\ 4\ 2\ 2 \end{matrix} \quad (392)$$

The foregoing counts were obtained with the first of the two experimental arrangements described with the following values for the various physical parameters:

$$\begin{array}{ll} h=6.56\mu; & D=3.95\times10^{-8}; \\ \tau=1.39\text{ sec.}; & a=49.5\mu\mu; \\ T=290.0^\circ\text{K}; & \nu=1.428. \end{array} \quad (393)$$

First of all, it is of interest to see how well the Poisson distribution (333) represents the observed frequencies of occurrence of the different values of n. Table III shows this comparison for the sequence of which (392) is an extract. It is seen that the representation is satisfactory. Also, the observed mean square deviation for this sequence is 1.35 while the value theoretically predicted is ν which is 1.43; again the agreement is satisfactory.

Turning next to questions relating to probability after-effects we may first note that each of the observed sequences can be used for several comparisons. For, by suitably selecting from a given sequence of sufficient length we can derive others with intervals between consecutive observations which are integral multiples of that characterizing the original sequence. Thus, by considering only the alternate numbers we obtain a new sequence in which the interval τ between two observations is twice that in the original sequence.

As we have already remarked, for any given sequence, we can compute theoretical values of P in terms of the physical parameters of the problem according to Eq. (388) or (391) depending on the experimental arrangement used. For the same sequences, we can also, using Eq. (382), derive values of P from the observed counts

TABLE III. The Poisson distribution for $W(n)$. $\nu=1.428$.

$n=$	0	1	2	3	4	5	6	7
$W(n)_{obs}$	381	568	357	175	67	28	5	2
$W(n)_{calc}$	380	542	384	184	66	19	5	2

[6] The functions $e^{-x}I_{0,1}(x)$ are tabulated in Watson's *Bessel functions* (Cambridge, 1922), pp. 698–713.

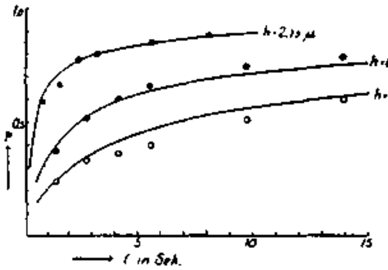


Fig. 5.

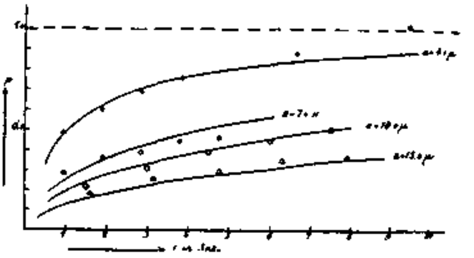


Fig. 6.

from the respective values of the mean square differences $(\Delta^2)_{Av}$. In Tables IV and V we have made, following Westgren, the comparison between the values P derived in this manner for two typical cases. The agreement is satisfactory. The confirmation of the theory is shown in a particularly striking manner in Figs. 5 and 6 where a comparison is made between the observed and the theoretical values of P in its dependence on τ for different values of h (or r_0).

It is now seen that an analysis of the data of colloid statistics actually provides us with a means of determining the Avogadro's constant N. For, from the mean square difference $(\Delta^2)_{Av}$ and the mean value of n (namely ν) we can determine P. On the other hand, according to Eqs. (380) and (381) P is determined by the geometry of the volume v *only, if* the unit of length is chosen to be $2(D\tau)^{\frac{1}{2}}$. Hence, from the empirically determined value of P we can deduce a value for this unit of length. In other words, a determination of the diffusion coefficient D is possible. But [cf. Eq. (378)]

$$D = kT/6\pi a\eta = (R/N)(T/6\pi a\eta), \qquad (394)$$

where R is the gas constant and N the Avogadro's number. Thus N can be determined. With the second of his two arrangements Westgren has used this method to determine N. As a mean of 50 determinations he finds $N = 6.09 \times 10^{23}$ with a probable error of 5 percent; this is in very satisfactory agreement with other independent determinations.

Turning next to the frequency of occurrence $H(n, m)$ of the pair of numbers (n, m) in a given sequence, we can predict this quantity according to Eq. (383); these predicted values can again be compared with those deduced directly from the counts. Such a comparison has also been made by Westgren whose results we give in Table VI.

Finally, we shall consider the experimental basis for the formulae (368) and (376) for the mean life and the average time of recurrence of a state of fluctuation. Using the counts of Svedberg, Smoluchowski has made a comparison between the values of T_n and Θ_n derived empirically from these counts and those predicted by Eqs. (368) and (376). The results of this comparison are shown in Table VII.

The long average times of recurrence for the states of large n are to be particularly noted (see §4 below). These long times are, however, a direct consequence of the "improbable"

Table IV. Comparison of the probability after-effect factor P derived from Eq. (388) and the experimental arrangement of Fig. 3 (Westgren). $h = 6.56\mu$; $a = 49.5\mu\mu$; $T = 290.0°K$; $D = 3.95 \times 10^{-9}$; $\nu = 1.428$.

τ(sec.)	$(\Delta^2)_{Av}$	P_{obs}	P_{calc}
1.39	1.068	0.374	0.394
2.78	1.452	0.513	0.517
4.17	1.699	0.600	0.587
5.56	1.859	0.656	0.634
9.73	2.125	0.744	0.713
13.90	2.265	0.793	0.760

Table V. Comparison of the probability after-effect factor P derived from Eq. (391) and the experimental arrangement of Fig. 4 (Westgren). $r_0 = 10.0\mu$; $a = 63.5\mu\mu$; $T = 290.1°K$; $D = 3.024 \times 10^{-9}$; $\nu = 1.933$.

τ(sec.)	$(\Delta^2)_{Av}$	P_{obs}	P_{calc}
1.50	0.836	0.217	0.238
3.00	1.200	0.310	0.332
4.50	1.512	0.391	0.401
6.00	1.718	0.444	0.456
7.50	1.939	0.502	0.503

TABLE VI. The observed and the theoretical frequencies of occurrence of the pairs (n, m) in a given sequence $(\nu=1.428;\ P=0.374)$. [In each case the top figure gives the observed number while the bottom figure (italicized), the number to be expected on the basis of Eq. (383).]

n	$m=0$	1	2	3	4	5	6
0	210	126	35	7	0	1	—
	221	*119*	*32*	*6*	*1*	—	—
1	134	281	117	29	1	1	—
	119	*262*	*122*	*31*	*5*	*1*	—
2	27	138	108	63	16	3	—
	32	*122*	*149*	*63*	*15*	*3*	—
3	10	20	76	38	24	6	0
	6	*31*	*63*	*56*	*22*	*5*	*1*
4	2	2	14	22	13	11	3
	1	*5*	*15*	*22*	*15*	*6*	*2*
5	—	0	2	10	10	1	3
	—	*1*	*3*	*5*	*6*	*3*	*1*

nature of these states. For, according to Eq. (376)

$$\Theta_n \sim (\tau/W(n)) = \tau(e^{\nu}n!/\nu^n) \quad (n \gg \nu). \quad (395)$$

which increases extremely rapidly for large values of n. For example, the number 7 was recorded only once in Svedberg's entire sequence of 518 counts; but the average time of recurrence for this state is 1105τ. Again, the number 17 (for instance) was never observed by Svedberg; and this is also understandable in view of the average time of recurrence for this state which is $\Theta_{17} \sim 10^{13}$. !

In concluding this discussion of the experimental verification of Smoluchowski's theory, we may remark on the inner relationships that have been disclosed to exist between the phenomena of Brownian motion, diffusion, and fluctuations in molecular concentration. But what is perhaps of even greater significance is that we have here the first example of a case in which it has been possible to follow in all its details, both theoretically and experimentally, the transition between the macroscopically irreversible nature of diffusion and the microscopically reversible nature of molecular fluctions. (These matters are further touched upon in §§3 and 4 below.)

3. Probability After-Effects for Continuous Observation

The theory of density fluctuations as developed in §1 is valid whenever the physical circumstances of the problem will permit us to introduce the probability after-effect factor P. It will be recalled that this factor $P(\tau)$ is defined as the probability that a particle, initially, somewhere inside a given element of volume will emerge from it before the elapse of a time τ. And, as we have seen in §1, we can express all the significant facts related to the phenomenon of the speed of fluctuations in terms of this single factor $P(\tau)$. But the theory as developed in §1 applies only when τ is finite, i.e., for the case of intermittent observations. We shall now show how this theory can be generalized to include the case of continuous observations.

First of all, it is clear that we should expect

$$P(\tau) \to 0 \quad \text{as} \quad \tau \to 0. \quad (396)$$

Hence, according to Eq. (364),

$$W(n;n) = e^{-\nu P}(1-P)^n + O(P^2) \qquad (\tau \to 0;\ P \to 0), \quad (397)$$

or

$$W(n;n) = 1-(n+\nu)P(\tau) + O(P^2) \qquad (\tau \to 0;\ P \to 0). \quad (398)$$

From this expression for $W(n;n)$ we can derive a formula for the probability $\phi_n(t)\Delta t$ that the state n will continue to be under observation for a time t and that during t and $t+\Delta t$ there will occur a transition to a state different from n. For this purpose, we divide the interval $(0, t)$ into a very large number of subintervals of duration Δt. Then, from the definition of $\phi_n(t)\Delta t$ it follows that

$$\phi_n(t)\Delta t = [W(n;n)]^{t/\Delta t}[1-W(n;n)], \quad (399)$$

or, using Eq. (398),

$$\phi_n(t)\Delta t = [1-(n+\nu)P(\Delta t) + O(P^2)]^{t/\Delta t} \times (n+\nu)P(\Delta t). \quad (400)$$

This last equation suggests that to obtain consistent results it would be necessary that

$$P(\Delta t) = O(\Delta t) \quad (\Delta t \to 0). \quad (401)$$

On general physical grounds, we may expect that this would in fact be the case. But it should not be concluded that Eq. (401) will be valid for *any* arbitrary idealization of the physical problem. For example, it is *not* true that $P(\Delta t)$ is $O(\Delta t)$ for the case of Brownian motions *idealised* as a problem in pure diffusion as we have done in §2. For, according to Eq. (379)

$$\langle|\Delta \mathbf{r}|^2\rangle_{Av}=6D\Delta t; \tag{402}$$

and hence, for P defined as in Eq. (380)

$$P=O[(\Delta t)^{\frac{1}{2}}] \quad (\Delta t\rightarrow 0), \tag{403}$$

contrary to Eq. (401). However, the reason for this disagreement is that the reduction of the problem of Brownian motions to one in diffusion can be achieved only when the intervals of time we are interested in are long compared to the time of relaxation β^{-1}. When this ceases to be the case, as in the present context, Eq. (379) is no longer true and we should strictly use the general distribution derived in Chapter II, §2 [see Eq. (171)]. And, according to Eq. (175)

$$\langle|\Delta \mathbf{r}|^2\rangle_{Av}=|\mathbf{u}_0|^2(\Delta t)^2, \quad (\Delta t\rightarrow 0). \tag{404}$$

On the basis of Eq. (404) we shall naturally be led to a formula for P consistent with (401) [see Eq. (413) below]. We shall therefore assume that

$$P(\Delta t)=P_0\Delta t+O(\Delta t^2) \quad (\Delta t\rightarrow 0), \tag{405}$$

where P_0 is a constant.

Combining Eqs. (400) and (405) we have

$$\phi_n(t)\Delta t=[1-(n+\nu)P_0\Delta t+O(\Delta t^2)]^{t/\Delta t} \times(n+\nu)P_0\Delta t, \tag{406}$$

or, passing to the limit $\Delta t=0$ we obtain

$$\phi_n(t)dt=\exp[-(n+\nu)P_0t](n+\nu)P_0dt. \tag{407}$$

Equation (407) expresses *a law of decay of a state of fluctuation* quite analogous to the law of decay of radioactive substances.

According to Eq. (407), the mean life, T_n, of the state n for continuous observation can be defined by

$$T_n=\int_0^\infty t\phi_n(t)dt; \tag{408}$$

in other words

$$T_n=1/(n+\nu)P_0. \tag{409}$$

Equation (409) is our present analogue of the formula (368) valid for intermittent observations.

Again, as in §1, we can also define the average time of recurrence of a state of fluctuation for continuous observation. This can be done by introducing the probability $W(Nn; Nn)$ and proceeding exactly as in the discussion of T_n. However, without going into details, it is evident that the relation (377) between T_n and Θ_n must continue to be valid, also for the case of continuous observation. Hence

$$\Theta_n=\frac{1}{(n+\nu)P_0}\,\frac{1-W(n)}{W(n)}. \tag{410}$$

We shall now derive for the case of Brownian motions, an explicit formula for P_0 which we formally introduced in Eq. (405). As we have already remarked, when dealing with continuous observation, the idealization of the phenomenon of Brownian motion as pure diffusion is not tenable. Instead, we should base our discussion on the exact distribution function $W(\mathbf{r},t;\mathbf{r}_0,\mathbf{u}_0)$ given by Eq. (171) and which is valid also for times of the order of the time of relaxation β^{-1}. However, since we are only interested in $P(\Delta t)$ for $\Delta t\rightarrow 0$ it would clearly be sufficient to consider the limiting form of the exact distribution $W(\mathbf{r},\mathbf{u},t;\mathbf{r}_0,\mathbf{u}_0)$ as $t\rightarrow 0$. On the other hand according to Eqs. (170)–(175) it follows that as $t\rightarrow 0$ we can regard the particles as describing linear trajectories with a Maxwellian distribution of the velocities. Hence, in our present context, $P(\Delta t)$ represents the probability that a particle initially inside a given element of volume v (with uniform probability) and with a velocity distribution governed by Maxwell's law will emerge from v before a time Δt. It is clear that formally, this is the same as the number of molecules striking the inner surface of the element of volume considered in a time Δt when the molecular concentration is $1/v$.

TABLE VII. The mean life T_n and the average time of recurrence Θ_n ($P=0.726$; $\nu=1.55$). (T_n and Θ_n are expressed in units of τ.)

n	T_n(obs.)	T_n(calc.)	Θ_n(obs.)	Θ_n(calc.)
0	1.67	1.47	6.08	5.54
1	1.50	1.55	3.13	3.16
2	1.37	1.38	4.11	4.05
3	1.25	1.23	7.85	8.07
4	1.23	1.12	18.6	20.9

Now, according to calculations familiar in the kinetic theory of gases, the number of molecules with velocities between $|\mathbf{u}|$ and $|\mathbf{u}|+d|\mathbf{u}|$ which strike unit area of any solid surface per unit time and in a direction with a solid angle $d\Omega$ at an angle ϑ with the normal to the surface is given by

$$N(m/2\pi kT)^{\frac{3}{2}} \exp(-m|\mathbf{u}|^2/2kT) \times |\mathbf{u}|^3 \cos\vartheta d\Omega d|\mathbf{u}|, \quad (411)$$

where N denotes the molecular concentration. Hence,

$$P(\Delta t)=\Delta t\frac{\sigma}{v}\left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}}\int_0^\infty\int_0^\pi \exp(-m|\mathbf{u}|^2/2kT) \times |\mathbf{u}|^3 \cos\vartheta d\Omega d|\mathbf{u}|, \quad (412)$$

where σ is the total surface area of the element of volume v. On evaluating the integrals in Eq. (412) we find that

$$P(\Delta t)=(\sigma/v)(kT/2\pi m)^{\frac{1}{2}}\Delta t. \quad (413)$$

Comparing this with Eq. (405) we conclude that for the case under consideration

$$P_0=(\sigma/v)(kT/2\pi m)^{\frac{1}{2}}. \quad (414)$$

The formulae (409) and (410) for the mean life and the average time of recurrence now take the forms

$$T_n=(v/\sigma(n+\nu))(2\pi m/kT)^{\frac{1}{2}}, \quad (415)$$

and

$$\Theta_n=(v/\sigma(n+\nu))(2\pi m/kT)^{\frac{1}{2}} \times([1-W(n)]/W(n)). \quad (416)$$

The case of greatest interest arises when the average number of particles, ν, contained in v is a very large number and the values of n considered are relatively close to ν. Then, the Poisson distribution $W(n)$ simplifies to (see Appendix III)

$$W(n)=[1/(2\pi\nu)^{\frac{1}{2}}]\exp[-(n-\nu)^2/2\nu]. \quad (417)$$

On this approximation, Eq. (416) becomes

$$\Theta_n\simeq\pi\frac{v}{\sigma}\left(\frac{m}{\nu kT}\right)^{\frac{1}{2}}\exp[(n-\nu)^2/2\nu]. \quad (418)$$

As an illustration of Eq. (418) we shall consider, following Smoluchowski, the average time of recurrence of a state of fluctuation in which the molecular concentration of oxygen in a sphere of air of radius a will differ from the average value by 1 percent. Table VIII gives Θ_n for different values of a.

TABLE VIII. The average time of recurrence of a state of fluctuation in which the molecular concentration in a sphere of air of radius a will differ from the average value by 1 percent. $T=300°\mathrm{K}$; $\nu=3\times10^{19}\times(4\pi a^3/3)$.

a(cm)	1	5×10^{-5}	3×10^{-5}	2.5×10^{-5}	1×10^{-5}
Θ(sec.)	$10^{10^{14}}$	10^{68}	10^{6}	1	10^{-11}

It is seen from Table VIII that under normal conditions, for volumes which are on the edge of visual perception even appreciable fluctuations in the molecular concentrations require such colossal average times of recurrence, that for all practical purposes the phenomenon of diffusion can be regarded as an irreversible process. On the other hand, for volumes which are just on the limit of microscopic vision, fluctuations in concentrations occur to such an extent and with such frequency that there can no longer be any question of irreversibility: under such conditions the notion of diffusion very largely loses its common meaning. For example, it would scarcely occur to one to illustrate the phenomenon of diffusion by the experiments of Svedberg and Westgren on colloid statistics though it is in fact true that *on the average* the results are in perfect accord with the principles of macroscopic diffusion [as is illustrated, for example, by Eq. (356) for $\langle\Delta_n\rangle_{\mathrm{Av}}$]. We shall return to these questions in the following section.

4. On the Reversibility of Thermodynamically Irreversible Processes, the Recurrence of Improbable States, and the Limits of Validity of the Second Law of Thermodynamics

If we formulate the second law of thermodynamics in any of its conventional forms, as, for example, that "heat cannot of itself be transferred from a colder to a hotter body" or, that "arbitrarily near to any given state there exist states which are inaccessible to the initial state by adiabatic processes" (Caratheodory), or that "the entropy of a closed system must never decrease," we, at once, get into contradiction with the kinetic molecular theory which demands the essential reversibility of all processes. Consequently, from the side of "dogmatic" thermodynamics two principal objections have been raised in the form of paradoxes and which are held to vitiate the entire outlook of the kinetic theory and statistical mechanics. We first state the two paradoxes.

(i) Loschmidt's Reversibility Paradox

Loschmidt first drew attention to the fact that in view of the essential symmetry of the laws of mechanics to the past and the future, all molecular processes must be reversible from the point of view of statistical mechanics. This is in apparent contradiction with the point of view held in thermodynamics that certain processes are irreversible.

(ii) Zermelo's Recurrence Paradox

There is a theorem in dynamics due to Poincaré which states that *in a system of material particles under the influence of forces which depend only on the spatial coordinates, a given initial state*[7] *must, in general, recur, not exactly, but to any desired degree of accuracy, infinitely often, provided the system always remains in the finite part of the phase space.* (For a proof of this theorem see Appendix V.) In other words, the trajectory described by the representative point in the phase space has a "quasi-periodic" character in the sense that after a finite interval of time (which can be specified) the system will return to the initial state to any desired degree of accuracy. Basing on this theorem of Poincaré, Zermelo has argued that the notion of irreversibility fundamental to macroscopic thermodynamics is incompatible with the standpoint of the kinetic theory.

As is well known, Boltzmann has tried to resolve these paradoxes of Loschmidt and Zermelo by probability considerations of a general nature. Thus, on the strength of certain rough estimates (see Appendix VI), Boltzmann concludes that the period of one of Poincaré's cycles is so enormously long, even for a cubic centimeter of gas, that the recurrence of an initially improbable state (i.e., the reversal to a state of lower entropy) while not strictly impossible, is yet so highly improbable that during the times normally available for observation, the chance of witnessing a thermodynamically irreversible process is *extremely* small.

Though Boltzmann's arguments and conclusions are fundamentally sound there are certain unsatisfactory features in basing on the period of a Poincaré cycle. For one thing, the period of such a cycle depends on how *nearly* we (arbitrarily) require the initial state to recur. Again, Poincaré's theorem refers to the return of the representative point in the $6N$-dimensional phase space (N denoting the number of particles in the system). Actually, in practice, we should treat two states of a gas as macroscopically distinct only if the numbers of molecules (considered indistinguishable) in the various limits of positions and velocities are different. Then, during a Poincaré cycle, the different macroscopically distinguishable states of the system will approximately recur a great many times. These recurrences of the different macroscopically distinct states, during a given Poincaré cycle, will be distributed very unequally among the states: thus, most of the recurrences will occur for the states of the system which are very close to what would be described as the thermodynamically "*normal state*." Moreover, it can also happen that during such a cycle, states deviating by arbitrarily large amounts from the normal state are assumed by the system. In other words, during a Poincare cycle we shall pass through many improbable states and indeed with equal frequency both in the directions of increasing and decreasing entropy.

Thus, while we may accept Boltzmann's point of view as fundamentally correct, it would clearly add to our understanding of the whole problem if we can explicitly demonstrate in a given instance how in spite of the essential reversibility of all molecular phenomena, we nevertheless get the impression of irreversibility.

Now, as we have already remarked in the preceding sections, Smoluchowski's theory of fluctuations in molecular concentrations allows us to bridge the gap between the regions of the

[7] This is defined by the positions and the velocities of all the particles, i.e., by the representative point in the phase space.

macroscopically irreversible diffusion and the microscopically reversible fluctuations. Consequently, a further discussion of this problem will enable us to follow explicitly how in this particular instance the Loschmidt and the Zermelo paradoxes resolve themselves.

(a) *The resolution of Loschmidt's paradox.*—Using Eqs. (333), (344), and (345) we readily verify that

$$\begin{aligned} H(n,\, n+k) &= W(n)\,W(n\,;\, n+k) \\ &= W(n+k)\,W(n+k\,;\, n) = H(n+k,\, n). \qquad (419) \end{aligned}$$

The quantity on the left-hand side in the foregoing equation represents the frequency of occurrence of the numbers n and $n+k$ on two successive occasions in a long sequence of observations; similarly, the quantity on the right-hand side gives the frequency of occurrence of the pair $(n+k,\, n)$. It therefore follows that under equilibrium conditions, the probability, that in a given length of time we observe a transition from the state n to the state m is equal to the probability that (in an equal length of time) we observe a transition from the state m to the state n. It is precisely the symmetry between the past and the future which guarantees this equality between $H(n\,;\, m)$ and $H(m\,;\, n)$. A glance at Table VI shows that this is amply confirmed by observations. [It may be further noted that, in accordance with Eq. (419) the numbers in italics on the opposite sides of the principal diagonal are equal.] All this, is, of course, in entire agreement with Loschmidt's requirements.

On the other hand, it is also evident from Table VI, that after a relatively large number like 5, 6, or 7 a number much smaller, generally follows; in other words, the probability that a number $n(\gg\nu)$ will further increase on the next observation is very small indeed. This circumstance illustrates how molecular concentrations differing appreciably from the average value will *almost* always tend to change in the direction indicated on the macroscopic notions concerning diffusion [cf. Eq. (356)]. This corresponds exactly to one of Boltzmann's statements that the negative entropy curve almost always decreases from any point. However this may be, in course of time, an abnormal initial state will again recur as a consequence of fluctuations, and we shall now see how in spite of this possibility for recurrence, the *apparently* irreversible nature of the phenomenon comes into being.

(b) *The resolution of Zermelo's paradox.*—Let us first consider the case of intermittent observations. As we have already remarked in §2, the number 17 never occurred in one of Svedberg's sequences for which ν had the value 1.55. But the average time of recurrence for this state [according to Eq. (376)] is $10^{13}\tau$; and since $\tau=1/39$ min., for the sequence considered, $\Theta\sim 500{,}000$ years. Hence, the diffusion from the state $n=17$ will have all the *appearances* of an irreversible process simply because the average time of recurrence is so very long compared to the times during which the system is under observation.

Turning next to the case of continuous observations, we shall return to the example considered in §3. As we have seen (cf. Table VIII) the average time of recurrence of a state in which the number of molecules of oxygen contained in a sphere of radius $a \geqslant 5\times 10^{-5}$ cm (and $T=300°$K and $\nu=3\times 10^{19}$ cm^{-3}) will differ from the average value by 1 percent is very long indeed ($\Theta>10^{68}$ seconds). The factor which is principally responsible for these large values for Θ is the exponential factor in Eq. (418). Accordingly, we may say, very roughly, that *the second law of thermodynamics is valid only for those diffusion processes in which the equalization of molecular concentrations which take place are by amounts appreciably greater than the root mean square relative fluctuation* (namely, $[\langle|n-\nu|^2\rangle_{Av}/\nu^2]^{\frac{1}{2}}=\nu^{-\frac{1}{2}}$). We have thus completely reconciled (at any rate, for the processes under discussion) the notion of irreversibility which is at the base of thermodynamics and the essential reversibility of all molecular phenomena demanded by statistical mechanics. This reconciliation has become possible only because we have been able to specify the limits of validity of the second law.

Quite generally, we may conclude with Smoluchowski that *a process appears irreversible (or reversible) according as whether the initial state is characterized by a long (or short) average time of recurrence compared to the times during which the system is under observation.*

5. The Effect of Gravity on the Brownian Motion: The Phenomenon of Sedimentation

The study of the effect of gravity on the Brownian motion provides an interesting illustration of the use to which Smoluchowski's equation [Eq. (312)]

$$(\partial w/\partial t) = \mathrm{div}_r\,(q\beta^{-2}\,\mathrm{grad}_r\,w - \mathbf{K}\beta^{-1}w) \tag{420}$$

can be put. In Eq. (420) $\mathbf{K}$ represents the acceleration caused by the external field of force. If the external field is that due to gravity, we can write

$$K_x = 0;\quad K_y = 0;\quad K_z = -(1-(\rho_0/\rho))g, \tag{421}$$

provided the coordinate system has been so chosen that the z axis is in the vertical direction. In Eq. (421), g denotes the value of gravity, ρ the density of the Brownian particle and $\rho_0(\leqslant\rho)$ that of the surrounding fluid. Hence, for the case (421), Eq. (420) becomes

$$(\partial w/\partial t) = (q/\beta^2)\nabla^2 w + (1-(\rho_0/\rho))(g/\beta)(\partial w/\partial z). \tag{422}$$

It is seen that Eq. (422) is of the same general form as Eq. (126). Accordingly, we can interpret the phenomenon described by Eq. (422) as a process of diffusion in which the number of particles crossing elements of area normal to x, y, and z directions, per unit area and per unit time, are given, respectively, by [cf. Eq. (127)]

$$-D(\partial w/\partial x),\quad -D(\partial w/\partial y), \tag{423}$$

and

$$-D(\partial w/\partial z) - cw, \tag{424}$$

where

$$D = (q/\beta^2) = (kT/m\beta);\quad c = (1-(\rho_0/\rho))(g/\beta). \tag{425}$$

Thus, while the diffusion in the (x, y) plane takes exactly as in the field free case, the situation in the z direction is modified. If we, therefore, limit ourselves to considering only the distribution in the z direction, of particles uniformly distributed in the (x, y) plane, the appropriate differential equation is

$$\frac{\partial w}{\partial t} = D\frac{\partial^2 w}{\partial z^2} + c\frac{\partial w}{\partial z}. \tag{426}$$

Let us now suppose that the particle is initially at a height z_0 measured from the bottom of the vessel containing the solution. Then, the probability of occurrence of the various values of z at later times will be governed by the solution of Eq. (426) which satisfies the boundary conditions

$$\left.\begin{aligned} &w \to \delta(z-z_0) \quad \text{as} \quad t\to 0,\\ &D(\partial w/\partial z) + cw = 0 \quad \text{at} \quad z=0 \quad \text{for all} \quad t>0. \end{aligned}\right\} \tag{427}$$

The second of two foregoing boundary conditions arises from the requirement that no particle shall cross the plane $z=0$ representing the bottom of the vessel [cf. Eq. (424)].

To obtain the solution of Eq. (426) satisfying the boundary conditions (427), we first introduce the following transformation of the variable [cf. Eq. (128)]

$$w = U(z, t)\exp\left[-\frac{c}{2D}(z-z_0) - \frac{c^2}{4D}t\right]. \tag{428}$$

Equation (426) reduces to the standard form

$$(\partial U/\partial t) = D(\partial^2 U/\partial z^2) \tag{429}$$

while the boundary conditions (427) become

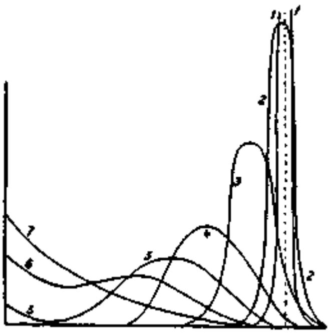


FIG. 7.

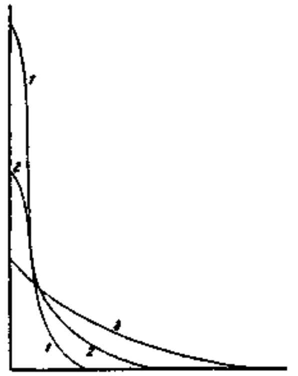


FIG. 8.

$$\left.\begin{aligned} &U \to \delta(z-z_0) \quad \text{as} \quad t \to 0, \\ &D(\partial U/\partial z) + (1/2)cU = 0 \quad \text{at} \quad z=0 \quad \text{for all} \quad t>0. \end{aligned}\right\} \tag{430}$$

Solving Eq. (429) with boundary conditions of the form (430) is a standard problem in the theory of heat conduction. We have

$$U = \frac{1}{2(\pi Dt)^{\frac{1}{2}}}\{\exp[-(z-z_0)^2/4Dt] + \exp[-(z+z_0)^2/4Dt]\} + \frac{c}{2D(\pi Dt)^{\frac{1}{2}}}\int_{z_0}^{\infty}\exp\left[-\frac{(\alpha+z)^2}{4Dt} + \frac{c(\alpha-z_0)}{2D}\right]d\alpha. \tag{431}$$

After some elementary transformations, Eq. (431) takes the form

$$U = \frac{1}{2(\pi Dt)^{\frac{1}{2}}}\{\exp[-(z-z_0)^2/4Dt] + \exp[-(z+z_0)^2/4Dt]\} + \frac{c}{D\sqrt{\pi}}\exp\left[\frac{c^2t}{4D} - \frac{c(z+z_0)}{2D}\right]\int_{\frac{z+z_0-ct}{2(Dt)^{\frac{1}{2}}}}^{\infty}\exp(-x^2)dx. \tag{432}$$

Returning to the variable w we have [cf. Eq. (428)]

$$w(t, z; z_0) = \frac{1}{2(\pi Dt)^{\frac{1}{2}}}\{\exp[-(z-z_0)^2/4Dt] + \exp[-(z+z_0)^2/4Dt]\} \times\exp\left[-\frac{c}{2D}(z-z_0) - \frac{c^2}{4D}t\right] + \frac{c}{D\sqrt{\pi}}e^{-cz/D}\int_{\frac{z+z_0-ct}{2(Dt)^{\frac{1}{2}}}}^{\infty}\exp(-x^2)dx \tag{433}$$

which is the required solution. In Fig. 7 we have illustrated according to Eq. (433) the distributions $w(z, t; z_0)$ for a given value of z_0 and various values of t.

If we suppose that at time $t=0$ we have a large number of particles distributed uniformly in the plane $z=z_0$ then in the first instance diffusion takes place as in the field free case (curves 1 and 2). However, gravity makes itself felt very soon (curves 3, 4, and 5) and the maximum begins to be displaced to lower values of z with the velocity c; at the same time, the maximum becomes flatter on account of the random motions experienced by the particles. Once the probability of finding particles near enough to the bottom of the vessel becomes appreciable, the curves again begin to rise upwards (curves 5 and 6) on account of the reflection which the particles suffer at $z=0$; and, finally as $t\to\infty$ we obtain the equilibrium distribution

$$w(z,\ \infty\,;z_0)=(c/D)e^{-cz/D}. \tag{434}$$

Since [cf. Eq. (425)]

$$(c/D)=(1-(\rho_0/\rho))(mg/kT), \tag{435}$$

we see that the equilibrium distribution (434) represents simply the law of isothermal atmospheres in its standard form.

The example we have just considered provides a further illustration of a case to which the conventional notions concerning entropy and the second law of thermodynamics cannot be applied. For the state of maximum entropy for the system consisting of the Brownian particles and the surrounding fluid, is that in which all the particles are at $z=0$; and, on strict thermodynamical principles we should conclude that with the continued operation of dissipative forces like dynamical friction, the state of maximum entropy will be attained. But according to Eq. (434), as $t\to\infty$ though the state of maximum entropy $z=0$ has the maximum probability, it is *not* true that the average value of the height at which the particles will be found is also zero. Actually, for the equilibrium distribution (434), we have

$$\langle z\rangle_{\text{Av}}=(D/c)=(kT/mg)[\rho/(\rho-\rho_0)], \tag{436}$$

which is the height of the equivalent homogeneous atmosphere. Moreover, even if the particles were initially at $z=0$, they will not continue to stay there. For, setting $z_0=0$ in Eq. (433) we find that

$$w(z,\,t;\,0)=(1/(\pi Dt)^{\frac{1}{2}})\exp\left[-(z+ct)^2/4Dt\right]+(c/D\sqrt{\pi})e^{-cz/D}\int_{\frac{z-ct}{2(Dt)^{\frac{1}{2}}}}^{\infty}\exp\left(-x^2\right)dx. \tag{437}$$

Equation (437) shows that as $t\to\infty$ we are again led to the equilibrium distribution (434) (see Fig. 8). Hence, the particles do a certain amount of mechanical work *at the expense of the internal energy of the surrounding fluid;* this is of course contrary to the strict interpretation of the second law of thermodynamics. The average work done in this manner is given by [if we use Eq. (436)]

$$\langle A\rangle_{\text{Av}}=m(1-(\rho_0/\rho))g\bar{z}=kT, \tag{438}$$

per particle. Hence, on the average there is a *decrease* in entropy of amount k per particle:

$$\langle S\rangle_{\text{Av}}=S_{\max}-Nk, \tag{439}$$

where N denotes the number of Brownian particles. However, as Smoluchowski has pointed out, this work done at the expense of the internal energy of the surrounding fluid cannot be utilized to run a heat engine with an efficiency higher than that of the Carnot cycle.

We may further note that except for values of $z\lesssim D/c$, a particle has a greater probability to descend than it has to ascend. As $z\to 0$ the converse is true. We may therefore say that the tendency for the entropy to *increase* (almost always) for particles at $z\gg D/c$ is compensated by the tendency of the entropy to *decrease* for particles very near $z=0$; so that, on the average, a steady state is maintained. Of course, we have a finite probability for particles, occasionally to ascend to very great heights; but in accordance with the conclusions of §4 we should expect that the average time of recurrence for such abnormal states must be very long indeed.

6. The Theory of Coagulation in Colloids

Smoluchowski discovered a very interesting application of the theory of Brownian motion in the phenomenon of coagulation exhibited by colloidal particles when an electrolyte is added to the solution. Smoluchowski's theory of this phenomenon is based on a suggestion of Zsigmondy that coagulation results as a consequence of each colloidal particle being surrounded (on the addition of an electrolyte) by a *sphere of influence* of a certain radius R such that the Brownian motion of a particle proceeds unaffected only so long as no other particle comes within its sphere of influence and that when the particles do come within a distance R they stick to one another to form a single unit. We are not concerned here with the physico-chemical basis for Zsigmondy's suggestion except perhaps to remark that the spheres of influence are supposed to originate in the formation of electric double layers around each particle; we are here interested only in the application of the principles of Brownian motion which is possible on the acceptance of Zsigmondy's suggestion. However, we may formulate somewhat more explicitly the problem we wish to investigate:

We imagine that initially the colloidal solution contains only single particles all similar to one another and of the same spherical size. We now suppose that at time $t=0$ an (appropriate) electrolyte is added to the solution in such a way that the resulting electrolytic concentration is uniform throughout the solution. The particles are now supposed to be all instantaneously surrounded by spheres of influence of radius R. From this instant onwards, each particle will continue to describe the original Brownian motion only so long as no other particle comes within its sphere of influence. Once two particles do approach to within this distance R they will coalesce to form a "*double particle*." This double particle will also describe Brownian motion but at a reduced rate consequent to its increased size. This double particle will, in turn, continue to remain as such only so long as it does not come within the appropriate spheres of influence of a single or another double particle: when this happens we shall have the formation of a triple or a quadruple particle; and, so on. The continuation of this process will eventually lead to the total coagulation of all the colloidal particles into one single mass.

The problem we wish to solve is the specification of the concentrations $\nu_1, \nu_2, \nu_3, \nu_4, \cdots$, of single, double, triple, quadruple, etc., particles at time t given that at time $t=0$ there were $\nu_0(=\nu_1[0])$ single particles.

As a preliminary to the discussion of the general problem formulated in the preceding paragraph we shall first consider the following more elementary situation:

A particle, assumed fixed in space, is in a medium of infinite extent in which a number of similar Brownian particles are distributed uniformly at time $t=0$. Further, if the stationary particle is assumed to be surrounded by a sphere of influence of radius R what is the rate at which particles arrive on the sphere of radius R surrounding the fixed particle?

We shall suppose that the stationary particle is at the origin of our system of coordinates. Then, in accordance with our definition of a sphere of influence, we can replace the surface $|\mathbf{r}|=R$ by a perfect absorber [cf. I, §5, see particularly Eq. (115)]. We have therefore to seek a solution of the diffusion equation [cf. Eqs. (173) and (306)]

$$(\partial w/\partial t)=D\nabla^2 w; \quad D=(q/\beta^2)=(kT/6\pi a\eta), \tag{440}$$

which satisfies the boundary conditions

$$\left.\begin{array}{l} w=\nu=\text{constant, at} \quad t=0, \quad \text{for} \quad |\mathbf{r}|>R, \\ w=0 \quad \text{at} \quad |\mathbf{r}|=R \quad \text{for} \quad t>0. \end{array}\right\} \tag{441}$$

In the first of the two foregoing boundary conditions ν denotes the average concentration of the particles exterior to $|\mathbf{r}|=R$ at time $t=0$.

Since w can depend only on the distance r from the center, the form of the diffusion equation (440) appropriate to this case is

$$(\partial/\partial t)(rw)=D(\partial^2/\partial r^2)(rw). \tag{442}$$

The solution of this equation satisfying the boundary conditions (441) is

$$w=\nu\left[1-\frac{R}{r}+\frac{2R}{r\sqrt{\pi}}\int_0^{(r-R)/2(Dt)^{\frac{1}{2}}}\exp(-x^2)dx\right]. \tag{443}$$

From Eq. (443) it follows that the rate at which particles arrive at the surface $|r|=R$ is given by [cf. Eq. (117)]

$$4\pi D\left(r^2\frac{\partial w}{\partial r}\right)_{r=R}=4\pi DR\nu\left(1+\frac{R}{(\pi Dt)^{\frac{1}{2}}}\right). \tag{444}$$

Equation (444) gives the rate at which particles describing Brownian motion will coalesce with a stationary particle surrounded by a sphere of influence of radius R. Suppose, now, that the particle we have assumed to be stationary is also describing Brownian motion. What is the corresponding generalization of (444)? In considering this generalization we shall not suppose that the diffusion coefficients characterizing the two particles which coalesce to form a multiple particle are necessarily the same. Under these circumstances we have clearly to deal with the *relative displacements* of the two particles; and it can be readily shown that the relative displacements between two particles describing Brownian motions independently of each other and with the diffusion coefficients D_1 and D_2 also follows the laws of Brownian motion with the diffusion coefficient $D_{12}=D_1+D_2$. For, the probability that the relative displacement of two particles, initially, together at $t=0$, lies between r and $r+dr$ is clearly

$$\left.\begin{aligned} W(r)dr&=dr\int_{-\infty}^{+\infty}W_1(r_1)W_2(r_1+r)dr_1\\ &=\frac{dr}{(4\pi D_1t)^{\frac{3}{2}}(4\pi D_2t)^{\frac{3}{2}}}\int_{-\infty}^{+\infty}\exp(-|r_1|^2/4D_1t)\exp(-|r_1+r|^2/4D_2t)dr_1 \end{aligned}\right\} \tag{445}$$

or, as may be readily verified [cf. the remarks following Eq. (62)]

$$W(r)=(1/[4\pi(D_1+D_2)t]^{\frac{3}{2}})\exp(-|r|^2/4(D_1+D_2)t). \tag{446}$$

On comparing this distribution of the relative displacements with the corresponding result for the individual displacements [see for example Eq. (172)] we conclude that the relative displacements do follow the laws of Brownian motion with the diffusion coefficient (D_1+D_2).

Thus, the required generalization of Eq. (444) is

$$4\pi(D_1+D_2)R\nu\left(1+\frac{R}{[\pi(D_1+D_2)t]^{\frac{1}{2}}}\right). \tag{447}$$

More generally, let us consider two sorts of particles with concentrations ν_i and ν_k. Let the respective diffusion coefficients be D_i and D_k. Further, let R_{ik} denote the distance to which two particles (one of each sort) must approach in order that they may coalesce to form a multiple particle. Then, the rate of formation of the multiple particles by the coagulation of the particles of the kind considered is clearly given by

$$J_{i+k}dt=4\pi D_{ik}R_{ik}\nu_i\nu_k\left(1+\frac{R_{ik}}{(\pi D_{ik}t)^{\frac{1}{2}}}\right)dt \tag{448}$$

where we have written

$$D_{ik}=D_i+D_k. \tag{449}$$

In our further discussions, we shall ignore the second term in the parenthesis on the right-hand side of Eq. (447); this implies that we restrict ourselves to time intervals $\Delta t\gg R^2/D$. In most cases of practical interest, this is justifiable as $R^2/D \sim 10^{-8}-10^{-4}$ second. With this understanding we can write

$$J_{i+k}dt \simeq 4\pi D_{ik}R_{ik}\nu_i\nu_k dt. \tag{450}$$

Using Eq. (450) we can now write down the fundamental differential equations which govern the variations of $\nu_1, \nu_2, \cdots, \nu_k, \cdots$ (of single, double, $\cdots$, k-fold, $\cdots$,) particles with time:

Thus, considering the variation of the number of k-fold particles with time, we have in analogy with the equations of chemical kinetics

$$\frac{d\nu_k}{dt} = 4\pi\left(\tfrac{1}{2}\sum_{i+j=k}\nu_i\nu_j D_{ij}R_{ij} - \nu_k\sum_{j=1}^{\infty}\nu_j D_{kj}R_{kj}\right) \quad (k=1, \cdots). \tag{451}$$

In this equation the first summation on the right-hand side represents the increase in ν_k due to the formation of k-fold particles by the coalescing of an i-fold and a j-fold particle (with $i+j=k$), while the second summation represents the decrease in ν_k due to the formation of $(k+j)$-fold particles in which one of the interacting particles is k-fold.

A general solution of the infinite system of Eq. (451) which will be valid under all circumstances does not seem feasible. But a special case considered by Smoluchowski appears sufficiently illustrative of the general solution.

First, concerning R_{ik}, the assumption is made that

$$R_{ik} = \tfrac{1}{2}(R_i + R_k), \tag{452}$$

where R_i and R_k are the radii of the spheres of influence of the i-fold and the k-fold particles. We can, if we choose, regard the assumption (452) as equivalent to Zsigmondy's suggestion concerning the basic cause of coagulation.

Again, according to Eq. (440), the diffusion coefficient is inversely proportional to the radius of the particle; and on the basis of experimental evidence it appears that the radii of the spheres of influence of various multiple particles are proportional to the radii of the respective particles. We therefore make the additional assumption that

$$D_iR_i = DR \quad (i=1, \cdots), \tag{453}$$

where D and R denote, respectively, the diffusion coefficient and the radius of the sphere of influence of the single particles.

Combining Eqs. (449), (452), and (453) we have

$$D_{ik}R_{ik} = \tfrac{1}{2}(D_i+D_k)(R_i+R_k) = \tfrac{1}{2}DR(R_i^{-1}+R_k^{-1})(R_i+R_k) = \tfrac{1}{2}DR(R_i+R_k)^2R_i^{-1}R_k^{-1}. \tag{454}$$

Finally, for the sake of mathematical simplicity we make the (not very plausible) assumption that

$$R_i = R_k. \tag{455}$$

Thus, with all these assumptions

$$D_{ik}R_{ik} = 2DR. \tag{456}$$

In view of (456), Eq. (451) becomes

$$\frac{d\nu_k}{dt} = 8\pi DR\left(\tfrac{1}{2}\sum_{i+j=k}\nu_i\nu_j - \nu_k\sum_{j=1}^{\infty}\nu_j\right) \quad (k=1, \cdots). \tag{457}$$

If we now let

$$\tau = 4\pi DRt, \tag{458}$$

Eq. (457) takes the more convenient form

$$\frac{d\nu_k}{d\tau} = \sum_{i+j=k}\nu_i\nu_j - 2\nu_k\sum_{j=1}^{\infty}\nu_j \quad (k=1, \cdots). \tag{459}$$

From Eq. (459) we readily find that

$$\left.\begin{aligned}\frac{d}{dt}\Big(\sum_{k=1}^{\infty}\nu_k\Big)&=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\nu_i\nu_j-2\sum_{k=1}^{\infty}\sum_{j=1}^{\infty}\nu_k\nu_j,\\ &=-\Big(\sum_{k=1}^{\infty}\nu_k\Big)^2,\end{aligned}\right\}\tag{460}$$

or,

$$\sum_{k=1}^{\infty}\nu_k=\frac{\nu_0}{1+\nu_0\tau},\tag{461}$$

remembering that at $t=0$, $\sum\nu_k=\nu_0$.

Using the integral (461) we can successively obtain the solutions for ν_1, ν_2, etc. Thus, considering the equation for ν_1 we have [cf. Eq. (459)]

$$d\nu_1/dt=-2\nu_1\sum_{k=1}^{\infty}\nu_k=-2\nu_1\nu_0/(1+\nu_0\tau);\tag{462}$$

in other words,

$$\nu_1=\frac{\nu_0}{(1+\nu_0\tau)^2},\tag{463}$$

again using the boundary condition that $\nu_1=\nu_0$ at $t=0$. Proceeding in this manner we can prove (by induction) that

$$\nu_k=\nu_0[(\nu_0\tau)^{k-1}/(1+\nu_0\tau)^{k+1}]\quad(k=1,2,\cdots).\tag{464}$$

In Fig. 9 we have illustrated the variations of $\sum\nu_k$, ν_1, ν_2, $\cdots$ with time. We shall not go into the details of the comparison of the predictions of this theory with the data of observations. Such comparisons have been made by Zsigmondy and others and the general conclusion is that Smoluchowski's theory gives a fairly satisfactory account of the broad features of the coagulation phenomenon.

7. The Escape of Particles over Potential Barriers

As a final illustration of the application of the principles of Brownian motion we shall consider, following Kramers, the problem of the escape of particles over potential barriers. The solution to this problem has important bearings on a variety of physical, chemical, and astronomical problems.

The situation we have in view is the following:

Limiting ourselves for the sake of simplicity to a one-dimensional problem, we consider a particle moving in a potential field $\mathfrak{V}(x)$ of the type shown in Fig. 10; more generally, we may consider an

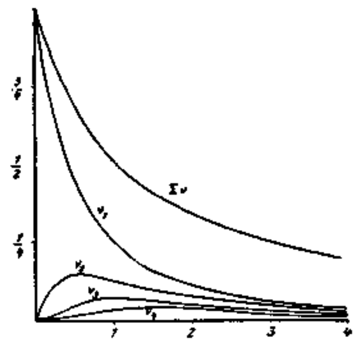


FIG. 9.

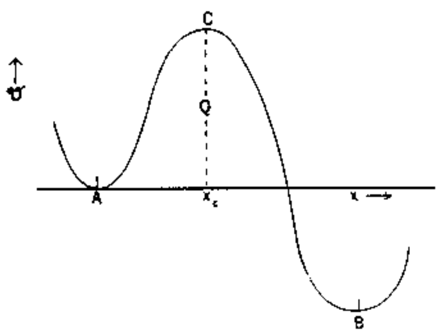


FIG. 10.

ensemble of particles moving in the potential field $\mathfrak{B}(x)$ without any mutual interference. We suppose that the particles are initially caught in the potential hole at A. The general problem we wish to solve concerns the rate at which particles will escape over the potential barrier in consequence of Brownian motion.

In the most general form, the solution to the problem formulated in the foregoing paragraph is likely to be beset with considerable difficulties. But a special case of interest arises when the height of the potential barrier is large compared to the energy of the thermal motions:

$$mQ \gg kT. \tag{465}$$

Under these circumstances, the problem can be treated as one in which the conditions are *quasi-stationary*. More specifically, we may suppose that to a high degree of accuracy a Maxwell-Boltzmann distribution obtains in the neighborhood of A. But the equilibrium distribution will not obtain for all values of x. For, by assumption, the density of particles beyond C is very small compared to the equilibrium values; and in consequence of this there will be a slow diffusion of particles (across C) tending to restore equilibrium conditions throughout. If the barrier were sufficiently high, this diffusion will take place as though stationary conditions prevailed.

Assuming first that we are interested only in time intervals that are long compared to the time of relaxation β^{-1} we can use Smoluchowski's Eq. (312). Under stationary conditions, Smoluchowski's equation predicts a current density j given by [cf. Eq. (316)]

$$j \cdot \int_A^B \beta e^{m\mathfrak{B}/kT} ds = (kT/m) w e^{m\mathfrak{B}/kT} \Big|_B^A, \tag{466}$$

where, in the integral on the right-hand side, the path of integration of s from A to B is arbitrary. In our present case β is a constant and, since further we are dealing with a one-dimensional problem, we can express Eq. (466) in the form

$$j = \frac{kT}{m\beta} \frac{w e^{m\mathfrak{B}(x)/kT}\Big|_B^A}{\int_A^B e^{m\mathfrak{B}(x)/kT} dx}. \tag{467}$$

Now, the number of particles ν_A in the vicinity of A can be calculated; for, in accordance with our earlier remarks we shall be justified in assuming that the Maxwell-Boltzmann distribution

$$d\nu_A = w_A e^{-m\mathfrak{B}/kT} dx \tag{468}$$

is valid in the neighborhood of A. If we now further suppose that

$$\mathfrak{B} \simeq \tfrac{1}{2}\omega_A^2 x^2 \quad (\omega_A = \text{constant};\ x \sim 0), \tag{469}$$

we obtain from Eq. (468)

$$\nu_A = w_A \int_{-\infty}^{+\infty} \exp(-m\omega_A^2 x^2/2kT) dx, \tag{470}$$

where the range of integration over x has been extended from $-\infty$ to $+\infty$ in view of the fact that the main contribution to the integral for ν_A must arise only from a small region near $x=0$. Hence,

$$\nu_A = (w_A/\omega_A)(2\pi kT/m)^{\frac{1}{2}}. \tag{471}$$

Returning to Eq. (467), we can write with sufficient accuracy [cf. Eq. (469)]:

$$j \simeq \frac{kT}{m\beta} w_A \left\{ \int_1^B e^{m\mathfrak{B}/kT} dx \right\}^{-1}. \tag{472}$$

In writing Eq. (472) we have assumed that the density of particles near B is very small: this is true to begin with anyway.

From Eqs. (471) and (472) we directly obtain for the rate at which a particle, initially caught in the potential hole at A, will escape over the barrier at C, the expression

$$P=\frac{j}{\nu_A}=\frac{\omega_A}{\beta}\left(\frac{kT}{2\pi m}\right)^{\frac{1}{2}}\left\{\int_A^B e^{m\mathfrak{B}/kT}dx\right\}^{-1}. \tag{473}$$

The principal contribution to the integral in the curly brackets in the foregoing equation arises from only a very small region near C [on account of the strong inequality (465)]. The value of the integral will therefore depend, very largely, only on the shape of the potential curve in the immediate neighborhood of C. If we now suppose that near $x=x_C$, $\mathfrak{B}(x)$ has a continuous curvature, we may write

$$\mathfrak{B}\simeq Q-\tfrac{1}{2}\omega_C^2(x-x_C)^2 \quad (\omega_C=\text{constant};\ x\sim x_C). \tag{474}$$

On this assumption, to a sufficient degree of accuracy we have

$$\left.\begin{aligned}\int_A^B e^{m\mathfrak{B}/kT}dx&\simeq e^{mQ/kT}\int_{-\infty}^{+\infty}\exp\left[-m\omega_C^2(x-x_C)^2/2kT\right]dx,\\ &=e^{mQ/kT}(2\pi kT/m\omega_C^2)^{\frac{1}{2}}.\end{aligned}\right\} \tag{475}$$

Combining Eqs. (473) and (475) we obtain

$$P=(\omega_A\omega_C/2\pi\beta)e^{-mQ/kT}, \tag{476}$$

which gives the probability, per unit time, that a particle originally in the potential hole at A, will escape to B crossing the barrier at C.

The formula (476) has been derived on the basis of Eq. (467) and this implies, as we have already remarked, that we are ignoring effects which take place in intervals of the order β^{-1}. Alternatively, we may say that the validity of Eq. (476) depends on how large the coefficient of dynamical friction β is: if β were sufficiently large, the formula (476) for P may be expected to provide an adequate approximation [see Eqs. (507) and (508) below]. On the other hand, if this should not be the case, we must, in accordance with our remarks in Chapter II, §4, subsection (*vi*), base our discussion of the generalized Liouville Eq. (249) in phase space; and in one dimension this equation has the form

$$\frac{\partial W}{\partial t}+u\frac{\partial W}{\partial x}+K\frac{\partial W}{\partial u}=\beta u\frac{\partial W}{\partial u}+\beta W+q\frac{\partial^2 W}{\partial u^2}, \tag{477}$$

where it may be recalled that

$$q=\beta(kT/m);\quad K=-(\partial\mathfrak{B}/\partial x). \tag{478}$$

In II, §5 we have shown that the Maxwell-Boltzmann distribution identically satisfies Eq. (249). Accordingly,

$$W=C\exp\left[-(mu^2+2m\mathfrak{B})/2kT\right], \tag{479}$$

where C is a constant, satisfies Eq. (477). However, under the conditions of our problem the equilibrium distribution (479) cannot be valid for *all* values of x; for, if it were, there would be no diffusion across the barrier at C and the conditions of the problem would not be met. On the other hand, we do expect the distribution (479) to be realized to a high degree of accuracy in the neighborhood of A. We, therefore, look for a stationary solution of Eq. (477) of the form

$$W=CF(x,u)\exp\left[-m(u^2+2\mathfrak{B})/2kT\right], \tag{480}$$

where $F(x, u)$ is very nearly unity in the neighborhood of $x=0$. Since we have further supposed that the density of particles in the region B is quite negligible, we should also require that $F(x, u)\rightarrow 0$ for values of x appreciably greater than $x=x_C$. We may express these conditions formally in the form

$$\begin{aligned} F(x, u) &\simeq 1 \quad \text{at} \quad x\sim 0, \\ F(x, u) &\simeq 0 \quad \text{for} \quad x\gg x_C. \end{aligned} \tag{481}$$

We shall now show how such a function $F(x, u)$ can be determined.

First of all it is evident that for the purposes of determining the rate of escape of particles across the barrier at C it is particularly important to determine F accurately in this region. Assuming that in the vicinity of C, $\mathfrak{B}$ has the form (474) and that stationary conditions prevail throughout, the equation for W in the neighborhood of $x=x_C$ becomes [cf. Eq. (477)]:

$$u\frac{\partial W}{\partial X}+\omega_C^2X\frac{\partial W}{\partial u}=\beta u\frac{\partial W}{\partial u}+\beta W+q\frac{\partial^2 W}{\partial u^2}, \tag{482}$$

where for the sake of brevity, we have used

$$X=x-x_C. \tag{483}$$

According to Eqs. (474), (480), and (483) the appropriate form for W valid in the region C, is

$$W=Ce^{-mQ/kT}F(X, u)\exp\left[-m(u^2-\omega_C^2X^2)/2kT\right]. \tag{484}$$

Substituting for W according to this equation in Eq. (482), we obtain

$$u\frac{\partial F}{\partial X}+\omega_C^2X\frac{\partial F}{\partial u}=q\frac{\partial^2 F}{\partial u^2}-\beta u\frac{\partial F}{\partial u}. \tag{485}$$

It is seen that $F=$constant satisfies this equation identically: this solution corresponds of course to the equilibrium distribution. However, the solution of Eq. (485) which we are seeking must satisfy the boundary conditions [cf. Eq. (481)]

$$\begin{aligned} F(X, u)&\rightarrow 1 \quad \text{as} \quad X\rightarrow -\infty, \\ F(X, u)&\rightarrow 0 \quad \text{as} \quad X\rightarrow +\infty. \end{aligned} \tag{486}$$

Assume for F the form

$$F=F(u-aX)=F(\xi) \quad \text{(say)}, \tag{487}$$

where a is, for the present, an unspecified constant. Substituting this form of F in Eq. (485) we obtain

$$-\left[(a-\beta)u-\omega_C^2X\right]\frac{dF}{d\xi}=q\frac{d^2F}{d\xi^2}. \tag{488}$$

In order that Eq. (488) be consistent it is clearly necessary that [cf. Eq. (487)]

$$\left[\omega_C^2/(a-\beta)\right]=a; \tag{489}$$

and in this case Eq. (488) becomes

$$-(a-\beta)\xi\frac{dF}{d\xi}=q\frac{d^2F}{d\xi^2}. \tag{490}$$

Equation (490) is readily integrated to give

$$F=F_0\int^{\xi}\exp\left[-(a-\beta)\xi^2/2q\right]d\xi, \tag{491}$$

where F_0 is a constant. On the other hand, according to Eq. (489) a is the root of the equation

$$a^2 - a\beta - \omega_C^2 = 0; \tag{492}$$

i.e.,

$$a = (\beta/2) \pm ([\beta^2/4] + \omega_C^2)^{\frac{1}{2}}. \tag{493}$$

If we choose for a the *positive root*, then

$$a - \beta = ([\beta^2/4] + \omega_C^2)^{\frac{1}{2}} - (\beta/2) \tag{494}$$

is also positive, and as we shall show presently, the solution (491) leads to an F which satisfies the required boundary conditions (486). For, by choosing

$$F_0 = [(a-\beta)/2\pi q]^{\frac{1}{2}}, \tag{495}$$

and setting the lower limit of integration in Eq. (491) as $-\infty$ we obtain the solution

$$F = \left(\frac{a-\beta}{2\pi q}\right)^{\frac{1}{2}} \int_{-\infty}^{\xi} \exp\left[-(a-\beta)\xi^2/2q\right] d\xi, \tag{496}$$

which satisfies the conditions

$$F \to 1 \quad \text{as} \quad \xi \to +\infty; \quad F \to 0 \quad \text{as} \quad \xi \to -\infty. \tag{497}$$

On the other hand, since $\xi = u - aX$ and $a(=[(\beta/2)^2 + \omega_C^2]^{\frac{1}{2}} + [\beta/2])$ is positive, $\xi \to +\infty$ or $-\infty$ is the same as $X \to -\infty$ or $+\infty$; in other words, the solution (496) for F satisfies the necessary boundary conditions (486).

Combining Eqs. (484) and (496) we have, therefore, the solution

$$W = C[(a-\beta)/2\pi q]^{\frac{1}{2}} e^{-mQ/kT} \exp\left[-m(u^2 - \omega_C^2 X^2)/2kT\right] \int_{-\infty}^{\xi} \exp\left[-(a-\beta)\xi^2/2q\right] d\xi. \tag{498}$$

Equation (498) is, of course, valid only in the neighborhood of C.

In the vicinity of A we have the solution [cf. Eqs. (469) and (479)]

$$W = C \exp\left[-m(u^2 + \omega_A^2 x^2)/2kT\right]. \tag{499}$$

Accordingly, the number of particles, ν_A, in the potential hole at A is given by

$$\left.\begin{aligned} \nu_A &\simeq C \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp\left[-m(u^2 + \omega_A^2 x^2)/2kT\right] dx du, \\ &= C(2\pi kT/m\omega_A). \end{aligned}\right\} \tag{500}$$

(This equation will enable us to normalize the distribution in such a way so as to correspond to one particle in the potential hole: for this purpose we need only choose $C = m\omega_A/2\pi kT$.)

Now, the diffusion current across C is given by

$$j = \int_{-\infty}^{+\infty} W(X=0; u) u du, \tag{501}$$

or, using the solution (498), we have

$$j = C[(a-\beta)/2\pi q]^{\frac{1}{2}} e^{-mQ/kT} \int_{-\infty}^{+\infty} du\, u \exp(-mu^2/2kT) \int_{-\infty}^{u} d\xi \exp\left[-(a-\beta)\xi^2/2q\right]. \tag{502}$$

After an integration by parts, we find

$$j = C[(a-\beta)/2\pi q]^{\frac{1}{2}} (kT/m) e^{-mQ/kT} \int_{-\infty}^{+\infty} \exp\{-u^2[m/2kT + (a-\beta)/2q]\} du. \tag{503}$$

But [cf. Eq. (478)]

$$(m/2kT)+[(a-\beta)/2q]=(a/2q). \tag{504}$$

From Eqs. (503) and (504) we now obtain

$$j=C(kT/m)[(a-\beta)/a]^{\frac{1}{2}}e^{-mQ/kT}. \tag{505}$$

Hence, the rate of escape of particles across C is given by

$$P=(j/\nu_A)=(\omega_A/2\pi)[(a-\beta)/a]^{\frac{1}{2}}e^{-mQ/kT}, \tag{506}$$

or, substituting for a and $a-\beta$ according to Eqs. (493) and (494), we find after some elementary reductions, that

$$P=(\omega_A/2\pi\omega_C)([\beta^2/4+\omega_C^2]^{\frac{1}{2}}-[\beta/2])e^{-mQ/kT}. \tag{507}$$

If

$$\beta\gg 2\omega_C \tag{508}$$

we readily verify that our present "exact" formula for P reduces to our earlier result (476) derived on the basis of the Smoluchowski equation. But (507) now provides in addition the precise condition for the approximate validity of (476). On the other hand, for $\beta\rightarrow 0$ we have

$$P=(\omega_A/2\pi)e^{-mQ/kT}\quad(\beta\rightarrow 0). \tag{509}$$

This last formula for P valid in the limit of vanishing dynamical friction, corresponds to what is sometimes called the approximation of the *transition-state method.*

CHAPTER IV

PROBABILITY METHODS IN STELLAR DYNAMICS: THE STATISTICS OF THE GRAVITATIONAL FIELD ARISING FROM A RANDOM DISTRIBUTION OF STARS

1. Fluctuations in the Force Acting on a Star; The Outline of the Statistical Method

One of the principal problems of stellar dynamics is concerned with the analysis of the nature of the force acting on a star which is a member of a stellar system. In a general way, it appears that we may broadly distinguish between the influence of the system as a whole and the influence of the immediate local neighborhood; the former will be a smoothly varying function of position and time while the latter will be subject to relatively rapid fluctuations (see below).

Considering first the influence of the system as a whole, it appears that we can express it in terms of the gravitational potential $\mathfrak{B}(r;t)$ derived from the density function $n(r, M;t)$ which governs the average spatial distribution of the stars of different masses at time t. Thus,

$$\mathfrak{B}(r;t)=-G\int_{-\infty}^{+\infty}\int_0^{\infty}\frac{Mn(r_1, M;t)}{|r_1-r|}dMdr_1, \tag{510}$$

where G denotes the constant of gravitation. The potential $\mathfrak{B}(r;t)$ derived in this manner may be said to represent the "smoothed out" distribution of matter in the stellar system. The force per unit mass acting on a star due to the "system as a whole" is therefore given by

$$K=-\operatorname{grad}\mathfrak{B}(r;t). \tag{511}$$

However, the fluctuations in the *complexion* of the local stellar distribution will make the instantaneous force acting on a star deviate from the value given by Eq. (511). To elucidate the nature and origin of these fluctuations, we surround the star under consideration by an element of volume σ, which we may suppose is small enough to contain, on the average, only a relatively few stars. The actual number of stars, which will be found in σ at any given instant, will not in general be the average number that will be expected to be in it, namely σn; it will be subject to fluctuations. These fluctuations will naturally be governed by a Poisson distribution with the variance σn [see Eq. (333)]. It is in direct consequence of this changing complexion of the local stellar distribution that the influence of the near neighbors on a star is variable. The average period of such a fluctuation is readily estimated: for the order of magnitude of the time involved is evidently that required for two stars to separate by a distance equal to the average distance D between the stars (see Appendix VII). We may, therefore, expect that the influence of the immediate neighborhood will fluctuate with an average period of the order of

$$T \simeq (D/(\langle |V|^2 \rangle_{Av})^{\frac{1}{2}}), \tag{512}$$

where $\langle |V|^2 \rangle_{Av}^{\frac{1}{2}}$ denotes the root mean square relative velocity between two stars.

In the neighborhood of the sun, $D \sim 3$ parsecs, $\langle |V|^2 \rangle_{Av}^{\frac{1}{2}} \sim 50$ km/sec. Hence

$$T \text{ (near the sun)} \simeq 6 \times 10^4 \text{ years.} \tag{513}$$

When we compare this time with the period of galactic rotation (which is about 2×10^8 years) we observe that in conformity with our earlier remarks, the fluctuations in the force acting on a star due to the changing local stellar distribution do occur with extreme rapidity compared to the rate at which any of the other physical parameters change. Accordingly we may write for the force per unit mass acting on a star, the expression

$$\mathfrak{F} = \boldsymbol{K}(\boldsymbol{r}; t) + \boldsymbol{F}(t), \tag{514}$$

where $\boldsymbol{K}$ is derived from the smoothed out distribution [as in Eqs. (510) and (511)] and $\boldsymbol{F}$ denotes the fluctuating force due to the near neighbors. Moreover, if Δt denotes an interval of time long compared to (512), we may write

$$\mathfrak{F}\Delta t = \boldsymbol{K}\Delta t + \boldsymbol{\delta}(t+\Delta t; t), \tag{515}$$

where

$$\boldsymbol{\delta}(t+\Delta t; t) = \int_t^{t+\Delta t} \boldsymbol{F}(\xi) d\xi \quad (\Delta t \gg T). \tag{516}$$

Under the circumstances stated (namely, $\Delta t \gg T$) the accelerations $\boldsymbol{\delta}(t+\Delta t; t)$ and $\boldsymbol{\delta}(t+2\Delta t; t+\Delta t)$ suffered during two successive intervals $(t+\Delta t, t)$ and $(t+2\Delta t, t+\Delta t)$ will not be expected to show any correlation. We may, therefore, anticipate the existence of a definite law of distribution which will govern the probability of occurrence of the different values of $\boldsymbol{\delta}(t+\Delta t; t)$. We thus see that the acceleration which a star suffers during an interval $\Delta t \gg T$ can be formally expressed as the sum of two terms: a *systematic* term $\boldsymbol{K}\Delta t$ due to the action of the gravitational field of the smoothed out distribution, and a *stochastic* term $\boldsymbol{\delta}(t+\Delta t; t)$ representing the influence of the near neighbors. Stated in this fashion, we recognize the similarity[8] between our present problems in stellar dynamics and those in the theory of Brownian motion considered in Chapters II and III. One important difference should however be noted: Under our present circumstances it is possible, as we shall presently see, to undertake an analysis of the statistical properties of $\boldsymbol{F}(t)$ and $\boldsymbol{\delta}(t+\Delta t; t)$ based on first principles and without appealing to any "intuitive" or *a priori* considerations as in the discussions of Brownian motion [see the remarks at the end of II, §1 and also those following Eq. (318)].

We shall now outline a general method which appears suitable for analyzing the statistical properties of $\boldsymbol{F}$.

The force $\boldsymbol{F}$ acting on a star, per unit mass, is given by

$$\boldsymbol{F} = G \sum_i \frac{M_i}{|\boldsymbol{r}_i|^3} \boldsymbol{r}_i, \tag{517}$$

where M_i denotes the mass of a typical "field" star and $\boldsymbol{r}_i$ its position vector relative to the star under consideration; further, in Eq. (517) the summation is to be extended over all the neighboring stars. The actual value of $\boldsymbol{F}$ given by Eq. (517) at any particular instant of time will depend on the instantaneous complexion of the local stellar distribution; it is in consequence subject to fluctuations. We can therefore ask only for the probability of occurrence,

$$W(\boldsymbol{F})dF_x dF_y dF_z = W(\boldsymbol{F})d\boldsymbol{F}, \tag{518}$$

of $\boldsymbol{F}$ in the range $\boldsymbol{F}$ and $\boldsymbol{F}+d\boldsymbol{F}$. In evaluating this probability distribution, we shall (consistent with the physical situations we have in view) suppose that fluctuations subject only to the restriction of a constant average density occur.

The probability distribution $W(\boldsymbol{F})$ of $\boldsymbol{F}$ can be obtained by a direct application of Markoff's method outlined in Chapter I, §3. We shall obtain the explicit form of this distribution (sometimes called the Holtsmark distribution) in §2 below, but we should draw attention, already at this stage, to the fact that the specification of $W(\boldsymbol{F})$ does *not* provide us with all the

[8] Cf. particularly Eq. (317) and Eq. (515) above.

necessary information concerning the fluctuating force F for an equally important aspect of F concerns the *speed of fluctuations.*

According to Eq. (517) the rate of change of F with time is given by

$$f=\frac{dF}{dt}=G\sum_i M_i\left\{\frac{V_i}{|r_i|^3}-3r_i\frac{(r_i\cdot V_i)}{|r_i|^5}\right\}, \quad (519)$$

where V_i denotes the velocity of a typical field star *relative* to the star under consideration. It is now clear that the speed of fluctuations in F can be specified in terms of the bivariate distribution

$$W(F, f) \quad (520)$$

which governs the probability of the simultaneous occurrence of prescribed values for both F and f. It is seen that this distribution function $W(F, f)$ will depend on the assignment of *a priori* probability in the *phase space* in contrast to the distribution $W(F)$ of F which depends only on a similar assignment in the *configuration space.* Again, it is possible by an application of Markoff's method *formally* to write down a general expression for $W(F, f)$; but it does not appear feasible to obtain the required distribution function in an explicit form. However, as Chandrasekhar and von Neumann have shown, explicit formulae for *all* the first and the second moments of f for a given F can be obtained; and it appears possible to make some progress in the specification of the statistical properties of F in terms of these moments.

2. The Holtsmark distribution $W(F)$

We shall now obtain the stationary distribution $W(F)$ of the force F acting on a star, per unit mass, due to the gravitational attraction of the neighboring stars.

Without loss of generality we can suppose that the star under consideration is at the origin O of our system of coordinates. About O describe a sphere of radius R and containing N stars. In the first instance we shall suppose that

$$F=G\sum_{i=1}^{N}\frac{M_i}{|r_i|^3}r_i=\sum_{i=1}^{N}F_i. \quad (521)$$

But we shall subsequently let R and N tend to infinity simultaneously in such a way that

$$(4/3)\pi R^3 n=N$$
$$(R\rightarrow\infty;\ N\rightarrow\infty;\ n=\text{constant}). \quad (522)$$

This limiting process is permissible, in view of what we shall later show to be the case, namely, that the dominant contribution to F is made by the nearest neighbor [cf. Eqs. (560) and (564) below]; consequently, the formal extrapolation to infinity of the density of stars obtaining only in a given region of a stellar system can hardly affect the results to any appreciable extent.

Considering first the distribution $W_N(F)$ at the center of a finite sphere of radius R and containing N stars, we seek the probability that

$$F_0\leqslant F\leqslant F_0+dF_0. \quad (523)$$

Applying Markoff's method to this problem we have [cf. Eqs. (51) and (52)]

$$W_N(F_0)=\frac{1}{8\pi^3}\int_{-\infty}^{+\infty}\exp(-i\varrho\cdot F_0)A_N(\varrho)d\varrho, \quad (524)$$

where

$$A_N(\varrho)=\prod_{i=1}^{N}\int_{M_i=0}^{\infty}\int_{|r_i|=0}^{R}\exp(i\varrho\cdot F_i)\times\tau_i(r_i, M_i)dr_i dM_i. \quad (525)$$

In Eq. (525) $\tau_i(r_i, M_i)$ governs the probability of occurrence of the ith star at the position r_i with a mass M_i. If we now suppose that only fluctuations which are compatible with a constant average density occur, then

$$\tau_i(r_i, M_i)=(3/4\pi R^3)\tau(M), \quad (526)$$

where $\tau(M)$ now governs the frequency of occurrence of the different masses among the stars. With the assumption (526) concerning the τ_i's Eq. (525) reduces to

$$A_N(\varrho)=\left[\frac{3}{4\pi R^3}\int_{M=0}^{\infty}\int_{|r|=0}^{R}\exp(i\varrho\cdot\phi)\times\tau(M)drdM\right]^N, \quad (527)$$

where we have written

$$\phi=GMr/|r|^3. \quad (528)$$

We now let R and N tend to infinity according to Eq. (522). We thus obtain

$$W(\mathbf{F})=\frac{1}{8\pi^3}\int_{-\infty}^{+\infty}\exp\,(-i\boldsymbol{\rho}\cdot\mathbf{F})A(\boldsymbol{\rho})d\boldsymbol{\rho}, \tag{529}$$

where

$$A(\boldsymbol{\rho})=\underset{R\to\infty}{\text{limit}}\left[\frac{3}{4\pi R^3}\int_{M=0}^{\infty}\int_{|\mathbf{r}|=0}^{R}\exp\,(i\boldsymbol{\rho}\cdot\boldsymbol{\phi})\times\tau(M)d\mathbf{r}dM\right]^{4\pi R^3 n/3}. \tag{530}$$

Since,

$$\frac{3}{4\pi R^3}\int_{M=0}^{\infty}\int_{|\mathbf{r}|=0}^{R}\tau(M)dMd\mathbf{r}=1, \tag{531}$$

we can rewrite our expression for $A(\rho)$ in the form

$$A(\boldsymbol{\rho})=\underset{R\to\infty}{\text{limit}}\left[1-\frac{3}{4\pi R^3}\int_{=0}^{\infty}\int_{|\mathbf{r}|=0}^{R}\tau(M)\times[1-\exp\,(i\boldsymbol{\rho}\cdot\boldsymbol{\phi})]d\mathbf{r}dM\right]^{4\pi R^3 n/3}. \tag{532}$$

The integral over $\mathbf{r}$ which occurs in Eq. (532) is seen to be absolutely convergent when extended over *all* $|\mathbf{r}|$, i.e., also for $|\mathbf{r}|\to\infty$. We can accordingly write

$$A(\boldsymbol{\rho})=\underset{R\to\infty}{\text{limit}}\left[1-\frac{3}{4\pi R^3}\int_{M=0}^{\infty}\int_{|\mathbf{r}|=0}^{\infty}\tau(M)\times[1-\exp\,(i\boldsymbol{\rho}\cdot\boldsymbol{\phi})]d\mathbf{r}dM\right]^{4\pi R^3 n/3}, \tag{533}$$

or

$$A(\boldsymbol{\rho})=\exp\,[-nC(\boldsymbol{\rho})], \tag{534}$$

where

$$C(\boldsymbol{\rho})=\int_{M=0}^{\infty}\int_{|\mathbf{r}|=0}^{\infty}\tau(M)[1-\exp\,(i\boldsymbol{\rho}\cdot\boldsymbol{\phi})]d\mathbf{r}dM. \tag{535}$$

In the integral defining $C(\boldsymbol{\rho})$ we shall introduce $\boldsymbol{\phi}$ as the variable of integration instead of $\mathbf{r}$. We readily verify that

$$d\mathbf{r}=-\tfrac{1}{2}(GM)^{3/2}|\boldsymbol{\phi}|^{-9/2}d\boldsymbol{\phi}. \tag{536}$$

Hence,

$$C(\boldsymbol{\rho})=\tfrac{1}{2}G^{3/2}\int_0^{\infty}dMM^{3/2}\tau(M)\int_{-\infty}^{+\infty}d\boldsymbol{\phi}|\boldsymbol{\phi}|^{-9/2}\times[1-\exp\,(i\boldsymbol{\rho}\cdot\boldsymbol{\phi})], \tag{537}$$

or, in an obvious notation

$$C(\boldsymbol{\rho})=\tfrac{1}{2}G^{3/2}\langle M^{3/2}\rangle_{\text{Av}}\int_{-\infty}^{+\infty}[1-\exp\,(i\boldsymbol{\rho}\cdot\boldsymbol{\phi})]\times|\boldsymbol{\phi}|^{-9/2}d\boldsymbol{\phi}. \tag{538}$$

The foregoing expression is clearly unaffected if we replace $\boldsymbol{\phi}$ by $-\boldsymbol{\phi}$. But this replacement changes $\exp\,(i\boldsymbol{\rho}\cdot\boldsymbol{\phi})$ into $\exp\,(-i\boldsymbol{\rho}\cdot\boldsymbol{\phi})$ under the integral sign; taking the arithmetic mean of the two resulting integrals, we obtain

$$C(\boldsymbol{\rho})=\tfrac{1}{2}G^{3/2}\langle M^{3/2}\rangle_{\text{Av}}\int_{-\infty}^{+\infty}[1-\cos\,(\boldsymbol{\rho}\cdot\boldsymbol{\phi})]|\boldsymbol{\phi}|^{-9/2}d\boldsymbol{\phi}. \tag{539}$$

Choosing polar coordinates with the z axis in the direction of $\boldsymbol{\rho}$ Eq. (539) can be transformed to

$$C(\boldsymbol{\rho})=\tfrac{1}{2}G^{3/2}\langle M^{3/2}\rangle_{\text{Av}}\int_0^{\infty}\int_{-1}^{+1}\int_0^{2\pi}\times[1-\cos\,(|\boldsymbol{\rho}||\boldsymbol{\phi}|t)]|\boldsymbol{\phi}|^{-5/2}d\omega dtd|\boldsymbol{\phi}|, \tag{540}$$

or, introducing further the variable $z=|\boldsymbol{\rho}||\boldsymbol{\phi}|$, we have

$$C(\boldsymbol{\rho})=\tfrac{1}{2}G^{3/2}\langle M^{3/2}\rangle_{\text{Av}}|\boldsymbol{\rho}|^{3/2}\times\int_0^{\infty}\int_{-1}^{+1}\int_0^{2\pi}[1-\cos\,(zt)]z^{-5/2}d\omega dtdz. \tag{541}$$

After performing the integrations over ω and t we obtain

$$C(\boldsymbol{\rho})=2\pi G^{3/2}\langle M^{3/2}\rangle_{\text{Av}}|\boldsymbol{\rho}|^{3/2}\times\int_0^{\infty}(z-\sin z)z^{-7/2}dz, \tag{542}$$

or after several integrations by parts

$$\begin{aligned}C(\boldsymbol{\rho})&=\frac{16}{15}\pi G^{3/2}\langle M^{3/2}\rangle_{\text{Av}}|\boldsymbol{\rho}|^{3/2}\int_0^{\infty}z^{-1/2}\cos z dz.\\&=\frac{4}{15}(2\pi G)^{3/2}\langle M^{3/2}\rangle_{\text{Av}}|\boldsymbol{\rho}|^{3/2}.\end{aligned} \tag{543}$$

Combining Eqs. (529), (534), and (543) we now obtain

$$W(\mathbf{F})=\frac{1}{8\pi^3}\int_{-\infty}^{+\infty}\exp\,(-i\boldsymbol{\rho}\cdot\mathbf{F}-a|\boldsymbol{\rho}|^{3/2})d\boldsymbol{\rho}, \tag{544}$$

where we have written

$$a=(4/15)(2\pi G)^{3/2}\langle M^{3/2}\rangle_{\text{Av}}n. \tag{545}$$

Using a frame of reference in which one of the principal axes is in the direction of $\mathbf{F}$ and changing to polar coordinates, the formula (544) for $W(F)$ can be reduced to

$$W(F)=\frac{1}{4\pi^2}\int_0^\infty\int_{-1}^{+1}\exp\left(-i|\varrho||F|t-a|\varrho|^{3/2}\right)\times|\varrho|^2dtd|\varrho|. \quad (546)$$

The integration over t is readily effected, and we obtain

$$W(F)=\frac{1}{2\pi^2|F|}\int_0^\infty\exp\left(-a|\varrho|^{3/2}\right)\times|\varrho|\sin\left(|\varrho||F|\right)d|\varrho|. \quad (547)$$

If we now put

$$x=|\varrho||F|, \quad (548)$$

Eq. (547) becomes

$$W(F)=\frac{1}{2\pi^2|F|^3}\int_0^\infty\exp\left(-ax^{3/2}/|F|^{3/2}\right)\times x\sin x dx. \quad (549)$$

We can rewrite the foregoing formula for $W(F)$ more conveniently if we introduce the *normal field* Q_H defined by

$$\left.\begin{aligned}Q_H&=a^{2/3}=(4/15)^{2/3}(2\pi G)(\langle M^{3/2}\rangle_{Av}n)^{2/3},\\&=2.6031G(\langle M^{3/2}\rangle_{Av}n)^{2/3}\end{aligned}\right\} \quad (550)$$

and express $|F|$ in terms of this unit:

$$|F|=\beta Q_H=\beta a^{2/3}. \quad (551)$$

Equation (549) now takes the form

$$W(F)=H(\beta)/4\pi a^2\beta^2, \quad (552)$$

where we have introduced the function $H(\beta)$ defined by

$$H(\beta)=\frac{2}{\pi\beta}\int_0^\infty\exp\left[-(x/\beta)^{3/2}\right]x\sin x dx. \quad (553)$$

Since,

$$W(|F|)=4\pi|F|^2W(F), \quad (554)$$

we obtain from Eqs. (551) and (552)

$$W(|F|)=H(\beta)/Q_H; \quad (555)$$

accordingly $H(\beta)$ defines the probability distribution of $|F|$ when it is expressed in units of Q_H. The function $H(\beta)$ has been evaluated numerically and is tabulated in Table IX.

The asymptotic behavior of the distribution $W(|F|)$ can be obtained from the formulae:

$$H(\beta)=4\beta^2/3\pi+O(\beta^4) \quad (\beta\to 0), \quad (556)$$

and

$$H(\beta)=(15/8)(2/\pi)^{1/2}\beta^{-5/2}+O(\beta^{-4}) \quad (\beta\to\infty). \quad (557)$$

We find [cf. Eqs. (551) and (555)]

$$W(|F|)\simeq(4/3\pi Q_H^3)|F|^2 \quad (|F|\to 0), \quad (558)$$

and

$$W(|F|)\simeq(15/8)(2/\pi)^{1/2}Q_H^{3/2}|F|^{-5/2} \quad (|F|\to\infty). \quad (559)$$

Substituting for Q_H from Eq. (550) in Eq. (559) we obtain

$$W(|F|)\simeq 2\pi G^{3/2}\langle M^{3/2}\rangle_{Av}n|F|^{-5/2} \quad (|F|\to\infty). \quad (560)$$

It is seen that while the frequency of occurrence of both the weak and the strong fields is quite small, it is only the fields of average intensity which have appreciable probabilities. In particular, the value of $|F|$ which has the maximum probability of occurrence is found to be (see Table IX) $\sim 1.6Q_H$.

Equations (552) and (553) provide, of course, the *exact* formula for the distribution of F for an *ideally* random distribution of stars. But an elementary treatment which leads to an approximate formula for $W(F)$ is of some interest and illuminates certain points in the theory. The treatment we refer to is based on the assumption that the force acting on a star is entirely due to its *nearest* neighbor.

Now, the law of distribution of the nearest neighbor is given by [see Appendix VII, Eq. (671)]

$$w(r)dr=\exp\left(-4\pi r^3n/3\right)4\pi r^2ndr, \quad (561)$$

and, since on the first neighbor approximation

$$|F|=GMr^{-2}, \quad (562)$$

we readily obtain the formula

$$W(|F|)d|F|=\exp\left[-4\pi(GM)^{3/2}n/3|F|^{3/2}\right]\times 2\pi(GM)^{3/2}n|F|^{-5/2}d|F|. \quad (563)$$

TABLE IX. The function $H(\beta)$.

β	$H(\beta)$	β	$H(\beta)$
0.0		5.0	0.04310
0.1	0.004225	5.2	0.03790
0.2	0.016666	5.4	0.03357
0.3	0.036643	5.6	0.02993
0.4	0.063084	5.8	0.02683
0.5	0.094601	6.0	0.02417
0.6	0.129598	6.2	0.02188
0.7	0.166380	6.4	0.01988
0.8	0.203270	6.6	0.01814
0.9	0.238704	6.8	0.01660
1.0	0.271322	7.0	0.01525
1.1	0.30003	7.2	0.01405
1.2	0.32402	7.4	0.01297
1.3	0.34281	7.6	0.01201
1.4	0.35620	7.8	0.01115
1.5	0.36426	8.0	0.01038
1.6	0.36726	8.2	0.00967
1.7	0.36566	8.4	0.00903
1.8	0.36004	8.6	0.00846
1.9	0.35101	8.8	0.00793
2.0	0.33918	9.0	0.00745
2.1	0.32519	9.2	0.00701
2.2	0.30951	9.4	0.00660
2.3	0.29266	9.6	0.00622
2.4	0.27485	9.8	0.00588
2.5	0.25667	10.0	0.00556
2.6	0.238	15.0	0.00188
2.7	0.222	20.0	0.00089
2.8	0.206	25.0	0.00050
2.9	0.190	30.0	0.00031
3.0	0.176	35.0	0.00021
3.2	0.150	40.0	0.00015
3.4	0.128	45.0	0.00011
3.6		50.0	0.00009
3.8		60.0	0.00005
4.0		70.0	0.00004
4.2		80.0	0.00003
4.4	0.06734	90.0	0.00002
4.6	0.05732	100.0	0.00002
4.8	0.04944		

According to the distribution (563)

$$W(|F|)\simeq 2\pi(GM)^{3/2}n|F|^{-5/2} \quad (|F|\rightarrow\infty), \quad (564)$$

which is seen to be in *exact* agreement with the formula (560) derived from the Holtsmark distribution (555). The physical meaning of this agreement, for $|F|\rightarrow\infty$ in the results derived from an exact and an approximate treatment of the same problem, is simply that the highest fields are in reality produced only by the nearest neighbor. More generally, it is found that the two distributions (555) and (563) agree over most of the range of $|F|$. Thus, the field which has the maximum frequency of occurrence on the basis of (563) is seen to differ from the corresponding value on the Holtsmark distribution by less than five percent. The region in which the two distributions (555) and (563) differ most markedly is when $|F|\rightarrow 0$: on the Holtsmark distribution $W(|F|)$ tends to zero as $|F|^2$ while on the nearest neighbor approximation $W(|F|)$ tends to zero as $\exp(-\text{const.}\,|F|^{-\frac{3}{2}})$ [cf. Eqs. (558) and (564)]. However, the fact that the nearest neighbor approximation should be seriously in error for the weak fields is, of course, to be expected: for, a weak field arises from a more or less symmetrical, average, complexion of the stars around the one under consideration and consequently F under these circumstances is the result of the action of several stars and not due to any one single star.

Finally, we may draw attention to one important difficulty in using the Holstmark distribution for *all* values of $|F|$: It predicts relatively too high probabilities for $|F|$ as $|F|\rightarrow\infty$. Thus, on the basis of the distribution (555), $\langle|F|^2\rangle_{Av}$ is divergent. [The same remark also applies to the distribution (564).] These relatively high probabilities for the high field strengths is a consequence of our assumption of complete randomness in stellar distribution for *all* elements of volume. It is, however, apparent that this assumption cannot be valid for the regions in the *very* immediate neighborhoods of the individual stars. For, if V denotes the relative velocity between two stars when separated by distances of the order of the average distance between the stars, the two stars cannot come closer together (on the approximation of linear trajectories) than a certain critical distance $r(|V|)$ such that

$$|V|^2/2=[G(M_1+M_2)/r(|V|)], \quad (565)$$

or

$$r(|V|)=[2G(M_1+M_2)/|V|^2]. \quad (566)$$

Otherwise the two stars should be strictly regarded as the components of a binary system and this is inconsistent with our original premises. This restriction therefore leads us to infer that departures from true randomness exist for $r\sim r(|V|)$. However, under the conditions we normally encounter in stellar systems, $r(|V|)$ is very small compared to the average distance between the stars. Thus, in our galaxy, in the general neighborhood of the sun, $r(|V|)\sim 2\times 10^{-6}$ parsec, and this is to be compared with an average distance between the stars of about three parsecs. Accordingly, we may expect the

Holtsmark distribution to be very close to the true distribution, except for the very highest values of $|F|$. More particularly, the deviations from the Holtsmark distribution are to be expected for field strengths of the order of

$$|F| \sim (GM_2/[r(|V|)]^2) \simeq (M_2[(|V|^2)_m]^2/4G(M_1+M_2)^2). \quad (567)$$

When $|F|$ becomes much larger than the quantity on the right-hand side of Eq. (567), the true frequencies of occurrence will very rapidly tend to zero as compared to what would be expected on the Holtsmark distribution, namely (560). A rigorous treatment of these deviations from the distribution (555) will require a reconsideration of the whole problem in *phase space* and is beyond the scope of the present investigation.

3. The Speed of Fluctuations in F

As we have already remarked the speed of fluctuations can be specified in terms of the distribution function $W(F, f)$ which gives the simultaneous probability of a given field strength F and an associated rate of change of F of amount f [cf. Eqs. (517) and (519)]. The general expression for this probability distribution can be readily written down using Markoff's method [I, §3, Eqs. (51), (52), and (53)]. We have [cf. Eqs. (529) and (530)]

$$W(F, f) = \frac{1}{64\pi^6}\int_{|\rho|=0}^{\infty}\int_{|\sigma|=0}^{\infty} \exp\left[-i(\rho\cdot F+\sigma\cdot f)\right]A(\rho, \sigma)d\rho d\sigma, \quad (568)$$

where

$$A(\rho, \sigma) = \underset{R\to\infty}{\text{limit}}\left[\frac{3}{4\pi R^3}\int_{0<M<\infty}\int_{|r|<R}\int_{|V|<\infty} \exp\left[i(\rho\cdot\phi+\sigma\cdot\psi)\right]\tau dr dV dM\right]^{4\pi R^3 n/3}. \quad (569)$$

In Eqs. (568) and (569) ρ and σ are two auxiliary vectors, n denotes the number of stars per unit volume, and

$$\phi = GM\frac{r}{|r|^3}; \quad \psi = GM\left\{\frac{V}{|r|^3} - 3\frac{r(r\cdot V)}{|r|^5}\right\}. \quad (570)$$

Further,

$$\tau dV dM = \tau(V; M)dV dM \quad (571)$$

gives the probability that a star with a relative velocity in the range $(V, V+dV)$ and with a mass between M and $M+dM$ will be found. It should also be noted that in writing down Eqs. (568) and (569) we have supposed (as in §2) that the fluctuations in the local stellar distribution which occur are subject only to the restriction of a constant average density.

Since

$$\frac{3}{4\pi R^3}\int_{M=0}^{\infty}\int_{|r|<R}\int_{|V|<\infty} \tau dr dV dM = 1, \quad (572)$$

we can rewrite (569) as

$$A(\rho, \sigma) = \underset{R\to\infty}{\text{limit}}\left\{1 - \frac{3}{4\pi R^3}\int_{M=0}^{\infty}\int_{|r|<R}\int_{|V|<\infty} \left(1-\exp\left[i(\rho\cdot\phi+\sigma\cdot\psi)\right]\right)\tau dr dV dM\right\}^{4\pi R^3 n/3}. \quad (573)$$

The integral over r which occurs in Eq. (573) is seen to be conditionally convergent when extended over all $|r|$, i.e., also for $|r|\to\infty$. Hence, we can write

$$A(\rho, \sigma) = \underset{R\to\infty}{\text{limit}}\left\{1 - \frac{3}{4\pi R^3}\int_{M=0}^{\infty}\int_{|r|=0}^{\infty}\int_{|V|=0}^{\infty} \left(1-\exp\left[i(\rho\cdot\phi+\sigma\cdot\psi)\right]\right)\tau dr dV dM\right\}^{4\pi R^3 n/3}, \quad (574)$$

or

$$A(\rho, \sigma) = \exp\left[-nC(\rho, \sigma)\right] \quad (575)$$

where

$$C(\rho, \sigma) = \int_0^\infty \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \{1 - \exp[i(\rho \cdot \phi + \sigma \cdot \psi)]\} \tau d\tau dV dM. \tag{576}$$

This formally solves the problem. It does not, however, appear that the integral representing $C(\rho, \sigma)$ can be evaluated explicitly in terms of any known functions. But if we are interested only in the moments of f for a given F and of F for a given f we need only the behavior of $A(\rho, \sigma)$and, therefore, also of $C(\rho, \sigma)$ for $|\sigma|$, respectively, $|\rho|$ tending to zero. For, considering the first and the second moments of the components f_ξ, f_η, and f_ζ of f along three directions ξ, η, and ζ at right angles to each other, we have

$$W(F)\langle f_\mu \rangle_{Av} = \int_{|f|=0}^\infty W(F, f) f_\mu df \quad (\mu = \xi, \eta, \zeta), \tag{577}$$

and

$$W(F)\langle f_\mu f_\nu \rangle_{Av} = \int_{|f|=0}^\infty W(F, f) f_\mu f_\nu df \quad (\mu, \nu = \xi, \eta, \zeta), \tag{578}$$

where $W(F)$ denotes the distribution of F for which we have already obtained an explicit formula in §2. Substituting now for $W(F, f)$ from Eq. (568) in the foregoing formulae for the moments we obtain

$$W(F)\langle f_\mu \rangle_{Av} = \frac{1}{64\pi^6} \int_{|f|=0}^\infty \int_{|\rho|=0}^\infty \int_{|\sigma|=0}^\infty \exp[-i(\rho \cdot F + \sigma \cdot f)] A(\rho, \sigma) f_\mu d\rho d\sigma df, \tag{579}$$

and

$$W(F)\langle f_\mu f_\nu \rangle_{Av} = \frac{1}{64\pi^6} \int_{|f|=0}^\infty \int_{|\rho|=0}^\infty \int_{|\sigma|=0}^\infty \exp[-i(\rho \cdot F + \sigma \cdot f)] A(\rho, \sigma) f_\mu f_\nu d\rho d\sigma df. \tag{580}$$

But

$$\left.\begin{aligned}
&\frac{1}{8\pi^3} \int_{|f|=0}^\infty \exp(-i\sigma \cdot f) f_\xi df = i\delta'(\sigma_\xi)\delta(\sigma_\eta)\delta(\sigma_\zeta), \\
&\frac{1}{8\pi^3} \int_{|f|=0}^\infty \exp(-i\sigma \cdot f) f_\xi^2 df = -\delta''(\sigma_\xi)\delta(\sigma_\eta)\delta(\sigma_\zeta), \\
&\frac{1}{8\pi^3} \int_{|f|=0}^\infty \exp(-i\sigma \cdot f) f_\xi f_\eta df = -\delta'(\sigma_\xi)\delta'(\sigma_\eta)\delta(\sigma_\zeta),
\end{aligned}\right\} \tag{581}$$

etc. In Eq. (581) δ denotes Dirac's δ-function and δ' and δ'' its first and second derivatives; and remembering also that

$$\int_{-\infty}^{+\infty} f(x)\delta(x)dx = f(0); \quad \int_{-\infty}^{+\infty} f(x)\delta'(x)dx = -f'(0); \quad \int_{-\infty}^{+\infty} f(x)\delta''(x)dx = f''(0), \tag{582}$$

Eqs. (579) and (580) for the moments reduce to

$$W(F)\langle f_\mu \rangle_{Av} = -\frac{i}{8\pi^3} \int_{|\rho|=0}^\infty \exp(-i\rho \cdot F) \left[\frac{\partial}{\partial \sigma_\mu} A(\rho, \sigma)\right]_{|\sigma|=0} d\rho, \tag{583}$$

and

$$W(F)\langle f_\mu f_\nu \rangle_{Av} = -\frac{1}{8\pi^3} \int_{|\rho|=0}^\infty \exp(-i\rho \cdot F) \left[\frac{\partial^2}{\partial \sigma_\mu \partial \sigma_\nu} A(\rho, \sigma)\right]_{|\sigma|=0} d\rho. \tag{584}$$

We accordingly see that the first and the second moments of f can be evaluated from a series expansion of $A(\rho, \sigma)$ or of $C(\rho, \sigma)$ which is correct up to the *second order* in $|\sigma|$. Such a series expansion has been found by Chandrasekhar and von Neumann and, quoting their final result, we have

$$C(\varrho,\sigma)=\frac{4}{15}(2\pi)^{\frac{3}{4}}G^{\frac{3}{2}}\langle M^{\frac{3}{2}}\rangle_M|\varrho|^{\frac{3}{2}}+\frac{2}{3}\pi iG(\sigma_1\langle MV_1\rangle_M+\sigma_2\langle MV_2\rangle_M-2\sigma_3\langle MV_3\rangle_M)$$
$$+\frac{3}{28}(2\pi)^{\frac{1}{4}}G^{\frac{1}{2}}|\varrho|^{-\frac{1}{2}}[(5\sigma_1^2+4\sigma_2^2-2\sigma_3^2)\langle M^{\frac{1}{2}}V_1^2\rangle_M+(4\sigma_1^2+5\sigma_2^2-2\sigma_3^2)\langle M^{\frac{1}{2}}V_2^2\rangle_M$$
$$+(4\sigma_3^2-2\sigma_1^2-2\sigma_2^2)\langle M^{\frac{1}{2}}V_3^2\rangle_M-8\sigma_2\sigma_3\langle M^{\frac{1}{2}}V_2V_3\rangle_M-8\sigma_3\sigma_1\langle M^{\frac{1}{2}}V_3V_1\rangle_M$$
$$+2\sigma_1\sigma_2\langle M^{\frac{1}{2}}V_1V_2\rangle_M]+O(|\sigma|^3)\quad(|\sigma|\to 0), \tag{585}$$

where $\langle\ \rangle_M$ indicates that the corresponding quantity has been averaged with the weight function $\tau(V;M)$ [cf. Eq. (571)]; further, in Eq. (585) $(\sigma_1,\sigma_2,\sigma_3)$ and (V_1,V_2,V_3) are the components of σ and V in a system of coordinates in which the z axis is in the direction of ϱ.

In Eq. (585) $V=(V_1,V_2,V_3)$ is of course the velocity of a field star relative to the one under consideration. If we now let u and v denote the velocities of the field star and the star under consideration in an appropriately chosen local standard of rest, then

$$V=u-v. \tag{586}$$

In their further discussion, Chandrasekhar and von Neumann introduce the assumption that the distribution of the velocities u among the stars is *spherical*, i.e., the distribution function $\Psi(u)$ has the form

$$\Psi(u)=\Psi(j^2(M)|u|^2), \tag{587}$$

where Ψ is an arbitrary function of the argument specified and the parameter j (of the dimensions of [velocity]$^{-1}$) can be a function of the mass of the star. This assumption for the distribution of the peculiar velocities u implies that the probability function $\tau(V;M)$ must be expressible as

$$\tau(V;M)=\Psi[j^2(M)|u|^2]\chi(M), \tag{588}$$

where $\chi(M)$ governs the distribution over the different masses. For a function τ of this form we clearly have

$$\begin{aligned}&\langle MV_i\rangle_M=-\langle M\rangle_M v_i;\quad \langle M^{\frac{1}{2}}V_i^2\rangle_M=\tfrac{1}{3}\langle M^{\frac{1}{2}}|u|^2\rangle_M+\langle M^{\frac{1}{2}}\rangle_M v_i^2\quad(i=1,2,3),\\&\langle M^{\frac{1}{2}}V_iV_j\rangle_M=\langle M^{\frac{1}{2}}\rangle_M v_iv_j\quad(i,j=1,2,3,\ i\neq j).\end{aligned} \tag{589}$$

Substituting these values in Eq. (577) we find after some minor reductions that

$$C(\varrho,\sigma)=\frac{4}{15}(2\pi)^{\frac{3}{4}}G^{\frac{3}{2}}\langle M^{\frac{3}{2}}\rangle_M|\varrho|^{\frac{3}{2}}-\frac{2}{3}\pi iG\langle M\rangle_M(\sigma_1v_1+\sigma_2v_2-2\sigma_3v_3)+\frac{1}{4}(2\pi)^{\frac{1}{4}}G^{\frac{1}{2}}\langle M^{\frac{1}{2}}|u|^2\rangle_M|\varrho|^{-\frac{1}{2}}(\sigma_1^2+\sigma_2^2)$$
$$+\frac{3}{28}(2\pi)^{\frac{1}{4}}G^{\frac{1}{2}}\langle M^{\frac{1}{2}}\rangle_M|\varrho|^{-\frac{1}{2}}[\sigma_1^2(5v_1^2+4v_2^2-2v_3^2)+\sigma_2^2(5v_2^2+4v_1^2-2v_3^2)$$
$$+\sigma_3^2(4v_3^2-2v_1^2-2v_2^2)+2\sigma_1\sigma_2v_1v_2-8\sigma_2\sigma_3v_2v_3-8\sigma_3\sigma_1v_3v_1]+O(|\sigma|^3). \tag{590}$$

With a series expansion of this form we can, as we have already remarked, evaluate all the first and the second moments of f for a given F.

Considering first the moment of f, Chandrasekhar and von Neumann find that

$$\langle f\rangle_M=\overline{\left(\frac{dF}{dt}\right)}_{F,v}=-\frac{2}{3}\pi G\langle M\rangle_M nB\left(\frac{|F|}{Q_H}\right)\left(v-3\frac{v\cdot F}{|F|^2}F\right), \tag{591}$$

where Q_H is the "normal field" introduced in §2 [Eqs. (550) and (551)] and

$$B(\beta)=3\left(\int_0^\beta H(\beta)d\beta\Big/\beta H(\beta)\right)-1. \tag{592}$$

We shall examine certain formal consequences of Eq. (592).

Multiplying Eq. (591) scalarly with $\boldsymbol{F}$ we obtain

$$\boldsymbol{F}\cdot\overline{\left(\frac{d\boldsymbol{F}}{dt}\right)}_{\boldsymbol{F},\boldsymbol{v}}=\frac{4}{3}\pi G(M)_m nB\left(\frac{|\boldsymbol{F}|}{Q_H}\right)(\boldsymbol{v}\cdot\boldsymbol{F}); \tag{593}$$

but

$$\boldsymbol{F}\cdot\overline{\left(\frac{d\boldsymbol{F}}{dt}\right)}_{\boldsymbol{F},\boldsymbol{v}}=|\boldsymbol{F}|\overline{\left(\frac{d|\boldsymbol{F}|}{dt}\right)}_{\boldsymbol{F},\boldsymbol{v}}. \tag{594}$$

Hence,

$$\overline{\left(\frac{d|\boldsymbol{F}|}{dt}\right)}_{\boldsymbol{F},\boldsymbol{v}}=\frac{4}{3}\pi G(M)_m nB\left(\frac{|\boldsymbol{F}|}{Q_H}\right)\frac{\boldsymbol{v}\cdot\boldsymbol{F}}{|\boldsymbol{F}|}. \tag{595}$$

On the other hand, if F_j denotes the component of $\boldsymbol{F}$ in an arbitrary direction at right angles at the direction of $\boldsymbol{v}$ then according to Eq. (591)

$$\overline{\left(\frac{dF_j}{dt}\right)}_{\boldsymbol{F},\boldsymbol{v}}=2\pi G(M)_m nB\left(\frac{|\boldsymbol{F}|}{Q_H}\right)\frac{\boldsymbol{v}\cdot\boldsymbol{F}}{|\boldsymbol{F}|^2}F_j. \tag{596}$$

Combining Eqs. (595) and (596) we have

$$\frac{1}{F_j}\overline{\left(\frac{dF_j}{dt}\right)}_{\boldsymbol{F},\boldsymbol{v}}=\frac{3}{2}\frac{1}{|\boldsymbol{F}|}\overline{\left(\frac{d|\boldsymbol{F}|}{dt}\right)}_{\boldsymbol{F},\boldsymbol{v}}. \tag{597}$$

Equation (597) is clearly equivalent to

$$\overline{\frac{d}{dt}(\log F_j-\tfrac{3}{2}\log|\boldsymbol{F}|)_{\boldsymbol{F},\boldsymbol{v}}}=0. \tag{598}$$

We have thus proved that

$$\overline{\left[\frac{d}{dt}\left(\frac{F_j}{|\boldsymbol{F}|^{\frac{3}{2}}}\right)\right]}_{\boldsymbol{F},\boldsymbol{v}}=0. \tag{599}$$

We shall now examine the physical consequences of Eq. (591) more closely. In words, the meaning of this equation is that the component of

$$-\frac{2}{3}\pi G(M)_m nB\left(\frac{|\boldsymbol{F}|}{Q_H}\right)\left(\boldsymbol{v}-3\frac{\boldsymbol{v}\cdot\boldsymbol{F}}{|\boldsymbol{F}|^2}\boldsymbol{F}\right) \tag{600}$$

along any particular direction gives the average value of the rate of change of $\boldsymbol{F}$ that is to be expected in the specified direction when the star is moving with a velocity $\boldsymbol{v}$. Stated in this manner we at once see the essential difference in the stochastic variations of $\boldsymbol{F}$ with time in the two cases $|\boldsymbol{v}|=0$ and $|\boldsymbol{v}|\neq 0$. In the former case $\langle\dot{\boldsymbol{F}}\rangle_m=0$; but this is not generally true when $|\boldsymbol{v}|\neq 0$. Or expressed differently, when $|\boldsymbol{v}|=0$ the changes in $\boldsymbol{F}$ occur with equal probability in all directions while this is not the case when $|v| \neq 0$. The true nature of this difference is brought out very clearly when we consider

$$\overline{\left(\frac{d|F|}{dt}\right)}_{F,\,v} \tag{601}$$

according to Eq. (595). Remembering that $B(\beta) \geqslant 0$ for $\beta \geqslant 0$, we conclude from Eq. (595) that

$$\overline{\left(\frac{d|F|}{dt}\right)}_{F,\,v} > 0 \quad \text{if} \quad (v \cdot F) > 0, \tag{602}$$

and

$$\overline{\left(\frac{d|F|}{dt}\right)}_{F,\,v} < 0 \quad \text{if} \quad (v \cdot F) < 0. \tag{603}$$

In other words, if F has a positive component in the direction of v, $|F|$ increases on the average, while if F has a negative component in the direction of v, $|F|$ decreases on the average. This essential asymmetry introduced by the direction of v may be expected to give rise to the phenomenon of ***dynamical friction***.

Considering next the second moments of f Chandrasekhar and von Neumann find that

$$\begin{aligned}(|f|^2{}_{F,\,v})_{Av} = 2ab\frac{\beta^{1/2}}{H(\beta)}\{2G(\beta)+7k[G(\beta)\sin^2\alpha - I(\beta)(3\sin^2\alpha-2)]\}\\ +\frac{g^2}{\beta H(\beta)}\{\beta H(\beta)(4-3\sin^2\alpha)+3K(\beta)(3\sin^2\alpha-2)\},\end{aligned} \tag{604}$$

where, α denote- the angle between the directions of F and v

$$a = \frac{4}{15}(2\pi)^{1/2}G^{1/2}\langle M^{1/2}\rangle_{Av}n;\quad b = \frac{1}{4}(2\pi)^{1/2}G^{1/2}\langle M^{3/2}|u|^2\rangle_{Av}n,\quad g = \frac{2}{3}\pi G\langle M\rangle_{Av}|v|n;\quad k = \frac{3}{7}\frac{\langle M^{3/2}\rangle_{Av}|v|^2}{\langle M^{3/2}|u|^2\rangle_{Av}}, \tag{605}$$

and

$$\left.\begin{aligned}&H(\beta) = \frac{2}{\pi\beta}\int_0^\infty \exp[-(x/\beta)^{3/2}]\beta\sin\beta d\beta,\\ &G(\beta) = \frac{3}{2}\int_0^\beta \beta^{-1/2}H(\beta)d\beta,\qquad I(\beta) = \beta^{-3/2}\int_0^\beta \beta^{1/2}G(\beta)d\beta,\qquad K(\beta) = \int_0^\beta H(\beta)d\beta.\end{aligned}\right\} \tag{606}$$

Averaging Eq. (604) for all possible mutual orientations of the two vectors F and v we readily find that

$$\langle\langle|f|^2{}_{F,\,|v|}\rangle\rangle_{Av} = 4ab\left\{\frac{\beta^{1/2}G(\beta)}{H(\beta)}\left(1+\frac{7}{3}k\right)+\frac{g^2}{2ab}\right\}, \tag{607}$$

or, substituting for k and $g^2/2ab$ from (605) we find

$$\langle\langle|f|^2{}_F\rangle\rangle_{Av} = 4ab\left\{\frac{\beta^{1/2}G(\beta)}{H(\beta)}\left(1+\frac{\langle M^{3/2}\rangle_{Av}|v|^2}{\langle M^{3/2}|u|^2\rangle_{Av}}\right)+\frac{5}{12\pi}\frac{\langle M\rangle_{Av}{}^2|v|^2}{\langle M^{1/2}\rangle_{Av}\langle M^{3/2}|u|^2\rangle_{Av}}\right\}. \tag{608}$$

In terms of Eq. (608) we can define an approximate formula for the mean life of the state F according to the equation

$$T_{|F|,\,|v|} = |F|/(\langle\langle|f|^2{}_F\rangle\rangle_{Av})^{1/2}. \tag{609}$$

Combining Eqs. (608) and (609) we find that

$$T_{|F|,\,|v|}=T_{|F|,\,0}\frac{1}{\left[1+\frac{\langle M^{\frac{1}{2}}\rangle_{Av}|v|^2}{\langle M^{\frac{1}{2}}|u|^2\rangle_{Av}}+\frac{5}{12\pi}\frac{\langle M\rangle_{Av}^2|v|^2}{\langle M^{\frac{3}{2}}\rangle_{Av}\langle M^{\frac{1}{2}}|u|^2\rangle_{Av}}\frac{H(\beta)}{\beta^{\frac{1}{2}}G(\beta)}\right]^{\frac{1}{2}}}, \tag{610}$$

where $T_{|F|,\,0}$ denotes the mean life when $|v|=0$:

$$T_{|F|,\,0}=\left[\frac{a^{\frac{1}{2}}}{4b}\frac{\beta^{\frac{3}{2}}H(\beta)}{G(\beta)}\right]^{\frac{1}{2}}. \tag{611}$$

From Eq. (610) we derive that

$$T\propto|F|\quad\text{as}\quad|F|\to 0;\quad T\propto|F|^{-\frac{1}{2}}\quad\text{as}\quad|F|\to\infty; \tag{612}$$

in other words the mean life tends to zero for both weak and strong fields.

I am greatly indebted to Mrs. T. Belland for her assistance in preparing the manuscript for the press. My thanks are also due to Dr. L. R. Henrich for his careful revision of the entire manuscript.

APPENDIXES

I. THE MEAN AND THE MEAN SQUARE DEVIATION OF A BERNOULLI DISTRIBUTION

Consider the Bernoulli distribution

$$w(x)=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}\quad(p<1;\ x\text{ a positive integer}\leqslant n). \tag{613}$$

An alternative form for $w(x)$ is

$$w(x)=C_x^n p^x q^{n-x}, \tag{614}$$

where C_x^n denotes the binomial coefficient and

$$q=1-p. \tag{615}$$

From Eq. (614) it is apparent that $w(x)$ is the coefficient of u^x in the expansion of $(pu+q)^n$:

$$w(x)=\text{coefficient of } u^x \text{ in } (pu+q)^n. \tag{616}$$

That $\sum w_x=1$ follows immediately from this remark:

$$\left.\begin{aligned}\sum_{x=1}^{n} w(x)&=\sum_{x=1}^{n}\text{coefficient of } u^x\text{ in }(pu+q)^n,\\&=[(pu+q)^n]_{u=1}=1.\end{aligned}\right\} \tag{617}$$

Consider now the mean and the mean square deviation of x. By definition

$$\langle x\rangle_{Av}=\sum_{x=1}^{n} xw(x) \tag{618}$$

and

$$\delta^2=\langle(x-\langle x\rangle_{Av})^2\rangle_{Av}=\langle x^2\rangle_{Av}-\langle x\rangle_{Av}^2=\sum_{x=1}^{n} x^2w(x)-\langle x\rangle_{Av}^2. \tag{619}$$

We have

$$\left.\begin{aligned}\langle x\rangle_{Av} &= \sum_{x=1}^{n} x \times \{\text{coefficient of } u^x \text{ in } (pu+q)^n\},\\ &= \sum_{x=1}^{n} \text{coefficient of } u^x \text{ in } \frac{d}{du}(pu+q)^n,\\ &= \left[\frac{d}{du}(pu+q)^n\right]_{u=1} = np(p+q).\end{aligned}\right\} \tag{620}$$

Hence

$$\langle x\rangle_{Av} = np. \tag{621}$$

Similarly,

$$\left.\begin{aligned}\langle x^2\rangle_{Av} &= \sum_{x=1}^{n} x^2 \times \{\text{coefficient of } u^x \text{ in } (pu+q)^n\},\\ &= \sum_{x=1}^{n} \text{coefficient of } u^x \text{ in } \frac{d}{du}\left(u\frac{d}{du}[pu+q]^n\right),\\ &= \left\{\frac{d}{du}\left(u\frac{d}{du}[pu+q]^n\right)\right\}_{u=1},\end{aligned}\right\} \tag{622}$$

or,

$$\langle x^2\rangle_{Av} = np + n(n-1)p^2. \tag{623}$$

Combining Eqs. (619), (621) and (623) we obtain

$$\delta^2 = np - np^2 = np(1-p) = npq. \tag{624}$$

II. A PROBLEM IN PROBABILITY: MULTIVARIATE GAUSSIAN DISTRIBUTIONS

In Chapter I (§4, subsection [a]) we considered the special case of the problem of random flights in which the N displacements which the particle suffers are all governed by Gaussian distributions but with different variances. We shall now consider a generalization of this problem which has important applications to the theory of Brownian motion (see Chapter II, §2, lemma II).

Let

$$\boldsymbol{\Psi} = \sum_{j=1}^{N} \psi_j \mathbf{r}; \quad \boldsymbol{\Phi} = \sum_{j=1}^{N} \phi_j \mathbf{r}, \tag{625}$$

where the ψ_j's and the ϕ_j's are two arbitrary sets of N real numbers each, and where further $\mathbf{r}$ is a stochastic variable the probability distribution of which is governed by

$$\tau(\mathbf{r}) = (1/(2\pi l^2)^{3/2}) \exp(-|\mathbf{r}|^2/2l^2), \tag{626}$$

where l is a constant. We require the probability $W(\boldsymbol{\Psi}, \boldsymbol{\Phi})d\boldsymbol{\Psi}d\boldsymbol{\Phi}$ that $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ shall lie, respectively, in the ranges $(\boldsymbol{\Psi}, \boldsymbol{\Psi}+d\boldsymbol{\Psi})$ and $(\boldsymbol{\Phi}, \boldsymbol{\Phi}+d\boldsymbol{\Phi})$. Applying Markoff's method to this problem, we have [cf. Eqs. (51) and (52)]

$$W(\boldsymbol{\Psi}, \boldsymbol{\Phi}) = \frac{1}{64\pi^6}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp[-i(\boldsymbol{\rho}\cdot\boldsymbol{\Psi}+\boldsymbol{\sigma}\cdot\boldsymbol{\Phi})]A_N(\boldsymbol{\rho}, \boldsymbol{\sigma})d\boldsymbol{\rho}d\boldsymbol{\sigma}, \tag{627}$$

where $\boldsymbol{\rho}$ and $\boldsymbol{\sigma}$ are two auxiliary vectors and

$$A_N(\boldsymbol{\rho}, \boldsymbol{\sigma}) = \prod_{j=1}^{N} \frac{1}{(2\pi l^2)^{3/2}}\int_{-\infty}^{+\infty} \exp[i(\boldsymbol{\rho}\cdot\psi_j\mathbf{r}+\boldsymbol{\sigma}\cdot\phi_j\mathbf{r})]\exp(-|\mathbf{r}|^2/2l^2)d\mathbf{r}. \tag{628}$$

To evaluate $A_N(\varrho, \sigma)$ we need the value of the typical integral

$$J=\frac{1}{(2\pi l^2)^{\frac{3}{2}}}\int_{-\infty}^{+\infty}\exp\left[i\mathbf{r}\cdot(\psi_j\varrho+\phi_j\sigma)-(|\mathbf{r}|^2/2l^2)\right]d\mathbf{r}. \tag{629}$$

We have

$$\left.\begin{aligned} J&=\prod_{x,y,z}\frac{1}{(2\pi l^2)^{\frac{1}{2}}}\int_{-\infty}^{+\infty}\exp\left\{-\left[x^2+2il^2x(\rho_1\psi_j+\sigma_1\phi_j)\right]/2l^2\right\}dx,\\ &=\exp\left\{-l^2\left[(\rho_1\psi_j+\sigma_1\phi_j)^2+(\rho_2\psi_j+\sigma_2\phi_j)^2+(\rho_3\psi_j+\sigma_3\phi_j)^2\right]/2\right\}.\end{aligned}\right\} \tag{630}$$

Hence

$$\left.\begin{aligned} A_N(\varrho, \sigma)&=\exp\left\{-l^2\sum_{j=1}^{N}\left[(\rho_1\psi_j+\sigma_1\phi_j)^2+(\rho_2\psi_j+\sigma_2\phi_j)^2+(\rho_3\psi_j+\sigma_3\phi_j)^2\right]/2\right\}\\ &=\exp\left[-(P|\varrho|^2+2R\varrho\cdot\sigma+Q|\sigma|^2)/2\right],\end{aligned}\right\} \tag{631}$$

where we have written

$$P=l^2\sum_{j=1}^{N}\psi_j^2;\quad R=l^2\sum_{j=1}^{N}\phi_j\psi_j;\quad Q=l^2\sum_{j=1}^{N}\phi_j^2. \tag{632}$$

Substituting for $A_N(\varrho, \sigma)$ from Eq. (632) in the formula for $W(\Psi, \Phi)$ [Eq. (627)] we obtain

$$W(\Psi, \Phi)=\frac{1}{64\pi^6}\prod_{i=1}^{3}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\exp\left\{-\left[P\rho_i^2+2R\rho_i\sigma_i+Q\sigma_i^2+2i(\rho_i\Psi_i+\sigma_i\Phi_i)\right]/2\right\}d\rho_id\sigma_i. \tag{633}$$

To evaluate the integrals occurring in the foregoing formula, we first perform a translation of the coordinate system according to

$$\rho_i=\xi_i+\alpha_i;\quad \sigma_i=\eta_i+\beta_i\quad (i=1, 2, 3), \tag{634}$$

where α_i and β_i are so chosen that

$$P\alpha_i+R\beta_i=-i\Psi_i;\quad R\alpha_i+Q\beta_i=-i\Phi_i\quad (i=1, 2, 3). \tag{635}$$

With this transformation of the variables we have

$$\left.\begin{aligned} P\rho_i^2+2R\rho_i\sigma_i+Q\sigma_i^2+2i(\rho_i\Psi_i+\sigma_i\Phi_i)&=P\xi_i^2+2R\xi_i\eta_i+Q\eta_i^2+i(\alpha_i\Psi_i+\beta_i\Phi_i),\\ &=P\xi_i^2+2R\xi_i\eta_i+Q\eta_i^2+\frac{1}{PQ-R^2}(P\Phi_i^2-2R\Phi_i\Psi_i+Q\Psi_i^2).\end{aligned}\right\} \tag{636}$$

Hence,

$$W(\Psi, \Phi)=\frac{1}{64\pi^6}\prod_{i=1}^{3}\exp\left[-(P\Phi_i^2-2R\Phi_i\Psi_i+Q\Psi_i^2)/2(PQ-R^2)\right]$$
$$\times\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\exp\left[-(P\xi_i^2+2R\xi_i\eta_i+Q\eta_i^2)/2\right]d\xi_id\eta_i. \tag{637}$$

From this equation we readily find that

$$W(\Psi, \Phi)=\left[1/8\pi^3(PQ-R^2)^{\frac{3}{2}}\right]\exp\left[-(P|\Phi|^2-2R\Psi\cdot\Phi+Q|\Psi|^2)/2(PQ-R^2)\right], \tag{638}$$

which gives the required probability distribution.

III. THE POISSON DISTRIBUTION AS THE LAW OF DENSITY FLUCTUATIONS

Consider an element of volume v which is a part of a larger volume V. Let there be N particles distributed in a random fashion inside the volume V. Under these conditions the probability that a particular particle will be found in the element of volume v is clearly v/V; similarly, the probability that it will *not* be found inside v is $(V-v)/V$. Hence, the probability $W_N(n)$ that *some* n particles will be found inside v is given by the Bernoulli distribution

$$W_N(n)=\frac{N!}{n!(N-n)!}\left(\frac{v}{V}\right)^n\left(1-\frac{v}{V}\right)^{N-n} \tag{639}$$

The average value of n is therefore given by [cf. Eq. (621)]

$$\langle n\rangle_{\text{Av}}=N(v/V)=\nu \quad \text{(say)}. \tag{640}$$

In terms of ν Eq. (639) can be expressed in the form

$$W_N(n)=\frac{N!}{n!(N-n)!}\left(\frac{\nu}{N}\right)^n\left(1-\frac{\nu}{N}\right)^{N-n}. \tag{641}$$

The case of greatest practical interest arises when both N and V tend to infinity but in such a way that ν remains constant [see Eq. (640)]. To obtain the corresponding limiting form of the distribution (641) we first rewrite it as

$$\left.\begin{aligned} W_N(n)&=\frac{1}{n!}N(N-1)(N-2)\cdots(N-n+1)\left(\frac{\nu}{N}\right)^n\left(1-\frac{\nu}{N}\right)^{N-n},\\ &=\frac{\nu^n}{n!}1\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)\left(1-\frac{\nu}{N}\right)^{N-n}, \end{aligned}\right\} \tag{642}$$

and then let $N\rightarrow\infty$ keeping both ν and n fixed. We have

$$\left.\begin{aligned} W(n)&=\lim_{N\to\infty} W_N(n),\\ &=\frac{\nu^n}{n!}\lim_{N\to\infty}\left\{\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\cdots\left(1-\frac{n-1}{N}\right)\left(1-\frac{\nu}{N}\right)^{N-n}\right\},\\ &=\frac{\nu^n}{n!}\lim_{N\to\infty}\left(1-\frac{\nu}{N}\right)^N. \end{aligned}\right\} \tag{643}$$

Hence,

$$W(n)=\nu^n e^{-\nu}/n!, \tag{644}$$

which is the required Poisson distribution.

In some applications of Eq. (644) (e.g., III, §3) ν is a very large number; and when this is the case, interest is attached to only those values of n which are relatively close to ν. We shall now show that under these conditions the Poisson distribution specializes still further to a Gaussian distribution.

Rewriting Eq. (644) in the form

$$\log W(n)=n\log\nu-\nu-\log n! \tag{645}$$

and adopting Stirling's approximation for $\log n$ [cf. Eq. (7)] we obtain

$$\log W(n)=n\log\nu-\nu-(n+\tfrac{1}{2})\log n+n-\tfrac{1}{2}\log 2\pi+O(n^{-1}). \tag{646}$$

Let

$$n=\nu+\delta. \tag{647}$$

Equation (646) becomes

$$\log W(n)=-(\nu+\delta+\tfrac{1}{2})\log\left(1+\frac{\delta}{\nu}\right)+\delta-\tfrac{1}{2}\log(2\pi\nu)+O(n^{-1}). \tag{648}$$

If we now suppose that $\delta/\nu \ll 1$ we can expand the logarithmic term in Eq. (648) as a power series in δ/ν. Retaining only the dominant term, we find

$$\log W(n) = -(\delta^2/2\nu) - \tfrac{1}{2}\log(2\pi\nu) \quad (\nu \to \infty\,;\ \delta/\nu \to 0). \tag{649}$$

Thus,

$$W(n) = [1/(2\pi\nu)^{\frac{1}{2}}] \exp[-(n-\nu)^2/2\nu], \tag{650}$$

which is the required Gaussian form.

IV. THE MEAN AND THE MEAN SQUARE DEVIATION OF THE SUM OF TWO PROBABILITY DISTRIBUTIONS

Let $w_1(x)$ and $w_2(y)$ represent two probability distributions. For the sake of definiteness we shall suppose that x and y take on only discrete values. A probability distribution which is said to be the *sum* of the two distributions is defined by

$$w(z) = \sum_{x+y=z} w_1(x)w_2(y), \tag{651}$$

where in the summation on the right-hand side we include all pairs of values of x and y (each in their respective domains) which satisfy the relation $x+y=z$. We may first verify that $w(z)$ defined according to Eq. (651) does in fact represent a probability distribution. To see this we have only to show that $\sum w(z) = 1$. Now,

$$\sum_z w(z) = \sum_z \sum_{x+y=z} w_1(x)w_2(y); \tag{652}$$

accordingly, in the summation on the right-hand side, x and y can now run through their respective ranges of values *independently* of each other. Hence,

$$\sum_z w(z) = [\sum_x w_1(x)][\sum_y w_2(y)] = 1. \tag{653}$$

We shall now prove that *the mean and the mean square deviation of the sum of two probability distributions is the sum of the means and the mean square deviations of the component distributions.*

To prove this theorem, we observe that by definitions

$$\langle z\rangle_{Av} = \sum_z z w(z) = \sum_z \sum_{x+y=z} (x+y)w_1(x)w_2(y), \tag{654}$$

or

$$\langle z\rangle_{Av} = \sum_x \sum_y [x w_1(x)w_2(y) + y w_1(x)w_2(y)], \tag{655}$$

where in the summations on the right-hand side we can again let x and y run their respective ranges of values independently of each other. Hence,

$$\langle z\rangle_{Av} = [\sum_x x w_1(x)][\sum_y w_2(y)] + [\sum_x w_1(x)][\sum_y y w_2(y)], \tag{656}$$

or

$$\langle z\rangle_{Av} = \langle x\rangle_{Av} + \langle y\rangle_{Av}. \tag{657}$$

Similarly,

$$\left.\begin{aligned}
\langle (z-\langle z\rangle_{Av})^2\rangle_{Av} &= \sum_z (z-\langle z\rangle_{Av})^2 w(z), \\
&= \sum_z \sum_{x+y=z} (x+y-\langle x\rangle_{Av}-\langle y\rangle_{Av})^2 w_1(x)w_2(y), \\
&= \sum_x \sum_y [(x-\langle x\rangle_{Av})^2 + 2(x-\langle x\rangle_{Av})(y-\langle y\rangle_{Av}) + (y-\langle y\rangle_{Av})^2] w_1(x)w_2(y), \\
&= [\sum_x (x-\langle x\rangle_{Av})^2 w_1(x)][\sum_y w_2(y)] + [\sum_x w_1(x)][\sum_y (y-\langle y\rangle_{Av})^2 w_2(y)] \\
&\qquad + 2[\sum_x (x-\langle x\rangle_{Av}) w_1(x)][\sum_y (y-\langle y\rangle_{Av}) w_2(y)].
\end{aligned}\right\} \tag{658}$$

Hence,

$$\langle(z-\langle z\rangle_{\text{Av}})^2\rangle_{\text{Av}}=\langle(x-\langle x\rangle_{\text{Av}})^2\rangle_{\text{Av}}+\langle(y-\langle y\rangle_{\text{Av}})^2\rangle_{\text{Av}}. \tag{659}$$

The theorem is now proved.

The extension of the foregoing results to include the case when x and y are continuously variable is, of course, obvious. Similarly the definitions and results can be further extended to include the sums of more than two probability distributions.

V. ZERMELO'S PROOF OF POINCARÉ'S THEOREM CONCERNING THE QUASI-PERIODIC CHARACTER OF THE MOTIONS OF A CONSERVATIVE DYNAMICAL SYSTEM

Consider a conservative dynamical system of n degrees of freedom and which is described by a Hamiltonian function H of the generalized coordinates $q_1, \cdots, q_n$ and momenta $p_1, \cdots, p_n$. The state of such a dynamical system can be represented by a point in the $2n$ dimensional phase space of the q's and p's. Similarly, the trajectory described by the representative point will describe the evolution of the dynamical system.

Through each point in the phase space there passes a unique trajectory which can be derived from the canonical equations of motion

$$\dot{q}_s=\frac{\partial H}{\partial p_s};\quad \dot{p}_s=-\frac{\partial H}{\partial q_s}\quad (s=1,\cdots,n). \tag{660}$$

More generally, consider any arbitrary continuous domain of points g_0 (of finite measure) in the phase space. Let the points g_0 be the representatives at time $t=0$ of an ensemble of dynamical systems all described by the same Hamiltonian function $H(p_1,\cdots,p_n;q_1,\cdots,q_n)$. At a later time t the representatives of the ensemble will occupy a continuous domain of points g_t which can be obtained by tracing through each point of g_0 the corresponding trajectory and following along the various trajectories for a time t. Because of the uniqueness, in general, of the trajectories passing through a given point in the phase space, the construction of the domain g_t from an initial domain g_0 is a unique process. We shall accordingly refer to g_t as the *future phase* (at time t) of the *initial phase* g_0 (at time $t=0$) of the given dynamical system.

Now, according to Liouville's theorem of classical dynamics, the density of any element of phase space remains constant during its motion according to the canonical Eqs. (660). Hence, if ω_t denotes the volume extension of the domain of points g_t introduced in the preceding paragraph, it follows from Liouville's theorem that ω_t remains constant as t varies.

We have already described how from an initial phase g_0 we can derive the future phase g_t at time t. The domain of points g_0 together with *all* its future phases g_t, $(0<t<\infty)$ clearly form a continuous domain of points which we shall denote by Γ_0: Γ_0 is accordingly the class of all states which at some finite past occupied states belonging to g_0. The extension of Γ_0 will be finite if we are considering a dynamical system which is enclosed—for, then, none of the coordinates or momenta can take on infinite values and the entire accessible region of the phase space remains finite. We shall suppose that this is the case and denote by Ω_0 the extension of Γ_0. Clearly $\Omega_0\geqslant\omega_0$. In a similar manner we can, quite generally, define the domain of points Γ_t which includes all the future phases of g_t. Let Ω_t denote the extension of Γ_t. It is evident that

$$\Omega_{t_1}\geqslant\Omega_{t_2}\quad\text{whenever}\quad t_1<t_2. \tag{661}$$

For, Ω_{t_1} denoting the extension of *all* the future phases of g_{t_1} must therefore necessarily include also the future phases of g_{t_2} if $t_1<t_2$. On the other hand, considering Γ_0 itself as a domain of points, we can construct *its* future phases in exactly the same way as the future phases g_t of g_0 were constructed. But the future phase of Γ_0 after a time t is clearly Γ_t. And therefore applying Liouville's theorem to the domain Γ_0 and its future phases Γ_t we conclude that

$$\Omega_t=\text{constant}. \tag{662}$$

Comparing this result with the inequality (661) we infer that *the domain of points Γ_t can differ from Γ_0 by at most a set of points of measure zero.* Hence, the future phases of g_t($t>0$ but arbitrary otherwise) must include g_0 apart, possibly, from a set of points of measure zero. But the points of g_t are themselves future phases of the points of g_0. Hence, the states belonging to g_0 (again, with the possible exception of a set of zero measure) must recur after the elapse of a sufficient length of time; and this is true no matter how small the extension ω_0 of g_0 is, provided it is only finite. From this, the deduction of Poincaré's theorem is immediate. (For a formal statement of Poincaré's theorem see Chapter III, §4).

VI. BOLTZMANN'S ESTIMATE OF THE PERIOD OF A POINCARÉ CYCLE

To estimate the order of magnitude of the period of a Poincaré cycle, Boltzmann has considered the following typical example:

A cubic centimeter of air containing 10^{18} molecules is considered in which all the molecules are initially supposed to have a speed of 500 meters per second. With a concentration of 10^{18} molecules, the average distance between the neighboring ones is of the order of 10^{-6} cm. Also, under normal conditions, each molecule will suffer something like 4×10^{9} collisions per second so that on the whole there will occur

$$b=2\times 10^{27} \text{ collisions per second.} \tag{663}$$

Since Poincaré's theorem asserts only the quasi-periodic character of the motions (see Chapter III, §4 and Appendix V) the period to be estimated clearly depends on the closeness to which we require the initial conditions to recur. For the case under discussion Boltzmann supposes that a molecule can be said to have approximately returned to its initial state if the differences in position (x, y, z) and velocity (u, v, w) in the initial and the final states are such that

$$|\Delta x|,\ |\Delta y|,\ |\Delta z| \leqslant 10^{-7} \text{ cm}, \tag{664}$$

and

$$|\Delta u|,\ |\Delta v|,\ |\Delta w| \leqslant 1 \text{ m/sec}. \tag{665}$$

In other words, we shall require the positions to agree to within 10 percent of the average distance between the molecules and the velocities to agree within one part in 500.

We shall first estimate the order of magnitude of the time required for the recurrence of an initial "abnormal" distribution in the velocities. According to Poincaré's theorem, an initial state need not recur earlier than the time necessary for all the molecules to take on all the possible values for the velocity. We can readily determine the number N of such possibilities with the understanding that we agree to distinguish between two velocities only if at least one of the components differ by more than 1 m/sec.

The first molecule can have all velocities ranging from zero to $a=500\times 10^{9}$ m/sec.—since we have supposed that in the initial state all the molecules have the same speed of 500 m/sec and that there are 10^{18} molecules in the system. Again, if the first molecule has a speed v_1 the second one can have speeds only in range 0 to $(a^2-v_1^2)^{\frac{1}{2}}$. Similarly, if the first and the second molecules have speeds v_1 and v_2, respectively, the third molecule can have speeds only in the range 0 to $(a^2-v_1^2-v_2^2)^{\frac{1}{2}}$; and so on. Accordingly, the required number of combinations N is

$$\left.\begin{aligned} N &= (4\pi)^{n-1}\int_0^a dv_1 v_1^2 \int_0^{(a^2-v_1^2)^{\frac{1}{2}}} dv_2 v_2^2 \int_0^{(a^2-v_1^2-v_2^2)^{\frac{1}{2}}} dv_3 v_3^2 \cdots \int_0^{(a^2-v_1^2-\cdots-v_{n-2}^2)^{\frac{1}{2}}} dv_{n-1} v_{n-1}^2, \\ &= (\pi^{(3n-3)/2}/2\cdot 3\cdot 4\cdots[3(n-1)/2])a^{3(n-1)} \quad (n, \text{ odd}), \\ &= (2(2\pi)^{(3n-4)/2}/3\cdot 5\cdot 7\cdots 3(n-1))a^{3(n-1)} \quad (n, \text{ even}), \end{aligned}\right\} \tag{666}$$

where

$$a=500\times 10^{9} \quad \text{and} \quad n=10^{18}. \tag{667}$$

Since each of these N combinations occurs on the average in a time $1/b$ seconds [cf. Eq. (663)] the total time required for the velocities to run through all the possible values is

$$N/b. \tag{668}$$

After this length of time we may expect the initial distribution of the velocities to recur to within the limits of accuracy specified except for one single molecule the direction of whose motion has been left unrestricted. On the other hand we have still left unspecified the positions of the centers of gravity of all the molecules. But in order that we may say that the initial state has recurred to a sufficient approximation, we must require the positions of the molecules in the final state also to agree with the initial values to some stated degree of accuracy. This would clearly require the time (668) to be multiplied by another number of order similar to N. However, the extremely large value already of N/b gives some indication of the enormous times which are involved Moreover, comparing these times with the time of relaxation of a gas which is of the order 10^{-8} second under normal conditions, we get an idea as to how extremely small the fraction of the total number of complexions is for which appreciable departures from a Maxwellian distribution occur. (For a further discussion of these and related matters see Chapter III, §4.)

VII. THE LAW OF DISTRIBUTION OF THE NEAREST NEIGHBOR IN A RANDOM DISTRIBUTION OF PARTICLES

This problem was first considered by Hertz (see reference 71 in the Bibliographical Notes for Chapter IV).

Let $w(r)dr$ denote the probability that the nearest neighbor to a particle occurs between r and $r+dr$. This probability must be clearly equal to the probability that no particles exist interior to r times the probability that a particle does exist in the spherical shell between r and $r+dr$. Accordingly, the function $w(r)$ must satisfy the relation

$$w(r)=\left[1-\int_0^r w(r)dr\right]4\pi r^2 n, \tag{669}$$

where n denotes the average number of particles per unit volume. From Eq. (669) we derive:

$$\frac{d}{dr}\left[\frac{w(r)}{4\pi r^2 n}\right]=-4\pi r^2 n\frac{w(r)}{4\pi r^2 n}. \tag{670}$$

Hence

$$w(r)=\exp\left(-4\pi r^3 n/3\right)4\pi r^2 n, \tag{671}$$

since, according to Eq. (669)

$$w(r)\rightarrow 4\pi r^2 n \quad \text{as} \quad r\rightarrow 0. \tag{672}$$

Equation (671) gives then the required law of distribution of the nearest neighbor.

Using the distribution (671) we can derive an *exact* formula for the "average distance" D between the particles. For, by definition

$$D=\int_0^\infty r w(r)dr, \tag{673}$$

or, if we use Eq. (671)

$$D=\int_0^\infty \exp\left(-4\pi r^3 n/3\right)4\pi r^3 n dr. \tag{674}$$

After some elementary reductions, Eq. (674) becomes

$$\left.\begin{aligned} D&=\frac{1}{(4\pi n/3)^{1/3}}\int_0^\infty e^{-x}x^{1/3}dx,\\ &=\Gamma(4/3)/(4\pi n/3)^{1/3}.\end{aligned}\right\} \quad (675)$$

Substituting for $\Gamma(4/3)$, we find

$$D=0.55396n^{-1/3}. \quad (676)$$

BIBLIOGRAPHICAL NOTES

Chapter I

§1.—We may briefly record here the history of the problem of random flights considered in this chapter:

Karl Pearson appears to have been the first to explicitly formulate a problem of this general type:

1. K. Pearson, Nature **77**, 294 (1905). Pearson's formulation of the problem was in the following terms: "A man starts from a point O and walks l yards in a straight line; he then turns through any angle whatever and walks another l yards in a second straight line. He repeats this process n times. I require the probability that after these n stretches he is at a distance between r and $r+dr$ from his starting point O." After Pearson had formulated this problem Lord Rayleigh pointed out that the problem is formally "the same as that of the composition of n isoperiodic vibrations of unit amplitude and of phases distributed at random" which he had considered as early as in 1880:

2. Lord Rayleigh, Phil. Mag. **10**, 73 (1880); see also ibid. **47**, 246 (1889). These papers are reprinted in *Scientific Papers of Lord Rayleigh*, Vol. I, p. 491, and Vol. IV, p. 370. In the foregoing papers Rayleigh obtains the asymptotic form of the solution as $n\to\infty$. But for finite values of n the general solution of Pearson's problem was given by

3. J. C. Kluyver, Konink. Akad. Wetenschap. Amsterdam **14**, 325 (1905). The general solution of the problem of random walk in one dimension was obtained by Smoluchowski apparently independently of the earlier investigators.

4. M. von Smoluchowski, Bull. Acad. Cracovie, p. 203 (1906). In its most general form the problem of random flights was formulated by A. A. Markoff who also outlined the method for obtaining the general solution.

5. A. A. Markoff, *Wahrscheinlichkeitsrechnung* (Leipzig, 1912), §§16 and 33.

§2.—The problem of the random walk with reflecting and absorbing barriers was first considered by Smoluchowski:

6. M. v. Smoluchowski, (a) Wien Ber. **124**, 263 (1915); also (b) "Drei Vortrage uber Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen," Physik. Zeits. **17**, 557, 585 (1916). See also

7. R. von Mises, *Wahrscheinlichkeitsrechnung* (Leipzig and Wien), pp. 479–518.

§3.—Markoff's method described in this section is a somewhat generalized version of what is given in Markoff (reference 5). See also

8. M. von Laue, Ann. d. Physik **47**, 853 (1915).

§4.—See A. A. Markoff (reference 5). The case of finite N considered in subsection (b) follows the treatment of

9. Lord Rayleigh, Phil. Mag. **37**, 321 (1919) (or *Scientific Papers*, Vol. VI, p. 604).

§5.—The passage to a differential equation for the case of the one-dimensional problem of the random walk was achieved by Rayleigh:

10. Lord Rayleigh, Phil. Mag. **47**, 246 (1899) (or *Scientific Papers*, Vol. IV, p. 370). See also Smoluchowski (reference 6). But the general treatment given in this section appears to be new.

We may also note the following further reference:

11. W. H. McCrea, Proc. Roy. Soc. Edinburgh **60**, 281 (1939).

Chapter II

The following general references may be noted.

12. The Svedberg, *Die Existenz der Molekule* (Leipzig, 1912).

13. G. L. de Haas-Lorentz, *Die Brownsche-Bewegung und einige verwandte Erscheinungen*, (Braunschweig, 1913).

14. M. v. Smoluchowski, see reference 6(b).

15. J. Perrin, *Atoms* (Constable, London, 1916).

16. R. Fürth, *Schwankungserscheinungen in der Physik* (Sammlung Vieweg, Braunschweig, 1920), Vol. 48.

§1.—As is well known the modern theory of Brownian motion was initiated by Einstein and Smoluchowski:

17. A. Einstein, Ann. d. Physik **17**, 549 (1905); also, ibid. **19**, 371 (1906).

18. M. v. Smoluchowski, Ann. d. Physik **21**, 756 (1906). In Einstein's and in Smoluchowski's treatment of the problem, Brownian motion is idealized as a problem in random flights; but as we have seen, this idealization is valid only when we ignore effects which occur in time intervals of order β^{-1}. For the general treatment of the problem we require to base our discussion on an equation of the type first introduced by Langevin:

19. P. Langevin, Comptes rendus **146**, 530 (1908). In this connection see

20. F. Zernike, *Handbuch der Physik* (Berlin, 1928), Vol. 3, p. 456.

§2.—The treatment of the Brownian motion of a free particle given in this section is derived from:

21. L. S. Ornstein and W. R. van Wijk, Physica **1**, 235 (1933). See also

22. W. R. van Wijk, Physica **3**, 1111 (1936). Earlier, but somewhat less general treatment along the same lines is contained in

23. G. E. Uhlenbeck and L. S. Ornstein, Phys. Rev. **36**, 823 (1930). In the foregoing papers the discussion has been carried out only for the case of one-dimensional motion. In the text we have treated the general three-dimensional problem; further, the arguments in references 21 and 22 have been rearranged considerably to make the presentation more direct and straightforward.

§3.—See Ornstein and Wijk (reference 21); also

24. G. E. Uhlenbeck and S. Goudsmidt, Phys. Rev. **34**, 145 (1929).

25. G. A. van Lear and G. E. Uhlenbeck, Phys. Rev. **38**, 1583 (1931).

§4.—The passage to a differential equation for the description of the Brownian motion of a free particle in the velocity space was achieved by

26. A. D. Fokker, Ann. d. Physik **43**, 812 (1914). A more general discussion of this problem is due to

27. M. Planck, Sitz. der preuss. Akad. p. 324 (1917). See also references 21 and 23; further,

28. Lord Rayleigh, *Scientific Papers*, Vol. III, p. 473.

29. L. S. Ornstein, Versl. Acad. Amst. **26**, 1005 (1917); also Konink. Akad. Wetenschap. Amsterdam **20**, 96 (1917).

30. H. C. Burger, Versl. Acad. Amst. **25**, 1482 (1917).

31. L. S. Ornstein and H. C. Burger, Versl. Acad. Amst. **27**, 1146 (1919); **28**, 183 (1919); also Konink. Akad. Wetenschap. Amsterdam **21**, 922 (1918).

Earlier attempts to generalize Liouville's equation of classical dynamics to include Brownian motion are contained in

32. O. Klein, Arkiv for Matematik, Astronomi, och Fysik **16**, No. 5 (1921); and

33. H. A. Kramers, Physica **7**, 284 (1940).

The passage to a differential equation in configuration space was first achieved by

34. M. v. Smoluchowski, Ann. d. Physik **48**, 1103 (1915); see also,

35. R. Fürth, Ann. d. Physik **53**, 177 (1917).

In the text the discussion of the various differential equations has been carried out more generally and more completely than in the references given above; this is particularly true of the discussion relating to the generalization of the Liouville equation of classical dynamics (subsections, ii–v).

§5.—See H. A. Kramers (reference 33).

Approaches to the problem of the Brownian motion somewhat different to the one we have adopted are contained in

36. G. Krutkov, Physik. Zeits. der Sowjetunion **5**, 287 (1934). See also the various articles by the same author in C. R. Acad. Sci. USSR during the years (1934) and (1935).

37. S. Bernstein, C. R. Acad. Sci. USSR, p. 1 (1934), and p. 361 (1934). A more particularly mathematical discussion of the problems of Brownian motion has been given by

38. J. L. Doob, Ann. Math. **43**, 351 (1942); see also the references given in this paper.

Chapter III

The following general references may be noted.

39. M. v. Smoluchowski, reference 6(b).

40. A. Sommerfeld, "Zum Andenken an Marian von Smoluchowski," Physik. Zeits. **18**, 533 (1917).

41. R. Fürth, Physik. Zeits. **20**, 303, 332, 350, 375 (1919); also reference 16.

42. H. Freundlich, Kapillarchemie (Leipzig, 1930–1932), Vols. I and II; see particularly pp. 485–510 in Vol. I and pp. 140–162 in Vol. II.

43. The Svedberg, *Die Existenz der Molekule* (Leipzig, 1912).

In reference 39 we have an extremely valuable account of the entire subject of Brownian motion and molecular fluctuations; there exists no better introduction to this subject than these lectures of Smoluchowski. In reference 40 Sommerfeld gives a fairly extensive bibliography of Smoluchowski's writings.

§1.—The theory of density fluctuations as developed by Smoluchowski represents one of the most outstanding achievements in molecular physics. Not only does it quantitatively account for and clarify a wide range of physical and physico-chemical phenomena, it also introduces such fundamental notions as the "probability after-effect" which are of very great significance in other connections (see Chapter IV).

44. M. v. Smoluchowski, Wien. Ber. **123**, 2381 (1914); see also Physik. Zeits. **16**, 321 (1915) and Kolloid Zeits. **18**, 48 (1916). For discussions of the problem of density fluctuations prior to the introduction of the notion of the "speed of fluctuations" see

45. M. v. Smoluchowski, *Boltzmann Festschrift* (1904), p. 626; *Bull. Acad. Cracovie*, p. 1057, 1907; Ann. d. Physik **25**, 205 (1908). Also

46. R. Lorenz and W. Eitel, Zeits. f. physik. Chemie **87**, 293, 434 (1914).

It is of some interest to recall that referring to his deviation of the formulae for $\langle\Delta_n\rangle_{Av}$ and $\langle\Delta_n^2\rangle_{Av}$ [Eqs. (356) and (358)] Smoluchowski says, "Aus diesem komplizierten Formeln [referring to the formula for $W(n; m)$] lassen sich mittels verwickelter summationen merkwurdigerweise recht einfache resultate fur die durschschnittliche Änderung der Teilchenzahl ableiten. . . . So wie fur das Anderungsquadrat bei unbestimmter Anfangszahl n [Eq. (363)]." This led to some heated discussion whether these formulae cannot be derived more simply; for example, see

47. L. S. Ornstein, Konink. Akad. Wetenschap. Amsterdam **21**, 92 (1917). But neither Ornstein nor Smoluchowski seems to have noticed that the formulae for $\langle\Delta_n\rangle_{Av}$ and $\langle\Delta_n^2\rangle_{Av}$ can be derived very directly from the fact that the transition probability $W(n; m)$ is the sum (in a technical sense) of a Bernoulli and a Poisson dis-

tribution; it is to this fact that the simplicity of the results are due.

§2.—Comparisons between the predictions of his theory with the data of colloid statistics were first made by Smoluchowski himself (reference 44). The experiments which were used for these first comparisons were those of

48. The Svedberg, Zeits. f. physik. Chemie **77**, 147 (1911); see also references 43 and 46. But precision experiments carried out with expressed intention of verifying Smoluchowski's theory are those of

49. A. Westgren, Arkiv for Matematik, Astronomi, och Fysik **11**, Nos. 8 and 14 (1916) and **13**, No. 14 (1918).

An interesting application of Smoluchowski's theory to a problem of rather different sort has been made by Fürth:

50. R. Fürth, Physik. Zeits. **19**, 421 (1918); **20**, 21 (1919). Fürth made systematic counts of the number of pedestrians in a block every five seconds. This interval of five seconds was chosen because the length of the block was such that a pedestrian observed in the block on one occasion has an appreciable probability of remaining in the same block when the next observation is made five seconds later. We can, accordingly, define a probability after-effect factor P ($=v\tau/a$, where v is the average speed of a pedestrian, τ the chosen interval of time and a the length of the block), and Smoluchowski's theory applies. A statistical analysis of this data showed that the agreement with the theory is excellent. It is amusing that by systematic counts of the kind made by Fürth it is possible actually to determine the average speed of a pedestrian!

§3.—The theory outlined in this section is derived from

51. M. v. Smoluchowski, Wien. Ber. **124**, 339 (1915); see also references 39 and 41.

§4.—Among the early discussions on the compatibility between the notions of conventional thermodynamics and the then new standpoint of the kinetic molecular theory, we may refer to

52. J. Loschmidt, Wien. Ber. **73**, 139 (1876); **75**, 67 (1877).

53. L. Boltzmann, Wien. Ber. **75**, 62 (1877); **76**, 373 (1877); also Nature **51**, 413 (1895) and *Vorlesungen über Gas Theorie* (Leipzig, 1895) Vol. I, p. 42 (or the reprinted edition of 1923).

54. E. Zermelo, Ann. d. Physik **57**, 485 (1896); **59**, 793 (1896).

55. L. Boltzmann, Ann. d. Physik **57**, 773 (1896); **60**, 392 (1897).

Smoluchowski's fundamental discussions of the limits of validity of the second law of thermodynamics are contained in

56. M. v. Smoluchowski, Physik. Zeits. **13**, 1069 (1912); **14**, 261 (1913). See also references 39 and 51.

It is somewhat disappointing that the more recent discussions of the laws of thermodynamics contain no relevant references to the investigations of Boltzmann and Smoluchowski [e.g., P. W. Bridgman, *The Nature of Thermodynamics* (Harvard University Press, 1941)]. The absence of references, particularly to Smoluchowski, is to be deplored since no one has contributed so much as Smoluchowski to a real clarification of the fundamental issues involved.

For an exhaustive discussion of the foundations of statistical mechanics, see

57. P. and T. Ehrenfest, *Begriffliche Grundlagen der Statistischen Auffassung in der Mechanik, Encyklopadie der Mathematischen Wissenschaften* (1911), Vol. 4, p. 4. And for Carathéodory's version of thermodynamics see

57a. S. Chandrasekhar, *An Introduction to the Study of Stellar Structure* (University of Chicago Press, 1939), Chap. I, pp. 11–37.

§5.—See Smoluchowski, reference 39; also

58. M. v. Smoluchowski, Ann. d. Physik **48**, 1103 (1915).

59. R. Fürth, Ann. d. Physik **53**, 177 (1917).

§6.—See Smoluchowski reference 39; also

60. M. v. Smoluchowski, Zeits. f. physik. Chemie **92**, 129 (1917).

61. R. Zsigmondy, Zeits. f. physik. Chemie **92**, 600 (1917). The papers 60 and 61 contain references to the earlier literature on the subject of coagulation. For the more recent literature see Freundlich (reference 42, particularly Vol. II, pp. 140–162).

§7.—See

62. H. A. Kramers, Physica **7**, 284 (1940). Also,

63. H. Pelzer and E. Wigner, Zeits. f. physik. Chemie, **B15**, 445 (1932).

An aspect of the theory of Brownian motion we have not touched upon concerns the natural limit set by it to all measuring processes. But an excellent review of this entire field exists:

64. R. B. Barnes and S. Silverman, Rev. Mod. Phys. **6**, 162 (1934).

Chapter IV

The ideas developed in this chapter are in the main taken from

65. S. Chandrasekhar, Astrophys. J. **94**, 511 (1941).

66. S. Chandrasekhar and J. von Neumann, Astrophys. J. **95**, 489 (1942).

67. S. Chandrasekhar and J. von Neumann, Astrophys. J. **97**, 1, (1943).

§1.—See references 65, 66, and 67; also

68. S. Chandrasekhar, *Principles of Stellar Dynamics* (University of Chicago Press, 1942), Chapters II and V.

§2.—The problem considered in this section is clearly equivalent to finding the probability of a given *electric* field strength at a point in a gas composed of simple ions. This latter problem was first considered by Holtsmark:

69. J. Holtsmark, Ann. d. Physik **58**, 577 (1919); also Physik. Zeits. **20**, 162 (1919) and **25**, 73 (1924). Among other papers on related subjects we may refer to

70. R. Gans, Ann. d. Physik **66**, 396 (1921).

71. P. Hertz, Math. Ann. **67**, 387 (1909).

72. R. Gans, Physik. Zeits. **23**, 109 (1922).

73. C. V. Raman, Phil. Mag. **47**, 671 (1924).

§3.—See references 66 and 67. See also three further papers on "Dynamical Friction" by Chandrasekhar in forthcoming issues of *The Astrophysical Journal* where further applications of the Fokker-Planck equation will be found.

A STATISTICAL THEORY OF STELLAR ENCOUNTERS

S. CHANDRASEKHAR

ABSTRACT

In this paper the principles of a statistical theory of stellar encounters are developed. The fundamental idea of this new method is to describe the fluctuating part of the gravitational field acting on a star in terms of two functions: a function $W(F)$, which gives the probability of occurrence of a field strength F, and a function $T(F)$ which gives the average time during which the field strength F acts. With regard to $W(F)$ it is shown that a probability-distribution function derived by Holtsmark to describe the interionic fields in a discharge tube can be adapted to suit the gravitational case. In a certain approximation this probability of a given field is directly related to the probability of finding the nearest neighbor to a given star at some prescribed distance. In this latter approximation the mean life of the state F can be obtained by using a formula due to Smoluchowski in the theory of Brownian motion. In terms of these functions $W(F)$ and $T(F)$ the probable accelerations which a star will undergo can be determined according to the principles of the theory of random walk.

As an application of the methods of this statistical theory, the problem of the time of relaxation, t_R, of a stellar system has been reconsidered. It is found that

$$t_R = \frac{1}{6}\left(\frac{3}{2\pi}\right)^{1/2} \frac{(\overline{v^2})^{3/2}}{G^2M^2N\left[\log_e\left(\frac{D\overline{v^2}}{3GM}\right) - 0.355\right]},$$

where G is the constant of gravitation, N the number of stars per unit volume, $\overline{v^2}$ their mean square speed, M the mass of a star, and, finally, $D = (4\pi/3N)^{-1/3}$. This formula for the time of relaxation is shown to be in agreement with that derived by the alternative method in which the individual encounters are analyzed in terms of the two-body problem.

1. *Introduction.*—In estimating the influence of a fluctuating stellar distribution on the motions of stars it has invariably been supposed that such effects can be considered as the cumulative result of a large number of separate events, each of which can be idealized as distinct two-body encounters.[1] But a closer examination of the problem along these lines reveals an essential inconsistency in the assumptions made. For, on evaluating any of the desired quantities (e.g., $\Sigma|\Delta E|^2$ or $\Sigma \sin^2 2\Psi$ [cf. I and II]), it appears that the most important of the effects arise from encounters with impact parameters of the same order as the average distance between the stars. In other words, the physical situations most relevant to the problem are precisely those for which the two-body idealization of stellar encounters fails as a satisfactory mode of description. While this results in a divergence of the appropriate integrals as the impact parameter D tends to infinity and has in consequence to be cut off arbitrarily at some value of D, the really serious drawback of the method consists, however, in the essential inadequacy of the model to take account of the inherent *physical* aspects of the problem. A consideration of this and other related difficulties suggests that we abandon the two-body approximation of stellar encounters altogether and devise a more satisfactory *statistical method.* It is the object of this paper to outline the principles of such a statistical theory and to show its practical feasibility by reconsidering the problem of the time of relaxation of a stellar system along these new lines.

[1] For the most recent version of the theory based on these ideas see S. Chandrasekhar, *Ap. J.*, 93, 285, 1941; R. E. Williamson and S. Chandrasekhar, *Ap. J.*, 93, 305, 1941; and S. Chandrasekhar, *Ap. J.*, 93, 323, 1941. These papers will be referred to as I, II, and III, respectively. References to earlier literature will be found in these papers.

2. *The general principles of the statistical method.*—Quite generally the force $\boldsymbol{F}$ acting on a star, per unit mass, is given by

$$\boldsymbol{F} = -G\Sigma \frac{M_n}{|\boldsymbol{r}_n|^3} \boldsymbol{r}_n , \tag{1}$$

where M_n denotes the mass of a typical field star and $\boldsymbol{r}_n$ its position vector relative to the star under consideration. Further, the summation in equation (1) is taken over all the neighboring stars. The actual value of $\boldsymbol{F}$ at any particular instant will depend on the instantaneous positions of all the other stars and is in consequence subject to *fluctuations*. It would therefore be practically impossible to specify the exact dependence of $\boldsymbol{F}$ on the position and/or the time for individual cases. But, on the other hand, we can ask the probability of occurrence of any particular field strength. In evaluating this probability, we can (consistent with the physical situations we have in view) suppose that fluctuations subject only to the restriction of a constant average density occur.

Let

$$W(X, Y, Z)dXdYdZ \tag{2}$$

be the probability that $\boldsymbol{F}$ occurs in the range

$$(X, Y, Z) \leqslant \boldsymbol{F} \leqslant (X + dX, Y + dY, Z + dZ) . \tag{3}$$

From the symmetry of the problem we should expect that

$$W(F) = 4\pi F^2 W(\boldsymbol{F}) , \tag{4}$$

where F stands for $|\boldsymbol{F}|$. The meaning of $W(F)$ is simply that it gives the fraction of a long-enough interval of time during which a force of intensity F acts. A knowledge of $W(F)$ is therefore essential to our problem. It does not, however, provide all the necessary information. For, in order that we may be able to follow the motion of any particular star statistically, we need to know in addition the average length of time during which a given field strength acts *once it has become operative*. In other words we also require a knowledge of the *mean life* of the statistical state defined by F.

Now the notions of the mean life and the *spontaneous decay* of a given state of fluctuation has been introduced by Smoluchowski in his general investigations on Brownian motion and fluctuation phenomena.[2] According to these ideas of Smoluchowski, the probability $\phi(t)dt$ that a state represented by a well-defined statistical fluctuation continues to exist for a time t and makes a transition to a state of different fluctuation during t and $t + dt$ is expressible in the form

$$\phi(t)dt = e^{-t/T} \frac{dt}{T} , \tag{5}$$

where T is a constant characteristic of the state. Accordingly, we may define T as the mean life of the state under consideration. In our case we need to specify the mean life

$$T(F) \tag{6}$$

[2] Marian von Smoluchowski, *Wien. Ber.*, **123**, IIa, 2381, 1915; *ibid.*, **124**, IIa, 339, 1915; see also his papers in *Phys. Zs.*, **16**, 321, 1915, and **17**, 557, 585, 1916. For a general account of Smoluchowski's ideas see R. Furth, *Schwankungserscheinungen in der Physik* (*Sammlung Vieweg*), Heft 48, Braunschweig, 1920.

of a state in which a force of intensity F acts on a particular star (per unit mass). General considerations would suggest that for the average field strengths $\bar{F}$ we should expect

$$T(\bar{F}) \sim \frac{\bar{D}}{\overline{|v|}} , \tag{7}$$

where $\bar{D}$ stands for the average distance between the stars and $\overline{|v|}$ their average speeds.

Now, when a state defined by $\boldsymbol{F}$ becomes realized in consequence of fluctuations, the star will be accelerated at the rate F, and, since the mean life of this state is $T(F)$, the average acceleration to be expected during such a state is

$$\Delta v = T(F)\boldsymbol{F} . \tag{8}$$

When this state of fluctuation gives place to another, the star will begin to be accelerated at a different rate and in a direction uncorrelated with that in the earlier state. Hence the net acceleration suffered by the star is formally given by

$$\Sigma \Delta v = \Sigma T(F)\boldsymbol{F} , \tag{9}$$

where, as we have already indicated, the frequency of occurrence of the different values of $\boldsymbol{F}$ will be governed by $W(\boldsymbol{F})$.

On the basis of equation (9) we cannot, of course, predict the actual acceleration suffered by a star in any specified length of time. On the other hand, according to the principles of the theory of *random walk*[3] we should be able to predict the probability of a star's having been accelerated by a specified amount in a given length of time. This is the principle of our method.

After this general statement of the fundamental ideas we proceed to a more detailed consideration of the various factors which are involved.

3. *The probability of a given field strength. The Holtsmark distribution.*—According to our remarks in § 2, our first problem is to determine the probability of occurrence of a given field strength at some definite point due to a random distribution of centers of inverse square field of forces. This problem is clearly equivalent to finding the probability of a given *electric* field strength acting at a point in a gas composed of simple ions. This latter problem has been considered by J. Holtsmark;[4] and, re-writing his probability function to be appropriate for the gravitational case, we have

$$W(F) = \frac{2F}{\pi}\int_0^\infty e^{-\frac{4}{15}\sqrt{2\pi}\,\pi(GM)^{3/2}N\rho^{3/2}}\,\rho \sin F\rho d\rho , \tag{10}$$

where N stands for the number of stars per unit volume. We can re-write the foregoing formula for $W(F)$ more conveniently if we introduce a normal field strength, defined by

$$Q_H = \left(\tfrac{8}{15}\sqrt{2}\right)^{2/3}\pi GMN^{2/3} = 2.603GMN^{2/3} , \tag{11}$$

[3] Lord Rayleigh, *Collected Papers*, **1**, 491, Cambridge, England, 1899; **2**, 370, 1903; M. von Smoluchowski, *Bull. Acad. Cracovie*, p. 203, 1906; J. H. Jeans, *An Introduction to the Kinetic Theory of Gases*, p. 219, Cambridge, England, 1940; E. H. Kennard, *Kinetic Theory of Gases*, chap. vii, New York: McGraw-Hill, 1938.

[4] *Ann. d. Phys.*, **58**, 577, 1919. See also the papers by the same author in *Phys. Zs.*, **20**, 162, 1919, and **25**, 73, 1924.

and express F in terms of this unit. According to equations (10) and (11), we have

$$W(F) = \frac{2}{\pi F}\int_0^\infty e^{-(Q_H/F)^{3/2}x^{3/2}} x \sin x dx \, , \tag{12}$$

or, if

$$F = \beta Q_H \, , \tag{13}$$

$$W(\beta)d\beta = \frac{2d\beta}{\pi\beta}\int_0^\infty e^{-(x/\beta)^{3/2}} x \sin x dx \, . \tag{14}$$

The function $W(\beta)$ has been evaluated numerically by Holtsmark and more recently by Verweij.[5]

We may note that, according to equation (12),

$$W(F) \to \frac{3}{2}\frac{Q_H^{3/2}}{F^{5/2}} \, , \qquad F \to \infty \, . \tag{15}$$

This corresponds to a relatively slow decrease of the probability for high field strengths. Indeed, the probability distribution (12) gives an infinite value for the mean square field, $\overline{F^2}$. For certain physical problems this is unsatisfactory, and Gans[6] and Holtsmark[7] have modified the law (12), in the electrical case, to take into account the finite sizes of the ions. For the astronomical applications we have in view, the finite sizes of the stars cannot clearly be of any relevance. However, a modification of a different nature must be introduced before we can use the Holtsmark distribution (12). We shall return to this question in § 5.

4. *The probable field strengths produced by the nearest neighbor.*—In a general way it is clear that the main contribution to the field acting on a star must be due to its nearest neighbor. Indeed, as we shall presently see, the probable field strengths produced by the nearest neighbor provides a sufficiently good first approximation to the probability distribution according to equation (12).

To show this, consider first the probability $w(r)dr$ of finding the nearest neighbor to a given star between r and $r + dr$. It is readily seen that if the distribution of the stars is perfectly random (subject only to the restriction of a constant average density N) then $w(r)$ must satisfy the equation[8]

$$\left[1 - \int_0^r w(r)dr\right] 4\pi r^2 N = w(r) \, . \tag{16}$$

From equation (16) we derive

$$\frac{d}{dr}\left[\frac{w(r)}{4\pi r^2 N}\right] = -4\pi r^2 N \frac{w(r)}{4\pi r^2 N} \, . \tag{17}$$

[5] S. Verweij, *Pub. Ap. Inst. Amsterdam*, No. 5, Table 3, 1936.

[6] *Ann. d. Phys.*, **66**, 396, 1921.

[7] *Phys. Zs.*, **25**, 73, 1924.

[8] P. Hertz, *Math. Ann.*, **67**, 387, 1909; R. Gans, *Phys. Zs.*, **23**, 109, 1922; C. V. Raman, *Phil. Mag.*, **47**, 671, 1924.

Hence

$$w(r) = e^{-4\pi r^3 N/3}\, 4\pi r^2 N \, , \tag{18}$$

since, according to equation (16),

$$w(r) \rightarrow 4\pi r^2 N \, , \qquad r \rightarrow 0 \, . \tag{19}$$

If we now suppose that the field acting on a star is entirely due to the nearest neighbor, then

$$F = \frac{GM}{r^2} \, , \tag{20}$$

and the law of distribution of the nearest neighbors (eq. [18]) becomes equivalent to

$$W(F)dF = e^{-4\pi(GM)^{3/2}N/3F^{3/2}} 2\pi N (GM)^{3/2} \frac{dF}{F^{5/2}} \, . \tag{21}$$

If we now introduce the normal field

$$Q = (\tfrac{4}{3}\pi)^{2/3} GMN^{2/3} = 2.599 GMN^{2/3} \, , \tag{22}$$

equation (21) becomes

$$W(F)dF = \tfrac{3}{2} Q^{3/2} e^{-Q^{3/2}/F^{3/2}} \frac{dF}{F^{5/2}} \, . \tag{23}$$

According to equation (23),

$$W(F)dF \rightarrow \tfrac{3}{2} Q^{3/2} \frac{dF}{F^{5/2}} \, , \qquad F \rightarrow \infty \, . \tag{24}$$

Comparing equations (11) and (15) with equations (22) and (24), respectively, we conclude that for all practical purposes we may regard them as identical. Moreover, a more detailed comparison of the distributions (12) and (23) shows that even as regards the general dependence on F the two agree sufficiently well. There is an appreciable disagreement between the two distributions only for very small values of F/Q; but, as we shall see later, the weak fields have no significant consequences for the statistical theory. Finally, we may remark that the agreement in the asymptotic behaviors of the two distributions for large values of F implies that the highest field strengths are produced by the nearest neighbor.

5. *The modification of the distribution function for high field strengths.*—As we have already remarked in § 3, the Holtsmark distribution (12) predicts too high probabilities for high field strengths. The same remark applies also to the distribution (23). In our present case the high probabilities result from the assumption of the randomness of the stellar distribution for all elements of volume. But it is clear that this assumption cannot be valid for the regions in the immediate neighborhoods of the individual stars. For a star with a linear velocity[9] v cannot come closer to another star than a certain critical distance $r(v)$ such that

$$\tfrac{1}{2} M v^2 = \frac{GM^2}{r(v)} \tag{25}$$

[9] At an average distance from the other stars.

or

$$r(v) = \frac{2GM}{v^2} . \tag{26}$$

For otherwise the star should strictly be regarded as the component of a binary system, and this is inconsistent with our original premises. This restriction naturally implies a departure from true randomness for these stars as $r \to r(v)$. However, it appears that under the conditions of our problem these departures become significant only as $r \to 0$. In any case it is apparent that the relatively high probabilities predicted by equation (12) or equation (23) for high field strengths will be reduced if proper account is taken of the increasing lack of randomness in stellar distribution as we approach the centers of attraction. A rigorous treatment of this effect will require a reconsideration of the whole problem in *phase space*[10] and is beyond the scope of the present investigation. However, an elementary treatment of the effect can be given, and this appears to be adequate for our purposes.

We shall first consider the problem along the lines of § 4. If $w(r)$ represents, as before, the probability of finding the nearest neighbor to a given star between r and $r + dr$, then the circumstance that stars with linear velocities v cannot come closer to the center than the limit given by (26) will modify equation (16) to

$$\left[1 - \int_0^r w(r)dr\right] 4\pi r^2 \chi(r) N = w(r) , \tag{27}$$

where the function $\chi(r)$ has been introduced to take account of the lack of randomness at close distances. Quite generally we should expect that

$$\chi(r) \to 0 , \quad r \to 0 ; \qquad \chi(r) \to 1 , \quad r \to \infty . \tag{28}$$

The formal solution of equation (27) can be readily written down. We have

$$w(r) = e^{-4\pi N \int_0^r r^2 \chi(r) dr} 4\pi r^2 \chi(r) N , \tag{29}$$

or, differently, as

$$w(r) = e^{-4\pi N r^3 \bar{\chi}(r)/3} 4\pi r^2 \chi(r) N , \tag{30}$$

where we have written

$$\bar{\chi}(r) = \frac{3}{r^3} \int_0^r r^2 \chi(r) dr . \tag{31}$$

According to equation (28),

$$\bar{\chi}(r) \to 0 , \qquad r \to 0 ; \qquad \bar{\chi}(r) \to 1 , \quad r \to \infty . \tag{32}$$

To make the law of distribution of the nearest neighbors according to equation (30) more definite, we need an explicit expression for $\chi(r)$. As we have already indicated, the exact specification of $\chi(r)$ will require a detailed consideration of the problem in phase

[10] In contrast to Holtsmark's treatment, in which the probability distribution of the centers of attraction in *configuration space* is assumed to be independent of the velocities of the particles.

space. But it appears that in a first approximation we may suppose that *the distribution of stars of any prescribed velocity* v *about a given star is perfectly random for all distances greater than the critical distance* $r(v) = 2GM/v^2$. Similarly, we may suppose that *no stars with velocity* v *occur within the sphere of radius* r(v).[11] On these assumptions we can readily write down an explicit expression for $\chi(r)$. We have

$$\chi(r) = \iiint_{|v|=\sqrt{2GM/r}}^{|v|=\infty} f(v)dv_x dv_y dv_z , \tag{33}$$

where $f(v)$ denotes the frequency function of the velocities among the stars. If, for the sake of definiteness, we suppose that $f(v)$ is Maxwellian,

$$f(v) = \frac{j^3}{\pi^{3/2}} e^{-j^2|v|^2} , \tag{34}$$

then

$$\chi(r) = \frac{4j^3}{\pi^{1/2}} \int_{\sqrt{2GM/r}}^{\infty} e^{-j^2v^2} v^2 dv . \tag{35}$$

The foregoing formula for $\chi(r)$ can be expressed more conveniently in the form

$$\chi(r) = \frac{4}{\pi^{1/2}} \int_{a/\sqrt{r}}^{\infty} e^{-y^2} y^2 dy , \tag{36}$$

where we have written

$$y = jv ; \qquad a = \sqrt{2GM}\, j . \tag{37}$$

An alternative form for $\chi(r)$ may be noted:

$$\chi(r) = 1 - \frac{2}{\pi^{1/2}} \left[\int_0^{a/\sqrt{r}} e^{-y^2} dy - \frac{a}{\sqrt{r}} e^{-a^2/r} \right] . \tag{38}$$

Again, according to equation (31), we have

$$\bar{\chi}(r) = \frac{12}{\pi^{1/2} r^3} \int_0^r r^2 \left(\int_{a/r^{1/2}}^{\infty} e^{-y^2} y^2 dy \right) dr \tag{39}$$

or, after an integration by parts,

$$\bar{\chi}(r) = \chi(r) - \frac{2a^3}{\pi^{1/2} r^3} \int_0^r r^{1/2} e^{-a^2/r} dr . \tag{40}$$

After some further reductions we find that

$$\bar{\chi}(r) = \chi(r) - \frac{4a^6}{3\pi^{1/2} r^3} \left[e^{-a^2/r} \left(\frac{r^{3/2}}{a^3} - 2\frac{r^{1/2}}{a} \right) + 4 \int_{a/r^{1/2}}^{\infty} e^{-y^2} dy \right] . \tag{41}$$

[11] This latter assumption is, however, *necessary* (see the remark immediately following equation [26]).

The functions χ and $\bar{\chi}$ are tabulated in Table 1 for different values of the argument $a/r^{1/2}$. An examination of this table shows that to a first approximation we may write

$$\chi(r) = 1 , \quad a \leqslant r^{1/2} ; \qquad \chi(r) = 0 , \quad a > r^{1/2} . \tag{42}$$

In this approximation equation (30) becomes

$$\left.\begin{aligned} w(r) &= e^{-4\pi N(r^3 - r_0^3)/3} 4\pi r^2 N \qquad && (r \geqslant r_0) , \\ &= 0 && (r < r_0) , \end{aligned}\right\} \tag{43}$$

where

$$r_0 = 2GMj^2 . \tag{44}$$

Returning to equation (30), we see that this law of distribution of the nearest neighbors implies a probability of occurrence of a field strength F (assuming that the field arises principally from the first neighbor) given by

$$W(F) = \tfrac{3}{2} Q^{3/2} e^{-Q^{3/2}\bar{\chi}(\sqrt{GM/F})/F^{3/2}} \frac{\chi(\sqrt{GM/F})}{F^{5/2}} , \tag{45}$$

where Q, χ, and $\bar{\chi}$ are defined as in equations (22), (38), and (41). The modification which we have thus effected in the distribution function (23) removes the principal objection to

TABLE 1

$\chi(r)$ AND $\bar{\chi}(r)$

$a/r^{1/2}$	χ	$\bar{\chi}$	$a/r^{1/2}$	χ	$\bar{\chi}$
0	1.000	1.000	1.2	0.410	0.275
0.2	0.994	0.989	1.4	.270	.161
0.4	0.956	0.922	1.6	.163	.086
0.6	0.868	0.787	1.8	.090	.043
0.8	0.734	0.610	2.0	.046	.019
1.0	0.572	0.430	3.0	0.000	0.000

it, namely, the prediction of a nonconvergent value for $\overline{F^2}$, for our present distribution function (45) yields a finite value for the mean square field.

In the approximation (42), equation (45) simplifies to

$$\left.\begin{aligned} W(F) &= \tfrac{3}{2} Q^{3/2} e^{-Q^{3/2}(F^{-3/2} - F_{\max}^{-3/2})} \frac{1}{F^{5/2}} \qquad && (F \leqslant F_{\max}) , \\ &= 0 && (F > F_{\max}) , \end{aligned}\right\} \tag{46}$$

where, according to equation (44),

$$F_{\max} = \frac{1}{4GMj^4} . \tag{47}$$

We shall now consider very briefly how the lack of randomness in the immediate neighborhoods of the centers of attraction can be incorporated into the Holtsmark distribution (12). It appears that it is not an altogether simple matter to modify the Holtsmark distribution rigorously even on the basis of the very simplified assumptions which led to the explicit expressions (38) and (41) for $\chi(r)$ and $\overline{\chi}(r)$. But, remembering that the highest fields are produced by the nearest neighbor and, further, that the lack of randomness becomes significant only as $r \to 0$, it appears that we may incorporate the main features by considering an approximation corresponding to equation (42), i.e., by supposing that no star has a first neighbor closer than $r_0 = 2GMj^2$ and that the distribution is random but for this restriction. In this last approximation the problem becomes formally the same as when the ions, in the electrical case, have finite dimensions. With suitable changes we can therefore use the results of Gans and Holtsmark,[12] who have modified the distribution function (12), in the electrical case, for the finite sizes of the ions. We have[13]

$$W(F) = \frac{2F}{\pi} e^{4\pi r_0^3 N/3} \int_0^\infty e^{-(Q_H \rho)^{3/2}} K(\rho) \rho \sin F\rho d\rho \,, \tag{48}$$

where $K(\rho)$ is a certain correction factor which is defined in Holtsmark's paper.[14]

6. *The mean life of the state* F.—Our next problem is to determine the mean life of a statistical state defined by F. The totality of statistical complexions which go to make up the state in question are not explicitly defined, and Smoluchowski's ideas cannot be applied without further deep generalizations of them. However, in the approximation in which the fluctuating fields are assumed to arise from the nearest neighbor, the statistical complexion is specified explicitly, and a formula due to Smoluchowski can be directly used.

Now, according to Smoluchowski, the mean life of a state in which n particles are found in an element of volume σ is given by[15]

$$T = \frac{\sqrt{6\pi}}{\sqrt{\overline{v^2}}\,(n+\nu)} \frac{\sigma}{S_\sigma} \,, \tag{49}$$

where $\overline{v^2}$ denotes the mean square speed of the particles, S_σ the surface area of the element σ, and ν the number of particles which the element σ would contain at the constant average density:

$$\nu = N\sigma \,. \tag{50}$$

For the particular case we have in view

$$\sigma = \tfrac{4}{3}\pi r^3 \,; \qquad S_\sigma = 4\pi r^2 \,; \qquad n = 1 \,; \qquad \nu = \tfrac{4}{3}\pi r^3 N \,, \tag{51}$$

[12] See the references given in nn. 6 and 7.

[13] J. Holtsmark, *Phys. Zs.*, 25, 73, 1924; see particularly eqs. (104) and (145).

[14] See eq. (124) in the paper referred to in n. 13. Holtsmark has not evaluated this correction factor explicitly for the case of an inverse square field. But an evaluation of this factor along the lines of Holtsmark's analysis for the dipole field is possible.

[15] Smoluchowski, *Phys. Zs.*, 17, 557, 1916; and see §§ 5, 6, and 7 in this paper and particularly eq. (30). See also Furth, *op. cit.*, pp. 34, 35, and 43.

since we need the mean life of a state in which a particular star continues to exist as the sole occupant of a sphere of radius r. Accordingly,

$$T(r) = \sqrt{\frac{2\pi}{3\overline{v^2}}}\,\frac{r}{\frac{4}{3}\pi r^3 N + 1}\,. \tag{52}$$

In the approximation of § 4

$$r = \sqrt{\frac{GM}{F}}\,, \tag{53}$$

and equation (52) implies for the state F the mean life

$$T(F) = \sqrt{\frac{2\pi GM}{3\overline{v^2}}}\,\frac{F}{Q^{3/2} + F^{3/2}}\,, \tag{54}$$

where Q is defined as in equation (22).

Formula (54) for $T(F)$ is clearly only an approximate one. But since, according to the Holtsmark distribution, the highest fields are produced by the nearest neighbor, the true expression for $T(F)$ must tend to equation (54) for high field strengths. Consequently, we may expect equation (54) to give as good an approximation to the true values of $T(F)$ as the $W(F)$ according to equation (23) or equation (30) provides an approximation to the Holtsmark distribution. This is probably quite sufficient for most purposes.

7. *The acceleration of a star in the fluctuating gravitational field.*—We shall begin our discussion of this problem by considering the following simplified case: Imagine a star's undergoing a series of accelerations during a large number of intervals of constant duration T, in such a way that during each interval it is accelerated at the same rate F but in directions which are uncorrelated from interval to interval. Under these circumstances the star experiences an increase of velocity of amount TF in each of the intervals; but these increments take place along uncorrelated directions in a random manner. We now ask the probability that at the end of s such intervals the star has undergone a net increase of velocity of mFT in some specified direction. According to the principles of the theory of random walk we have[16]

$$P_m = \sqrt{\frac{3}{2\pi s}}\,e^{-3m^2/2s}\,, \tag{55}$$

when s is sufficiently large. Since the net increase in velocity Δv and the time t during which this increase has taken place are related to m and s by

$$\Delta v = mFT\,; \qquad t = sT\,, \tag{56}$$

we have

$$P(\Delta v) = \sqrt{\frac{3T}{2\pi t}}\,e^{-3|\Delta v|^2/(2F^2Tt)}\,. \tag{57}$$

[16] See the references given in n. 3, particularly Kennard, *op. cit.*, pp. 269–72.

Hence, the probability that there occurs an increase in velocity in the range $[\Delta v, \Delta v + d(\Delta v)]$ during a time t in some prescribed direction is given by

$$P(\Delta v)d(\Delta v) = \sqrt{\frac{3}{2\pi F^2 T t}}\, e^{-3|\Delta v|^2/(2F^2Tt)} d(\Delta v) \,. \tag{58}$$

Accordingly,

$$\overline{\Delta v^2} = F^2 T t \,. \tag{59}$$

We shall now generalize the foregoing problem to the case when F does not have a unique value but occurs according to a definite frequency function $W(F)$ and when the average duration of an acceleration at the rate F is given by a function $T(F)$. In view of the addition theorem for the Gaussian error functions, equation (58) becomes modified under these more general circumstances to

$$P(\Delta v) = \sqrt{\frac{3}{2\pi \overline{F^2 T} t}}\, e^{-3|\Delta v|^2/(2\overline{F^2T}t)} \,, \tag{60}$$

where

$$\overline{F^2 T} = \int_0^\infty W(F) F^2 T(F) dF \,. \tag{61}$$

Hence, instead of equation (59) we now have

$$\overline{\Delta v^2} = \overline{F^2 T} t \,. \tag{62}$$

8. *The evaluation of* $\overline{\Delta v^2}$.—According to equations (45), (54), and (61) we have

$$\overline{F^2 T} = \frac{3}{2}\sqrt{\frac{2\pi GM}{3\overline{v^2}}}\, Q^{3/2} \int_0^\infty \frac{F^{1/2}}{Q^{3/2} + F^{3/2}}\, e^{-Q^{3/2}\bar{\chi}(\sqrt{GM/F})/F^{3/2}} \chi(\sqrt{GM/F}) dF \,. \tag{63}$$

When we introduce a new variable x defined by

$$\frac{Q^{3/2}}{F^{3/2}} = x \,, \tag{64}$$

equation (63) becomes

$$\overline{F^2 T} = 2\left(\frac{2\pi}{3}\right)^{3/2} \frac{G^2 M^2 N}{\sqrt{\overline{v^2}}} \int_0^\infty e^{-x\bar{\chi}(\sqrt{GM/Q}\, x^{1/3})} \left(\frac{1}{x} - \frac{1}{x+1}\right) \chi(\sqrt{GM/Q}\, x^{1/3}) dx \,. \tag{65}$$

Substituting for Q from equation (22) in the argument for the functions χ and $\bar{\chi}$ in the foregoing equation, we obtain

$$\sqrt{\frac{GM}{Q}}\, x^{1/3} = \left(\frac{x}{\frac{4}{3}\pi N}\right)^{1/3} = D x^{1/3} \,, \tag{66}$$

where we have written

$$D = \frac{1}{(\frac{4}{3}\pi N)^{1/3}}. \tag{67}$$

Now, according to equations (38) and (41), the functions χ and $\bar{\chi}$ depend on r only through the combination $a/r^{1/2}$. From equations (37) and (66) we now find that

$$\frac{a}{r^{1/2}} = \sqrt{\frac{2GMj^2}{D}}\, x^{-1/6}. \tag{68}$$

Since

$$\frac{D}{2GM} = 2.33 \times 10^4 \frac{(D/\text{parsec})}{(M/\odot)(10\ \text{km/sec})^2}, \tag{69}$$

it follows that under most stellar conditions $2\,GMj^2/D \sim 10^{-4}$. Hence, only for values of $x < 10^{-11}$ do the functions χ and $\bar{\chi}$ deviate appreciably from unity (see Table 1). We can therefore replace χ by unity whenever it does not occur multiplied by a factor which diverges at $x = 0$. Similarly, we can replace $\bar{\chi}$ also by unity; but this we can always do since $\bar{\chi}$ occurs in the exponent multiplied with x. Thus, to a high degree of accuracy, the integral on the right-hand side of equation (63) is the same as

$$\int_0^\infty \frac{e^{-x}}{x}\chi(Dx^{1/3})dx - \int_0^\infty \frac{e^{-x}}{x+1}\,dx = J \quad (\text{say}). \tag{70}$$

Substituting for χ according to (36) in the foregoing equation we obtain

$$J = \frac{4}{\pi^{1/2}}\int_0^\infty \frac{e^{-x}}{x}\left[\int_{\sqrt{2GMj^2/D}\,x^{-1/6}}^\infty e^{-y^2}y^2dy\right]dx - \int_1^\infty \frac{e^{-(x-1)}}{x}\,dx. \tag{71}$$

Writing

$$-E(-x) = \int_x^\infty \frac{e^{-x}}{x}\,dx, \tag{72}$$

we have

$$J = \frac{4}{\pi^{1/2}}\int_0^\infty \frac{d}{dx}(E(-x))\int_{\sqrt{2GMj^2/D}\,x^{-1/6}}^\infty e^{-y^2}y^2dy\,dx + eE(-1). \tag{73}$$

Integrating by parts the integral on the right-hand side of equation (73), we find

$$J = -\frac{4}{\pi^{1/2}}\int_0^\infty E\left(-\left[\frac{2GMj^2}{D}\right]^3 z^{-6}\right)e^{-z^2}z^2dz - 0.5963. \tag{74}$$

The argument for the exponential integral occurring under the integral sign in equation (74) is seen to be extremely small for the values of x which are at all relevant to the value of the integral. Hence we can use the asymptotic expansion for $E(-x)$ valid for $x \to 0$. We have

$$E(-x) = \log x + 0.5772 + O(x), \tag{75}$$

where the constant on the right-hand side is the Euler-Mascheroni constant. Using the foregoing expansion for $E(-x)$ in equation (74), we readily find that

$$J = 3 \log \left(\frac{D}{2GMj^2}\right) - 0.5772 + \frac{6}{\pi^{1/2}} \int_0^\infty e^{-x} x^{1/2} \log x dx - 0.5963 . \tag{76}$$

On the other hand, we have

$$\left.\begin{aligned} \int_0^\infty e^{-x} x^{1/2} \log x dx &= \Gamma(\tfrac{3}{2}) \left[\frac{d \log \Gamma(x)}{dx}\right]_{x=3/2} , \\ &= \frac{\pi^{1/2}}{2} \times 0.03649 . \end{aligned}\right\} \tag{77}$$

Hence,

$$J = 3 \log \left(\frac{D}{2GMj^2}\right) - 1.0640 . \tag{78}$$

Finally, substituting for $\overline{F^2T}$ according to equations (65) and (78) in equation (62) we obtain

$$\overline{\Delta v^2} = 6\left(\frac{2\pi}{3}\right)^{3/2} \frac{GM^2N}{\sqrt{\overline{v^2}}} \left[\log \left(\frac{D\overline{v^2}}{3GM}\right) - 0.355\right] t . \tag{79}$$

We may note that if we had used the approximation (46) for $W(F)$ (instead of the more accurate formula [45]) we should have obtained

$$\overline{F^2T} = 2\left(\frac{2\pi}{3}\right)^{3/2} \frac{G^2M^2N}{\sqrt{\overline{v^2}}} \int_{(2GMj^2/D)^3}^\infty \frac{e^{-x}}{x(x+1)} dx \tag{80}$$

instead of equation (65). On evaluating the integral on the right-hand side of (80), we find

$$\int_{(2GMj^2/D)^3}^\infty \frac{e^{-x}}{x(x+1)} dx = 3 \log \left(\frac{D}{2GMj^2}\right) - 1.1735 , \tag{81}$$

which should be compared with equation (78). We thus see that approximations based on the assumption (43) (or their equivalents) are likely to provide sufficient accuracy for most purposes. In particular, the modification of the Holtsmark distribution suggested on page 519 to take account of the lack of randomness in stellar distribution in the immediate neighborhoods of stars can be justified on these grounds.

9. *The time of relaxation of a stellar system.*—An immediate application of the fundamental formula (79) is to the problem of the *time of relaxation* of a stellar system. According to the general ideas outlined in I, § 1, we can definite this as the time required for $\overline{\Delta v^2}$ to become of the same order as $\overline{v^2}$. Thus, if t_R denotes this time, we have

$$t_R = \frac{1}{6}\left(\frac{3}{2\pi}\right)^{3/2} \frac{(\overline{v^2})^{3/2}}{G^2M^2N\left[\log \left(\frac{D\overline{v^2}}{3GM}\right) - 0.355\right]} . \tag{82}$$

We can now compare this formula with that obtained on the basis of the two-body idealization of stellar encounters. We have[17]

$$t_E = \frac{1}{16}\left(\frac{3}{\pi}\right)^{1/2} \frac{(\overline{v^2})^{3/2}}{G^2M^2N \log\left(\frac{\overline{D}\overline{v^2}}{3GM}\right)}, \tag{83}$$

where $\overline{D}$ is the average distance between the stars.[18] We notice that the two equations (82) and (83) are of identical forms; further it is found that the numerical factors in the two formulae differ only by a factor 1.11. This agreement, while confirming the general correctness of our statistical method, exhibits also its immense superiority over the earlier treatments of the same problem both in the appropriateness of the physical ideas and in the simplicity of the mathematical treatment.

10. *Concluding remarks.*—The perfectly natural way in which the solution to the problem of the time of relaxation appears on the present theory suggests the extension of these methods to solve other problems of stellar dynamics. Thus the evolution of wide binaries in a fluctuating gravitational field is a problem to which the principles of the statistical theory are particularly well adapted. For, while on the classical methods the treatment of this problem would require the analysis of individual encounters considered strictly as three-body problems, on the statistical theory all such detailed considerations would be eliminated. Again, the application of the fundamental theorem of statistical dynamics due to Planck and Fokker[19] to problems of stellar dynamics is another field to which the method of the present paper can be used. We shall consider these problems in later papers.

In conclusion I wish to record my indebtedness to Messrs. G. Randers and R. E. Williamson for valuable discussions.

YERKES OBSERVATORY
July 24, 1941

NOTE ADDED IN PROOF.—Since the foregoing paper was written it has been found possible to solve rigorously the question of the half-life treated approximately in section 6. While this exact treatment leads to substantially the same results, it enables a more complete visualization of the phenomenon in question. It is hoped to publish these newer results in the near future.

[17] See. S. Chandrasekhar, *The Principles of Stellar Dynamics*, chap. ii, University of Chicago Press. (In Press.)

[18] According to eq. (18), the average distance $\overline{D}$ between the stars is given by

$$\overline{D} = \int_0^\infty e^{-4\pi r^3 N/3} 4\pi r^3 N dr \tag{84}$$

or, after some elementary reductions,

$$\overline{D} = \frac{1}{(\frac{4}{3}\pi N)^{1/3}} \int_0^\infty e^{-x} x^{1/3} dx . \tag{85}$$

Hence, comparing eqs. (67) and (83), we have

$$\overline{D} = \Gamma(\tfrac{4}{3}) D = 0.8930 D , \tag{86}$$

a result due to Hertz (see the reference in n. 8).

[19] M. Planck, *Sitzungsber. der preuss. Akad.*, p. 324, Berlin, 1917; A. Fokker, *Ann. d. Phys.*, 45, 812, 1914.

NEW METHODS IN STELLAR DYNAMICS*

By

S. Chandrasekhar †

CONTENTS

* Awarded an A. Cressy Morrison Prize in Natural Science in 1942 by The New York Academy of Sciences. Publication made possible through a grant from the income of the Ralph Winifred Tower Memorial Fund.

† Yerkes Observatory, Williams Bay, Wisconsin.

PREFACE

The present paper gives a condensed version of certain new methods which the author has recently been developing for investigating the dynamics of stellar systems. In presenting the subject, it was thought desirable that the emphasis be placed throughout on the physical aspects of the problems and whenever this has required the suppression of the mathematical details I have not avoided doing so. This is particularly true in the more technical parts of the subject.

Since the original version of this paper was submitted to the New York Academy of Sciences in September, 1942, the subject has advanced along several directions. The author is therefore greatly indebted to the Council of the Academy for permission to drastically revise and recast the entire article.

S. C.

July, 1943.

I. THE STATISTICS OF THE GRAVITATIONAL FIELD ARISING FROM A RANDOM DISTRIBUTION OF STARS

The Outline of the Statistical Method

One of the principal problems of stellar dynamics is concerned with the analysis of the nature of the force acting on a star which is a member of a stellar system.[1] In a general way, it is clear that we may broadly distinguish between the influence of the system as a whole and the influence of the immediate neighborhood. The former will be a smoothly varying function of position and time while the latter will be subject to relatively rapid fluctuations (see below).

Considering first the influence of the system as a whole, it appears that we can express it in terms of the gravitational potential $\mathbf{V}(\boldsymbol{r};t)$ derived from the function $n(\boldsymbol{r}, M; t)$ which governs the average spatial distribution of the stars of different masses at time t. Thus

$$\mathbf{V}(\boldsymbol{r};t) = -G\int_{-\infty}^{+\infty}\int_{0}^{\infty}\frac{Mn(\boldsymbol{r}_1,M,t)dMd\boldsymbol{r}_1}{|\boldsymbol{r}_1 - \boldsymbol{r}|}, \qquad (1)$$

where G denotes the constant of gravitation. The potential $\mathbf{V}(\boldsymbol{r};t)$ derived in this manner may be said to represent the "smoothed out"

[1] See for example, Chandrasekhar, S., "Principles of Stellar Dynamics." Chapter II. University of Chicago Press. 1942.

distribution of matter in the stellar system. The force per unit mass acting on a star due to the "system as a whole" is therefore given by

$$K = -\operatorname{grad}\mathfrak{V}(r;t). \tag{2}$$

However, the fluctuations in the *complexion* of the local stellar distribution will make the instantaneous force acting on a star to deviate from the value given by (2). To elucidate the nature and origin of these fluctuations we surround the star under consideration by an element of volume σ which we shall suppose is small enough to contain, on the average, only a relatively few stars. The actual number of stars, which will be found in σ at any given instant of time, will not in general be the average number that will be expected to be in it, namely, $n\sigma$; it will be subject to fluctuations. These fluctuations will naturally be governed by a Poisson distribution with variance $n\sigma$. It is in direct consequence of this changing complexion of the local stellar distribution that the influence of the near neighbors on a star is variable. The average period of such a fluctuation is readily estimated. The order of magnitude of the time involved is evidently that required for two stars to separate by a distance D equal to the average distance between the stars. We may therefore expect that the influence of the immediate neighborhood will fluctuate with an average period of the order of

$$T \backsimeq \frac{D}{\sqrt{\overline{|\mathbf{V}|^2}}}, \tag{3}$$

where $[\overline{|V|^2}]^{1/2}$ denotes the root mean square relative velocity between two stars.

In the neighborhood of the sun, $D \sim 3$ parsecs, $[\overline{|V|^2}]^{1/2} \sim 50$ km/sec. Hence,

$$T \text{ (near the sun)} \sim 6 \times 10^4 \text{ years}. \tag{4}$$

When we compare this time with the period of galactic rotation (which is about 2×10^8 years), we observe that, in conformity with our earlier remarks, the fluctuations in the force acting on a star due to the changing local stellar distribution does in fact occur with extreme rapidity compared to the rate at which any of the other physical parameters change. Accordingly, we may write for the force per unit mass acting on a star, the expression

$$\mathbf{F} = K(r;t) + F(t), \tag{5}$$

where K is derived from the smoothed out distribution, as in equations (1) and (2), and F denotes the fluctuating force due to the near neighbors. Moreover, if Δt denotes an interval of time, long compared to (3), we may write

$$\boldsymbol{F}\Delta t = \boldsymbol{K}\Delta t + \delta\boldsymbol{u}(t;t+\Delta t), \tag{6}$$

where

$$\delta\boldsymbol{u}(t;t+\Delta t) = \int_t^{t+\Delta t} \boldsymbol{F}(\xi)d\xi \qquad (\Delta t \gg T). \tag{7}$$

Under the circumstances stated ($\Delta t \gg T$), the accelerations $\delta\boldsymbol{u}(t;t+\Delta t)$ and $\delta\boldsymbol{u}(t + \Delta t; t + 2\Delta t)$ suffered during two successive intervals $(t,t+\Delta t)$ and $(t + \Delta t,t + 2\Delta t)$ will not be expected to show any correlation. We may therefore anticipate the existence of a definite law of distribution which will govern the probability of occurrence of the different values of $\delta\boldsymbol{u}(t;t+\Delta t)$. We thus see that the acceleration which a star suffers during an interval $\Delta t \gg T$ can be formally expressed as the sum of two terms: a *systematic* term, $\boldsymbol{K}\Delta t$, due to the action of the gravitational field of the smoothed out distribution and a *stochastic* term, $\delta\boldsymbol{u}(t; t + \Delta t)$, representing the influence of the near neighbors. Stated in this fashion, we recognize the similarity between our present problems in stellar dynamics with those which occur in the modern theories of Brownian motion.[2]

We proceed now to the outline of a general method which appears suitable for analyzing the statistical properties of $\boldsymbol{F}(t)$. The force $\boldsymbol{F}$ acting on a star, per unit mass, is given by

$$\boldsymbol{F} = G\sum_i \frac{M_i}{|\boldsymbol{r}_i|^3}\boldsymbol{r}_i, \tag{8}$$

where M_i denotes the mass of a typical "field" star and $\boldsymbol{r}_i$ its position vector relative to the star under consideration; further, in equation (8) the summation is to be extended over all the neighboring stars. The actual value of $\boldsymbol{F}$ given by equation (8) at any particular instant of time will depend on the instantaneous complexion of the local stellar distribution. It is in consequence subject to fluctuations. We can therefore ask only for the probability of occurrence

$$W(\boldsymbol{F})dF_x dF_y dF_z = W(\boldsymbol{F})d\boldsymbol{F} \tag{9}$$

of $\boldsymbol{F}$ in the range $\boldsymbol{F}$ and $\boldsymbol{F} + d\boldsymbol{F}$. In evaluating this probability distribution we shall suppose, consistent with the physical situations we have in view, that fluctuations subject only to the restriction of a constant average density occur. However, the specification of $W(\boldsymbol{F})$ does *not* provide us with all the necessary information concerning the fluctuating force $\boldsymbol{F}$. An equally important aspect of the problem concerns the *speed of fluctuations*.

[2] See a forthcoming article by the author in the "Reviews of Modern Physics."

According to equation (8) the rate of change of $\boldsymbol{F}$ with time is given by

$$\boldsymbol{f} = \frac{d\boldsymbol{F}}{dt} = G \sum_i M_i \left\{ \frac{\boldsymbol{V}_i}{|\boldsymbol{r}_i|^3} - 3\boldsymbol{r}_i \frac{(\boldsymbol{r}_i \cdot \boldsymbol{V}_i)}{|\boldsymbol{r}_i|^5} \right\}, \tag{10}$$

where $\boldsymbol{V}_i$ denotes the velocity of a typical field star *relative* [3] to the star under consideration. It is now apparent that the speed of fluctuations in $\boldsymbol{F}$ can be specified in terms of the bivariate distribution

$$W(\boldsymbol{F}, \boldsymbol{f}), \tag{11}$$

which governs the probability of the simultaneous occurrence of prescribed values for both $\boldsymbol{F}$ and $\boldsymbol{f}$. It is seen that this distribution function $W(\boldsymbol{F}, \boldsymbol{f})$ will depend on the assignment of *a priori* probability in the *phase space* in contrast to the distribution $W(\boldsymbol{F})$ of $\boldsymbol{F}$, which depends only on a similar assignment in the *configuration space.* While it is possible by an application of a general method, due to Markoff, to write down a general formula for $W(\boldsymbol{F}, \boldsymbol{f})$, it does not appear feasible to obtain the required distribution function in an explicit form. However, it is possible to obtain explicit formulae for all the first and second moments of $\boldsymbol{f}$ for a given $\boldsymbol{F}$; and it appears possible to make some progress in the specification of the statistical properties of $\boldsymbol{F}$ in terms of these moments.

The Statistical Properties of F

We require the stationary distribution of $\boldsymbol{F}$ and its simultaneous rate of change $\boldsymbol{f}$ acting on a given star. Without loss of generality we can suppose that the point under consideration is at the origin, O, of our system of coordinates. About O describe a sphere of radius R and containing N stars. In the first instance we shall suppose that

$$\boldsymbol{F} = G \sum_{i=1}^{N} \frac{M_i}{|\boldsymbol{r}_i|^3} \boldsymbol{r}_i \tag{12}$$

and

$$\boldsymbol{f} = G \sum_{i=1}^{N} M_i \left\{ \frac{\boldsymbol{V}_i}{|\boldsymbol{r}_i|^3} - 3\boldsymbol{r}_i \frac{(\boldsymbol{r}_i \cdot \boldsymbol{V}_i)}{|\boldsymbol{r}_i|^5} \right\}; \tag{13}$$

but we shall later let R and N tend to infinity simultaneously in such a way that

$$\frac{4}{3}\pi R^3 n = N; \qquad (R \to \infty; \qquad N \to \infty; \qquad n = \text{constant}). \tag{14}$$

This limiting process is permissible, in view of the fact that the dominant contribution to $\boldsymbol{F}$ is made by the nearest neighbor [4]; consequently, the

[3] It is in this respect that the analysis which follows differs from that contained in **Chandrasekhar, S., & von Neumann, J.** Astrophysical Jour. **95**: 489, 1942, where the speed of fluctuations in $\boldsymbol{F}$ acting at some *fixed* point in space is considered.

[4] Cf. **Chandrasekhar, S.,** Astrophysical Jour. **94**: 511, 1941 (see particularly § 4).

formal extrapolation to infinity of the density of stars obtained only in a given region of stellar system can hardly affect the results to any appreciable extent.

Using a general method due to Markoff, we can readily write down a general formula for the distribution function $W(\boldsymbol{F},\boldsymbol{f})$. We have [5]

$$W(\boldsymbol{F},\boldsymbol{f}) = \frac{1}{64\pi^6}\int_{|\boldsymbol{\varrho}|=0}^{\infty}\int_{|\boldsymbol{\sigma}|=0}^{\infty} e^{-i(\boldsymbol{\varrho}\cdot\boldsymbol{F}+\boldsymbol{\sigma}\cdot\boldsymbol{f})}\, A(\boldsymbol{\varrho},\boldsymbol{\sigma})d\boldsymbol{\varrho}d\boldsymbol{\sigma}, \qquad (15)$$

where

$$A(\boldsymbol{\varrho},\boldsymbol{\sigma}) = \underset{R\to\infty}{\text{Limit}}\left[\frac{3}{4\pi R^3}\int_{M=0}^{\infty}\int_{|\boldsymbol{V}|=0}^{\infty}\int_{|\boldsymbol{r}|=0}^{R} e^{i(\boldsymbol{\varrho}\cdot\boldsymbol{\phi}+\boldsymbol{\sigma}\cdot\boldsymbol{\psi})}\, \tau d\boldsymbol{r}d\boldsymbol{V}dM\right]^{4\pi R^3 n/3}. \qquad (16)$$

In equations (15) and (16) $\boldsymbol{\varrho}$ and $\boldsymbol{\sigma}$ are two auxiliary vectors; n denotes the number of stars per unit volume;

$$\boldsymbol{\phi} = GM\frac{\boldsymbol{r}}{|\boldsymbol{r}|^3};\quad \boldsymbol{\psi} = GM\left(\frac{\boldsymbol{V}}{|\boldsymbol{r}|^3} - 3\boldsymbol{r}\,\frac{(\boldsymbol{r}\cdot\boldsymbol{V})}{|\boldsymbol{r}|^5}\right). \qquad (17)$$

Further,

$$\tau d\boldsymbol{V}dM = \tau(\boldsymbol{V};M)d\boldsymbol{V}dM \qquad (18)$$

gives the probability that a star with a relative velocity in the range $(\boldsymbol{V}, \boldsymbol{V} + d\boldsymbol{V})$ and with a mass between M and $M + dM$ will be found. It should also be noted that in writing equations (15) and (16) we have supposed that the fluctuations in the local stellar distribution which occur are subject only to the restriction of a constant average density.

Since

$$\frac{3}{4\pi R^3}\int_{M=0}^{\infty}\int_{|\boldsymbol{r}|=0}^{R}\int_{|\boldsymbol{V}|=0}^{\infty}\tau d\boldsymbol{V}d\boldsymbol{r}dM = 1, \qquad (19)$$

we can rewrite (16) as

$$A(\boldsymbol{\varrho},\boldsymbol{\sigma}) = \underset{R\to\infty}{\text{Limit}}\left[1 - \frac{3}{4\pi R^3}\int_{M=0}^{\infty}\int_{|\boldsymbol{r}|=0}^{R}\int_{|\boldsymbol{V}|=0}^{\infty}[1 - e^{i(\boldsymbol{\varrho}\cdot\boldsymbol{\phi}+\boldsymbol{\sigma}\cdot\boldsymbol{\psi})}]\tau d\boldsymbol{V}d\boldsymbol{r}dM\right]^{4\pi R^3 n/3}. \qquad (20)$$

The integral over $\boldsymbol{r}$ which occurs in equation (20) is seen to be absolutely convergent when extended over all $|\boldsymbol{r}|$ i.e., also for $|\boldsymbol{r}| \to \infty$. Hence, we can write

[5] Cf. **Chandrasekhar, S., & von Neumann, J.**, Astrophysical Jour. **95**: 489. 1942. (§ 2).

$$A(\boldsymbol{\rho},\boldsymbol{\sigma}) = \underset{R\to\infty}{\text{Limit}} \left[1 - \frac{3}{4\pi R^3}\int_{M=0}^{\infty}\int_{|\boldsymbol{r}|=0}^{\infty}\int_{|\boldsymbol{V}|=0}^{\infty}[1 - e^{i(\boldsymbol{\rho}\cdot\boldsymbol{\phi}+\boldsymbol{\sigma}\cdot\boldsymbol{\psi})}]\tau dVdrdM\right]^{4\pi R^3 n/3}, \tag{21}$$

or

$$A(\boldsymbol{\rho},\boldsymbol{\sigma}) = e^{-nC(\boldsymbol{\rho},\boldsymbol{\sigma})}, \tag{22}$$

where

$$C(\boldsymbol{\rho},\boldsymbol{\sigma}) = \int_0^{\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[1 - e^{i(\boldsymbol{\rho}\cdot\boldsymbol{\phi}+\boldsymbol{\sigma}\cdot\boldsymbol{\psi})}]\tau drdVdM. \tag{23}$$

This formally solves the problem. It does not, however, appear that the integral representing $C(\boldsymbol{\rho},\boldsymbol{\sigma})$ can be evaluated explicitly in terms of any of the known functions. But if we are interested only in the distribution $W(\boldsymbol{F})$ of $\boldsymbol{F}$ and in the moments of $\boldsymbol{f}$ for a given $\boldsymbol{F}$ then we need only the behavior of $A(\boldsymbol{\rho},\boldsymbol{\sigma})$, and therefore also of $C(\boldsymbol{\rho},\boldsymbol{\sigma})$, for $|\boldsymbol{\sigma}| \to 0$, for the distribution $W(\boldsymbol{F})$ is clearly given by

$$W(\boldsymbol{F}) = \int_{-\infty}^{+\infty} W(\boldsymbol{F},\boldsymbol{f})d\boldsymbol{f}. \tag{24}$$

Similarly, the first and the second moments of the components f_ξ, f_η, and f_ζ of $\boldsymbol{f}$ along three directions ξ, η and ζ at right angles to each other are given by

$$W(\boldsymbol{F})\overline{f_\mu} = \int_{-\infty}^{+\infty} W(\boldsymbol{F},\boldsymbol{f})f_\mu d\boldsymbol{f} \qquad (\mu = \xi,\eta,\zeta), \tag{25}$$

and

$$W(\boldsymbol{F})\overline{f_\mu f_\nu} = \int_{-\infty}^{+\infty} W(\boldsymbol{F},\boldsymbol{f})f_\mu f_\nu d\boldsymbol{f} \qquad (\mu,\nu = \xi,\eta,\zeta). \tag{26}$$

Substituting now for $W(\boldsymbol{F},\boldsymbol{f})$ according to equation (15) in the foregoing equations we obtain

$$\left.\begin{aligned}
W(\boldsymbol{F}) &= \frac{1}{64\pi^6}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-i(\boldsymbol{\rho}\cdot\boldsymbol{F}+\boldsymbol{\sigma}\cdot\boldsymbol{f})}\, A(\boldsymbol{\rho},\boldsymbol{\sigma})d\boldsymbol{f}d\boldsymbol{\rho}d\boldsymbol{\sigma},\\
W(\boldsymbol{F})\overline{f_\mu} &= \frac{1}{64\pi^6}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-i(\boldsymbol{\rho}\cdot\boldsymbol{F}+\boldsymbol{\sigma}\cdot\boldsymbol{f})}\, A(\boldsymbol{\rho},\boldsymbol{\sigma})f_\mu d\boldsymbol{f}d\boldsymbol{\rho}d\boldsymbol{\sigma},\\
W(\boldsymbol{F})\overline{f_\mu f_\nu} &= \frac{1}{64\pi^6}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-i(\boldsymbol{\rho}\cdot\boldsymbol{F}+\boldsymbol{\sigma}\cdot\boldsymbol{f})}\, A(\boldsymbol{\rho},\boldsymbol{\sigma})f_\mu f_\nu d\boldsymbol{f}d\boldsymbol{\rho}d\boldsymbol{\sigma}.
\end{aligned}\right\} \tag{27}$$

But

$$
\left.
\begin{aligned}
\frac{1}{8\pi^3}\int_{-\infty}^{+\infty} e^{-i\boldsymbol{\sigma}\cdot\boldsymbol{f}} d\boldsymbol{f} &= \delta(\sigma_\xi)\delta(\sigma_\eta)\delta(\sigma_\zeta),\\
\frac{1}{8\pi^3}\int_{-\infty}^{+\infty} e^{-i\boldsymbol{\sigma}\cdot\boldsymbol{f}} f_\xi d\boldsymbol{f} &= i\delta'(\sigma_\xi)\delta(\sigma_\eta)\delta(\sigma_\zeta),\\
\frac{1}{8\pi^3}\int_{-\infty}^{+\infty} e^{-i\boldsymbol{\sigma}\cdot\boldsymbol{f}} f_\xi^2 d\boldsymbol{f} &= -\delta''(\sigma_\xi)\delta(\sigma_\eta)\delta(\sigma_\zeta),\\
\frac{1}{8\pi^3}\int_{-\infty}^{+\infty} e^{-i\boldsymbol{\sigma}\cdot\boldsymbol{f}} f_\xi f_\eta d\boldsymbol{f} &= -\delta'(\sigma_\xi)\delta'(\sigma_\eta)\delta(\sigma_\zeta),
\end{aligned}
\right\}
\tag{28}
$$

etc. In equations (28), δ denotes Dirac's δ-function and δ' and δ'' its first and second derivatives; and remembering also that

$$
\left.
\begin{aligned}
\int_{-\infty}^{+\infty} f(x)\delta(x)dx = f(0); \qquad \int_{-\infty}^{+\infty} f(x)\delta'(x)dx = -f'(0);\\
\int_{-\infty}^{+\infty} f(x)\delta''(x)dx = f''(0),
\end{aligned}
\right\}
\tag{29}
$$

equations (27) reduce to

$$
W(\boldsymbol{F}) = \frac{1}{8\pi^3}\int_{-\infty}^{+\infty} e^{-i\boldsymbol{\rho}\cdot\boldsymbol{F}}\left[A(\boldsymbol{\rho},\boldsymbol{\sigma})\right]_{|\boldsymbol{\sigma}|=0} d\boldsymbol{\rho},
\tag{30}
$$

$$
W(\boldsymbol{F})\overline{f_\mu} = -\frac{i}{8\pi^3}\int_{-\infty}^{+\infty} e^{-i\boldsymbol{\rho}\cdot\boldsymbol{F}}\left[\frac{\partial}{\partial\sigma_\mu}A(\boldsymbol{\rho},\boldsymbol{\sigma})\right]_{|\boldsymbol{\sigma}|=0} d\boldsymbol{\rho},
\tag{31}
$$

and

$$
W(\boldsymbol{F})\overline{f_\mu f_\nu} = -\frac{1}{8\pi^3}\int_{-\infty}^{+\infty} e^{-i\boldsymbol{\rho}\cdot\boldsymbol{F}}\left[\frac{\partial^2}{\partial\sigma_\mu\partial\sigma_\nu}A(\boldsymbol{\rho},\boldsymbol{\sigma})\right]_{|\boldsymbol{\sigma}|=0} d\boldsymbol{\rho}.
\tag{32}
$$

We accordingly see that the distribution function $W(\boldsymbol{F})$ and all the first and the second moments of $\boldsymbol{f}$ for a given $\boldsymbol{F}$ can be evaluated from a series expansion of $A(\boldsymbol{\rho},\boldsymbol{\sigma})$ [or of $C(\boldsymbol{\rho},\boldsymbol{\sigma})$] which is correct up to the second order in $|\boldsymbol{\sigma}|$. The development of such a series is long and tedious

Omitting, therefore, all the details of the calculations, we give only the final result. It is found that

$$\left.\begin{aligned}
C(\boldsymbol{\varrho},\boldsymbol{\sigma}) = {} & \frac{4}{15}(2\pi)^{3/2}G^{3/2}\overline{M^{3/2}}|\boldsymbol{\varrho}|^{3/2} \\
& + \frac{2}{3}\pi i G_1(\sigma_1\overline{MV_1} + \sigma_2\overline{MV_2} - 2\sigma_3\overline{MV_3}) \\
& + \frac{3}{28}(2\pi)^{3/2}G^{1/2}|\boldsymbol{\varrho}|^{-3/2}[(5\sigma_1^2 + 4\sigma_2^2 - 2\sigma_3^2)\overline{M^{1/2}V_1^2} \\
& + (4\sigma_1^2 + 5\sigma_2^2 - 2\sigma_3^2)\overline{M^{1/2}V_2^2} \\
& + (4\sigma_3^2 - 2\sigma_1^2 - 2\sigma_2^2)\overline{M^{1/2}V_3^2} - 8\sigma_2\sigma_3\overline{M^{1/2}V_2V_3} \\
& - 8\sigma_1\sigma_3\overline{M^{1/2}V_1V_3} + 2\sigma_1\sigma_2\overline{M^{1/2}V_1V_2}] + O(|\boldsymbol{\sigma}|^3), \quad (|\boldsymbol{\sigma}|\to 0),
\end{aligned}\right\} \tag{33}$$

where a bar indicates that the corresponding quantity has been averaged with the weight function $\tau(\boldsymbol{V};M)$ (see equation [18]); further, in equation (33), $(\sigma_1,\sigma_2,\sigma_3)$ and (V_1,V_2,V_3) are the components of $\boldsymbol{\sigma}$ and $\boldsymbol{V}$ in a system of coordinates in which the Z-axis is in the direction of $\boldsymbol{\varrho}$.

In equation (33) $\boldsymbol{V} = (V_1,V_2,V_3)$ denotes of course the velocity of a field star relative to the one under consideration. If we now let $\boldsymbol{u}$ and $\boldsymbol{v}$ denote respectively the velocities of the field star and the star under consideration in an appropriately chosen local standard of rest, then

$$\boldsymbol{V} = \boldsymbol{u} - \boldsymbol{v}. \tag{34}$$

In our further discussion we shall introduce the assumption that the distribution of the velocities $\boldsymbol{u}$ among the stars is *spherical*; i.e., the distribution function $\Psi(\boldsymbol{u})$ has the form

$$\Psi(\boldsymbol{u}) \equiv \Psi[j^2(M)|\boldsymbol{u}|^2], \tag{35}$$

where Ψ is an arbitrary function of the argument specified and the parameter j (of the dimensions of [velocity]$^{-1}$) can be a function of the mass of the star. This assumption for the distribution of the peculiar velocities $\boldsymbol{u}$ implies that the probability function $\tau(\boldsymbol{V};M)$ must be expressible as

$$\tau(\boldsymbol{V};M) = \Psi[j^2(M)|\boldsymbol{u}|^2]\chi(M) \tag{36}$$

where $\chi(M)$ governs the distribution over the different masses. For a function τ of this form we clearly have

$$\left.\begin{aligned}
\overline{MV_i} = -\overline{M}v_i; \qquad \overline{M^{1/2}V_i^2} = \frac{1}{3}\overline{M^{1/2}|\boldsymbol{u}|^2} + \overline{M^{1/2}}v_i^2 \qquad (i = 1,2,3) \\
\overline{M^{1/2}V_iV_j} = \overline{M^{1/2}}v_iv_j, \qquad [i,j = 1,2,3; \qquad i \neq j)].
\end{aligned}\right\} \tag{37}$$

Substituting these values in equation (33) we find, after some minor reductions, that

$$
\left.\begin{aligned}
C(\boldsymbol{\varrho},\boldsymbol{\sigma}) = {} & \frac{4}{15}(2\pi)^{3/2}G^{3/2}\overline{M^{3/2}}|\boldsymbol{\varrho}|^{3/2} - \frac{2}{3}\pi i G\overline{M}(\sigma_1 v_1 + \sigma_2 v_2 - 2\sigma_3 v_3) \\
& + \frac{1}{4}(2\pi)^{3/2}G^{1/2}\overline{M^{1/2}|\boldsymbol{u}|^2}|\boldsymbol{\varrho}|^{-3/2}(\sigma_1^2 + \sigma_2^2) \\
& + \frac{3}{28}(2\pi)^{3/2}G^{1/2}\overline{M^{1/2}}|\boldsymbol{\varrho}|^{-3/2}\{\sigma_1^2(5v_1^2 + 4v_2^2 - 2v_3^2) \\
& + \sigma_2^2(4v_1^2 + 5v_2^2 - 2v_3^2) + \sigma_3^2(4v_3^2 - 2v_1^2 - 2v_2^2) \\
& - 8\sigma_2\sigma_3 v_1 v_3 - 8\sigma_3\sigma_1 v_3 v_1 + 2\sigma_1\sigma_2 v_1 v_2\} + O(|\boldsymbol{\sigma}|^3), \quad (|\boldsymbol{\sigma}| \to 0).
\end{aligned}\right\} \tag{38}
$$

With a series expansion of this form, we can, as we have already remarked, evaluate the distribution $W(\boldsymbol{F})$ and all the first and the second moments of $\boldsymbol{f}$ for a given $\boldsymbol{F}$.

THE DISTRIBUTION $W(\boldsymbol{F})$

According to equations (30) and (38) we have

$$
W(\boldsymbol{F}) = \frac{1}{8\pi^3}\int_{-\infty}^{+\infty} e^{-i\boldsymbol{\varrho}\cdot\boldsymbol{F} - a|\boldsymbol{\varrho}|^{3/2}}d\boldsymbol{\varrho}, \tag{39}
$$

where we have written

$$
a = \frac{4}{15}(2\pi G)^{3/2}\overline{M^{3/2}}n. \tag{40}
$$

From equation (39) we derive the formula[6]

$$
W(\boldsymbol{F}) = \frac{1}{4\pi a^2}\frac{H(\beta)}{\beta^2}, \tag{41}
$$

where

$$
H(\beta) = \frac{2}{\pi\beta}\int_0^{\infty} e^{-(x/\beta)^{3/2}}x \sin x\, dx, \tag{42}
$$

and β measures $|\boldsymbol{F}|$ in units of Q_H where

$$
Q_H = a^{2/3} = 2.6031G(\overline{M^{3/2}}n)^{2/3}. \tag{43}
$$

The function $H(\beta)$ has been numerically evaluated and tabulated in Chandrasekhar and von Neumann's paper.

THE FIRST MOMENT OF $\boldsymbol{f}$: DYNAMICAL FRICTION

Turning next to the first moment of $\boldsymbol{f}$ it is found after some lengthy calculations that

$$
\bar{\boldsymbol{f}} = \overline{\left(\frac{d\boldsymbol{F}}{dt}\right)}_{\boldsymbol{F},\boldsymbol{v}} = -\frac{2}{3}\pi G\overline{M}nB\left(\frac{|\boldsymbol{F}|}{Q_H}\right)\left(\boldsymbol{v} - 3\frac{\boldsymbol{v}\cdot\boldsymbol{F}}{|\boldsymbol{F}|^2}\boldsymbol{F}\right), \tag{44}
$$

where Q_H is the normal field defined in equation (43) and

[6] Cf. **Chandrasekhar, S., & von Neumann, J.** Astrophysical Jour. **95**: 489. 1942. (§ 7).

$$B(\beta) = 3\frac{\int_0^\beta H(\beta)d\beta}{\beta H(\beta)} - 1. \tag{45}$$

The function $B(\beta)$ has the following asymptotic forms:

$$\left.\begin{aligned} B(\beta) &\to \frac{1}{15}\Gamma\left(\frac{10}{3}\right)\beta^2 \qquad (\beta \to 0), \\ B(\beta) &\to \frac{8}{5}\sqrt{\frac{\pi}{2}}\,\beta^{3/2} \qquad (\beta \to \infty). \end{aligned}\right\} \tag{46}$$

We shall first examine certain formal consequences of equation (44).

Multiplying equation (44) scalarly with $\boldsymbol{F}$ we obtain

$$\boldsymbol{F}.\overline{\left(\frac{d\boldsymbol{F}}{dt}\right)}_{\boldsymbol{F},\boldsymbol{v}} = \frac{4}{3}\pi G\overline{M}nB\left(\frac{|\boldsymbol{F}|}{Q_H}\right)(\boldsymbol{v}.\boldsymbol{F}) \tag{47}$$

but

$$\boldsymbol{F}.\overline{\left(\frac{d\boldsymbol{F}}{dt}\right)}_{\boldsymbol{F},\boldsymbol{v}} = |\boldsymbol{F}|\overline{\left(\frac{d|\boldsymbol{F}|}{dt}\right)}_{\boldsymbol{F},\boldsymbol{v}}. \tag{48}$$

Hence,

$$\overline{\left(\frac{d|\boldsymbol{F}|}{dt}\right)}_{\boldsymbol{F},\boldsymbol{v}} = \frac{4}{3}\pi G\overline{M}nB\left(\frac{|\boldsymbol{F}|}{Q_H}\right)\frac{\boldsymbol{v}.\boldsymbol{F}}{|\boldsymbol{F}|}. \tag{49}$$

On the other hand, if F_j denotes the component of $\boldsymbol{F}$ in an arbitrary direction at right angles to the direction of $\boldsymbol{v}$ then, according to equation (44),

$$\overline{\left(\frac{dF_j}{dt}\right)}_{\boldsymbol{F},\boldsymbol{v}} = 2\pi G\overline{M}nB\left(\frac{|\boldsymbol{F}|}{Q_H}\right)\frac{\boldsymbol{v}.\boldsymbol{F}}{|\boldsymbol{F}|^2}F_j. \tag{50}$$

Combining equations (49) and (50) we have

$$\frac{1}{F_j}\overline{\left(\frac{dF_j}{dt}\right)}_{\boldsymbol{F},\boldsymbol{v}} = \frac{3}{2}\frac{1}{|\boldsymbol{F}|}\overline{\left(\frac{d|\boldsymbol{F}|}{dt}\right)}_{\boldsymbol{F},\boldsymbol{v}}. \tag{51}$$

Equation (51) is clearly equivalent to

$$\overline{\left[\frac{d}{dt}\left(\log F_j - \frac{3}{2}\log|\boldsymbol{F}|\right)\right]}_{\boldsymbol{F},\boldsymbol{v}} = 0. \tag{52}$$

We have thus proved that

$$\overline{\left[\frac{d}{dt}\left(\frac{F_j}{|\boldsymbol{F}|^{3/2}}\right)\right]}_{\boldsymbol{F},\boldsymbol{v}} = 0. \tag{53}$$

We shall now examine the physical consequences of equation (44) more closely. In words, the meaning of this equation is that the component of

$$-\frac{2}{3}\pi G\overline{M}nB\left(\frac{|F|}{Q_H}\right)\left(\nu - 3\frac{\nu\cdot F}{|F|^2}F\right) \tag{54}$$

along any particular direction gives the average value of the rate of change in the force F per unit mass acting on a star that is to be expected in the specified direction, when the star is moving with a velocity ν in an appropriately chosen local standard of rest. Stated in this manner, we at once see the essential difference in the stochastic variations of F with time in the two cases $|\nu| = 0$ and $|\nu| \neq 0$. In the former case, $\overline{\dot{F}} \equiv 0$; but this is not generally true when $|\nu| \neq 0$. Or expressed differently, when $|\nu| = 0$ the changes in F occur with equal probability in all directions, while this is not the case when $|\nu| \neq 0$. The true nature of this difference is brought out very clearly when we consider

$$\overline{\left(\frac{d|F|}{dt}\right)}_{F,\nu} \tag{55}$$

according to equation (49). Remembering that $B(\beta) \geq 0$ for $\beta \geq 0$, we conclude from equation (49) that

$$\overline{\left(\frac{d|F|}{dt}\right)}_{F,\nu} > 0 \quad \text{if} \quad \nu\cdot F > 0, \tag{56}$$

and

$$\overline{\left(\frac{d|F|}{dt}\right)}_{F,\nu} < 0 \quad \text{if} \quad \nu\cdot F < 0. \tag{57}$$

In other words, if F has a positive component in the direction of ν, F increases on the average, while if F has a negative component in the direction of ν, $|F|$ decreases on the average. This essential asymmetry introduced by the direction of ν may be expected to give rise to the phenomenon of dynamical friction.

The characteristic aspects of the situation governed by equation (44) are best understood when we contrast it with the case $|\nu| = 0$. Under these circumstances, we can visualize the motion of the representative point in the velocity space somewhat as follows.[7] The representative point suffers small random displacements in a manner that can be adequately described by the problem of random flights or more generally as Brownian motion. More specifically, the star may be assumed to suffer a large number of discrete increases in velocity of amounts $T(|F|)F$, where T denotes the mean life of the state $|F|$ (see subsection below). Moreover, these increases may be assumed to take place in

[7] Cf. Chandrasekhar, S., Astrophysical Jour. 94: 511. 1941. (§§ 2 and 7.)

random directions. Accordingly, we may conclude that the mean square increase $\overline{|\Delta \boldsymbol{v}|^2}$ in the velocity to be expected in time t is given by

$$\overline{|\Delta \boldsymbol{v}|^2} = \overline{|\boldsymbol{F}|^2 T} t. \tag{58}$$

An alternative way of describing the same situation is that if we denote by $W(\boldsymbol{v};t)$ the probability that the star has a velocity $\boldsymbol{v}$ at time t when the velocity at $t = 0$ is $\boldsymbol{v}_0$, then W satisfies the *diffusion equation*

$$\frac{\partial W}{\partial t} = q\left(\frac{\partial^2 W}{\partial v_1^2} + \frac{\partial^2 W}{\partial v_2^2} + \frac{\partial^2 W}{\partial v_3^2}\right), \tag{59}$$

with the *"coefficient of diffusion"* q having the value

$$q = \frac{1}{6}\overline{|\boldsymbol{F}|^2 T}. \tag{60}$$

The solution of equation (59) for our purposes then is

$$W(\boldsymbol{v},t;\boldsymbol{v}_0) = \frac{1}{(4\pi q t)^{3/2}} e^{-|\boldsymbol{v}-\boldsymbol{v}_0|^2/4qt}. \tag{61}$$

The formula (58) is seen to be an immediate consequence of the solution (61).

Returning to the discussion of the case governed by equations (44) and (49), it is at once clear that the idealization of the motion of the representative point in the velocity space, as a problem in random flights, can no longer be valid. For, according to (56) and (57), during a given state of fluctuation of $\boldsymbol{F}$ a star is likely to suffer a greater absolute amount of acceleration if $(\boldsymbol{v}\cdot\boldsymbol{F})$ is negative than if $(\boldsymbol{v}\cdot\boldsymbol{F})$ is positive. But the *a priori* probability for $(\boldsymbol{v}\cdot\boldsymbol{F})$ to be positive or negative is equal. Hence, when integrated over a large number of fluctuations the star must suffer cumulatively a larger absolute amount of acceleration in a direction opposite to its own direction of motion than in the direction of motion. In other words we may expect a net tendency for the star to be relatively decelerated in the direction of its motion; further, this tendency is proportional to $|\boldsymbol{v}|$. But these are exactly what are implied by the existence of dynamical friction. (See Part II where the question of dynamical friction is considered in greater detail.)

THE SECOND MOMENT OF $[f]$ AND THE MEAN LIFE OF THE STATE $[F]$

According to equation (32)

$$W(F)\overline{|\boldsymbol{f}|^2} = -\frac{1}{8\pi^3}\int_{-\infty}^{+\infty} e^{-i\boldsymbol{\varrho}\cdot\boldsymbol{F}}\left[\nabla_{\sigma}^2 A(\boldsymbol{\varrho},\boldsymbol{\sigma})\right]_{|\sigma|=0} d\boldsymbol{\varrho}. \tag{62}$$

Using the expansion (33) for $C(\boldsymbol{\rho},\boldsymbol{\vartheta})$ we find after some lengthy calculations that

$$\left.\begin{aligned} \overline{|\boldsymbol{f}|^2}_{\boldsymbol{F},\boldsymbol{\nu}} = 2ab\,\frac{\beta^{1/2}}{H(\beta)}\Big\{ 2G(\beta) + 7k[\sin^2\alpha G(\beta) - (3\sin^2\alpha - 2)I(\beta)]\Big\} \\ + \frac{g^2}{\beta H(\beta)}\Big\{(4 - 3\sin^2\alpha)\beta H(\beta) + 3(3\sin^2\alpha - 2)K(\beta)\Big\}. \end{aligned}\right\} \tag{63}$$

where α denotes the angle between the directions of $\boldsymbol{F}$ and $\boldsymbol{\nu}$,

$$\left.\begin{aligned} a &= \frac{4}{15}(2\pi G)^{3/2}\overline{M^{3/2}}n, & b &= \frac{1}{4}(2\pi)^{3/2}G^{1/2}\overline{M^{1/2}|\boldsymbol{u}|^2}n, \\ g &= \frac{2}{3}\pi G\overline{M}|\boldsymbol{\nu}|n, & k &= \frac{3}{7}\frac{\overline{M^{1/2}}|\boldsymbol{\nu}|^2}{\overline{M^{1/2}|\boldsymbol{u}|^2}}, \end{aligned}\right\} \tag{64}$$

and

$$\left.\begin{aligned} H(\beta) &= \frac{2}{\pi\beta}\int_0^\infty e^{-(x/\beta)^{3/2}}x\sin x\,dx, \\ G(\beta) &= \frac{3}{2}\int_0^\beta \beta^{-3/2}H(\beta)d\beta, \\ I(\beta) &= \beta^{-3/2}\int_0^\beta \beta^{1/2}G(\beta)d\beta, \\ K(\beta) &= \int_0^\beta H(\beta)d\beta. \end{aligned}\right\} \tag{65}$$

Averaging equation (63) for all possible mutual orientations of the two vectors $\boldsymbol{F}$ and $\boldsymbol{\nu}$ we readily find that

$$\overline{\overline{|\boldsymbol{f}|^2}}_{|\boldsymbol{F}|,|\boldsymbol{\nu}|} = 4ab\left\{\frac{\beta^{1/2}G(\beta)}{H(\beta)}\left(1 + \frac{7}{3}k\right) + \frac{g^2}{2ab}\right\}, \tag{66}$$

or substituting for k and $g^2/2ab$ from (64), we find that

$$\overline{\overline{|\boldsymbol{f}|^2}}_{|\boldsymbol{F}|,|\boldsymbol{\nu}|} = 4ab\left\{\frac{\beta^{1/2}G(\beta)}{H(\beta)}\left(1 + \frac{\overline{M^{1/2}}|\boldsymbol{\nu}|^2}{\overline{M^{1/2}|\boldsymbol{u}|^2}}\right) + \frac{5}{12\pi}\frac{\overline{M}^2|\boldsymbol{\nu}|^2}{\overline{M^{3/2}}\,\overline{M^{1/2}|\boldsymbol{u}|^2}}\right\}. \tag{67}$$

In terms of equation (67) we can define an approximate formula for the mean life of the state $|\boldsymbol{F}|$ according to [8]

$$T_{|\boldsymbol{F}|,|\boldsymbol{\nu}|} = \frac{|\boldsymbol{F}|}{\sqrt{\overline{\overline{|\boldsymbol{f}|^2}}_{|\boldsymbol{F}|,|\boldsymbol{\nu}|}}}. \tag{68}$$

[8] Cf. **Chandrasekhar, S., & von Neumann, J.** Astrophysical Jour. **95**: 489. 1942. Equation (167).

Combining equations (67) and (68) we find that

$$T_{|\boldsymbol{F}|,|\boldsymbol{v}|} = T_{|\boldsymbol{F}|,0}\frac{1}{\left[1+\dfrac{\overline{M^{1/2}|\boldsymbol{v}|^2}}{\overline{M^{1/2}|\boldsymbol{u}|^2}}+\dfrac{5}{12\pi}\dfrac{\overline{M^2}\,|\boldsymbol{v}|^2}{\overline{M^{3/2}}\,\overline{M^{1/2}|\boldsymbol{u}|^2}}\dfrac{H(\beta)}{\beta^{1/2}G(\beta)}\right]^{1/2}}, \tag{69}$$

where $T_{|\boldsymbol{F}|,0}$ denotes the mean life when $|\boldsymbol{v}| = 0$:

$$T_{|\boldsymbol{F}|,0} = \sqrt{\frac{a^{1/3}}{4b}\frac{\beta^{3/2}H(\beta)}{G(\beta)}}. \tag{70}$$

Equation (70) suggests that we measure T in terms of the following unit, t_0, which appears natural to this problem:

$$t_0 = \sqrt{\frac{a^{1/3}}{4b}}. \tag{71}$$

Substituting for a and b from equation (64), we find that

$$\left.\begin{aligned} t_0 &= \frac{1}{(30)^{1/6}\pi^{1/2}}\left(\frac{[\overline{M^{3/2}}]^{1/3}}{[\overline{M^{1/2}|\boldsymbol{u}|^2}]}\right)^{1/2}\frac{1}{n^{1/3}} \\ &= \frac{0.3201}{n^{1/3}}\left(\frac{[\overline{M^{3/2}}]^{1/3}}{\overline{M^{1/2}|\boldsymbol{u}|^2}}\right)^{1/2}. \end{aligned}\right\} \tag{72}$$

And, finally, if we denote by $\tau(\beta;|\boldsymbol{v}|)$ the mean life expressed in this unit, we have

$$\tau(\beta;|\boldsymbol{v}|) = \tau(\beta;0)\frac{1}{\left[1+\dfrac{\overline{M^{1/2}|\boldsymbol{v}|^2}}{\overline{M^{1/2}|\boldsymbol{u}|^2}}+\dfrac{5}{12\pi}\dfrac{\overline{M^2}|\boldsymbol{v}|^2}{\overline{M^{1/2}|\boldsymbol{u}|^2}}\dfrac{H(\beta)}{\beta^{1/2}G(\beta)}\right]^{1/2}}. \tag{73}$$

From equation (73) we derive the asymptotic formulae

$$\tau \to \beta\frac{1}{\left[1+\dfrac{\overline{M^{1/2}|\boldsymbol{v}|^2}}{\overline{M^{1/2}|\boldsymbol{u}|^2}}+\dfrac{5}{12\pi}\dfrac{\overline{M^2}|\boldsymbol{v}|^2}{\overline{M^{3/2}}\,\overline{M^{1/2}|\boldsymbol{u}|^2}}\right]^{1/2}}\quad(\beta\to 0), \tag{74}$$

and

$$\tau \to \sqrt{\frac{15}{8}}\frac{1}{\left[1+\dfrac{\overline{M^{1/2}|\boldsymbol{v}|^2}}{\overline{M^{1/2}|\boldsymbol{u}|^2}}\right]^{1/2}}\beta^{-1/2}\qquad(\beta\to\infty). \tag{75}$$

The function $\tau(\beta;0)$ is tabulated in Chandrasekhar and von Neumann's paper. Our present results show that approximately

$$\tau(\beta;|\boldsymbol{v}|) \sim \tau(\beta;0)\frac{1}{\left[1+\dfrac{\overline{M^{1/2}|\boldsymbol{v}|^2}}{\overline{M^{1/2}|\boldsymbol{u}|^2}}\right]^{1/2}}. \tag{76}$$

According to equations (74) and (75), the approximate formula (76) may be expected to give values of τ correct to within 15 per cent over the entire range of β.

We may particularly draw attention to the very short lives of the weak fields.

II. DYNAMICAL FRICTION AND THE PRINCIPLES OF STATISTICAL DYNAMICS

General Considerations

As we have seen in Part I, in a first approximative discussion of the fluctuating part of the gravitational field acting on a star, we may suppose that the probability function $W(\boldsymbol{u},t)$ governing the occurrence of the velocity $\boldsymbol{u}$ at time t satisfies the diffusion equation (see equation [59]),

$$\frac{\partial W}{\partial t} = q\nabla_u^2 W \cdot \tag{77}$$

According to this equation, the probability distribution of the velocities $\boldsymbol{u}$ at time t when it is known with certainty that the star had the velocity $\boldsymbol{u}_0$ at time $t = 0$ is given by

$$W(\boldsymbol{u},t;\boldsymbol{u}_0) = \frac{1}{(4\pi qt)^{3/2}} e^{-|\boldsymbol{u}-\boldsymbol{u}_0|^2/4qt}. \tag{78}$$

We shall now indicate why the considerations outlined above cannot be valid for times which are long compared to $|\boldsymbol{u}|^2/q$ where $|\boldsymbol{u}|^2$ denotes the mean square velocity of the stars in an appropriately chosen local standard of rest. For, if $W(\boldsymbol{u},t;\boldsymbol{u}_0)$ according to equation (78) were valid for all times, then the probability that a star may have suffered any arbitrarily assigned large acceleration can be made as close to unity as we may choose by letting t approach infinity. This is, however, contrary to what we should expect on general grounds, namely, that $W(\boldsymbol{u},t;\boldsymbol{u}_0)$ approaches a Maxwellian distribution independently of $\boldsymbol{u}_0$ as $t \to \infty$. Expressed somewhat differently, we should strictly suppose that the stochastic variations in the velocity which a star experiences must be such as to leave an initial Maxwellian distribution of the velocities invariant. This is evidently not the case with the process described by equation (77). And the question now arises as to how we can generalize our earlier approximate considerations leading to equation (77) so that the underlying stochastic process may satisfy the criterion stated above. We shall now show how this can be achieved by the introduction of *dynamical friction*. More specifically, we shall suppose that the acceleration $\Delta\boldsymbol{u}$ which a star experiences in an interval of time Δt (long compared to the periods of the elementary fluctuations in $\boldsymbol{F}$ but short compared to the intervals during which $\boldsymbol{u}$ may be expected to change appreciably) can be expressed as the sum of two terms in the form

$$\Delta \boldsymbol{u} = \delta \boldsymbol{u}(\Delta t) - \eta \boldsymbol{u} \Delta t \tag{79}$$

where the first term on the right-hand side is governed by the probability distribution

$$\psi(\delta \boldsymbol{u}[\Delta t]) = \frac{1}{(4\pi q \Delta t)^{3/2}} e^{-|\delta \boldsymbol{u} - \mathrm{grad}_{\boldsymbol{u}} q\, \Delta t|^2/4q\Delta t}, \tag{80}$$

and where the second term $-\eta \boldsymbol{u} \Delta t$ represents a *deceleration* of the star in the direction of its motion by an amount depending on $|\boldsymbol{u}|$. The constant of proportionality η can therefore be properly called the *coefficient of dynamical friction.*

With the underlying stochastic process defined as in equations (79) and (80), the probability distribution $W(\boldsymbol{u},t + \Delta t)$ of $\boldsymbol{u}$ at time $t + \Delta t$ can be derived from the distribution $W(\boldsymbol{u},t)$ at the earlier time t by means of the integral equation

$$W(\boldsymbol{u},t+\Delta t) = \int_{-\infty}^{+\infty} W(\boldsymbol{u}-\Delta \boldsymbol{u},t)\psi(\boldsymbol{u}-\Delta \boldsymbol{u};\Delta \boldsymbol{u})d(\Delta \boldsymbol{u}), \tag{81}$$

where $\psi(\boldsymbol{u};\Delta \boldsymbol{u})$ denotes the transition probability (see equation [80])

$$\psi(\boldsymbol{u};\Delta \boldsymbol{u}) = \frac{1}{(4\pi q \Delta t)^{3/2}} e^{-|\Delta \boldsymbol{u} - \mathrm{grad}_{\boldsymbol{u}} q\Delta t + \eta \boldsymbol{u} \Delta t|^2/4q\Delta t} \tag{82}$$

Expanding $W(\boldsymbol{u},t + \Delta t)$, $W(\boldsymbol{u} - \Delta \boldsymbol{u},t)$ and $\psi(\boldsymbol{u} - \Delta \boldsymbol{u};\Delta \boldsymbol{u})$ which occur in equation (81) in the form of Taylor series, evaluating the various moments of $\Delta \boldsymbol{u}$ according to equation (82) and passing finally to the limit we obtain the following equation

$$\frac{\partial W}{\partial t} = \mathrm{div}_{\boldsymbol{u}}\, (q \,\mathrm{grad}_{\boldsymbol{u}}\, W + \eta W \boldsymbol{u}). \tag{83}$$

Finally, the condition that the Maxwellian distribution

$$\left(\frac{3}{2\pi \overline{|\boldsymbol{u}|^2}}\right)^{3/2} e^{-3|\boldsymbol{u}|^2/2\overline{|\boldsymbol{u}|^2}} \tag{84}$$

satisfy equation (83) *identically* requires that q and η be related according to

$$q = \frac{1}{3}\, \overline{|\boldsymbol{u}|^2} \eta. \tag{85}$$

Summarizing the conclusion reached, we may say that *general considerations, such as the invariance of the Maxwellian distribution to the underlying stochastic process, require that stars experience dynamical friction during their motion.*

The Resolution of Certain Fallacies and an Elementary Derivation of the Coefficient of Dynamical Friction

The conclusion we have reached in the preceding paragraph appears contrary to what might be expected on first sight. For we *might* argue in the following manner:

(a) Suppose we consider a star with a velocity $|\boldsymbol{u}|$ appreciably less than the root mean square velocity $(\overline{|\boldsymbol{u}|^2})^{1/2}$. We would then expect it to encounter oftener stars with velocities greater than its own than stars with velocities less than its own. And, consequently, we might be led to believe that stars with velocities less than the average would be systematically accelerated and, similarly, that stars with velocities greater than the average would be systematically decelerated.

How then does dynamical friction come to operate on *all* stars? Before we answer this question we shall state the second paradox.

(b) We might go farther and even argue that the conclusion reached in (a) is "reasonable." For, it might be supposed that systematically different effects on stars with relatively large, respectively small velocities, are required for the statistical maintenance of the average (i.e., normal) conditions.

In view of the great importance of dynamical friction for statistical dynamics, we shall analyze the questions raised above in some detail and expose the fallacies involved in (a) and (b).

First, it is easy to show that (b) is a plain misunderstanding. For, there is nothing obvious in the requirement that for the statistical maintenance of the average conditions stars differing from the average conditions should be affected differently according to the *sense* of their departure from the normal state. Indeed, the requirement that the normal conditions are self perpetuating is to state in a different form one of two things: Either, that starting from any arbitrary initial state we approach the normal state (e.g., the Maxwellian distribution of the velocities) as $t \to \infty$; or, that once the normal state has been attained it continues to be maintained. It is now apparent that these conditions can be met only if a given star behaves at later times in a manner less and less dependent on an initial state as time goes on; or expressing the same thing somewhat differently, we should much rather expect a star to gradually lose all trace of its initial state as the time progresses. Such a gradual loss of "memory" can be achieved only by the operation of a dissipative force like dynamical friction which will gradually damp out any given initial velocity. Thus, if we assume for the sake of simplicity, that η is independent of $|\boldsymbol{u}|$, then the *average* velocity at later times will tend to zero according as

$$\bar{\boldsymbol{u}} = \boldsymbol{u}_0 e^{-\eta t}. \tag{86}$$

But this is not to imply that the mean square velocity tends to zero. Indeed, the restoration of a Maxwellian distribution of velocities from an arbitrary initial state requires that

$$\bar{\boldsymbol{u}} \to 0 \quad \text{and} \quad \overline{|\boldsymbol{u}|^2} \to \text{a constant as } t \to \infty. \tag{87}$$

To achieve the first of the two foregoing conditions we need dynamical friction. Thus, the conclusions reached in (a), if valid, are contrary to the requirement for the restoration and maintenance of the normal state. It is therefore necessary to show wherein the argumentation of (a) is in error, and this we now proceed to do.

The way to refute arguments such as (a) is, of course, to actually verify directly whether or not a star with a given initial velocity is decelerated on the average independent of the magnitude of its velocity. For this purpose, it is perhaps simplest and most instructive to examine the problem on an approximation in which the fluctuations in $\boldsymbol{F}$ are analyzed in terms of single encounters each idealized as a two-body problem. On this approximation the increments in velocity $\Delta u_{||}$ and $\Delta u_{\perp}$ which a star with a velocity $u = |\boldsymbol{u}|$ and mass m suffers as the result of an encounter with another star, in directions which are respectively parallel to and perpendicular to the direction of motion, can be specified in terms of the parameters defining the encounter. We have[9]

$$\Delta u_{||} = -\frac{2m_1}{m_1 + m}[(u - v_1 \cos\theta)\cos\psi + v_1 \sin\theta\cos\Theta\sin\psi]\cos\psi, \tag{88}$$

and

$$\left.\begin{aligned}\Delta u_{\perp} = \pm\frac{2m_1}{m_1 + m}\,[&v_1^2 + u^2 - 2uv_1\cos\theta \\ &- \{(u - v_1\cos\theta)\cos\psi + v_1\sin\theta\cos\Theta\sin\psi\}^2]^{1/2}\cos\psi\end{aligned}\right\} \tag{89}$$

where m_1 and v_1 denote the mass and velocity of a typical field star and the rest of the symbols have the same meanings as in "Stellar Dynamics," Chapter ii (see particularly, pp. 51–64).

According to equation (89), and as can indeed be expected on general symmetry grounds, $\Delta u_{\perp}$ when summed over a large number of encounters vanishes identically. But this is not the case with $\Delta u_{||}$, for the net increase in the velocity which the star suffers in the direction of its motion during a time Δt (long compared to the periods of the elementary fluctuations in $\mathbf{F}$, but short compared to the time intervals during which $|\boldsymbol{u}|$ may be expected to change appreciably) is given by

[9] Cf. Chandrasekhar, S., "Principles of Stellar Dynamics," p. 229 (equation 5.721). University of Chicago Press. 1942. This monograph will be referred to hereafter as "Stellar Dynamics."

$$\Sigma\Delta u_{||} = \Delta t \int_0^\infty dv_1 \int_0^\pi d\theta \int_0^{2\pi} d\varphi \int_0^{D_0} dD \int_0^{2\pi} d\Theta N(v_1,\theta,\varphi) V D \Delta u_{||}, \quad (90)$$

when V denotes the relative velocity between the two stars, D the impact parameter, and where, further, the various integrations are, with respect to the different parameters, defining the single encounters. We shall not go into the details here of the evaluation of the multiple integral (90),[10] but only state that on carrying out the various integrations the remarkable result emerges that *to a sufficient accuracy only stars with velocities less than the one under consideration contribute to* $\Sigma\Delta u_{||}$. This result conclusively establishes the fallacy in the assertions made in (a) and, moreover, accounts for the appearance of dynamical friction on our present analysis. Omitting then the details of the analysis we find that

$$\Sigma\Delta u_{||} = -4\pi N m_1(m_1 + m) \frac{G^2}{|\boldsymbol{u}|^2}\left(\log_e\left[\frac{D_0\overline{|\boldsymbol{u}|^2}}{G(m_1+m)}\right]\right) \times [\Phi(j|\boldsymbol{u}|) - j|\boldsymbol{u}|\Phi'(j|\boldsymbol{u}|)]\Delta t, \quad (91)$$

where N denotes the number of stars per unit volume, D_0 the average distance between the stars, Φ and Φ' the error function

$$\Phi(x) = \frac{2}{\pi^{1/2}}\int_0^x e^{-x^2}dx, \quad (92)$$

and its derivative, respectively, and j the parameter which occurs in the assumed Maxwellian distribution of the velocities

$$\frac{j^3}{\pi^{3/2}} e^{-j^2|\boldsymbol{u}|^2} d\boldsymbol{u}; \qquad j = \left(\frac{3}{2\overline{|\boldsymbol{u}|^2}}\right)^{1/2}. \quad (93)$$

Remembering that $\Sigma\Delta u_{\perp} = 0$ we can write

$$\Sigma\Delta\boldsymbol{u} = -\eta\boldsymbol{u}\Delta t \quad (94)$$

where the coefficient of dynamical friction η has now the value

$$\eta = 4\pi N m_1(m_1 + m) \frac{G^2}{|\boldsymbol{u}|^3}\left(\log_e\left[\frac{D_0\overline{|\boldsymbol{u}|^2}}{G(m_1+m)}\right]\right) \times [\Phi(j|\boldsymbol{u}|) - j|\boldsymbol{u}|\Phi'(j|\boldsymbol{u}|)]. \quad (95)$$

In order next to verify directly the existence of a relation of the form (85) we evaluate the sum

$$\Sigma\Delta u_{||}^2. \quad (96)$$

We find that [11]

[10] The details have since been published in Astrophysical Jour. 97: 255. 1943.
[11] "Stellar Dynamics." Equations (2.356) and (5.724).

$$\Sigma \Delta u_{\|}{}^2 = \frac{8}{3}\pi N m_1{}^2 \frac{G^2}{|\boldsymbol{u}|^3}\left(\log_e\left[\frac{D_0\overline{|\boldsymbol{u}|^2}}{G(m_1+m)}\right]\right) \times [\Phi(j|\boldsymbol{u}|) - j|\boldsymbol{u}|\Phi'(j|\boldsymbol{u}|)]\overline{|\boldsymbol{u}|^2}\Delta t. \quad (97)$$

Hence,

$$\frac{\Sigma \Delta u_{\|}{}^2}{\eta \Delta t} = \frac{2}{3}\frac{m_1}{m+m_1}\overline{|\boldsymbol{u}|^2}, \quad (98)$$

which is to be compared with equation (85). It is thus seen that a detailed analysis of the fluctuating field of the nearby stars in terms of individual stellar encounters idealized, as two body problems, fully confirms the conclusions reached earlier on the basis of certain general principles. In addition we now have an explicit evaluation of the coefficients q and η.

The Principles of Statistical Dynamics

In the two preceding sections we have seen how we can take into account the effect of the near neighbors on the motion of a star statistically through the two coefficients q and η. In thus representing the effect of the near neighbors in terms of the diffusion coefficient q (in the velocity space) and the frictional coefficient η we have abandoned all attempts to describe in detail the motion of any single star and have agreed instead to follow its motion through the distribution function $W(\boldsymbol{u},t)$ governing the probability of occurrence of the velocity $\boldsymbol{u}$ at time t. And as we have already shown, this probability function $W(\boldsymbol{u},t)$ satisfies the equation

$$\frac{\partial W}{\partial t} = \text{div}_u\,(q\,\text{grad}_u\,W + \eta W\boldsymbol{u}), \quad (99)$$

where it may be recalled that q and η are related according to equation (85). This differential equation satisfied by W leads to an important interpretation of the stochastic process which takes place in the velocity space. For, according to equation (99) we can visualize the motion of the representative points in the velocity space as a *process of diffusion* in which the rate of flow across an element of surface $d\sigma$ is given by

$$-\,(q\,\text{grad}_u\,W + \eta W\boldsymbol{u}).\mathbf{1}_{d\sigma}d\sigma, \quad (100)$$

where $\mathbf{1}_{d\sigma}$ is a unit vector normal to the element of surface considered. We shall find that this interpretation of the stochastic process which takes place in the velocity space as a diffusion process has important consequences for the applications of the theory (see Part III).

So far, we have restricted ourselves to what happens in the velocity space. We have, moreover, assumed that no external forces were acting. The question now arises as to how we can incorporate in a rational system of dynamics the stochastic variations in the velocity which a star suffers on account of the fluctuating force acting on it. It is evident that to build such a system of dynamics what we need is essentially a differential equation which will be appropriate for discussing the probability distribution in the six dimensional phase space in contrast to equation (99) which operates only in the velocity space. In other words, we require a proper generalization of Liouville's equation of classical dynamics to include terms corresponding to the stochastic variations in $\boldsymbol{u}$. Such a generalization can be readily found.

Quite generally we may write

$$\left.\begin{aligned} \Delta\boldsymbol{u} &= \boldsymbol{K}\Delta t + \delta\boldsymbol{u}(\Delta t) - \eta\boldsymbol{u}\Delta t \\ \Delta\boldsymbol{r} &= \boldsymbol{u}\Delta t \end{aligned}\right\} \tag{101}$$

where $\boldsymbol{K}$ denotes the external force per unit mass acting on the star and $\Delta\boldsymbol{u}$ and $\Delta\boldsymbol{r}$ the increments in the velocity and position experienced by the star in a time Δt. The interval which is chosen must again be such that it is long compared to the periods of the elementary fluctuations but short compared to the intervals during which $\boldsymbol{u}$ and $\boldsymbol{r}$ may be expected to change appreciably. Then analogous to the integral equation (81) we now have

$$W(\boldsymbol{r},\boldsymbol{u},t+\Delta t) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W(\boldsymbol{r}-\Delta\boldsymbol{r},\boldsymbol{u}-\Delta\boldsymbol{u},t) \\ \Psi(\boldsymbol{r}-\Delta\boldsymbol{r},\boldsymbol{u}-\Delta\boldsymbol{u};\Delta\boldsymbol{r},\Delta\boldsymbol{u})d(\Delta\boldsymbol{r})d(\Delta\boldsymbol{u}), \tag{102}$$

where $\Psi(\boldsymbol{r},\boldsymbol{u};\Delta\boldsymbol{r},\Delta\boldsymbol{u})$ denotes the transition probability in the phase space. According to equations (82) and (101) we now have

$$\Psi(\boldsymbol{r},\boldsymbol{u};\Delta\boldsymbol{r},\Delta\boldsymbol{u}) = \frac{1}{(4\pi q\Delta t)^{3/2}}\, e^{-|\Delta\boldsymbol{u}-\boldsymbol{K}\Delta t-q\,\mathrm{grad}_u\Delta t+\eta\boldsymbol{u}\Delta t|^2/4q\Delta t} \\ \times\, \delta(\Delta x - u_x\Delta t)\delta(\Delta y - u_y\Delta t)\delta(\Delta z - u_z\Delta t). \tag{103}$$

Expanding the various terms in equation (102) in the form of Taylor series and proceeding as in the derivation of equation (83) we obtain

$$\frac{\partial W}{\partial t} + \boldsymbol{u}\cdot\mathrm{grad}_r\, W + \boldsymbol{K}\cdot\mathrm{grad}_u\, W = \mathrm{div}_u\,(q\,\mathrm{grad}_u\, W + \eta W\boldsymbol{u}). \tag{104}$$

The foregoing equation represents the complete generalization of Liouville's theorem of classical dynamics for a single particle. On the left-hand side we have the usual Stokes' operator D/Dt operating on W while on the right-hand side we have the terms incorporating the fluctuations caused by the neighboring stars. It should, however, be noticed that the Liouville equation now operates in the six dimensional phase space. This is because we have taken into account the effect of the neighboring stars statistically through the terms involving q and η. Further, it should be noticed, too, that the relation (85) between q and η ensures that the Maxwell-Boltzmann distribution in the phase space is invariant to the stochastic process considered (see the section below).

Analytical Dynamics versus Statistical Dynamics

In the preceding sections we have outlined the general principles of a statistical theory of stellar dynamics. In order that we may emphasize and further amplify the basic ideas which are involved, we shall contrast the outlook of statistical dynamics with the point of view familiar in analytical dynamics.

ANALYTICAL DYNAMICS	STATISTICAL DYNAMICS
1. In analytical dynamics we follow in *detail* the motion of each of the degrees of freedom of the dynamical system.	1. In statistical dynamics we follow instead the motion of each of the particles *statistically* when under the fluctuating influence of a large number of other particles belonging also to the system.
2. The notion of acceleration is fundamental to analytical dymanics.	2. For the success of the methods of statistical dynamics it is essential that time intervals Δt exist with the property that they are long compared to the periods of the elementary fluctuations but which are at the same time short compared to the times necessary for $\boldsymbol{u}$ to change appreciably. Moreover, during such an interval Δt the mean square increment in $\boldsymbol{u}$ is given by $$\overline{\lvert\Delta \boldsymbol{u}\rvert^2} = 2q\Delta t.$$ Accordingly $$\frac{\sqrt{\overline{\lvert\Delta \boldsymbol{u}\rvert^2}}}{\Delta t} \to \infty \quad \text{as} \quad \Delta t \to 0.$$ In other words, we cannot properly define acceleration within the framework of statistical dynamics.

3. The equations of motion governing a conservative dynamical system can be thrown into the canonical forms

$$\dot{p}_r = -\frac{\partial H}{\partial q_r}; \qquad \dot{q}_r = \frac{\partial H}{\partial p_r} \qquad (r = 1, \cdots, N),$$

where H is the Hamiltonian function. These equations can be interpreted by the statement that the development of a conservative dynamical system represents "the gradual unfolding of a contact transformation" (Whittaker).

3. In statistical dynamics the fundamental assumption is generally made that the stochastic process which takes place can be described as a *Markoff chain*. More explicitly, we suppose that the probability distribution

$$W(\boldsymbol{r}, \boldsymbol{u}, t + \Delta t)$$

at the time $t + \Delta t$ can be derived from the distribution $(W(\boldsymbol{r}, \boldsymbol{u}, t)$ at the slightly earlier time t through an integral equation of the form

$$W(\boldsymbol{r}, \boldsymbol{u}, t + \Delta t) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} W(\boldsymbol{r} - \Delta\boldsymbol{r}, \boldsymbol{u} - \Delta\boldsymbol{u}, t) \times \Psi(\boldsymbol{r} - \Delta\boldsymbol{r}, \boldsymbol{u} - \Delta\boldsymbol{u}; \Delta\boldsymbol{r}, \Delta\boldsymbol{u}) \times d(\Delta\boldsymbol{r})d(\Delta\boldsymbol{u})$$

where $\Psi(\boldsymbol{r}, \boldsymbol{u}; \Delta\boldsymbol{r}, \Delta\boldsymbol{u})$ denotes the transition probability. (The foregoing integral relation connecting $W(\boldsymbol{r}, \boldsymbol{u}, t + \Delta t)$ and $W(\boldsymbol{r}, \boldsymbol{u}, t)$ can be regarded as defining a Markoff chain.) Analogous to the interpretation of the canonical equations in analytical dynamics we may describe a Markoff process as "the gradual unfolding of a transition probability."

4. In the $2N$ dimensional phase space the hydrodynamical flow which can be set up by following each point in this space according to the canonical equations is described by Liouville's theorem. According to this theorem, an initially assigned density

$$W(q_1, \cdots, q_N, p_1, \cdots, p_N)$$

in the phase space varies according to the equation

$$\frac{\partial W}{\partial t} + \sum_{r=1}^{N}\left(\frac{\partial H}{\partial p_r}\frac{\partial W}{\partial q_r} - \frac{\partial H}{\partial q_r}\frac{\partial W}{\partial p_r}\right) = 0.$$

4. The probability distribution in the 6-dimensional phase space (i.e., the phase space of a *single* particle) is governed by the equation

$$\frac{\partial W}{\partial t} + \boldsymbol{u}\cdot\operatorname{grad}_r W + \boldsymbol{K}\cdot\operatorname{grad}_u W = \operatorname{div}_u(q \operatorname{grad}_u W + \eta W\boldsymbol{u}),$$

where $\boldsymbol{K}$ denotes the external force per unit mass acting on the particle, and q and η the diffusion and the frictional coefficients describing the stochastic process which takes place in the velocity space (see 5 below).

5. The order of the system of equations governing the development of a dynamical system equals twice the number of degrees of freedom of the system.

5. In statistical dynamics *almost all* the coordinates are ignored. This ignoration of the coordinates of all the neighboring particles becomes possible only because we are able to represent their influence on the statistical motion of any single particle through the two coefficients q and η. More particularly, the stochastic variations which take place in the velocity space can be described as a general process of diffusion in which the rate of flow across an element of surface $d\sigma$ is given by

$$-\left(q\,\mathrm{grad}_u W + \eta W \boldsymbol{u} - \boldsymbol{K} W\right) \cdot \mathbf{1}_{d\sigma}\, d\sigma,$$

where $\mathbf{1}_{d\sigma}$ is a unit vector normal to the element of surface considered.

6. The equations of motion of a conservative dynamical system possess the energy integral

$$H = \text{Constant}.$$

6. The generalized Liouville equation in the 6-dimensional phase space governing the probability distribution $W(\boldsymbol{r}, \boldsymbol{u}, t)$ is satisfied identically by the Maxwell-Boltzmann distribution

$$W = \text{Constant } e^{-3(|u|^2 + 2\mathfrak{V})/2\overline{|u|^2}}$$

where

$$\boldsymbol{K} = -\mathrm{grad}\, \mathfrak{V}.$$

It is this circumstance which enables the restoration of a Maxwell-Boltzmann distribution from any arbitrary initial state.

7. When dealing with conservative dynamical systems, dissipative forces are foreign to the notions of analytical dynamics. However, dissipative forces may appear in the discussion of dynamical systems in their *non-natural* forms, i.e., when the system is considered in a reduced number of coordinates after the process of the ignoration of coordinates (Whittaker, *Analytical Dynamics*, p. 57).

7. The occurrence of dissipative forces like dynamical friction in the stochastic variations in the velocity experienced by a particle is essential for the success of statistical dynamics. For, it is precisely on account of the occurrence of the term involving dynamical friction that the restoration and maintenance, for example, of a Maxwellian distribution of the velocities among the particles is made possible. Alternatively, we may express the same thing by saying that the operation of a dissipative force like dynamical friction is exactly what is needed to conserve the energy of the assembly as a whole. This may sound paradoxical at first sight, but it is intimately connected with the fact that in statistical dynamics we have essentially performed an ignoration of the coordinates of the neighboring particles.

III. THE RATE OF ESCAPE OF STARS FROM CLUSTERS AND THE EVIDENCE FOR THE OPERATION OF DYNAMICAL FRICTION

The General Theory of the Rate of Escape of Stars from Clusters

One of the most important factors in the evolution of the galactic and the globular clusters is their gradual impoverishment due to the escape of stars.[12] Essentially, the mechanism underlying this escape of stars is as follows:

On account of the fluctuating gravitational field acting on a star we should expect that there exists a finite probability for a star to acquire a velocity sufficient to escape from the cluster during any specified length of time t. And if a star should acquire the necessary velocity it would naturally escape from the cluster. We shall now show how, on the basis of the statistical theory developed in Part II, we can evaluate this factor in the evolution of clusters quite rigorously.

To be specific, we shall suppose that in order that a star may escape from a cluster it is only necessary that it acquire a velocity greater than or equal to a certain critical velocity which we shall denote by v_∞. On this assumption the probability that a star will have acquired the necessary velocity for escape during a certain time t can be evaluated quite simply from the probability $p(v_0,t)\Delta t$ that a star having initially a velocity $|\boldsymbol{u}| = v_0$ at time $t = 0$ will acquire for the *first time* the velocity $|\boldsymbol{u}| = v_\infty$ during t and $t + dt$. For, on integrating $p(v_0,t)$ over t from 0 to t we shall obtain the total probability $Q(v_0,t)$ that the star will have acquired the velocity v_∞ during the entire interval from 0 to t. And finally averaging $Q(v_0,t)$ over the relevant range of the initial velocities, we shall obtain the *expectation* $Q(t)$ that a star will have acquired the velocity v_∞ during the time t.

The advantage of formulating the problem in the manner described above is that the function $p(v_0,t)$ can be determined in terms of a solution of the equation (see equation [99])

$$\frac{\partial W}{\partial t} = \mathrm{div}_u\,(q\,\mathrm{grad}_u\,W + \eta W\boldsymbol{u}) \tag{105}$$

which satisfies certain appropriate boundary conditions. For, remembering that the stochastic process described by the foregoing equation has a simple interpretation in terms of general type of diffusion process, it is evident that $p(v_0,t)$ will be given by

[12] This fact was first clearly recognized by Ambarzumian and Spitzer. For an account of these earlier discussions see "Stellar Dynamics," §§ 5.3 and 5.4, pp. 250–213.

$$p(v_0,t) = -\left(4\pi q|\boldsymbol{u}|^2 \frac{\partial W(|\boldsymbol{u}|,t)}{\partial|\boldsymbol{u}|}\right)_{|\boldsymbol{u}|=v_\infty}, \tag{106}$$

where $W(|\boldsymbol{u}|,t)$ denotes a spherically symmetric solution of equation (105) which satisfied the boundary conditions

$$W(|\boldsymbol{u}|,t) = 0 \text{ for } |\boldsymbol{u}| = v_\infty \text{ for all } t > 0, \tag{107}$$

and

$$W(|\boldsymbol{u}|,t) \to \frac{1}{4\pi v_0^2}\,\delta(|\boldsymbol{u}| - v_0) \quad \text{as} \quad t \to 0, \tag{108}$$

where δ stands for the δ-function of Dirac. We shall now show how we can obtain such a solution.

For the case under discussion we have (see equation [95])

$$\eta = 8\pi N m^2 G^2\left(\log_e\left[\frac{D_0\overline{|\boldsymbol{u}|^2}}{2Gm}\right]\right)\frac{1}{|\boldsymbol{u}|^3}[\Phi(j|\boldsymbol{u}|) - j|\boldsymbol{u}|\Phi'(j|\boldsymbol{u}|)]. \tag{109}$$

This formula for η can be written more conveniently as

$$\eta = \eta_0\nu(j|\boldsymbol{u}|) \tag{110}$$

where

$$\eta_0 = 8\pi N m^2 G^2\left(\log_e\left[\frac{D_0\overline{|\boldsymbol{u}|^2}}{2Gm}\right]\right)\left(\frac{3}{2\overline{|\boldsymbol{u}|^2}}\right)^{3/2}\frac{4}{3\pi^{1/2}} \tag{111}$$

and

$$\nu(\rho) = \frac{3\pi^{1/2}}{4}\,\rho^{-3}[\Phi(\rho) - \rho\Phi'(\rho)]; \tag{112}$$

with $\nu(\rho)$ defined in this manner

$$\left.\begin{aligned} &\nu(\rho) \to 1 \quad \text{as} \quad \rho \to 0, \\ &\nu(\rho) \to \frac{3\pi^{1/2}}{4}\,\rho^{-3} \quad \text{as} \quad \rho \to \infty. \end{aligned}\right\} \tag{113}$$

Again, since q and η are generally related according to equation (85), we have

$$q = \frac{1}{3}\,\overline{|\boldsymbol{u}|^2}\eta_0\nu(j|\boldsymbol{u}|). \tag{114}$$

Returning to equation (105) we introduce the following change of the independent variables:

$$\tau = \eta_0 t; \qquad \boldsymbol{u} = \left(\frac{2}{3}\,\overline{|u|^2}\right)^{1/2}\boldsymbol{\varrho}. \tag{115}$$

Equation (105) now takes the dimensionless form

$$\frac{\partial W}{\partial \tau} = \operatorname{div}_{\varrho}\left[\frac{1}{2}\,\nu(|\boldsymbol{\varrho}|)\operatorname{grad}_{\varrho} W + \nu(|\boldsymbol{\varrho}|)W\boldsymbol{\varrho}\right]. \tag{116}$$

For a spherically symmetric solution $W(|\boldsymbol{\varrho}|,\tau)$ equation (116) reduces to

$$\rho\frac{\partial w}{\partial \tau} = \frac{\partial}{\partial \rho}\left[\nu(\rho)\left\{\frac{1}{2}\rho\frac{\partial w}{\partial \rho} + \left(\rho^2 - \frac{1}{2}\right)w\right\}\right] \tag{117}$$

where we have written

$$\rho = |\boldsymbol{\varrho}|; \qquad w = W\rho. \tag{118}$$

According to equations (107) and (108) we require a solution of equation (117) which satisfies the boundary conditions

$$w(\rho,\tau) = 0 \text{ for both } \rho = 0 \text{ and } \rho = \rho_\infty \text{ for all } \tau > 0, \tag{119}$$

and

$$w(\rho,\tau) \to \frac{1}{4\pi\rho_0}\,\delta(\rho - \rho_0) \quad \text{as} \quad \tau \to 0. \tag{120}$$

Now equation (117) is separable in the variables ρ and τ. Accordingly we write

$$w = e^{-\lambda\tau}\phi(\rho) \tag{121}$$

where λ is for the present an unspecified constant; we then obtain for ϕ the differential equation

$$\frac{d}{d\rho}\left[\nu(\rho)\left\{\frac{1}{2}\rho\frac{d\phi}{d\rho} + \left(\rho^2 - \frac{1}{2}\right)\phi\right\}\right] + \lambda\rho\phi = 0 \tag{122}$$

If we now let

$$\phi = e^{-\rho^2/2}\psi \tag{123}$$

we obtain

$$\frac{d^2\psi}{d\rho^2} + \frac{d\log\nu}{d\rho}\frac{d\psi}{d\rho} + \left[2\frac{\lambda}{\nu(\rho)} + 3 - \rho^2 - \frac{d\log\nu}{d\rho}\left(\frac{1}{\rho} - \rho\right)\right]\psi = 0. \tag{124}$$

It is now seen that in order that a solution of the foregoing equation may vanish at $\rho = 0$ and at $\rho = \rho_\infty$, it is necessary that λ take one of an infinite enumerable set of discrete values

$$\lambda_1, \lambda_2, \cdots, \lambda_n, \cdots \tag{125}$$

which we may properly call the "characteristic values" of the problem. Further if

$$\psi_1, \psi_2, \cdots, \psi_n, \cdots \tag{126}$$

denote the solutions of equation (124) which satisfy the boundary conditions (119) at $\rho = 0$ and $\rho = \rho_\infty$ and belong, respectively, to the values $\lambda_1, \lambda_2, \cdots, \lambda_n, \cdots$ then it can be readily verified that these solutions form a complete set of orthogonal functions. Without loss of generality we can therefore suppose that these functions are all properly normalized. Consequently, in terms of the fundamental solutions

$$w_n = e^{-\lambda_n\tau}e^{-\rho^2/2}\psi_n(\rho) \tag{127}$$

which satisfy the boundary conditions (119) we can construct solutions which will satisfy any further arbitrary boundary condition for $\tau = 0$. Thus, remembering that a δ-function can always be constructed in terms of any complete set of orthogonal functions according to

$$\delta(\rho - \rho_0) = \sum_{n=1}^{\infty} \psi_n(\rho)\psi_n(\rho_0), \tag{128}$$

it is evident that

$$w = \frac{1}{4\pi\rho_0} e^{-(\rho^2-\rho_0^2)/2} \sum_{n=1}^{\infty} e^{-\lambda_n\tau}\, \psi_n(\rho)\psi_n(\rho_0) \tag{129}$$

satisfies all the boundary conditions of our problem. Corresponding to the solution (129) for w we have

$$W = \frac{1}{4\pi\rho\rho_0} e^{-(\rho^2-\rho_0^2)/2} \sum_{n=1}^{\infty} e^{-\lambda_n\tau}\, \psi_n(\rho)\psi_n(\rho_0). \tag{130}$$

Using the foregoing solution for W we can write down the probability function $p(\rho_0,\tau)$. For, since

$$p(\rho_0,\tau) = -2\pi\rho_\infty^2\nu(\rho_\infty)\left(\frac{\partial W}{\partial\rho}\right)_{\rho=\rho_\infty}, \tag{131}$$

we have

$$p(\rho_0,\tau) = \frac{\rho_\infty}{2\rho_0}\nu(\rho_\infty)e^{-(\rho_\infty^2-\rho_0^2)/2}\sum_{n=1}^{\infty} e^{-\lambda_n\tau}\left(-\frac{\partial\psi_n}{\partial\rho}\right)_{\rho=\rho_\infty}\psi_n(\rho_0). \tag{132}$$

To obtain the probability $Q(\rho_0,\tau)$ that a star having an initial velocity corresponding to ρ_0 will have acquired a velocity corresponding to ρ_∞ during the time τ we have simply to integrate equation (132) from 0 to τ. Thus we find that

$$\left.\begin{aligned} Q(\rho_0,\tau) = \frac{\rho_\infty}{2\rho_0}\nu(\rho_\infty)e^{-(\rho_\infty^2-\rho_0^2)/2}\sum_{n=1}^{\infty}\frac{1}{\lambda_n}(1-e^{-\lambda_n\tau}) \\ \left(-\frac{d\psi_n}{d\rho}\right)_{\rho=\rho_\infty}\psi_n(\rho_0). \end{aligned}\right\} \tag{133}$$

Finally to obtain, $Q(\tau)$, that an average star will have acquired the necessary velocity for escape during a time τ, we must average the foregoing expression over all ρ_0. The final result can therefore be expressed in the form

$$Q(\tau) = \sum_{n=1}^{\infty} Q_n(\tau), \tag{134}$$

where

$$Q_n(\tau) = A_n(1 - e^{-\lambda_n\tau}), \tag{135}$$

and

$$A_n = \frac{1}{2\lambda_n}\rho_\infty\nu(\rho_\infty)e^{-\rho_\infty^2/2}\left(-\frac{d\psi_n}{d\rho}\right)_{\rho=\rho_\infty}\overline{\left[\frac{e^{\rho_0^2/2}}{\rho_0}\psi_n(\rho_0)\right]}. \tag{136}$$

The Evidence for the Operation of Dynamical Friction

We shall now illustrate with some numerical results the theory developed in the preceding section.

Now, since in a star cluster the root mean square velocity of escape is twice the mean square velocity of the stars in the system,[13] it is clear that the values of ρ_∞ which come under discussion are in the general neighborhood of

$$\rho_\infty = \sqrt{6} \sim 2.45. \tag{137}$$

For values of ρ_∞ in this neighborhood it was found that $Q(\tau)$ can be represented with ample accuracy by the first term on the right-hand side of equation (134). Accordingly it was sufficient to specify the lowest characteristic value λ_1 (for a given ρ_∞) and the normalized characteristic function ψ_1 belonging to it. In this manner it was found that

$$Q(\tau) = 1 - e^{-0.0075\tau} \qquad (\rho_\infty = 2.4518). \tag{138}$$

(The foregoing equation provides sufficient accuracy for $\tau > 5$).

Since $Q(\tau)$ gives the expectation that an average star will have escaped during a time τ (in units of η_0^{-1}) we can properly regard $(0.0075\eta_0)^{-1}$ as a measure of the half-life of the cluster. Thus

$$\text{Half-life of a cluster} \simeq 133\ \eta_0^{-1}. \tag{139}$$

For the Pleiades η_0^{-1} is of the order of 2×10^7 years, so that its half-life is of the order of 3×10^9 years. In judging this value it should be remembered that (as may be readily verified) when dynamical friction is ignored, a half-life for the Pleiades of the order of only 5×10^7 years is predicted. There can thus be hardly any doubt that dynamical friction provides the principal mechanism for the continued existence of the galactic clusters like the Pleiades for times of the order of 3×10^9 years. But, even with dynamical friction properly allowed for (as we have done), it will be hard to account for such clusters half-lives of the order 10^{10} years. This, in turn, provides another strong argument in favor of the now currently adopted "short time scale" of the order of 3×10^9 years.

In concluding this essay, we might draw attention to the far reaching analogy which exists between these newer methods in stellar dynamics and methods long familiar in the theory of Brownian movement. However, while parts of the theory of Brownian motion are heuristic and appeal to intuitive considerations, it appears that in stellar dynamics the entire problem can be analyzed explicitly in all its phases.

[13] Cf. "Stellar Dynamics," pp. 206–207.

THE THEORY OF THE FLUCTUATIONS IN BRIGHTNESS OF THE MILKY WAY. I

S. CHANDRASEKHAR AND G. MÜNCH
Yerkes Observatory
Received July 11, 1950

ABSTRACT

In this paper the following statistical problem is considered. Let stars and interstellar clouds occur with a uniform distribution. Let the system extend to a linear distance L in the direction of a line of sight. Let a cloud reduce the intensity of the light of the stars immediately behind it by a factor q. Let the occurrence of clouds with a transparency factor q be governed by a frequency function $\psi(q)$. Given all this, it is required to find the probability distribution, $g(I, L)$, of the observed brightness, I. From a consideration of this problem it is shown that the following integral equation governs the distribution of brightness:

$$g(u, \xi) + \frac{\partial g}{\partial u} + \frac{\partial g}{\partial \xi} = \int_0^1 g\left(\frac{u}{q}, \xi\right) \psi(q) \frac{dq}{q},$$

where u is the observed brightness measured in suitable units and ξ is the average number of clouds in the direction of the line of sight. It is further shown that the foregoing integral equation enables us to obtain explicit formulae for all the moments of g as functions of ξ and the moments of $\psi(q)$. As an illustration of the use of these general formulae for the moments, an example investigated by Markarian has been reconsidered in an attempt to derive the mean and mean-square deviation of the optical thicknesses of the interstellar clouds.

1. *Introduction.*—The fact, now generally recognized, that interstellar matter occurs in the form of clouds and that the average number of such clouds intersected by a line of sight is of the order of 5 per kiloparsec requires a reorientation of the problems and objectives of stellar statistics. That such a reorientation is needed is brought out most clearly by the definiteness and the precision of the conclusions reached by Ambarzumian and his collaborators in three relatively brief investigations,[1] in which the cloud structure of interstellar matter is explicitly introduced as an essential element of the problem. Thus, in the investigation by Ambarzumian and Gordeladse, in which the observed association of emission and reflection nebulae is quantitatively accounted for on the hypothesis of a random distribution of stars and interstellar clouds by considering the volumes of space illuminated by stars of various spectral types, estimates are obtained for the average number of clouds per unit volume ($\sim 1.2 \times 10^{-4}$ per cubic parsec) and the average number of clouds intersected by a straight line (5–7 per kiloparsec). Similarly, from a simple analysis of the statistics of extragalactic nebulae, Ambarzumian deduced that the photographic absorption per cloud is of the order of 0.2 mag.; this, combined with the earlier estimate of the number of clouds in a line of sight, leads to a photographic extinction coefficient of 1.0–1.5 mag. per kiloparsec, which is in general agreement with other determinations of this quantity. The far-reaching nature of these conclusions—they were revolutionary at the time that they were drawn—should convince one that stellar statistics will gain enormously by making the distribution and the properties of interstellar clouds more immediate objectives of the investigation than has been customary so far. For example, the known fluctuations in the brightness of the Milky Way can be interpreted most readily in terms of the fluctuations in the numbers of the absorbing clouds in the line of sight; for, while other factors doubtless contribute to the observed fluctuations, these must be secondary to the effect of the fluctuations in the numbers of

[1] V. A. Ambarzumian and S. G. Gordeladse, *Bull. Abastumani Obs.*, No. 2, p. 37, 1938; V. A. Ambarzumian, *Bull. Abastumani Obs.*, No. 4, p. 17, 1940; and B. E. Markarian, *Contr. Burakan Obs. Acad. Sci. Armenian S.S.R.*, No. 1, 1946.

clouds, since so few of these are generally involved. Indeed, in a short note published in 1944, Ambarzumian[2] formulated the following problem which he considered basic for such an analysis:

Let stars and absorbing clouds occur with a uniform distribution in a plane of infinite extent. Further, let a cloud reduce the intensity of the light of the stars immediately behind it by a factor q. Let the occurrence of clouds with a "transparency" q be governed by a frequency function $\psi(q)$. What is, then, the probability distribution of the observed brightness?

Ambarzumian derived an integral equation for the required probability distribution and showed how its first and second moments can be expressed quite simply in terms of $\bar{q}$ and $\overline{q^2}$. However, when Markarian[1] came to applying this theory to observations, he found that Ambarzumian's assumption of the infinite extent of the system in the direction of the line of sight was too restrictive and that the problem must be considered for the case in which the average number of clouds in the line of sight is finite. The need for this generalization is apparent when we remember that the average number of clouds in the direction of galactic latitude β is n cosec β, where n is the corresponding number in the direction of the galactic pole; thus in our own galaxy $n \sim 3$ and n cosec $\beta \sim 10$ for $\beta = 20°$. Moreover, this dependence of the average number of clouds on the galactic latitude will provide a valuable check on the analysis.

Markarian did not derive the integral equation governing the distribution of brightness for the more general problem; but he did obtain explicit expressions for the first and the second moments for the case in which all the clouds are equally transparent. In this paper we shall derive the general integral equation governing the distribution of brightness and show how all its moments can be found. And we shall illustrate the use of these general relations for the moments by an example.

2. *The basic integral equation.*—Let I denote the observed brightness and L the distance of the observer to the limits of the system in the direction of the line of sight. Then

$$I = \int_0^L \prod_{i=1}^{n(s)} q_i \eta \, d s \,, \tag{1}$$

where η is the emission per unit volume by the stars assumed to be uniformly distributed, $n(s)$ is the number of clouds in the distance interval $(0, s)$ in the line of sight and is a chance variable, and q_i $[i = 1, 2, \ldots, n(s)]$ is the factor by which the ith cloud cuts down the intensity of the light from the stars immediately behind it. As we have already stated in § 1, we shall assume that the q's occur with a known frequency, $\psi(q)$.

If ν is the average number of clouds per unit distance, then $n(s)$ will be governed by the Poisson distribution,

$$e^{-\nu s} \frac{(\nu s)^n}{n!} \qquad (n = 0, 1, \ldots) \,, \tag{2}$$

having the variance νs.

The problem is to determine the probability distribution of I. It is convenient to reformulate this problem in dimensionless variables. For this purpose we shall let

$$u = I \frac{\nu}{\eta} \qquad \text{and} \qquad r = \nu s \,. \tag{3}$$

Then

$$u = \int_0^\xi \prod_{i=1}^{n(r)} q_i d r \,, \tag{4}$$

where

$$\xi = \nu L \tag{5}$$

[2] *C.R. (Doklady) Acad. Sci. URSS*, **14**, 223, 1944.

is the average number of clouds to be expected in the distance L. Also, according to formula (2), the occurrence of a particular number of clouds, n, in the interval $(0, r)$ will be governed by the Poisson distribution,

$$e^{-r}\frac{r^n}{n!} \qquad (n = 0, 1, \ldots) . \tag{6}$$

Let $g(u, \xi)$ denote the frequency distribution of u for a given ξ. Since the q_i's are all less than, or equal to, 1, it follows from the definition of u as the integral (4) that u *can never exceed* ξ. Hence

$$g(u, \xi) = 0 \qquad \text{for} \qquad u > \xi . \tag{7}$$

In addition to $g(u, \xi)$, it is convenient to define the probability that u exceeds a specified value. Let

$$f(u, \xi) = \int_u^\xi g(u, \xi)\, du \tag{8}$$

denote this probability. An integral equation governing $f(u, \xi)$ can be derived in the following manner:

By definition

$$f(u, \xi) = \text{Probability that} \int_0^\xi \prod_{i=1}^{n(r)} q_i\, dr \geqslant u . \tag{9}$$

Equivalently, we may also write

$$f(u, \xi) = \text{Probability that}\left\{\int_0^a \prod_{i=1}^{n(r)} q_i\, dr + \prod_{j=1}^{n(a)} q_j \int_a^\xi \prod_{i=1}^{n(r)-n(a)} q_i\, dr\right\} \geqslant u , \tag{10}$$

where a is an arbitrary positive constant $\leqslant \xi$. Replacing $r - a$ by r in the second integral on the right-hand side, we have

$$f(u, \xi) = \text{Probability that}\left\{\int_0^a \prod_{i=1}^{n(r)} q_i\, dr + \prod_{j=1}^{n(a)} q_j \int_0^{\xi-a} \prod_{i=1}^{n(r)} q_i\, dr\right\} \geqslant u . \tag{11}$$

Now let $a \ll 1$. Then up to $O(a^2)$, we have only two possibilities: either there is no cloud in the interval $(0, a)$, or there is just exactly one cloud. The probabilities of these two occurrences—again, up to $O(a^2)$—are $1 - a$ and a, respectively. Hence

$$\begin{aligned} \prod_{i=1}^{n(a)} q_i &= 1 \text{ with probability } 1 - a , \\ &\geqslant q \text{ and } \leqslant q + dq \text{ with probability } a\psi(q)\, dq . \end{aligned} \tag{12}$$

We may, therefore, rewrite equation (11) in the form

$$\begin{aligned} f(u, \xi) = {} & a\int_0^1 dq\,\psi(q) \times \text{probability that}\left\{\theta a + q\int_0^{\xi-a} \prod_{i=1}^{n(r)} q_i\, dr\right\} \geqslant u \\ & + (1-a) \times \text{probability that}\left\{a + \int_0^{\xi-a} \prod_{i=1}^{n(r)} q_i\, dr\right\} \geqslant u + O(a^2) , \end{aligned} \tag{13}$$

where $\theta \leqslant 1$ is some positive constant. Since we are neglecting all quantities of $O(a^2)$ and higher, it is clearly sufficient to evaluate the integral in equation (13) which occurs with the factor a to zero order in a. Thus

$$f(u,\xi)=a\int_0^1 dq\psi(q)\times\text{probability that}\left\{\int_0^{\xi}\prod_{i=1}^{n(r)} q_i\,dr\right\}\geqslant\frac{u}{q}$$
$$+(1-a)\times\text{probability that}\left\{\int_0^{\xi-a}\prod_{i=1}^{n(r)} q_i\,dr\right\}\geqslant u-a+O(a^2)\,; \tag{14}$$

or, remembering the definition of $f(u, \xi)$, we have

$$f(u,\xi)=a\int_0^1 dq\psi(q)\,f\left(\frac{u}{q},\,\xi\right)+(1-a)\,f(u-a,\,\xi-a)+O(a^2)\,. \tag{15}$$

Hence

$$f(u,\xi)=a\int_0^1 dq\psi(q)\,f\left(\frac{u}{q},\,\xi\right)+f(u,\xi)$$
$$-af(u,\xi)-a\frac{\partial f}{\partial u}-a\frac{\partial f}{\partial\xi}+O(a^2)\,. \tag{16}$$

The function $f(u, \xi)$ must therefore satisfy the integral equation,

$$f+\frac{\partial f}{\partial u}+\frac{\partial f}{\partial\xi}=\int_0^1\psi(q)\,f\left(\frac{u}{q},\xi\right)dq\,. \tag{17}$$

Now, differentiating this equation with respect to u, we obtain the integral equation governing $g(u, \xi)$,

$$g(u,\xi)+\frac{\partial g}{\partial u}+\frac{\partial g}{\partial\xi}=\int_0^1\psi(q)\,g\left(\frac{u}{q},\,\xi\right)\frac{dq}{q}\,. \tag{18}$$

We have already pointed out that u cannot take values *exceeding* ξ. But it can take the value ξ itself with exactly the probability $e^{-\xi}$, since this is the probability that no cloud will occur in the interval $(0, \xi)$. The distribution of u has therefore a delta function, $\delta(u-\xi)$, at $u=\xi$ with an "amplitude" $e^{-\xi}$. Therefore, writing

$$g(u,\xi)=\phi(u,\xi)+e^{-\xi}\delta(u-\xi)\,,$$

we find that the differential equation for ϕ is

$$\phi(u,\xi)+\frac{\partial\phi}{\partial u}+\frac{\partial\phi}{\partial\xi}=\psi\left(\frac{u}{\xi}\right)\frac{e^{-\xi}}{\xi}+\int_{u/\xi}^1\psi(q)\,\phi\left(\frac{u}{q},\,\xi\right)\frac{dq}{q}\,.$$

In deriving this equation the assumption has been made that since

$$\left(1+\frac{\partial}{\partial u}+\frac{\partial}{\partial\xi}\right)e^{-\xi}\times\text{a function of }(u-\xi)\equiv 0\,,$$

the same is also true of $e^{-\xi}\delta(u-\xi)$.

For the case in which the system extends to infinity in the direction of the line of sight, equation (18) reduces to the one given by Ambarzumian,[2] namely,

$$g(u)+\frac{dg}{du}=\int_0^1\psi(q)\,g\left(\frac{u}{q}\right)\frac{dq}{q}\,. \tag{19}$$

Equation (18) is the basic equation of the problem.

3. *The moments of the distribution function* $g(u, \xi)$.—We shall now show how the integral equation (18) enables us to determine all the moments,

$$\mu_n = \int_0^\xi g(u, \xi)\, u^n du \qquad (n = 0, 1, \ldots), \tag{20}$$

of the distribution function $g(u, \xi)$. But first we may note that, by definition,

$$\mu_0 = \int_0^\xi g(u, \xi)\, du \equiv 1; \tag{21}$$

all the other moments will, however, be functions of ξ.

Now, multiplying equation (18) by u^n and integrating over the range of u, we obtain

$$\mu_n + \int_0^\xi \frac{\partial g}{\partial u} u^n du + \frac{d\mu_n}{d\xi} = \int_0^1 dq\, q^n \psi(q) \int_0^\xi \frac{du}{q}\left(\frac{u}{q}\right)^n g\left(\frac{u}{q}, \xi\right). \tag{22}$$

An integration by parts reduces the integral on the left-hand side to $-n\mu_{n-1}$ if use is made of equation (7). Also, the integral on the right-hand side can be reduced in the following manner:

$$\begin{aligned}\int_0^1 dq\, q^n \psi(q) \int_0^\xi \frac{du}{q}\left(\frac{u}{q}\right)^n g\left(\frac{u}{q}, \xi\right)& \\ &= \int_0^1 dq\, q^n \psi(q) \int_0^{\xi/q} dx\, x^n g(x, \xi) \\ &= \int_0^1 dq\, q^n \psi(q) \int_0^\xi dx\, x^n g(x, \xi) = \mu_n \int_0^1 dq\, q^n \psi(q).\end{aligned} \tag{23}$$

In the foregoing reductions use has again been made of the fact that $g(u, \xi) = 0$ for $u > \xi$. Thus equation (22) becomes

$$\frac{d\mu_n}{d\xi} + (1 - q_n)\mu_n = n\mu_{n-1} \qquad (n = 0, 1, \ldots), \tag{24}$$

where, for the sake of brevity, we have written

$$q_n = \overline{q^n} = \int_0^1 dq\, q^n \psi(q). \tag{25}$$

It may be noticed here that, by writing

$$\mu_n = \int_0^\xi \phi(u, \xi)\, u^n du + \xi^n e^{-\xi} = u_n + \xi^n e^{-\xi}$$

in equation (24), we find that the equation satisfied by u_n is

$$\frac{du_n}{d\xi} + (1 - q_n) u_n = q_n \xi^n e^{-\xi} + n u_{n-1}.$$

But this same differential equation also follows directly from the equation satisfied by ϕ.

It is evident that all the moments μ_n must vanish at $\xi = 0$. On the other hand, from equation (22) for $n = 2$, namely,

$$\frac{d\mu_2}{d\xi} + (1 - q_2)\mu_2 = 2\mu_1, \tag{26}$$

we conclude that $d\mu_2/d\xi$ also vanishes at $\xi = 0$. And by induction it follows quite generally from equation (24) that μ_n *and its first* $(n-1)$ *derivatives must vanish at* $\xi = 0$. Also, the μ_n's must be bounded for $\xi \to \infty$. As we shall see presently, these boundary conditions suffice to solve the system of equations (24) uniquely.

By successive applications of equation (24) for $n = 1, 2$, etc., we obtain

$$\left[\prod_{j=1}^{n}\left(\frac{d}{d\xi}+a_j\right)\right]\mu_n = n!\,, \tag{27}$$

where

$$a_j = 1 - q_j\,. \tag{28}$$

The solution of equation (27) which satisfies the boundary condition at $\xi = \infty$ is

$$\mu_n = \sum_{k=1}^{n} A_k e^{-a_k\xi} + \frac{n!}{\prod_{j=1}^{n} a_j}\,, \tag{29}$$

where the A_k's$(k = 1, \ldots, n)$ are n constants of integration.

The boundary conditions,

$$\mu_n = 0 \quad \text{and} \quad \frac{d^j\mu_n}{d\xi^j} = 0 \qquad (j = 1, \ldots, n-1)\,, \tag{30}$$

now require that

$$\sum_{k=1}^{n} A_k = -\frac{n!}{\prod_{j=1}^{n} a_j}$$

and

$$\sum_{k=1}^{n} A_k a_k^j = 0 \qquad (j = 1, \ldots, n-1)\,. \tag{31}$$

The matrix of the system of equations (30) is seen to be the Vandermondie matrix;[3] its reciprocal is the matrix[4]

$$\left\| \frac{S_{n-l-1,\,r}}{\prod_{j\neq r}^{(1,\,n)} (a_r - a_j)} \right\|, \tag{32}$$

in which the $S_{n-l-1,\,r}$'s $(l = 0, 1, \ldots, n-1;\ r = 1, \ldots, n)$ are the independent symmetric functions in the $(n-1)$ variables $(a_1, \ldots, a_{r-1}; a_{r+1}, \ldots, a_n)$,

$$S_{0,\,r} = 1\,; \qquad S_{1,\,r} = -\sum_{j\neq r}^{(1,\,n)} a_r; \ldots; \qquad S_{n-1,\,r} = (-1)^n \prod_{j\neq r}^{(1,\,n)} a_j\,. \tag{33}$$

The constants A_k are therefore given by

$$A_k = -\frac{S_{n-1,\,k}}{\prod_{j\neq k}^{(1,\,n)} (a_k - a_j)}\,\frac{n!}{\prod_{j=1}^{n} a_j} = \frac{(-1)^n n!}{a_k \prod_{j\neq k}^{(1,\,n)} (a_k - a_j)} \qquad (k = 1, \ldots, n)\,. \tag{34}$$

[3] Cf. O. Perron, *Algebra* (Leipzig: De Gruyter, 1932), 1, No. 22, 92–94.

[4] Cf. S. Chandrasekhar, *Ap. J.*, 101, 328, 1945, esp. eqs. (75) and (81).

Reverting to the variables $1 - q_j$ (eq. [28]), we have

$$A_k = \frac{n!}{\prod_{j \neq k}^{(0,\, n)} (q_k - q_j)} \qquad (k = 1, \ldots, n)\,. \tag{35}$$

Extending definition (35) also to $k = 0$, we can write the solution for μ_n in the form

$$\mu_n = n! \sum_{k=0}^{n} \frac{e^{-(1-q_k)\xi}}{\prod_{j \neq k}^{(0,\, n)} (q_k - q_j)} \qquad (n = 1, \ldots)\,. \tag{36}$$

In particular, for $n = 1$ and 2 we have

$$\mu_1 = \frac{1}{1 - q_1} [1 - e^{-(1-q_1)\xi}]$$

and

$$\mu_2 = \frac{2}{(1 - q_1)(1 - q_2)} + \frac{2\, e^{-(1-q_1)\xi}}{(q_1 - 1)(q_1 - q_2)} + \frac{2\, e^{-(1-q_2)\xi}}{(q_2 - 1)(q_2 - q_1)}\,. \tag{37}$$

If all the clouds are characterized by the same value of q, then

$$q_j = q^j\,, \tag{38}$$

and equations (37) reduce to the ones given by Markarian.[5]

4. *A direct proof of the relations satisfied by the moments.*—It is of interest to verify that the relations between the moments of g obtained in § 3 and, in particular, the differential equation (eq. [24]) governing them are deducible directly from the definition of u as the integral (4). For this purpose we first establish the following lemma:

Lemma.—Let $f(x_1, \ldots, x_n)$ be a symmetrical function in the n variables $x_1, \ldots, x_n$. Then

$$\begin{aligned} I_n &= \int_b^a dx_n \int_b^a dx_{n-1} \int_b^a dx_{n-2} \ldots \int_b^a dx_1 f(x_1, \ldots, x_n) \\ &= n! \int_b^a dx_n \int_{x_n}^a dx_{n-1} \int_{x_{n-1}}^a dx_{n-2} \ldots \int_{x_2}^a dx_1 f(x_1, \ldots, x_n)\,. \end{aligned} \tag{39}$$

Proof.—First, we verify that the lemma is true for two variables; this is very readily done. Next we show that, on the assumption that the lemma is true for all multiple integrals of $(n - 1)$ and lower folds, the truth of the lemma for n-fold integrals can be deduced. The general truth of the lemma would then follow by induction.

Considering, then, the n-fold integral I_n and transforming the $(n - 1)$-fold integral over $x_{n-1}, x_{n-2}, \ldots, x_1$ in accordance with the lemma, we have

$$I_n = (n-1)! \int_b^a dx_n \int_b^a dx_{n-1} \int_{x_{n-1}}^a dx_{n-2} \ldots \int_{x_2}^a dx_1 f(x_1, \ldots, x_n)\,. \tag{40}$$

Splitting the range of integration over x_{n-1} from a to x_n and x_n to b, we have

$$I_n = (n-1)! \int_b^a dx_n \int_{x_n}^a dx_{n-1} \int_{x_{n-1}}^a dx_{n-2} \ldots \int_{x_2}^a dx_1 f(x_1, \ldots, x_n) + J\,, \tag{41}$$

[5] *Op. cit.*, eqs. (10) and (13).

where

$$J = (n-1)!\int_b^a dx_n \int_b^{x_n} dx_{n-1} \int_{x_{n-1}}^a dx_{n-2} \ldots \int_{x_2}^a dx_1 f(x_1, \ldots, x_n)\,. \tag{42}$$

Now, inverting the order of the integration over x_n and x_{n-1} in J and using the symmetry of $f(x_1, \ldots, x_n)$ in x_{n-1} and x_n, we have

$$\begin{aligned} J &= (n-1)!\int_b^a dx_{n-1} \int_{x_{n-1}}^a dx_n \int_{x_{n-1}}^a dx_{n-2} \ldots \int_{x_2}^a dx_1 f(x_1, \ldots, x_n) \\ &= (n-1)!\int_b^a dx_n \int_{x_n}^a dx_{n-1} \int_{x_n}^a dx_{n-2} \int_{x_{n-2}}^a dx_{n-3} \ldots \int_{x_2}^a dx_1 f(x_1, \ldots, x_n)\,. \end{aligned} \tag{43}$$

The $(n-2)$-fold integral over $x_{n-2}, x_{n-3}, \ldots, x_1$ in this last expression for J can be transformed in accordance with the converse form of the lemma for $(n-2)$ variables and leads to

$$J = (n-1)\int_b^a dx_n \int_{x_n}^a dx_{n-1} \int_{x_n}^a dx_{n-2} \int_{x_n}^a dx_{n-3} \ldots \int_{x_n}^a dx_1 f(x_1, \ldots, x_n)\,. \tag{44}$$

By a further application of the lemma for the $(n-1)$ variables $x_{n-1}, x_{n-2}, \ldots, x_1$, we obtain

$$J = (n-1)(n-1)!\int_b^a dx_n \int_{x_n}^a dx_{n-1} \int_{x_{n-1}}^a dx_{n-2} \ldots \int_{x_2}^a dx_1 f(x_1, \ldots, x_n)\,. \tag{45}$$

With J given by equation (45), equation (41) becomes

$$I_n = n!\int_b^a dx_n \int_{x_n}^a dx_{n-1} \int_{x_{n-1}}^a dx_{n-2} \ldots \int_{x_2}^a dx_1 f(x_1, \ldots, x_n)\,; \tag{46}$$

and this verifies the lemma for n variables. The general truth of the lemma therefore follows by induction.

An immediate corollary of the lemma is

$$I_n = n\int_b^a dx_n \int_{x_n}^a dx_{n-1} \int_{x_n}^a dx_{n-2} \ldots \int_{x_n}^a dx_1 f(x_1, \ldots, x_n)\,. \tag{47}$$

This alternative expression for I_n follows from an application of the lemma in the converse form to the $(n-1)$-fold integral over $x_{n-1}, x_{n-2}, \ldots, x_1$ in equation (46).

Returning to the problem of deriving the differential equation (24) connecting the moments of g, we first observe that, by definition,

$$\mu_m = \left[\int_0^\xi \prod_{i=1}^{n(r)} q_i\, dr\right]^m_{\text{average}}\,. \tag{48}$$

Alternatively, we can write

$$\mu_m = \int_0^\xi dr_m \int_0^\xi dr_{m-1} \int_0^\xi dr_{m-2} \ldots \int_0^\xi dr_1 \left\{\prod_{j=1}^{m} \prod_{i=1}^{n(r_j)} q_i\right\}_{\text{average}}\,. \tag{49}$$

The integrand of this m-fold integral is clearly a symmetrical function of the variables.

Accordingly, using the lemma in the form (47), we can rewrite the foregoing expression for μ_m in the form

$$\mu_m = m\int_0^\xi d r_m \int_{r_m}^\xi d r_{m-1} \int_{r_m}^\xi d r_{m-2} \cdots \int_{r_m}^\xi d r_1 \Big\{ \prod_{j=1}^{m} \prod_{i=1}^{n(r_j)} q_i \Big\}_{\text{average}} . \tag{50}$$

With the integration over the variables carried out in this fashion,

$$r_j \geqslant r_m \qquad \text{for all} \qquad j \leqslant m-1 . \tag{51}$$

Under these circumstances

$$n(r_j) - n(r_m) = n(r_j - r_m) , \tag{52}$$

and

$$\prod_{j=1}^{m} \prod_{i=1}^{n(r_j)} q_i = \Big\{ \prod_{i=1}^{n(r_m)} q_i^m \Big\} \Big\{ \prod_{j=1}^{m-1} \prod_{i=1}^{n(r_j - r_m)} q_i \Big\} . \tag{53}$$

In view of the inequality (51), it is evident that the occurrence of clouds in the interval $(0, r_m)$ is uncorrelated with the occurrence of clouds in any of the intervals $r_j - r_m$ $(j = m-1, \ldots, 1)$. Hence

$$\Big\{ \prod_{j=1}^{m} \prod_{i=1}^{n(r_j)} q_i \Big\}_{\text{average}} = \Big\{ \prod_{i=1}^{n(r_m)} q_i^m \Big\}_{\text{average}} \times \Big\{ \prod_{j=1}^{m-1} \prod_{i=1}^{n(r_j - r_m)} q_i \Big\}_{\text{average}} . \tag{54}$$

Using this result in equation (50), we have

$$\mu_m = m\int_0^\xi d r_m \Big\{ \prod_{i=1}^{n(r_m)} q_i^m \Big\}_{\text{average}} \times \int_{r_m}^\xi d r_{m-1} \int_{r_m}^\xi d r_{m-2} \cdots \int_{r_m}^\xi d r_1 \times \Big\{ \prod_{j=1}^{m-1} \prod_{i=1}^{n(r_j - r_m)} q_i \Big\}_{\text{average}} . \tag{55}$$

Now, writing r_j in place of $r_j - r_m (j = m-1, \ldots, 1)$ in equation (55), we have

$$\mu_m = m\int_0^\xi d r \Big\{ \prod_{i=1}^{n(r)} q_i^m \Big\}_{\text{average}} \times \int_0^{\xi - r} d r_{m-1} \int_0^{\xi - r} d r_{m-2} \cdots \int_0^{\xi - r} d r_1 \times \Big\{ \prod_{j=1}^{m-1} \prod_{i=1}^{n(r_j)} q_i \Big\}_{\text{average}} , \tag{56}$$

where, for brevity, we have suppressed the subscript m in r_m. The $(m-1)$-fold integral in equation (56) is clearly $\mu_{m-1}(\xi - r)$. Accordingly, we may rewrite equation (56) in the form

$$\mu_m = m\int_0^\xi d r \Big\{ \prod_{i=1}^{n(r)} q_i^m \Big\}_{\text{average}} \mu_{m-1}(\xi - r) . \tag{57}$$

On the other hand (cf. eqs. [6] and [25]),

$$\left\{\prod_{i=1}^{n(r)} q_i^m\right\}_{\text{average}} = \sum_{n=0}^{\infty} \frac{e^{-r} r^n}{n!} \prod_{i=1}^{n} \int_0^1 d q_i q_i^m \psi(q_i)$$
$$= \sum_{n=0}^{\infty} e^{-r} \frac{(r q_m)^n}{n!} = e^{-r(1-q_m)} . \tag{58}$$

Hence

$$\mu_m = m \int_0^{\xi} d r e^{-r(1-q_m)} \mu_{m-1} (\xi - r) . \tag{59}$$

But this is the integrated form of equation (24) when the boundary condition $\mu_m = 0$ at $\xi = 0$ is also satisfied. The equations and boundary conditions from which solution (36) for the moments was derived in § 3 have now been obtained directly from the definition of u.

5. *An application of the formulae for the moments μ_n to derive certain statistical properties of the interstellar clouds.*—As an illustration of the application of the formulae for the moments derived in § 3, we shall reconsider an example investigated by Markarian,[1] on the assumption that all the interstellar clouds are equally transparent.

Markarian's example is based principally on van Rhijn's tabulation of the counts[6] of the number of stars $N_m(\beta, \lambda)$, per square degree, brighter than a given apparent magnitude m, and in a region of the sky centered at galactic latitude β and galactic longitude λ. In terms of these numbers[7] Markarian evaluated the quantity

$$I(\beta, \lambda) = \sum_m \{ N_{m+1}(\beta, \lambda) - N_m(\beta, \lambda) \} \times 10^{-0.4m} , \tag{60}$$

as a function of β and λ. This quantity, $I(\beta, \lambda)$, is clearly a measure of the brightness of the Milky Way in the region considered. It is, therefore, comparable, apart from a constant of proportionality, to I as we have defined it in equation (1).

In a detailed investigation it will be necessary to compare the observed fluctuations of the quantity (60) from the mean with the theoretical distributions derived on the basis of the integral equation (18). However, in a first attempt, it may suffice to restrict ourselves to the dispersion of the brightness about the mean.

Since we may expect the average number of absorbing clouds in the direction of galactic latitude β to vary with β as cosec β, it is clear that, in determining the dispersion of $I(\beta, \lambda)$ about the mean, we must treat the regions with different β's, separately. Thus, denoting the dispersion in the brightness of the Milky Way at galactic latitude β as $\delta^2(\beta)$, we have

$$\delta^2(\beta) = \frac{\text{Mean}\{ I^2(\beta, \lambda) \}}{[\text{Mean}\{ I(\beta, \lambda) \}]^2} - 1 , \tag{61}$$

where, in taking the means, β is kept constant.

In the investigation we have quoted, Markarian has derived values for the dispersion $\delta^2(\beta)$, according to equation (61) for those values of β for which van Rhijn's (Groups I–IV) and Baker and Kiefer's (Groups V–VII) tables permit a determination. His final results are summarized in Table 1.

Now, on the model of the distribution of stars and clouds adopted in this paper (§ 1), the value of $\delta^2(\beta)$ should be compared with the theoretical quantity,

$$\delta^2(\xi) = \frac{\mu_2}{\mu_1^2} - 1 . \tag{62}$$

[6] *Groningen Pub.*, No. 43, 1924. Markarian has also used the data given in R. H. Baker and L. Kiefer, *Ap. J.*, **94**, 482, 1941.

[7] Since the areas actually used in van Rhijn's tabulation extend over an appreciable range of β, Markarian had to reduce the observed numbers to a mean latitude β by applying a correction based on the observed mean variation of $N_m(\beta, \lambda)$ with β.

With μ_1 and μ_2 given by equations (37), we have

$$\delta^2(\xi) = \frac{2(1-q_1)}{(1-q_2)(q_1-q_2)} \frac{(1-q_2)[1-e^{-(1-q_1)\xi}] - (1-q_1)[1-e^{-(1-q_2)\xi}]}{[1-e^{-(1-q_1)\xi}]^2} - 1 \,, \tag{63}$$

where it may be recalled that ξ is the average number of clouds in the direction of the line of sight and q_1 and q_2 are the mean and the mean square of the transparency factor q of the clouds:

$$q_1 = \int_0^1 q\psi(q)\,dq \qquad \text{and} \qquad q_2 = \int_0^1 q^2\psi(q)\,dq \,. \tag{64}$$

While q_1 and q_2 occur as two independent parameters in equation (63), it is evident that,

TABLE 1

RESULTS OF MARKARIAN'S ANALYSIS OF $N_m(\beta, \lambda)$

	GROUP						
	I	II	III	IV	V	VI	VII
β	0°	±10°	±30°	±40°	−10°	±10°	0°
λ	100°	100°	100°	100°	160°	130°	130°
No. of regions used......	11	12	10	11	10	9	11
$\delta^2(\beta)$	0.092	0.075	0.030	0.020	0.082	0.100	0.126

by virtue of definitions (64), our freedom in assigning values to q_2 for a given q_1 is strictly limited; for the inequality

$$q_1^2 \leqslant q_2 < q_1 \tag{65}$$

is a consequence of the definitions of these quantities only. And, moreover, the equality between q_1^2 and q_2 can occur only when all the clouds are equally transparent with a factor q_1. It may also be noted in this connection that, according to equation (63),

$$\delta^2(\xi) \to \frac{(1-q_1)(q_1-q_2)}{1-q_2} \qquad \text{as} \qquad \xi \to \infty$$

and

$$\delta^2(\xi) \to 0 \qquad \text{as} \qquad \xi \to 0 \,. \tag{66}$$

For a comparison of the observed values of $\delta^2(\beta)$ with the theoretical predictions based on equation (63), we require a knowledge of the three parameters, ξ, q_1, and q_2, which the theoretical expression for $\delta^2(\xi)$ involves. However, of the three parameters, ξ and q_1 are not independent if we make use of the results of the counts of extragalactic nebulae.[8] According to these latter counts, the mean photographic absorption, $\Delta m(\beta)$, in the direction of galactic latitude β is given by

$$\Delta m(\beta) = 0^{\mathrm{m}}.25 \operatorname{cosec} \beta \,. \tag{67}$$

If τ is the mean absorption, in magnitudes, per cloud, the number of clouds to be expected, on the average, in the direction β is

$$\xi = \frac{0^{\mathrm{m}}.25\,|\operatorname{cosec} \beta|}{\tau} \,. \tag{68}$$

[8] E. P. Hubble, *Ap. J.*, 79, 8, 1934.

But, by definition, τ is related to $\bar{q} = q_1$ by the equation

$$\tau = -2.5 \log q_1 . \tag{69}$$

Hence

$$\xi(\beta) = -0.1 \frac{|\operatorname{cosec} \beta|}{\log q_1} . \tag{70}$$

Consequently, for any assigned value of q_1, we can determine the average number of clouds in the direction β. Values of ξ derived in this fashion for three assigned values of q_1 (0.75, 0.80, and 0.85) are listed in Table 2.

TABLE 2

AVERAGE NUMBER OF CLOUDS IN DIRECTION OF GALACTIC LATITUDE β FOR THREE ASSIGNED VALUES OF q_1

q_1	τ (MAG.)	ξ			
		$\beta=0°$	$\beta=\pm 10°$	$\beta=\pm 30°$	$\beta=\pm 40°$
0.75	0.31	∞	4.61	1.60	1.24
.80	.24	∞	5.94	2.06	1.61
0.85	0.18	∞	8.15	2.83	2.20

In interpreting his deduced values of $\delta^2(\beta)$, Markarian made the assumption that all the clouds are equally transparent. On this assumption, $q_2 = q_1^2$, and $\delta^2(\beta)$ becomes determinate when q_1 is given. However, the expression for $\delta^2(\xi)$, which allows for an arbitrary distribution of q, involves the additional parameter q_2. Accordingly, using equation (63), we have computed $\delta^2(\xi)$ for various values of q_2 and for $q_1 = 0.75$, 0.80, and 0.85, respectively. The results of the calculation are illustrated in the form of curves in Figure 1, a ($q_1 = 0.75$), b($q_1 = 0.80$), and c($q_1 = 0.85$).

To appreciate what latitude we have for changing q_2 for a given q_1 and what a difference in q_2 from q_1^2 means in terms of the distribution of q, we may note that, for the frequency function,

$$\psi(q) = (n+1) q^n \qquad (0 \leqslant q \leqslant 1) , \tag{71}$$

$$q_1 = \frac{n+1}{n+2}, \qquad q_2 = \frac{n+1}{n+3}, \qquad \text{and} \qquad \frac{q_2}{q_1^2} - 1 = \frac{1}{(n+3)(n+1)} . \tag{72}$$

In particular, for $n = 5$,

$$q_1 = 0.857 , \qquad q_1^2 = 0.735 , \qquad q_2 = 0.75 , \qquad \text{and} \qquad \frac{q_2}{q_1^2} - 1 = 0.021 . \tag{73}$$

Accordingly, for $q_1 = 0.85$, a change of q_2 from $(0.85)^2 = 0.7225$ to 0.7325 is by no means an unduly large change. Bearing this in mind, we conclude from an examination of the curves in Figure 1 that the predicted variation of $\delta^2(\xi)$ for a given q_1 depends rather sensitively on q_2; indeed, it would appear that relatively slight changes in q_2 (keeping q_1 fixed) affect $\delta^2(\xi)$ almost as much as quite appreciable changes in q_1 (keeping $q_2 = q_1^2$). This is a somewhat unexpected result disclosed by the present analysis.

In Figure 1, a, b, and c we have plotted the observed values of $\delta^2(\beta)$ against the ξ's appropriate for the values of q_1 to which each of the figures refers (cf. Table 2). It is seen that, with the present data, we cannot distinguish in a unique manner the different effects of q_1 and q_2. However, it does appear that

$$q_1 = 0.85 \qquad \text{and} \qquad q_2 = 0.7325 \tag{74}$$

give the best fit with the observations. It is of interest to recall in this connection that, on the balance of all the evidence available, Markarian favored[9] the acceptance of a value of $q_1 = 0.85$, though the agreement of his observed points with the curve $q_1 = 0.75$ and $q_2 = q_1^2 = 0.5625$ is definitely better than with the curve $q_1 = 0.85$ and $q_2 = q_1^2 = 0.7225$ (cf. Fig. 1, *a* and *c*).

While the values derived for q_1 and q_2 (eq. [74]) are uncertain, it is nevertheless of some interest to observe that these values correspond to a root-mean-square deviation of 0.1 in q. A variation in q of this amount (i.e., ± 0.1) about the mean value $q_1 = 0.85$ corresponds to a variation in the true optical thicknesses of interstellar clouds in the range

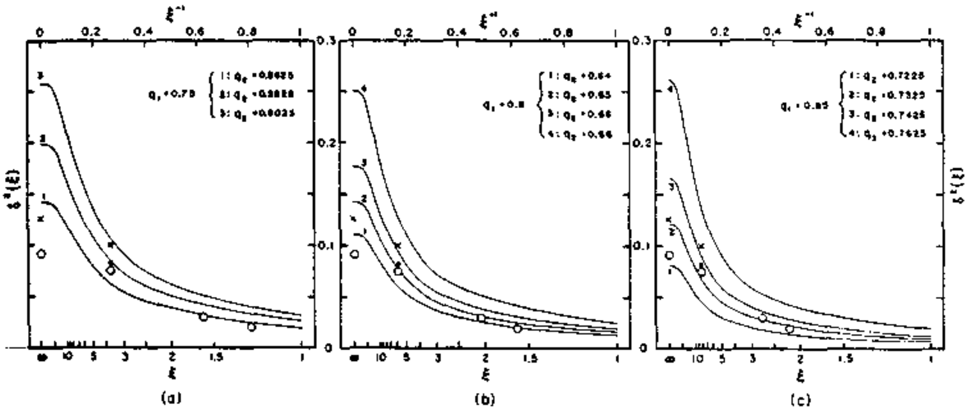


FIG. 1.—The transparency factor q of the interstellar clouds as derived from the observed dispersion in the brightness of the Milky Way at various galactic latitudes. The curves represent the predicted variation of the dispersion $\delta^2(\xi)$ with the average number of absorbing clouds, ξ, in the line of sight for various values of the mean (q_1) and mean square (q_2) of the transparency factor of the clouds. The different sets of curves are for different values of q_1 (0.75 in *a*; 0.80 in *b*; and 0.85 in *c*), the parameter distinguishing the curves in each set being q_2. The lowest curve in each is for the case in which all the clouds are equally transparent and $q_2 = q_1^2$.

The dispersion of the observed brightness of the Milky Way, $\delta^2(\beta)$, at various galactic latitudes can be represented as a variation with ξ if a value of q_1 is assumed (cf. eq. [70]). The values of $\delta^2(\beta)$ deduced by Markarian are plotted in the figure against the ξ's appropriate for each figure (cf. Table 2). The crosses, the open circles, and the solid dots refer to the regions centered at $\lambda = 130°$, $\lambda = 100°$, and $\lambda = 160°$, respectively.

0.29 and 0.05; a variation in the absorptive power of clouds of this amount is entirely reasonable.

Again, if we assume that the average photographic extinction coefficient is 2 mag. per kiloparsec,[10] then we shall need an average of 10–11 clouds per kiloparsec. This estimate is not necessarily in discord with the usual estimate[11] of 7 clouds per kiloparsec; for it may be estimated that a dispersion of 0.1 in q means that about three-fourths of all the clouds (i.e., 7 or 8 in the present instance) will have values in the range 0.75–0.95 and it is possible that the few dense clouds, recognized as such, are not included in the general survey. In any case the present rediscussion of Markarian's example would seem to indicate that a great deal of information concerning the interstellar clouds can be derived by an extension of the basic observational material and their discussion along the lines of this paper.

[9] This value of q_1 is also favored by L. Spitzer, *Ap. J.*, **108**, 276, 1948, esp. p. 278.

[10] H. van de Hulst, *Rech. Astr. Obs. Utrecht*, Vol. **11**, Part II, 1949.

[11] Cf. Spitzer, *op. cit.*; and B. Strömgren, *Ap. J.*, **108**, 242, 1948.

This page is intentionally left blank

IV. Turbulence and Hydromagnetic Problems in Astrophysics

After completing the writing of the book *Radiative Transfer* in 1948, Chandra, following a brief vacation, decided to embark on the study of turbulence. "Turbulence is a phenomenon of the large scale," he said, "and the essence of astrophysical and, indeed, also of geophysical phenomena is the scale." Having decided on the topic for his entry into a new area, Chandra began his studies by making turbulence the subject of his Monday evening seminars at the Yerkes Observatory. Heisenberg had just then published two papers developing an elementary theory based on an easily visualizable picture of what was taking place in a turbulent medium. He had provided an integral equation for determining the spectrum of turbulence encoded in a function $F(k)$, where k denoted the wave number of eddies in the turbulence.

Chandra found that in the case of statistically stationary turbulence, he could obtain an explicit solution to Heisenberg's equation; furthermore, in the case of decay, by a suitable choice of variables, the integral equation could be reduced to a linear first order differential equation and one quadrature. This is the subject matter of Paper 1. Paper 2, the Third Henry Norris Lecture, gives a masterly historical review of the subject of turbulence in the context of astrophysics and offers a brief survey of the various problems in astrophysics in which turbulence may play an important role.

Heisenberg's theory of turbulence led Chandra to the study of the statistical theory of homogeneous isotropic turbulence and to its application in magneto-hydrodynamics. Noting that isotropic turbulence presents the subject in its simplest manifestation, and that it is a general feature of turbulence that there is always a preferred direction defined by the mean flow, Chandra generalized the theory of isotropic turbulence to axisymmetric turbulence. Paper 3 contains a review of the underlying physical theory as well as a summary of his work in magneto-hydrodynamics.

The next two papers are joint papers with Enrico Fermi. Speaking about this collaborative work, Chandra said that it was one of the most exciting experiences in his entire scientific career. Paper 4 deals with the nature of the magnetic field in the spiral arm of the galaxy and Paper 5 is concerned with problems related to the gravitational stability of cosmic masses of infinite electrical conductivity in the presence of a magnetic field.

IV. Turbulence and Hydromagnetic Problems in Astrophysics

On Heisenberg's elementary theory of turbulence

By S. Chandrasekhar, F.R.S., *Yerkes Observatory*

(*Received* 22 *July* 1949)

In this paper the spectrum of turbulence is considered on the basis of an elementary theory recently developed by Heisenberg. Explicit solutions for the spectrum have been obtained both when the conditions are stationary and an equilibrium spectrum obtains and when the conditions are non-stationary and the turbulence is decaying. In the former case the problem admits of an explicit solution. In the latter case the problem reduces to determining a one-parametric family of solutions of a certain second-order differential equation. The decay spectra for various values of the Reynolds number (which remains constant during the decay) are illustrated.

1. Introduction

In two recent papers Heisenberg (1948 *a*, *b*) has developed an elementary theory for determining the spectrum of turbulence on an easily visualizable picture of what is taking place in a turbulent medium. The physical ideas underlying Heisenberg's theory are stated very simply:

Considering the rate of dissipation of energy by eddies with wave numbers less than a particular k, Heisenberg distinguishes between the energy directly dissipated in the form of molecular motion and thermal energy and the energy communicated in the form of kinetic energy to all eddies with wave numbers exceeding the specified k. According to the equations of hydrodynamics we can write for the dissipation into thermal energy, per unit volume, the expression

$$\epsilon_k(\text{thermal}) = 2\rho\nu \int_0^k F(k)\, k^2 dk, \tag{1}$$

where $F(k)$ denotes the spectrum of turbulence, ν the kinematic viscosity (i.e. the coefficient of viscosity as ordinarily defined divided by the density ρ). Equation (1) follows from the usual expression for the total dissipation of energy by viscosity in an isotropically turbulent medium, namely

$$\epsilon = \rho\nu\,\overline{|\operatorname{curl}\mathbf{v}|^2} = 2\rho\nu\int_0^\infty F(k)\,k^2dk. \tag{2}$$

In analogy with (1) Heisenberg writes for the energy transferred to all eddies with wave numbers exceeding a given k by eddies with wave numbers less than this k, the expression

$$\epsilon_k(\text{mechanical}) = 2\rho\nu_k\int_0^k F(k)\,k^2dk. \tag{3}$$

In writing this expression, the assumption is made that the process of energy transfer between the sets of eddies considered can be visualized in terms of a suitably defined coefficient of '*eddy viscosity*' ν_k. Heisenberg further assumes that ν_k can be expressed as an integral over the spectrum in the form

$$\nu_k = \kappa\int_k^\infty \frac{dk}{k^{\frac{3}{2}}}\sqrt{F(k)}, \tag{4}$$

where κ is a numerical constant of order unity. The justification for (4), apart from the fact that the quantity on the right-hand side is of the right dimensions, is that for a power law, $F(k)\propto k^{-n}$, equation (4) is equivalent to the expression

$$\nu_k \sim l_k v_k \sim v_k/k,$$

derived on the picture of the eddies with wave numbers exceeding k, describing a mean free path l_k of order $1/k$ with a certain mean speed $v_k \propto k^{-\frac{1}{2}(n-1)}$.

Combining equations (1), (3) and (4) we have

$$\epsilon_k = 2\rho\left\{\nu + \kappa\int_k^\infty \frac{dk}{k^{\frac{3}{2}}}\sqrt{F(k)}\right\}\int_0^k F(k)\,k^2dk. \tag{5}$$

This is the equation in terms of which Heisenberg has discussed the spectrum of turbulence, both when the conditions are stationary and an equilibrium spectrum obtains and when the conditions are nonstationary and the turbulence is decaying.

When the equilibrium spectrum obtains

$$\epsilon_k = \text{constant}, \tag{6}$$

independent of k. On the other hand, when there is no external source of energy and the turbulence is decaying,

$$\epsilon_k = -\frac{\partial}{\partial t}\int_0^k F(k,t)\,dk. \tag{7}$$

Heisenberg has considered these two cases and derived the principal features of the spectrum. However, it appears that the mathematical problem presented by equations (5), (6) and (7) can be solved very much more completely than was attempted by Heisenberg. It is the object of this paper to present the complete solutions of the two problems discussed by Heisenberg (cf. Chandrasekhar 1949).

2. The spectrum of stationary turbulence

When the conditions are stationary ϵ_k should be independent of k, and the explicit solution of equation (5) can be found in the following manner:

Differentiating equation (5) with respect to k, we obtain

$$\frac{\nu}{\kappa}+\int_k^\infty \frac{dk}{k^{\frac{3}{2}}}\sqrt{F(k)} = \frac{1}{k^2[F(k)\,k^3]^{\frac{1}{2}}}\int_0^k F(k)\,k^2\,dk. \tag{8}$$

It is now convenient to introduce the auxiliary functions

$$g = k^3F(k) \quad \text{and} \quad y = \int_0^k F(k)\,k^2\,dk. \tag{9}$$

Relations which readily follow from these definitions are

$$\frac{dk}{dy} = \frac{k}{g}, \tag{10}$$

and

$$\log k = \text{constant} + \int_0^y dy/g(y). \tag{11}$$

Using equations (9) and (10), we can rewrite equation (8) in the form

$$\frac{\nu}{\kappa}+\int_y^\infty \frac{dy}{k^2g^{\frac{1}{2}}}-\frac{y}{k^2g^{\frac{1}{2}}} = 0. \tag{12}$$

Now differentiating this equation with respect to y and making use of (10), we obtain

$$-\frac{1}{k^2g^{\frac{1}{2}}}+\frac{2y}{k^2g^{\frac{3}{2}}}-\frac{1}{k^2}\frac{d}{dy}\left(\frac{y}{g^{\frac{1}{2}}}\right) = 0. \tag{13}$$

On further reduction, this equation becomes

$$\frac{dg}{dy}-\frac{4}{y}g+4 = 0. \tag{14}$$

This equation is readily integrated, and the general solution is given by

$$g = \tfrac{4}{3}y(1-cy^3), \tag{15}$$

where c is a constant.

Inserting the solution (15) for $g(y)$ in (11), and evaluating the integral on the right-hand side we obtain

$$k = \alpha\left(\frac{y^3}{1-cy^3}\right)^{\frac{1}{4}}, \tag{16}$$

where α is a constant.

Alternative forms of the solution represented by equations (15) and (16) are

$$y^3 = \frac{k^4}{\alpha^4+ck^4} \quad \text{and} \quad g = \frac{4}{3}\left(\frac{\alpha y}{k}\right)^4. \tag{17}$$

The spectrum now follows from equations (9), (15) and (17). We have

$$F(k) = \frac{4}{3}\frac{\alpha^4}{k^{\frac{5}{3}}(\alpha^4+ck^4)^{\frac{4}{3}}}. \tag{18}$$

We can rewrite this in the form

$$F(k) = F_0\left(\frac{k_0}{k}\right)^{\frac{5}{3}} \frac{1}{[1+(k/k_s)^4]^{\frac{4}{3}}}, \tag{19}$$

where

$$k_s = \alpha/c^{\frac{1}{4}} \quad \text{and} \quad F_0 = 4/3\alpha^{\frac{4}{3}}k_0^{\frac{5}{3}}. \tag{20}$$

It remains to relate the constants α and c with the parameter ν/κ which occurs in the original equation of the problem. For this purpose we use equation (8). According to this equation (cf. equations (9) and (10))

$$\begin{aligned}\frac{\nu}{\kappa} &= \frac{y}{k^2\sqrt{g}} - \frac{\alpha^2}{2\sqrt{3}}\int_\infty^k \frac{1}{(c+\alpha^4/k^4)^{\frac{2}{3}}}\frac{d}{dk}\left(\frac{1}{k^4}\right)dk \\ &= \frac{y}{k^2\sqrt{g}} - \frac{\sqrt{3}}{2\alpha^2}\left[\left(c+\frac{\alpha^4}{k^4}\right)^{\frac{1}{3}} - c^{\frac{1}{3}}\right],\end{aligned} \tag{21}$$

or, using equations (17), we find

$$\frac{\nu}{\kappa} = \frac{\sqrt{3}}{2}\frac{c^{\frac{1}{3}}}{\alpha^2}. \tag{22}$$

This is the required relation.

Returning to the spectrum of stationary turbulence given by equation (19), we may first observe that according to equations (20) and (22)

$$k_s = \frac{1}{\sqrt{\alpha}}\left(\frac{\sqrt{3}}{2}\frac{\kappa}{\nu}\right)^{\frac{3}{4}}. \tag{23}$$

From this relation it is seen that when $\nu \to 0$ (or, equivalently, when the Reynolds number tends to infinity) $k_s \to \infty$, and the spectrum given by (19) reduces to the equilibrium Kolmogoroff spectrum. On the other hand, for any finite ν, the spectrum follows a k^{-7}-law when $k \gg k_s$; k_s is accordingly the wave number at which ϵ_k (thermal) becomes comparable to ϵ_k(mechanical) in the transaction between the eddies. These results, obtained from the explicit solution of the problem, are in agreement with Heisenberg's qualitative discussion of this same problem.

For $\nu \to 0$, the wave number k_s is related to the Reynolds number in a simple way; for, since $c \to 0$ as $\nu \to 0$, we may write in a sufficient approximation:

$$F_0 = 4/3\alpha^{\frac{4}{3}}k_0^{\frac{5}{3}} \simeq F(k_0). \tag{24}$$

Using (24) to eliminate α from (23), we find

$$k_s = k_0\left(\frac{3}{4}\frac{\kappa}{\nu}\sqrt{\frac{F(k_0)}{k_0}}\right)^{\frac{3}{4}}. \tag{25}$$

Heisenberg defines the Reynolds number R_0 of the entire motion by the equation

$$R_0 = \frac{\pi}{\nu}\sqrt{\frac{3F(k_0)}{k_0}}, \tag{26}$$

which is ν^{-1} times the product of the root-mean-square velocity of turbulent motion and the diameter of the largest eddies present. In terms of this Reynolds number the expression for k_s becomes

$$k_s = 0{\cdot}2211k_0(R_0\kappa)^{\frac{3}{4}}. \tag{27}*$$

* The use of the explicit solution for the spectrum given here changes the numerical constants in Heisenberg's paper (1948*a*). The most important of these changes are that in his equations (27), (30), (31), (73) and (76) the numerical coefficients should be 0·3164, 0·2211, 4·52, 0·0756 and 0·349 instead of 0·0496, 0·16, 6·25, 0·05 and 0·42 respectively.

3. The decay of turbulence

On Heisenberg's theory, the decay of turbulence is governed by the equation (cf. equations (5) and (7))

$$-\frac{\partial}{\partial t}\int_0^k F(k,t)\,dk = 2\left\{\nu+\kappa\int_k^\infty \frac{dk}{k^{\frac{3}{2}}}\sqrt{F(k,t)}\right\}\int_0^k F(k,t)\,k^2\,dk. \tag{28}$$

If the initial spectrum $F(k,0)$ is given, the foregoing equation will determine the spectrum for all later times. A case of some importance which arises in this connexion is the following: Suppose that we have initially an equilibrium spectrum governed by equation (19) and that at a certain instant the agency maintaining the turbulence is cut off. Then in the decay of turbulence which will ensue we may distinguish three stages: an early stage during which the larger eddies ($k \sim k_0$) adjust themselves to the fact that no energy is being communicated to them; an intermediate stage during which there is a sufficient store of energy among the larger eddies to maintain an equilibrium distribution among the smaller eddies and when the Reynolds number remains approximately constant; and finally, a last stage during which the store of energy among the larger eddies is getting exhausted and the Reynolds number decreases to zero. While a unified discussion of all three stages of decay is a difficult problem, it appears that on the basis of Heisenberg's theory we can follow the second stage quite completely and in an explicit fashion. For, from the constancy of the Reynolds number which we expect during this stage, we conclude that the spectrum must be 'self-preserving' in the sense that it keeps the same form though the scale may change with time. From equation (28), the condition of self-preservation is seen to be equivalent to seeking solutions of this equation of the form

$$F(k,t) = \frac{1}{\kappa^2 k_0^2 t_0^2}\left(\frac{t_0}{t}\right)^{\frac{1}{2}} f\left(\frac{k\sqrt{t}}{k_0\sqrt{t_0}}\right), \tag{29}$$

where k_0 and t_0 are certain constants and f is a function of the argument specified.

For $F(k,t)$ of the form (29), equation (28) becomes

$$\frac{1}{4}\int_0^x \left[f(x) - x\frac{df}{dx}\right]dx = \left\{\frac{1}{R}+\int_x^\infty \frac{dx}{x^{\frac{3}{2}}}\sqrt{f(x)}\right\}\int_0^x f(x)\,x^2\,dx, \tag{30}$$

where

$$R = 1/\nu k_0^2 t_0 \tag{31}$$

is a measure of the Reynolds number which remains constant during the stage of the decay considered.

From equation (30) it readily follows that

$$f(x) \to \text{constant}\, x \quad (x \to 0). \tag{32}$$

Heisenberg (1948*b*) has discussed in general terms the solutions of equation (30) which have the behaviour (32) at the origin and has shown that for the case of infinite Reynolds number ($\nu = 0$) the solution must have the asymptotic form

$$f(x) \to \text{constant}\, x^{-\frac{5}{3}} \quad (x \to \infty;\ \nu = 0), \tag{33}$$

while for a finite Reynolds number

$$f(x) \to \text{constant}\, x^{-7} \quad (x \to \infty;\ \nu > 0). \tag{34}$$

As Heisenberg has pointed out these behaviours at infinity are to be expected on general grounds. But it is perhaps of some interest to obtain the explicit solutions of the problem. This can be achieved in the following manner:

Differentiating equation (30) with respect to x and rearranging the terms, we obtain

$$\frac{1}{4}\left(1-\frac{x}{f}\frac{df}{dx}\right) = x^2\left\{\frac{1}{R}+\int_x^\infty \frac{dx}{x^{\frac{5}{2}}}\sqrt{f(x)}\right\} - \frac{1}{\sqrt{(fx^3)}}\int_0^x f(x)\, x^2 dx. \tag{35}$$

It is now convenient to introduce the auxiliary functions

$$g = x^3 f(x) \quad \text{and} \quad y = \int_0^x f(x)\, x^2 dx. \tag{36}$$

Relations which follow from these definitions are

$$\frac{dx}{dy} = \frac{x}{g}, \tag{37}$$

and

$$\frac{dg}{dy} = \left(3x^2 f + x^3\frac{df}{dx}\right)\frac{x}{g} = 3+\frac{x}{f}\frac{df}{dx}. \tag{38}$$

Using these relations, we can rewrite equation (35) in the form

$$\frac{1}{R}-\frac{1}{x^2}\left\{\frac{1}{4}\left(4-\frac{dg}{dy}\right)+\frac{y}{\sqrt{g}}\right\}+\int_y^\infty \frac{dy}{x^2\sqrt{g}} = 0. \tag{39}$$

Now differentiating this equation with respect to y we obtain

$$\frac{2}{x^3}\frac{dx}{dy}\left\{\frac{1}{4}\left(4-\frac{dg}{dy}\right)+\frac{y}{\sqrt{g}}\right\} - \frac{1}{x^2}\frac{d}{dy}\left\{\frac{1}{4}\left(4-\frac{dg}{dy}\right)+\frac{y}{\sqrt{g}}\right\} - \frac{1}{x^2\sqrt{g}} = 0. \tag{40}$$

Inserting for dx/dy in value (37), we find that equation (40) reduces to

$$g^{\frac{3}{2}}g'' + 2y(4+g') + 2g^{\frac{1}{2}}(4-g') - 8g = 0, \tag{41}$$

where we have used primes to denote differentiation with respect to y.

Once equation (41) has been integrated, the solution can be completed by a further quadrature; for, according to equation (37),

$$\log(x/x_0) = \int^y \frac{dy}{g(y)}. \tag{42}$$

In (42), x_0 is a constant of integration. There is, however, no loss of generality in setting

$$x_0 = 1, \tag{43}$$

since this represents only a scale factor and can be absorbed in the definition of k_s (cf. equation (29)). With this understanding, we shall adopt

$$\log x = \int^y \frac{dy}{g(y)} \tag{44}$$

as the equation relating x and y.

3·1. *The physical solutions*

As we have already stated (cf. equation (32)) the physical solutions must be such that $f(x)$ becomes proportional to x as $x \to 0$; according to equations (36) and (38) this means that we must seek solutions of equation (41) which start tangentially to the line $g = 4y$ at the origin:

$$g' = 4 \quad \text{and} \quad g = 4y \quad (y \to 0). \tag{45}$$

From equation (41) it now follows that there exists a one parametric family of solutions which have this behaviour at $y = 0$. Indeed, they arise from the continuation of the solutions represented by the following series expansion valid at $y = 0$:

$$\begin{aligned} g(y) = {} & 4y + y^{\frac{3}{2}}(a + \tfrac{4}{3}\log y) \\ & + y^2(a_2 + b_2 \log y - \tfrac{1}{3}\log^2 y) \\ & + y^{\frac{5}{2}}(a_3 + b_3 \log y + c_3 \log^2 y + \tfrac{11}{90}\log^3 y) + O(y^2 \log^4 y), \end{aligned} \tag{46}$$

where a is an arbitrary constant and

$$\left. \begin{aligned} a_2 &= -2{\cdot}861111 + 0{\cdot}958333a - 0{\cdot}1875a^2, \\ b_2 &= +1{\cdot}277777 - 0{\cdot}5a, \\ a_3 &= -2{\cdot}071525 + 1{\cdot}5889722a - 0{\cdot}383542a^2 + 0{\cdot}0515625a^3, \\ b_3 &= +2{\cdot}118629 - 1{\cdot}022777a + 0{\cdot}20625a^2, \\ c_3 &= -0{\cdot}6818519 + 0{\cdot}275a. \end{aligned} \right\} \tag{47}$$

Corresponding to this behaviour of g near the origin, the relation between x and y is (cf. equation (44))

$$\begin{aligned} \log x = {} & \tfrac{1}{4}\log y - y^{\frac{1}{2}}\left(\frac{a}{8} - \tfrac{1}{3} + \tfrac{1}{6}\log y\right) \\ & + y[(\alpha_2 - \beta_2 + 2\gamma_2) + (\beta_2 - 2\gamma_2)\log y + \gamma_2 \log^2 y] \\ & - y^{\frac{3}{2}}[(\tfrac{2}{3} - \tfrac{4}{9}\beta_3 + \tfrac{16}{27}\gamma_3 - \tfrac{32}{27}\delta_3) + (\tfrac{2}{3}\beta_3 - \tfrac{8}{9}\gamma_3 + \tfrac{16}{9}\delta_3)\log y \\ & + (\tfrac{2}{3}\gamma_3 - \tfrac{4}{3}\delta_3)\log^2 y + \tfrac{2}{3}\delta_3 \log^3 y] + O(y^2 \log^4 y), \end{aligned} \tag{48}$$

where

$$\left. \begin{aligned} \alpha_2 &= \frac{1}{4}\left(\frac{a^2}{16} - \frac{a_2}{4}\right), & \alpha_3 &= \frac{1}{4}\left(\frac{a_3}{4} - \frac{aa_2}{8} + \frac{a^3}{64}\right), \\ \beta_2 &= \frac{1}{4}\left(\frac{a}{6} - \frac{b_2}{4}\right), & \beta_3 &= \frac{1}{4}\left(\frac{b_3}{4} - \frac{a_2}{6} - \frac{ab_2}{8} + \frac{a^2}{16}\right), \\ \gamma_2 &= \tfrac{7}{144}, & \gamma_3 &= \frac{1}{4}\left(\frac{c_3}{4} - \frac{b_2}{6} + \frac{a}{8}\right), \\ \text{and} & & \delta_3 &= \tfrac{133}{4320}. \end{aligned} \right\} \tag{49}$$

The solutions which have the behaviours described by the foregoing equations have the following arrangement in the (g, y) plane (figure 1).

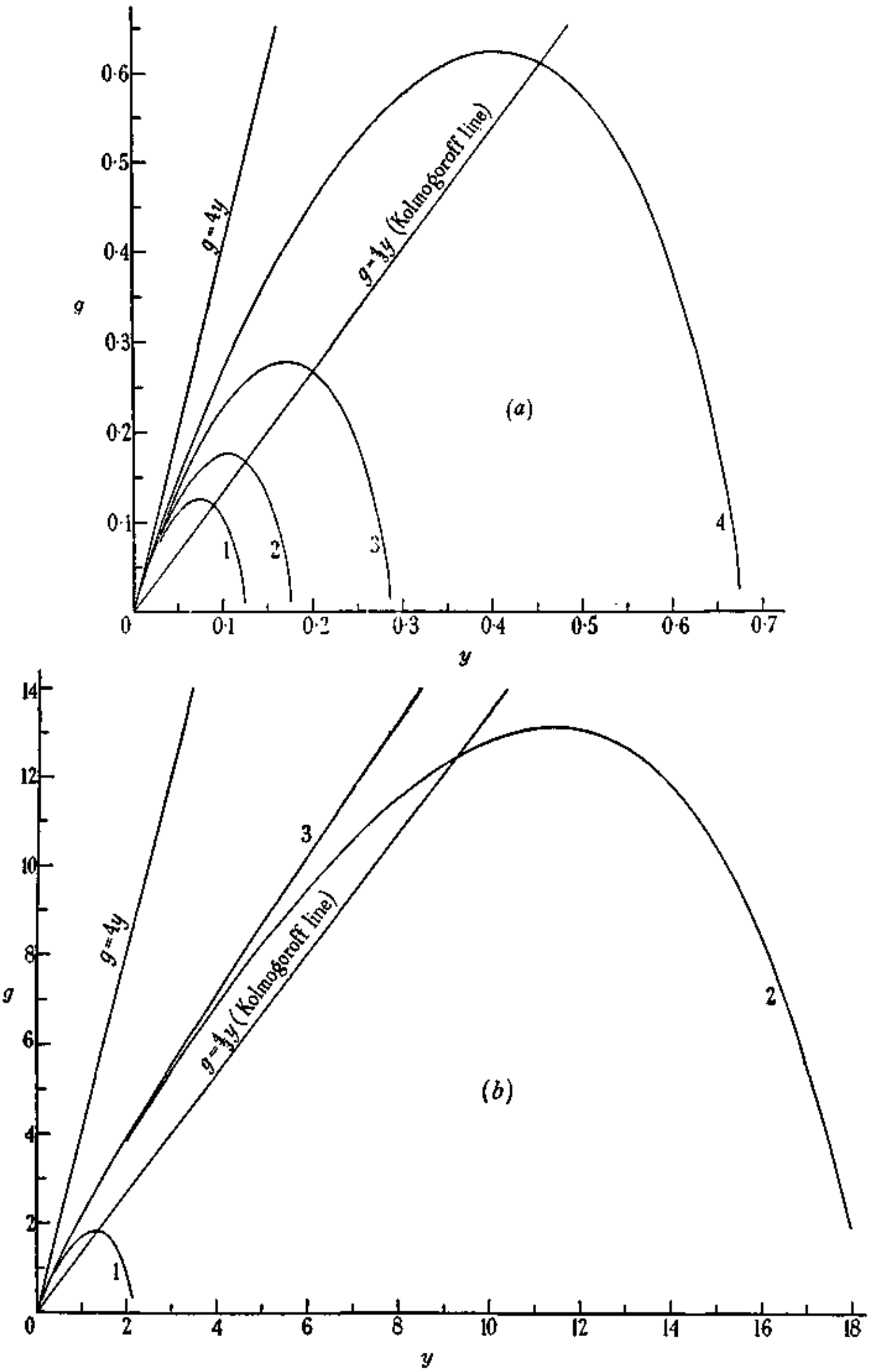


FIGURE 1. The solution curves in the (g, y)-plane. In (a), the curves marked 1, 2, 3 and 4 are for $a = 0$, 0·5, 1·0 and 1·5 respectively. In (b), the curves marked 1, 2 and 3 are for $a = 1{\cdot}75$, 1·81 and 1·8104739 respectively; for $y \to \infty$, the curve 3 tends to the line $g = \frac{4}{3}y$ asymptotically.

[27]

There exists a value of $a = a^*$ ($\sim 1{\cdot}810\ldots$) such that for all $a < a^*$ the solutions are bounded and have a single maximum and reach the y-axis for a finite y ($= y_0$, say) with the behaviour

$$g(y) = 4(y_0-y) + \frac{32}{3y_0}(y_0-y)^{\frac{3}{2}} + \frac{8}{y_0}\left(\frac{16}{3y_0}-1\right)(y_0-y)^2 + O([y_0-y]^{\frac{5}{2}}) \quad (a<a^*;\ y\to y_0;\ y<y_0). \qquad (50)\dagger$$

These solutions are 'bounded' by the solution for $a = a^*$ which increases monotonically, approaching the line $g = \frac{4}{3}y$ for $y \to \infty$ with the behaviour

$$g = \tfrac{4}{3}y[1 + 0{\cdot}659829y^{-\frac{1}{2}} + 0{\cdot}0634921y^{-1} - 0{\cdot}0674957y^{-\frac{3}{2}} - 0{\cdot}0387767y^{-2} + 0{\cdot}0314475y^{-\frac{5}{2}} + \ldots]. \qquad (51)$$

The solutions for $a > a^*$ have no physical meaning; they correspond to negative values for the coefficient of viscosity (see equation (57) below).

According to equation (38), the behaviours (50) and (51) correspond to

$$\frac{x}{f}\frac{df}{dx} = -7 \quad \text{and} \quad -\tfrac{5}{3} \text{ respectively.} \qquad (52)$$

$$\left.\begin{array}{ll} \text{Thus} & f(x) \propto x^{-7} \quad (x\to\infty,\ a<a^*) \\ \text{and} & f(x) \propto x^{-\frac{5}{3}} \quad (x\to\infty,\ a=a^*). \end{array}\right\} \qquad (53)$$

The solutions for $a < a^*$ therefore correspond to finite values of the Reynolds number, while the solution for $a = a^*$ is that appropriate for zero viscosity (or infinite Reynolds number). In the former case we have the k^{-7} behaviour as $k \to \infty$, while in the latter case we have the $k^{-\frac{5}{3}}$ behaviour of the Kolmogoroff spectrum.

When $a < a^*$ but close to a^*, the solution curves in the (g, y) plane run parallel to the line $g = \frac{4}{3}y$ for a long while before turning down; this corresponds to the fact that for large Reynolds numbers there exists a portion of the spectrum in which the $k^{-\frac{5}{3}}$ law is closely followed before the k^{-7} law becomes operative.

It remains to relate the parameter a with the Reynolds number R. For this purpose we consider equation (39) for $y \to 0$. Using the relations (cf. equations (46) and (47))

$$\left.\begin{aligned} g &= 4y + y^{\frac{3}{2}}(a + \tfrac{4}{3}\log y) + O(y^2\log^2 y), \\ g' &= 4 + y^{\frac{1}{2}}(\tfrac{3}{2}a + \tfrac{4}{3} + 2\log y) + O(y\log^2 y), \\ x^2 &= y^{\frac{1}{2}} + O(y\log y), \end{aligned}\right\} \qquad (54)$$

valid near $y = 0$, we find that equation (39) becomes

$$y^{\frac{1}{2}}(\tfrac{1}{6} - \tfrac{3}{8}a - \tfrac{1}{2}\log y) = y^{\frac{1}{2}}\left(\frac{1}{R} + \int_y^{y_0} \frac{dy}{x^2\sqrt{g}}\right) + O(y\log^2 y). \qquad (55)$$

† This behaviour applies only to the *left* of y_0. Solutions intersecting the y-axis from the right have an entirely different behaviour, namely,

$$g(y) = [6y_0(y-y_0)]^{\frac{1}{2}} + b(y-y_0) + O([y-y_0]^{\frac{3}{2}}) \quad (y\to y_0;\ y>y_0),$$

where b is an arbitrary constant. In consequence of this very singular behaviour of the solutions at y_0, the series expansion (50) is very poorly convergent and does not represent the solution even when y is only very slightly different from y_0.

Now letting $y \to 0$, we obtain

$$\frac{1}{R} = \tfrac{1}{6} - \tfrac{3}{8}a - \lim_{y \to 0}\left(\int_y^{y_0} \frac{dy}{x^2\sqrt{g}} + \tfrac{1}{2}\log y\right). \tag{56}$$

The limit on the right-hand side exists, the logarithmic divergence of the integral as $y \to 0$ exactly cancelling the term $\frac{1}{2}\log y$. We may accordingly write

$$\frac{1}{R} = \tfrac{1}{6} - \tfrac{3}{8}a - \tfrac{1}{2}\log y_0 + \int_0^{y_0}\left(\frac{1}{2y} - \frac{1}{x^2\sqrt{g}}\right)dy. \tag{57}$$

This is the required relation between a and R.

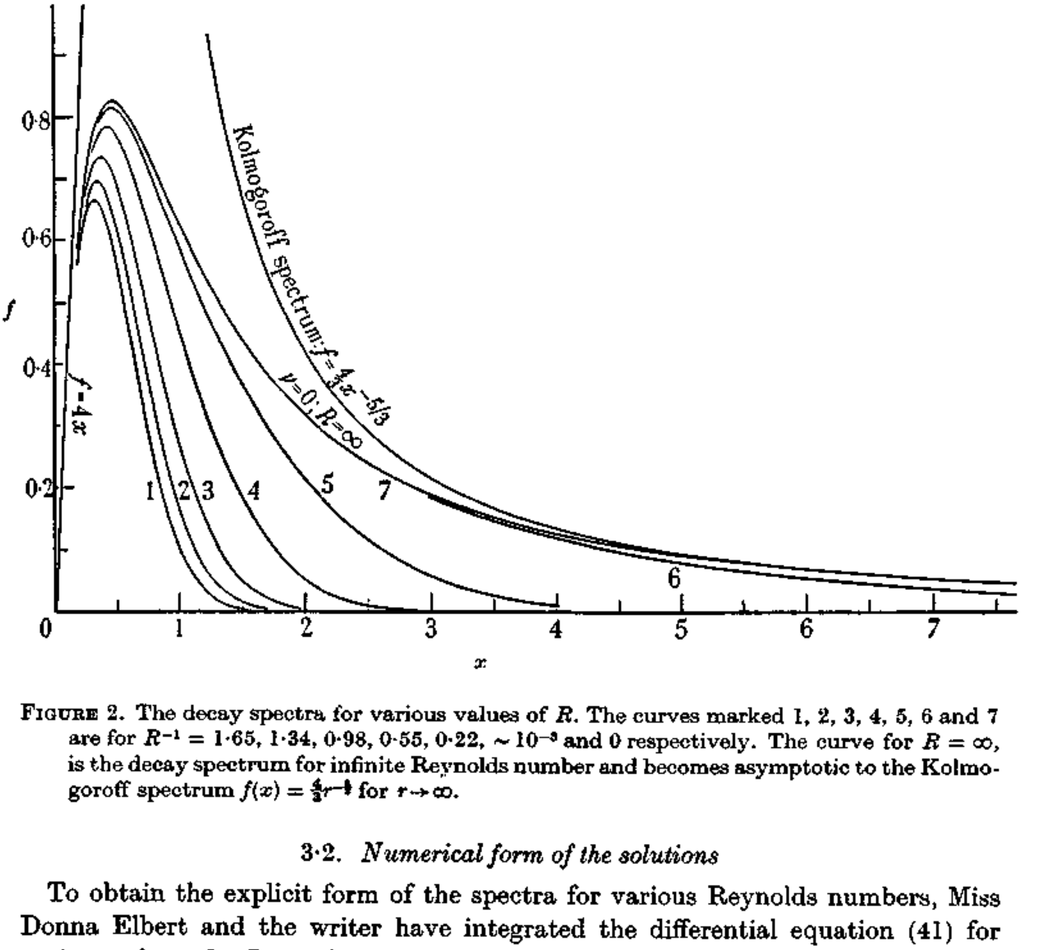


FIGURE 2. The decay spectra for various values of R. The curves marked 1, 2, 3, 4, 5, 6 and 7 are for $R^{-1} = 1{\cdot}65$, $1{\cdot}34$, $0{\cdot}98$, $0{\cdot}55$, $0{\cdot}22$, $\sim 10^{-3}$ and 0 respectively. The curve for $R = \infty$, is the decay spectrum for infinite Reynolds number and becomes asymptotic to the Kolmogoroff spectrum $f(x) = \frac{4}{3}r^{-\frac{5}{3}}$ for $r \to \infty$.

3·2. *Numerical form of the solutions*

To obtain the explicit form of the spectra for various Reynolds numbers, Miss Donna Elbert and the writer have integrated the differential equation (41) for various values of a. It was found that $a = 1{\cdot}8104\ldots$ leads to the critical solution which approaches the line $g = \frac{4}{3}y$ for $y \to \infty$.

With g determined by numerical integration, a further quadrature according to equation (44) gave the relation between x and y. The spectra $f(x)$ determined in this fashion are given in table 1 and are further illustrated in figure 2. The values of R determined according to equation (57) for the different solutions are also included in table 1.

TABLE 1. THE SPECTRUM OF TURBULENCE DURING DECAY

$a = 1{\cdot}8104739;\ R^{-1} = 0$

x	f	x	f	x	f	x	f
0·186	0·5854	0·450	0·8228	1·735	0·3755	4·365	0·1097
0·224	0·6558	0·456	0·8232	1·905	0·3369	4·460	0·1063
0·250	0·6956	0·462	0·8236	2·064	0·3058	4·554	0·1030
0·271	0·7224	0·489	0·8233	2·213	0·2803	4·647	0·1000
0·288	0·7421	0·535	0·8179	2·355	0·2589	4·739	0·0971
0·303	0·7569	0·575	0·8091	2·492	0·2406	4·830	0·0944
0·317	0·7686	0·611	0·7989	2·624	0·2247	4·921	0·0919
0·329	0·7780	0·643	0·7881	2·751	0·2109	5·010	0·0894
0·341	0·7859	0·672	0·7770	2·874	0·1987	5·099	0·0871
0·351	0·7922	0·700	0·7659	2·995	0·1878	5·187	0·0849
0·361	0·7977	0·726	0·7551	3·112	0·1781	5·274	0·0828
0·371	0·8022	0·817	0·7143	3·227	0·1693	5·361	0·0808
0·379	0·8061	0·895	0·6780	3·339	0·1614	5·446	0·0789
0·388	0·8091	0·965	0·6458	3·449	0·1541	5·532	0·0771
0·396	0·8121	1·029	0·6171	3·557	0·1475	5·616	0·0753
0·404	0·8144	1·088	0·5912	3·663	0·1415	5·700	0·0737
0·411	0·8164	1·144	0·5679	3·768	0·1359	5·825	0·0713
0·418	0·8180	1·196	0·5466	3·871	0·1307	5·989	0·0684
0·425	0·8194	1·246	0·5271	3·972	0·1259	6·151	0·0657
0·432	0·8205	1·294	0·5092	4·072	0·1214	6·311	0·0632
0·438	0·8215	1·340	0·4927	4·171	0·1172	6·469	0·0608
0·444	0·8222	1·550	0·4253	4·268	0·1134	6·625	0·0586

$a = 1{\cdot}81;\ R^{-1} \sim 10^{-3}$

x	f	x	f	x	f	x	f
0·186	0·5854	0·450	0·8227	1·738	0·3728	4·909	0·0818
0·224	0·6558	0·456	0·8230	1·908	0·3337	5·212	0·0730
0·250	0·6956	0·462	0·8234	2·068	0·3023	5·514	0·0653
0·271	0·7224	0·489	0·8231	2·218	0·2763	5·817	0·0584
0·288	0·7421	0·535	0·8176	2·362	0·2544	6·122	0·0522
0·303	0·7568	0·575	0·8088	2·501	0·2356	6·431	0·0466
0·317	0·7686	0·611	0·7986	2·634	0·2194	6·746	0·0414
0·329	0·7780	0·643	0·7876	2·764	0·2051	7·069	0·0367
0·341	0·7858	0·672	0·7766	2·890	0·1925	7·402	0·0323
0·351	0·7921	0·700	0·7655	3·013	0·1812	7·748	0·0282
0·361	0·7976	0·726	0·7546	3·133	0·1711	8·112	0·0244
0·371	0·8021	0·817	0·7136	3·251	0·1619	8·497	0·0209
0·379	0·8060	0·896	0·6771	3·367	0·1536	8·910	0·0176
0·388	0·8092	0·966	0·6448	3·481	0·1459	9·360	0·0145
0·396	0·8120	1·029	0·6159	3·593	0·1389	9·860	0·0116
0·404	0·8142	1·088	0·5900	3·704	0·1324	10·430	0·0090
0·411	0·8163	1·144	0·5666	3·813	0·1264	11·103	0·0066
0·418	0·8178	1·197	0·5452	3·922	0·1208	11·942	0·0044
0·425	0·8194	1·247	0·5256	4·029	0·1155	13·085	0·0025
0·432	0·8204	1·295	0·5076	4·135	0·1106	14·935	0·0011
0·438	0·8214	1·341	0·4909	4·293	0·1039	—	—
0·444	0·8220	1·551	0·4230	4·603	0·0920	—	—

TABLE 1 (*cont.*)

$a = 1{\cdot}75$; $R^{-1} = 0{\cdot}22$

x	f	x	f	x	f	x	f
0·186	0·5846	0·457	0·8142	0·924	0·6296	1·675	0·3053
0·224	0·6545	0·463	0·8143	0·946	0·6180	1·731	0·2877
0·250	0·6954	0·480	0·8136	0·967	0·6068	1·786	0·2710
0·271	0·7201	0·500	0·8116	0·988	0·5959	1·842	0·2551
0·288	0·7394	0·519	0·8086	1·008	0·5853	1·897	0·2400
0·304	0·7537	0·545	0·8029	1·028	0·5751	1·953	0·2255
0·317	0·7651	0·570	0·7963	1·048	0·5652	2·009	0·2116
0·330	0·7741	0·592	0·7891	1·067	0·5555	2·066	0·1982
0·341	0·7816	0·613	0·7816	1·086	0·5461	2·151	0·1792
0·352	0·7876	0·633	0·7739	1·104	0·5370	2·239	0·1612
0·362	0·7927	0·652	0·7662	1·129	0·5251	2·330	0·1442
0·371	0·7968	0·676	0·7558	1·152	0·5137	2·424	0·1280
0·380	0·8004	0·698	0·7454	1·193	0·4946	2·522	0·1127
0·388	0·8033	0·720	0·7352	1·226	0·4790	2·625	0·0981
0·396	0·8058	0·740	0·7250	1·259	0·4642	2·735	0·0842
0·404	0·8078	0·760	0·7152	1·292	0·4499	2·854	0·0710
0·412	0·8095	0·779	0·7056	1·323	0·4363	2·985	0·0586
0·419	0·8108	0·797	0·6961	1·354	0·4232	3·132	0·0468
0·426	0·8120	0·815	0·6868	1·385	0·4106	3·302	0·0357
0·432	0·8127	0·832	0·6778	1·445	0·3866	3·508	0·0254
0·439	0·8134	0·853	0·6668	1·504	0·3644	3·775	0·0160
0·445	0·8138	0·878	0·6539	1·562	0·3435	4·177	0·0078
0·451	0·8141	0·901	0·6415	1·619	0·3238	—	—

$a = 1{\cdot}5$; $R^{-1} = 0{\cdot}55$

x	f	x	f	x	f	x	f
0·186	0·5812	0·461	0·7797	0·897	0·5288	1·503	0·1817
0·224	0·6488	0·467	0·7788	0·925	0·5091	1·532	0·1702
0·251	0·6862	0·484	0·7750	0·952	0·4901	1·563	0·1590
0·272	0·7107	0·510	0·7690	0·979	0·4714	1·594	0·1480
0·289	0·7284	0·534	0·7586	1·005	0·4533	1·627	0·1373
0·305	0·7412	0·561	0·7472	1·031	0·4357	1·660	0·1267
0·318	0·7511	0·585	0·7352	1·057	0·4186	1·695	0·1164
0·331	0·7586	0·609	0·7228	1·083	0·4020	1·732	0·1064
0·343	0·7647	0·630	0·7104	1·108	0·3857	1·771	0·0965
0·353	0·7693	0·651	0·6980	1·133	0·3699	1·812	0·0869
0·363	0·7732	0·671	0·6856	1·158	0·3544	1·855	0·0775
0·373	0·7760	0·690	0·6733	1·184	0·3393	1·902	0·0683
0·382	0·7783	0·708	0·6612	1·209	0·3245	1·952	0·0594
0·391	0·7799	0·726	0·6492	1·234	0·3101	2·008	0·0508
0·399	0·7812	0·744	0·6373	1·260	0·2960	2·070	0·0425
0·407	0·7820	0·760	0·6257	1·286	0·2822	2·141	0·0344
0·414	0·7826	0·777	0·6143	1·311	0·2687	2·224	0·0267
0·422	0·7827	0·793	0·6030	1·337	0·2555	2·325	0·0194
0·429	0·7828	0·808	0·5918	1·364	0·2425	2·459	0·0126
0·436	0·7824	0·824	0·5810	1·391	0·2299	2·658	0·0065
0·442	0·7821	0·839	0·5702	1·418	0·2174	2·890	0·0030
0·449	0·7813	0·854	0·5596	1·446	0·2053	—	—
0·455	0·7807	0·868	0·5492	1·474	0·1933	—	—

TABLE 1 (*cont.*)

$a = 1{\cdot}0;\ R^{-1} = 0{\cdot}98$

x	f	x	f	x	f	x	f
0·186	0·5744	0·469	0·7108	0·839	0·4080	1·182	0·1606
0·225	0·6375	0·475	0·7078	0·853	0·3957	1·201	0·1514
0·252	0·6710	0·494	0·6981	0·867	0·3836	1·219	0·1423
0·273	0·6920	0·516	0·6844	0·880	0·3716	1·238	0·1334
0·291	0·7065	0·537	0·6703	0·894	0·3598	1·258	0·1246
0·307	0·7160	0·558	0·6558	0·908	0·3481	1·279	0·1159
0·321	0·7229	0·577	0·6412	0·922	0·3365	1·301	0·1073
0·334	0·7276	0·595	0·6264	0·936	0·3251	1·323	0·0989
0·346	0·7309	0·613	0·6118	0·950	0·3138	1·347	0·0906
0·357	0·7328	0·630	0·5972	0·964	0·3027	1·372	0·0825
0·367	0·7340	0·647	0·5827	0·979	0·2917	1·398	0·0745
0·377	0·7343	0·663	0·5683	0·993	0·2808	1·427	0·0666
0·386	0·7341	0·679	0·5540	1·008	0·2701	1·458	0·0589
0·395	0·7332	0·694	0·5399	1·022	0·2595	1·491	0·0514
0·404	0·7322	0·709	0·5260	1·037	0·2490	1·528	0·0441
0·412	0·7306	0·724	0·5122	1·052	0·2387	1·569	0·0370
0·420	0·7289	0·739	0·4986	1·067	0·2285	1·616	0·0300
0·428	0·7268	0·754	0·4852	1·083	0·2184	1·672	0·0234
0·435	0·7246	0·768	0·4719	1·099	0·2085	1·741	0·0170
0·442	0·7220	0·783	0·4588	1·115	0·1986	1·832	0·0110
0·449	0·7196	0·797	0·4459	1·131	0·1890	1·972	0·0055
0·456	0·7167	0·811	0·4331	1·148	0·1794	—	—
0·463	0·7139	0·825	0·4205	1·165	0·1699	—	—

$a = 0{\cdot}5;\ R^{-1} = 1{\cdot}34$

x	f	x	f	x	f
0·187	0·5677	0·504	0·6214	0·881	0·2565
0·226	0·6264	0·523	0·6058	0·896	0·2439
0·253	0·6560	0·541	0·5898	0·912	0·2316
0·275	0·6725	0·558	0·5740	0·928	0·2194
0·293	0·6832	0·574	0·5580	0·944	0·2073
0·309	0·6896	0·590	0·5422	0·960	0·1954
0·323	0·6936	0·606	0·5264	0·977	0·1838
0·337	0·6953	0·621	0·5108	0·994	0·1722
0·349	0·6959	0·636	0·4952	1·012	0·1608
0·360	0·6952	0·651	0·4798	1·031	0·1496
0·371	0·6938	0·666	0·4646	1·050	0·1386
0·381	0·6916	0·680	0·4496	1·070	0·1277
0·391	0·6890	0·695	0·4346	1·091	0·1171
0·400	0·6858	0·709	0·4200	1·113	0·1065
0·409	0·6824	0·723	0·4053	1·137	0·0963
0·418	0·6785	0·737	0·3910	1·162	0·0861
0·426	0·6746	0·751	0·3767	1·188	0·0762
0·434	0·6702	0·766	0·3628	1·217	0·0665
0·442	0·6659	0·780	0·3488	1·249	0·0570
0·450	0·6612	0·794	0·3352	1·285	0·0478
0·457	0·6567	0·808	0·3216	1·326	0·0388
0·464	0·6517	0·822	0·3083	1·374	0·0302
0·471	0·6469	0·837	0·2950	1·433	0·0219
0·478	0·6419	0·851	0·2821	1·512	0·0140
0·485	0·6368	0·866	0·2691	1·634	0·0068

TABLE 1 (*cont.*)

$a = 0$; $R^{-1} = 1{\cdot}65$

x	f	x	f	x	f
0·187	0·5608	0·509	0·5555	0·816	0·2312
0·227	0·6151	0·522	0·5414	0·829	0·2198
0·254	0·6410	0·536	0·5274	0·843	0·2086
0·276	0·6551	0·549	0·5134	0·856	0·1975
0·295	0·6625	0·562	0·4994	0·870	0·1865
0·311	0·6658	0·574	0·4855	0·885	0·1756
0·326	0·6669	0·587	0·4717	0·899	0·1650
0·339	0·6658	0·599	0·4580	0·915	0·1544
0·352	0·6637	0·611	0·4443	0·930	0·1440
0·364	0·6603	0·623	0·4307	0·947	0·1337
0·375	0·6564	0·635	0·4173	0·964	0·1236
0·385	0·6518	0·647	0·4040	0·982	0·1136
0·396	0·6468	0·659	0·3909	1·000	0·1038
0·405	0·6412	0·670	0·3778	1·020	0·0941
0·415	0·6356	0·682	0·3649	1·041	0·0846
0·424	0·6294	0·694	0·3521	1·064	0·0753
0·432	0·6233	0·706	0·3394	1·088	0·0661
0·441	0·6168	0·718	0·3268	1·115	0·0572
0·449	0·6104	0·730	0·3144	1·145	0·0484
0·457	0·6036	0·742	0·3022	1·179	0·0399
0·465	0·5970	0·754	0·2900	1·218	0·0316
0·472	0·5901	0·766	0·2780	1·265	0·0237
0·480	0·5833	0·778	0·2661	1·326	0·0160
0·487	0·5763	0·791	0·2543	1·414	0·0089
0·495	0·5695	0·803	0·2427	—	—

In conclusion, I wish to record my indebtedness to Miss Donna Elbert for valuable assistance with the various numerical integrations involved in the preparation of this paper.

REFERENCES

Chandrasekhar, S. 1949 *Phys. Rev.* **75**, 896, 1454; **76**, 158, 1949.
Heisenberg, W. 1948*a* *Z. Phys.* **124**, 628.
Heisenberg, W. 1948*b* *Proc. Roy. Soc.* A, **195**, 402.

TURBULENCE—A PHYSICAL THEORY OF ASTROPHYSICAL INTEREST*

S. CHANDRASEKHAR
Yerkes Observatory, University of Chicago
Received August 4, 1949

May I say at the outset that I have found myself deeply inadequate for the task of giving this third Henry Norris Russell Lecture. I am afraid that I have not discovered or paved a Royal Road that I can describe to you in the manner of Dr. Russell; neither have I the excellence of the material which Dr. Adams presented in his second Henry Norris Russell Lecture. And I am aware that no general interest attaches to matters in which I may claim some degree of competence. I have therefore chosen, after considerable hesitation, to describe to you the recent advances in our understanding of the phenomenon of turbulence, in the belief that these advances are relevant to the progress of astrophysics. Perhaps it is premature to take an occasion like this to describe a physical theory which has yet to establish its relations to astronomical developments. But the history of astronomy and astrophysics shows that major advances in our understanding of astrophysical phenomena have coincided with and depended upon advances in fundamental physical theory. While many examples illustrating this can be given, there is none more conspicuous or notable in recent history than that provided by the work of Henry Norris Russell; thus, during the great period in which the foundations both of atomic spectra and of stellar spectroscopy were laid, Russell was a great exponent of both subjects. As is well known, the main features of the theory of complex spectra emerged for the first time from the pioneering investigations of Drs. Russell and Saunders on the alkaline earths. The main conclusion of these investigations, stated by the authors in the words "both valency electrons may jump at once from outer to inner orbits, while the net energy lost is radiated as a single quantum," has since been incorporated into the analysis of stellar spectra as the "Russell-Saunders" coupling and is one of the keystones of atomic theory. In these early papers of Dr. Russell all the steps preliminary to the formulation of the exclusion principle were taken, and I do not believe that it is a misstatement of history to say that the honor of the discovery of the exclusion principle would have gone to Russell had his concern with the applications of the principles of atomic spectra to astrophyiscal problems been a little less. However that may be, as astronomers we may count ourselves fortunate that Russell's concern with astrophysical problems was as earnest then as it has always been, for otherwise, we should not have had so immediate or so complete an integration of physical and astrophysical theories as was, in fact, achieved when Russell's great work on the quantitative analysis of the solar spectrum and the first determination of the composition of the sun appeared in 1929. I have referred to this example of Russell's work to emphasize the interdependence of physical and astrophysical thought. And, as I have stated, it seems to me probable that the recent advances in the physics of turbulence, due in large measure to G. I. Taylor, von Karman, Kolmogoroff, and Heisenberg, may play an important part in the future developments of astrophysics. But, before I describe the nature of these advances in physical theory, I may perhaps indicate briefly the astrophysical contexts in which they may find their most fruitful applications.

The first person clearly to draw attention to the importance for astrophysics of tur-

* Third Henry Norris Russell Lecture of the American Astronomical Society delivered at Ottawa, Canada, on June 21, 1949.

bulence with its correct hydrodynamical meaning was Rosseland. In a paper published in 1928, Rosseland[1] pointed out that if differential motions—i.e., motions of one part relative to another—occur in cosmical gas masses, then the motions should be turbulent in the sense that we should not expect to describe them in terms of the classical equations of motion of Stokes and Navier. In drawing this inference, Rosseland was guided by the experience in meteorology and oceanography and by the following reasoning.

We are all familiar with the fact that a linear flow of water in a tube can be obtained only for velocities below a certain critical limit and that, when the velocity exceeds this limit, laminar flow ceases and a complex, irregular, and fluctuating motion sets in. More generally than in this context of flow through a tube, it is known that motions governed by the equations of Stokes and Navier change into turbulent motion when a certain nondimensional constant called the "Reynolds number" exceeds a certain value of the order of 1000. This Reynolds number depends upon the linear dimension, L, of the system, the coefficient of viscosity μ, the density ρ, and the velocity v in the following manner:

$$R = \frac{\rho v L}{\mu}. \tag{1}$$

Since R depends directly on the linear dimension of the system, Rosseland argued that motions in the oceans, in terrestrial and planetary atmospheres, and still more in stellar atmospheres, once they occur, must become turbulent in this sense. Rosseland further pointed out that, if turbulence develops, the coefficients of viscosity and heat conduction may be expected to increase a million fold. And the importance of this enhanced efficiency of heat and momentum transport in a turbulent medium cannot be exaggerated.

Stimulated by Rosseland's ideas, McCrea[2] suggested in the same year that the solar chromosphere must be in a state of turbulence and that this turbulence may, in part, contribute to its support against gravity.

About a year later Harold Jeffreys[3] drew attention to a fact which had been ignored until then, namely, that if the generation of energy inside stars is confined to a small region at the center, then the radiation will not be able to dispose of it at a gradient under the adiabatic and that, if a superadiabatic gradient comes into being, vertical currents will be generated which will effectively restore the adiabatic gradient, leaving, however, a slight superadiabatic gradient to make possible the transport of heat. The condition for the occurrence of such convective transport of heat can be written down, and it follows from this condition that even a relatively mild concentration of the energy sources toward the center will lead to its occurrence near the center. Indeed, with the clarification of the source of stellar energy as due to nuclear transformations, it is now generally recognized that all stars must have convective cores in which turbulence prevails. And, as was shown, particularly by Cowling,[4] the existence of turbulence is of primary importance in all considerations relating to the stability of stars.

Returning to the role of turbulence in the atmospheres of the stars, we next observe that the investigations of Struve and Elvey[5] established the occurrence of large-scale motions in the atmospheres of stars like 17 Leporis, ϵ Aurigae, and α Persei. In investigations, which, it may be noted, were also the first to apply the then new method of the curve of growth to the analysis of stellar atmospheres—the method had already been applied to the solar atmosphere by Minnaert—Struve and Elvey showed that the linear portion of the curve of growth, as well as the line profiles themselves, cannot be explained in terms of the Doppler effect due to thermal motions alone and that large-scale

[1] *M.N.*, **89**, 49, 1929.

[2] *M.N.*, **89**, 718, 1929.

[3] *Nature*, **127**, 162, 1931; also *M.N.*, **91**, 121, 1931.

[4] *M.N.*, **94**, 768, 1934; **96**, 42, 1935.

[5] *Ap. J.*, **79**, 409, 1934; see also O. Struve, *Proc. Nat. Acad. Sci*, **18**, 585, 1932.

motions of a turbulent nature must be postulated. This conclusion has since been confirmed and extended by various other investigators.

That turbulence must play a part also in the solar atmosphere became clear after Unsöld[6] had shown that in the deeper layers of the solar photosphere, where hydrogen begins to get ionized, the radiative gradient must become unstable. Since that time the view first advanced by Siedentopf[7] and Biermann,[8] that the solar granulation must, in some way, be related to this hydrogen convection zone, has been steadily gaining ground.

Again the investigations of Struve and his associates during the past few years have shown that the shells surrounding early-type stars and the gaseous envelopes in which spectroscopic binaries are frequently imbedded must also be turbulent, the turbulence in these contexts arising, in the first instance, from the different parts of the shell or medium rotating with different angular velocities.

And, finally, it would appear that the interstellar clouds must also be in a state of turbulence; for, assuming that a typical cloud is 10 parsecs in diameter and that relative motions to the extent of 10 km/sec occur, we find that the Reynolds number must be of the order of 10^5; and the motions inside the cloud must therefore be turbulent. The even larger question now occurs whether we may not indeed regard the clouds of various dimensions in interstellar space as eddies in a medium occupying the whole of galactic space.

From this brief survey of the various problems in which turbulence may play a role, it would almost appear that, if we are in the mood for it, we may encounter turbulence no matter where we turn. But what is the picture of turbulence in terms of which we wish to interpret such a wide diversity of phenomena? It is that in a turbulent medium there are eddies which spontaneously form and disintegrate; that this process goes on continuously; that each eddy travels a certain average distance with a certain average speed before it loses its identity—a specific enough picture but not one derived from, or justified by, a physical theory. Thus, while the basic concepts of "mean free path" and "root-mean-square velocity" which underlie the picture are plausible enough, it was not known how these quantities were to be related with the physical conditions of the problem. Indeed, from the point of view of a rational physical theory, the situation has been so unsatisfactory that, in a recent conversation, Dr. Russell recalled that E. W. Brown, referring to the frequency with which appeals were being made to the action of a resisting medium to account for this or that anomaly in the motions of celestial bodies, once remarked: "What fifty years ago used to be attributed to the direct intervention of the Deity are now being attributed to a resisting medium." Dr. Russell added that he sometimes felt the same way about the frequency with which turbulence is currently being invoked to account for astrophysical phenomena. Nevertheless, it would seem that the application of the newer developments in the theory of turbulence may help to remove this element of the miraculous in astrophysics.

As I indicated at the outset, the study of turbulence in hydrodynamics started with investigations on the stability of laminar flow. In general, these investigations began with simple patterns of flow, like axial flow through a tube or plane-parallel flow over an infinite plate, and examined the stability of these flows to perturbations of particular types with a view to determining the critical value of the Reynolds number at which laminar flow becomes unstable. The mathematical analysis required for the investigation of stability along these lines is of a very treacherous kind, and, in spite of the enormous effort which has been expended on this problem by various authors, including Heisenberg, Tollmien, Lin, and Pekeris, no positive or general conclusions seem to have been reached. However, as Heisenberg has recently emphasized,[9] investigations of sta-

[6] *Zs. f. Ap.*, **1**, 138, and **2**, 209, 1931.

[7] *A.N.*, **247**, 297, 1933; **249**, 53, 1933; **255**, 157, 1935.

[8] *Zs. f. Ap.*, **22**, 244, 1943.

[9] *Zs. f. Naturforsch.*, **3**, 434, 1948.

bility along these lines, even if successful, cannot, in principle, lead to an understanding of the phenomenon of turbulence itself; for the basic problem of turbulence is of an entirely different character. That this is the case becomes apparent when we ask ourselves the very elementary question, "What is the reason that a phenomenon like turbulence can occur at all?" The answer must be that an ideal fluid is a mechanical system with a very large number of degrees of freedom and that, in consequence, it is theoretically capable of a very large number of different types of motions. Laminar motion is only one of the many possible motions that the system is capable of, and to expect that it will always be realized is as futile as to expect that in a gas we shall find all the molecules moving with the same velocity parallel to one another. It is far more likely that all the possible motions will be simultaneously present. The fundamental problem of turbulence would therefore appear to be a statistical one of specifying the probability with which the various types of motion may occur and are present. Stated in this way, it is clear that the problem of turbulence has an analogy with the problem of analyzing a continuous spectrum of radiation. In the latter case, the greatest interest is generally attached to the distribution of intensity in the spectrum and only secondarily to the phase relationships. Similarly, when we consider the motions in a turbulent fluid, we may make a harmonic analysis of the instantaneous velocity field $v\ (r,\ t)$ in the form

$$v\ (r,\ t) = \sum_{k} v_k\,(t)\ e^{ik\cdot r} \tag{2}$$

and ask for the average energy stored in the various wave lengths. We can visualize this formal procedure in the following manner.

Considering the state of motion at a given instant, we may analyze the fluctuating velocity field as the result of superposition of periodic variations with all possible wave lengths. We may picture the component with a wave length λ as corresponding to an eddy of size λ, and, since many wave lengths are needed to represent a general velocity field, we may speak of a "hierarchy of eddies." This hierarchy of eddies will be limited on the side of long wave lengths by the fact that no eddy of size larger than the dimension of the medium in which we analyze the turbulence can occur.

Instead of the wave length λ, it is often more convenient to speak of a wave number $k = 2\pi/\lambda$.

Analyzing the motion into eddies in this manner, we can ask: What is the energy per unit volume stored in eddies with wave numbers between k and $k + dk$? If $\rho F(k)dk$ denotes this energy, $F(k)$ is said to define the *spectrum* of turbulence.

It can be shown that most of the interesting features of turbulent motion can be deduced from its spectrum. For example, the correlations $\overline{u_i u_j'}$ between the instantaneous velocity components u_i and u_j' at two different points of the medium can be expressed simply in terms of the spectrum. Such correlations were first introduced by G. I. Taylor[10] as a basis for a phenomenological theory of turbulence, and they have since been studied extensively, both theoretically and experimentally. It is therefore natural that the question of the spectrum of turbulence should be in the forefront in all recent discussions of turbulence, since most of the available experimental data on the subject are capable of being interpreted in terms of the spectrum.

Now, returning to the optical analogy I referred to earlier, we know that under conditions of equilibrium the distribution of energy in the continuous spectrum will be that given by Planck's law. We may ask whether a similar equilibrium spectrum exists for turbulence. In answering this question, we must keep in mind one important distinction between the optical analogue and turbulence. In the optical case the equilibrium Planck

[10] *Proc. R. Soc. London, A*, **164**, 476, 1938. For a general account of these investigations see H. L. Dryden, *Quar. Appl. Math.*, 1, 7, 1943.

spectrum will be reached, no matter what the initial distribution is. In contrast, turbulence can be maintained only by an external agency, like continuous stirring, the energy available from thermal instability, or rotation in a differentially rotating atmosphere. In other words, energy is required for the maintenance of turbulence; in the absence of such an agency, turbulence will *decay*, and the spectrum will be a function of time. In discussing the spectrum of turbulence, we must therefore distinguish between two cases: the case in which the agency maintaining turbulence is communicating energy to the medium at a constant rate and a stationary condition prevails and the case in which there is no external agency maintaining turbulence and the turbulence, in consequence, is decaying.

In the stationary case it is clear that energy must be dissipated in the form of thermal energy at the same rate at which energy is being supplied.

According to the laws of hydrodynamics, the rate of dissipation of energy by viscosity is given by

$$\epsilon = \rho\nu\overline{|\mathrm{curl}\, v\,|^2} \tag{3}$$

per unit volume. In equation (3) ν is the *kinematic viscosity* and is μ/ρ. In terms of the spectrum this expression for ϵ becomes

$$\epsilon = 2\rho\nu\int_0^\infty F(k)\, k^2 dk\,. \tag{4}$$

Under stationary conditions this must be the rate at which energy is being communicated to the medium by the external agency.

Considering, now, the problem of determining the spectrum of turbulence under stationary conditions, we may first remark that the presence of an external agency maintaining turbulence requires us to distinguish between the region of the spectrum in which the eddy sizes are comparable to the linear dimension —l_0, say—of the system and the region of the spectrum in which the eddy sizes are small compared to the linear dimension of the system. In the first region the nature of the spectrum must clearly depend on the external agency. We should not, therefore, expect to give a theory of the spectrum in this region which will be universally valid. Each situation will have to be analyzed separately. On the other hand, it does not seem unreasonable to suppose that the distribution of energy among the eddies which are small compared to the dimension of the system will be largely independent of the particular mechanism maintaining the turbulence and will depend only on the rate ϵ at which energy is being supplied. In terms of wave numbers we may express this in the following way.

Let $k_0 \sim 1/l_0$ denote the wave number of the largest eddies present. Then for $k \gg k_0$ we may expect the spectrum to approach a universal one, depending only on ϵ and ν. We may further expect that, as the Reynolds number tends to infinity, more and more of the spectrum will follow a universal law. When this is the case, we say that we have the equilibrium spectrum for a fully developed turbulence. In astronomical contexts turbulence, when it occurs, may be expected to be fully developed in this sense.

Turning, now, to the specification of the spectrum, we may suppose that the energy supplied by the external agency is communicated principally to the largest eddies, i.e., for $k \sim k_0$. Let ϵ denote this rate. As we have seen, energy is being dissipated by viscosity at this same rate; it is evident that this dissipation into thermal energy will be effected principally by the smallest eddies, in which the motions may be expected to be laminar. Consequently, energy at this same rate must flow through the entire hierarchy of eddies, and the equilibrium spectrum will be determined by this condition of constant flow of energy through the hierarchy. To translate this condition of constant flow of energy through the hierarchy into a quantitative expression, we consider the rate ϵ_k at which energy flows from eddies of all wave numbers less than a particular k to eddies of all wave numbers greater than this k. In a general way it is clear that we must distinguish between two different types of contributions to ϵ_k: First, there is the dissipation directly into thermal energy:

$$\epsilon_k\,(\text{thermal}) = 2\,\rho\nu \int_0^k F(k)\,k^2 dk\,. \tag{5}$$

Then there is the energy communicated to the eddies of smaller sizes in the form of kinetic energy of motion. We shall give an expression for this later, but we may note meantime that, for any given k, the relative importance of the two contributions will depend on k and the Reynolds number of the entire motion. If k_s denotes the wave number of the eddies in which the motion begins to be laminar, then we should expect that, for $k \gg k_s$, the transfer of energy into the kinetic energy of motion will be negligible. On the other hand, if the Reynolds number is sufficiently large, then in a significant portion of the spectrum the inequality $k_0 \ll k \ll k_s$ will be valid; this inequality means that there exists a range of sizes which is small compared to the largest eddies present but large compared to the eddies in which the dissipation by kinetic viscosity occurs; and this will certainly happen if we let the Reynolds number tend to infinity. Now let the Reynolds number be sufficiently large for the inequality $k_0 \ll k \ll k_s$ to be valid over a portion of the spectrum. In this portion of the spectrum, in contrast to the portion of the spectrum where $k \gg k_s$, the thermal contribution to ϵ_k must be negligible; when this is the case, the spectrum may be expected to become independent of the viscosity as well and depend only on ϵ. These ideas, which underlie the recent developments in the theory of turbulence, were first clearly recognized by L. F. Richardson, to whom the following rhyme is attributed:

Big whirls have little whirls,
That feed on their velocity;
And little whirls have lesser whirls,
And so on to viscosity.

However, mathematical expression was first given to these ideas by Kolmogoroff,[11] in the form of two principles. In our present context we may state the principles of Kolmogoroff in the following form:

1. The spectrum of turbulence for all k much greater than a certain k_0 must be determined uniquely by $e = \epsilon/\rho$ and the kinematic viscosity $\nu = \mu/\rho$.

2. For infinite Reynolds number the spectrum must, in addition, become independent of ν and depend only on e.

We shall see that in these forms the principles are valid also for the problem of decay.

We shall now show how the two principles of Kolmogoroff enable us to determine the form of the spectrum in the region $k_0 \ll k \ll k_s$.

Now $F(k)$ is of dimension (velocity)^2 $\times$ length, while k itself is of dimension (length)$^{-1}$. Quantities of these dimensions which can be constructed out of e and ν are

$$(\text{Velocity})^2 \times \text{length} = [(\nu^5 e)^{1/4}]\,,$$

$$\text{Length} = \left[\left(\frac{\nu^3}{e}\right)^{1/4}\right]. \tag{6}$$

Consequently, Kolmogoroff's first principle requires that

$$F(k) \equiv (\nu^5 e)^{1/4} f\,(k\nu^{3/4} e^{-1/4})\,, \tag{7}$$

where f is a universal function of the argument specified. According to the second principle, $F(k)$ should be independent of ν in the region $k_0 \ll k \ll k_s$. Accordingly, in this region $f(x)$ must be of the form

$$f(x) \doteq C x^{-5/3}\,, \tag{8}$$

[11] 1941, *C.R. Acad. Sci. U.S.S.R.*, **30**, 301, and **32**, 16, 1942; see also G. K. Batchelor, *Proc. Cambridge Phil. Soc.*, **43**, 533, 1947.

where C is a constant; for then

$$F(k) \doteq C(\nu^5 e)^{1/4}\left(\frac{\nu^3}{e}\right)^{-5/12} k^{-5/3} \tag{9}$$
$$= C e^{2/3} k^{-5/3} ;$$

and the requirement that $F(k)$ be independent of ν is satisfied. Hence, when the Reynolds number tends to infinity, the spectrum will follow more and more closely a $k^{-5/3}$-law—this is the *Kolmogoroff spectrum.* I should perhaps mention at this stage that the $k^{-5/3}$-law was discovered independently, also by Onsager[12] and von Weizsäcker,[13] but several years later.

It is sometimes convenient to think of all eddies with wave numbers exceeding a certain k (i.e., with wave lengths less than a certain λ) as having a certain mean velocity, v_k. For this purpose we may adopt as definition the equation

$$v_k^2 = \int_k^\infty F(k)\, dk . \tag{10}$$

When the equilibrium Kolmogoroff spectrum prevails, this equation gives the law

$$v_k \propto k^{-1/3} \propto \lambda^{1/3} . \tag{11}$$

While Kolmogoroff's method of determining the form of the equilibrium spectrum of fully developed turbulence is very elegant, it does not, one must admit, give any real insight into the physical nature of turbulence. Also, even under equilibrium conditions, it does not give the part of the spectrum in which the dissipation by viscosity begins to be an important factor. An elementary theory which visualizes clearly the phenomenon of turbulence and which gives, at the same time, the complete equilibrium spectrum is due to Heisenberg.[14] The ideas underlying Heisenberg's development can be explained very simply.

Considering the rate at which eddies with wave numbers between 0 and k transfer energy to eddies with wave numbers exceeding k, Heisenberg writes

$$\epsilon_k\,(\text{mechanical}) = 2\,\rho\nu_k \int_0^k F(k)\, k^2 dk , \tag{12}$$

in analogy with expression (4) for the thermal part of the energy transfer. In writing ϵ_k (mechanical) in the form (12), we are assuming that the process of energy transfer between the sets of eddies $(0, k)$ and (k, ∞) can be visualized in terms of a suitably defined coefficient of viscosity, ν_k. We are, of course, familiar with the concept of eddy viscosity derived from the picture of eddies describing a certain mean free path l_k with a certain root-mean-square velocity, v_k. On this picture

$$\nu_k \sim l_k v_k \sim \frac{v_k}{k} , \tag{13}$$

since we may expect l_k to be of the order of $1/k$. However, for our purposes this is not a suitable expression for ν_k; to be useful, it must be expressed in terms of the spectrum. As the simplest of possible expressions, Heisenberg assumes that

$$\nu_k = \kappa \int_k^\infty \sqrt{\frac{F(k)}{k^3}}\, dk , \tag{14}$$

[12] *Phys. Rev.*, **68**, 286, 1945.

[13] *Zs. f. Phys.*, **124**, 614, 1948.

[14] *Zs. f. Phys.*, **124**, 628, 1948, and *Proc. R. Soc. London, A*, **195**, 402, 1948.

where κ is a certain numerical constant. Apart from the fact that the expression on the right-hand side is of the correct dimension, the justification for writing it in this particular form is the following.

If $F(k)$ follows a simple power law of the form k^{-n}, then

$$v_k = \left[\int_k^\infty F(k)\, dk\right]^{1/2} \propto k^{-(n-1)/2},$$

and

$$\nu_k \propto \frac{v_k}{k} \propto k^{-(n+1)/2} \propto \int_k^\infty \frac{dk}{k^{(n+3)/2}} \propto \int_k^\infty \sqrt{\frac{F(k)}{k^3}}\, dk. \tag{15}$$

In other words, formula (14) is a valid form when $F(k)$ follows a power law; we assume that we may use the same expression even when this is not the case.

According to equations (12) and (14), we have

$$\epsilon_k\,(\text{mechanical}) = 2\rho \int_k^\infty \sqrt{\frac{F(k'')}{k''^3}}\, dk'' \int_0^k F(k')\, k'^2 dk'. \tag{16}$$

This expression for ϵ_k admits of a simple interpretation. In a unit volume of the medium, eddies with wave numbers between k' and $k' + dk'$ transfer energy to eddies with wave numbers between k'' and $k'' + dk''$ ($k'' \geq k'$) at the rate

$$\epsilon(k'; k'')\, dk' dk'' = 2\rho F(k')\, k'^2 \sqrt{\frac{F(k'')}{k''^3}}\, dk' dk''. \tag{17}$$

We may think of $\epsilon(k'; k'')$ as a transition probability governing the process of energy transfer between eddies. The possibility of defining such a transition probability is, of course, implicit in our concept of the existence of a hierarchy of eddies.

Now, combining the expressions for ϵ_k(thermal) and ϵ_k(mechanical), we have

$$\epsilon_k = 2\rho \left\{\nu + \kappa \int_k^\infty \sqrt{\frac{F(k'')}{k''^3}}\, dk''\right\} \int_0^k F(k')\, k'^2 dk'. \tag{18}$$

This is the fundamental equation of Heisenberg's theory. It combines in a single expression the ideas underlying the picture of turbulence in terms of eddies describing mean free paths and the principle expressed by Richardson's rhyme.

For stationary turbulence, ϵ_k must be a constant, independent of k, and this condition suffices to determine the spectrum. Indeed, the exact solution of equation (18) for the case ϵ_k = Constant can be given explicitly.[15] We have

$$\begin{aligned} F(k) &= \text{Constant}\left(\frac{k_0}{\kappa}\right)^{5/3} \frac{1}{[1 + (k/k_s)^4]^{4/3}} \qquad (k > k_0) \\ &= 0 \qquad\qquad\qquad\qquad\qquad\qquad\quad (k < k_0), \end{aligned} \tag{19}$$

where

$$k_s \doteq 0.2211\, k_0 (R_0 \kappa)^{3/4}, \tag{20}$$

and R_0 denotes the Reynolds number

$$\begin{aligned} R_0 = \frac{1}{\nu}\,&(\text{Root-mean-square velocity of all eddies present}) \\ &\times \text{diameter } (= \pi/k_0) \text{ of the largest eddies}. \end{aligned} \tag{21}$$

[15] S. Chandrasekhar, *Phys. Rev.*, 75, 896, 1949.

We observe that, according to these equations, when $R_0 \to \infty$, $k_s \to \infty$, and we recover the Kolmogoroff spectrum. More generally, when the Reynolds number is finite, but sufficiently large, there will be a region of the spectrum in which the inequality $k_0 < k \ll k_s$ will be valid; and in this region the spectrum will follow closely the $k^{-5/3}$-law. However, the solution of Heisenberg's equation also shows that, for any finite Reynolds number, no matter how large, we must get departures from the $k^{-5/3}$-law when k approaches k_s and that for $k \gg k_s$ the spectrum follows an inverse seventh-power law:

$$F(k) \propto k^{-7} \qquad (k \gg k_s) . \tag{22}$$

Evidently this is the region of the spectrum where the dissipation by viscosity into thermal energy is the dominant factor. Accordingly, we may take k_s as defining the wave number of the eddies at which the dissipation by kinetic viscosity becomes comparable to the kinetic energy transferred to smaller eddies by eddy viscosity.

In astronomical contexts we shall probably be mostly concerned with stationary turbulence. But it is an important aspect of Heisenberg's theory that it also enables the treatment of the problem of the decay of turbulence. I shall therefore spend a few moments on this aspect of the subject.

Now, if there is no external agency maintaining turbulence, then clearly

$$\epsilon_k = -\rho \frac{\partial}{\partial t} \int_0^k F(k, t)\, dk , \tag{23}$$

since, by definition, ϵ_k is the net energy dissipated by the eddies with wave number between 0 and k, either in the form of molecular motion and thermal energy or in the form of the motions of the smaller eddies and kinetic energy. The decay of turbulence will therefore be described by the equation

$$-\frac{\partial}{\partial t} \int_0^k F(k, t)\, dk = 2 \left\{ \nu + \kappa \int_k^\infty \sqrt{\frac{F(k'', t)}{k''^3}}\, dk'' \right\} \times \int_0^k F(k', t)\, k'^2 dk' . \tag{24}$$

A case of some importance in this connection is the following: Suppose that we have initially an equilibrium spectrum and that, at a certain instant, the agency maintaining the turbulence is cut off. Then, in the decay of turbulence which will ensue, we may distinguish three stages—an early stage, during which the larger eddies ($k \sim k_0$) adjust themselves to the fact that no energy is being communicated to them; an intermediate stage, during which there is a sufficient store of energy among the larger eddies to maintain an equilibrium distribution among the lower members of the hierarchy and the Reynolds number remains constant; and, finally, a last stage, during which the store of energy among the larger eddies is getting exhausted and the Reynolds number decreases to zero.

While a unified discussion of all three stages of decay is a difficult problem, it appears that on the basis of Heisenberg's equation we can follow the second stage quite completely and in an explicit fashion; for, from the constancy of the Reynolds number which we expect during this stage, we conclude that the spectrum must be "self-preserving" in the sense that it keeps the same form, though the scale may change with time. From equation (24) the condition of self-preservation is seen to be equivalent to seeking solutions of this equation of the form

$$F(k, t) \equiv \frac{1}{\sqrt{t}} f(k\sqrt{t}) , \tag{25}$$

where f is a function of the argument $k\sqrt{t}$. The physical meaning of a solution of this form is that during the decay the eddies grow in size like $\sqrt{t}$ and that the total energy stored in turbulence decays like $1/t$:

$$\epsilon = \int_0^\infty F(k, t)\, dk = \frac{1}{t} \int_0^\infty f(x)\, dx . \tag{26}$$

However, the *form* of the spectrum remains unchanged.

With $F(k, t)$ given by equation (25), the equation determining $f(x)$ can be reduced to a second-order nonlinear differential equation which can be studied by standard methods.[16] And the discussion shows that, for any finite Reynolds number,

$$F(k, t) \propto k \qquad (k \to 0),$$

and

$$F(k, t) \propto k^{-7} \qquad (k \to \infty). \quad (27)$$

Moreover, when the Reynolds number is sufficiently large, there is a part of the spectrum which approximately follows the $k^{-5/3}$-law. And when the Reynolds number actually becomes infinite (or equivalently $\nu = 0$), the spectrum approaches the $k^{-5/3}$-law exactly as $k \to \infty$.

As I stated earlier, the decay spectrum predicted by these curves is valid only during the second stage, when there is a sufficient store of energy among the larger eddies to maintain an equilibrium distribution for $k \to \infty$. When this ceases to be the case, the decay will proceed much more rapidly and, as Batchelor and Townsend[17] have shown, the $1/t$-law is then replaced by $1/t^{5/2}$-law, and the Reynolds number, instead of remaining constant, starts decreasing to zero.

I think that that about describes the present state of the theory of turbulence. Having spent so much time on the physical theory, I should like to conclude by a brief reference to an application which von Weizsäcker has made of these ideas on turbulence.[18]

As is probably generally known, von Weizsäcker has outlined a general cosmogony, the essential feature of which is the prominent role which he ascribes to the interplay between turbulence and rotation.

It is the usual fate of cosmogonical theories not to survive. I do not suppose that von Weizsäcker's theory will prove the exception to this rule. However, I have been personally attracted by his writings for two reasons, first, because he expresses himself with a restraint and a modesty which is unusual among writers in this field and, second, because I think that we may accede to the importance he ascribes to turbulence without, at the same time, subscribing to his detailed picture of the manner in which he expects turbulence to operate. From one point of view he may be said to have scored already; for it was his emphasis on the role of turbulence in cosmogony that led Heisenberg to examine the basic physical theory, with the result that we have today the beginnings of a foundation on which we may build.

As I have said, von Weizsäcker's cosmogony rests on the effects which he expects from the interplay of rotation and turbulence. More particularly, the effects which he expects can be described in the following terms.

Consider, for example, a sheet of gas at very low density in the equatorial plane of a central mass, which we may identify with a star or with the nucleus of a galaxy. If we ignore, in the first instance, the effects of pressure and viscosity, each element of gas will describe a Keplerian orbit in the field of the central mass. If the system is assumed to have an axial symmetry, the orbits must be circular, and the angular velocity will vary with distance, s, from the axis, according to the law

$$\omega^2 = \frac{GM}{s^3}. \quad (28)$$

The successive rings of gas in the medium will therefore have motions relative to one another, and turbulence will ensue. As a result of this turbulence, viscous stresses will come into play and will perturb the motions, both in the radial and in the transverse direc-

[16] *Ibid.*, **76**, 1454, 1949, and *Proc. R. Soc. London, A* (in press).

[17] *Proc. R. Soc. London, A*, **193**, 539; **194**, 527, 1948.

[18] *Zs. f. Ap.*, **22**, 319, 1944; **24**, 181, 1947, and *Zs. f. Naturforsch.*, **3A**, 524, 1948.

tions. Examining the sense of these perturbations, von Weizsäcker concludes that all matter interior to a certain critical radius will fall toward the center, while the matter outside this radius will tend to move outward and dissipate into space. He invokes this mechanism in a variety of different contexts: for the dissipation of the gaseous envelope which, he imagines, once surrounded the sun and in which he presumes the solar system was formed; for interpreting the ring structure in extragalactic nebula. And, going further back to the beginning of things, he believes the linear dimensions of the present galaxies must have been determined by the condition that at these distances the forces which led to the expansion of the universe and the forces which result from turbulence compete on about equal terms; he infers the existence of such a distance from the fact that, while the mean velocities increase linearly with the distance in an expanding universe, the mean velocities increase only as the one-third power of the distance in a turbulent medium. I should emphasize again that it is not necessary to subscribe to all these speculations of von Weizsäcker to grant the importance of turbulence for the purpose of cosmogony. It is, indeed, entirely possible that the theory of turbulence which I have described may bear its first fruits in a much less spectacular way in the solution of more specific problems. Thus Martin Schwarzschild has already extended Heisenberg's theory to include the agency maintaining turbulence for the case when turbulence results from thermal instability. It is clear that the extension of Heisenberg's theory to the case of turbulence induced by thermal instability must have important applications to the interpretation of the solar granules.

While we shall have to wait for these and similar developments before we can finally pass on the importance of the recent developments in our understanding of turbulence for astrophysics, I think we may be sure of at least one thing.

"We cannot make bricks without straw"; that is a common enough saying. It is equally true that we cannot construct a rational astrophysical theory without an adequate base of physical knowledge. It would therefore seem to me that we cannot expect to incorporate the concept of turbulence in astrophysical theories in any essential manner without a basic physical theory of the phenomenon of turbulence itself. It appears that the first outlines of such a physical theory are just emerging.

SOME ASPECTS OF THE STATISTICAL THEORY OF TURBULENCE

BY

S. CHANDRASEKHAR

1. Isotropic tensors and the equations of isotropic turbulence. The central idea in the statistical theory of homogeneous isotropic turbulence initiated by G. I. Taylor [1] is that of isotropy, which requires the time average of any function of the velocity components, defined with respect to a particular set of axes, to be invariant under arbitrary rotations and reflections of the axes of reference. A phenomenological theory of turbulence incorporating this idea of isotropy divides itself into two parts: a *kinematical* part, which consists in setting up correlations between velocity components at two different points of the medium and in reducing the forms of the associated tensors to meet the requirements of isotropy; and a *dynamical* part, which consists in deriving the consequences of the equations of motion and continuity for the fundamental scalar functions defining the correlation tensors. When dealing with turbulence in an incompressible fluid, it is convenient to include in the kinematical part the restrictions on the correlation tensors introduced by the equation of continuity and to reserve for the dynamical part alone the implications of the Stokes-Navier equation. The principal equations of this theory were derived by von Kármán and Howarth [2], but a concise and elegant treatment of the subject is due to H. P. Robertson [3], who developed for this purpose a theory of *isotropic tensors* and *forms*. The basic idea underlying this new formal development can be explained quite simply.

Consider the fundamental correlation tensor

$$Q_{ij} = \overline{u_i u_j'} \tag{1}$$

of the components u_i and u_j' of the velocity at two different points, say, $P\ (x_i)$ and $P'\ (x_i')$ in the medium. Let ξ denote the vector joining the points x_i and x_i', that is, $\xi_i = x_i' - x_i$. The correlation $\overline{u_{\mathbf{a}} u_{\mathbf{b}}'}$ of the velocities along two directions specified by the unit vectors $\mathbf{a}$ at P and $\mathbf{b}$ at P' can be expressed in terms of Q_{ij}; for, clearly,

$$\overline{u_{\mathbf{a}} u_{\mathbf{b}}'} = a_i b_j \overline{u_i u_j'} = Q_{ij} a_i b_j, \tag{2}$$

where, here and elsewhere, summation over repeated indices is to be understood. Now when we say that the turbulence is homogeneous and isotropic, we mean, in particular, that $Q_{ij} a_i b_j$ is invariant to all translations and proper rotations of the vector configuration ξ, $\mathbf{a}$, and $\mathbf{b}$ as a rigid body and also to reflections at the origin.

Similarly, the correlation $\overline{u_{\mathbf{a}} u_{\mathbf{b}} u_{\mathbf{c}}'}$ of the velocities $u_{\mathbf{a}}$, $u_{\mathbf{b}}$, and $u_{\mathbf{c}}'$ along directions specified by the unit vectors $\mathbf{a}$ and $\mathbf{b}$ at P and $\mathbf{c}$ at P' can be expressed in terms of the triple correlation tensor

$$T_{ijk} = \overline{u_i u_j u_k'}; \tag{3}$$

thus,

$$\overline{u_a u_b u_c'} = T_{ijk} a_i b_j c_k. \tag{4}$$

In isotropic turbulence $T_{ijk} a_i b_j c_k$ should again be a scalar invariant to all rotations of the vector configuration $\boldsymbol{\xi}$, $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ as a rigid body and to reflections at the origin. Robertson pointed out that, in virtue of this invariance of the scalar products like $Q_{ij} a_i b_j$ and $T_{ijk} a_i b_j c_k$, we can write down quite readily the forms of the various tensors by appealing to the following result from the theory of invariants: Any invariant function of any number of vectors $\boldsymbol{\xi}$, $\mathbf{a}$, $\mathbf{b}$, $\cdots$, etc., can be expressed in terms of the fundamental invariants of the following two types: (1) the scalar products such as $(\boldsymbol{\xi} \cdot \mathbf{a}) = \xi_i a_i$ and $(\mathbf{a} \cdot \mathbf{b}) = a_i b_i$ of any two vectors including the scalar squares $(\boldsymbol{\xi} \cdot \boldsymbol{\xi})$, $(\mathbf{a} \cdot \mathbf{a})$, etc., and (2) the determinants such as $[\mathbf{a}\mathbf{b}\boldsymbol{\xi}] = \epsilon_{ijk} a_i b_j \xi_k$ of any three of the vectors, where ϵ_{ijk} is the usual alternating symbol.

Among the scalar invariants constructed in accordance with the foregoing theorem we must distinguish between two classes: the class which can be expressed as sums and products of the scalar products only (*i.e.*, the scalar invariants of type 1 only); and the class which can be expressed as sums and products of an odd number of determinants (*i.e.*, the scalar invariants of type 2) and the scalar products. The invariants of the first class are unchanged under the full rotation group, *i.e.*, not only for the proper rotations exemplified by the motions of a rigid body but also for the "improper" rotations such as reflections in a point. We shall call the invariants of the second class *skew invariants* and the corresponding tensors *skew tensors*. In the theory of isotropic turbulence we shall have to deal with both types of invariants and tensors. Thus the correlations Q_{ij} and T_{ijk} are true isotropic tensors, and the scalar products $Q_{ij} a_i b_j$ and $T_{ijk} a_i b_j c_k$ should be expressible in terms of the scalar invariants of type 1 only. On the other hand, correlations such as $\overline{u_i \omega_j'}$ and $\overline{u_i u_j \omega_j'}$ which include an odd number of the components of the vorticity $\boldsymbol{\omega}$, are skew isotropic tensors which transform as tensors under proper rotations but which take the opposite sign to true tensors on reflection in the origin. The corresponding skew forms should be expressible as sums and products of an odd number of the available determinants and, of course, the scalar products as well.

As we have already remarked, by appealing to the foregoing results from the theory of invariants, we can write down the forms of the various correlation tensors. Thus, considering $Q_{ij} a_i b_j$, we observe that it must be expressible in terms of $(\boldsymbol{\xi} \cdot \mathbf{a})$, $(\boldsymbol{\xi} \cdot \mathbf{b})$, and $(\mathbf{a} \cdot \mathbf{b})$ only. And since $Q_{ij} a_i b_j$ is bilinear, it can only be of the form

$$Q_{ij} a_i b_j = Q_1 (\boldsymbol{\xi} \cdot \mathbf{a})(\boldsymbol{\xi} \cdot \mathbf{b}) + Q_2 (\mathbf{a} \cdot \mathbf{b}), \tag{5}$$

where Q_1 and Q_2 are two arbitrary functions of the distance $r = (\xi \cdot \xi)^{\frac{1}{2}}$ between the two points considered. Since (5) must be true for all **a** and **b**, it follows that the general form of an isotropic tensor of the second order is

$$Q_{ij} = Q_1 \xi_i \xi_j + Q_2 \delta_{ij}. \tag{6}$$

Similarly, by considering the scalar products available for the construction of the scalar invariant $T_{ijk} a_i b_j c_k$ [namely, $(\xi \cdot \mathbf{a})$, $(\xi \cdot \mathbf{b})$, $(\xi \cdot \mathbf{c})$, $(\mathbf{a} \cdot \mathbf{b})$, $(\mathbf{b} \cdot \mathbf{c})$, and $(\mathbf{c} \cdot \mathbf{a})$], we readily verify that the most general form of an isotropic tensor of the third order is

$$T_{ijk} = T_1 \xi_i \xi_j \xi_k + T_2 \xi_i \delta_{jk} + T_3 \xi_j \delta_{ki} + T_4 \xi_k \delta_{ij}, \tag{7}$$

where T_1, T_2, T_3, and T_4 are four arbitrary functions of r. In the particular case of the triple correlation $\overline{u_i u_j u'_k}$, it is clear that it must be symmetrical in the indices i and j, and this requires that $T_2 = T_3$ in (7).

In the statistical theory of turbulence the correlation tensors one considers are generally solenoidal in one or more of their indices. This solenoidal property results from an application of the equation of continuity. For, in an incompressible fluid (to which most attention has been given so far; see, however, Sec. 4 of this paper) the equation of continuity,

$$\frac{\partial u_i}{\partial x_i} = 0, \tag{8}$$

requires that the velocity field at any point be divergence-free. Consequently, the tensor $\overline{u_i u'_j}$ must be solenoidal in both its indices, whereas the triple correlation tensor $\overline{u_i u_j u'_k}$ must be solenoidal in the last index k. We can impose these solenoidal conditions on the tensors given by equations (6) and (7) and obtain certain differential equations (one in the case of Q_{ij} and two in the case of T_{ijk}) connecting the coefficients of these tensors. These differential equations will reduce the number of independent scalars required for the definition of Q_{ij} and T_{ijk}. This is the procedure which Robertson [3] has followed in his paper. However, it appears that the representation of a tensor that is solenoidal in one or more of its indices in terms of a minimal number of scalars can be achieved most simply by expressing the tensors as the curl with respect to those indices of a *skew tensor*. Thus, expressing

$$Q_{ij} = \epsilon_{jlm} \frac{\partial q_{im}}{\partial \xi_l}, \tag{9}$$

where q_{ij} is a skew isotropic tensor of the second order, we can satisfy the solenoidal and the isotropic conditions simultaneously. [Since Q_{ij} will be symmetrical in its indices, it will also be solenoidal in i.] Since out of three vectors ξ, **a**, and **b** we can form one and only one determinant, it is clear that the most general form of q_{ij} is

$$q_{ij} = Q \epsilon_{ijk} \xi_k \tag{10}$$

where Q is an arbitrary function of r. Taking the curl of q_{ij} with respect to the index j, we find that

$$Q_{ij} = \frac{Q'}{r}\xi_i\xi_j - (rQ' + 2Q)\delta_{ij}, \tag{11}$$

where a prime attached to a scalar function (such as Q) denotes differentiation with respect to r. This then is the most general form of a solenoidal isotropic tensor of the second order. We may say that Q is the *defining scalar* of the tensor Q_{ij}. This representation of Q_{ij} in terms of a defining scalar is unique; for, $Q = 0$ implies that $Q_{ij} \equiv 0$, and conversely from $Q_{ij} \equiv 0$ we can conclude that $Q = 0$ (*i.e.*, provided we do not allow any singularity for Q at $r = 0$).

If Q_{ij} is isotropic and solenoidal, then so is $\nabla^2 Q_{ij}$ since the Laplacian is a scalar operator. The defining scalar of $\nabla^2 Q_{ij}$ can be obtained by applying the Laplacian directly to (9) since the operation of the Laplacian and the curl are permutable. In this manner we find that

$$\nabla^2 q_{ij} = \left(\frac{\partial^2 Q}{\partial r^2} + \frac{4}{r}\frac{\partial Q}{\partial r}\right)\epsilon_{ijl}\xi_l. \tag{12}$$

Accordingly, the defining scalar of $\nabla^2 Q_{ij}$ is

$$\frac{\partial^2 Q}{\partial r^2} + \frac{4}{r}\frac{\partial Q}{\partial r}. \tag{13}$$

As an application of the foregoing results we may note that, if Q_{ij} is isotropic and solenoidal, then the tensor equation

$$\frac{\partial Q_{ij}}{\partial t} = \nabla^2 Q_{ij} \tag{14}$$

is equivalent to the scalar equation

$$\frac{\partial Q}{\partial t} = \frac{\partial^2 Q}{\partial r^2} + \frac{4}{r}\frac{\partial Q}{\partial r} \tag{15}$$

for the defining scalar.

By an accidental circumstance peculiar to the case, the foregoing discussion of the second-order isotropic solenoidal tensor Q_{ij} does not bring out one important consideration which must be taken into account in representing solenoidal tensors as the curl of suitably defined skew tensors: that is the question of gauge invariance. The importance of this consideration will become clear when we try to express the triple correlation $T_{ijk} = \overline{u_i u_j u'_k}$ in the form

$$T_{ijk} = \epsilon_{klm}\frac{\partial t_{ijm}}{\partial \xi_l}, \tag{16}$$

where t_{ijk} is a third-order isotropic skew tensor. Now, to construct a skew trilinear form, $t_{ijk}a_i b_j c_k$, we have the four determinants $[\mathbf{abc}]$, $[\mathbf{ab}\boldsymbol{\xi}]$, $[\mathbf{bc}\boldsymbol{\xi}]$, and $[\mathbf{ca}\boldsymbol{\xi}]$ available. From this it would appear that a general skew trilinear form must be a linear combination of

(17) $\quad [\mathbf{abc}], \quad (\mathbf{c}\cdot\boldsymbol{\xi})[\mathbf{ab}\boldsymbol{\xi}], \quad (\mathbf{a}\cdot\boldsymbol{\xi})[\mathbf{bc}\boldsymbol{\xi}], \quad$ and $\quad (\mathbf{b}\cdot\boldsymbol{\xi})[\mathbf{ca}\boldsymbol{\xi}]$

with coefficients which are arbitrary functions of r. Correspondingly, it would appear that the general skew tensor of the third order must be a linear combination of the *four* tensors

(18) $$\epsilon_{ijk}, \quad \xi_k\epsilon_{ijl}\xi_l, \quad \xi_i\epsilon_{jkl}\xi_l, \quad \text{and} \quad \xi_j\epsilon_{kil}\xi_l.$$

However, in virtue of the identity

(19) $$\xi_i\epsilon_{jkl}\xi_l + \xi_j\epsilon_{kil}\xi_l + \xi_k\epsilon_{ijl}\xi_l = r^2\epsilon_{ijk},$$

only *three* of the four tensors (18) are linearly independent. Also, since

(20) $$\frac{\partial}{\partial \xi_k} Q\epsilon_{ijl}\xi_l = \frac{Q'}{r}\,\xi_k\epsilon_{ijl}\xi_l + Q\epsilon_{ijk},$$

where Q is an arbitrary function of r, it follows that, for the purposes of defining a solenoidal tensor according to (16), the tensor $Q\epsilon_{ijk}$ is equivalent to the tensor $-(Q'/r)\xi_k\epsilon_{ijl}\xi_l$, since the two tensors differ only by a gradient of a vector and the curl of this difference taken in the fashion (16) is zero. Accordingly, for defining (in a gauge-invariant fashion) a tensor T_{ijk} solenoidal in k, the most general skew tensor we need consider is

(21) $$t_{ijk} = T_1\xi_i\epsilon_{jkl}\xi_l + T_2\xi_j\epsilon_{ikl}\xi_l,$$

where T_1 and T_2 are two arbitrary functions of r. If the tensor T_{ijk} is in addition symmetrical in the indices i and j (as is the case with $\overline{u_iu_ju_k'}$), then

(22) $$T_1 = T_2 = T \text{ (say)}$$

and

(23) $$\begin{aligned} T_{ijk} &= \operatorname{curl} T(\xi_i\epsilon_{jkl}\xi_l + \xi_j\epsilon_{ikl}\xi_l) \\ &= \frac{2}{r}\,T'\xi_i\xi_j\xi_k - (rT' + 3T)(\xi_i\delta_{jk} + \xi_j\delta_{ik}) + 2T\xi_k\delta_{ij}. \end{aligned}$$

Again it should be noted that the foregoing representation of the tensor T_{ijk} in terms of a single defining scalar T is unique in that $T = 0$ implies that $T_{ijk} \equiv 0$, and conversely from $T_{ijk} \equiv 0$ we can conclude that $T = 0$ (*i.e.*, provided T has no singularity at $r = 0$).

By operating the Laplacian on t_{ijk} given by equation (21), we find that the defining scalar of $\nabla^2 T_{ijk}$ is

(24) $$\frac{\partial^2 T}{\partial r^2} + \frac{6}{r}\frac{\partial T}{\partial r}.$$

Thus the equation

(25) $$\frac{\partial T_{ijk}}{\partial t} = \nabla^2 T_{ijk},$$

where T_{ijk} is an isotropic tensor symmetrical in i and j and solenoidal in k, is equivalent to the equation

$$\frac{\partial T}{\partial t} = \frac{\partial^2 T}{\partial r^2} + \frac{6}{r}\frac{\partial T}{\partial r} \tag{26}$$

for the defining scalar T.

The contracted tensor

$$T_{ij} = \frac{\partial T_{ikj}}{\partial \xi_k} \tag{27}$$

of the triple correlation $\overline{u_i u_j u_k'}$ occurs in the treatment of the Stokes-Navier equation [cf. equation (37) below]. The tensor T_{ij} is clearly solenoidal, and its defining scalar can be found by similarly contracting the skew tensor t_{ijk}; thus

$$\frac{\partial}{\partial \xi_k} T(\xi_i \epsilon_{kjl} \xi_l + \xi_k \epsilon_{ijl} \xi_l). \tag{28}$$

We find this is equal to

$$\left(r\frac{\partial T}{\partial r} + 5T\right)\epsilon_{ijl}\xi_l. \tag{29}$$

Hence the defining scalar of T_{ij} is

$$\left(r\frac{\partial}{\partial r} + 5\right)T. \tag{30}$$

The foregoing theory of isotropic tensors adapts itself quite readily to the treatment of the Stokes-Navier equation:

$$\frac{\partial u_i}{\partial t} + \frac{\partial}{\partial x_k} u_i u_k = -\frac{1}{\rho}\frac{\partial p}{\partial x_i} + \nu \nabla^2 u_i. \tag{31}$$

In this equation ν is the kinematic viscosity. Multiplying equation (31) by the velocity component u_j' at x_i' and averaging the resulting equation, we have

$$\overline{u_j' \frac{\partial u_i}{\partial t}} + \frac{\partial}{\partial x_k}\overline{u_i u_k u_j'} = \nu \nabla^2 \overline{u_i u_j'}. \tag{32}$$

[The term in the pressure in the equation of motion does not appear in this equation, since the correlation $\overline{pu_j'}$, being a solenoidal isotropic vector, is identically zero; for, if such a vector existed, it should be expressible as the curl of a skew isotropic vector; such a vector clearly does not exist, since with two vectors $\boldsymbol{\xi}$ and $\mathbf{a}$ we cannot form a determinant to represent the corresponding skew linear form.]

Interchanging the indices i and j and also the primed and the unprimed quantities in equation (32), we have

$$\overline{u_i \frac{\partial u_j'}{\partial t}} + \frac{\partial}{\partial x_k'}\overline{u_j' u_k' u_i} = \nu \nabla'^2 \overline{u_i u_j'}. \tag{33}$$

Adding equations (32) and (33) and remembering that

$$\frac{\partial}{\partial \xi_i} = \frac{\partial}{\partial x'_i} = -\frac{\partial}{\partial x_i}, \tag{34}$$

we have

$$\frac{\partial}{\partial t}\overline{u_i u'_j} = 2\frac{\partial}{\partial \xi_k}\overline{u_i u_k u'_j} + 2\nu\,\nabla^2\overline{u_i u'_j}. \tag{35}$$

In deriving this last equation we have made further use of the relation

$$\overline{u_i u_j u'_k} = -\overline{u'_i u'_j u_k}, \tag{36}$$

which is a consequence of the assumed homogeneity.

In terms of the tensors Q_{ij} and T_{ijk} equation (35) can be rewritten in the form

$$\frac{\partial Q_{ij}}{\partial t} = 2\frac{\partial}{\partial \xi_k}T_{ikj} + 2\nu\,\nabla^2 Q_{ij}. \tag{37}$$

The scalars defining the various second-order tensors occurring in this equation are [cf. equations (15) and (30)]

$$\frac{\partial Q}{\partial t}, \qquad \left(r\frac{\partial}{\partial r} + 5\right)T, \qquad \text{and} \qquad \left(\frac{\partial^2}{\partial r^2} + \frac{4}{r}\frac{\partial}{\partial r}\right)Q, \tag{38}$$

respectively. Hence equation (37) is equivalent to the scalar equation

$$\frac{\partial Q}{\partial t} = 2\left(r\frac{\partial}{\partial r} + 5\right)T + 2\nu\left(\frac{\partial^2}{\partial r^2} + \frac{4}{r}\frac{\partial}{\partial r}\right)Q. \tag{39}$$

This is the equation of von Kármán and Howarth derived in a different way.

Equation (39) admits an integral: multiplying it by r^4 and integrating from 0 to r, we have

$$\frac{\partial}{\partial t}\int_0^r Qr^4\,dr = 2r^5T + 2\nu r^4\frac{\partial Q}{\partial r}. \tag{40}$$

If we now assume that $r^5T \to 0$ and $r^4Q' \to 0$ as $r \to \infty$, then it follows that

$$\int_0^\infty Q(r,t)r^4\,dr = \text{const.} \tag{41}$$

This is the so-called Loitsiansky invariant. As Batchelor [4] has shown, the meaning of this invariant is that "the spectral density at very low wave numbers is permanent and is determined by the initial conditions." We shall see that a similar invariant exists also in the theory of turbulence of a compressible fluid.

2. The theory of axisymmetric tensors and the equations of axisymmetric turbulence. While the theory of isotropic turbulence which we have described in part in the preceding section introduces the subject in its simplest context, it is, nevertheless, almost invariably true that, whenever turbulence is present, there is also present a preferred direction defined by the direction of the mean flow. Indeed, since the fluctuations in the velocity field are generally defined with respect to the local mean values, it would seem that a more natural starting point for the theory will be provided by the concept of *axisymmetry*, which will require the mean value of any function of the velocities to be invariant not for the full rotation group but only for rotations about the preferred direction $\boldsymbol{\lambda}$ (say) and for reflections in planes containing $\boldsymbol{\lambda}$ and perpendicular to $\boldsymbol{\lambda}$. A corresponding theory of axisymmetric turbulence is best developed in terms of a theory of *axisymmetric tensors* analogous to the theory of isotropic tensors.

Let $F_{ijk}\ldots$ denote a Cartesian tensor; further let **a**, **b**, **c**, etc., denote arbitrary unit vectors. Consider the scalar product

$$F(\mathbf{a},\mathbf{b},\mathbf{c}, \cdots) = F_{ijk}\ldots a_i b_j c_k \cdot \tag{42}$$

We shall say that the tensor $F_{ijk}\ldots$ is axially symmetric about a direction specified by a unit vector $\boldsymbol{\lambda}$ if $F(\mathbf{a},\mathbf{b},\mathbf{c}, \cdots)$ is invariant for arbitrary rotations about the direction $\boldsymbol{\lambda}$ and for reflections in planes containing $\boldsymbol{\lambda}$ and normal to $\boldsymbol{\lambda}$.

It is of interest to see the relation between axisymmetric tensors and forms and isotropic tensors and forms. In the theory of isotropic tensors, the form F defined as in equation (42) is invariant for rotations and reflections of the configurations formed by the vectors **a**, **b**, **c**, etc., and $\boldsymbol{\xi}$, where $\xi_i = x_i' - x_i$ and x_i' and x_i are the coordinates of the two points at which correlations of the field variables are considered. On the other hand in the theory of axisymmetric tensors the form F must be invariant for all rotations and reflections of the vector configuration formed by $\boldsymbol{\xi}$, **a**, **b**, **c**, etc., *and* $\boldsymbol{\lambda}$. The general problem in the theory of axisymmetric tensors is, therefore, to determine the form $F(\boldsymbol{\xi},\boldsymbol{\lambda};\mathbf{a},\mathbf{b},\mathbf{c}, \cdots)$ which will be invariant under the full rotation group of the vector configuration $\boldsymbol{\xi}$, $\boldsymbol{\lambda}$, **a**, **b**, **c**, etc. The formal problem can be readily solved by appealing to the same theorem in the theory of invariants quoted in the last section.

As in the theory of isotropic tensors we must distinguish between axisymmetric tensors and skew axisymmetric tensors, which take the opposite sign to true tensors on reflection in the origin. Again, it is not difficult to write down the general expressions for such skew tensors, since the corresponding skew forms must be expressible as sums of products of an odd number of determinants such as $[\mathbf{ab}\boldsymbol{\xi}]$, $[\mathbf{ab}\boldsymbol{\lambda}]$, $[\mathbf{a}\boldsymbol{\lambda}\boldsymbol{\xi}]$, $[\mathbf{abc}]$, etc., formed by any three of the vectors $\boldsymbol{\xi}$, $\boldsymbol{\lambda}$, **a**, **b**, **c**, etc. By choosing combinations of the scalar products and an odd number of the available determinants which will be in conformity with (42), we can write down the corresponding skew forms and tensors.

In the theory of axisymmetric turbulence (as in the case of isotropic turbulence) particular interest attaches to tensors which are solenoidal in one or more of their indices. In these cases the tensors can be expressed as the curl of certain suitably defined skew tensors, and defining scalars can be introduced, as in the theory of isotropic solenoidal tensors. However, in representing the solenoidal tensors in terms of certain defining scalars, particular attention must be given to selecting a minimal set of linearly independent skew tensors and also to the matter of gauge invariance. The theory of axisymmetric tensors of the first, second, and third orders along these lines has been completed by Chandrasekhar [5] (see also Batchelor [6]). The following results on axisymmetric tensors of the first and second orders are taken from [5].

(i) *Axisymmetric vector.* An axisymmetric vector L_i must be of the form

$$L_i = M\xi_i + N\lambda_i, \tag{43}$$

where M and N are arbitrary functions of r and $r\mu$, where

$$r^2 = (\boldsymbol{\xi}\cdot\boldsymbol{\xi}) \quad \text{and} \quad r\mu = (\boldsymbol{\xi}\cdot\boldsymbol{\lambda}). \tag{44}$$

If L_i is solenoidal, it can be expressed as

$$\begin{aligned} L_i &= \operatorname{curl} L(r,\mu)\epsilon_{ilm}\lambda_l\xi_m \\ &= -\left(\mu\frac{\partial L}{\partial r} + \frac{1-\mu^2}{r}\frac{\partial L}{\partial\mu}\right)\xi_i + \left(r\frac{\partial L}{\partial r} + 2L\right)\lambda_i. \end{aligned} \tag{45}$$

The representation of an axisymmetric solenoidal vector L_i in terms of a single defining scalar L in the manner (45) is unique.

The defining scalar of $\nabla^2 L_i$ is

$$\Delta L = \left(\frac{\partial^2}{\partial r^2} + \frac{4}{r}\frac{\partial}{\partial r} + \frac{1-\mu^2}{r^2}\frac{\partial^2}{\partial\mu^2} - \frac{4\mu}{r^2}\frac{\partial}{\partial\mu}\right)L. \tag{46}$$

(ii) *Axisymmetric tensors of the second order.* An axisymmetric tensor Q_{ij} must be of the form

$$Q_{ij} = A\xi_i\xi_j + B\delta_{ij} + C\lambda_i\lambda_j + D\lambda_i\xi_j + E\xi_i\lambda_j, \tag{47}$$

where A, B, C, D, and E are arbitrary functions of r and $r\mu$. [Note that, in contrast to an isotropic tensor, Q_{ij} is now not necessarily symmetrical in i and j.] If Q_{ij} is solenoidal in j, it can be expressed in terms of three defining scalars Q_1, Q_2, and Q_3 in the manner

$$Q_{ij} = \epsilon_{jlm}\frac{\partial q_{im}}{\partial\xi_l}, \tag{48}$$

where

$$q_{ij} = Q_1\epsilon_{ijl}\xi_l + Q_2\lambda_j\epsilon_{ilm}\lambda_l\xi_m + Q_3\xi_j\epsilon_{ilm}\lambda_l\xi_m. \tag{49}$$

Thus

$$\begin{aligned} Q_{ij} = {} & \{\xi_i\xi_jD_r - \delta_{ij}(r^2D_r + r\mu D_\mu + 2) + \lambda_i\xi_jD_\mu\}Q_1 \\ & + \{\xi_i\xi_jD_r - \delta_{ij}[r^2(1-\mu^2)D_r + 1] + \lambda_i\lambda_j(r^2D_r + 1) \\ & - (\lambda_i\xi_j + \xi_i\lambda_j)r\mu D_r\}Q_2 + \{-\xi_i\xi_jD_\mu + \delta_{ij}[r^2(1-\mu^2)D_\mu - r\mu] \\ & - \lambda_i\lambda_jr^2D_\mu + (\lambda_i\xi_j + \xi_i\lambda_j)r\mu D_\mu + \xi_i\lambda_j\}Q_3, \end{aligned} \tag{50}$$

where for the sake of brevity we have written

$$D_r = \frac{1}{r}\frac{\partial}{\partial r} - \frac{\mu}{r^2}\frac{\partial}{\partial\mu} \quad \text{and} \quad D_\mu = \frac{1}{r}\frac{\partial}{\partial\mu}. \tag{51}$$

The foregoing representation of Q_{ij} in terms of Q_1, Q_2, and Q_3 is unique.

If Q_{ij} is in addition symmetrical in i and j, then

$$Q_3 = D_\mu Q_1 = \frac{1}{r}\frac{\partial Q_1}{\partial \mu}, \tag{52}$$

and we have only two defining scalars.

The defining scalars of the Laplacian of Q_{ij}, which is solenoidal in j, are

$$\Delta Q_1, \qquad \Delta Q_2 + 2D_{\mu\mu}Q_1, \qquad \text{and} \qquad \Delta Q_3 + 2D_{r\mu}Q_1, \tag{53}$$

where $D_{\mu\mu} = D_\mu D_\mu$ and $D_{r\mu} = D_r D_\mu = D_\mu D_r$.

In the theory of axisymmetric turbulence, one obtains from the Stokes-Navier equation a pair of equations governing the two scalars Q_1 and Q_2 defining the fundamental correlation tensor Q_{ij}. They are (cf. Chandrasekhar [5], Secs. 8 and 9):

$$\begin{aligned}\frac{\partial Q_1}{\partial t} &= 2\nu\,\Delta Q_1 + S_1,\\ \frac{\partial Q_2}{\partial t} &= 2\nu(\Delta Q_2 + 2D_{\mu\mu}Q_1) + S_2,\end{aligned} \tag{54}$$

where S_1 and S_2 are the scalars defining

$$S_{ij} = \frac{\partial}{\partial \xi_k}\left(\overline{u_i u_k u_j'} - \overline{u_i u_k' u_j'}\right) + \frac{1}{\rho}\left(\frac{\partial}{\partial \xi_i}\overline{p u_j'} - \frac{\partial}{\partial \xi_j}\overline{p' u_i}\right). \tag{55}$$

Equations (54) replace the equation of von Kármán and Howarth in the theory of isotropic turbulence.

Equations (54) have been used to discuss the decay of turbulence during the last stages of the decay when the inertial term in the Stokes-Navier equation can be neglected [7].

3. Turbulence in magneto-hydrodynamics. A problem which has, in recent years, gained considerable importance in various geophysical and astrophysical contexts is the question of the growth of stray magnetic fields in a highly conducting turbulent fluid.

If we consider an incompressible fluid with a high electrical conductivity σ, then with suitable approximations the equations governing the velocity and the magnetic fields can be written in the forms [8–10]

$$\frac{\partial u_i}{\partial t} + \frac{\partial}{\partial x_k}(u_i u_k - h_i h_k) = -\frac{1}{\rho}\frac{\partial}{\partial x_i}\left(p + \frac{1}{2}\rho|\mathbf{h}|^2\right) + \nu\,\nabla^2 u_i \tag{56}$$

and

$$\frac{\partial h_i}{\partial t} + \frac{\partial}{\partial x_k}(h_i u_k - u_i h_k) = \lambda\,\nabla^2 h_i, \tag{57}$$

where $\lambda = 1/(4\pi\mu\sigma)$ (μ being the magnetic permeability) and $\mathbf{h}$ is related to the magnetic field $\mathbf{H}$ by

$$\mathbf{h} = \left(\frac{\mu}{4\pi\rho}\right)^{\frac{1}{2}}\mathbf{H}. \tag{58}$$

Defined in this manner, $\mathbf{h}$ has the dimensions of a velocity.

Both **u** and **h** are solenoidal vectors; however, **h**, unlike **u** but like the vorticity $\boldsymbol{\omega}$, is an axial vector.

It will be seen that equation (57) is formally the same as that governing the vorticity in an incompressible fluid with λ playing the role of ν. Therefore if $\lambda = 0$ (*i.e.*, in the case of infinite conductivity), the circulation theorems of Kelvin and Helmholtz on the vorticity in an inviscid fluid can be translated in terms of **H**: *e.g.*, the statement that "the vortex lines move with the fluid" has now the counterpart that "the magnetic lines move with the fluid."

It is evident that by defining various double and triple correlations we can treat equations (56) and (57) in a manner analogous to the Stokes-Navier equation in Secs. 1 and 2. (For the sake of simplicity we shall restrict ourselves to the case of homogeneous isotropic turbulence; the extension to the case of axisymmetric turbulence is straightforward and can be carried out if necessary.) Thus, from equation (56) we can derive [cf. equations (31) to (37)]

$$\frac{\partial Q_{ij}}{\partial t} = 2\frac{\partial}{\partial \xi_k}(T_{ikj} - S_{ikj}) + 2\nu\,\nabla^2 Q_{ij}, \tag{59}$$

where Q_{ij} and T_{ijk} denote the double $(\overline{u_i u_j'})$ and the triple $(\overline{u_i u_j u_k'})$ correlations, respectively, and

$$S_{ijk} = \overline{h_i h_j u_k'}. \tag{60}$$

Since S_{ijk} is symmetrical in i and j and solenoidal in k, it can be defined, like T_{ijk}, in terms of a single scalar; thus,

$$S_{ijk} = \operatorname{curl} S(\xi_i \epsilon_{jkl}\xi_l + \xi_j \epsilon_{ikl}\xi_l). \tag{61}$$

In terms of the defining scalars, equation (59) becomes [cf. equation (39)]

$$\frac{\partial Q}{\partial t} = 2\left(r\frac{\partial}{\partial r} + 5\right)(T - S) + 2\nu\left(\frac{\partial^2}{\partial r^2} + \frac{4}{r}\frac{\partial}{\partial r}\right)Q. \tag{62}$$

This is the generalization of the von Kármán-Howarth equation (39) to magneto-hydrodynamics. And like the von Kármán-Howarth equation, equation (62) also admits the invariant

$$\int_0^\infty Q(r,t)r^4\,dr = \text{const.} \tag{63}$$

From the continued existence of this invariant in magneto-hydrodynamics, we can conclude that no transfer of energy from the velocity field to the magnetic field takes place among the largest eddies present.

Considering equation (57), we can derive two further equations governing the defining scalars, H and R, of the correlations $\overline{h_i h_j'}$ and $\overline{u_i h_j'}$. The former correlation should clearly be of the form

$$\overline{h_i h_j'} = \operatorname{curl} H\epsilon_{ijl}\xi_l = \frac{H'}{r}\xi_i\xi_j - (rH' + 2H)\delta_{ij}. \tag{64}$$

On the other hand, $\overline{u_i h'_j}$ must be a skew tensor since, if $\mathbf{h}$ is assumed to be an axial vector, accordingly, on this assumption,[1] it must be of the form

$$\overline{u_i h'_j} = R\epsilon_{ijl}\xi_l. \tag{65}$$

The equations governing H and R are found to be [9]

$$\frac{\partial H}{\partial t} = 2P + 2\lambda\left(\frac{\partial^2}{\partial r^2} + \frac{4}{r}\frac{\partial}{\partial r}\right)H \tag{66}$$

and

$$\frac{\partial R}{\partial t} = \left(r\frac{\partial}{\partial r} + 5\right)(U - V) + \left(\frac{\partial^2}{\partial r^2} + \frac{4}{r}\frac{\partial}{\partial r}\right)[(\lambda + \nu)R - W], \tag{67}$$

where P, U, V, and W are the defining scalars of the following correlations:

$$\begin{aligned} \overline{u_i u_j h'_k} &= U(\xi_i\epsilon_{jkl}\xi_l + \xi_j\epsilon_{ikl}\xi_l), \\ \overline{h_i h_j h'_k} &= V(\xi_i\epsilon_{jkl}\xi_l + \xi_j\epsilon_{ikl}\xi_l), \\ \overline{(h_i u_j - h_j u_i)h'_k} &= P(\xi_i\delta_{jk} - \xi_j\delta_{ik}), \\ \overline{u_i(h'_j u'_k - h'_k u'_j)} &= (2W + rW')\epsilon_{ijk} - \frac{W'}{r}\xi_i\epsilon_{jkl}\xi_l. \end{aligned} \tag{68}$$

It will be seen that the equation for R allows an invariant of the Loitsiansky type. Also, from the fact that Q admits a Loitsiansky invariant, it can be concluded that H must also admit a Loitsiansky invariant, which must in fact be zero [10].

By considering the behavior of equations (62) and (66) for $r \to 0$, we can derive the equations governing the rate of dissipation of energy. Thus, we find [8, 9]

$$\begin{aligned} \frac{1}{2}\frac{d}{dt}\overline{|\mathbf{u}|^2} &= -7.5\,\overline{h_1^2\frac{\partial u_1}{\partial x_1}} - 60\nu Q_2, \\ \frac{1}{2}\frac{d}{dt}\overline{|\mathbf{h}|^2} &= +7.5\,\overline{h_1^2\frac{\partial u_1}{\partial x_1}} - 60\lambda H_2, \end{aligned} \tag{69}$$

where Q_2 and H_2 define the curvatures of the longitudinal correlations $\overline{u_1 u'_1}$ and $\overline{h_1 h'_1}$ at $r = 0$. According to equations (69) the dissipation of the kinetic energy consists of two terms: the dissipation into heat by viscosity and the transformation of the kinetic energy into magnetic energy by the stretching of the magnetic lines of force; and this gain in the magnetic energy appears in the equation for $\overline{|\mathbf{h}|^2}$. The term in λ in the latter equation corresponds to the loss of magnetic energy by Joule heating.

[1] This assumption is not essential; for if on the contrary $\mathbf{h}$ is assumed to be a polar vector, then $\overline{u_i h'_j}$ must be of the form curl $R\epsilon_{ijl}\xi_l$, and similarly the correlations on the left-hand sides of (68) will be expressible as curls of the quantities on the right-hand side. With these redefinitions equation (67) governing R will continue to be valid. This is as it should be, since none of the physical results derived can depend on whether $\mathbf{h}$ is axial or polar.

By adding the equations (69) we have

$$\frac{1}{2}\frac{d}{dt}(\overline{|\mathbf{u}|^2} + \overline{|\mathbf{h}|^2}) = -60(\nu Q_2 + \lambda H_2). \tag{70}$$

The physical meaning of this equation is that the rate of loss of energy from the system is entirely due to dissipation either by viscosity in the form of molecular motion or by conductivity in the form of Joule heat.

4. The fluctuations of density in homogeneous isotropic turbulence in a compressible fluid. With few exceptions current discussions relating to the statistical theory of turbulence have been restricted to incompressible fluids. The principal simplification which this assumption introduces is, of course, the solenoidal property of the velocity and therefore also of the various correlations considered in the theory. This essential simplification is lost when we go to compressible fluids, and this fact also explains why attempts to treat turbulence in terms of the same types of velocity correlations have not proved very successful. However, it appears that, by shifting our attention to correlations in the fluctuations of density, we can get some insight into the processes taking place in turbulence in a compressible fluid (see Chandrasekhar [11, 12]). For example, from the equation of continuity alone we can derive an invariant analogous to the Loitsiansky invariant in the theory of turbulence of an incompressible fluid.

Now the equation of continuity for a compressible fluid is

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_i}(\rho u_i) = 0. \tag{71}$$

From this equation we can readily derive by the usual methods that

$$\frac{\partial}{\partial t}\overline{\rho\rho'} + \frac{\partial}{\partial x_i}\overline{\rho' \rho u_i} + \frac{\partial}{\partial x_i'}\overline{\rho\rho' u_i'} = 0. \tag{72}$$

In homogeneous isotropic turbulence $\overline{\rho\rho'}$ is a scalar function depending, apart from time, only on the distance r between the points x_i and x_i' considered. However, instead of $\overline{\rho\rho'}$ it is convenient to define

$$\overline{\rho\rho'} - \bar{\rho}^2 = \bar{\omega}(r,t), \tag{73}$$

since defined in this manner $\bar{\omega} \to 0$ as $r \to \infty$. Also, since in homogeneous turbulence the mean density $\bar{\rho}$ is a constant independent of position and time, $\bar{\omega}$ differs from $\overline{\rho\rho'}$ only by an additive constant. An equivalent definition of $\bar{\omega}$ is

$$\bar{\omega} = \overline{(\rho - \bar{\rho})(\rho' - \bar{\rho})} = \overline{\delta\rho\, \delta\rho'}, \tag{74}$$

where $\delta\rho$ is the instantaneous fluctuation in the density from the mean.

Again, in isotropic turbulence

$$\overline{\rho' \rho u_i} = -\overline{\rho\rho' u_i'} = L(r,t)\xi_i, \tag{75}$$

where L is a function depending on r and t.

Combining equations (72), (73), and (75) and remembering (34), we obtain

$$\frac{\partial \bar{\omega}}{\partial t} = 2\frac{\partial}{\partial \xi_i} L\xi_i = 2\left(r\frac{\partial L}{\partial r} + 3L\right). \tag{76}$$

Equation (76) allows us to derive an invariant. Rewriting it in the form

$$r^2\frac{\partial \bar{\omega}}{\partial t} = \frac{\partial}{\partial r}(r^3 L), \tag{77}$$

and integrating it from 0 to r, we have

$$\frac{\partial}{\partial t}\int_0^r r^2\bar{\omega}\, dr = 2r^3 L. \tag{78}$$

If $L \to 0$ faster than r^{-3} does as $r \to \infty$, it follows from (78) that

$$\frac{\partial}{\partial t}\int_0^\infty r^2\bar{\omega}(r,t)\, dr = 0, \tag{79}$$

or,

$$\int_0^\infty r^2\bar{\omega}(r,t)\, dr = \text{const.} \tag{80}$$

As in the case of the Loitsiansky invariant, the meaning of this new invariant is that the large-scale components of the density fluctuations are permanent features of the system. More particularly, if a Fourier analysis of the density fluctuations is made and if $\Pi(k)$ denotes the spectrum of $\overline{|\delta\rho|^2}$, then for $k \to 0$, $\Pi(k)$ has the behavior $\Pi_0 k^2$, where Π_0 is a constant depending only on the initial conditions of the problem [11].

An equation of motion for $\bar{\omega}(r,t)$ can be derived from the equation

$$\frac{\partial^2 \rho}{\partial t^2} = \frac{\partial^2}{\partial x_i\, \partial x_j}(\rho u_i u_j) + \nabla^2(p - \bar{p}) - \frac{4}{3}\mu\nabla^2\frac{\partial u_j}{\partial x_j}, \tag{81}$$

which follows from the Stokes-Navier equation for a compressible fluid. Thus, using equations (71) and (81) in the relation

$$\frac{\partial^2 \bar{\omega}}{\partial t^2} = \frac{\partial^2}{\partial t^2}\overline{\rho\rho'} = \overline{\rho'\frac{\partial^2 \rho}{\partial t^2}} + \overline{\rho\frac{\partial^2 \rho'}{\partial t^2}} + 2\overline{\frac{\partial \rho}{\partial t}\frac{\partial \rho'}{\partial t}}, \tag{82}$$

we readily obtain

$$\frac{\partial^2 \bar{\omega}}{\partial t^2} = 2\frac{\partial^2}{\partial \xi_i\, \partial \xi_j}[\overline{\rho\rho'(u_i u_j - u_i u_j')}] + 2\nabla^2[\overline{\rho'(p - \bar{p})}] - \frac{8}{3}\mu\nabla^2\frac{\partial}{\partial \xi_j}\overline{\rho u_j'}. \tag{83}$$

where now

$$\nabla^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right). \tag{84}$$

Now we may suppose that the fourth-order correlations occurring in equation (83) can be expressed in terms of the second-order correlations, as in a joint

Gaussian distribution. A similar assumption has been made in evaluating the fluctuation of pressure in a turbulent incompressible fluid, with apparently satisfactory results [13]. On this assumption, we can replace the two fourth-order correlations in equation (83) by

$$\overline{\rho\rho' u_i u_j} = \overline{\rho\rho'}\,\overline{u_i u_j} + \overline{\rho' u_i}\,\overline{\rho u_j} + \overline{\rho' u_j}\,\overline{\rho u_i} = \tfrac{1}{3}\overline{\rho\rho'}\,\overline{u^2}\delta_{ij}, \tag{85}$$

where $\overline{u^2}$ denotes the mean-square velocity of turbulence and

$$\overline{\rho\rho' u_i u_j'} = \overline{\rho\rho'}\,\overline{u_i u_j'} + \overline{\rho' u_i}\,\overline{\rho u_j'} + \overline{\rho u_i}\,\overline{\rho' u_j'} = \overline{\rho\rho'}\,\overline{u_i u_j'} - \overline{\rho u_i'}\,\overline{\rho u_j'}. \tag{86}$$

With the foregoing substitutions equation (83) reduces to

$$\frac{\partial^2 \bar{\omega}}{\partial t^2} = 2\,\nabla^2\overline{(p-\bar{p})\rho'} + \frac{2}{3}\overline{u^2}\,\nabla^2\bar{\omega} - 2\frac{\partial^2}{\partial\xi_i\,\partial\xi_j}\left(\overline{\rho\rho'}\,\overline{u_i u_j'} - \overline{\rho u_i'}\,\overline{\rho u_j'}\right) - \frac{8}{3}\mu\,\nabla^2\frac{\partial}{\partial\xi_j}\overline{\rho u_j'}. \tag{87}$$

We shall now further suppose that the variations in pressure and density which are continually taking place in the medium take place, at each point, adiabatically, *i.e.*,

$$\frac{\delta p}{\bar{p}} = \gamma\frac{\delta\rho}{\bar{\rho}}, \tag{88}$$

where γ is the ratio of the specific heats. Consistent with this assumption, we should strictly ignore the term in viscosity in equation (87). Also we can replace

$$\overline{(p-\bar{p})\rho'} \qquad \text{by} \qquad \gamma\frac{\bar{p}}{\bar{\rho}}\,\overline{\delta\rho(\bar{\rho}+\delta\rho')} = c^2\bar{\omega}(r,t), \tag{89}$$

where c denotes the velocity of sound appropriate to the mean pressure and density. With these simplifications equation (83) reduces to

$$\frac{\partial^2\bar{\omega}}{\partial t^2} = 2\left(c^2 + \frac{1}{3}\overline{u^2}\right)\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\bar{\omega}}{\partial r}\right) - 2\frac{\partial^2}{\partial\xi_i\,\partial\xi_j}\left[\overline{\rho\rho'}\,\overline{u_i u_j'} - \overline{\rho u_i'}\,\overline{\rho u_j'}\right]. \tag{90}$$

On the assumption that $\overline{u_i u_j'}$ and $\overline{\rho u_i'}$ tend to zero sufficiently rapidly as $r \to \infty$, it would follow from equation (90) that, under conditions of local isotropy [*i.e.*, under *conditions* when mean-square relative velocities such as $\overline{(u_i - u_j')^2}$ are independent of time], it will allow periodic solutions whose behavior at infinity will be given by

$$\bar{\omega}(r,t) = \frac{\text{const.}}{r}\exp i\sigma\left[t + \frac{r}{(2c^2 + \frac{2}{3}\overline{u^2})^{\frac{1}{2}}}\right]. \tag{91}$$

In other words the fluctuations in density are propagated over large distances in the medium with a velocity $(2c^2 + \frac{2}{3}\overline{u^2})^{\frac{1}{2}}$, and each scale of density fluctuation varies periodically with its own characteristic period, independently of the others.

5. The gravitational instability of an infinite homogeneous turbulent medium. The manner of treating density fluctuations described in the preceding section has an application to the problem of the gravitational stability of an infinite homogeneous turbulent medium (cf. Chandrasekhar [12]; for the earlier treatment of the subject ignoring turbulence, see Jeans [14]). We shall briefly describe this application in this section.

We picture to ourselves an infinite homogeneous medium in which the gravitational effects of the fluctuations in density are important. Under these circumstances equation (81) of Sec. 4 is replaced by

$$\text{(92)} \qquad \frac{\partial^2 \rho}{\partial t^2} = \frac{\partial^2}{\partial x_i \, \partial x_j} (\rho u_i u_j) + \nabla^2 (p - \bar{p}) - \frac{\partial}{\partial x_i} \left(\rho \frac{\partial V}{\partial x_i} \right) - \frac{4}{3} \mu \nabla^2 \frac{\partial u_j}{\partial x_j},$$

where V denotes the gravitational potential.

In the problem we are considering, we may write

$$\text{(93)} \qquad p = \bar{p} + \delta p, \qquad \rho = \bar{\rho} + \delta\rho, \qquad \text{and} \qquad V = \bar{V} + \delta V,$$

where $\bar{p}$, $\bar{\rho}$, and $\bar{V}$ are certain constants. With these substitutions the term in the gravitational potential in equation (92) becomes

$$\text{(94)} \qquad \frac{\partial}{\partial x_i} \left[(\bar{\rho} + \delta\rho) \frac{\partial}{\partial x_i} \delta V \right] \simeq \bar{\rho} \nabla^2 \delta V = -4\pi G \bar{\rho} \, \delta\rho,$$

where we have neglected quantities of the second order in $\delta\rho$ and further made use of the variation of Poisson's equation governing the gravitational potential. Using the result of equation (94) in equation (92), we have

$$\text{(95)} \qquad \frac{\partial^2 \rho}{\partial t^2} = \frac{\partial^2}{\partial x_i \, \partial x_j} (\rho u_i u_j) + \nabla^2 \delta p + 4\pi G \bar{\rho} \, \delta\rho - \frac{4}{3} \mu \nabla^2 \frac{\partial u_j}{\partial x_j}.$$

Introducing the correlation function $\bar{\omega}(r,t)$ [cf. equation (74)], using equation (95) instead of equation (81), making the same two simplifying assumptions expressed by equations (85), (86), and (89), and ignoring the term in viscosity, we obtain

$$\text{(96)} \quad \frac{\partial^2 \bar{\omega}}{\partial t^2} = 2 \left(c^2 + \frac{1}{3} \overline{u^2} \right) \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \bar{\omega}}{\partial r} \right) + 8\pi G \bar{\rho} \bar{\omega} - 2 \frac{\partial^2}{\partial \xi_i \, \partial \xi_j} (\overline{\rho\rho'} \, \overline{u_i u_j'} - \overline{\rho u_i'} \, \overline{\rho u_j'}).$$

For sufficiently large values of r the terms in the velocity correlations in equation (96) will become negligible, and the equation will tend to

$$\text{(97)} \qquad \frac{\partial^2 \bar{\omega}}{\partial t^2} = 2 \left(c^2 + \frac{1}{3} \overline{u^2} \right) \frac{1}{r^2} \frac{\partial}{\partial r} \left(r^2 \frac{\partial \bar{\omega}}{\partial r} \right) + 8\pi G \bar{\rho} \bar{\omega}.$$

But this equation admits spherical wave solutions of the form

$$\tilde{\omega}(r,t) = \frac{A(t)}{r} e^{ikr}. \tag{98}$$

Accordingly, we may expect that a superposition of these solutions will represent the asymptotic behavior of the solution of equation (96). On the other hand, the equation determining the amplitudes (at infinity) of these waves is

$$\frac{d^2A}{dt^2} = -2\left[\left(c^2 + \frac{1}{3}\overline{u^2}\right)k^2 - 4\pi G\bar{\rho}\right]A. \tag{99}$$

Consequently if

$$k^2 < \frac{4\pi G\bar{\rho}}{c^2 + \frac{1}{3}\overline{u^2}}, \tag{100}$$

the amplitudes of the corresponding spherical wave in the superposition will increase exponentially with time. In other words, eddies in the density fluctuations which are larger than a certain critical size will be amplified; in this we may see a tendency toward disintegration and formation of condensation in the medium.

Bibliography

1. G. I. Taylor, *Statistical theory of turbulence*, I–IV, Proc. Roy. Soc. London Ser. A vol. 151 (1935) pp. 421–478.

2. T. von Kármán and L. Howarth, *On the statistical theory of isotropic turbulence*, Proc. Roy. Soc. London Ser. A vol. 164 (1938) pp. 192–215.

3. H. P. Robertson, *The invariant theory of isotropic turbulence*, Proc. Cambridge Philos. Soc. vol. 36 (1940) pp. 209–223.

4. G. K. Batchelor, *The role of big eddies in homogeneous turbulence*, Proc. Roy. Soc. London Ser. A vol. 195 (1949) pp. 513–532.

5. S. Chandrasekhar, *The theory of axisymmetric turbulence*, Philos. Trans. Roy. Soc. London Ser. A vol. 242 (1950) pp. 557–577.

6. G. K. Batchelor, *The theory of axisymmetric turbulence*, Proc. Roy. Soc. London Ser. A vol. 186 (1946) pp. 480–502.

7. S. Chandrasekhar, *The decay of axisymmetric turbulence*, Proc. Roy. Soc. London Ser. A vol. 203 (1950) pp. 358–364.

8. G. K. Batchelor, *On the spontaneous magnetic field in a conducting liquid in turbulent motion*, Proc. Roy. Soc. London Ser. A vol. 201 (1950) pp. 405–416.

9. S. Chandrasekhar, *The invariant theory of isotropic turbulence in magneto-hydrodynamics*, I, Proc. Roy. Soc. London Ser. A vol. 204 (1950) pp. 435–449.

10. ———, *The invariant theory of isotropic turbulence in magneto-hydrodynamics*, II, Proc. Roy. Soc. London Ser. A vol. 207 (1951) pp. 301–306.

11. ———, *The fluctuations of density in isotropic turbulence*, Proc. Roy. Soc. London Ser. A vol. 210 (1951) pp. 18–25.

12. ———, *The gravitational instability of an infinite homogeneous turbulent medium*, Proc. Roy. Soc. London Ser. A vol. 210 (1951) pp. 26–29.

13. G. K. Batchelor, *Pressure fluctuations in isotropic turbulence*, Proc. Cambridge Philos. Soc. vol. 47 (1951) pp. 359–374.

14. J. H. Jeans, *Astronomy and cosmogony*, Cambridge University Press, London, 1929, pp. 345–349.

Yerkes Observatory, University of Chicago,
Williams Bay, Wis.

MAGNETIC FIELDS IN SPIRAL ARMS

S. Chandrasekhar and E. Fermi
University of Chicago
Received March 23, 1953

ABSTRACT

In this paper two independent methods are described for estimating the magnetic field in the spiral arm in which we are located. The first method is based on an interpretation of the dispersion (of the order of 10°) in the observed planes of polarization of the light of the distant stars; it leads to an estimate of $H = 7.2 \times 10^{-6}$ gauss. The second method is based on the requirement of equilibrium of the spiral arm with respect to lateral expansion and contraction: it leads to an estimate of $H = 6 \times 10^{-6}$ gauss.

The hypothesis of the existence of a magnetic field in galactic space[1] has received some confirmation by Hiltner's[2] observation of the polarization of the light of the distant stars. It seems plausible that this polarization is due to a magnetic orientation of the interstellar dust particles;[3] for such an orientation would lead to different amounts of absorption of light polarized parallel and perpendicular to the magnetic field and, therefore, to a polarization of the light reaching us. On this interpretation of the interstellar polarization we should expect to observe no polarization in the general direction of the magnetic lines of force and a maximum polarization in a direction normal to the lines of force. And if we interpret from this point of view the maps[4] of the polarization effect as a function of the direction of observation, it appears that the direction of the galactic magnetic field is roughly parallel to the direction of the spiral arm in which we are located. In this paper we shall discuss some further consequences of this interpretation of interstellar polarization, in an attempt to arrive at an estimate of the strength of the interstellar magnetic field.

As we observe distant stars in a direction approximately perpendicular to the spiral arm, it appears that the direction of polarization is only approximately parallel to the arm. There are indeed quite appreciable and apparently irregular fluctuations in the direction of polarization of the distant stars.[4] This would indicate that the magnetic lines of force are not strictly straight and that they may be better described as "wavy" lines. The mean angular deviation of the plane of polarization from the direction of the spiral arm appears to be about $\alpha = 0.2$ radians.[4] There must clearly be a relation between this angle, α, and the strength of the magnetic field, H. For, if the magnetic field were sufficiently strong, the lines of force would be quite straight and α would be very small; on the other hand, if the magnetic field were sufficiently weak, the lines of force would be dragged around in various directions by the turbulent motions of the gas masses in the spiral arm and α would be large. To obtain the general relation between α and H, we proceed as follows:

The velocity of the transverse magneto-hydrodynamic wave is given by

$$V = \frac{H}{\sqrt{(4\pi\rho)}}, \tag{1}$$

[1] E. Fermi, *Phys. Rev.*, **75**, 1169, 1949.

[2] W. A. Hiltner, *Ap. J.*, **109**, 471, 1949.

[3] Of the two theories which have been proposed (L. Spitzer and J. W. Tukey, *Ap. J.*, **114**, 187, 1951, and L. Davis and J. L. Greenstein, *Ap. J.*, **114**, 206, 1951), that by Davis and Greenstein appears to be in better accord with the facts.

[4] W. A. Hiltner, *Ap. J.*, **114**, 241, 1951.

where ρ is the density of the diffused matter. In computing the velocity, V, we should not include in ρ the average density due to the stars, since the stars may be presumed to move across the lines of force without appreciable interaction with them, whereas the diffused matter in the form of both gas and dust has a sufficiently high electrical conductivity to be effectively attached to the magnetic lines of force in such a way that only longitudinal relative displacements are possible.

According to equation (1), the transverse oscillations of a particular line of force can be described by an equation of the form

$$y = a \cos k (x - Vt), \tag{2}$$

where x is a longitudinal co-ordinate and y represents the lateral displacement. We take the derivatives of y with respect to x and t and obtain

$$y' = -ak \sin k (x - Vt)$$

and (3)

$$\dot{y} = -akV \sin k (x - Vt).$$

From these equations it follows that

$$V^2\overline{y'^2} = \overline{\dot{y}^2}. \tag{4}$$

The lateral velocity of the lines of force must be equal to the lateral velocity of the turbulent gas. If v denotes the root-mean-square velocity of the turbulent motion, we should have

$$\overline{\dot{y}^2} = \tfrac{1}{3} v^2. \tag{5}$$

The factor $\frac{1}{3}$ arises from the fact that only one component of the velocity is effective in shifting the lines of force in the y-direction. The quantity y', on the other hand, represents the deviation of the line of force from a straight line projected on the plane of view. Hence,

$$\overline{y'^2} = a^2. \tag{6}$$

Now, combining equations (1), (4), (5), and (6), we obtain

$$H = (\tfrac{4}{3}\pi\rho)^{1/2} \frac{v}{a}. \tag{7}$$

In equation (7) we shall substitute the following numerical values, which appear to describe approximately the conditions prevailing in the spiral arm in which we are located:[5]

$$\rho = 2 \times 10^{-24}\ \text{gm/cm}^3,\ v = 5 \times 10^5\ \text{cm/sec, and } a = 0.2 \text{ radians.} \tag{8}$$

With these values equation (7) gives

$$H = 7.2 \times 10^{-6}\ \text{gauss.} \tag{9}$$

An alternative procedure for estimating the intensity of the magnetic field is based on the requirement of equilibrium of the spiral arm with respect to lateral expansion and contraction. As an order of magnitude, we may expect to obtain the condition for this equilibrium by equating the gravitational pressure in the spiral arm to the sum of the material pressure and the pressure due to the magnetic field. In computing the gravita-

[5] For ρ, the estimate of J. H. Oort (cf. *Ap. J.*, **116**, 233, 1952) from observations of the 21-cm line is used; while the value of v adopted is that of A. Blaauw, *B.A.N.*, **11**, 405, 1952.

tional pressure, we should allow for the gravitational force due to all the mass present, i.e., of the stars as well as of the diffused matter. We are interested, however, in computing the gravitational pressure exerted on the diffused matter only. Assuming for simplicity that the spiral arm is a cylinder of radius R with uniform density, one finds for the gravitational pressure:

$$p_{\text{grav}} = \pi G \rho \rho_t R^2 , \tag{10}$$

where G denotes the constant of gravitation, ρ is the density of the diffused matter only, and ρ_t is the total mean density, including the contribution of the stars. The kinetic pressure of the turbulent gas is given by

$$p_{\text{kin}} = \tfrac{1}{3} \rho v^2 \tag{11}$$

while the magnetic pressure is given by

$$p_{\text{mag}} = \frac{H^2}{8\pi} . \tag{12}$$

And for the equilibrium we must have

$$p_{\text{grav}} = p_{\text{kin}} + p_{\text{mag}} . \tag{13}$$

In computing p_{grav} we shall assume a radius of the spiral arm of 250 parsecs or $R = 7.7 \times 10^{20}$ cm. As before, we shall take $\rho = 2 \times 10^{-24}$ gm/cm^3; and for ρ_t we shall assume[6] 6×10^{-24} gm/cm^3. For these values of R, ρ, and ρ_t equation (9) gives $p_{\text{grav}} = 1.5 \times 10^{-12}$ dynes, while p_{kin} computed with the values already given is 0.2×10^{-12} dynes. We attribute the difference to the magnetic pressure. Hence

$$\frac{H^2}{8\pi} = 1.3 \times 10^{-12} , \tag{14}$$

or

$$H = 6 \times 10^{-6} \text{ gauss.} \tag{15}$$

The two independent methods of estimating H therefore agree in giving essentially the same value for the field strength. A field of about 7×10^{-6} gauss indicated by these estimates is ten times smaller than that which Davis and Greenstein[3] have estimated as necessary for producing an adequate orientation of the dust particles to account for the interstellar polarization. If the present estimate of 7×10^{-6} gauss is correct, one should conclude that the mechanism of orientation is somewhat more effective than has been assumed by Davis and Greenstein.

Since this paper was written, our attention has been drawn to the fact that the idea underlying the first of the two methods by which we estimate the magnetic field in the spiral arm is contained in an earlier paper by Leverett Davis, Jr. (*Phys. Rev.*, **81**, 890, 1951). We are sorry that we were not aware of this paper when we wrote ours. However, since with the better estimates of the astronomical parameters now available the value of H derived is a great deal different from Davis' value and since further the value we have derived is in accord with our second independent estimate, we have allowed the paper to stand in its original form.

[6] Cf. J. H. Oort, *Ap. J.*, **116**, 233, 1952.

PROBLEMS OF GRAVITATIONAL STABILITY IN THE PRESENCE OF A MAGNETIC FIELD

S. Chandrasekhar and E. Fermi
University of Chicago
Received March 23, 1953

ABSTRACT

In this paper a number of problems are considered which are related to the gravitational stability of cosmical masses of infinite electrical conductivity in which there is a prevalent magnetic field. In Section I the virial theorem is extended to include the magnetic terms in the equations of motion, and it is shown that when the magnetic energy exceeds the numerical value of the gravitational potential energy, the configuration becomes dynamically unstable. It is suggested that the relatively long periods of the magnetic variables may be due to the magnetic energy of these stars approaching the limit set by the virial theorem. In Section II the adiabatic radial pulsations of an infinite cylinder along the axis of which a magnetic field is acting is considered. An explicit expression for the period is obtained. Section III is devoted to an investigation of the stability for transverse oscillations of an infinite cylinder of incompressible fluid when there is a uniform magnetic field acting in the direction of the axis. It is shown that the cylinder is unstable for all periodic deformations of the boundary with wave lengths exceeding a certain critical value, depending on the strength of the field. The wave length of maximum instability is also determined. It is found that the magnetic field has a stabilizing effect both in increasing the wave length of maximum instability and in prolonging the time needed for the instability to manifest itself. For a cylinder of radius $R = 250$ parsecs and $\rho = 2 \times 10^{-24}$ gm/cm^3 a magnetic field in excess of 7×10^{-6} gauss effectively removes the instability. In Section IV it is shown that a fluid sphere with a uniform magnetic field inside and a dipole field outside is not a configuration of equilibrium and that it will tend to become oblate by contracting in the direction of the field. Finally, in Section V the gravitational instability of an infinite homogeneous medium in the presence of a magnetic field is considered, and it is shown that Jeans's condition is unaffected by the presence of the field.

1. *Introduction.*—In this paper we shall consider a number of problems relating to the dynamical and gravitational stability of cosmical masses in which there is a prevalent magnetic field. In the discussion of these problems, the assumption will be made that the medium is effectively of infinite electrical conductivity. This latter assumption implies only that the conductivity is large enough for the magnetic lines of force to be considered as practically attached to the matter during the length of time under consideration; it has been known for some time that this is the case in most astronomical connections.[1]

The abstract gives an adequate summary of the paper.

I. THE VIRIAL THEOREM AND THE CONDITION FOR DYNAMICAL STABILITY

2. *The virial theorem.*—In a subject such as this it is perhaps best that we start by establishing theorems of the widest possible generality. The extension of the virial theorem to include the forces derived from the prevailing magnetic field provides such a starting point. We shall see that under conditions of equilibrium this extension of the virial theorem leads to the relation

$$2T + 3(\gamma - 1)\mathfrak{U} + \mathfrak{M} + \Omega = 0 \tag{1}$$

between the kinetic energy (T) of mass motion, the heat energy ($\mathfrak{U}$) of molecular motion, the magnetic energy ($\mathfrak{M}$) of the prevailing field, and the gravitational potential energy (Ω), where γ denotes the ratio of the specific heats. That a relation of the form (1) should exist is readily understood: For the balance between the pressures p_{kin}, p_{gas}, and p_{mag} due

[1] Cf. L. Biermann, *Annual Review of Nuclear Science*, 2 (Stanford: Annual Reviews, Inc., 1953) 349.

to the visible motions, the molecular motions, and the magnetic field, on the one hand, and the gravitational pressure, $p_{\rm grav}$, on the other, requires

$$p_{\rm kin} + p_{\rm gas} + p_{\rm mag} = p_{\rm grav}\,, \tag{2}$$

while the order of magnitudes of these pressures are given by

$$p_{\rm kin} = c_1 \frac{T}{V}\,, \qquad p_{\rm gas} = c_2 \frac{\mathfrak{U}}{V}\,, \qquad p_{\rm mag} = \frac{H^2}{8\pi} = c_3 \frac{\mathfrak{M}}{V}\,, \tag{3}$$

and

$$p_{\rm grav} = \text{Density} \times \text{gravity} \times \text{linear dimension} = -\,c_4 \frac{\Omega}{V}\,, \tag{3a}$$

where V denotes the volume of the configuration and c_1, c_2, c_3, and c_4 are numerical constants. A relation of the form (1) is therefore clearly implied. We now proceed to establish the exact relation (1).

With the usual assumptions of hydromagnetics, the equations of motion governing an inviscid fluid can be written in the form

$$\rho \frac{du_i}{dt} = -\frac{\partial}{\partial x_i}\left(p + \frac{|\boldsymbol{H}|^2}{8\pi}\right) + \rho \frac{\partial V}{\partial x_i} + \frac{1}{4\pi} \frac{\partial}{\partial x_j} H_i H_j\,, \tag{4}$$

where ρ denotes the density, p the pressure, V the gravitational potential, and $\boldsymbol{H}$ the intensity of the magnetic field. (In eq. [1] and in the sequel, summation over repeated indices is to be understood.)

Multiply equation (4) by x_i and integrate over the volume of the configuration. Reducing the left-hand side of the equation in the usual manner, we find

$$\begin{aligned} \iiint \rho x_i \frac{du_i}{dt}\, dx_1 dx_2 dx_3 &= \int_0^M x_i \frac{d^2 x_i}{dt^2}\, dm \\ &= \frac{1}{2} \frac{d^2}{dt^2} \int_0^M r^2 dm - \int_0^M |u|^2 dm\,, \end{aligned} \tag{5}$$

where $dm = \rho dx_1 dx_2 dx_3$ and the integration is effected over the entire mass, M, of the configuration. Letting

$$I = \int_0^M r^2 dm \qquad \text{and} \qquad T = \frac{1}{2} \int |u|^2 dm \tag{6}$$

denote the moment of inertia and the kinetic energy of mass motion, respectively, we have

$$\begin{aligned} \frac{1}{2} \frac{d^2 I}{dt^2} - 2T = &-\iiint x_i \frac{\partial}{\partial x_i}\left(p + \frac{|\boldsymbol{H}|^2}{8\pi}\right) dx_1 dx_2 dx_3 \\ &+ \frac{1}{4\pi} \iiint x_i \frac{\partial}{\partial x_j} H_i H_j dx_1 dx_2 dx_3 + \int_0^M x_i \frac{\partial V}{\partial x_i}\, dm\,. \end{aligned} \tag{7}$$

The last of the three integrals on the right-hand side of this equation represents the gravitational potential energy, Ω, of the configuration. The remaining two volume integrals can be reduced by integration by parts. Thus the first of the two integrals gives

$$\begin{aligned} &-\iiint x_i \frac{\partial}{\partial x_i}\left(p + \frac{|\boldsymbol{H}|^2}{8\pi}\right) dx_1 dx_2 dx_3 \\ &\quad = -\int \left(p + \frac{|\boldsymbol{H}|^2}{8\pi}\right) \boldsymbol{r} \cdot d\boldsymbol{S} + 3 \iiint \left(p + \frac{|\boldsymbol{H}|^2}{8\pi}\right) dx_1 dx_2 dx_3\,. \end{aligned} \tag{8}$$

The surface integral (over dS) vanishes, since the pressure (including the magnetic pressure $|H|^2/8\pi$) must vanish on the boundary of the configuration; and the volume integral over p and $|H|^2/8\pi$ is readily expressible in terms of the internal energy ($\mathfrak{U}$) and the magnetic energy ($\mathfrak{M}$) of the configuration. Thus we have

$$-\iiint x_i \frac{\partial}{\partial x_i}\left(p+\frac{|H|^2}{8\pi}\right) dx_1dx_2dx_3 = 3(\gamma-1)\,\mathfrak{U}+3\,\mathfrak{M}\,, \tag{9}$$

where γ denotes the ratio of the specific heats. In the same way the second volume integral in equation (7) gives

$$\frac{1}{4\pi}\iiint x_i \frac{\partial}{\partial x_j} H_iH_j dx_1dx_2dx_3 = -2\,\mathfrak{M}\,. \tag{10}$$

Now, combining equations (7), (9), and (10), we have

$$\frac{1}{2}\frac{d^2I}{dt^2} = 2T+3(\gamma-1)\,\mathfrak{U}+\mathfrak{M}+\Omega\,. \tag{11}$$

This is the required generalization of the virial theorem; it differs from the usual one only in the appearance of $\mathfrak{M}+\Omega$ in place of Ω.

3. *The condition for dynamical stability.*—If the configuration is one of equilibrium, then it follows from the virial theorem that

$$3(\gamma-1)\mathfrak{U}+\mathfrak{M}+\Omega=0\,. \tag{12}$$

On the other hand, the total energy, $\mathfrak{E}$, of the configuration is given by

$$\mathfrak{E}=\mathfrak{U}+\mathfrak{M}+\Omega\,. \tag{13}$$

Eliminating $\mathfrak{U}$ between equations (12) and (13), we obtain

$$\mathfrak{E}=-\frac{3\gamma-4}{3(\gamma-1)}(|\Omega|-\mathfrak{M})\,. \tag{14}$$

From this equation for the total energy it follows that *a necessary condition for the dynamical stability of an equilibrium configuration is*

$$(3\gamma-4)(|\Omega|-\mathfrak{M})>0\,. \tag{15}$$

Thus, even when $\gamma>\frac{4}{3}$ (the condition for dynamical stability in the absence of a magnetic field) a sufficiently strong internal magnetic field can induce dynamical instability in the configuration. In fact, according to formula (15), the condition for dynamical stability, when $\gamma>\frac{4}{3}$, is

$$\mathfrak{M}=\frac{1}{8\pi}\iiint |H|^2dx_1dx_2dx_3=\tfrac{1}{6}R^3(H^2)_{\text{av}}<|\Omega|\,, \tag{16}$$

where $(H^2)_{\text{av}}$ denotes the mean square magnetic field.

For a spherical configuration of uniform density,

$$\Omega=-\frac{3}{5}\frac{GM^2}{R}\,, \tag{17}$$

where M is its mass, R is its radius, and G is the constant of gravitation. We can use this expression for Ω to estimate the limit imposed by the virial theorem on the magnetic fields which can prevail. On expressing M and R in units of the solar mass and the solar radius, we find from equations (16) and (17) that

$$\sqrt{(H^2)_{\text{av}}} < 2.0 \times 10^8 \frac{M}{R^2} \text{ gauss} . \tag{18}$$

For the peculiar A stars for which Babcock has found magnetic fields of the order of 10^4 gauss, we may estimate that

$$M \simeq 4\odot \quad \text{and} \quad R \simeq 5R_\odot \quad \text{(A star)}; \tag{19}$$

and expression (18) gives

$$\sqrt{(H^2)_{\text{av}}} < 3 \times 10^7 \text{ gauss} \quad \text{(A stars)}. \tag{19a}$$

Of greater interest is the limit set by expression (18) for an S-type star for which Babcock has found a variable magnetic field of the order of 1000 gauss. For an S-type star we can estimate that

$$M \simeq 15\odot \quad \text{and} \quad R \simeq 300R_\odot \quad \text{(S star)}; \tag{20}$$

and (18) now gives

$$\sqrt{(H^2)_{\text{av}}} < 3 \times 10^4 \text{ gauss} \quad \text{(S star)}. \tag{20a}$$

Thus the limit set by (18) is seen to be two to three orders of magnitude larger than the surface fields observed by Babcock. However, the fields in the interior may be much stronger than the surface fields; and it is even possible that the actual root-mean-square fields in these stars are near their maximum values. Indeed, from the fact that the periods of the magnetic variables are long compared with the adiabatic pulsation periods they would have if they were nonmagnetic, we may surmise that $\sqrt{(H^2)_{\text{av}}}$ is near the limit set by (18); for, as is well known, we may lengthen the period of the lowest mode of oscillation of a system by approaching the limit of dynamical stability; and we can accomplish this by letting $\mathfrak{M} \to |\Omega|$.

Note added June 17.—Since we wrote this paper, Dr. Babcock has informed us that he has measured a variable magnetic field (+2000 to −1200 gauss) for the star VV Cephei. It has been estimated that for this star $M = 100\odot$ and $R = 2600R_\odot$. With these values, inequality (18) gives $\sqrt{(H^2)_{\text{av}}} < 3000$ gauss. We may conclude that this star must be on the verge of dynamical stability and, anticipating the result established in Section IV, probably highly oblate.

4. *The virial theorem for an infinite cylindrical distribution of matter.*—Some care is needed in applying the results of §§ 2 and 3 to a distribution of matter which can be idealized as an infinite cylinder (such as, for example, a spiral arm); for the potential energy per unit length of an infinite cylinder is infinite. For this reason it is perhaps best that we consider the problem *de novo.*

We shall consider, then, an infinite cylinder in which the prevailing magnetic field is in the direction of the axis of the cylinder; and we shall suppose that all the variables are functions only of the distance r from the axis of the cylinder. Under these conditions the equations of motion reduce to the single one

$$\rho \frac{du_r}{dt} = -\frac{\partial}{\partial r}\left(p + \frac{H^2}{8\pi}\right) - \frac{2Gm(r)}{r}\rho , \tag{21}$$

where $m(r)$ is the mass per unit length interior to r.

Multiplying equation (21) by $2\pi r^2 dr$ and integrating over the entire range of r, we find in the usual manner that

$$\frac{1}{2}\frac{d^2}{dt^2}\int_0^M r^2 dm - \int_0^M \left(\frac{dr}{dt}\right)^2 dm = 2(\gamma - 1)\,\mathfrak{U} + 2\mathfrak{M} - GM^2 , \tag{22}$$

where M denotes the mass per unit length of the cylinder and $\mathfrak{U}$ and $\mathfrak{M}$ are the kinetic and the magnetic energies per unit length of the cylinder, respectively.

From equation (22) it follows that, under equilibrium conditions, we should have

$$2(\gamma-1)\mathfrak{U}+2\mathfrak{M}-GM^2=0\ . \tag{23}$$

A necessary condition for equilibrium to obtain is, therefore,

$$\mathfrak{M}<\tfrac{1}{2}GM^2\ . \tag{24}$$

We can rewrite the condition (24) alternatively in the form

$$\mathfrak{M}=\frac{1}{4}\int_0^R H^2 r\,d r=\tfrac{1}{8}(H^2)_{\text{av}}R^2<\tfrac{1}{2}\pi^2R^4\bar{\rho}^2G\ , \tag{25}$$

or

$$\sqrt{(H^2)_{\text{av}}}<2\pi R\bar{\rho}G\ , \tag{25a}$$

This last condition on the root-mean-square field is essentially equivalent to one of the formulae used in the preceding paper[2] (eq. [13]) for estimating the magnetic field in the spiral arm; the difference between the two formulae arises from the fact that in that paper the gravitational attraction was not limited to the interstellar gas only; allowance was also made for the stars contributing to the gravitational force acting on the gas.

II. THE RADIAL PULSATIONS OF AN INFINITE CYLINDER

5. *The pulsation equation.*—In view of the inconclusive nature of the current treatments[3] of the adiabatic pulsations of magnetic stars, it is perhaps of interest to see how the corresponding problem in infinite cylinders can be fully solved. We consider, then, the radial pulsation of an infinite cylinder, along the axis of which there is a prevailing magnetic field.

Choosing the time t and the mass per unit length, $m(r)$, interior to r, as the independent variables, we can write the equations of continuity and motion in the forms

$$\frac{\partial}{\partial m}(\pi r^2)=\frac{1}{\rho} \tag{26}$$

and

$$\frac{\partial^2 r}{\partial t^2}=-2\pi r\frac{\partial P}{\partial m}-\frac{2Gm(r)}{r}\ , \tag{27}$$

where

$$P=p+\frac{H^2}{8\pi} \tag{28}$$

denotes the total pressure. Distinguishing the values of the various parameters for the equilibrium configuration by a subscript zero and writing

$$r=r_0+\delta r,\qquad P=P_0+\delta P,\qquad \rho=\rho_0+\delta\rho,\ \text{etc.} \tag{29}$$

we find that the equations governing radial oscillations of small amplitudes are

$$\frac{\partial}{\partial m}(2\pi r_0\delta r)=-\frac{\delta\rho}{\rho_0^2} \tag{30}$$

and

$$\frac{\partial^2}{\partial t^2}\delta r=-2\pi\delta r\frac{\partial P_0}{\partial m}-2\pi r_0\frac{\partial}{\partial m}\delta P+\frac{2Gm}{r_0^2}\delta r\ . \tag{31}$$

[2] *Ap. J.*, **118**, 113, 1953.

[3] M. Schwarzschild, *Ann. d'ap.*, **12**, 148, 1949; G. Gjellestad, *Rep. No. 1, Inst. Ther. Ap.* (Oslo, 1950), and *Ann. d'ap.*, **15**, 276, 1952; V. C. A. Ferraro and D. J. Memory, *M.N.*, 112, 361, 1952; T. G. Cowling, *M.N.*, 112, 527, 1952.

Using the equation

$$2\pi r_0 \frac{\partial P_0}{\partial m} = -\frac{2Gm}{r_0}, \tag{32}$$

which must obtain in equilibrium, we can rewrite equation (31) in the form

$$\frac{\partial^2}{\partial t^2}\delta r = -2\pi r_0 \frac{\partial}{\partial m}\delta P + \frac{4Gm}{r_0^2}\delta r. \tag{33}$$

We shall now evaluate δP. For an adiabatic pulsation,

$$\delta P = \delta p + \frac{\boldsymbol{H}_0 \cdot \delta \boldsymbol{H}}{4\pi} = \gamma \frac{p_0}{\rho_0}\delta\rho + \frac{\boldsymbol{H}_0 \cdot \delta \boldsymbol{H}}{4\pi}. \tag{34}$$

Now when the medium is of infinite electrical conductivity, the change, $\Delta \boldsymbol{H}$, *at a given point* in a prevailing magnetic field, $\boldsymbol{H}_0$, caused by a displacement $\delta \boldsymbol{r}$, is given quite generally by

$$\Delta \boldsymbol{H} = \operatorname{curl}(\delta \boldsymbol{r} \times \boldsymbol{H}_0). \tag{35}$$

This relation is derived in § 14 below (see eq. [130]); but we may note here that it merely expresses the fact that the changes in the magnetic field are simply a consequence of the lines of force being pushed aside. According to equation (35), the change in the magnetic field, $\delta \boldsymbol{H}$, *as we follow the motion*, is given by

$$\delta \boldsymbol{H} = \operatorname{curl}(\delta \boldsymbol{r} \times \boldsymbol{H}_0) + (\delta \boldsymbol{r} \cdot \operatorname{grad})\boldsymbol{H}_0. \tag{36}$$

When $\boldsymbol{H}_0$ is in the z-direction and $\delta \boldsymbol{r}$ is radial, the only nonvanishing component of $\delta \boldsymbol{H}$ is

$$\delta H_z = -\frac{1}{r}\frac{\partial}{\partial r}(H_0 r \delta r) + \delta r \frac{\partial H_0}{\partial r} = -\frac{H_0}{r}\frac{\partial}{\partial r}(r\delta r), \tag{37}$$

in the z-direction. Hence in the case under consideration

$$\frac{\boldsymbol{H}_0 \cdot \delta \boldsymbol{H}}{4\pi} = -\frac{H_0^2 \rho_0}{4\pi}\frac{\partial}{\partial m}(2\pi r_0 \delta r), \tag{38}$$

and the expression for δP becomes

$$\delta P = -\left(\gamma p_0 + \frac{H_0^2}{4\pi}\right)\rho_0 \frac{\partial}{\partial m}(2\pi r_0 \delta r), \tag{39}$$

where we have substituted for $\delta\rho$ in equation (34) in accordance with equation (30).

With δP given by equation (38), equation (33) takes the form

$$\frac{\partial^2}{\partial t^2}\delta r = 4\pi^2 r \frac{\partial}{\partial m}\left\{\left(\gamma p + \frac{H^2}{4\pi}\right)\rho \frac{\partial}{\partial m}(r\delta r)\right\} + \frac{4Gm}{r^2}\delta r. \tag{40}$$

In writing equation (39), we have suppressed the subscripts zero distinguishing the equilibrium configuration, since there is no longer any cause for ambiguity.

When all the physical variables vary with time like $e^{i\sigma t}$, equation (39) reduces to

$$\left(\sigma^2 + \frac{4Gm}{r^2}\right)\delta r = -4\pi^2 r \frac{d}{dm}\left\{\left(\gamma p + \frac{H^2}{4\pi}\right)\rho \frac{d}{dm}(r\delta r)\right\}, \tag{41}$$

where δr has now the meaning of an amplitude.

The boundary conditions,

$$\delta r = 0 \quad \text{at} \quad m = 0 \quad \text{and} \quad \delta P = 0 \quad \text{at} \quad m = M \,, \tag{42}$$

in conjunction with the pulsation equation (40) will determine for σ^2 a sequence of possible characteristic values, σ_k^2. And it can be readily shown that the solutions, δr_k, belonging to the different characteristic values, are orthogonal:

$$\int_0^M \delta r_k \delta r_l dm = 0 \qquad (k \neq l) \,. \tag{43}$$

In view of this orthogonality of the functions δr_k, we should expect that the characteristic values themselves could be determined by a variational method. The basis for this method is developed in the following section.

6. *An integral formula for σ^2 and a variational method for determining it.*—Multiply equation (41) by δr and integrate over the range of m, i.e., from 0 to M. We obtain

$$\sigma^2 \int_0^M (\delta r)^2 dm + 4G \int_0^M \left(\frac{\delta r}{r}\right)^2 dm = -4\pi^2 \int_0^M r \delta r \frac{d}{dm} \left\{ \left(\gamma p + \frac{H^2}{4\pi}\right) \rho \frac{d}{dm} (r \delta r) \right\} dm \,. \tag{44}$$

By integrating by parts the integral on the right-hand side, we obtain

$$\sigma^2 \int_0^M (\delta r)^2 dm + 4G \int_0^M \left(\frac{\delta r}{r}\right)^2 dm = 4\pi^2 \int_0^M \left(\gamma p + \frac{H^2}{4\pi}\right) \rho \left[\frac{d}{dm} (r \delta r)\right]^2 dm \,. \tag{45}$$

Writing $p = (P - H^2/8\pi)$ in equation (45), we obtain, after some elementary reductions,

$$\sigma^2 \int_0^M (\delta r)^2 dm = \gamma \int_0^M \frac{P}{\rho} \left[\frac{1}{r} \frac{d}{dr} (r \delta r)\right]^2 dm - 4G \int_0^M \left(\frac{\delta r}{r}\right)^2 m\, dm + \frac{1}{8\pi} (2 - \gamma) \int_0^M \frac{H^2}{\rho} \left[\frac{1}{r} \frac{d}{dr} (r \delta r)\right]^2 dm \,. \tag{46}$$

It can be shown that the foregoing equations give a minimum value for σ^2 when the true solution δr belonging to the lowest characteristic value of the pulsation equation is substituted; and any other function δr (satisfying the boundary conditions) will give a larger value for σ^2. These facts can clearly be made the basis of a variational procedure for determining σ^2.

In the theory of the adiabatic pulsations of ordinary stars, it is known[4] that we get a very good estimate of σ^2 (for the fundamental mode) by setting

$$\delta r = \text{Constant } r \,, \tag{47}$$

in an integral formula for σ^2 similar to equation (46). We shall assume that this will continue to be the case in our present problem. Therefore, making the substitution (47) in equation (46), we obtain

$$\sigma^2 \int_0^M r^2 dm = 4\gamma \int_0^M \frac{P}{\rho} dm - 2GM^2 + (2 - \gamma) \int_0^R H^2 r\, dr \,. \tag{48}$$

[4] P. Ledoux and C. L. Pekeris, *Ap. J.*, **94**, 124 1941.

On the other hand,

$$\int_0^M \frac{P}{\rho}\,dm = 2\pi\int_0^R P\,r\,dr = -\pi\int_0^R r^2\frac{dP}{dr}\,dr = G\int_0^M m\,dm = \tfrac{1}{2}GM^2. \qquad (49)$$

Hence

$$\sigma^2\int_0^M r^2 dm = 2(\gamma-1)GM^2 + (2-\gamma)\int_0^R H^2 r\,dr. \qquad (50)$$

An alternative form of this equation is (cf. eq. [23])

$$\begin{aligned}\sigma^2\int_0^M r^2 dm &= 4(\gamma-1)[\tfrac{1}{2}GM^2 - \mathfrak{M}] + 4\mathfrak{M}\\ &= 4[(\gamma-1)^2\mathfrak{U} + \mathfrak{M}].\end{aligned} \qquad (51)$$

III. THE GRAVITATIONAL INSTABILITY OF AN INFINITELY LONG CYLINDER WHEN A CONSTANT MAGNETIC FIELD IS ACTING IN THE DIRECTION OF THE AXIS

7. *The formulation of the problem.*—In Section II we have seen that an infinitely long cylinder in which there is a prevalent magnetic field in the direction of the axis is stable for radial oscillations. But the question was left open as to whether the cylinder may not be unstable for transverse or for longitudinal oscillations. In Section III we shall take up the discussion of the transverse oscillations; however, in order not to complicate an already difficult problem, we shall restrict ourselves to the case when the medium is incompressible in addition to being an infinitely good electrical conductor.

We picture to ourselves, then, an infinite cylinder of uniform circular cross-section of radius R_0, along the axis of which a constant magnetic field of intensity H_0 is acting. Since any transverse perturbation can be expressed as a superposition of waves of different wave lengths, the question of stability can be investigated by considering, individually, perturbations of different wave lengths. We suppose, then, that the cylinder is subject to a perturbation, the result of which is to deform the boundary into

$$r = R + a\cos kz. \qquad (52)$$

Since the fluid is assumed to be incompressible, the mass per unit length must be the same before and after the deformation; this, clearly, requires that

$$R_0^2 = R^2 + \tfrac{1}{2}a^2. \qquad (53)$$

We shall see that, as a result of the deformation, the mean field in the z-direction is also changed by an amount of order a^2 (see eq. [87] below).

The investigation of the stability of the cylinder consists of two parts. First, we must calculate the change in the potential energy, $\Delta\Omega$, and the magnetic energy, $\Delta\mathfrak{M}$, per unit length resulting from the perturbation. Then, depending on whether $\Delta\Omega + \Delta\mathfrak{M}$ is positive or negative, we shall have stability or instability. We shall see presently that $\Delta\Omega + \Delta\mathfrak{M} < 0$ for all k less than a certain determinate value depending on H_0. In other words, the cylinder is unstable for all wave lengths exceeding a certain critical value λ_*. The determination of λ_* is the first problem in the investigation of stability. The second problem concerns the specification of the wave number k_m (say) for which the instability will develop at the maximum rate. We can determine this mode of maximum instability by considering the amplitude of the deformation (cf. eq. [52]) as a function of time, constructing a Lagrangian for the cylinder and determining the manner of increase of the amplitudes of the unstable modes. We shall find that whenever $\lambda > \lambda_*$ (or $k < k_*$), the amplitude increases like e^{qt}, where q is a function of k (and H_0). The mode of maximum instability is clearly the one which makes q (for a given H) a maximum.

Before proceeding to the details of the calculations, we may state that the method we have described derives from an early investigation of Rayleigh's[5] on the stability of liquid jets.

8. *The change in the potential energy per unit length caused by the deformation.*—Following the outline given in § 7, we shall first calculate the change in the potential energy, $\Delta\Omega$, per unit length caused by the deformation which makes the cross-section change from one of a constant radius R_0 to one whose boundary is given by equation (52). Since the potential energy per unit length of an infinite cylinder is infinite, the evaluation of $\Delta\Omega$ requires some care. We proceed as follows:

Let U and V denote the external and the internal gravitational potentials of the deformed cylinder. They satisfy the equations

$$\nabla^2 U = 0 \quad \text{and} \quad \nabla^2 V = -4\pi G\rho \,. \tag{54}$$

We shall first solve these equations to the first order in the amplitude a appropriately for the problem on hand. The solutions must clearly be of the forms

$$U = -2\pi G\rho R^2 \log r + aAK_0(kr) \cos kz + c_0 \tag{55}$$

and

$$V = -\pi G\rho r^2 + aBI_0(kr) \cos kz \,, \tag{55a}$$

where c_0 is an additive constant (with which we need not further concern ourselves), A and B are constants to be determined, and I_n and K_n are the Bessel functions of order n for a purely imaginary argument, which have no singularity at the origin and at infinity, respectively.

The constants A and B in solutions (55) are to be determined by the condition that U and V and $\partial U/\partial r$ and $\partial V/\partial r$ must be continuous on the boundary (52). Carrying out the calculations consistently to the first order in a, we find that the continuity conditions require

$$A\,K_0\,(kR) = B\,I_0\,(kR) \tag{56}$$

and

$$A\,K_1\,(kR) + B\,I_1\,(kR) = \frac{4\pi G\rho}{k} \,. \tag{56a}$$

Solving these equations, we find

$$A = 4\pi G\rho R I_0(kR) \quad \text{and} \quad B = 4\pi G\rho R K_0(kR) \,. \tag{57}$$

The required solution for V is, therefore,

$$V = -\pi G\rho r^2 + 4\pi G\rho R\, aK_0(kR) I_0(kr) \cos kz + O(a^2). \tag{58}$$

Now suppose that the amplitude of the deformation is increased by an *infinitesimal* amount from a to $a + \delta a$. The change in the potential energy, $\delta\Delta\Omega$, consequent to this infinitesimal increase in the amplitude, can be determined by evaluating the work done in the redistribution of the matter required to increase the amplitude. For evaluating this latter work, it is necessary to specify in a quantitative manner the redistribution which takes place; and we shall now do this.

An arbitrary deformation of an incompressible fluid can be thought of as resulting from a displacement $\boldsymbol{\xi}$ applied to each point of the fluid. The assumed incompressibility of the medium requires that div $\boldsymbol{\xi} = 0$; and, since no loss of generality is implied by supposing that the displacement is irrotational, we shall write

$$\boldsymbol{\xi} = \text{grad}\, \psi, \tag{59}$$

[5] Lord Rayleigh, *Scientific Papers* (Cambridge: At the University Press, 1900), 2, 361; also *Theory of Sound* ("Dover Reprints" [New York, 1945]), 2, 350–362.

and require that

$$\nabla^2\psi = 0. \tag{60}$$

A solution of equation (60) which is suitable for considering the deformation of a uniform cylinder into one whose boundary is given by (52) is

$$\psi = AI_0(kr)\cos kz, \tag{61}$$

where A is a constant. The corresponding radial and z-components of $\boldsymbol{\xi}$ are

$$\xi_r = AkI_1(kr)\cos kz \quad \text{and} \quad \xi_z = -AkI_0(kr)\sin kz. \tag{62}$$

Since at $r = R$, ξ_r must reduce to $a\cos kz$ (cf. eq. [52]), we must have

$$A = \frac{a}{kI_1(kR)}. \tag{63}$$

The displacements,

$$\xi_r = a\frac{I_1(kr)}{I_1(kR)}\cos kz \quad \text{and} \quad \xi_z = -a\frac{I_0(kr)}{I_1(kR)}\sin kz, \tag{64}$$

applied to each point of the cylinder will deform it into the required shape. The displacement $\delta\boldsymbol{\xi}$, which must be applied to increase the amplitude from a to $a + \delta a$, is therefore,

$$\delta\xi_r = \delta a\frac{I_1(kr)}{I_1(kR)}\cos kz \quad \text{and} \quad \delta\xi_z = -\delta a\frac{I_0(kr)}{I_1(kR)}\sin kz. \tag{65}$$

The change in the potential energy, $\delta\Delta\Omega$, per unit length involved in the infinitesimal deformation (65) can be obtained by integrating over the whole cylinder the work done by the displacement $\delta\boldsymbol{\xi}$ in the force field specified by the gravitational potential (58). It is therefore given by

$$\delta\Delta\Omega = -2\pi\rho\left\{\int_0^{R+a\cos kz}\delta\boldsymbol{\xi}\cdot\text{grad }V\,r\,dr\right\}_{\text{av}}, \tag{66}$$

where the averaging is to be done with respect to z. Substituting for V and $\delta\boldsymbol{\xi}$ from equations (58) and (65), we obtain

$$\begin{aligned}\delta\Delta\Omega = -2\pi\rho\delta a\Bigg\{\int_0^{R+a\cos kz}&\cos kz\frac{I_1(kr)}{I_1(kR)}[-2\pi G\rho r\\ &+4\pi\rho GRakK_0(kR)I_1(kr)\cos kz]\,r\,dr\\ +\int_0^{R+a\cos kz}&\sin kz\frac{I_0(kr)}{I_1(kR)}[4\pi\rho GRakK_0(kR)I_0(kr)\sin kz]\,r\,dr\Bigg\}_{\text{av}}.\end{aligned} \tag{67}$$

Evaluating the foregoing expression consistently to the first order in a, we find

$$\delta\Delta\Omega = 2\pi^2\rho^2GR^2a\delta a - 4\pi^2\rho^2GRa\delta a\frac{kK_0(kR)}{I_1(kR)}\int_0^R[I_1^2(kr)+I_0^2(kr)]\,r\,dr, \tag{68}$$

or, using the readily verifiable result,

$$\int_0^R[I_1^2(kr)+I_0^2(kr)]\,r\,dr = \frac{R}{k}I_0(kR)\,I_1(kR), \tag{69}$$

we have

$$\delta\Delta\Omega = 4\pi^2\rho^2GR^2[\tfrac{1}{2} - I_0(x)K_0(x)]a\delta a, \tag{70}$$

where for the sake of brevity we have written

$$x = kR\,. \tag{71}$$

Finally, integrating equation (70) over a from 0 to a, we obtain

$$\Delta\Omega = 2\pi^2\rho^2 G R^2[\tfrac{1}{2} - I_0(x)K_0(x)]a^2\,. \tag{72}$$

This is the required expression for the change in the potential energy per unit length caused by the deformation.

9. *The change in the magnetic energy per unit length caused by the deformation.*—The changes in the magnetic field inside the cylinder can best be determined from the condition that the magnetic induction across any section normal to the axis of the cylinder must remain unaffected by the deformation. This condition follows from the assumed infinite electrical conductivity of the medium. Thus, if

$$H\mathbf{1}_z + \mathbf{h} \tag{73}$$

represents the magnetic field inside the cylinder (where $\mathbf{1}_z$ is a unit vector in the z-direction, $\mathbf{h}$ is a field, of order a, varying periodically with z, and H is the mean field), we must require that

$$N = \int_0^R H_0 r\,dr = \int_0^{R+a\cos kz} (H + h_z)\, r\,dr = \text{Constant}\,. \tag{74}$$

Turning to the determination of H and $\mathbf{h}$, we may first observe that $\mathbf{h}$ can be derived from a magnetostatic potential ϕ satisfying the equation $\nabla^2\phi = 0$. For the problem on hand we can represent ϕ as a series in powers of a of the form

$$\phi = \sum_{n=1}^{\infty} \frac{a^n A_n}{nk} I_0(nkr) \sin nkz\,, \tag{75}$$

where the A_n's are constants to be determined. Retaining terms up to the second order in a, we have

$$h_r = aA_1 I_1(kr)\sin kz + a^2 A_2 I_1(2kr)\sin 2kz \tag{76}$$

and

$$h_z = aA_1 I_0(kr)\cos kz + a^2 A_2 I_0(2kr)\cos 2kz\,, \tag{77}$$

for the components of $\mathbf{h}$.

With h_z given by equation (77), the magnetic induction across a normal section of the cylinder is given by

$$N = \int_0^{R+a\cos kz} \{H + aA_1 I_0(kr)\cos kz + a^2 A_2 I_0(2kr)\cos 2kz\}\, r\,dr\,. \tag{78}$$

Evaluating N correct to the second order in a, we obtain

$$\begin{aligned} N = \tfrac{1}{2}H(R^2 + \tfrac{1}{2}a^2) + \tfrac{1}{2}a^2 A_1 R I_0(kR) + a\Big[HR + \frac{A_1}{k} I_1(kR)R\Big]\cos kz \\ + a^2\Big[\tfrac{1}{4}H + \tfrac{1}{2}A_1 R I_0(kR) + \frac{A_2}{2k} R I_1(2kR)\Big]\cos 2kz\,; \end{aligned} \tag{79}$$

and according to equation (74) this must be identically equal to (cf. eq. [53])

$$\tfrac{1}{2}H_0 R_0^2 = \tfrac{1}{2}H_0(R^2 + \tfrac{1}{2}a^2)\,. \tag{80}$$

Hence we must require that

$$\tfrac{1}{2}H(R^2 + \tfrac{1}{2}a^2) + \tfrac{1}{2}a^2A_1RI_0(kR) = \tfrac{1}{2}H_0(R^2 + \tfrac{1}{2}a^2)\,, \tag{81}$$

$$HR + \frac{A_1}{k}I_1(kR)R = 0\,, \tag{82}$$

and

$$\tfrac{1}{4}H + \tfrac{1}{2}A_1RI_0(kR) + \frac{A_2}{2k}RI_1(2kR) = 0\,. \tag{83}$$

From equations (82) and (83) we find:

$$A_1 = -\frac{H}{R}\frac{x}{I_1(x)} \tag{84}$$

and

$$A_2 = \frac{H}{R^2}\frac{x}{I_1(2x)}\left\{\frac{xI_0(x)}{I_1(x)} - \frac{1}{2}\right\}, \tag{85}$$

where $x = kR$ (cf. eq. [71]).

With A_1 given by equation (84), equation (81) gives (correct to order a^2)

$$H_0 = H\left\{1 - \frac{a^2}{R^2}\frac{xI_0(x)}{I_1(x)}\right\}, \tag{86}$$

or, equivalently,

$$H = H_0\left\{1 + \frac{a^2}{R^2}\frac{xI_0(x)}{I_1(x)}\right\}. \tag{87}$$

This equation shows that the mean field inside the deformed cylinder is larger than that in the undeformed cylinder; the difference is of order a^2 and depends on the wave number of the deformation.

Equations (76), (77), (84), (85), and (87) determine the field inside the cylinder correct to the second order in a. It may be noted here that the same solution can also be derived from the alternative (but equivalent) condition that the magnetic lines of force follow the boundary of the cylinder (52).

With the field inside the cylinder determined, we can now evaluate the magnetic energy, $\mathfrak{M}$, per unit length. We have

$$\begin{aligned}\mathfrak{M} &= \frac{1}{4}\left\{\int_0^{R+a\cos kz}|H|^2 r\,dr\right\}_{\text{av}} \\ &= \frac{1}{4}\left\{\int_0^{R+a\cos kz}(H^2 + 2Hh_z + h_z^2 + h_r^2)\,r\,dr\right\}_{\text{av}}\,,\end{aligned} \tag{88}$$

where the averaging is to be done with respect to z. Substituting for h_r and h_z from equations (76) and (77) and evaluating $\mathfrak{M}$ correct to the second order in a, we obtain (cf. eq. [69])

$$\begin{aligned}\mathfrak{M} &= \tfrac{1}{8}H^2(R^2 + \tfrac{1}{2}a^2) + \tfrac{1}{2}aH\left\{\cos kz\int_0^{R+a\cos kz}A_1I_0(kr)\,r\,dr\right\}_{\text{av}} \\ &\quad + \tfrac{1}{2}a^2HA_2\left\{\cos 2kz\int_0^{R+a\cos kz}I_0(2kr)\,r\,dr\right\}_{\text{av}} \\ &\quad + \tfrac{1}{8}a^2A_1^2\int_0^R[I_0^2(kr) + I_1^2(kr)]\,r\,dr \\ &= \tfrac{1}{8}H^2(R^2 + \tfrac{1}{2}a^2) + \tfrac{1}{4}a^2HA_1RI_0(kR) + \tfrac{1}{8}a^2A_1^2\frac{R}{k}I_0(kR)\,I_1(kR)\,.\end{aligned} \tag{89}$$

On making further use of equations (53), (84), and (87), we can reduce this last expression for $\mathfrak{M}$ to the form

$$\mathfrak{M} = \tfrac{1}{8} H_0^2 R_0^2 + \tfrac{1}{8} a^2 H^2 \frac{x I_0(x)}{I_1(x)} . \tag{90}$$

But the magnetic energy per unit length of the undeformed cylinder is $\frac{1}{8}H_0^2R_0^2$. Hence

$$\Delta \mathfrak{M} = \tfrac{1}{8} a^2 H^2 \frac{x I_0(x)}{I_1(x)} . \tag{91}$$

10. *The modes of deformation which are unstable.*—Combining the results of §§ 8 and 9, we have

$$\Delta\Omega + \Delta\mathfrak{M} = \left\{ 2\pi^2 \rho^2 G R^2 [\tfrac{1}{2} - I_0(x) K_0(x)] + \tfrac{1}{8} H^2 \frac{x I_0(x)}{I_1(x)} \right\} a^2 . \tag{92}$$

Letting

$$H_s^2 = 16\pi^2 \rho^2 R^2 G \qquad \text{or} \qquad H_s = 4\pi\rho R\sqrt{G} , \tag{93}$$

we can rewrite equation (92) more conveniently in the form

$$\Delta\Omega + \Delta\mathfrak{M} = 2\pi^2 \rho^2 R^2 G \left\{ [\tfrac{1}{2} - I_0(x) K_0(x)] + \frac{x I_0(x)}{I_1(x)} \left(\frac{H}{H_s}\right)^2 \right\} a^2 . \tag{94}$$

Whether the mode of deformation considered is stable or unstable will depend upon the sign of the quantity in braces in the foregoing expression.

Now the asymptotic behaviors of the Bessel functions which appear in equation (94) are:

$$I_0(x) \to 1, \quad I_1(x) \to \tfrac{1}{2}x, \quad \text{and} \quad K_0(x) \to -(\gamma + \log \tfrac{1}{2}x) \quad (x \to 0) , \tag{95}$$

where γ (not to be confused with the ratio of the specific heats) is Euler's constant 0.5772 . . . , and

$$I_0(x) \to \frac{e^x}{(2\pi x)^{1/2}}, \quad I_1(x) \to \frac{e^x}{(2\pi x)^{1/2}}, \quad \text{and} \quad K_0(x) \to \left(\frac{\pi}{2x}\right)^{1/2} e^{-x} \quad (x \to \infty) . \tag{96}$$

Hence $\Delta\Omega$ (cf. eq. [72]) tends to minus infinity logarithmically as $x \to 0$ and tends monotonically to the positive limit $\pi^2\rho^2R^2G$ as $x \to \infty$, while $\Delta\mathfrak{M}$ (cf. eq. [91]) tends to the positive limit $\frac{1}{4}a^2H^2$ as $x \to 0$ and increases monotonically to infinity (linearly) as $x \to \infty$. These behaviors of $\Delta\Omega$ and $\Delta\mathfrak{M}$ are illustrated in Figure 1, in which the functions $[\frac{1}{2} - I_0(x)K_0(x)]$ and $xI_0(x)/I_1(x)$ are plotted.

From the asymptotic behaviors of $\Delta\Omega$ and $\Delta\mathfrak{M}$ it follows that the equation

$$\tfrac{1}{2} - I_0(x) K_0(x) + \frac{x I_0(x)}{I_1(x)} \left(\frac{H}{H_s}\right)^2 = 0 \tag{97}$$

allows a single positive root. Let $x = x_*$ denote this root. Then

$$\Delta\Omega + \Delta\mathfrak{M} > 0 \qquad \text{for} \qquad x > x_* , \tag{98}$$

and

$$\Delta\Omega + \Delta\mathfrak{M} < 0 \qquad \text{for} \qquad x < x_* . \tag{98a}$$

Hence *all modes of deformation with $x < x_*$ are unstable.* Since $x = kR$, x_* specifies the minimum wave number (in units of $1/R$) for a stable deformation; alternatively, we could also say that *all modes of deformation with wave lengths exceeding*

$$\lambda_* = \frac{2\pi R}{x_*} \tag{99}$$

are unstable.

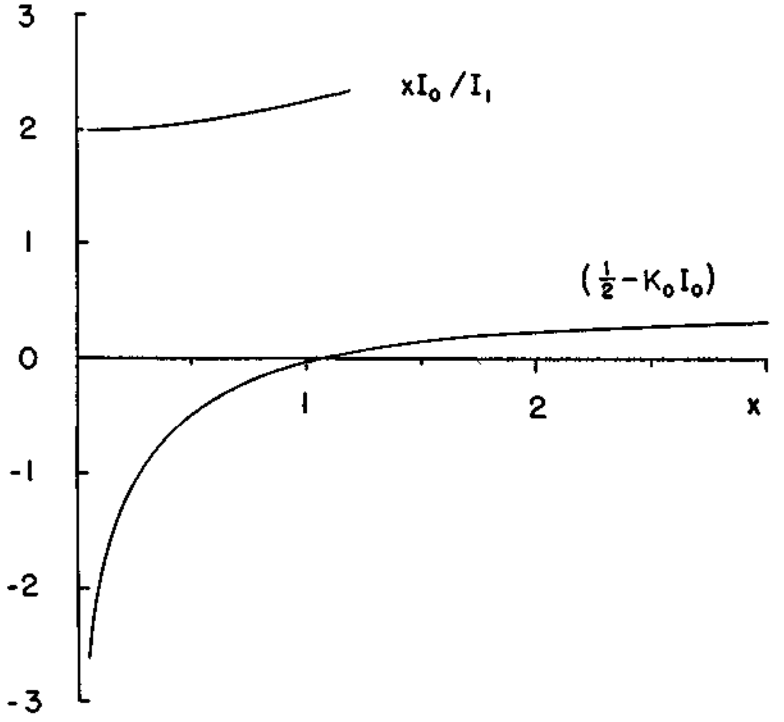


FIG. 1.—The dependence of the changes in the potential energy, $\Delta\Omega$, and magnetic energy, $\Delta\mathfrak{M}$, per unit length of an infinitely long cylinder on the wave number of the deformation; $\Delta\Omega$ is proportional to $[\frac{1}{2} - I_0(x)K_0(x)]$, while $\Delta\mathfrak{M}$ is proportional to $xI_0(x)/I_1(x)$, where x is the wave number measured in the unit $1/R$.

TABLE 1

DEPENDENCE OF WAVE NUMBERS x_* AND x_m AT WHICH INSTABILITY FIRST SETS IN AND AT WHICH IT IS MAXIMUM, ON PREVAILING MAGNETIC FIELD

H/H_s	x_*	x_m	$q_m/(4\pi G\rho)^{1/2}$
0	1.067	0.58	0.246
0.25	0.832	.47	.208
0.50	0.480	.28	.133
0.75	0.232	.14	.0685
1.00	0.092	.057	.0281
1.25	0.0299	.0182	.0091
1.50	0.00757	.00459	.00229
2.00	0.000228	0.000139	0.0000693

In Table 1 we have listed x_* for a few values of H/H_s. This table exhibits the strong stabilizing effect of the magnetic field: this is shown in the present connection by the very rapid increase, with increasing H, of the wave length at which instability sets in. In fact, for $H > H_s$ this increase becomes exponential; this can be shown in the following way:

Since $x_* = 0.092$ already for $H = H_s$, for $H > H_s$ we may replace the Bessel functions which occur in equation (97) by their dominant terms for $x \rightarrow 0$; thus,

$$\tfrac{1}{2} + \gamma + \log \tfrac{1}{2} x_* + 2 \left(\frac{H}{H_s}\right)^2 = 0 \qquad (H > H_s) \,. \tag{100}$$

Hence

$$x_* = 2 \exp \left\{ - \left[\gamma + \tfrac{1}{2} + 2 \left(\frac{H}{H_s}\right)^2 \right] \right\} \qquad (H > H_s) \,, \tag{101}$$

or, numerically,

$$x_* = 0.6811 \, e^{-2(H/H_s)^2} \qquad (H > H_s) \,. \tag{102}$$

11. *The mode of maximum instability.*—In the preceding section we have seen that an infinite cylinder is gravitationally unstable for all modes of deformation with wave lengths exceeding a certain critical value. We shall now show that there exists a wave length for which the instability is a maximum. For this purpose we shall suppose that the amplitude, a, of the deformation is a function of time and seek an equation of motion for it.

We have already seen that the potential energy (gravitational plus magnetic) per unit length of the cylinder measured from the equilibrium state is

$$\mathfrak{B} = \Delta\mathfrak{M} + \Delta\Omega = -2\pi^2\rho^2 R^2 G F(x) a^2 \,, \tag{103}$$

where

$$F(x) = I_0(x) K_0(x) - \tfrac{1}{2} - \left(\frac{H}{H_s}\right)^2 \frac{x I_0(x)}{I_1(x)} \,. \tag{104}$$

Defined in this manner, $F(x) > 0$ for $x < x_*$, i.e., it is positive for all unstable modes and negative for all stable modes.

To obtain the Lagrangian function for the cylinder, we must find the kinetic energy of the motion resulting from the varying amplitude. Since we have assumed that the fluid is incompressible, a velocity potential, ψ, exists which satisfies Laplace's equation. And the solution for the velocity potential appropriate to the problem on hand is

$$\psi = B I_0(kr) \cos kz \,, \tag{105}$$

where B is a constant to be determined. The components of the velocity derived from the foregoing potential are

$$u_r = \frac{\partial \psi}{\partial r} = +B k I_1(kr) \cos kz \tag{106}$$

and

$$u_z = \frac{\partial \psi}{\partial z} = -B k I_0(kr) \sin kz \,. \tag{106a}$$

The constant of proportionality, B, in the foregoing equations must be determined from the condition that the radial velocity, u_r, at $r = R$ must agree with that implied by equation (52); i.e., we should have

$$B k I_1(kR) \cos kz = \frac{da}{dt} \cos kz \,. \tag{107}$$

Hence

$$B = \frac{1}{k I_1(x)} \frac{da}{dt} \,. \tag{108}$$

From equations (106) we obtain, for the kinetic energy per unit length, the expression (cf. eq. [69])

$$\mathfrak{T} = \tfrac{1}{2}\pi\rho B^2 k^2 \int_0^R [I_0^2(kr) + I_1^2(kr)]\, r\, dr$$
$$= \tfrac{1}{2}\pi\rho B^2 k^2 \frac{R}{k} I_0(x)\, I_1(x)\,, \tag{109}$$

or, substituting for B from equation (108), we have

$$\mathfrak{T} = \tfrac{1}{2}\pi\rho R^2 \frac{I_0(x)}{x I_1(x)} \left(\frac{da}{dt}\right)^2. \tag{110}$$

The Lagrangian function (per unit length) for the infinite cylinder is therefore given by

$$\mathfrak{L} = \mathfrak{T} - \mathfrak{B} = \tfrac{1}{2}\pi\rho R^2 \frac{I_0(x)}{x I_1(x)} \left(\frac{da}{dt}\right)^2 + 2\pi^2\rho^2 R^2 G F(x)\, a^2. \tag{111}$$

The equation of motion for a derived from the Lagrangian (111) is

$$\pi\rho R^2 \frac{I_0(x)}{x I_1(x)} \frac{d^2 a}{dt^2} = 4\pi^2\rho^2 R^2 G F(x)\, a\,, \tag{112}$$

or, alternatively,

$$\frac{d^2 a}{dt^2} = 4\pi G\rho \left\{\frac{x I_1(x)}{I_0(x)} [I_0(x)\, K_0(x) - \tfrac{1}{2}] - \left(\frac{H}{H_s}\right)^2 x^2\right\} a\,, \tag{113}$$

where we have substituted for $F(x)$ in accordance with equation (104). The solution for a is therefore of the form

$$a = \text{Constant}\; e^{\pm qt}\,, \tag{114}$$

where

$$q^2 = 4\pi G\rho \left\{\frac{x I_1(x)}{I_0(x)} [I_0(x)\, K_0(x) - \tfrac{1}{2}] - \left(\frac{H}{H_s}\right)^2 x^2\right\}. \tag{115}$$

Accordingly, q is purely imaginary for $x > x_*$ and is real for $x < x_*$; this is in agreement with the fact that all modes with $x > x_*$ are stable, while all modes with $x < x_*$ are unstable.

As defined by equation (115), $q = 0$ both for $x = 0$ and for $x = x_*$. There is, therefore, a determinate intermediate value of x—say, x_m—for which q attains a maximum—say, q_m. *The wave number x_m clearly represents the mode of maximum instability;* for it is the mode for which the amplitude of the deformation increases most rapidly. The wave length

$$\lambda_m = \frac{2\pi R}{x_m}\,, \tag{116}$$

corresponding to the wave number x_m, gives approximately the length of the "pieces" into which the cylinder will ultimately break up: for the component with the wave length λ_m, in the Fourier analysis of an arbitrary perturbation, is the one whose amplitude will increase most rapidly with time and, therefore, represents the mode in which the instability will first assert itself. Finally, it is clear that $1/q_m$ gives a measure of the time needed for the instability to make itself manifest.

In Table 1 the values of x_m and $q_m/(4\pi G\rho)^{1/2}$ are also listed. As in the case of x_* (§ 10), we can give explicit formulae for x_m and q_m for $H > H_s$. Since for $H > H_s$ we are concerned only with values of $x \ll 1$, we may replace the Bessel functions which occur in the expression for q^2 by their dominant terms for $x \to 0$. Thus

$$q^2 = 4\pi G\rho \left\{ -\tfrac{1}{2}x^2(\gamma + \tfrac{1}{2} + \log \tfrac{1}{2}x) - x^2 \left(\frac{H}{H_s}\right)^2 \right\} \qquad (H > H_s) . \tag{117}$$

The expression on the right-hand side attains its maximum when

$$(\gamma + \tfrac{1}{2} + \log \tfrac{1}{2}x) + \tfrac{1}{2} + 2\left(\frac{H}{H_s}\right)^2 = 0 . \tag{118}$$

Hence

$$x_m = 2 \exp \left\{ -(\gamma + 1) - 2\left(\frac{H}{H_s}\right)^2 \right\} = 0.4131\, e^{-2(H/H_s)^2} \qquad (H > H_s) . \tag{119}$$

The corresponding expression for q_m is

$$q_m = \tfrac{1}{2}x_m(4\pi G\rho)^{1/2} \qquad (H > H_s) . \tag{120}$$

These formulae emphasize the fact, apparent from an examination of Table 1, that, as the strength of the magnetic field increases, not only does the wave length of the mode of maximum instability increase exponentially, but the time needed for the instability to manifest itself also increases exponentially.

TABLE 2

WAVE LENGTHS λ_* AND λ_m AT WHICH INSTABILITY SETS IN AND AT WHICH IT IS MAXIMUM AND CHARACTERISTIC TIME, q_m^{-1}, NEEDED FOR INSTABILITY TO MANIFEST ITSELF FOR CASE $R = 250$ PARSECS AND $\rho = 2 \times 10^{-24}$ GM/CM^3

H (Gauss)	λ_* (Parsecs)	λ_m (Parsecs)	q_m^{-1} (Years)
0	1.5×10^3	2.7×10^3	1.0×10^8
1.25×10^{-6}	1.9×10^3	3.3×10^3	1.2×10^8
$2.5\ \times10^{-6}$	3.3×10^3	5.6×10^3	1.8×10^8
3.75×10^{-6}	6.8×10^3	1.1×10^4	3.6×10^8
$5.0\ \times10^{-6}$	1.7×10^4	2.8×10^4	8.7×10^8
6.25×10^{-6}	5.2×10^4	8.6×10^4	2.7×10^9
$7.5\ \times10^{-6}$	2.1×10^5	3.4×10^5	1.1×10^{10}
$10.0\ \times10^{-6}$	6.9×10^6	1.1×10^7	3.5×10^{11}

12. *Numerical illustrations.*—To illustrate the theory developed in the preceding sections we shall take, as typical of a spiral arm of a galaxy,

$$R = 250 \text{ parsecs} \qquad \text{and} \qquad \rho = 2 \times 10^{-24} \text{ gm/cm}^3 . \tag{121}$$

The corresponding value of H_s is (cf. eq. [93])

$$H_s = 5.0 \times 10^{-6} \text{ gauss} . \tag{122}$$

For these values of the physical parameters, the nondimensional results given in Table 1 can be converted into astronomical measures; they are given in Table 2. From the values given in this table it follows that between $H = H_s$ and $H = 2H_s$ the characteristic time of the instability becomes so long that, for all practical purposes, the instability is effectively removed by the presence of the magnetic field.

IV. THE FLATTENING OF A GRAVITATING FLUID SPHERE UNDER THE INFLUENCE OF A MAGNETIC FIELD

13. *The formulation of the problem.*—In this section we shall consider the gravitational equilibrium of an incompressible fluid sphere with a uniform magnetic field inside and a dipole field outside. We shall show that under these circumstances the sphere is not a configuration of equilibrium and that it will become oblate by contracting along the axis of symmetry.

We suppose, then, that initially the magnetic field in the interior of the sphere is uniform and of intensity H in the z-direction. In spherical polar co-ordinates (r, θ, φ) the components of H in the radial (r) and the transverse (θ) directions are

$$H_r^{(i)} = H\mu \quad \text{and} \quad H_\theta^{(i)} = -H(1-\mu^2)^{1/2} \qquad (r < R), \tag{123}$$

where $\mu = \cos\theta$ and the superscript i indicates that these are the components of the field *inside* the sphere.

When the field inside the sphere is uniform, that outside the sphere must be a dipole field given by

$$H_r^{(e)} = H\left(\frac{R}{r}\right)^3 \mu \quad \text{and} \quad H_\theta^{(e)} = \tfrac{1}{2}H\left(\frac{R}{r}\right)^3 (1-\mu^2)^{1/2}, \tag{124}$$

where R denotes the radius of the sphere.

The energy, $\mathfrak{M}$, of the magnetic field specified by equations (123) and (124) is given by

$$\begin{aligned}\mathfrak{M} &= \frac{H^2}{8\pi}\left(\tfrac{4}{3}\pi R^3\right) + \tfrac{1}{4}H^2\int_R^\infty\int_{-1}^{+1}\left(\frac{R}{r}\right)^6\left\{\mu^2 + \tfrac{1}{4}(1-\mu^2)\right\} r^2 dr\, d\mu \\ &= \tfrac{1}{4}H^2R^3 .\end{aligned} \tag{125}$$

Let the sphere be now deformed in such a way that the equation of the bounding surface is

$$r(\mu) = R + \epsilon P_l(\mu), \tag{126}$$

where $\epsilon \ll R$, $\mu = \cos\theta$ (θ being the polar angle), and $P_l(\mu)$ denotes, as usual, the Legendre polynomial of order l. We shall call such a deformation of the sphere a "P_l-deformation." We shall investigate the stability of the sphere by examining whether or not it is stable to a P_l-deformation.

14. *The change in the magnetic energy of the sphere due to a P_l-deformation.*—As we have already pointed out in § 8, an arbitrary deformation of an incompressible body can be thought of as the result of applying to each point of the body a displacement $\boldsymbol{\xi}$. And if, as in § 8 (eqs. [59] and [60]), we express $\boldsymbol{\xi}$ as the gradient of a scalar function, ψ, the solution of Laplace's equation satisfied by ψ appropriate to a P_l-deformation of a sphere is

$$\psi = A r^l P_l(\mu), \tag{127}$$

where A is a constant. The corresponding expressions for the components of $\boldsymbol{\xi}$ are

$$\xi_r = \frac{\partial\psi}{\partial r} = A l r^{l-1} P_l(\mu) \tag{128}$$

and

$$\xi_\theta = \frac{1}{r}\frac{\partial\psi}{\partial\theta} = -A r^{l-1}(1-\mu^2)^{1/2} P_l'(\mu), \tag{128a}$$

where a prime is used to denote differentiation with respect to μ. According to equation (126), at $r = R$, $\xi_r = \epsilon P_l(\mu)$; this determines A, and we have

$$\xi_r = \epsilon \left(\frac{r}{R}\right)^{l-1} P_l(\mu) \text{ and } \xi_\theta = -\frac{\epsilon}{l}\left(\frac{r}{R}\right)^{l-1} (1-\mu^2)^{1/2} P_l'(\mu) . \tag{129}$$

Now the deformation of a body will alter the prevailing magnetic field; and, since in a medium of infinite electrical conductivity a change in the existing magnetic field can be effected only by bodily pushing aside the lines of force, it follows that

$$\delta \boldsymbol{H} = \text{curl}\,(\boldsymbol{\xi} \times \boldsymbol{H}) . \tag{130}$$

[The truth of this last relation can be established in the following way: Suppose that the displacement $\boldsymbol{\xi}$ takes place as a slow continuous movement so that if $\boldsymbol{u}$ denotes the velocity, $\boldsymbol{u} = \partial\boldsymbol{\xi}/\partial t$ (i.e., if quantities of the second order of smallness are neglected). On the other hand, when the electrical conductivity is infinite,

$$\delta \boldsymbol{E} = -\boldsymbol{u} \times \boldsymbol{H} ,$$

where $\delta\boldsymbol{E}$ is the electrical field resulting from the changing magnetic field $\delta\boldsymbol{H}$ in accordance with Maxwell's equation,

$$\text{curl}\ \delta\boldsymbol{E} = -\frac{\partial}{\partial t}\,\delta\boldsymbol{H} .$$

Combining the last two equations, we have

$$\text{curl}\left(\frac{\partial\boldsymbol{\xi}}{\partial t}\times\boldsymbol{H}\right) = \frac{\partial}{\partial t}(\delta\boldsymbol{H}) .$$

The relation (130) is simply the integrated form of this equation.]

When the fluid is incompressible (i.e., when div $\boldsymbol{\xi} = 0$ in addition to div $\boldsymbol{H} = 0$), equation (130) can be written alternatively in the form

$$\delta\boldsymbol{H} = (\boldsymbol{H}\cdot\text{grad})\boldsymbol{\xi} - (\boldsymbol{\xi}\cdot\text{grad})\boldsymbol{H} . \tag{131}$$

And when the initial field is homogeneous, equation (131) simplifies still further to

$$\delta\boldsymbol{H} = (\boldsymbol{H}\cdot\text{grad})\boldsymbol{\xi} . \tag{132}$$

In spherical polar co-ordinates the foregoing equation is equivalent to

$$\delta H_r = \left(H_r\frac{\partial}{\partial r} + \frac{H_\theta}{r}\frac{\partial}{\partial\theta}\right)\xi_r - \frac{H_\theta\xi_\theta}{r} \tag{133}$$

and

$$\delta H_\theta = \left(H_r\frac{\partial}{\partial r} + \frac{H_\theta}{r}\frac{\partial}{\partial\theta}\right)\xi_\theta + \frac{H_\theta\xi_r}{r} . \tag{133a}$$

These equations in conjunction with equations (123) and (129) give

$$\delta H_r^{(i)} = \epsilon H\,(l-1)\frac{r^{l-2}}{R^{l-1}}P_{l-1}(\mu) \tag{134}$$

and

$$\delta H_\theta^{(i)} = -\epsilon H\frac{r^{l-2}}{R^{l-1}}(1-\mu^2)^{1/2}P_{l-1}'(\mu) . \tag{134a}$$

The corresponding change in the internal magnetic energy density is given by

$$\delta\left(\frac{|\boldsymbol{H}|^2}{8\pi}\right)=\frac{1}{4\pi}\boldsymbol{H}^{(i)}\cdot\delta\boldsymbol{H}^{(i)}$$
$$=\epsilon\frac{H^2}{4\pi}\frac{r^{l-2}}{R^{l-1}}\{(l-1)\mu P_{l-1}(\mu)+(1-\mu^2)P'_{l-1}(\mu)\}\,. \tag{135}$$

On further simplification this reduces to

$$\delta\left(\frac{|\boldsymbol{H}|^2}{8\pi}\right)=\epsilon(l-1)\frac{H^2}{4\pi}\frac{r^{l-2}}{R^{l-1}}P_{l-2}(\mu)\,. \tag{136}$$

Hence, when averaged over all directions, this is zero except when $l = 2$, in which case

$$\delta\left(\frac{|\boldsymbol{H}|^2}{8\pi}\right)=\frac{\epsilon}{4\pi}\frac{H^2}{R}\qquad(l=2)\,; \tag{137}$$

the corresponding change in the internal magnetic energy, $\Delta\mathfrak{M}^{(i)}$, is given by

$$\Delta\mathfrak{M}^{(i)}=\tfrac{1}{3}\epsilon H^2R^2\,. \tag{138}$$

15. *The change in the external magnetic energy of the sphere due to a P_l-deformation.*— Writing the magnetic field outside the deformed sphere in the form

$$H_r^{(e)}=H\left(\frac{R}{r}\right)^3\mu+\delta H_r^{(e)} \tag{139}$$

and

$$H_\theta^{(e)}=\tfrac{1}{2}H\left(\frac{R}{r}\right)^3(1-\mu^2)^{1/2}+\delta H_\theta^{(e)}\,, \tag{139a}$$

we shall suppose that $\delta H_r^{(e)}$ and $\delta H_\theta^{(e)}$ are derivable from a magnetic potential $\delta\phi^{(e)}$. Since the magnetic potential satisfies Laplace's equation, the solution for $\delta\phi^{(e)}$ must be expressible as a linear combination of the fundamental solutions $P_j(\mu)/r^{j+1}$, which vanish at infinity.

We shall find it convenient to write the solution for $\delta\phi^{(e)}$ in the form

$$\delta\phi^{(e)}=-\epsilon H\left\{\frac{l-1}{l}\left(\frac{R}{r}\right)^l P_{l-1}(\mu)+\Sigma A_j\left(\frac{R}{r}\right)^{j+1}P_j(\mu)\right\}, \tag{140}$$

where the A_j's are coefficients to be determined. The expressions for $\delta H_r^{(e)}$ and $\delta H_\theta^{(e)}$ derived from this potential are

$$\delta H_r^{(e)}=\epsilon H\left\{(l-1)\frac{R^l}{r^{l+1}}P_{l-1}(\mu)+\Sigma A_j(j+1)\frac{R^{j+1}}{r^{j+2}}P_j(\mu)\right\} \tag{141}$$

and

$$\delta H_\theta^{(e)}=\epsilon H\left\{\frac{l-1}{l}\frac{R^l}{r^{l+1}}P^1_{l-1}(\mu)+\Sigma A_j\frac{R^{j+1}}{r^{j+2}}P^1_j(\mu)\right\}. \tag{141a}$$

The coefficients A_j in equations (141) and (141a) can be determined from the condition that the component of the magnetic field normal to a bounding surface must be continuous. To the first order in ϵ this condition requires that

$$\{H_r^{(e)}\}_{R+\epsilon P_l}+\{H_\theta^{(e)}\}_R\frac{\epsilon}{R}(1-\mu^2)^{1/2}\frac{\partial P_l}{\partial\mu}$$
$$=\{H_r^{(i)}\}_{R+\epsilon P_l}+\{H_\theta^{(i)}\}_R\frac{\epsilon}{R}(1-\mu^2)^{1/2}\frac{\partial P_l}{\partial\mu}\,, \tag{142}$$

where $-(\epsilon/R)(1-\mu^2)^{1/2}\partial P_l/\partial\mu$ is the angle (to the first order in ϵ) which the deformed boundary makes with the θ-direction; the terms in H_θ in the foregoing equation arise from this latter circumstance. Now, according to equations (124), (139), and (140),

$$\{H_r^{(e)}\}_{R+\epsilon P_l}+\{H_\theta^{(e)}\}_R\frac{\epsilon}{R}(1-\mu^2)^{1/2}\frac{\partial P_l}{\partial\mu}=H\mu\left(1-3\frac{\epsilon}{R}P_l\right)$$
$$+\tfrac{1}{2}H\frac{\epsilon}{R}(1-\mu^2)\frac{\partial P_l}{\partial\mu}+\frac{\epsilon}{R}H\{(l-1)P_{l-1}+\Sigma A_j(j+1)P_j\}\,, \tag{143}$$

while, according to equations (123) and (134),

$$\{H_r^{(i)}\}_{R+\epsilon P_l}+\{H_\theta^{(i)}\}_R\frac{\epsilon}{R}(1-\mu^2)^{1/2}\frac{\partial P_l}{\partial\mu}=H\mu$$
$$+\frac{\epsilon}{R}H(l-1)P_{l-1}-\frac{\epsilon}{R}H(1-\mu^2)\frac{\partial P_l}{\partial\mu}; \tag{143a}$$

and the equality of the expressions on the right-hand sides of equations (143) and (143a) requires

$$\Sigma A_j(j+1)P_j=3\mu P_l-\tfrac{3}{2}(1-\mu^2)\frac{\partial P_l}{\partial\mu}$$
$$=\frac{3}{2(2l+1)}\{(l+1)(l+2)P_{l+1}-l(l-1)P_{l-1}\}\,. \tag{144}$$

Hence

$$A_{l-1}=-\frac{3(l-1)}{2(2l+1)}\,,\qquad A_{l+1}=\frac{3(l+1)}{2(2l+1)}\,, \tag{145}$$

and

$$A_j=0\quad\text{for}\quad j\neq l-1\quad\text{or}\quad l+1\,. \tag{145a}$$

Inserting these values of A in equations (141) and (141a), we obtain

$$\delta H_r^{(e)}=\epsilon H\left\{\frac{(l-1)(l+2)}{2(2l+1)}\frac{R^l}{r^{l+1}}P_{l-1}(\mu)+\frac{3(l+1)(l+2)}{2(2l+1)}\frac{R^{l+2}}{r^{l+3}}P_{l+1}(\mu)\right\} \tag{146}$$

and

$$\delta H_\theta^{(e)}=\epsilon H\left\{\frac{(l-1)(l+2)}{2l(2l+1)}\frac{R^l}{r^{l+1}}P_{l-1}^1(\mu)+\frac{3(l+1)}{2(2l+1)}\frac{R^{l+2}}{r^{l+3}}P_{l+1}^1(\mu)\right\}. \tag{146a}$$

Returning to equations (139) and (139a), we can write the change in the external magnetic energy, $\Delta\mathfrak{M}^{(e)}$, to the first order in ϵ, in the form

$$\Delta\mathfrak{M}^{(e)}=\frac{H^2}{8\pi}\iiint_{R+\epsilon P_l\geq r\geq R}\left(\frac{R}{r}\right)^6\{\mu^2+\tfrac{1}{4}(1-\mu^2)\}r^2dr\,d\mu\,d\varphi$$
$$+\frac{H}{8\pi}\iiint_{r>R}\left(\frac{R}{r}\right)^3\{2P_1(\mu)\,\delta H_r^{(e)}+P_1^1(\mu)\,\delta H_\theta^{(e)}\}r^2dr\,d\mu\,d\varphi\,. \tag{147}$$

After some minor reductions we find

$$\Delta\mathfrak{M}^{(e)} = \tfrac{1}{4}\epsilon H^2R^2 \int_{-1}^{+1} \{\tfrac{1}{2}P_2(\mu) + \tfrac{1}{2}\}P_l(\mu)\,d\mu$$
$$+ \tfrac{1}{2}\epsilon H^2 \int_R^\infty dr\, r^2 \int_{-1}^{+1} d\mu \left(\frac{R}{r}\right)^3 P_1(\mu) \left\{ \frac{(l-1)\,(l+2)}{2\,(2l+1)}\,\frac{R^l}{r^{l+1}}\,P_{l-1}(\mu) + \frac{3\,(l+1)\,(l+2)}{2\,(2l+1)}\,\frac{R^{l+2}}{r^{l+3}}\,P_{l+1}(\mu)\right\} \tag{148}$$
$$+ \tfrac{1}{4}\epsilon H^2 \int_R^\infty dr\, r^2 \int_{-1}^{+1} d\mu \left(\frac{R}{r}\right)^3 P_1^1(\mu) \left\{ \frac{(l-1)\,(l+2)}{2l\,(2l+1)}\,\frac{R^l}{r^{l+1}}\,P_{l-1}^1(\mu) + \frac{3\,(l+1)}{2\,(2l+1)}\,\frac{R^{l+2}}{r^{l+3}}\,P_{l+1}^1(\mu)\right\}.$$

From this equation it is evident that $\Delta\mathfrak{M}^{(e)}$ vanishes (to the first order in ϵ) for all deformations except a P_2-deformation. And for a P_2-deformation we have

$$\Delta\mathfrak{M}^{(e)} = \tfrac{1}{8}\epsilon H^2R^2 \int_{-1}^{+1} [P_2(\mu)]^2 d\mu + \tfrac{1}{5}\epsilon H^2R^5 \int_R^\infty \int_{-1}^{+1} \frac{dr}{r^4}\{\tfrac{1}{2}P_2(\mu) + \tfrac{1}{2}\}\,d\mu \tag{149}$$

or

$$\Delta\mathfrak{M}^{(e)} = \tfrac{7}{60}\,\epsilon H^2R^2 \qquad (l = 2)\ . \tag{150}$$

Finally, combining equations (138) and (150), we obtain

$$\Delta\mathfrak{M} = \Delta\mathfrak{M}^{(i)} + \Delta\mathfrak{M}^{(e)} = \tfrac{9}{20}\,\epsilon H^2R^2\,, \tag{151}$$

for the total change in the magnetic energy due to a P_2-deformation; it vanishes to this order for all higher deformations.

We have, therefore, shown that *the change in the magnetic energy is of the second order in ϵ for all deformations of the sphere except a P_2-deformation; and for a P_2-deformation it is of the first order in ϵ and is given by (151).* Moreover, for a P_2-deformation $\Delta\mathfrak{M} > 0$ when the deformation is in the sense of making the sphere into a prolate spheroid; and $\Delta\mathfrak{M} < 0$ when the deformation is in the sense of making the sphere into an oblate spheroid.

16. *The change in the gravitational potential energy and the instability of the sphere to a P_2-deformation.*—The change in the potential energy, $\Delta\Omega$, due to a P_l-deformation can also be computed. The result is well known for a P_2-deformation. For a general P_l-deformation we can evaluate $\Delta\Omega$ by following the procedure used in § 8. We shall not give here the details of the calculations, which lead to the result

$$\Delta\Omega = \frac{3\,(l-1)}{(2l+1)^2}\left(\frac{\epsilon}{R}\right)^2 \frac{GM^2}{R}. \tag{152}$$

The change in the potential energy is therefore always positive and is of the second order in ϵ. This is in contrast to $\Delta\mathfrak{M}$, which, as we have seen, is of the first order in ϵ for a P_2-deformation and is negative for a deformation which tends to make it oblate. We can therefore conclude that the sphere is unstable and that it will tend to collapse toward an oblate spheroidal shape. To estimate the extent to which this collapse may proceed, let us consider $\Delta\Omega + \Delta\mathfrak{M}$ for a P_2-deformation. We have (cf. eqs. [151] and [152])

$$\Delta\Omega + \Delta\mathfrak{M} = \frac{3}{25}\,\frac{GM^2}{R^3}\,\epsilon^2 + \tfrac{9}{20}H^2R^2\epsilon \qquad (l = 2)\,. \tag{153}$$

As a function of ϵ, $\Delta\Omega + \Delta\mathfrak{M}$ has a minimum which it takes when

$$\frac{6}{25}\frac{GM^2}{R^3}\epsilon + \tfrac{9}{20}H^2R^2 = 0\,, \tag{154}$$

or

$$\frac{\epsilon}{R} = -\frac{15}{8}\frac{H^2R^4}{GM^2}\,. \tag{155}$$

If H_* denotes the value of the constant magnetic field inside the sphere for which $\mathfrak{M}$ (given by eq. [125]) is equal to the numerical value of the gravitational potential energy Ω ($= -3GM^2/5R$), then

$$\tfrac{1}{4}H_*^2R^3 = \frac{3}{5}\frac{GM^2}{R}\,. \tag{156}$$

In terms of H_* defined in this manner, we can rewrite equation (155) in the form

$$\frac{\epsilon}{R} = -4.5\left(\frac{H}{H_*}\right)^2. \tag{157}$$

We may interpret this relation by saying that when a star has a magnetic field approaching the limit set by the virial theorem (cf. Sec. I), then it tends to become highly oblate; in this respect the magnetic field has the same effect as a rotation.

V. THE GRAVITATIONAL INSTABILITY OF AN INFINITE HOMOGENEOUS MEDIUM IN THE PRESENCE OF A MAGNETIC FIELD

17. *The statement of the problem.*—It is well known that, by considering the propagation of a wave in an infinite homogeneous medium and allowing for the gravitational effects of the density fluctuations, Jeans[6] showed that the velocity of wave propagation is given by

$$V_J = \sqrt{(c^2 - 4\pi G\rho/k^2)}\,, \tag{158}$$

where $c = \sqrt{(\gamma p/\rho)}$ denotes the convectional velocity of sound and k is the wave number. Accordingly, when

$$k < c(4\pi\rho G)^{-1/2}\,, \tag{159}$$

the velocity of wave propagation becomes imaginary; and under these circumstances the amplitude of the wave will increase exponentially with time. The inequality (159) is therefore the condition for gravitational instability; this is Jeans's result. In Section V we shall show that Jeans's condition (159) is unaffected by the presence of a magnetic field. The physical reason for this is evident for a deformation in which the density waves are perpendicular to the lines of force because the motion of the particles in this case will be parallel to the lines of force and therefore will not be impeded by the magnetic field. But also a density wave forming an angle with the lines of force may be obtained by particle motions parallel to the lines of force, as shown in Figure 2.

18. *The three modes of wave propagation in the presence of a magnetic field and the condition for gravitational instability.*—Consider an extended homogeneous gaseous medium of infinite electrical conductivity, and suppose that there is present a uniform magnetic

[6] *Astronomy and Cosmogony* (Cambridge: At the University Press, 1929), pp. 345–347.

field of intensity $\boldsymbol{H}$. Then the fluctuations in density ($\delta\rho$), pressure (δp), magnetic field ($\boldsymbol{h}$), and gravitational potential (δV) are governed by the equations

$$\rho \frac{\partial \boldsymbol{u}}{\partial t} = \frac{1}{4\pi}(\text{curl } \boldsymbol{h} \times \boldsymbol{H}) - \text{grad } \delta p + \rho \text{ grad } \delta V ,$$

$$\frac{\partial \boldsymbol{h}}{\partial t} = \text{curl } (\boldsymbol{u} \times \boldsymbol{H}) , \tag{160}$$

$$\frac{\partial}{\partial t} \delta \rho = - \rho \text{ div } \boldsymbol{u} ,$$

and

$$\nabla^2 \delta V = - 4\pi G \delta \rho .$$

If the changes in pressure and density are assumed to take place adiabatically, then

$$\delta p = c^2 \delta \rho . \tag{161}$$

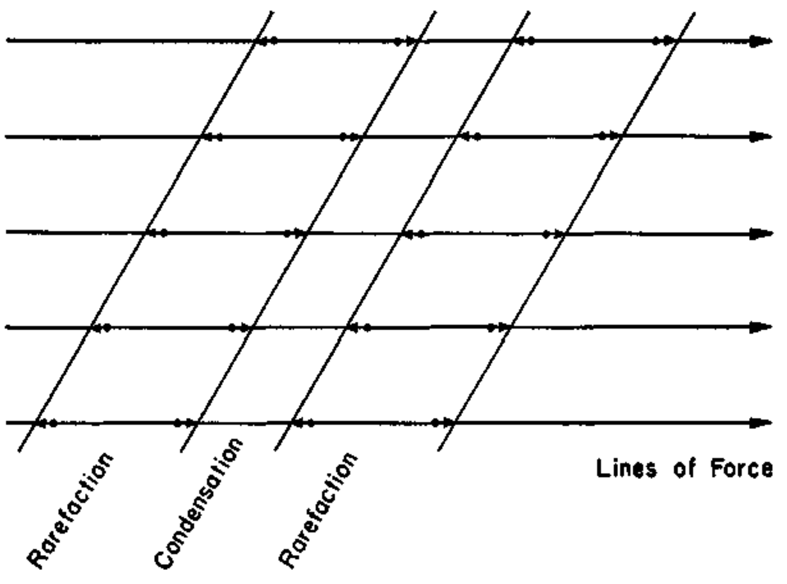


FIG. 2.—Illustrating why the presence of a magnetic field does not affect Jeans's condition for the gravitational instability of an infinite homogeneous medium.

We shall seek the solutions of equations (160) and (161) which correspond to the propagation of waves in the z-direction. Then $\partial/\partial z$ is the only nonvanishing component of the gradient. And if we further suppose that the orientation of the co-ordinate axes is so chosen that

$$\boldsymbol{H} = (0, H_y, H_z) , \tag{162}$$

it readily follows that $h_z = 0$; and we find that equations (160) and (161) break up into the two noncombining systems:

$$\rho \frac{\partial u_x}{\partial t} = \frac{H_z}{4\pi} \frac{\partial h_x}{\partial z} , \qquad \frac{\partial h_x}{\partial t} = H_z \frac{\partial u_x}{\partial z} ; \tag{163}$$

and

$$
\begin{aligned}
&\rho \frac{\partial u_y}{\partial t} - \frac{H_z}{4\pi} \frac{\partial h_y}{\partial z} = 0 , \\
&\rho \frac{\partial u_z}{\partial t} + \frac{H_y}{4\pi} \frac{\partial h_y}{\partial z} + c^2 \frac{\partial}{\partial z}\, \delta\rho - \rho \frac{\partial}{\partial z}\, \delta V = 0 , \\
&\frac{\partial h_y}{\partial t} + H_y \frac{\partial u_z}{\partial z} - H_z \frac{\partial u_y}{\partial z} = 0 , \\
&\frac{\partial}{\partial t}\, \delta\rho + \rho \frac{\partial u_z}{\partial z} = 0 , \\
&\frac{\partial^2}{\partial z^2}\, \delta V + 4\pi G \delta\rho = 0 .
\end{aligned} \tag{164}
$$

Equations (163) can be combined to give

$$
\frac{\partial^2 h_x}{\partial t^2} = \frac{H_z^2}{4\pi\rho} \frac{\partial^2 h_x}{\partial z^2} \quad \text{and} \quad \frac{\partial^2 u_x}{\partial t^2} = \frac{H_z^2}{4\pi\rho} \frac{\partial^2 u_x}{\partial t^2} . \tag{165}
$$

These equations are the same as those leading to the ordinary hydromagnetic waves of Alfvén propagated with the velocity

$$
V_A = \frac{H_z}{\sqrt{(4\pi\rho)}} . \tag{166}
$$

This mode of wave propagation is therefore unaffected by gravitation and compressibility.

Turning next to solutions of equations (164), which also represent the propagation of waves in the z-direction, we can write

$$
\frac{\partial}{\partial t} = i\omega \quad \text{and} \quad \frac{\partial}{\partial z} = -ik , \tag{167}
$$

where ω denotes the frequency and k the wave number. Making the substitutions (167) in equations (164), we obtain a system of linear homogeneous equations which can be written in matrix notation in the following form:

$$
\begin{Vmatrix}
\rho\omega & k \dfrac{H_z}{4\pi} & 0 & 0 & 0 \\
0 & -k \dfrac{H_y}{4\pi} & \rho\omega & -k c^2 & k\rho \\
k H_z & \omega & -k H_y & 0 & 0 \\
0 & 0 & -k\rho & \omega & 0 \\
0 & 0 & 0 & 4\pi G & -k^2
\end{Vmatrix}
\begin{Vmatrix}
u_y \\ h_y \\ u_z \\ \delta\rho \\ \delta V
\end{Vmatrix}
= 0 . \tag{168}
$$

The condition that equation (168) has a nontrivial solution is that the determinant of the matrix on the left-hand side should vanish. Expanding the determinant, we find that it can be reduced to the form

$$\left(\frac{\omega}{k}\right)^4 - \left\{\frac{H^2}{4\pi\rho} + \left(c^2 - \frac{4\pi G\rho}{k^2}\right)\right\}\left(\frac{\omega}{k}\right)^2 + \frac{H_z^2}{4\pi\rho}\left(c^2 - \frac{4\pi G\rho}{k^2}\right) = 0\,. \tag{169}$$

In terms of the velocity of wave propagation, $V = \omega/k$, we can rewrite equation (169) in the form

$$V^4 - (V_A^2 \sec^2\vartheta + V_J^2)V^2 + V_A^2 V_J^2 = 0\,, \tag{170}$$

where ϑ denotes the angle between the directions of $\boldsymbol{H}$ and of wave propagation and V_J and V_A have the same meanings as in equations (158) and (166).

It is seen that equation (170) allows two modes of wave propagation. If V_1 and V_2 denote the velocities of propagation of these two modes, we conclude from equation (170) that

$$V_1 V_2 = V_A V_J$$

and

$$V_1^2 + V_2^2 = V_A^2 \sec^2\vartheta + V_J^2\,. \tag{171}$$

Accordingly, *if V_J is imaginary, then either V_1 or V_2 must be imaginary.* In other words, one of the two modes of wave propagation will be unstable if Jeans's condition (159) is satisfied. The condition for gravitational instability is therefore unaffected by the presence of the magnetic field. However, as to which of the two modes will become unstable will depend on the strength of the magnetic field. Thus for $H \to 0$, the two modes given by equation (170) approach, respectively, Jeans's mode and Alfvén's mode. And if we suppose that

$$V_1 \to V_J \quad \text{and} \quad V_2 \to V_A \quad \text{as} \quad H \to 0\,, \tag{172}$$

then it follows from equation (170) that so long as $V_J^2 > 0$,

$$V_1 \to V_A \sec\vartheta \quad \text{and} \quad V_2 \to V_J \cos\vartheta \quad \text{as} \quad H \to \infty\,. \tag{173}$$

Hence, for $H \to \infty$, the mode which will become unstable when Jeans's condition is satisfied will be the mode which for $H \to 0$ is Alfvén's mode; and the mode which for $H \to 0$ is Jeans's mode becomes a hydromagnetic wave for $H \to \infty$ and is unaffected by gravitation. This "crossing-over" of the two modes with increasing strength of the magnetic field is in agreement with what is known[7] from the theory of wave propagation in a compressible medium in the absence of gravitation.

[7] Cf. H. van de Hulst, *Symposium: Problems of Cosmical Aerodynamics* (Dayton, Ohio: Central Air Documents Office, 1951), chap. vi; also N. Herlofson, *Nature*, **165**, 1020, 1950.

V. Hydrodynamic and Hydromagnetic Stability

Norman Lebovitz, in his preface to *Selected Papers*, Vol. 4, says that in the early fifties Chandra began a reconsideration of the fundamental problems of hydrodynamics from several standpoints, including the adequacy of their formulation for drawing inferences in natural settings. Regarding the adequacy of formulation, whereas one knew that rotation is persuasive and important in astrophysical and geophysical settings, the persuasive character of magnetic fields in astrophysical settings (specially on the galactic scale) had only recently been appreciated. Chandra's reconsideration generalized the fundamental problems to include the magnetic fields. Regarding mathematical formulation, the novelty of Chandra's approach was the systematic exploitation of variational methods. Related matters, such as constructing appropriate basis functions and the use of the system of equations adjoint to the one under consideration, augmented the store of mathematical techniques available for solving these problems.

Chandra, following his unique pattern, summarized the status of the field with an emphasis on the new mathematical developments in his classic monograph *Hydrodynamic and Hydromagnetic Stability* (Oxford University Press, Oxford, 1961). Seminal papers devoted to applications to astrophysics are in Part IV. The papers in this part concern applications to problems arising out of laboratory experiments.

Paper 1 extends the previous theory of stability of viscous flow between two coaxial cylinders to the case where the fluid is an electrical conductor subjected to an external magnetic field along the axis of the cylinders. Paper 2 was inspired by the laboratory experiments of D. Fultz and R. Hide on the patterns of fluid motion that occur when a radial temperature gradient is maintained in the space between two rotating coaxial cylinders. The next two papers concern the hydrodynamic stability of helium II between rotating cylinders.

Paper 5 (Rumford Medal Lecture, 1957) has at the beginning a historical introduction to the subject of thermal convection, including classic experiments of Benard and their interpretation by Lord Rayleigh. It goes on to discuss the effects of rotation and magnetic fields on convection and the instabilities that set in. Paper 6 covers essentially the same subject with more details on the theory. The last paper is an illustration of the exploitation of variational methods (alluded to by Lebovitz). It describes how variational methods have proved successful in solving characteristic-value problems in differential equations of high order that arise in the studies of hydrodynamic and hydromagnetic stability.

V. Hydrodynamic and Hydromagnetic Stability

Reprinted without change of pagination from the
Proceedings of the Royal Society, A, *volume* 216, p. 293, 1953

The stability of viscous flow between rotating cylinders in the presence of a magnetic field

By S. Chandrasekhar, F.R.S., *Yerkes Observatory*

(*Received* 4 *September* 1952)

In this paper the theory of the stability of viscous flow between two rotating coaxial cylinders which has been developed by Taylor, Jeffreys and Meksyn is extended to the case when the fluid considered is an electrical conductor and a magnetic field along the axis of the cylinders is present. A differential equation of order eight is derived which governs the situation in marginal stability; and a significant set of boundary conditions for the problem is formulated. The case when the two cylinders are rotating in the same direction and the difference (d) in their radii is small compared to their mean (R_0) is investigated in detail. A variational procedure for solving the underlying characteristic value problem and determining the critical Taylor numbers for the onset of instability is described. As in the case of thermal instability of a horizontal layer of fluid heated below, the effect of the magnetic field is to inhibit the onset of instability, the inhibiting effect being the greater, the greater the strength of the field and the value of the electrical conductivity. In both cases, the inhibiting effect of the magnetic field depends on the strength of the field (H), the density (ρ) and the coefficients of electrical conductivity (σ), kinematic viscosity (ν) and magnetic permeability (μ) through the same non-dimensional combination $Q=\mu^2H^2d^2\sigma/\rho\nu$; however, the effect on rotational stability is more pronounced than on thermal instability. A table of the critical Taylor numbers for various values of Q is provided.

1. Introduction

In a recent paper the author (Chandrasekhar 1952; this paper will be referred to hereafter as I; see also Thompson 1951) has considered the problem of the thermal instability of a layer of fluid heated below when the fluid considered is an electrical conductor and an external magnetic field is impressed on the fluid. This paper will be devoted to the related problem of the stability of viscous flow between two rotating coaxial cylinders when a magnetic field in the direction of the axis of the cylinders is present and the liquid considered is again an electrical conductor. As is well known, this latter problem, in the absence of a magnetic field, was first successfully investigated by Taylor (1923). That the two problems of thermal instability and rotational instability should be related in the framework of classical hydrodynamics was first pointed out by Low (1925, 1929), who indicated how Taylor's criterion for the onset of rotational instability in the case when the two cylinders are rotating in the same direction can be deduced from Rayleigh's criterion for thermal instability by replacing the product ($\kappa\nu$) of the coefficients of thermometric conductivity and kinematic viscosity by the square (ν^2) of the kinematic viscosity. The mathematical justification for this analogy between the two problems was given by Jeffreys (1928) who established the close similarity (and, under certain conditions, the identity) of the underlying characteristic value problems. We shall see that in the framework of magneto-hydrodynamics the same close analogy between the two problems does not exist, though in both cases the effect of the magnetic field is to inhibit the onset of instability and the extent of the inhibition depends (under certain conditions) on the strength of the impressed field

(H) and the coefficients of electrical conductivity (σ) and magnetic permeability (μ) through the same non-dimensional combination (cf. I, equation (52))

$$Q = \frac{\mu^2 H^2 \sigma}{\rho \nu} d^2, \tag{1}$$

where ρ denotes the density and $d = R_2 - R_1$ is the difference in radii of the two cylinders.

2. The equations of the problem

The basic equations of the problem are (cf. I, equations (5), (20) and (21))

$$\frac{\partial u_i}{\partial t} + u_j \frac{\partial u_i}{\partial x_j} - \frac{\mu}{4\pi\rho} H_j \frac{\partial H_i}{\partial x_j} = -\frac{\partial \varpi}{\partial x_i} + \nu \nabla^2 u_i, \tag{2}$$

$$\frac{\partial H_i}{\partial t} + u_j \frac{\partial H_i}{\partial x_j} - H_j \frac{\partial u_i}{\partial x_j} = \eta \nabla^2 H_i, \tag{3}$$

$$\frac{\partial u_i}{\partial x_i} = 0 \quad \text{and} \quad \frac{\partial H_i}{\partial x_i} = 0, \tag{4}$$

where

$$\varpi = \frac{1}{\rho}\left(p + \mu \frac{|\mathbf{H}|^2}{8\pi}\right), \tag{5}$$

$$\eta = \frac{1}{4\pi\mu\sigma}, \tag{6}$$

and the other symbols have their usual meanings. (In equations (2) to (4) summation over the repeated indices is to be understood.) For discussing the problem of the stability of viscous flow between two rotating coaxial cylinders, it is convenient to have the foregoing equations written in cylindrical polar co-ordinates (r, θ, z). Denoting by u_r, u_θ and u_z the components of the velocity and by H_r, H_θ and H_z the components of the magnetic vector in the radial, the transverse and the z-directions respectively, we find that equations (2) and (3) take the forms

$$\frac{\partial u_r}{\partial t} + (\mathbf{u}\,.\,\mathrm{grad})\,u_r - \frac{u_\theta^2}{r} - \frac{\mu}{4\pi\rho}\left\{(\mathbf{H}\,.\,\mathrm{grad})\,H_r - \frac{H_\theta^2}{r}\right\} = -\frac{\partial \varpi}{\partial r} + \nu\left(\nabla^2 u_r - \frac{2}{r^2}\frac{\partial u_\theta}{\partial \theta} - \frac{u_r}{r^2}\right), \tag{7}$$

$$\frac{\partial u_\theta}{\partial t} + (\mathbf{u}\,.\,\mathrm{grad})\,u_\theta + \frac{u_\theta u_r}{r} - \frac{\mu}{4\pi\rho}\left\{(\mathbf{H}\,.\,\mathrm{grad})\,H_\theta + \frac{H_\theta H_r}{r}\right\} = -\frac{1}{r}\frac{\partial \varpi}{\partial \theta} + \nu\left(\nabla^2 u_\theta + \frac{2}{r^2}\frac{\partial u_r}{\partial \theta} - \frac{u_\theta}{r^2}\right), \tag{8}$$

$$\frac{\partial u_z}{\partial t} + (\mathbf{u}\,.\,\mathrm{grad})\,u_z - \frac{\mu}{4\pi\rho}(\mathbf{H}\,.\,\mathrm{grad})\,H_z = -\frac{\partial \varpi}{\partial z} + \nu \nabla^2 u_z, \tag{9}$$

$$\frac{\partial H_r}{\partial t} + (\mathbf{u}\,.\,\mathrm{grad})\,H_r - (\mathbf{H}\,.\,\mathrm{grad})\,u_r = \eta\left(\nabla^2 H_r - \frac{2}{r^2}\frac{\partial H_\theta}{\partial \theta} - \frac{H_r}{r^2}\right), \tag{10}$$

$$\frac{\partial H_\theta}{\partial t} + (\mathbf{u}\,.\,\mathrm{grad})\,H_\theta - (\mathbf{H}\,.\,\mathrm{grad})\,u_\theta + \frac{1}{r}(u_\theta H_r - u_r H_\theta) = \eta\left(\nabla^2 H_\theta + \frac{2}{r^2}\frac{\partial H_r}{\partial \theta} - \frac{H_\theta}{r^2}\right), \tag{11}$$

and

$$\frac{\partial H_z}{\partial t} + (\mathbf{u}\,.\,\mathrm{grad})\,H_z - (\mathbf{H}\,.\,\mathrm{grad})\,u_z = \eta \nabla^2 H_z, \tag{12}$$

where for the sake of brevity (**u**.grad) and (**H**.grad) have been written for the operations

$$\mathbf{u}\,.\,\mathrm{grad} = u_r\frac{\partial}{\partial r}+\frac{u_\theta}{r}\frac{\partial}{\partial\theta}+u_z\frac{\partial}{\partial z} \tag{13}$$

and

$$\mathbf{H}\,.\,\mathrm{grad} = H_r\frac{\partial}{\partial r}+\frac{H_\theta}{r}\frac{\partial}{\partial\theta}+H_z\frac{\partial}{\partial z}. \tag{14}$$

Further, equations (4) in cylindrical co-ordinates are

$$\frac{\partial u_r}{\partial r}+\frac{u_r}{r}+\frac{1}{r}\frac{\partial u_\theta}{\partial\theta}+\frac{\partial u_z}{\partial z} = 0 \tag{15}$$

and

$$\frac{\partial H_r}{\partial r}+\frac{H_r}{r}+\frac{1}{r}\frac{\partial H_\theta}{\partial\theta}+\frac{\partial H_z}{\partial z} = 0. \tag{16}$$

It may be readily verified that the foregoing equations admit the stationary solution

$$u_r = u_z = 0, \quad u_\theta = V(r) = Ar+\frac{B}{r}, \quad \frac{\partial\varpi}{\partial r} = \frac{V^2}{r}, \tag{17}$$

$$H_r = H_\theta = 0 \quad \text{and} \quad H_z = H = \text{constant}, \tag{18}$$

where A and B are two constants related to the angular velocities of rotation, Ω_1 and Ω_2, of the inner and the outer cylinders (of radii R_1 and R_2, $R_2 > R_1$) by

$$A = \Omega_1\frac{1-mR_2^2/R_1^2}{1-R_2^2/R_1^2} \quad \text{and} \quad B = \Omega_1\frac{R_1^2(1-m)}{1-R_1^2/R_2^2}, \tag{19}$$

where

$$m = \Omega_2/\Omega_1. \tag{20}$$

We shall now suppose that the solution represented by equations (17) and (18) is slightly perturbed and that the components of the velocity and of the magnetic field of the perturbed motion are

$$u_r, \; V+v_\theta, \; u_z \quad \text{and} \quad h_r, \; h_\theta \text{ and } H+h_z, \tag{21}$$

respectively, where u_r, v_θ and u_z are small compared to V, and h_r, h_θ and h_z are small compared to H. And, as in Taylor's investigation, we shall further suppose that the perturbation is symmetrical about the z-axis so that u_r, v_θ, u_z, h_r, h_θ and h_z are all functions of r, z and t only. With these assumptions the linearized form of equations (7) to (12) are

$$\frac{\partial\varpi}{\partial r}-\frac{V^2}{r} = -\frac{\partial u_r}{\partial t}+\nu\left(\nabla^2u_r-\frac{u_r}{r^2}\right)+2\left(A+\frac{B}{r^2}\right)v_\theta+\frac{\mu H}{4\pi\rho}\frac{\partial h_r}{\partial z}, \tag{22}$$

$$0 = -\frac{\partial v_\theta}{\partial t}+\nu\left(\nabla^2v_\theta-\frac{v_\theta}{r^2}\right)-2Au_r+\frac{\mu H}{4\pi\rho}\frac{\partial h_\theta}{\partial z}, \tag{23}$$

$$\frac{\partial\varpi}{\partial z} = -\frac{\partial u_z}{\partial t}+\nu\nabla^2u_z+\frac{\mu H}{4\pi\rho}\frac{\partial h_z}{\partial z}, \tag{24}$$

$$-H\frac{\partial u_r}{\partial z} = -\frac{\partial h_r}{\partial t}+\eta\left(\nabla^2h_r-\frac{h_r}{r^2}\right), \tag{25}$$

$$-H\frac{\partial v_\theta}{\partial z} = -\frac{\partial h_\theta}{\partial t}+\eta\left(\nabla^2h_\theta-\frac{h_\theta}{r^2}\right)-\frac{2B}{r^2}h_r \tag{26}$$

and

$$-H\frac{\partial u_z}{\partial z} = -\frac{\partial h_z}{\partial t}+\eta\nabla^2h_z; \tag{27}$$

while for symmetrical perturbations the conditions requiring **u** and **h** to be solenoidal are

$$\frac{\partial u_r}{\partial r}+\frac{u_r}{r}+\frac{\partial u_z}{\partial z}=0 \quad\text{and}\quad \frac{\partial h_r}{\partial r}+\frac{h_r}{r}+\frac{\partial h_z}{\partial z}=0. \tag{28}$$

It may be noted that in equations (22) to (27), ∇^2 has now the meaning

$$\nabla^2=\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{\partial^2}{\partial z^2}. \tag{29}$$

We shall seek solutions of equations (22) to (28) which are of the form

$$\left.\begin{aligned} u_r &= e^{\omega t}u(r)\cos\lambda z, & h_r &= e^{\omega t}\phi(r)\sin\lambda z,\\ v_\theta &= e^{\omega t}v(r)\cos\lambda z, & h_\theta &= e^{\omega t}\psi(r)\sin\lambda z,\\ u_z &= e^{\omega t}w(r)\sin\lambda z, & h_z &= e^{\omega t}\chi(r)\cos\lambda z,\end{aligned}\right\} \tag{30}$$

where, as the notation implies, u, v, etc., are all functions of r only. For solutions of the form (30), equations (22) to (28) become

$$\frac{\partial\varpi}{\partial r}-\frac{V^2}{r}=\left[\nu\left(DD^*-\lambda^2-\frac{\omega}{\nu}\right)u+2\left(A+\frac{B}{r^2}\right)v+\frac{\mu H\lambda}{4\pi\rho}\phi\right]\cos\lambda z, \tag{31}$$

$$0=\nu\left(DD^*-\lambda^2-\frac{\omega}{\nu}\right)v-2Au+\frac{\mu H\lambda}{4\pi\rho}\psi, \tag{32}$$

$$\frac{\partial\varpi}{\partial z}=\left[\nu\left(D^*D-\lambda^2-\frac{\omega}{\nu}\right)w-\frac{\mu H\lambda}{4\pi\rho}\chi\right]\sin\lambda z, \tag{33}$$

$$+\lambda Hu=\eta\left(DD^*-\lambda^2-\frac{\omega}{\eta}\right)\phi, \tag{34}$$

$$+\lambda Hv=\eta\left(DD^*-\lambda^2-\frac{\omega}{\eta}\right)\psi-\frac{2B}{r^2}\phi, \tag{35}$$

$$-\lambda Hw=\eta\left(D^*D-\lambda^2-\frac{\omega}{\eta}\right)\chi, \tag{36}$$

$$D^*u=-\lambda w \quad\text{and}\quad D^*\phi=\lambda\chi, \tag{37}$$

where

$$D=\frac{d}{dr} \quad\text{and}\quad D^*=D+\frac{1}{r}=\frac{d}{dr}+\frac{1}{r}. \tag{38}$$

The operators D and D* satisfy the commutation relation

$$DD^*=D^*D-\frac{1}{r^2}=\frac{d^2}{dr^2}+\frac{1}{r}\frac{d}{dr}-\frac{1}{r^2}; \tag{39}$$

this relation is useful in the manipulation of equations (31) to (37).

Considering equations (31) to (37), we may first observe that equations (34), (36) and (37) are not all independent; for by operating equation (34) by D* and making use of equations (37) we recover equation (36). Next, we may eliminate ϖ from the equations by differentiating equation (31) with respect to z and equation (33) with respect to r. In this way, we obtain

$$-\frac{\nu}{\lambda}\left(DD^*-\lambda^2-\frac{\omega}{\nu}\right)Dw+\frac{\mu H}{4\pi\rho}D\chi=\nu\left(DD^*-\lambda^2-\frac{\omega}{\nu}\right)u+2\left(A+\frac{B}{r^2}\right)v+\frac{\mu H\lambda}{4\pi\rho}\phi. \tag{40}$$

Replacing w and χ in this equation by $-\mathrm{D}^*u/\lambda$ and $\mathrm{D}^*\phi/\lambda$ in accordance with equation (37) we are left with

$$\frac{\nu}{\lambda^2}\left(\mathrm{DD}^*-\lambda^2-\frac{\omega}{\nu}\right)(\mathrm{DD}^*-\lambda^2)\,u+\frac{\mu H}{4\pi\rho\lambda}(\mathrm{DD}^*-\lambda^2)\,\phi = 2\left(A+\frac{B}{r^2}\right)v. \tag{41}$$

Equation (41) together with equations (32), (34) and (35) provide a system of equations of order ten for u, v, ϕ and ψ. The fact that there are no terms in z in these equations shows that a general disturbance can be analyzed into normal modes which are simple harmonic in z and characterized by the wave-length $(2\pi/\lambda)$ of their periodicities in this direction.

3. The boundary conditions

We have seen in § 2 how the equations governing the normal modes of perturbation of the stationary solution (17) and (18) can be reduced to a system of order ten for u, v, ϕ and ψ. To make the problem determinate we need ten boundary conditions, five at each of the two bounding cylindrical surfaces.

If we assume that the rotating cylinders between which the liquid is contained are of rigid material and that no slip occurs at the surfaces of contact, then three pairs of boundary conditions follow at once; they are

$$u = v = w = 0 \quad \text{at} \quad r = R_1 \text{ and } r = R_2. \tag{42}$$

In virtue of the relation $\mathrm{D}^*u = -\lambda w$ (cf. equation (37)), the foregoing conditions are equivalent to

$$u = v = 0 \text{ and } \mathrm{D}^*u = 0 \quad \text{at} \quad r = R_1 \text{ and } R_2. \tag{43}$$

The six boundary conditions (43) must be supplemented by four others which must arise from the conditions on the components of the magnetic field h_r and h_θ at $r = R_1$ and $r = R_2$. Now, at a surface of discontinuity (such as at $r = R_1$ and $r = R_2$) the only conditions on a magnetic vector which electromagnetic theory provides are that the normal component of the magnetic induction and the tangential component of the magnetic intensity are continuous. It is important to realize that compatible with these conditions a wide range of solutions is possible. For, the perturbations in the magnetic field in the region of the rotating liquid will in general penetrate into the material of the confining cylinders for distances of the order of the wave-length $(2\pi/\lambda)$ of the perturbation; a wide range of effects must therefore be allowed for before the problem can be fully specified. However, since we are primarily interested only in the manner which the presence of a magnetic field affects the onset of instability, it would be most significant if we can in some way limit the magnetic perturbations to the liquid between the cylinders only. As Professor Fermi has pointed out to the writer,† we can accomplish this by supposing that the material of which the cylinders are made is a perfect electrical conductor; for in this case the magnetic perturbations in the liquid will not in fact penetrate into the material of the confining cylinders. However, the surfaces of the cylinders

† I should like to acknowledge here my indebtedness to Professors E. Fermi and G. Wentzel for valuable discussions on these and related matters.

in contact with the liquid will be electrically charged and there will be surface currents as well. The effect of these surface charges and surface currents will be to prevent any electric field in the radial direction and any magnetic field in the z-direction from penetrating into the cylinders. And since the surface charges and the surface currents cannot prevent the penetration of an electric field in the z-direction or a magnetic field in the radial direction, we must require that

$$E_z = 0 \text{ and } h_r = 0 \quad \text{at} \quad r = R_1 \text{ and } r = R_2. \tag{44}$$

These are the conditions proposed by Fermi; it will be convenient to refer to them as the Fermi boundary conditions.

The implication of the condition $E_z = 0$ for the components of the magnetic field can be deduced in the following way. Quite generally, the electric and the magnetic fields are related by (cf. I, equations (3))

$$\operatorname{curl} \mathbf{H} = 4\pi\sigma(\mathbf{E} + \mu \mathbf{u} \times \mathbf{H}). \tag{45}$$

The z-component of this equation (in cylindrical polar co-ordinates) is

$$\frac{\partial h_\theta}{\partial r} + \frac{h_\theta}{r} - \frac{1}{r}\frac{\partial h_r}{\partial \theta} = 4\pi\sigma[E_z + \mu\{u_r h_\theta - h_r(V + v_\theta)\}]. \tag{46}$$

Applying this equation to the boundaries ($r = R_1$ and $r = R_2$) and remembering that here u_r and h_r vanish (cf. equations (43) and (44)), we infer that the condition $E_z = 0$ is equivalent to

$$\frac{\partial h_\theta}{\partial r} + \frac{h_\theta}{r} - \frac{1}{r}\frac{\partial h_r}{\partial \theta} = 0 \quad (r = R_1 \text{ and } r = R_2). \tag{47}$$

For the particular form of the solution we have adopted (equations (30)) the condition (47) is the same as

$$\mathrm{D}^*\psi = 0 \quad (r = R_1 \text{ and } r = R_2). \tag{48}$$

A significant set of boundary conditions for the problem is, therefore,

$$u,\ v,\ \mathrm{D}^*u,\ \phi \text{ and } \mathrm{D}^*\psi \quad \text{all vanish at} \quad r = R_1 \text{ and } r = R_2. \tag{49}$$

4. The equations governing marginal stability

It is apparent from the form of the solutions sought (equations (30)) that what distinguishes stability from instability is the sign of the real part, $\mathscr{R}(\omega)$, of ω. The situation in marginal stability (i.e. on the verge of stability) must therefore be characterized by $\mathscr{R}(\omega) = 0$. Now in the case when no magnetic field is present, it has been shown by Meksyn (1946*a*) that ω is necessarily real for all unstable modes of disturbance and that the state of marginal stability obtained as the limit of these unstable modes is characterized by $\omega = 0$ (and not merely $\mathscr{R}(\omega) = 0$); in other words, the principle of the exchange of stabilities is valid. In this paper we shall assume that the principle is valid also when a magnetic field is present. We shall return to a rigorous justification of the principle in a later paper.

On the assumption then that the principle of the exchange of stabilities is valid, the equations governing marginal stability are (cf. equations (32), (34), (35) and (41))

$$\frac{\nu}{\lambda^2}(\mathrm{DD}^*-\lambda^2)^2u+\frac{\mu H}{4\pi\rho\lambda}(\mathrm{DD}^*-\lambda^2)\phi = 2\left(A+\frac{B}{r^2}\right)v, \tag{50}$$

$$\nu(\mathrm{DD}^*-\lambda^2)v = 2Au-\frac{\mu H\lambda}{4\pi\rho}\psi, \tag{51}$$

$$\lambda Hu = \eta(\mathrm{DD}^*-\lambda^2)\phi \tag{52}$$

and

$$\lambda Hv = \eta(\mathrm{DD}^*-\lambda^2)\psi-\frac{2B}{r^2}\phi. \tag{53}$$

Using equation (52) we can rewrite equation (50) in the form

$$\left[\frac{\nu}{\lambda^2}(\mathrm{DD}^*-\lambda^2)^2+\frac{\mu H^2}{4\pi\rho\eta}\right]u = 2\left(A+\frac{B}{r^2}\right)v. \tag{54}$$

And the boundary conditions with respect to which these equations must be solved are the same as before (equation (49)).

5. The case when the cylinders are rotating in the same direction and $(R_2-R_1)\ll\frac{1}{2}(R_2+R_1)$. The reduction to a characteristic value problem

A general discussion of the boundary value problem presented by equations (51) to (54) will clearly be a very difficult matter. Indeed, even in the case of zero field the only circumstance under which the problem has proved amenable to treatment is when the difference in radii of the two cylinders is small compared with their mean, i.e. when

$$d = R_2-R_1\ll\tfrac{1}{2}(R_2+R_1) = R_0 \quad \text{(say)}. \tag{55}$$

In the discussion of equations (51) to (54) we shall, therefore, limit ourselves to this case.

When (55) holds, we need not distinguish between D and D* (cf. Jeffreys 1928 and Meksyn 1946*a*), and the equations of the preceding section become

$$\left[\frac{\nu}{\lambda^2}(\mathrm{D}^2-\lambda^2)^2+\frac{\mu H^2}{4\pi\rho\eta}\right]u = 2\left(A+\frac{B}{r^2}\right)v, \tag{56}$$

$$\nu(\mathrm{D}^2-\lambda^2)v = 2Au-\frac{\mu H\lambda}{4\pi\rho}\psi, \tag{57}$$

$$\lambda Hu = \eta(\mathrm{D}^2-\lambda^2)\phi \tag{58}$$

and

$$\lambda Hv = \eta(\mathrm{D}^2-\lambda^2)\psi-\frac{2B}{r^2}\phi. \tag{59}$$

The corresponding (modified) boundary conditions are (cf. equation (49))

$$u,\ \mathrm{D}u,\ v,\ \phi \text{ and } \mathrm{D}\psi \quad \text{all vanish at} \quad r = R_1 \text{ and } r = R_2. \tag{60}$$

A further simplification is possible when the two cylinders are rotating in the same direction: for, it follows from the analysis of Taylor and Meksyn in the field-free case that in the framework of the approximation in which we do not distinguish between D and D*, we can replace $(A+B/r^2)$ and (B/r^2) by the values they take at $r = \frac{1}{2}(R_2+R_1) = R_0$. Thus, in equations (56) and (59) we shall let

$$A+\frac{B}{r^2} \simeq A+\frac{B}{R_0^2} = \tfrac{1}{2}\Omega_1\left[(1+m)-\frac{3d}{4R_0}(1-m)\right] = \Omega_0 \quad \text{(say)} \tag{61}$$

and

$$\frac{B}{r^2} \simeq \frac{B}{R_0^2}, \tag{62}$$

where it may be recalled that Ω_1 is the angular velocity of the inner cylinder and $m = \Omega_2/\Omega_1$.

In rewriting equations (56) to (59) in the framework of the foregoing approximations, we shall find it convenient to measure the radial distances from R_0 in units of d. Thus, letting

$$\zeta = \frac{1}{d}[r-\tfrac{1}{2}(R_2+R_1)] \quad \text{and} \quad \lambda = \frac{a}{d}, \tag{63}$$

we find that equations (56) to (59) can be brought to the forms

$$[(\mathrm{D}^2-a^2)^2+Qa^2]\,u = \frac{2\Omega_0 d^2}{\nu}a^2 v, \tag{64}$$

$$(\mathrm{D}^2-a^2)\,v = \frac{2Ad^2}{\nu}u-\frac{\mu H}{4\pi\rho\nu}ad\psi, \tag{65}$$

$$(\mathrm{D}^2-a^2)\,\phi = \frac{H}{\eta}adu, \tag{66}$$

and

$$(\mathrm{D}^2-a^2)\,\psi = \frac{H}{\eta}adv+\frac{2Bd^2}{R_0^2\eta}\phi, \tag{67}$$

where D now stands for $\mathrm{d}/\mathrm{d}\zeta$ and (cf. equations (1) and (6))

$$Q = \frac{\mu H^2}{4\pi\rho\eta\nu}d^2 = \frac{\mu^2 H^2\sigma}{\rho\nu}d^2. \tag{68}$$

The corresponding form of the boundary conditions is

$$u,\ \mathrm{D}u,\ v,\ \phi \ \text{and}\ \mathrm{D}\psi \quad \text{all vanish for} \quad \zeta = \pm\tfrac{1}{2}. \tag{69}$$

By successive elimination we can readily derive from equations (64) to (67) a single differential equation of order ten for any one of the variables u, v, ϕ or ψ. But a practical circumstance enables the reduction of the order of the equation we need consider to eight. The circumstance is this:

In the field-free case investigated by Taylor, the boundary value problem was reduced to a characteristic value problem for the 'Taylor number'

$$T = -\frac{2\Omega_0 d^2}{\nu}\frac{2Ad^2}{\nu}. \tag{70}$$

Since in equations (64) to (67) the combination $2Bd^2/R_0^2\eta$ occurs in addition to $2\Omega_0 d^2/\nu$ and $2Ad^2/\nu$, it is evident that the boundary value problem will now involve the further non-dimensional number

$$\tau = -\frac{2\Omega d^2}{\nu}\frac{2Bd^2}{R_0^2\eta}. \tag{71}$$

Now, under terrestrial conditions η is a million or more times larger than ν. Thus, considering mercury at room temperature as typical of the fluids to which we may want to apply the theory under terrestrial conditions, we have

$$\eta = 7{\cdot}5\times 10^3\,\mathrm{cm}^2/\mathrm{s}, \quad \text{while} \quad \nu = 1{\cdot}2\times 10^{-3}\,\mathrm{cm}^2/\mathrm{s}. \tag{72}$$

Similarly, under what we may consider as typical of astrophysical conditions (cf. I, equations (165)),

$$\eta = 7\times 10^6\,\mathrm{cm}^2/\mathrm{s}, \quad \text{while} \quad \nu = 5\times 10^3\,\mathrm{cm}^2/\mathrm{s}. \tag{73}$$

Consequently all conditions under which we may contemplate applying the present theory, τ will be a thousand to a million times smaller than T since A and B/R_0^2 should be considered as of comparable magnitudes. This is the practical circumstance to which we referred; and in view of this we may neglect the term in B/R_0^2 in equation (67) and rewrite it in the form

$$v = \frac{\eta}{Had}(\mathrm{D}^2-a^2)\psi. \tag{74}$$

Using equation (74) to eliminate v from equations (64) and (65), we have

$$[(\mathrm{D}^2-a^2)^2+Qa^2]\,u = \frac{2\Omega_0 d^2a^2}{\nu}\frac{\eta}{Had}(\mathrm{D}^2-a^2)\psi, \tag{75}$$

and
$$[(\mathrm{D}^2-a^2)^2+Qa^2]\,\psi = \frac{2Ad^2}{\nu}\frac{Had}{\eta}u. \tag{76}$$

Now combining these two equations we obtain

$$[(\mathrm{D}^2-a^2)^2+Qa^2]^2\,\psi = -Ta^2(\mathrm{D}^2-a^2)\psi, \tag{77}$$

where
$$T = -\frac{4\Omega_0 A}{\nu^2}d^4. \tag{78}$$

According to equations (74) and (76) the boundary conditions $u = 0$, $\mathrm{D}u = 0$ and $v = 0$ at $\zeta = \pm\frac{1}{2}$ are equivalent to

$$[(\mathrm{D}^2-a^2)^2+Qa^2]\,\psi = 0, \quad \mathrm{D}[(\mathrm{D}^2-a^2)^2+Qa^2]\,\psi = 0$$

and
$$(\mathrm{D}^2-a^2)\psi = 0 \quad (\zeta = \pm\tfrac{1}{2}); \tag{79}$$

in addition to these we have the further boundary condition

$$\mathrm{D}\psi = 0 \quad (\zeta = \pm\tfrac{1}{2}). \tag{80}$$

Equations (79) and (80) provide eight boundary conditions, and the requirement that a solution of equation (77) satisfy these conditions will lead to a determinate sequence of possible values for T for any given a^2. For, equation (77) being a linear equation with constant coefficients, its general solution must be of the form

$$\psi = \sum_{i=1}^{8} A_i\,\mathrm{e}^{q_i\zeta}, \tag{81}$$

where the q_i's occurring in pairs are the roots of the characteristic equation

$$[(q^2-a^2)^2+Qa^2]^2 = -Ta^2(q^2-a^2), \tag{82}$$

and the A_i's are constants of integration. The eight boundary conditions (79) and (80) (four at $\zeta = +\frac{1}{2}$ and four at $\zeta = -\frac{1}{2}$) will lead to a set of eight linear homogeneous equations for the A_i's, and the condition that the determinant of this system vanishes (so that we may have a non-trivial solution) will determine a sequence of possible values for T. Among these possible values for T (for a given a^2) there will be a lowest value; the minimum of these lowest values (as a function of a^2) will finally specify the critical Taylor number, T_c, at which instability will first set in.

While there is, of course, no difficulty of principle in carrying out the procedure outlined in the preceding paragraph, it is evident that the actual carrying out of the procedure for various assigned values of Q and a^2 will involve a prohibitive amount of work. However, it will appear that a variational principle can be established which leads to a convenient practical method of determining T_c without an undue amount of labour. The variational principle is developed in the following section.

6. A variational procedure for solving the characteristic value problem associated with equation (77)

First we shall obtain an integral expression for T. Letting

$$P = [(\mathrm{D}^2 - a^2)^2 + Qa^2]\,\psi \quad \text{and} \quad G = (\mathrm{D}^2 - a^2)\,\psi, \tag{83}$$

we can rewrite the differential equation governing ψ in the form

$$[(\mathrm{D}^2 - a^2)^2 + Qa^2]\,P = -Ta^2G. \tag{84}$$

Further, in terms of P and G the boundary conditions (79) and (80) are equivalent to

$$P = \mathrm{D}P = G = \mathrm{D}\psi = 0 \quad \text{at} \quad \zeta = \pm\tfrac{1}{2}. \tag{85}$$

Now multiply equation (84) by P and integrate over the range of ζ. The left-hand side of the equation gives

$$\int_{-\frac{1}{2}}^{+\frac{1}{2}} P(\mathrm{D}^2 - a^2)^2 P\,\mathrm{d}\zeta + Qa^2 \int_{-\frac{1}{2}}^{+\frac{1}{2}} P^2\,\mathrm{d}\zeta. \tag{86}$$

Since both P and $\mathrm{D}P$ vanish at $\zeta = \pm\frac{1}{2}$ it readily follows after two integrations by parts that

$$\int_{-\frac{1}{2}}^{+\frac{1}{2}} P(\mathrm{D}^2 - a^2) f\,\mathrm{d}\zeta = \int_{-\frac{1}{2}}^{+\frac{1}{2}} f(\mathrm{D}^2 - a^2)\,P\,\mathrm{d}\zeta, \tag{87}$$

where f is any continuous function in the interval $(+\frac{1}{2}, -\frac{1}{2})$. Using this result in (86) we have

$$\int_{-\frac{1}{2}}^{+\frac{1}{2}} P[(\mathrm{D}^2 - a^2)^2 + Qa^2]\,P\,\mathrm{d}\zeta = \int_{-\frac{1}{2}}^{+\frac{1}{2}} \{[(\mathrm{D}^2 - a^2)\,P]^2 + Qa^2P^2\}\,\mathrm{d}\zeta. \tag{88}$$

Turning next to the right-hand side of equation (84) we have to consider (cf. equations (83))

$$\int_{-\frac{1}{2}}^{+\frac{1}{2}} PG\,\mathrm{d}\zeta = \int_{-\frac{1}{2}}^{+\frac{1}{2}} G\mathrm{D}^2G\,\mathrm{d}\zeta + Qa^2 \int_{-\frac{1}{2}}^{+\frac{1}{2}} \psi\mathrm{D}^2\psi\,\mathrm{d}\zeta - a^2 \int_{-\frac{1}{2}}^{+\frac{1}{2}} (G^2 + Qa^2\psi^2)\,\mathrm{d}\zeta. \tag{89}$$

By integrating by parts the first two integrals on the right-hand side of (89) and remembering that G and $\mathrm{D}\psi$ vanish at the limits of integration, we obtain

$$\int_{-\frac{1}{2}}^{+\frac{1}{2}} PG\,\mathrm{d}\zeta = -\int_{-\frac{1}{2}}^{+\frac{1}{2}} \{(\mathrm{D}G)^2 + a^2G^2 + Qa^2[(\mathrm{D}\psi)^2 + a^2\psi^2]\}\,\mathrm{d}\zeta. \tag{90}$$

The result of multiplying equation (84) by P and integrating over ζ is, therefore,

$$T = \frac{\displaystyle\int_{-\frac{1}{2}}^{+\frac{1}{2}} \{[(\mathrm{D}^2-a^2)\,P]^2 + Qa^2P^2\}\,\mathrm{d}\zeta}{\displaystyle a^2\int_{-\frac{1}{2}}^{+\frac{1}{2}} \{(\mathrm{D}G)^2 + a^2G^2 + Qa^2[(\mathrm{D}\psi)^2 + a^2\psi^2]\}\,\mathrm{d}\zeta} = \frac{I_1}{a^2I_2} \quad \text{(say)}. \tag{91}$$

This formula expresses T as the ratio of two positive definite integrals.

Consider now the effect on T of a variation $\delta\psi$ in ψ compatible with the boundary conditions on ψ. We have, to the first order,

$$\delta T = \frac{1}{a^2I_2}\left(\delta I_1 - \frac{I_1}{I_2}\,\delta I_2\right) = \frac{1}{a^2I_2}\,(\delta I_1 - a^2T\delta I_2), \tag{92}$$

where δI_1 and δI_2 denote the corresponding variations in I_1 and I_2:

$$\delta I_1 = 2\int_{-\frac{1}{2}}^{+\frac{1}{2}} \{[(\mathrm{D}^2-a^2)\,P]\,[(\mathrm{D}^2-a^2)\,\delta P] + Qa^2P\delta P\}\,\mathrm{d}\zeta \tag{93}$$

and
$$\delta I_2 = 2\int_{-\frac{1}{2}}^{+\frac{1}{2}} \{(\mathrm{D}G)\,(\mathrm{D}\delta G) + a^2G\delta G + Qa^2[(\mathrm{D}\psi)\,(\mathrm{D}\delta\psi) + a^2\psi\delta\psi]\}\,\mathrm{d}\zeta. \tag{94}$$

Making use of the boundary conditions (85) and

$$\delta P = \mathrm{D}\delta P = \delta G = \mathrm{D}\delta\psi = 0 \quad \text{at} \quad \zeta = \pm\tfrac{1}{2}, \tag{95}$$

we can reduce the expressions for δI_1 and δI_2 by one or more integrations by parts. Thus

$$\delta I_1 = 2\int_{-\frac{1}{2}}^{+\frac{1}{2}} \delta P\{[(\mathrm{D}^2-a^2)^2 + Qa^2]\,P\}\,\mathrm{d}\zeta \tag{96}$$

and
$$\begin{aligned}\delta I_2 &= -2\int_{-\frac{1}{2}}^{+\frac{1}{2}} \{G[(\mathrm{D}^2-a^2)\,\delta G] + Qa^2[(\mathrm{D}^2-a^2)\,\psi]\,\delta\psi\}\,\mathrm{d}\zeta \\ &= -2\int_{-\frac{1}{2}}^{+\frac{1}{2}} G\{[(\mathrm{D}^2-a^2)^2 + Qa^2]\,\delta\psi\}\,\mathrm{d}\zeta = -2\int_{-\frac{1}{2}}^{+\frac{1}{2}} G\delta P\,\mathrm{d}\zeta.\end{aligned} \tag{97}$$

Now combining equations (92), (96) and (97), we obtain

$$\delta T = \frac{2}{a^2I_2}\int_{-\frac{1}{2}}^{+\frac{1}{2}} \delta P\{[(\mathrm{D}^2-a^2)^2 + Qa^2]\,P + Ta^2G\}\,\mathrm{d}\zeta. \tag{98}$$

Hence to the first order $\delta T \equiv 0$ *for all small arbitrary variations* δP provided

$$[(\mathrm{D}^2-a^2)^2 + Qa^2]\,P + Ta^2G = 0, \tag{99}$$

i.e. if the differential equation governing ψ is satisfied. It is evident that the converse of this proposition is also true. Moreover, it follows from (98) that the true solution of the problem leads to minimum value for T when evaluated according to formula (91). This last fact enables us to formulate the following variational procedure for solving equation (77) (for any assigned a^2) and satisfying the boundary conditions of the problem.

Assume for DP an expression involving one or more parameters A_k and which vanishes for $\zeta = \pm\frac{1}{2}$. Integrating the expression for DP, obtain P and adjust the constant of integration so that P may also vanish for $\zeta = \pm\frac{1}{2}$. (The fact that both P and DP must vanish for $\zeta = \pm\frac{1}{2}$ implies a certain restriction in the form for DP we can assume.) Next solve the equation (cf. equation (83))

$$[(\mathrm{D}^2-a^2)^2+Qa^2]\,\psi = P \tag{100}$$

for ψ and determine the constants of integration so as to satisfy the boundary conditions $\mathrm{D}\psi = (\mathrm{D}^2-a^2)\,\psi = 0$ for $\zeta = \pm\frac{1}{2}$. And, finally, evaluate T according to the formula (91) and minimize it with respect to the parameters A_k. In this way we shall obtain the 'best' value of T for the chosen form of DP.

7. The critical Taylor number as a function of Q for the case when the two cylinders are rotating in the same direction and $d \ll R_0$

We shall obtain the solution of the characteristic value problem formulated in § 5 by an application of the variational procedure described in § 6. But first we may observe that the identity of the boundary conditions to be satisfied at $\zeta = \pm\frac{1}{2}$ requires that the proper solutions, ψ, of equation (77) fall into two non-combining groups of odd and even solutions, respectively. And it can be readily verified that the even solutions give lower values for T. We must therefore choose the even solutions.

As a simple trial function for DP which is odd and vanishes for $\zeta = \pm\frac{1}{2}$, we shall assume

$$\mathrm{D}P = \tfrac{1}{2}\sin\pi\zeta - \zeta + \mathscr{A}\sin 2\pi\zeta, \tag{101}$$

where $\mathscr{A}$ is a variational parameter. On integration, equation (101) yields

$$P = \tfrac{1}{8} - \tfrac{1}{2}\zeta^2 - \frac{1}{2\pi}\cos\pi\zeta - \frac{\mathscr{A}}{2\pi}(1+\cos 2\pi\zeta), \tag{102}$$

where the constant of integration has been adjusted to make P vanish at $\zeta = \pm\frac{1}{2}$. With P given by equation (102), the equation governing ψ is

$$[(\mathrm{D}^2-a^2)^2+Qa^2]\,\psi = \tfrac{1}{8} - \tfrac{1}{2}\zeta^2 - \frac{1}{2\pi}\cos\pi\zeta - \frac{\mathscr{A}}{2\pi}(1+\cos 2\pi\zeta). \tag{103}$$

The solution of this equation appropriate to the problem on hand is

$$\psi = \tfrac{1}{8}\gamma_0 - \tfrac{1}{2}\gamma_0(\zeta^2+4a^2\gamma_0) - \frac{\gamma_1}{2\pi}\cos\pi\zeta - \frac{\mathscr{A}}{2\pi}(\gamma_0+\gamma_2\cos 2\pi\zeta) + B\cosh\beta_1\zeta + C\cosh\beta_2\zeta, \tag{104}$$

where B and C are constants of integration,

$$\gamma_n = [(n^2\pi^2+a^2)^2+Qa^2]^{-1}, \tag{105}$$

$$\left.\begin{aligned} &\beta_1 = \alpha_1 + \mathrm{i}\alpha_2, && \beta_2 = \alpha_1 - \mathrm{i}\alpha_2,\\ &\alpha_1 = \{\tfrac{1}{2}\sqrt{(a^4+Qa^2)}+\tfrac{1}{2}a^2\}^{\frac{1}{2}} \quad\text{and} && \alpha_2 = \{\tfrac{1}{2}\sqrt{(a^4+Qa^2)}-\tfrac{1}{2}a^2\}^{\frac{1}{2}}.\end{aligned}\right\} \tag{106}$$

Since β_1 and β_2 are complex conjugates, B and C must also be complex conjugates. From (104) it follows that

$$\mathrm{D}\psi = -\gamma_0\zeta + \tfrac{1}{2}\gamma_1 \sin \pi\zeta + \mathscr{A}\gamma_2 \sin 2\pi\zeta + B\beta_1 \sinh \beta_1\zeta + C\beta_2 \sinh \beta_2 \zeta \tag{107}$$

and

$$(\mathrm{D}^2 - a^2)\,\psi = -\gamma_0(\tfrac{1}{8}c_0 + q) + \tfrac{1}{2}\gamma_0 c_0 \zeta^2 + \frac{\gamma_1 c_1}{2\pi}\cos \pi\zeta + \frac{\mathscr{A}}{2\pi}(\gamma_0 c_0 + \gamma_2 c_2 \cos 2\pi\zeta) + \mathrm{i}a\sqrt{Q}\,(B\cosh \beta_1\zeta - C\cosh \beta_2\zeta), \tag{108}$$

where we have introduced the further abbreviations

$$c_n = n^2\pi^2 + a^2 \quad \text{and} \quad q = (Q-a^2)/(Q+a^2). \tag{109}$$

The vanishing of $\mathrm{D}\psi$ and $(\mathrm{D}^2-a^2)\,\psi$ at $\zeta = \pm\frac{1}{2}$ require that

$$B\beta_1 \sinh \tfrac{1}{2}\beta_1 + C\beta_2 \sinh \tfrac{1}{2}\beta_2 = \tfrac{1}{2}(\gamma_0 - \gamma_1) \tag{110}$$

and

$$B\cosh \tfrac{1}{2}\beta_1 - C\cosh \tfrac{1}{2}\beta_2 = -\frac{\mathrm{i}}{a\sqrt{Q}}\left[\gamma_0 q + \frac{\mathscr{A}}{2\pi}(\gamma_2 c_2 - \gamma_0 c_0)\right]. \tag{111}$$

The solution of these equations can be expressed in the forms

$$\left.\begin{aligned} B &= \delta\left[\tfrac{1}{2}(\gamma_0-\gamma_1) - \frac{\mathrm{i}}{a\sqrt{Q}}\left\{\gamma_0 q + \frac{\mathscr{A}}{2\pi}(\gamma_2 c_2 - \gamma_0 c_0)\right\}\beta_2 \tanh \tfrac{1}{2}\beta_2\right]\operatorname{sech} \tfrac{1}{2}\beta_1, \\ C &= \delta\left[\tfrac{1}{2}(\gamma_0-\gamma_1) + \frac{\mathrm{i}}{a\sqrt{Q}}\left\{\gamma_0 q + \frac{\mathscr{A}}{2\pi}(\gamma_2 c_2 - \gamma_0 c_0)\right\}\beta_1 \tanh \tfrac{1}{2}\beta_1\right]\operatorname{sech} \tfrac{1}{2}\beta_2, \end{aligned}\right\} \tag{112}$$

where

$$\delta = (\beta_1 \tanh \tfrac{1}{2}\beta_1 + \beta_2 \tanh \tfrac{1}{2}\beta_2)^{-1}. \tag{113}$$

With P and $(\mathrm{D}^2-a^2)\,\psi$ given by equations (102) and (108), we find, after some lengthy reductions, that

$$\begin{aligned} \int_{-\frac{1}{2}}^{+\frac{1}{2}} \{[(\mathrm{D}^2-a^2)\,P]^2 + Qa^2P^2\}\,\mathrm{d}\zeta = {} & \Big[-\tfrac{1}{8}(a^2+8)\left\{1 + a^2\left(\frac{2}{\pi^2} - \frac{1}{24}\right)\right\} \\ & + \frac{a^4}{320} + \frac{1}{8\pi^4}(\pi^2+a^2)(\pi^4+3\pi^2a^2-16a^2) + Qa^2\left(\frac{1}{120}+\frac{1}{8\pi^2}-\frac{2}{\pi^4}\right)\Big] \\ & + \frac{1}{\pi}\left[a^4\left(\frac{1}{\pi^2}-\frac{1}{12}\right) + \frac{1}{12\pi^2}(4\pi^2+a^2)^2 + \frac{Qa^2}{12}\left(\frac{13}{\pi^2}-1\right)\right]\mathscr{A} \\ & + \frac{1}{4\pi^2}[a^4 + \tfrac{1}{2}(4\pi^2+a^2)^2 + \tfrac{3}{2}Qa^2]\,\mathscr{A}^2 \end{aligned} \tag{114}$$

and

$$\begin{aligned} -\int_{-\frac{1}{2}}^{+\frac{1}{2}} P(\mathrm{D}^2-a^2)\,\psi\,\mathrm{d}\zeta = {} & \Big[-\left(\frac{1}{\pi^2}-\frac{1}{12}\right)\gamma_0 q - \left(\frac{1}{\pi^4}-\frac{1}{120}\right)\gamma_0 c_0 + \frac{\pi^2-8}{8\pi^4}\gamma_1 c_1 \\ & -\gamma_0(\gamma_0-\gamma_1)\,\Delta a\sqrt{Q}\,\{2\gamma_0^{\frac{1}{2}}F(\alpha) + \pi^2\gamma_1(c_0+c_1)\,f(\alpha)\,a\sqrt{Q}\} \\ & \qquad - 2q\gamma_0^{2}\Delta\{2q\gamma_0^{-\frac{1}{2}}g(\alpha) + \pi^2\gamma_1\mathfrak{F}(\alpha)\}\Big] \\ & + \frac{1}{2\pi}\Big[\left(\frac{3}{4\pi^2}-\frac{1}{6}\right)\gamma_0 c_0 + \frac{4}{3\pi^2}\gamma_1 c_1 + \frac{1}{12\pi^2}\gamma_2 c_2 - \gamma_0 q \\ & - 8\pi^2\gamma_0^{\frac{1}{2}}\gamma_2(\gamma_0-\gamma_1)\,\Delta\mathfrak{G}(\alpha)\,a\sqrt{Q} + 16\pi^2\gamma_0^{\frac{1}{2}}\gamma_2 c_0(c_2-Q)\,\Delta g(\alpha)\,q\gamma_0 \\ & - 2\Delta\gamma_0(\gamma_2 c_2 - \gamma_0 c_0)\,\{2q\gamma_0^{-\frac{1}{2}}g(\alpha) + \pi^2\gamma_1\mathfrak{F}(\alpha)\}\Big]\mathscr{A} \\ & + \frac{1}{4\pi^2}[\gamma_0 c_0 + \tfrac{1}{2}\gamma_2 c_2 + 16\pi^2\gamma_0^{\frac{1}{2}}\gamma_2 c_0(c_2-Q)\,\Delta g(\alpha)\,(\gamma_2 c_2 - \gamma_0 c_0)]\,\mathscr{A}^2, \end{aligned} \tag{115}$$

$$\left.\begin{aligned}
\text{where}\quad \Delta &= [2(\alpha_1 \sinh\alpha_1 - \alpha_2 \sin\alpha_2)]^{-1},\\
f(\alpha) &= \cosh\alpha_1 + \cos\alpha_2, \quad g(\alpha) = \cosh\alpha_1 - \cos\alpha_2,\\
F(\alpha) &= \alpha_1(\alpha_1^2 - 3\alpha_2^2)\sin\alpha_2 + \alpha_2(\alpha_2^2 - 3\alpha_1^2)\sinh\alpha_1,\\
\mathfrak{F}(\alpha) &= [c_1\alpha_1(\alpha_1^2 - 3\alpha_2^2) + \alpha_2(\alpha_2^2 - 3\alpha_1^2)\, a\sqrt{Q}]\sinh\alpha_1\\
&\quad + [c_1\alpha_2(\alpha_2^2 - 3\alpha_1^2) - \alpha_1(\alpha_1^2 - 3\alpha_2^2)\, a\sqrt{Q}]\sin\alpha_2\\
\text{and}\quad \mathfrak{G}(\alpha) &= (\alpha_1 c_2 - \alpha_2 a\sqrt{Q})\sin\alpha_2 - (c_2\alpha_2 + \alpha_1 a\sqrt{Q})\sinh\alpha_1.
\end{aligned}\right\} \tag{116}$$

The resulting expression for T is therefore of the form

$$T = -\frac{\int_{-\frac{1}{2}}^{+\frac{1}{2}} \{[(D^2-a^2)P]^2 + Qa^2P^2\}\, d\zeta}{a^2\int_{-\frac{1}{2}}^{+\frac{1}{2}} P(D^2-a^2)\psi\, d\zeta} = \frac{f + g\mathscr{A} + h\mathscr{A}^2}{l + m\mathscr{A} + n\mathscr{A}^2}. \tag{117}$$

As a function of $\mathscr{A}$, T attains its minimum value when $\mathscr{A}$ is a root of the equation

$$(gl - mf) + 2(hl - fn)\mathscr{A} + (hm - gn)\mathscr{A}^2 = 0. \tag{118}$$

With $\mathscr{A}$ determined by this equation, equation (117) will give the best value of T for any assigned a^2. By repeating the calculations for various assigned values of a^2 we can determine T_c. The results of such calculations are summarized in table 1 and further illustrated in figures 1 and 2.

TABLE 1. CRITICAL TAYLOR NUMBERS FOR THE CASE WHEN THE TWO CYLINDERS ARE ROTATING IN THE SAME DIRECTION AND $d \ll R_0$

Q	a	$\mathscr{A}$	T_c	T_c/T_0
0	3·15	−0·02697	$1{\cdot}708\times10^3$†	1·000
5	3·2	−0·02683	$2{\cdot}184\times10^3$	1·279
10	3·3	−0·02724	$2{\cdot}691\times10^3$	1·575
20	3·4	−0·02657	$3{\cdot}808\times10^3$	2·229
50	3·45	−0·00097	$7{\cdot}997\times10^3$	4·681
100	3·35	+0·01496	$1{\cdot}757\times10^4$	10·29
200	2·9	+1·309	$4{\cdot}472\times10^4$	26·18
400	2·2	−0·1979	$1{\cdot}179\times10^5$	69·02
1000	1·45	−0·1446	$3{\cdot}786\times10^5$	221·6
4000	0·77	−0·1313	$1{\cdot}738\times10^6$	1018
10000	0·50	−0·1289	$4{\cdot}459\times10^6$	2610

† The 'exact value' of this number is 1707·8 (cf. Pellew & Southwell 1940).

It is apparent from figure 1 that a magnetic field has, depending on the value of Q, an inhibiting effect on the onset of rotational instability; and this inhibiting effect is greater, the greater H and σ are. In a general way the reason for this inhibiting effect is clear. In an electrically conducting fluid, the magnetic lines of force have a tendency to be dragged with the fluid, the attachment of the fluid to the lines of force being the stronger, the stronger the magnetic field. Consequently, as the strength of the field is increased, motions at right angles to **H** become increasingly 'difficult' and this prevents an 'easy' closing in of the stream lines required for the onset of instability. However, this picture should not be taken too literally, for, if it were strictly true, we should expect that the wave-length of the mode of disturbance for which instability first sets in should steadily increase with H. But this is not the case, since, according to the results of table 1 (see also figure 2 in this connexion), a ($= 2\pi d$/wave-length) first increases with Q; and it is only for

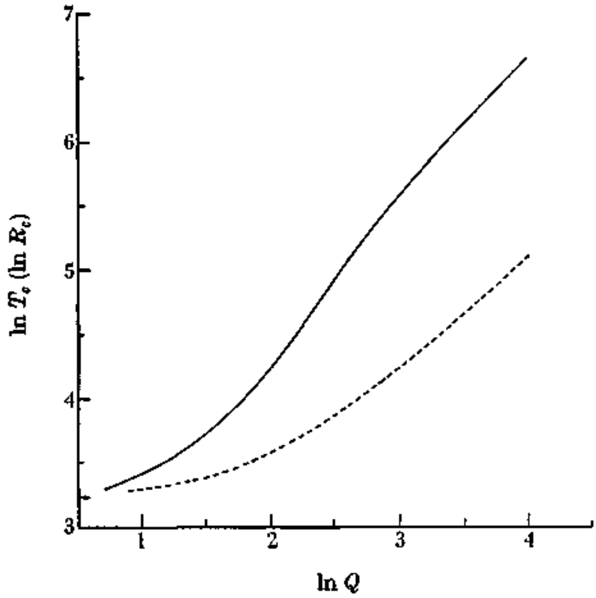


FIGURE 1. The variations of the critical Taylor number T_c (full-line curve) for the onset of rotational instability and the critical Rayleigh number R_c (dashed curve) for the onset of thermal instability as functions of Q. (The Rayleigh numbers are for the case when the layer of fluid heated below is confined between two rigid planes (Chandrasekhar 1952, table 2).)

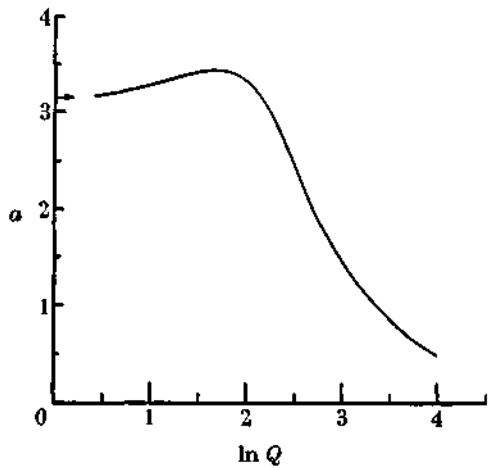


FIGURE 2. The variation of the wave number a (in the unit $1/d$) of the disturbance at which instability first sets in as a function of Q.

$Q > 50$ that it starts decreasing. It is possible that the initial increase in a has to be traced to the fact that the magnetic field has to compete with angular momentum as well

In order to see how effective the magnetic field will be in practical cases in inhibiting rotational instability, we shall consider mercury at room temperature. We have

$$\sigma = 1{\cdot}1\times10^{-5},\quad \rho\nu = 1{\cdot}7\times10^{-2}\quad\text{and}\quad Q = 6\times10^{-4}H^2d^2, \tag{119}$$

where H is measured in gauss. For $d = 1$ cm, $H = 10^3$ gauss, $Q = 600$, and we find from figure 1 that the critical Taylor number for instability is $1{\cdot}4\times10^5$, i.e. about eighty times the value for zero field. The predicted effect is therefore very pronounced and should be detectable in the laboratory. Indeed, it would appear that the effect should be detectable even with ordinary acids and acid solutions. Thus, for nitric acid at room temperature,

$$\sigma = 8\times10^{-10},\quad \rho\nu = 2\times10^{-2}\quad\text{and}\quad Q = 4\times10^{-8}H^2d^2. \tag{120}$$

Accordingly, under ordinary laboratory conditions the effect should become measurable for $H \sim 10^4$ gauss.

Finally, we may compare the inhibiting effect of a magnetic field on rotational and thermal instabilities. In the absence of a magnetic field it is known (Low 1925; Jeffreys 1928) that the critical Rayleigh number for thermal instability in a horizontal layer of fluid heated below and confined between rigid planes is the same as the critical Taylor number for rotational instability under the circumstances in which the problem has been investigated (i.e. when the cylinders are rotating in the same direction and $d \ll R_0$). When a magnetic field is present this identity between the Taylor and the Rayleigh numbers no longer exists, though both continue to be functions of the same non-dimensional parameter Q. Also, the inhibition of rotational instability by a magnetic field is very much more pronounced than the inhibition of thermal instability. To illustrate the extent to which the two effects differ, the Rayleigh number (R_c) is also plotted in figure 1 as a function of Q. It will be seen how rapidly the two curves diverge for increasing Q. Thus, for $Q = 10^4$, R_c and T_c differ by a factor 36. A further difference between the two problems is that while in the case of thermal instability the cells which appear at marginal stability get progressively elongated as the field increases, in the case of rotational instability the cells are at first a little compressed and they begin to get elongated only after the field has increased beyond a certain strength.

8. Concluding remarks

The present paper completes the discussion of the stability of viscous flow between cylinders rotating in the same direction and in the presence of a magnetic field to the same extent that the problem has been discussed in the absence of a field by Taylor (1923) and Meksyn (1946*a*). If the cylinders rotate in opposite directions, a new situation arises, in that the angular velocity has then a node in the fluid and approximations which can be made when the cylinders rotate in the same direction cease to be valid. However, by making suitable approximations Meksyn (1946*b*, *c*) has succeeded in obtaining an analytical solution for this case also. The extension of Meksyn's analysis to the case when a magnetic field is present will be given in a later paper.

A related problem which the writer hopes to consider soon is the modification of the analysis of this paper to the case when a pressure gradient parallel to the axes of the cylinders is present; the corresponding analysis in the absence of a field has been carried out by Goldstein (1937).

In conclusion I wish to record my indebtedness to Miss Donna Elbert for valuable assistance with the numerical work.

References

Chandrasekhar, S. 1952 *Phil. Mag.* [7], **43**, 501.
Goldstein, S. 1937 *Proc. Camb. Phil. Soc.* **33**, 41.
Jeffreys, H. 1928 *Proc. Roy. Soc.* A, **118**, 170.
Low, A. R. 1925 *Nature, Lond.*, **115**, 299.
Low, A. R. 1929 *Proc. Roy. Soc.* A, **125**, 180.
Meksyn, D. 1946*a* *Proc. Roy. Soc.* A, **187**, 115.
Meksyn, D. 1946*b* *Proc. Roy. Soc.* A, **187**, 480.
Meksyn, D. 1946*c* *Proc. Roy. Soc.* A, **187**, 492.
Pellew, A. & Southwell, R. V. 1940 *Proc. Roy. Soc.* A, **176**, 312.
Taylor, G. I. 1923 *Phil. Trans.* A, **223**, 289.
Thompson, W. B. 1951 *Phil. Mag.* [7], **42**, 1417.

PRINTED IN GREAT BRITAIN AT THE UNIVERSITY PRESS, CAMBRIDGE
(BROOKE CRUTCHLEY, UNIVERSITY PRINTER)

The Stability of Viscous Flow between Rotating Cylinders in the Presence of a Radial Temperature Gradient

S. CHANDRASEKHAR

Yerkes Observatory, University of Chicago

CONTENTS

The research reported in this paper has in part been supported by the Geophysics Research Directorate of the Air Force Cambridge Research Center, Air Research and Development Command, under Contract AF 19(604)-299 with the University of Chicago.

1. Introduction. Recent experimental studies by D. Fultz [1] and R. Hide [2] on the patterns of fluid motion which occur in the space between two rotating co-axial cylinders in the presence of a radial temperature gradient have disclosed remarkable analogies with the phenomenon of the jet stream in the upper atmosphere. It is the opinion of Fultz and Hide that under the conditions of their experiments the effect of gravity was at least as important as that of the Coriolis force. Nevertheless, in this paper we shall ignore the effect of gravity and consider the simpler problem of the stability of the two-dimensional viscous flow between rotating cylinders when a radial temperature gradient is present. This simpler problem is of some interest in itself: it discloses a type of rotationally induced thermal instability which appears to be novel; and as we shall see this new type of thermal instability has several features in common with the more familiar Bénard type of gravitationally induced thermal instability.

It may be stated here that the present paper is one of a series which is devoted to a systematic study of the various problems of stability in hydrodynamics and hydromagnetics (Chandrasekhar [3]–[9]; also Chandrasekhar & Fermi [10]).

2. The equations of the problem. As we have stated we shall consider the problem as two-dimensional. The equations of motion and heat conduction appropriate to the problem on hand are in cylindrical polar co-ordinates (r, θ),

$$(1)\qquad \rho\left(\frac{\partial}{\partial t} + u_r\frac{\partial}{\partial r} + \frac{u_\theta}{r}\frac{\partial}{\partial \theta}\right)u_r - \rho\frac{u_\theta^2}{r} = -\frac{\partial p}{\partial r} + \rho\nu\left(\nabla^2 u_r - \frac{2}{r^2}\frac{\partial u_\theta}{\partial \theta} - \frac{u_r}{r^2}\right),$$

$$(2)\qquad \rho\left(\frac{\partial}{\partial t} + u_r\frac{\partial}{\partial r} + \frac{u_\theta}{r}\frac{\partial}{\partial \theta}\right)u_\theta + \rho\frac{u_\theta u_r}{r} = -\frac{1}{r}\frac{\partial p}{\partial \theta} + \rho\nu\left(\nabla^2 u_\theta + \frac{2}{r^2}\frac{\partial u_r}{\partial \theta} - \frac{u_\theta}{r^2}\right)$$

and

$$(3)\qquad \left(\frac{\partial}{\partial t} + u_r\frac{\partial}{\partial r} + \frac{u_\theta}{r}\frac{\partial}{\partial \theta}\right)T = \kappa\nabla^2 T,$$

where ρ, p and T denote the density, pressure and temperature, u_r and u_θ the components of the velocity in the radial and the transverse (θ) directions and ν and κ are the coefficients of kinematic viscosity and thermometric conductivity, respectively. Also, we have the "equation of state"

$$(4)\qquad \rho = \rho_0(1 - \alpha\Delta T), \qquad \Delta T = T - T_0,$$

where α denotes the coefficient of volume expansion, ρ_0 the density corresponding to a mean temperature T_0 and ΔT is the deviation of the local temperature from T_0.

It may be readily verified that equations (1) and (2) admit the stationary solution

$$(5)\qquad u_r = 0, \qquad u_\theta = V(r) = Ar + \frac{B}{r} \qquad \text{and} \qquad p = p_0(r),$$

where A and B are constants (related to the angular velocities of rotation, Ω_1 and Ω_2 of the inner and the outer cylinders of radii R_1 and R_2, $R_2 > R_1$) and $p_0(r)$ is determined by

$$\frac{\partial p_0}{\partial r} = \rho_0\{1 - \alpha\Delta T_0(r)\} \frac{V^2}{r}. \tag{6}$$

In equation (6), $\Delta T_0(r)$ represents a stationary solution of equation (3). Regarding this latter, we shall make two assumptions:

$$\begin{aligned} &\text{case I:} && \kappa\nabla^2\{\Delta T_0(r)\} = \epsilon, \\ (7)\quad \text{and} \\ &\text{case II:} && \kappa\nabla^2\{\Delta T_0(r)\} = 0. \end{aligned}$$

Case I will be appropriate to a situation when there is a uniform distribution of heat sources such that in the absence of conduction the temperature at each point will rise at the rate ϵ; while case II will be appropriate to a situation when the two cylinders confining the liquid are maintained at constant, different, temperatures. The stationary temperature gradients maintained in the two cases will be

$$\text{case I:} \qquad \frac{d}{dr}\Delta T_0(r) = \beta r \qquad \left(\beta = \frac{\epsilon}{2\kappa} = \text{constant}\right) \tag{8}$$

and

$$\text{case II:} \qquad \frac{d}{dr}\Delta T_0(r) = \frac{\beta}{r} \qquad (\beta = \text{constant}). \tag{9}$$

In general we shall write

$$\frac{d}{dr}\Delta T_0(r) = \beta(r). \tag{10}$$

We shall now suppose that the solution represented by equations (5) and (6) is slightly perturbed. Let this perturbed motion be characterized by

$$u_r = u, \qquad u_\theta = V + v, \qquad \Delta T = \Delta T_0(r) + \tau \qquad \text{and} \qquad p = p_0 + \delta p, \tag{11}$$

where u and v are small compared to V while τ and δp are small compared to ΔT_0 and p_0, respectively. With these assumptions the linearized form of equations (1) and (2) are

$$\rho_0\left(\frac{\partial u}{\partial t} + \frac{V}{r}\frac{\partial u}{\partial \theta}\right) + \rho_0\alpha\tau\frac{V^2}{r} - \rho_0\frac{2Vv}{r} = -\frac{\partial}{\partial r}\delta p + \rho_0\nu\left(\nabla^2 u - \frac{2}{r^2}\frac{\partial v}{\partial \theta} - \frac{u}{r^2}\right) \tag{12}$$

and

$$\rho_0\left(\frac{\partial v}{\partial t} + u\frac{\partial V}{\partial r} + \frac{V}{r}\frac{\partial v}{\partial \theta}\right) + \rho_0\frac{Vu}{r} = -\frac{1}{r}\frac{\partial}{\partial \theta}\delta p + \rho_0\nu\left(\nabla^2 v + \frac{2}{r^2}\frac{\partial u}{\partial \theta} - \frac{v}{r^2}\right), \tag{13}$$

where it should be noted that in obtaining these equations the further assumption has been made that α, except when multiplied by V^2, leads to a quantity of the second order of smallness. The assumption in particular that $\alpha\tau V^2$ is a quantity of the first order of smallness is a necessary one: on it depends the onset of instability when the temperature gradient which is maintained exceeds a certain critical value. It may be recalled in this connection that a similar assumption is made in the theory of the gravitationally induced thermal instability (Rayleigh [**11**], Jeffreys [**12**]) when the effect of the variation in density (due to thermal expansion) is ignored in all terms in the equations of motion except the one in gravity.

In the framework of the approximations leading to equations (12) and (13) the equation of continuity reduces simply to the statement that the velocity is a solenoidal vector. Thus,

$$\frac{\partial u}{\partial r}+\frac{u}{r}+\frac{1}{r}\frac{\partial v}{\partial \theta}=0. \tag{14}$$

On making use of the particular form of V (equation (5)) we find that equations (12) and (13) can be reduced to the forms

$$\frac{\partial \varpi}{\partial r}=-\frac{\partial u}{\partial t}-\frac{V}{r}\frac{\partial u}{\partial \theta}-\alpha\tau\frac{V^2}{r}+2\frac{Vv}{r}+\nu\left(\nabla^2 u-\frac{2}{r^2}\frac{\partial v}{\partial \theta}-\frac{u}{r^2}\right) \tag{15}$$

and

$$\frac{1}{r}\frac{\partial \varpi}{\partial \theta}=-\frac{\partial v}{\partial t}-2Au-\frac{V}{r}\frac{\partial v}{\partial \theta}+\nu\left(\nabla^2 v+\frac{2}{r^2}\frac{\partial u}{\partial \theta}-\frac{v}{r^2}\right), \tag{16}$$

where for the sake of brevity we have written

$$\varpi=\delta p/\rho_0. \tag{17}$$

Eliminating ϖ between equations (15) and (16) we obtain

$$\begin{aligned}\frac{\partial}{\partial \theta}&\left\{-\frac{\partial u}{\partial t}-\frac{V}{r}\frac{\partial u}{\partial \theta}-\alpha\tau\frac{V^2}{r}+2\frac{Vv}{r}+\nu\left(\nabla^2 u-\frac{2}{r^2}\frac{\partial v}{\partial \theta}-\frac{u}{r^2}\right)\right\}\\&=\frac{\partial}{\partial r}r\left\{-\frac{\partial v}{\partial t}-2Au-\frac{V}{r}\frac{\partial v}{\partial \theta}+\nu\left(\nabla^2 v+\frac{2}{r^2}\frac{\partial u}{\partial \theta}-\frac{v}{r^2}\right)\right\}.\end{aligned} \tag{18}$$

Returning to the equation of heat conduction (3), we find that the corresponding linearized form of this equation is

$$\frac{\partial \tau}{\partial t}+\beta(r)u+\frac{V}{r}\frac{\partial \tau}{\partial \theta}=\kappa\nabla^2\tau \tag{19}$$

We shall now seek solutions of equations (14), (18) and (19) which are of the forms

$$u=u(r)e^{i(n\theta+\sigma t)},\qquad v=v(r)e^{i(n\theta+\sigma t)},\qquad \tau=\tau(r)e^{i(n\theta+\sigma t)}, \tag{20}$$

where n is an integer and σ is a constant unspecified for the present. For solutions of the form (20) equations (14), (18) and (19) become

$$rD^* u = D(ru) = -inv, \tag{21}$$

$$\begin{aligned} &in\left\{-i\sigma u - in\frac{Vu}{r} - \alpha\tau\frac{V^2}{r} + 2\frac{Vv}{r} + \nu\left[\left(DD^* - \frac{n^2}{r^2}\right)u - 2in\frac{v}{r^2}\right]\right\} \\ &\qquad = rD^*\left\{-i\sigma v - 2Au - in\frac{Vv}{r} + \nu\left[\left(DD^* - \frac{n^2}{r^2}\right)v + 2in\frac{u}{r^2}\right]\right\} \end{aligned} \tag{22}$$

and

$$i\sigma\tau + \beta(r)u + in\frac{V\tau}{r} = \kappa\left(D^* D - \frac{n^2}{r^2}\right)\tau, \tag{23}$$

where

$$D = \frac{d}{dr} \quad \text{and} \quad D^* = D + \frac{1}{r}. \tag{24}$$

The operators D and D^* satisfy the commutation relation

$$DD^* = D^* D - \frac{1}{r^2} = \frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} - \frac{1}{r^2}; \tag{25}$$

this relation is useful in the subsequent reductions.

Equation (22) can be rearranged in the manner

$$\begin{aligned} &rD^*\left[\nu\left(DD^* - \frac{n^2}{r^2}\right) - in\Omega - i\sigma\right]v - in\left[\nu\left(DD^* - \frac{n^2}{r^2}\right) - in\Omega - i\sigma\right]u \\ &\qquad - 2\left[rD^*\left(A - in\frac{\nu}{r^2}\right)u + in\left(\frac{V}{r} - in\frac{\nu}{r^2}\right)v\right] = -in\alpha\frac{V^2}{r}\tau, \end{aligned} \tag{26}$$

where

$$\Omega = \frac{V}{r} = A + \frac{B}{r^2}, \tag{27}$$

denotes the angular velocity. We can now eliminate v from equation (26) by making use of equation (21). We find

$$\begin{aligned} &\left[rD^*\left\{\nu\left(DD^* - \frac{n^2}{r^2}\right) - in\Omega - i\sigma\right\}rD^* - n^2\left\{\nu\left(DD^* - \frac{n^2}{r^2}\right) - in\Omega - i\sigma\right\}\right. \\ &\qquad \left. - 2in\frac{B}{r}D^* - \frac{4n^2\nu}{r^2}\right]u = -(n^2\Omega^2\alpha)r\tau; \end{aligned} \tag{28}$$

while equation (23) can be written in the form

$$\left\{\kappa\left(D^* D - \frac{n^2}{r^2}\right) - in\Omega - i\sigma\right\}\tau = \beta(r)u. \tag{29}$$

Equations (28) and (29) must be solved together with the boundary conditions which state that at $r = R_1$ and $r = R_2$ the fluctuations in the velocity and in the temperature must vanish, *i.e.*,

$$u = v = 0 \quad \text{and} \quad \tau = 0 \quad \text{at} \quad r = R_1 \quad \text{and} \quad r = R_2. \tag{30}$$

3. The case Ω = constant; the equations governing marginal stability. In the remaining part of this paper the assumption will be made that the two cylinders are rotated at the same constant rate. As may be expected on general grounds and as is also apparent from equations (28) and (29) this assumption will lead to considerable simplification.

When

$$\Omega = A = \text{constant and } B = 0, \tag{31}$$

it is permissible to introduce a *phase angular velocity*

$$\omega = n\Omega + \sigma. \tag{32}$$

In terms of ω equations (28) and (29) take the simpler forms

$$\begin{aligned}\Bigg[rD^*\left\{\nu\left(DD^* - \frac{n^2}{r^2}\right) - i\omega\right\}rD^* \\ - n^2\left\{\nu\left(DD^* - \frac{n^2}{r^2}\right) - i\omega\right\} - \frac{4n^2\nu}{r^2}\Bigg]u = -(n^2\Omega^2\alpha)r\tau\end{aligned} \tag{33}$$

and

$$\left[\kappa\left(D^*D - \frac{n^2}{r^2}\right) - i\omega\right]\tau = \beta(r)u, \tag{34}$$

while the form of the solutions sought is (*cf.* equations (20))

$$u = u(r)e^{in(\theta-\Omega t)+i\omega t} \text{ and } \tau = \tau(r)e^{in(\theta-\Omega t)+i\omega t} \tag{35}$$

Accordingly, in a frame of reference rotating with the angular velocity Ω, what distinguishes stability from instability of a pattern of fluid motion with $2n$ vortices, is the real part, $\Re(\omega)$, of ω. Now, it is known that in several related problems (Meksyn [13]; also Pellew & Southwell [14]) the situation in marginal stability (*i.e.*, one on the verge of stability) is characterized by $\omega = 0$ (and not merely by $\Re(\omega) = 0$); in other words, the principle of the exchange of stabilities is valid. In this paper we shall assume that this principle is valid also for the problem on hand. In a later paper we shall return to a justification of the principle.

On the assumption then that the principle of the exchange of stabilities is valid, the equations governing marginal stability are

$$\left\{rD^*\left(DD^* - \frac{n^2}{r^2}\right)rD^* - n^2\left(DD^* - \frac{n^2}{r^2}\right) - \frac{4n^2}{r^2}\right\}u = -\frac{n^2\,\Omega^2\,\alpha}{\nu}\,r\tau \tag{36}$$

and

$$\kappa\left(D^*\,D - \frac{n^2}{r^2}\right)\tau = \beta(r)u. \tag{37}$$

On expanding the differential operator on the left-hand side of equation (36), we find that the equation is in fact equivalent to

$$\left(D^*\,D - \frac{n^2}{r^2}\right)^2(ru) = -\frac{n^2\,\Omega^2\,\alpha}{\nu}\,\tau. \tag{38}$$

From equations (37) and (38) it now follows that

$$\left(D^*\,D - \frac{n^2}{r^2}\right)^3(ru) = -\frac{n^2\,\Omega^2\,\alpha}{\kappa\nu}\,\beta(r)u, \tag{39}$$

where it may be noted that (*cf.* equation (25))

$$D^*\,D - \frac{n^2}{r^2} = \frac{d^2}{dr^2} + \frac{1}{r}\,\frac{d}{dr} - \frac{n^2}{r^2}\,. \tag{40}$$

According to equations (21) and (38) the boundary conditions with respect to which we must solve equation (39) are

$$u = 0, \qquad D(ru) = 0 \qquad \text{and} \qquad \left(D^*\,D - \frac{n^2}{r^2}\right)^2(ru) = 0 \tag{41}$$
$$\text{at} \qquad r = R_1 \qquad \text{and} \qquad r = R_2.$$

It is evident that solving equation (39) together with the six boundary conditions (41), three at each of the two boundaries, is equivalent to a characteristic value problem. In the following sections it will be shown how we can solve this problem for the two cases (8) and (9). It is the manner of solving this characteristic value problem that constitutes the essential mathematical content of this paper.

4. A variational procedure for solving the characteristic value problem in case $\beta(r) = \beta r$ (case I). In this case, letting

$$W = ru \tag{42}$$

and measuring r in units of the radius, R_2, of the outer cylinder, we may first observe that the characteristic value problem we have to solve is one of determining

$$S_n = \frac{\Omega^2\,\alpha\beta}{\kappa\nu}\,R_2^6, \tag{43}$$

such that the equation

$$\left(\frac{d^2}{dr^2}+\frac{1}{r}\frac{d}{dr}-\frac{n^2}{r^2}\right)^3 W = -n^2 S_n\,W, \tag{44}$$

may have a non-trivial solution which satisfies the boundary conditions

$$W = 0, \qquad DW = 0 \quad\text{and}\quad \left(\frac{d^2}{dr^2}+\frac{1}{r}\frac{d}{dr}-\frac{n^2}{r^2}\right)^2 W = 0$$
$$\text{at}\quad r = 1 \quad\text{and}\quad r = R_1/R_2 = \eta \text{ (say)}. \tag{45}$$

Reduced in this manner, the problem is seen to be very similar to the one encountered in the theory of the thermal instability of fluid spheres and spherical shells (*cf.* Chandrasekhar [4], [5] and [6]). And as in this latter theory the solution for the lowest characteristic number, S_n, can be effected by the application of a variational principle which we shall now proceed to formulate.

Letting

$$G = \left(\frac{d^2}{dr^2}+\frac{1}{r}\frac{d}{dr}-\frac{n^2}{r^2}\right) W = \frac{1}{r}\frac{d}{dr}\left(r\frac{dW}{dr}\right)-\frac{n^2\,W}{r^2},$$
$$F = \left(\frac{d^2}{dr^2}+\frac{1}{r}\frac{d}{dr}-\frac{n^2}{r^2}\right)^2 W = \frac{1}{r}\frac{d}{dr}\left(r\frac{dG}{dr}\right)-\frac{n^2\,G}{r^2}, \tag{46}$$

we can rewrite the differential equation governing W in the form

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dF}{dr}\right)-\frac{n^2\,F}{r^2} = -n^2 S_n\,W. \tag{47}$$

The boundary conditions (45) now require that

$$F = W = DW = 0 \quad\text{at}\quad r = 1 \quad\text{and}\quad r = \eta. \tag{48}$$

Multiply equation (47) by rF and integrate over the range of r. The left-hand side of the equation gives

$$\int_\eta^1 F\frac{d}{dr}\left(r\frac{dF}{dr}\right) r\,d- n^2\int_\eta^1 F^2\frac{dr}{r}. \tag{49}$$

By integrating by parts the first of the two integrals in (49) and remembering that F vanishes at both limits, we obtain

(50) $$-\int_\eta^1 \left\{ r\left(\frac{dF}{dr}\right)^2 + \frac{n^2F^2}{r} \right\} dr.$$

Turning next to the right-hand side of equation (47) we have (*cf.* equation (46))

(51) $$\int_\eta^1 rWF\,dr = \int_\eta^1 W\frac{d}{dr}\left(r\frac{dG}{dr}\right)dr - n^2\int_\eta^1 WG\frac{dr}{r}.$$

After two integrations by parts the foregoing becomes

(52) $$\int_\eta^1 rWF\,dr = \left(rW\frac{dG}{dr} - rG\frac{dW}{dr}\right)_\eta^1 + \int_\eta^1 rG\left\{\frac{1}{r}\frac{d}{dr}\left(r\frac{dW}{dr}\right) - \frac{n^2}{r^2}W\right\}dr.$$

The integrated parts vanish on account of the boundary conditions (*cf.* equation (48)) and we are left with

(53) $$\int_\eta^1 rWF\,dr = \int_\eta^1 rG^2\,dr.$$

The result of multiplying equation (47) by rF and integrating is, therefore,

(54) $$n^2S_n = \frac{\int_\eta^1 r\{(dF/dr)^2 + n^2F^2/r^2\}\,dr}{\int_\eta^1 rG^2\,dr}.$$

This formula expresses S_n as the ratio of two positive definite integrals.

If we now consider the effect on S_n of a variation δW in W compatible with the boundary conditions, we readily find that

(55) $$n^2\,\delta S_n = -\frac{2}{\int_\eta^1 rG^2\,dr}\int_\eta^1 r\left\{\frac{d^2F}{dr^2} + \frac{1}{r}\frac{dF}{dr} - \frac{n^2}{r^2}F + n^2S_nW\right\}\delta F\,dr.$$

Hence to the first order, $\delta S_n \equiv 0$ *for all small arbitrary variations* δF. Further, it follows from (55) that the true solution of the problem gives a minimal value for S_n. This last fact enables us to formulate the following variational procedure of solving equation (47) and satisfying the boundary conditions of the problem:

Assume for F an expression involving one or more parameters A_j which vanishes at $r = 1$ and $r = \eta$. With the chosen form of F determine W as a solution of the equation

(56) $$\left(\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} - \frac{n^2}{r^2}\right)^2 W = F,$$

which satisfies the boundary conditions

(57) $$W = 0 \quad \text{and} \quad \frac{dW}{dr} = 0 \quad \text{at} \quad r = 1 \quad \text{and} \quad r = \eta.$$

Then evaluate S_n according to formula (54) and minimize it with respect to the parameters A_j. In this way we shall obtain the "best" value of S_n for the chosen form of F.

5. The variational solution of the characteristic value problem for case I. One confining cylinder. We shall first consider the case $\eta = 0$ *i.e.*, when there is only one confining cylinder. In this case the continuity of the solutions at $r = 0$ and the form of the equation to be solved for W suggest that we assume for F the trial function

(58) $$F = \Sigma_j A_j J_n(\alpha_j r),$$

where J_n denotes the Bessel function of order n, the α_j's $(j = 1, 2, \cdots)$ are its zeros and the A_j's are the variational parameters. With this choice, F vanishes at $r = 1$ as required. It may be recalled here that the functions $J_n(\alpha_j r)$ (for a given n) satisfy the orthogonality relations

(59) $$\int_0^1 rJ_n(\alpha_j r)J_n(\alpha_k r)\, dr = \tfrac{1}{2}[J_n'(\alpha_j)]^2\, \delta_{jk},$$

where primes denote differentiation with respect to the argument of the Bessel function and δ_{jk} is the usual Kronecker symbol.

With F given by equation (58) the equation to be solved for W is

(60) $$\left(\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} - \frac{n^2}{r^2}\right)^2 W = \Sigma_j A_j J_n(\alpha_j r).$$

The general solution of this equation which has no singularity at the origin is

(61) $$W = \Sigma_j(A_j/\alpha_j^4)J_n(\alpha_j r) + Br^n + Cr^{n+2},$$

where B and C are constants of integration. The condition $W = 0$ at $r = 1$ requires $B = -C$ and we have

(62) $$W = \Sigma_j(A_j/\alpha_j^4)J_n(\alpha_j r) + B(r^n - r^{n+2}).$$

The constant B is determined by the remaining condition at $r = 1$, namely that here dW/dr must vanish. This leads to

(63) $$B = \tfrac{1}{2}\Sigma_j(A_j/\alpha_j^3)J_n'(\alpha_j).$$

Turning next to the evaluation of S_n according to formula (54), we find that for F and W given by equations (58) and (62)

$$\int_0^1 r\left\{\left(\frac{dF}{dr}\right)^2 + \frac{n^2 F^2}{r^2}\right\} dr$$

$$(64) \qquad \begin{aligned} &= -\int_0^1 rF\left\{\frac{d^2 F}{dr^2} + \frac{1}{r}\frac{dF}{dr} - \frac{n^2 F}{r^2}\right\} dr \\ &= \int_0^1 r\Sigma_j A_j J_n(\alpha_j r)\Sigma_k A_k \alpha_k^2 J_n(\alpha_k r)\, dr \\ &= \tfrac{1}{2}\Sigma_j A_j^2 \alpha_j^2 [J_n'(\alpha_j)]^2 \end{aligned}$$

and

$$\int_0^1 rG^2\, dr = \int_0^1 rWF\, dr$$

$$(65)^* \qquad \begin{aligned} &= \tfrac{1}{2}\Sigma_j(A_j^2/\alpha_j^4)[J_n'(\alpha_j)]^2 + B\Sigma_j A_j\int_0^1 (r^{n+1} - r^{n+3})J_n(\alpha_j r)\, dr \\ &= \tfrac{1}{2}\Sigma_j(A_j^2/\alpha_j^4)[J_n'(\alpha_j)]^2 + 2B\Sigma_j(A_j/\alpha_j^2)J_{n+2}(\alpha_j). \end{aligned}$$

The resulting expression for S_n is, therefore

$$(66) \qquad n^2 S_n = \frac{\Sigma_j A_j^2 \alpha_j^2 [J_n'(\alpha_j)]^2}{\Sigma_j(A_j^2/\alpha_j^4)[J_n'(\alpha_j)]^2 + 2\Sigma_j(A_j/\alpha_j^2)J_{n+2}(\alpha_j)\Sigma_k(A_k/\alpha_k^3)J_n'(\alpha_j)},$$

where we have substituted for B from equation (63). By minimizing this last expression with respect to the A_j's we shall obtain the "best" value for the lowest characteristic number S_n for the chosen form of F.

The simplest trial function for F of the chosen form is

$$(67) \qquad F = J_n(\alpha_1 r),$$

where α_1 is the first zero of J_n. For this choice of F there is no variational parameter with respect to which we have to minimize and equation (66) directly gives

$$(68) \qquad n^2 S_n = \frac{\alpha_1^7 J_n'(\alpha_j)}{\alpha_1 J_n'(\alpha_1) + 2J_{n+2}(\alpha_1)}.$$

Values of S_n obtained with the aid of this formula are listed in Table 1.

The values given by (68) can be improved by including a second term in F. Thus, with the assumption

$$(69) \qquad F = J_n(\alpha_1 r) + AJ_n(\alpha_2 r),$$

* The transformations used in going from (49) to (50) and similarly from (51) to (53) can be used in the reverse fashion since these depend (apart from definitions) only on the boundary conditions of the problem.

TABLE 1

The Lowest Characteristic Numbers S_n for Case I and One Confining Cylinder

n	FIRST APPROXIMATION	A	SECOND APPROXIMATION
1	6954	0.0811	6873
2	8416	0.0944	8297
3	1.235×10^4	0.1017	1.216×10^4
4	1.828×10^4	0.1059	1.800×10^4
5	2.648×10^4	0.1083	2.607×10^4
6	3.731×10^4	0.1097	3.675×10^4
7	5.123×10^4	0.1104	5.048×10^4
8	6.871×10^4	0.1107	6.773×10^4
9	9.027×10^4	0.1107	8.902×10^4
10	1.165×10^5	0.1105	1.149×10^5
11	1.478×10^5	0.1101	1.459×10^5
12	1.851×10^5	0.1097	1.828×10^5
13	2.287×10^5	0.1092	2.260×10^5
14	2.795×10^5	0.1086	2.762×10^5
15	3.381×10^5	0.1080	3.343×10^5

where α_2 denotes the second zero of J_n and A is a variational parameter, equation (66) gives

$$\begin{aligned} n^2 S_n = \alpha_1^7 \{[J'_n(\alpha_1)]^2 + A^2(\alpha_2/\alpha_1)^2[J'_n(\alpha_2)]^2\} \times [J'_n(\alpha_1)\{\alpha_1 J'_n(\alpha_1) + 2J_{n+2}(\alpha_1)\} \\ + 4A(\alpha_1/\alpha_2)^3 J_{n+2}(\alpha_1)J'_n(\alpha_2) + A^2(\alpha_1/\alpha_2)^5 J'_n(\alpha_2)\{\alpha_2 J'_n(\alpha_2) + 2J_{n+2}(\alpha_2)\}]^{-1}. \end{aligned} \tag{70}$$

The values of S_n obtained after minimizing the foregoing expression with respect to A are listed in Table 1 together with the values of A which give the minimum values.

From a comparison of the results obtained in the first and the second approximations it would appear that the second approximation gives values which are probably correct to one part in 10^4.

6. The variational solution of the characteristic value problem for case I. Two confining cylinders. Turning next to the solution of the characteristic value problem for case I when there are two confining cylinders, we shall assume as a trial function for F a linear combination of the Bessel functions, $J_n(\alpha r)$ and $Y_n(\alpha r)$, of the two kinds which vanishes at $r = 1$ and $r = \eta$. For this latter purpose we first define the cylinder function (of order ν)

$$\mathcal{C}_{n,\nu}(z) = Y_n(\alpha\eta)J_\nu(z) - J_n(\alpha\eta)Y_\nu(z), \tag{71}$$

where α is a constant which we shall leave unspecified for the present. Then,

$$\mathcal{C}_{n,n}(\alpha r) = Y_n(\alpha\eta)J_n(\alpha r) - J_n(\alpha\eta)Y_n(\alpha r), \tag{72}$$

clearly vanishes for $r = \eta$; it will also vanish for $r = 1$ provided

$$Y_n(\alpha\eta)J_n(\alpha) - J_n(\alpha\eta)Y_n(\alpha) = 0. \tag{73}$$

It is known (*cf.* Gray & Mathews [15] p. 82, theorem X) that equation (73) admits an infinite number of roots all of which are real and simple; and that if $\alpha_j (j = 1, 2, \cdots)$ are the distinct roots of the equation, the functions $\mathcal{C}_{n,n}(\alpha_j r)$ $(j = 1, 2, \cdots)$ form an orthogonal set with the integral property

$$\int_\eta^1 r\, \mathcal{C}_{n,n}(\alpha_j r)\, \mathcal{C}_{n,n}(\alpha_k r)\, dr = N_{j,n}\, \delta_{jk}, \tag{74}$$

where

$$N_{j,n} = \frac{2}{\pi^2 \alpha_j^2}\left\{\frac{J_n^2(\alpha_j \eta)}{J_n^2(\alpha_j)} - 1\right\}. \tag{75}$$

For later use we may note here that the derivatives of $\mathcal{C}_{n,n}(\alpha_j r)$ at $r = 1$ and $r = \eta$ (which we shall denote by $\mathcal{C}'_n(\alpha_j)$ and $\mathcal{C}'_n(\alpha_j \eta)$, respectively) are given by

$$\mathcal{C}'_n(\alpha_j) = \left[\frac{d}{dr}\mathcal{C}_{n,n}(\alpha_j r)\right]_{r=1} = -\frac{2}{\pi}\frac{J_n(\alpha_j \eta)}{J_n(\alpha_j)} \tag{76}$$

and

$$\mathcal{C}'_n(\alpha_j \eta) = \left[\frac{d}{dr}\mathcal{C}_{n,n}(\alpha_j r)\right]_{r=\eta} = -\frac{2}{\eta\pi}. \tag{77}$$

Also, since $\mathcal{C}_{n,\nu}(z)$ is a cylinder function of ν it satisfies (with respect to ν) the same recurrence relations as the Bessel functions J_ν and Y_ν.

Returning to the variational solution of the characteristic value problem we assume for F the trial function

$$F(r) = \Sigma_j A_j\, \mathcal{C}_{n,n}(\alpha_j r), \tag{78}$$

where the A_j's are the variational parameters. With this choice, F vanishes at $r = 1$ and $r = \eta$. With F given by (78) the equation governing W (equation (56)) can be explicitly solved and for $n > 1$ we have (*cf.* equation (61))

$$W = \Sigma_j (A_j/\alpha_j^4)\mathcal{C}_{n,n}(\alpha_j r) + B_1 r^n + B_2 r^{n+2} + B_3 r^{-n} + B_4 r^{-n+2}, \tag{79}$$

where B_1, B_2, B_3 and B_4 are constants of integration to be determined by the boundary conditions (57). (It may be explicitly noted here that the solution given by (79) does not apply for $n = 1$; we shall return to this case presently.) The boundary conditions (57) lead to the equations

$$B_1 + B_2 + B_3 + B_4 = 0, \tag{80}$$

$$B_1 \eta^n + B_2 \eta^{n+2} + B_3 \eta^{-n} + B_4 \eta^{-n+2} = 0, \tag{81}$$

$$nB_1 + (n + 2)B_2 - nB_3 - (n - 2)B_4 = -\Sigma_j (A_j/\alpha_j^4)\mathcal{C}'_n(\alpha_j). \tag{82}$$

$$\begin{aligned} nB_1 \eta^{n-1} + (n + 2)B_2 \eta^{n+1} - nB_3 \eta^{-n-1} - (n - 2)B_4 \eta^{-n+1} \\ = -\Sigma_j (A_j/\alpha_j^4)\mathcal{C}'_n(\alpha_j \eta), \end{aligned} \tag{83}$$

where $\mathcal{C}'_n(\alpha_j)$ and $\mathcal{C}'_n(\alpha_j \eta)$ are defined as in equations (76) and (77). On solving the foregoing equations we find that the solution can be expressed in the form

$$B_i = \Sigma_j (A_j/\alpha_j^4) b_{ij} \qquad (i = 1, 2, 3 \text{ and } 4), \tag{84}$$

where the constants b_{1j}, b_{2j}, b_{3j} and b_{4j} depend only on α_j. Thus

$$b_{2j} = K_n \{(n - 1)(1 - \eta^2)\Delta_n(\alpha_j) + (1 - \eta^{-2n+2})\delta_n(\alpha_j)\} \tag{85}$$

and

$$b_{4j} = K_n \{(1 - \eta^{2n+2})\Delta_n(\alpha_j) - (n + 1)(1 - \eta^2)\delta_n(\alpha_j)\}, \tag{86}$$

where

$$\Delta_n(\alpha_j) = -\tfrac{1}{2}\{\mathcal{C}'_n(\alpha_j) - \eta^{-n+1}\mathcal{C}'_n(\alpha_j \eta)\}, \tag{87}$$

$$\delta_n(\alpha_j) = -\tfrac{1}{2}\{\mathcal{C}'_n(\alpha_j) - \eta^{n+1}\mathcal{C}'_n(\alpha_j \eta)\} \tag{88}$$

and

$$K_n = [(n^2 - 1)(1 - \eta^2)^2 + (1 - \eta^{2n+2})(1 - \eta^{-2n+2})]^{-1}. \tag{89}$$

With b_{2j} and b_{4j} given by equations (85) and (86), b_{1j} and b_{3j} follow from the equations:

$$b_{1j} = -\frac{1}{n}\{\tfrac{1}{2}\mathcal{C}'_n(\alpha_j) + (n + 1)b_{2j} + b_{4j}\} \tag{90}$$

and

$$b_{3j} = +\frac{1}{n}\{\tfrac{1}{2}\mathcal{C}'_n(\alpha_j) + b_{2j} - (n - 1)b_{4j}\}. \tag{91}$$

Turning next to the evaluation of S_n we find that for F and W given by equations (78) and (79) the integrals occurring in (54) can be reduced by steps similar to those indicated in (64) and (65). Thus, we now find

$$\int_\eta^1 r\left\{\left(\frac{dF}{dr}\right)^2 + \frac{n^2 F^2}{r^2}\right\} dr = \Sigma_j A_j^2 \alpha_j^2 N_{j,n} \tag{92}$$

and

$$(93)\qquad \int_\eta^1 rG^2\,dr = \Sigma_j(A_j^2/\alpha_j^4)N_{j,n} + \Sigma_j A_j \int_\eta^1 \mathcal{C}_{n,n}(\alpha_j r)[B_1 r^{n+1} + B_2 r^{n+3} + B_3 r^{-n+1} + B_4 r^{-n+3}]\,dr.$$

The remaining integrals on the right-hand side of (93) can be evaluated if proper use is made of the various recurrence relations satisfied by the cylinder functions $\mathcal{C}_{n,\nu}$ and also of equations (80) and (81); and we find after some lengthy reductions that

$$(94)\qquad \int_\eta^1 rG^2\,dr = \Sigma_j(A_j^2/\alpha_j^4)N_{j,n} - 2B_2\,\Sigma_j(A_j/\alpha_j^2)\mathfrak{M}_{n,n+2}^{(j)} - 2B_4\,\Sigma_j(A_j/\alpha_j^2)\mathfrak{N}_{n,n-2}^{(j)},$$

where

$$(95)\qquad \mathfrak{M}_{n,\nu}^{(j)} = \mathcal{C}_{n,\nu}(\alpha_j) - \eta^{\nu}\mathcal{C}_{n,\nu}(\alpha_j\,\eta)$$

and

$$(96)\qquad \mathfrak{N}_{n,\nu}^{(j)} = \mathcal{C}_{n,\nu}(\alpha_j) - \eta^{-\nu}\,\mathcal{C}_{n,\nu}(\alpha_j\,\eta).$$

The resulting expression for S_n is, therefore,

$$(97)\qquad n^2 S_n = \frac{\Sigma_j\,A_j^2\,\alpha_j^2\,N_{j,n}}{\Sigma_j\{(A_j^2/\alpha_j^4)N_{j,n} - 2B_2(A_j/\alpha_j^2)\mathfrak{M}_{n,n+2}^{(j)} - 2B_4(A_j/\alpha_j^2)\mathfrak{N}_{n,n-2}^{(j)}\}}$$

where B_2 and B_4 are given by equations (84) to (89).

As we have already stated the foregoing solution does not apply for the case $n = 1$. The reason why we must distinguish this case is that the complementary function in the solution (79) is no longer the general one as the terms in B_1 and B_4 become identical when $n = 1$. However, it is readily found that in the case $n = 1$ the general solution for W is

$$(98)\qquad W = \Sigma_j(A_j/\alpha_j^4)\mathcal{C}_{1,1}(\alpha_j r) + B_1 r + B_2 r^3 + B_3 r^{-1} + B_4\,r\log r,$$

and the equations determining the constants of integration are:

$$(99)\qquad \begin{aligned} B_1 + B_2 + B_3 &= 0,\\ B_1\,\eta + B_2\,\eta^3 + B_3\,\eta^{-1} + B_4\,\eta\log\eta &= 0,\\ B_1 + 3B_2 - B_3 + B_4 &= -\Sigma_j(A_j/\alpha_j^4)\mathcal{C}_1'(\alpha_j),\\ B_1 + 3B_2\,\eta^2 - B_3\,\eta^{-2} + B_4(1 + \log\eta) &= -\Sigma_j(A_j/\alpha_j^4)\mathcal{C}_1'(\alpha\,\eta). \end{aligned}$$

On solving these equations we find that the solutions for B_1, B_2, B_3 and B_4 can be expressed in the same way as before (*i.e.* by equations of the form (84)); but equations (85), (86), (89), (90) and (91) are now replaced by

$$b_{2j} = K_1\{\tfrac{1}{2}(1 - \eta^2)\Delta_1(\alpha_j) + (\log \eta)\delta_1(\alpha_j)\},$$

$$b_{4j} = K_1\{-(1 - \eta^4)\Delta_1(\alpha_j) + 2(1 - \eta^2)\delta_1(\alpha_j)\},$$

$$K_1 = [(1 - \eta^2)^2 + (1 - \eta^4)\log \eta]^{-1}, \tag{100}$$

$$b_{1j} = -[\tfrac{1}{2}\mathfrak{C}_1'(\alpha_j) + 2b_{2j} + \tfrac{1}{2}b_{4j}]$$

and

$$b_{3j} = +[\tfrac{1}{2}\mathfrak{C}_1'(\alpha_j) + b_{2j} + \tfrac{1}{2}b_{4j}].$$

The reduction of the formula for S_1 proceeds somewhat differently but essentially along the same lines as for $n > 1$. We finally obtain

$$S_1 = \frac{\Sigma_j A_j^2 \alpha_j^2 N_{j,1}}{\Sigma_j\{(A_j/\alpha_j^4)N_{j,1} - 2B_2(A_j/\alpha_j^2)\mathfrak{M}_{1,3}^{(j)} + 2B_4(A_j/\alpha_j^3)\mathfrak{M}_{1,0}^{(j)}\}}. \tag{101}$$

(In equations (100) and (101) the general definitions of $\Delta_n(\alpha_j)$, $\delta_n(\alpha_j)$, $\mathfrak{M}_{n,\nu}^{(j)}$ and $\mathfrak{N}_{n,\nu}^{(j)}$ are retained without modifications.)

The simplest trial function for F of the chosen form is

$$F = \mathfrak{C}_{n,n}(\alpha_1 r), \tag{102}$$

where α_1 is the first zero of the expression on the left-hand side of equation (73). For this choice of F there is no variational parameter with respect to which we have to minimize and equations (97) and (101) directly give (after substituting for B_2 and B_4):

$$\begin{aligned} S_n = \frac{\alpha_1^8}{n^2} N_{1,n}\{\alpha_1^2 N_{1,n} \\ - 2K_n[(n-1)(1-\eta^2)\Delta_n(\alpha_1) + (1-\eta^{-2n+2})\delta_n(\alpha_1)]\mathfrak{M}_{n,n+2}^{(1)} \\ - 2K_n[(1-\eta^{2n+2})\Delta_n(\alpha_1) - (n+1)(1-\eta^2)\delta_n(\alpha_1)]\mathfrak{N}_{n,n-2}^{(1)}\}^{-1} \\ (n > 1) \end{aligned} \tag{103}$$

and

$$\begin{aligned} S_1 = \alpha_1^8 N_{1,1}\{\alpha_1^2 N_{1,1} - 2K_1[\tfrac{1}{2}(1-\eta^2)\Delta_1(\alpha_1) + (\log\eta)\delta_1(\alpha_1)]\mathfrak{M}_{1,3}^{(1)} \\ + 2K_1[-(1-\eta^4)\Delta_1(\alpha_1) + 2(1-\eta^2)\delta_1(\alpha_1)]\mathfrak{M}_{1,0}^{(1)}/\alpha_1\}^{-1} \quad (n = 1). \end{aligned} \tag{104}$$

In Table 2 are listed the values of S_n obtained with the aid of the foregoing formulae and the recent tabulation of Chandrasekhar & Donna Elbert [16]

TABLE 2

The Lowest Characteristic Numbers S_n for Case I and Two Confining Cylinders: η denotes the Ratio of the Radius of the Inner to that of the Outer Cylinder

n	$\eta = 0.2$	$\eta = 0.3$	$\eta = 0.4$	$\eta = 0.5$	$\eta = 0.6$	$\eta = 0.8$
1	2.399×10^4	4.945×10^4	1.184×10^5	3.418×10^5	1.275×10^6	7.974×10^7
2	1.225×10^4	2.006×10^4	4.057×10^4	1.038×10^5	3.553×10^5	2.038×10^7
3	1.343×10^4	1.736×10^4	2.856×10^4	6.204×10^4	1.886×10^5	9.395×10^6
4	1.855×10^4	2.052×10^4	2.786×10^4	5.065×10^4	1.332×10^5	5.557×10^6
5	2.649×10^4	2.725×10^4	3.223×10^4	4.960×10^4	1.118×10^5	3.789×10^6
6		3.765×10^4	4.069×10^4	5.436×10^4	1.053×10^5	2.837×10^6
7					1.075×10^5	2.274×10^6
8					1.160×10^5	1.918×10^6
9						1.686×10^6
10						1.532×10^6
11						1.445×10^6
12						1.368×10^6

of the roots of equation (73). The dependence of S_n on n is further illustrated in Figure 1. From Table 2 and Figure 1 it is apparent how the pattern of convection which first appears at marginal stability shifts progressively to systems with a large number of vortices as η approaches unity.

The values given by (103) and (104) could be improved by including a second term in F. However, in view of the fact that for $\eta = 0$ the first approximation already gives values accurate to a fraction of a per cent (*cf.* Table 1) it was not considered necessary to carry out this improvement.

7. The solution of the characteristic value problem for case II. One confining cylinder. In case II the confining cylinders are maintained at constant, but different, temperatures. From a laboratory standpoint this is therefore the more important of the two cases considered.

In case II (*cf.* equation (9))

$$\beta(r) = \beta/r \qquad (\beta = \text{a constant}), \tag{105}$$

and equation (39) becomes

$$\left(D^* D - \frac{n^2}{r^2}\right)^3 W = -\frac{n^2 \Omega^2 \alpha\beta}{\kappa\nu} \frac{W}{r^2}, \tag{106}$$

where, as before, $W = ru$. Again, if we measure r in units of the radius R_2 of the outer cylinder the problem reduces to one of determining

$$S_n = \frac{\Omega^2 \alpha\beta}{\kappa\nu} R_2^4, \tag{107}$$

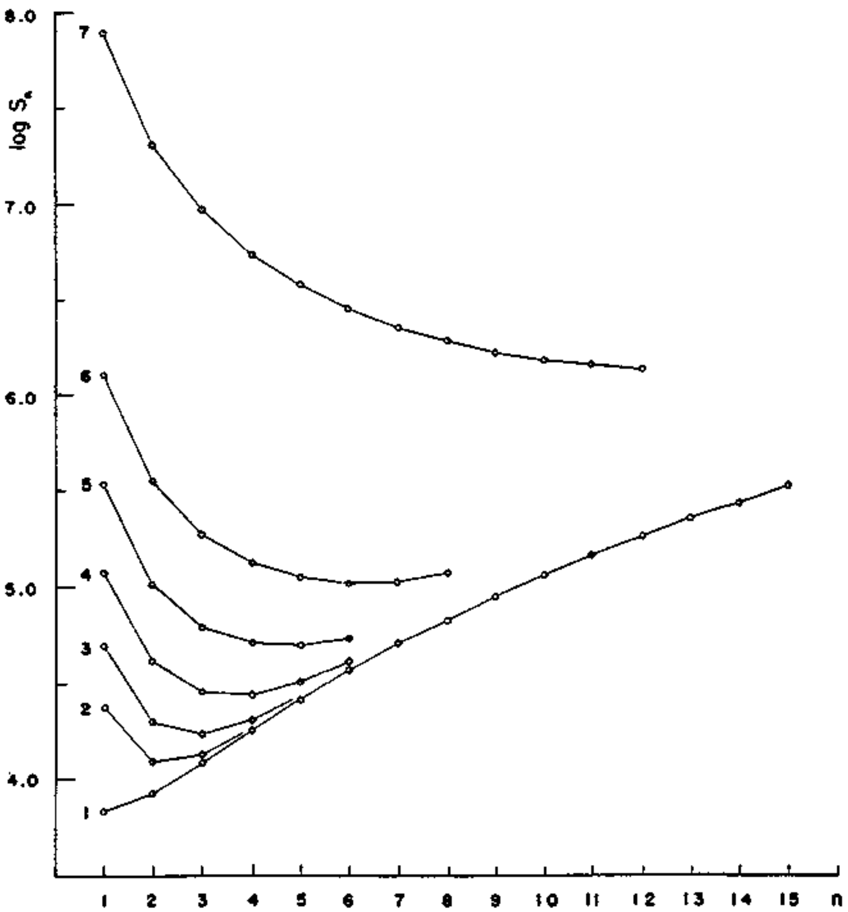


FIGURE 1. The criterion (S_n) for the onset of instability with a convection pattern with $2n$ vortices in the viscous flow between two rotating cylinders in the presence of a radial temperature gradient, for various values of the ratio (η) of the radii of the two cylinders and for case I. The curves labelled 1, 2, $\cdots$, 7 refer to $\eta = 0, 0.2, 0.3, 0.4, 0.5, 0.6$ and 0.8, respectively.

such that the differential equation

$$\left(\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} - \frac{n^2}{r^2}\right)^3 W = -n^2 S_n \frac{W}{r^2}, \tag{108}$$

allows a non-trivial solution which satisfies the boundary conditions (45).

Now, it can be readily verified that the occurrence of $1/r^2$ on the right-hand side of equation (108) prevents the formulation of a variational procedure for solving the underlying characteristic value problem. We shall therefore describe a different method of solution which, it will appear, is quite satisfactory for the purposes of determining the lowest characteristic number S_n. We shall illustrate the method by considering first the special case when there is only one confining cylinder. Strictly, the consideration of this special case under the conditions postulated is improper: For, the admission of a temperature gradient which varies as $1/r$ implies a logarithmic singularity in the temperature distribution at the origin if r should take the value zero. However, in spite of this singularity at the origin, the characteristic value problem itself has a meaning and the consideration of this case as a limit of permissible situations may be allowed.

As in §5 we shall let

$$\left(\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} - \frac{n^2}{r^2}\right)^2 W = F, \tag{109}$$

and rewrite equation (108) in the form

$$\left(\frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} - \frac{n^2}{r^2}\right) F = -n^2 S_n \frac{W}{r^2}. \tag{110}$$

We shall now assume that in the interval (0, 1), F can be expanded as a series in the form

$$F(r) = \Sigma_j A_j J_n(\alpha_j r), \tag{111}$$

where (*cf.* equation (59))

$$A_j = 2[J_n'(\alpha_j)]^{-2} \int_0^1 rF(r) J_n(\alpha_j r)\, dr. \tag{112}$$

With F given by (111) the solution of equation (109) which is continuous at the origin and satisfies the boundary conditions $W = dW/dr = 0$ at $r = 1$, is the same as that given in §5. We have (*cf.* equations (62) and (63))

$$W = \Sigma_j (A_j/\alpha_j^4) J_n(\alpha_j r) + B(r^n - r^{n+2}), \tag{113}$$

where

$$B = \tfrac{1}{2}\Sigma_j (A_j/\alpha_j^3) J_n'(\alpha_j). \tag{114}$$

Substituting (111) and (113) in equation (110), we are left with

$$\Sigma_j A_j \alpha_j^2 J_n(\alpha_j r) = \frac{n^2 S_n}{r^2}\{\Sigma_j (A_j/\alpha_j^4) J_n(\alpha_j r) + B(r^n - r^{n+2})\}. \tag{115}$$

Now multiplying equation (115) by $rJ_n(\alpha_k r)$ and integrating over the range r, we obtain

$$\tfrac{1}{2} A_k \alpha_k^2 [J_n'(\alpha_k)]^2 = n^2 S_n \int_0^1 \frac{1}{r}\left\{\Sigma_j \frac{A_j}{\alpha_j^4} J_n(\alpha_j r) + B(r^n - r^{n+2})\right\} J_n(\alpha_k r)\, dr. \tag{116}$$

Letting

$$\left(j\left|\frac{1}{r^2}\right|k\right) = \int_0^1 J_n(\alpha_j r)\frac{1}{r} J_n(\alpha_k r)\,dr \tag{117}$$

and

$$D_k = \int_0^1 (r^{n-1} - r^{n+1})J_n(\alpha_k r)\,dr, \tag{118}$$

we can rewrite equation (116) in the form

$$\tfrac{1}{2}A_k\,\alpha_k^2[J'_n(\alpha_k)]^2\,Q = \Sigma_j(A_j/\alpha_j^4)\,(j\,|\,r^{-2}\,|\,k) + BD_k, \tag{119}$$

where

$$Q = \frac{1}{n^2}S_n. \tag{120}$$

Finally, substituting for B from equation (114) we obtain the following infinite system of homogeneous equations for the A_j's:

$$\Sigma_j\frac{A_j}{\alpha_j^4}\left\{\left(j\left|\frac{1}{r^2}\right|k\right) + \tfrac{1}{2}\alpha_j\,J'_n(\alpha_j)D_k - \tfrac{1}{2}\alpha_j^6\,[J'_n(\alpha_j)]^2\,Q\delta_{jk}\right\} = 0. \tag{121}$$

Hence Q is to be determined as a root of the infinite determinantal equation

$$\left\|\left(j\left|\frac{1}{r^2}\right|k\right) + \tfrac{1}{2}\alpha_j J'_n(\alpha_j)D_k - \tfrac{1}{2}\alpha_j^6\,[J'_n(\alpha_j)]^2\,\delta Q_{jk}\right\| = 0. \tag{122}$$

By using the known properties of the Bessel functions, we can show that

$$\begin{aligned} D_k = \frac{1}{\alpha_k^n}\{2^{n-1}(n-1)! &- 2(n-1)\alpha_k^{n-2}J_{n-2}(\alpha_k) \\ &- 2^2(n-1)(n-2)\alpha_k^{n-3}J_{n-3}(\alpha_k) \\ - 2^3(n-1)(n-2)(n-3)&\alpha_k^{n-4}J_{n-4}(\alpha_k) - \cdots - 2^{n-1}(n-1)!J_0(\alpha_k)\}. \end{aligned} \tag{123}$$

Also the diagonal elements of the matrix $(j|\,r^{-2}\,|k)$ can be explicitly evaluated. We find (*cf.* Watson [17] p. 137) that

$$\left(j\left|\frac{1}{r^2}\right|j\right) = \frac{1}{2n}\left\{1 - 2\sum_{m=1}^{n-1} J_m^2(\alpha_j) - J_0^2(\alpha_j)\right\}. \tag{124}$$

But it does not appear that the non-diagonal elements can be similarly evaluated.

Now a method of solving equation (122) of infinite order would be to set the determinant formed by the first j rows and columns equal to zero and let j take increasingly larger values. In practise the success of this method will depend on how rapidly the lowest root of the equation of order j tends to its limit as $j \to \infty$. It will appear that for the problem on hand the process converges sufficiently rapidly.

On the method of solution outlined in the preceding paragraph the *first approximation* for the lowest characteristic number S_n will be given by setting the $(1,1)$ element of the matrix equal to zero. Thus

$$\tfrac{1}{2}\alpha_1^2\,[J_n'(\alpha_1)]^2\,Q = \frac{1}{\alpha_1^4}\left(1\,\middle|\,\frac{1}{r^2}\,\middle|\,1\right) + \frac{J_n'(\alpha_1)}{2\alpha_1^3}\,D_1, \tag{125}$$

where α_1 denotes the first zero of $J_n(z)$. Substituting for $(1|\,r^{-2}\,|1)$ and D_1 in accordance with equations (123) and (124), we obtain

$$\begin{aligned} n^2 S_n = \alpha_1^6\,[J_n'(\alpha_1)]^2 \Bigg[\frac{1}{n}\left\{1 - 2\sum_{m=1}^{n-1} J_m^2(\alpha_1) - J_0^2(\alpha_1)\right\} \\ + \frac{J_n'(\alpha_1)}{\alpha_1^{n-1}}\,\{2^{n-1}(n-1)! - 2(n-1)\alpha_k^{n-2}\,J_{n-2}(\alpha_1) \\ - 2^2(n-1)(n-2)\alpha_1^{n-3}\,J_{n-3}(\alpha_1) - \cdots - 2^{n-1}(n-1)!\,J_0(\alpha_1)\}\Bigg]^{-1}. \end{aligned} \tag{126}$$

The values of S_n obtained with the aid of the foregoing formula for $n = 1, 2, \cdots, 6$ are given in Table 3. For these same values of n a second approximation was also carried out by setting the determinant formed by the first two rows and columns of (122) equal to zero; and for $n = 1$ a third approximation was carried out as well. In carrying out these higher approximations the non-diagonal elements $(j|\,r^{-2}\,|k)(j \neq k)$ had to be evaluated numerically. This part of the work was enormously lightened by tables of the functions

$$J_n(\alpha_j)\text{ for } j = 1(1)5, \qquad n = 1(1)6 \text{ and } r = 0(0.01)1.00, \tag{127}$$

calculated for the purpose by the Electronic Computer Project of the Institute for Advanced Study at Princeton. (A fuller acknowledgement is made at the end of the paper.)

A comparison of the values of S_n given in Table 3 shows that on going from the first to the second approximation we reduce the value of S_n by a nearly

TABLE 3

The Lowest Characteristic Numbers S_n for Case II and One Confining Cylinder

n	FIRST APPROXIMATION	SECOND APPROXIMATION	THIRD APPROXIMATION
1	1180	1075	1077
2	2340	2128	
3	4390	4000	
4	7560	6910	
5	12180	11170	
6	18570	17100	

constant factor: in fact, the factor varies only between 1.100 and 1.086 in the tabulated range of n. Also, in the one case for which the result of a third approximation is available, the value of S_n was changed from that given by the second approximation by only one part in 500. From these facts we may conclude that already the second approximation provides the required characteristic numbers with errors not probably exceeding one per cent.

8. The solution of the characteristic value problem for case II. Two confining cylinders. Turning finally to the solution for case II when there are two confining cylinders, we shall assume that F can be expanded as a series in the form (*cf.* equations (74) and (75))

$$F = \Sigma_j A_j \mathcal{C}_{n,n}(\alpha_j r), \tag{128}$$

where

$$A_j = \frac{1}{N_{j,n}} \int_\eta^1 rF(r)\mathcal{C}_{n,n}(\alpha_j r)\, dr. \tag{129}$$

With F given by (128) the solution of equation (109) which satisfies the boundary conditions (57) is evidently the same as that given in §6. Combining equations (79) and (84) (and similarly (98) and (84)) we can write the required solution for W in the form

$$\begin{aligned} W = \Sigma_j(A_j/\alpha_j^4)\, \{\mathcal{C}_{n,n}(\alpha_j r) + b_{1j} r^n + b_{2j} r^{n+2} + b_{3j} r^{-n} \\ + b_{4j} r^{-n+2} \text{ (or } r \log r \text{ in case } n = 1)\}, \end{aligned} \tag{130}$$

where the coefficients b_{1j}, *etc.*, have the same meanings as in §6 (equations (85) to (91) and (100)).

Now substituting for F and W according to equations (128) and (130) in equation (109) we obtain

$$\begin{aligned} \Sigma_j A_j \alpha_j^2 \mathcal{C}_{n,n}(\alpha_j r) = \frac{n^2 S_n}{r^2} \Sigma_j(A_j/\alpha_j^4)\{\mathcal{C}_{n,n}(\alpha_j r) \\ + b_{1j} r^n + b_{2j} r^{n+2} + b_{3j} r^{-n} + b_{4j} r^{-n+2} \quad \text{(or } r \log r \text{ in case } n = 1)\}. \end{aligned} \tag{131}$$

Next multiplying this equation by $r\mathcal{C}_{n,n}(\alpha_k r)$ and integrating over the range of r we obtain (*cf.* equations (74) and (75))

$$\begin{aligned} A_k \alpha_k^2 N_{k,n} Q = \Sigma_j(A_j/\alpha_j^4) \\ \times \{(j|\, r^{-2}\, |k) + b_{1j}\langle r^{n-1}\rangle_k + b_{2j}\langle r^{n+1}\rangle_k + b_{3j}\langle r^{-n-1}\rangle_k \\ + b_{4j}\langle r^{-n+1}\rangle_k \text{ (or } \langle\log r\rangle_k \text{ in case } n = 1)\}, \end{aligned} \tag{132}$$

where $Q = 1/n^2 S_n$,

$$\left(j \left| \frac{1}{r^2} \right| k\right) = \int_\eta^1 \mathcal{C}_{n,n}(\alpha_j r)\, \frac{1}{r}\, \mathcal{C}_{n,n}(\alpha_k r)\, dr \tag{133}$$

and

$$\langle X\rangle_k = \int_\eta^1 X\mathfrak{C}_{n,n}(\alpha_k r)\, dr. \tag{134}$$

Equation (132) represents the following infinite system of homogeneous equations for the constants A_j:

$$\begin{aligned}\Sigma_j(A_j/\alpha_j^4)\,\{(j|\,r^{-2}\,|k) + b_{1j}\langle r^{n-1}\rangle_k + b_{2j}\langle r^{n+1}\rangle_k + b_{3j}\langle r^{-n-1}\rangle_k \\ + b_{4j}\langle r^{-n+1}\rangle_k \text{ (or } \langle\log r\rangle_k \text{ in case } n = 1) - Q\alpha_j^6 N_{j,n}\,\delta_{jk}\} = 0.\end{aligned} \tag{135}$$

Hence Q is to be determined as a root of the infinite determinantal equation

$$\begin{aligned}\|\,(j|\,r^{-2}\,|k) + b_{1j}\langle r^{n-1}\rangle_k + b_{2j}\langle r^{n+1}\rangle_k + b_{3j}\langle r^{-n-1}\rangle_k \\ + b_{4j}\langle r^{-n+1}\rangle_k \text{ (or } \langle\log r\rangle_k \text{ in case } n = 1) - Q\alpha_j^6 N_{j,n}\,\delta_{jk}\,\| = 0.\end{aligned} \tag{136}$$

By using the known properties of cylinder functions, we can show that

$$\begin{aligned}\langle r^{n-1}\rangle_k = -\frac{1}{\alpha_k^n}\,\{\alpha_k^{n-1}\,\mathfrak{M}_{n,n-1}{}^{(k)} + 2(n-1)\alpha_k^{n-2}\,\mathfrak{M}_{n,n-2}{}^{(k)} \\ + 2^2(n-1)(n-2)\alpha_k^{n-3}\,\mathfrak{M}_{n,n-3}{}^{(k)} + \cdots + 2^{n-1}(n-1)!\,\mathfrak{M}_{n,0}{}^{(k)}\},\end{aligned} \tag{137}$$

$$\langle r^{n+1}\rangle_k = \mathfrak{M}_{n,n+1}{}^{(k)}/\alpha_k, \tag{138}$$

$$\langle r^{-n+1}\rangle_k = -\mathfrak{N}_{n,n-1}{}^{(k)}/\alpha_k \tag{139}$$

and

$$\begin{aligned}\left(k\left|\frac{1}{r^2}\right|k\right) = -\frac{1}{2n}\Big\{[\mathfrak{C}_{n,0}{}^2(\alpha_k) - \mathfrak{C}_{n,0}{}^2(\alpha_k\eta)] \\ + 2\sum_{m=1}^{n-1}[\mathfrak{C}_{n,m}{}^2(\alpha_k) - \mathfrak{C}_{n,m}{}^2(\alpha_k\eta)]\Big\}.\end{aligned} \tag{140}$$

But it does not appear that

$$\begin{aligned}\langle r^{-n-1}\rangle_k &= \int_\eta^1 \frac{dr}{r^{n+1}}\,\mathfrak{C}_{n,n}(\alpha_k r), \\ \langle\log r\rangle_k &= \int_\eta^1 \log r\mathfrak{C}_{1,1}(\alpha_k r)\,dr \qquad (n = 1),\end{aligned} \tag{141}$$

or the non-diagonal elements of the matrix (133) can be expressed similarly in terms of known quantities; in these cases numerical integration would seem unavoidable.

As in §7 a first approximation to the value of the lowest characteristic number S_n can be obtained by setting the (1, 1) element of the matrix on the left-hand

TABLE 4

The Lowest Characteristic Numbers S_n for Case II and Two Confining Cylinders: η denotes the Ratio of the Radius of the Inner to the Outer Cylinder

n	First Approximation				Second Approximation
	$\eta = 0.2$	$\eta = 0.3$	$\eta = 0.4$	$\eta = 0.5$	$\eta = 0.5$
1	7478			1.855×10^4	1.897×10^4
2	4109	7936		5.652×10^4	5.813×10^4
3	5013	7188	1.356×10^4	3.399×10^4	3.505×10^4
4	7750	9025	1.299×10^4	2.799×10^4	2.893×10^4
5			1.620×10^4	2.772×10^4	2.868×10^4
6				3.081×10^4	3.171×10^4

side of equation (136) equal to zero. The values of S_n obtained in such a first approximation are given in Table 4 for various values of n and η.

We already know (§7) that for $\eta = 0$, the error of the first approximation is to give values of S_n which are too *high* by ten per cent. To estimate the errors in the present calculations, a second approximation was carried out for $\eta = 0.5$. The values obtained in the second approximation (by setting the determinant formed by the first two rows and columns of the secular matrix equal to zero) are also listed in Table 4. And a comparison of the values of $S_n(n = 1, 2, \cdots, 6)$ obtained in the first and in the second approximations indicates that now the error of the first approximation is to give values of S_n which are too *low* by three per cent. From these facts we may conclude that no entry in Table 4 is likely to be in error by more than ten per cent; indeed, for $\eta = 0.3$, 0.4 and 0.5 the errors may be very much less.

The results of the calculations of this as well as of the preceding section are illustrated in Figure 2. The remarkable similarity of these results with those obtained for case I (compare particularly Figures 1 and 2) is especially noteworthy.

9. An illustrative example. As an example illustrative of the theory developed in this paper we shall consider the case of water (at room temperatures) enclosed between two long cylinders maintained at constant different temperatures and of radii 5.0 cm and 2.5 cm, respectively. Let the entire system be rotated at a constant angular velocity of 1 radian per second. Then according to the results of Table 4, the lowest mode of instability is one with ten vortices ($n = 5$) and occurs when

$$S = \frac{\Omega^2 \alpha\beta}{\kappa\nu} R_2^4 = 2.87 \times 10^4. \tag{142}$$

(Actually, the values of S at which convection patterns with eight or twelve vortices can occur are not substantially higher.) Inserting the numerical values

$$\Omega = 1 \text{ sec}^{-1}; \qquad R_2 = 5 \text{ cm}; \qquad \alpha = 2.1 \times 10^{-4} \text{ degree}^{-1},$$

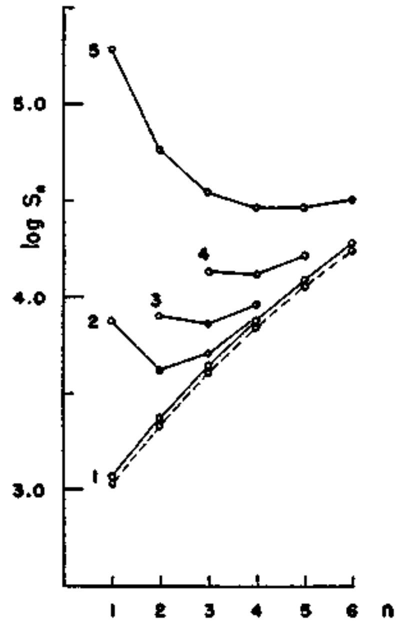


FIGURE 2. Same as Figure 1 but for case II. The curves labelled 1, 2, 3, 4 and 5 refer to $\eta = 0, 0.2, 0.3, 0.4$ and 0.5, respectively; of the two curves labelled 1 the dotted one refers to the results of the second approximation while the other to the results of the first approximation. For $\eta = 0.5$ (curve 5) the results of the second approximation are plotted; the remaining curves refer to the results of the first approximation.

(143)

$$\kappa = 1.4 \times 10^{-3}\ \text{cm}^2/\text{sec} \quad \text{and} \quad \nu = 1.0 \times 10^{-2}\ \text{cm}^2/\text{sec},$$

we find from (142) that the critical value of β at which instability will set in is

(144) $$\beta_{\text{critical}} = 3.1\ \text{degree/cm}.$$

Under the circumstances considered (namely that of case II)

(145) $$\frac{dT}{dr} = \frac{\beta}{r};$$

accordingly, β is related to the difference in temperature of the two cylinders by

(146) $$\Delta T = T_2 - T_1 = \beta \log (R_2/R_1) = -\beta \log \eta.$$

Hence (144) implies a difference in temperature of

(147) $$\Delta T = 3.1 \log_e 2 = 2.1°.$$

In other words, for water enclosed between two cylinders of radii 5.0 and 2.5 cm and rotated at the rate 1 radian/sec, instability of the type considered will arise when the difference in temperature between the two cylinders increases above 2.1°. On account of the occurrence of R_2^4 in the definition of S, this difference in temperature predicted for $R_2 = 5.0$ cm and $R_2 = 2.5$ cm will be reduced by a factor 16 if the scale of the apparatus is increased by a factor 2 and the rate of rotation is unaltered *i.e.*, $\Delta T = 0.13°$ under these latter circumstances. Now the experiments of Fultz and Hide seem to have been performed with cylinders of radii 10 cm or more and rates of rotation much larger than 1 radian/sec; moreover, the patterns of motion they have studied are for differences of temperature generally in excess of one degree. Consequently, in their experiments factors not included in this theory must have been operative. Nevertheless, it would be of interest to repeat their experiments with cylinders of radii 5 cm or less (or with smaller rates of rotation) to see if the effects predicted here can be observed.

10. Acknowledgements. As has already been acknowledged the calculations for case II when there is only one confining cylinder (§7) were greatly facilitated by the Electronic Computer Project of the Institute for Advanced Study at Princeton having provided tables of the functions

$$J_n(\alpha_j r) \text{ for } n = 1(1)6, \qquad j = 1(1)5 \text{ and } r = 0(0.01)1.00. \tag{148}$$

Similarly the calculations presented in §8 for two confining cylinders were facilitated by the Project having provided tables of the functions

$$\mathcal{C}_{n,n}(\alpha_j r) \text{ for } \eta = 0.2(0.1)0.6 \text{ and } 0.8, \qquad n = 1(1)6 \text{ and } j = 1 \tag{149}$$

and, also

$$\mathcal{C}_{n,n}(\alpha_j r) \text{ for } \eta = 0.5, \qquad n = 1(1)6 \text{ and } j = 2 \text{ and } 3. \tag{150}$$

All these functions have been tabulated at intervals of 0.01 to eight significant figures of which at least seven are believed to be reliable.

It must be evident that without the tables provided by the Electronic Computer Project of the Institute for Advanced Study the task of evaluating numerically the various integrals needed in the calculation of S_n (for case II) would have been too formidable to have undertaken. It is therefore a particular pleasure to acknowledge here my indebtedness to Professor J. von Neumann and Dr. Herman H. Goldstine of the Institute for Advanced Study for their most generous co-operation. My thanks are also due to Miss Donna Elbert who gave valuable assistance in all the remaining calculations.

References

[1] Fultz, D., "Experimental analogies to atmospheric motion", *Compendium of Meteorology*, Amer. Met. Soc. (1951) 1235–1248.

[2] Hide, R., (unpublished) Reported in Geophysical Discussion of the Royal Astronomical Society, *Observatory* **72** (1952) 231–232.

[3] Chandrasekhar, S., "On the inhibition of convection by a magnetic field", *Phil. Mag.* (7) **43** (1952) 501–532.

[4] ———, "The thermal instability of a fluid sphere heated within", *Phil. Mag.* (7) **43** (1952) 1317–1329.

[5] ———, "The onset of convection by thermal instability in spherical shells", *Phil. Mag.* (7) **44** (1953) 233–241.

[6] ———, "The onset of convection by thermal instability in spherical shells (a correction)", *Phil. Mag.* (7) **44** (1953) 1129–1130.

[7] ———, "The stability of viscous flow between rotating cylinders in the presence of a magnetic field", *Proc. Roy. Soc.* A, **216** (1953) 293–309.

[8] ———, "The instability of a layer of fluid heated below and subject to Coriolis forces", *Proc. Roy. Soc.* A, **217** (1953) 306–327.

[9] ———, "The gravitational instability of an infinite homogeneous medium when Coriolis force is acting and a magnetic field is present", *Ap. J.* (in press).

[10] Chandrasekhar, S. & Fermi, E., "Problems of gravitational stability in the presence of a magnetic field", *Ap. J.* 118 (1953) 116–141.

[11] Lord Rayleigh, "On convection currents in a horizontal layer of fluid, when the higher temperature is on the under side", *Scientific Papers*, Vol. 6, Cambridge University Press, London (1920) 432–446.

[12] Jeffreys, H., "The stability of a layer of fluid heated below", *Phil. Mag.* (7) **2** (1926) 833–844.

[13] Meksyn, D., "Stability of viscous flow between rotating cylinders. I", *Proc. Roy. Soc.* A, **187** (1946) 115–128.

[14] Pellew, A. & Southwell, R. V., "On maintained convective motion in a fluid heated from below", *Proc. Roy. Soc.* A, **176** (1940) 312–343.

[15] Gray, A. & Mathews, G. B., *A treatise on Bessel functions*, 2nd ed. Macmillan, London (1922).

[16] Chandrasekhar, S. & Elbert, D., "The roots of $Y_n(\lambda\eta)J_n(\lambda) - J_n(\lambda\eta)Y_n(\lambda) = 0$", *Camb. Phil. Soc.* (in press).

[17] Watson, G. N., *A treatise on the theory of Bessel functions*, 2nd ed., Cambridge (1944) 137.

Reprinted without change of pagination from the
Proceedings of the Royal Society, A, *volume* 241, pp. 9–28, 1957

The hydrodynamic stability of helium II between rotating cylinders. I

BY S. CHANDRASEKHAR, F.R.S.

Enrico Fermi Institute for Nuclear Studies, University of Chicago

AND R. J. DONNELLY

Institute for the Study of Metals, University of Chicago

(*Received* 6 *March* 1957)

The hydrodynamic instability of helium II between rotating cylinders is investigated on two assumptions regarding the mutual friction force, **F**, between the normal and the superfluid components of the liquid. On both assumptions **F** is proportional to the constant vorticity which prevails in the stationary state and to the difference in the velocities between the two fluids; however, on one assumption the effect of **F** is confined entirely to the transverse plane, while on the other it is allowed to be isotropic (with respect to the difference in the velocities). The hydrodynamic problem is solved for the case when the two cylinders (of radii R_1 and R_2) are rotated in the same direction and $(R_2-R_1) \ll \frac{1}{2}(R_2+R_1)$. It follows from the theory that when $\partial(r^2\Omega)/\partial r < 0$ (where Ω denotes the angular velocity and r the distance from the axis) the flow becomes (eventually) unstable along two branches: the first of these is the normal (Taylor) instability of a viscous fluid inhibited by its coupling with an inviscid fluid, and the second is the (Rayleigh) instability of the superfluid inhibited, in turn, by its coupling with a viscous fluid. Further, in all cases the critical Taylor number of instability (suitably defined) becomes asymptotic to a relation which is equivalent to $\Gamma^2 = \frac{1}{2}(R_2^2-R_1^2)/R_1^2$, where Γ is the coupling constant. From an experiment of Kolm & Herlin's (1956), to which the present theory appears applicable, a value of $\Gamma = 0{\cdot}52$ is deduced; this is in very good accord with the value $\Gamma = 0{\cdot}55$ which Hall & Vinen (1956*a*) have deduced from an unrelated experiment.

1. INTRODUCTION

On the two-fluid model of helium II, hydrodynamical equations governing the behaviour of the normal and the super-fluid components of the liquid have been written down; these equations are in accord with many experiments performed at low velocities. On the other hand, the nature of the mutual friction between the two fluids which one invokes to explain certain other effects observed at higher velocities is not entirely clear. However, Hall & Vinen (1956*a, b*) have recently deduced a particular form for the mutual friction from Feynman's (1955) theory of liquid helium; and on its basis they are able to account for their observations on the absorption of second sound in rotating helium II. It is the object of this and later papers to show how certain aspects of the mutual friction can be inferred from a study of the hydrodynamic stability of helium II between rotating cylinders along the lines of Sir Geoffrey Taylor's (1923) classic investigation.

2. THE PROPOSED FORMS FOR THE MUTUAL FRICTION

Consider liquid helium below the λ point. Let ρ_n and ρ_s denote the densities, and **u** and **v** the velocities, of the normal and the super-fluid components, respectively.

The normal fluid is characterized by a certain kinematic viscosity† ν $(=\mu/\rho_n)$, while the superfluid is inviscid. Let the mutual friction between the two fluids be given by the force

$$\frac{\rho_n \rho_s}{\rho} \mathbf{F}(\mathbf{u}, \mathbf{v}) \tag{1}$$

per unit volume, where ρ $(=\rho_n+\rho_s)$ is the total density.

The equations governing the two fluids may then be written in the forms (cf. Daunt & Smith 1954)

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}\,.\operatorname{grad})\,\mathbf{u} = -\frac{1}{\rho}\operatorname{grad} p - \frac{\rho_s}{\rho_n} S \operatorname{grad} T + \nu(\tfrac{4}{3}\operatorname{grad}\operatorname{div}\mathbf{u} - \operatorname{curl}\operatorname{curl}\mathbf{u}) - \frac{\rho_s}{\rho}\mathbf{F} \tag{2}$$

and

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\,.\operatorname{grad})\,\mathbf{v} = -\frac{1}{\rho}\operatorname{grad} p + S \operatorname{grad} T + \frac{\rho_n}{\rho}\mathbf{F}, \tag{3}$$

where T denotes the absolute temperature and S the entropy per gram of the total fluid. In addition to equations (2) and (3) we have the equations expressing the conservation of mass and entropy:

$$\left.\begin{aligned} &\frac{\partial \rho_n}{\partial t} + \operatorname{div}(\rho_n \mathbf{u}) = 0, \quad \frac{\partial \rho_s}{\partial t} + \operatorname{div}(\rho_s \mathbf{v}) = 0 \\ \text{and} \quad &\frac{\partial}{\partial t}(\rho S) + \operatorname{div}(\rho S \mathbf{u}) = 0. \end{aligned}\right\} \tag{4}‡$$

In hydrodynamical experiments at constant temperatures and sufficiently low velocities ($u, v \ll$ velocity of sound) we may treat both components of the fluid as incompressible, in which case the equations take the simplified forms

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}\,.\operatorname{grad})\,\mathbf{u} = -\operatorname{grad} \varpi_n + \nu\nabla^2\mathbf{u} - \alpha_s \mathbf{F}, \tag{5}$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\,.\operatorname{grad})\,\mathbf{v} = -\operatorname{grad} \varpi_s + \alpha_n \mathbf{F} \tag{6}$$

and

$$\operatorname{div}\mathbf{u} = \operatorname{div}\mathbf{v} = 0, \tag{7}$$

where we have introduced the abbreviations

$$\varpi_n = \frac{p}{\rho} + \frac{\rho_s}{\rho_n} ST, \quad \varpi_s = \frac{p}{\rho} - ST, \tag{8}$$

$$\alpha_s = \rho_s/\rho \quad \text{and} \quad \alpha_n = \rho_n/\rho \quad (\alpha_s + \alpha_n = 1). \tag{9}$$

To complete the hydrodynamical equations we need an assumption concerning **F**.

In selecting a form for **F**, the following general considerations are perhaps relevant.

† It is to be noted that we are defining the kinematic viscosity with respect to the density of the normal fluid. It is customary to define it with respect to the total density ρ; but the definition we have adopted is more convenient in the context.

‡ In writing these equations, we are ignoring a possible 'source-term' representing the conversion of the normal fluid into the superfluid and vice versa; also, the equation expressing the conservation of entropy is strictly true only if the terms dependent on viscosity in equation (2) are neglected.

In a stationary state, it is natural to suppose that

$$\mathbf{u} = \mathbf{v} \quad \text{and} \quad \mathbf{F} = 0; \tag{10}$$

and for slight departures from a stationary state, we may expect that $\mathbf{F}$ is some linear function of $\mathbf{u}-\mathbf{v}$; thus

$$F_i = \Gamma_{ij}(u_j - v_j), \tag{11}$$

where the tensor Γ_{ij} depends on the state of common motion which prevails in the stationary state. It is not clear how Γ_{ij} might depend, generally, on the initial state of motion. However, for the problem of the stability of viscous flow between rotating cylinders, the selection of a proper form for Γ_{ij} is less arbitrary. For, it is apparent, that in this case, equations (5) to (7) admit for $\mathbf{u}$ and $\mathbf{v}$ the same stationary solution as does an ordinary incompressible fluid, i.e. the motion is entirely transverse and the velocity in this direction is given by (see equations (25) and (26) below):

$$V = Ar + B/r, \tag{12}$$

where r is the radial distance from the axis of rotation and A and B are two constants which are determined by the angular velocities and the radii of the rotating cylinders. This state of motion is characterized by a constant vorticity; we have

$$\boldsymbol{\omega} = \operatorname{curl} \mathbf{u} = 2A\mathbf{1}_z, \tag{13}$$

where $\mathbf{1}_z$ is a unit vector along the axis of rotation. It is, therefore, natural to suppose that Γ_{ij} depends on the vorticity in some simple way. We might, for example, consider two assumptions for Γ_{ij}, both of which are closely related to the form for the mutual friction deduced by Hall & Vinen (1956*a*): thus

$$\mathbf{F} = \frac{\Gamma}{|\boldsymbol{\omega}|}[\boldsymbol{\omega} \times (\mathbf{u}-\mathbf{v})] \times \boldsymbol{\omega}, \tag{14}$$

or

$$\mathbf{F} = \Gamma\, |\boldsymbol{\omega}|\, (\mathbf{u}-\mathbf{v}), \tag{15}$$

where Γ is a constant of proportionality depending only on the temperature. The first of these assumptions makes the mutual friction anisotropic and confines its effect entirely to the transverse plane. This appears to be in better accord with the experiments than the second assumption, which makes the mutual friction isotropic.

On the basis of a detailed calculation based on Feynman's model of quantized vortex lines, Hall & Vinen suggest the somewhat more general form

$$\mathbf{F} = \frac{\Gamma}{|\boldsymbol{\omega}|}[\boldsymbol{\omega} \times (\mathbf{u}-\mathbf{v})] \times \boldsymbol{\omega} + \Lambda\boldsymbol{\omega} \times (\mathbf{u}-\mathbf{v}), \tag{16}$$

where Λ is a further constant. In this paper, we shall restrict ourselves to the two forms (14) and (15); we shall consider Hall & Vinen's more general form for $\mathbf{F}$ in a later paper.

3. The perturbation equations for the case of transverse mutual friction

We shall consider first the case when the mutual friction has the form (14).

For discussing the stability of viscous flow between rotating cylinders, it is convenient to have the fundamental equations written in cylindrical polar coordinates. Denoting by (u_r, u_θ, u_z) and (v_r, v_θ, v_z) the components of the normal and the superfluids in the radial (r), the transverse (θ), and the z directions, respectively, we find that equations (5) and (6), for $\mathbf{F}$ given by equation (14), take the forms

$$\frac{\partial u_r}{\partial t}+(\mathbf{u}\,.\,\mathrm{grad})\,u_r-\frac{u_\theta^2}{r}=-\frac{\partial \varpi_n}{\partial r}+\nu\left(\nabla^2 u_r-\frac{2}{r^2}\frac{\partial u_\theta}{\partial \theta}-\frac{u_r}{r^2}\right)-\Gamma\,|\,\boldsymbol{\omega}\,|\,\alpha_s(u_r-v_r), \quad (17)$$

$$\frac{\partial u_\theta}{\partial t}+(\mathbf{u}\,.\,\mathrm{grad})\,u_\theta+\frac{u_\theta u_r}{r}=-\frac{1}{r}\frac{\partial \varpi_n}{\partial \theta}+\nu\left(\nabla^2 u_\theta+\frac{2}{r^2}\frac{\partial u_r}{\partial \theta}-\frac{u_\theta}{r^2}\right)-\Gamma\,|\,\boldsymbol{\omega}\,|\,\alpha_s(u_\theta-v_\theta), \quad (18)$$

$$\frac{\partial u_z}{\partial t}+(\mathbf{u}\,.\,\mathrm{grad})\,u_z=-\frac{\partial \varpi_n}{\partial z}+\nu\nabla^2 u_z, \quad (19)$$

$$\frac{\partial v_r}{\partial t}+(\mathbf{v}\,.\,\mathrm{grad})\,v_r-\frac{v_\theta^2}{r}=-\frac{\partial \varpi_s}{\partial r}+\Gamma\,|\,\boldsymbol{\omega}\,|\,\alpha_n(u_r-v_r), \quad (20)$$

$$\frac{\partial v_\theta}{\partial t}+(\mathbf{v}\,.\,\mathrm{grad})\,v_\theta+\frac{v_\theta v_r}{r}=-\frac{1}{r}\frac{\partial \varpi_s}{\partial \theta}+\Gamma\,|\,\boldsymbol{\omega}\,|\,\alpha_n(u_\theta-v_\theta) \quad (21)$$

and

$$\frac{\partial v_z}{\partial t}+(\mathbf{v}\,.\,\mathrm{grad})\,v_z=-\frac{\partial \varpi_s}{\partial z}, \quad (22)$$

where we have used the abbreviations

$$\left.\begin{aligned}(\mathbf{u}\,.\,\mathrm{grad})&=u_r\frac{\partial}{\partial r}+\frac{u_\theta}{r}\frac{\partial}{\partial \theta}+u_z\frac{\partial}{\partial z},\\ (\mathbf{v}\,.\,\mathrm{grad})&=v_r\frac{\partial}{\partial r}+\frac{v_\theta}{r}\frac{\partial}{\partial \theta}+v_z\frac{\partial}{\partial z};\end{aligned}\right\} \quad (23)$$

also ∇^2 now stands for the conventional Laplacian in cylindrical polar co-ordinates. Further, the equations of continuity in these co-ordinates are

$$\left.\begin{aligned}&\frac{\partial u_r}{\partial r}+\frac{u_r}{r}+\frac{1}{r}\frac{\partial u_\theta}{\partial \theta}+\frac{\partial u_z}{\partial z}=0\\ \text{and}\qquad &\frac{\partial v_r}{\partial r}+\frac{v_r}{r}+\frac{1}{r}\frac{\partial v_\theta}{\partial \theta}+\frac{\partial v_z}{\partial z}=0.\end{aligned}\right\} \quad (24)$$

It is evident (as we have already stated in § 2) that the foregoing equations allow the stationary solution

$$u_r=v_r=u_z=v_z=0, \quad \frac{\partial \varpi_n}{\partial r}=\frac{\partial \varpi_s}{\partial r}=\frac{V^2}{r}, \quad (25)$$

where

$$u_\theta=v_\theta=V=Ar+B/r. \quad (26)$$

In equation (26) A and B are two constants related to the angular velocities of rotation Ω_1 and Ω_2 of the inner and outer cylinders (of radii R_1 and R_2, $R_2>R_1$) by

$$A=\Omega_1\frac{1-\mu R_2^2/R_1^2}{1-R_2^2/R_1^2} \quad \text{and} \quad B=\Omega_1\frac{R_1^2(1-\mu)}{1-R_1^2/R_2^2}, \quad (27)$$

where $\mu=\Omega_2/\Omega_1$.

We shall now suppose that the stationary solution represented by equations (25) and (26) is slightly perturbed. Let the components of the velocity in the perturbed state be

$$u_r, \quad V+u_\theta, \quad u_z \quad \text{and} \quad v_r, \quad V+v_\theta, \quad v_z; \tag{28}$$

also, let ϖ_n and ϖ_s suffer the increments $\delta\varpi_n$ and $\delta\varpi_s$.

As in Taylor's original investigation, we shall suppose that the perturbation is symmetrical about the z axis, so that u_r, v_r, etc., are all functions of r, z and t only. On this assumption, the linearized forms of equations (17), (18) and (19) are

$$\frac{\partial}{\partial r}\delta\varpi_n = -\frac{\partial u_r}{\partial t}+\nu\left(\nabla^2 u_r-\frac{u_r}{r^2}\right)+2\Omega u_\theta-K\alpha_s(u_r-v_r), \tag{29}$$

$$0 = -\frac{\partial u_\theta}{\partial t}+\nu\left(\nabla^2 u_\theta-\frac{u_\theta}{r^2}\right)-2Au_r-K\alpha_s(u_\theta-v_\theta) \tag{30}$$

and

$$\frac{\partial}{\partial z}\delta\varpi_n = -\frac{\partial u_z}{\partial t}+\nu\nabla^2 u_z, \tag{31}$$

where

$$K = 2\Gamma\,|A|, \tag{32}$$

$$\Omega = A+Br^{-2} \tag{33}$$

is a function of r, and

$$\nabla^2 = \frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{\partial^2}{\partial z^2}. \tag{34}$$

The corresponding equations for $\mathbf{v}$ can be obtained by suppressing in equations (29) to (31) the terms in ν, and also interchanging $\mathbf{u}$ and $\mathbf{v}$ and the subscripts s and n.

Finally, we have the equation of continuity

$$\frac{\partial u_r}{\partial r}+\frac{u_r}{r}+\frac{\partial u_z}{\partial z} = 0, \tag{35}$$

and a similar equation for $\mathbf{v}$.

We shall seek solutions of equations (29) to (31) and (35) which are of the forms

$$\left.\begin{aligned} u_r(r,z,t) &= \mathrm{e}^{pt}u_r(r)\cos\lambda z, & v_r(r,z,t) &= \mathrm{e}^{pt}v_r(r)\cos\lambda z;\\ u_\theta(r,z,t) &= \mathrm{e}^{pt}u_\theta(r)\cos\lambda z, & v_\theta(r,z,t) &= \mathrm{e}^{pt}v_\theta(r)\cos\lambda z;\\ u_z(r,z,t) &= \mathrm{e}^{pt}u_z(r)\sin\lambda z, & v_z(r,z,t) &= \mathrm{e}^{pt}v_z(r)\sin\lambda z.\end{aligned}\right\} \tag{36}$$

Equations (29) to (31) and (35) then become

$$\frac{\partial}{\partial r}\delta\varpi_n = \left[\nu\left(\mathrm{DD}^*-\lambda^2-\frac{p}{\nu}\right)u_r+2\Omega u_\theta-K\alpha_s(u_r-v_r)\right]\cos\lambda z, \tag{37}$$

$$0 = \nu\left(\mathrm{DD}^*-\lambda^2-\frac{p}{\nu}\right)u_\theta-2Au_r-K\alpha_s(u_\theta-v_\theta), \tag{38}$$

$$\frac{\partial}{\partial z}\delta\varpi_n = \left[\nu\left(\mathrm{D}^*\mathrm{D}-\lambda^2-\frac{p}{\nu}\right)u_z\right]\sin\lambda z, \tag{39}$$

$$\mathrm{D}^*u_r = -\lambda u_z \quad \text{and} \quad \mathrm{D}^*v_r = -\lambda v_z, \tag{40}$$

where

$$\mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}r} \quad \text{and} \quad \mathrm{D}^* = \mathrm{D}+\frac{1}{r} = \frac{\mathrm{d}}{\mathrm{d}r}+\frac{1}{r}. \tag{41}$$

The operators D and D* satisfy the commutation relation

$$\mathrm{DD}^* = \mathrm{D}^*\mathrm{D} - \frac{1}{r^2} = \frac{\mathrm{d}^2}{\mathrm{d}r^2} + \frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r} - \frac{1}{r^2}. \tag{42}$$

Eliminating $\delta\varpi_n$ from equations (37) and (39) and making use of equation (40), we obtain

$$-\lambda\left[\nu\left(\mathrm{DD}^* - \lambda^2 - \frac{p}{\nu}\right)u_r + 2\Omega u_\theta - K\alpha_s(u_r - v_r)\right]$$
$$= \nu\mathrm{D}\left(\mathrm{D}^*\mathrm{D} - \lambda^2 - \frac{p}{\nu}\right)u_z = -\frac{\nu}{\lambda}\left(\mathrm{DD}^* - \lambda^2 - \frac{p}{\nu}\right)\mathrm{DD}^*u_r. \tag{43}$$

Rearranging this last equation, we have

$$\frac{\nu}{\lambda^2}(\mathrm{DD}^* - \lambda^2)\left(\mathrm{DD}^* - \lambda^2 - \frac{p}{\nu}\right)u_r = 2\Omega u_\theta - K\alpha_s(u_r - v_r). \tag{44}$$

We also have (cf. equation (38))

$$\nu\left(\mathrm{DD}^* - \lambda^2 - \frac{p}{\nu}\right)u_\theta = 2Au_r + K\alpha_s(u_\theta - v_\theta). \tag{45}$$

The corresponding equations governing v_r and v_θ are, clearly,

$$-\frac{p}{\lambda^2}(\mathrm{DD}^* - \lambda^2)v_r = 2\Omega v_\theta + K\alpha_n(u_r - v_r) \tag{46}$$

and

$$-pv_\theta = 2Av_r - K\alpha_n(u_\theta - v_\theta). \tag{47}$$

Solutions of equations (44) to (47) must be sought which satisfy the boundary conditions

$$u_r = u_\theta = u_z = \mathrm{D}^*u_r = v_r = 0 \quad \text{for} \quad r = R_1 \text{ and } R_2. \tag{48}$$

That u_r and v_r must vanish on the boundaries is clearly necessary; that u_θ and u_z must also vanish follows from the requirement that there is no slip between the normal fluid (which is viscous) and the walls of the cylinder.

4. The equations governing marginal stability; the case of transverse mutual friction (*continued*)

It can be shown (see the next paper (part II) Chandrasekhar 1957), that in the present problem, as in Taylor's original problem, the principle of the exchange of stabilities is valid. Assuming that this is the case, we can write down the equations which govern the state of marginal stability by putting $p = 0$ in equations (44) to (47). We thus obtain

$$\frac{\nu}{\lambda^2}(\mathrm{DD}^* - \lambda^2)^2 u_r = 2\Omega u_\theta - K\alpha_s(u_r - v_r), \tag{49}$$

$$\nu(\mathrm{DD}^* - \lambda^2)u_\theta = 2Au_r + K\alpha_s(u_\theta - v_\theta), \tag{50}$$

$$0 = 2\Omega v_\theta + K\alpha_n(u_r - v_r), \tag{51}$$

and

$$0 = 2Av_r - K\alpha_n(u_\theta - v_\theta). \tag{52}$$

Combining equations (49) and (51), we have

$$\frac{\nu}{\lambda^2}(\mathrm{DD}^*-\lambda^2)^2 u_r = 2\Omega\left(u_\theta+\frac{\alpha_s}{\alpha_n}v_\theta\right). \tag{53}$$

Rewriting equation (50) in the manner

$$[\nu(\mathrm{DD}^*-\lambda^2)-K\alpha_s]\,u_\theta-2Au_r = -K\alpha_s v_\theta, \tag{54}$$

and eliminating v_θ between equations (53) and (54), we obtain after some reductions (in which use is made of the fact that $\alpha_s+\alpha_n = 1$)

$$\left[4\Omega A-\frac{K\alpha_n\nu}{\lambda^2}(\mathrm{DD}^*-\lambda^2)^2\right]u_r = 2\Omega[\nu(\mathrm{DD}^*-\lambda^2)-K]\,u_\theta. \tag{55}$$

Next, combining equations (49) and (52), we have

$$\begin{aligned}\frac{\nu}{\lambda^2}\left[(\mathrm{DD}^*-\lambda^2)^2+\frac{K\alpha_s\lambda^2}{\nu}\right]u_r &= 2\Omega u_\theta+K\alpha_s v_r\\ &= 2\Omega u_\theta+\frac{K^2\alpha_s\alpha_n}{2A}(u_\theta-v_\theta);\end{aligned} \tag{56}$$

and substituting for $(u_\theta-v_\theta)$ from equation (50) we obtain after some further reductions

$$\left[(\mathrm{DD}^*-\lambda^2)^2+\frac{K\lambda^2}{\nu}\right]u_r = \left[\frac{2\Omega\lambda^2}{\nu}+\frac{K\alpha_n\lambda^2}{2A}(\mathrm{DD}^*-\lambda^2)\right]u_\theta \tag{57}$$

or

$$\frac{2A\nu}{\lambda^2}\left[(\mathrm{DD}^*-\lambda^2)^2+\frac{K\lambda^2}{\nu}\right]u_r = [4\Omega A+K\alpha_n\nu(\mathrm{DD}^*-\lambda^2)]\,u_\theta. \tag{58}$$

This equation and (cf. equation (55))

$$\left[4\Omega A-\frac{K\alpha_n\nu}{\lambda^2}(\mathrm{DD}^*-\lambda^2)^2\right]u_r = 2\Omega[\nu(\mathrm{DD}^*-\lambda^2)-K]\,u_\theta \tag{59}$$

govern the state of marginal stability.

Considering next the boundary conditions (48), we observe that, according to equation (51), the vanishing of u_r and v_r on $r = R_1$ and R_2 requires that the same be true of v_θ; and from the vanishing of u_r, u_θ and v_θ on the boundaries we conclude from equation (50) that DD^*u_θ also vanishes. Hence the boundary conditions with respect to which equations (58) and (59) must be solved are

$$u_r = \mathrm{D}^*u_r = u_\theta = \mathrm{DD}^*u_\theta = 0 \quad \text{for} \quad r = R_1 \text{ and } R_2. \tag{60}$$

5. The case when the cylinders are rotating in the same direction and $(R_2-R_1)\ll\frac{1}{2}(R_2+R_1)$; the case of transverse mutual friction (*continued*)

It is known from investigations on the original Taylor problem (cf. Chandrasekhar 1954) that, when the difference in the radii of the two cylinders is small compared with their mean radius, we need not distinguish between D and D*. Also, if the two cylinders are rotating in the same direction, we may further replace $\Omega(r)$ in equations (58) and (59) by the mean value $\bar{\Omega}$. In rewriting equations (58) and (59) in the framework of these approximations, we shall find it convenient to measure radial distances from the inner cylinder in units of $d = R_2 - R_1$. Thus, letting

$$\zeta = \frac{1}{d}(r - R_1) \quad \text{and} \quad \lambda = \frac{a}{d}, \tag{61}$$

we find that equations (58) and (59) can be brought to the forms

$$[(\mathrm{D}^2 - a^2)^2 + Ca^2]\, u_r = \frac{2\bar{\Omega} d^2}{\nu} a^2 \left[1 - \frac{C\alpha_n}{T}(\mathrm{D}^2 - a^2)\right] u_\theta \tag{62}$$

and
$$(\mathrm{D}^2 - a^2 - C)\, u_\theta = \frac{2Ad^2}{\nu}\left[1 + \frac{C\alpha_n}{Ta^2}(\mathrm{D}^2 - a^2)^2\right] u_r, \tag{63}$$

where
$$T = -\frac{4\bar{\Omega} A}{\nu^2} d^4 \tag{64}$$

is the Taylor number and
$$C = \frac{Kd^2}{\nu} = 2\Gamma \frac{|A|\, d^2}{\nu}. \tag{65}$$

Eliminating u_θ between equations (62) and (63) we obtain

$$(\mathrm{D}^2 - a^2 - C)\,[(\mathrm{D}^2 - a^2)^2 + Ca^2]\, u_r = -Ta^2 \left[1 - \frac{C\alpha_n}{T}(\mathrm{D}^2 - a^2)\right]\left[1 + \frac{C\alpha_n}{Ta^2}(\mathrm{D}^2 - a^2)^2\right] u_r; \tag{66}$$

u_θ satisfies an identical equation.

For assigned values of a^2, C and α_n, solutions of equations (62) and (63) must be sought which satisfy the boundary conditions (cf. equation (60))

$$u_r = \mathrm{D}u_r = u_\theta = \mathrm{D}^2 u_\theta = 0 \quad \text{for} \quad \zeta = 0 \text{ and } 1. \tag{67}$$

This constitutes a characteristic value problem for T; and we are particularly interested in determining the range of Taylor numbers (for assigned C and α_n) for which the flow is stable.

The foregoing characteristic value problem for T can be solved by a method which has proved successful in solving Taylor's original problem very simply (Chandrasekhar 1954). The problem is solved by this method in part II. It will appear that the regions of stability in the (a, T) plane have some unexpected features. Insight into the origin of these can be obtained if the essential features of the solution can be deduced in an elementary way. This we can do for a slightly altered characteristic value problem which admits (in contrast to the one specified by the correct boundary conditions (67)) an explicit solution. The alteration consists in replacing the boundary conditions (67) by

$$u_r = \mathrm{D}^2 u_r = u_\theta = \mathrm{D}^2 u_\theta = 0 \quad \text{for} \quad \zeta = 0 \text{ and } 1. \tag{68}$$

Experience with similar characteristic value problems assures us that the nature of the dependence of the required characteristic values on the various parameters of the problem is the same independently of the boundary conditions with respect to which the problem may have been solved. (We shall verify that this is also the case for the problem on hand.) For this reason, we shall restrict ourselves in this paper to the solution of the problem for the boundary conditions (68), though numerical results derived from a solution satisfying the correct boundary conditions will be quoted.

(a) The solution of the characteristic value problem for the boundary conditions (68)

It follows from equations (62) and (63) that, when the boundary conditions (68) are satisfied, u_r and u_θ together with all their even derivatives vanish for $\zeta = 0$ and 1; thus

$$\mathrm{D}^{(2n)}u_r = \mathrm{D}^{(2n)}u_\theta = 0 \quad \text{for} \quad \zeta = 0 \text{ and } 1 \quad \text{and} \quad n = 0, 1, 2, \ldots. \tag{69}$$

Accordingly, the solutions for u_r and u_θ must be multiples of $\sin m\pi\zeta$, where m is an integer; this is self-evident, but a formal proof could be given if one is desired. The various fundamental modes of u_r are, therefore, given by

$$u_r = \sin m\pi\zeta \quad (m = 1, 2, \ldots). \tag{70}$$

The substitution of the solution (70) in equation (66) leads to the characteristic equation

$$(m^2\pi^2 + a^2 + C)\,[(m^2\pi^2 + a^2)^2 + Ca^2] = Ta^2\left[1 + \frac{C\alpha_n}{T}(m^2\pi^2 + a^2)\right]\left[1 + \frac{C\alpha_n}{Ta^2}(m^2\pi^2 + a^2)^2\right], \tag{71}$$

or, equivalently

$$[Ta^2 + C\alpha_n(m^2\pi^2 + a^2)^2]\,[T + C\alpha_n(m^2\pi^2 + a^2)] = T(m^2\pi^2 + a^2 + C)\,[(m^2\pi^2 + a^2)^2 + Ca^2]. \tag{72}$$

Expanding this equation, we obtain the following quadratic equation for T:

$$T^2a^2 - TQ_m + C^2\alpha_n^2(m^2\pi^2 + a^2)^3 = 0, \tag{73}$$

where
$$\begin{aligned} Q_m &= (m^2\pi^2 + a^2 + C)\,[(m^2\pi^2 + a^2)^2 + Ca^2] - C\alpha_n(m^2\pi^2 + a^2)\,(m^2\pi^2 + 2a^2) \\ &= (m^2\pi^2 + a^2)^3 + C\alpha_s(m^2\pi^2 + a^2)\,(m^2\pi^2 + 2a^2) + C^2a^2. \end{aligned} \tag{74}$$

Hence
$$T = \frac{1}{2a^2}\{Q_m \pm [Q_m^2 - 4C^2\alpha_n^2 a^2(m^2\pi^2 + a^2)^3]^{\frac{1}{2}}\}. \tag{75}$$

Thus for assigned C and α_n there are two solutions. The arrangement of these solutions in the (a, T) plane for different values of m is of crucial importance for stability considerations.

(b) The higher branch

Considering first the solution with the positive sign in equation (75), we have

$$T = \frac{1}{2a^2}\{Q_m + [Q_m^2 - 4C^2\alpha_n^2 a^2(m^2\pi^2 + a^2)^3]^{\frac{1}{2}}\}. \tag{76}$$

For given α_n and C this equation defines a one-parameter family of an ascending sequence of non-intersecting curves (see figure 1; the upper two curves). Also, each of these curves is characterized by a single minimum. The fundamental mode $m = 1$ provides the lowest of these curves. As far as this branch is concerned, therefore, the critical Taylor number for the onset of instability is given by the minimum of the lowest mode described by the equations

$$T = \frac{1}{2a^2}\{Q^2 + [Q - 4C^2\alpha_n^2 a^2(\pi^2+a^2)^3]^{\frac{1}{2}}\}, \tag{77}$$

where
$$Q = (\pi^2+a^2)^3 + C\alpha_s(\pi^2+a^2)(\pi^2+2a^2) + C^2a^2. \tag{78}$$

Critical Taylor numbers T_c, evaluated in this manner for various values of C and $\alpha_n = 0$, 0·5 and 0·83, are given in table 1 and illustrated in figure 2.

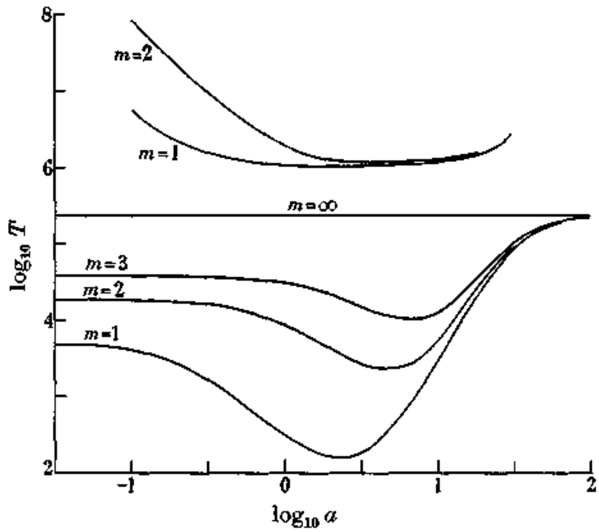


FIGURE 1. The dependence of the Taylor number (T) on the wave number a (in units of $1/d$) for the two branches and various modes for the case of transverse mutual friction. The curves refer to a value of $C = 1000$ and $\alpha_n = 0{\cdot}5$.

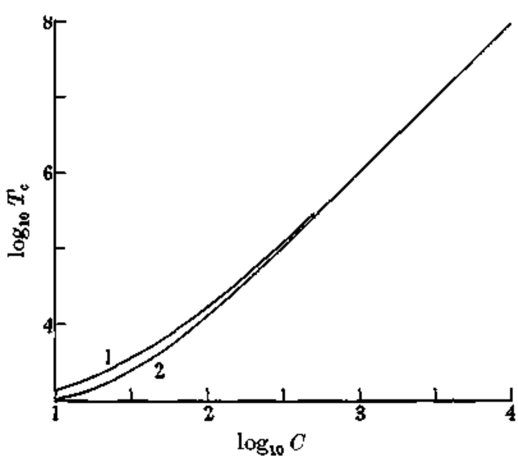


FIGURE 2. The log C-log T_c relation for the case of transverse mutual friction derived from the solution of the modified boundary value problem. The curves labelled 1 and 2 refer to values of $\alpha_n = 0$ and 0·5, respectively, and represent the instability arising from the higher branch.

TABLE 1. THE CRITICAL TAYLOR NUMBERS APPROPRIATE TO THE HIGHER BRANCH AND FOR THE TRANSVERSE FORM OF THE MUTUAL FRICTION FOR VARIOUS VALUES OF C AND α_n

	the solution for the modified boundary value problem				the solution for the correct boundary value problem					
	$\alpha_n = 0$		$\alpha_n = 0{\cdot}5$		$\alpha_n = 0$		$\alpha_n = 0{\cdot}5$		$\alpha_n = 0{\cdot}83$	
C	a	T_c	a	T_c		T_c	a	T_c	a	T_c
0	2·2	$6{\cdot}575 \times 10^2$	2·2	$6{\cdot}575 \times 10^2$	3·13	$1{\cdot}715 \times 10^3$†	3·13	$1{\cdot}715 \times 10^3$	—	—
10	2·4	$1{\cdot}343 \times 10^3$	2·3	$1{\cdot}036 \times 10^3$	3·39	$2{\cdot}865 \times 10^3$	3·28	$2{\cdot}329 \times 10^3$	—	—
20	2·5	$2{\cdot}223 \times 10^3$	2·4	$1{\cdot}602 \times 10^3$	3·53	$4{\cdot}193 \times 10^3$	3·39	$3{\cdot}116 \times 10^3$	—	—
40	2·5	$4{\cdot}578 \times 10^3$	2·4	$3{\cdot}343 \times 10^3$	3·67	$7{\cdot}433 \times 10^3$	3·54	$5{\cdot}275 \times 10^3$	—	—
60	2·6	$7{\cdot}731 \times 10^3$	2·5	$5{\cdot}899 \times 10^3$	3·75	$1{\cdot}147 \times 10^4$	3·64	$8{\cdot}248 \times 10^3$	—	—
100	2·6	$1{\cdot}643 \times 10^4$	2·6	$1{\cdot}343 \times 10^4$	3·8	$2{\cdot}199 \times 10^4$	3·8	$1{\cdot}665 \times 10^4$	—	—
200	2·6	$5{\cdot}219 \times 10^4$	2·6	$4{\cdot}629 \times 10^4$	3·9	$6{\cdot}245 \times 10^4$	3·9	$5{\cdot}182 \times 10^4$	3·8	$4{\cdot}422 \times 10^4$
500	2·6	$2{\cdot}794 \times 10^5$	2·6	$2{\cdot}649 \times 10^5$	4·1	$3{\cdot}053 \times 10^5$	4·0	$2{\cdot}782 \times 10^5$	4·0	$2{\cdot}598 \times 10^5$
1000	2·6	$1{\cdot}058 \times 10^6$	2·6	$1{\cdot}029 \times 10^6$	4·2	$1{\cdot}113 \times 10^6$	4·2	$1{\cdot}057 \times 10^6$	4·2	$1{\cdot}020 \times 10^6$
3000	2·6	$9{\cdot}173 \times 10^6$	2·6	$9{\cdot}087 \times 10^6$	4·4	$9{\cdot}365 \times 10^6$	4·4	$9{\cdot}183 \times 10^6$	—	—
10000	2·6	$1{\cdot}006 \times 10^8$	2·6	$1{\cdot}003 \times 10^8$	4·8	$1{\cdot}013 \times 10^8$	4·7	$1{\cdot}007 \times 10^8$	—	—

† It is known that the exact value of T_c for $C = 0$ is 1708; the difference between this and the entry 1715 is a measure of the inaccuracy introduced by the approximate method of solving the problem described in part II.

It is apparent from figures 1 and 2 and it can also be deduced from equations (74) and (76) that

$$T \to C^2 \text{ as } C \to \infty \quad \text{for all } \alpha_n \ (\neq 1), a, \text{ and } m. \tag{79}$$

Using the expression for T and C given by equations (64) and (65), we find that this asymptotic relation is equivalent to

$$\Gamma^2 = -\overline{\Omega}/A. \tag{80}$$

We shall return to this relation in § 7.

On the other hand, when $C \to 0$, T tends to the Taylor number for the classical problem: 1708 when the problem is solved with the correct boundary conditions (67) and 658 when the problem is solved with the altered boundary conditions (68).

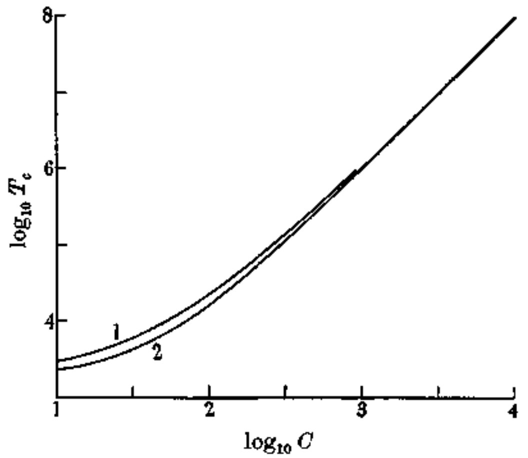


FIGURE 3. The log C-log T_c relation for the case of transverse mutual friction derived from the solution of the correct boundary value problem. The curves labelled 1 and 2 refer to values of $\alpha_n = 0$ and 0·5, respectively, and represent the instability arising from the higher branch.

Table 1 also includes the results derived from the correct solution of the boundary value problem obtained in part II; these latter results are illustrated in figure 3. In agreement with expectations, the correct solution exhibits the same general features as the solution (77); in particular, the asymptotic behaviour (79) obtains also for the exact solution.

(c) The lower branch

Considering next the branch with the negative sign in equation (75), we have

$$T = \frac{1}{2a^2}\{Q_m - [Q_m^2 - 4C^2\alpha_n^2 a^2(m^2\pi^2 + a^2)^3]^{\frac{1}{2}}\}. \tag{81}$$

For different values of m these curves again form a family of non-intersecting curves (see figure 1; the lower set of curves). However, they all tend to the same constant value, $C^2\alpha_n^2$, as $a \to \infty$:

$$T \to C^2\alpha_n^2 \quad \text{for} \quad a \to \infty. \tag{82}$$

For $a \to 0$, they tend to different constant values; thus

$$T \to \frac{C^2\alpha_n^2 m^6\pi^6}{Q_m(a=0)} = \frac{C^2\alpha_n^2}{1 + C\alpha_s/m^2\pi^2} \quad (a \to 0). \tag{83}$$

Consequently, *the solutions belonging to the lower branch are all bounded by the line* $T = C^2\alpha_n^2$ *in the* (a, T) *plane.*

In general, the critical Taylor number for the onset of instability on this lower branch is very much smaller than for the higher branch (see table 3, § 6, where the analogous results for isotropic mutual friction are given).

(*d*) *The relative amplitudes of the different velocity components at marginal instability*

The relative amplitudes of the different velocity components at marginal instability is of some interest. These can be obtained from equations (51), (52), (62) and (63) as follows:

For the lowest mode (to which we shall restrict ourselves) the different velocity components are all proportional to $\sin \pi\zeta$ and equations (62) and (63) give

$$u_\theta = -\frac{2Ad^2}{\nu}\frac{Ta^2 + C\alpha_n(\pi^2+a^2)^2}{Ta^2(\pi^2+a^2+C)}u_r$$

$$= \frac{T[(\pi^2+a^2)^2+Ca^2]}{(2\bar{\Omega}d^2/\nu)\,a^2[T+C\alpha_n(\pi^2+a^2)]}u_r. \tag{84}$$

On the other hand, according to equations (51) and (52)

$$\left.\begin{aligned} C\alpha_n v_r - \frac{2\bar{\Omega}d^2}{\nu}v_\theta &= C\alpha_n u_r, \\ \frac{2Ad^2}{\nu}v_r + C\alpha_n v_\theta &= C\alpha_n u_\theta. \end{aligned}\right\} \tag{85}$$

From these equations we derive

$$\left.\begin{aligned} v_r &= C\alpha_n \frac{(C\alpha_n)\,u_r + (2\bar{\Omega}d^2/\nu)\,u_\theta,}{C^2\alpha_n^2 - T} \\ v_\theta &= C\alpha_n \frac{-(2Ad^2/\nu)\,u_r + (C\alpha_n)\,u_\theta}{C^2\alpha_n^2 - T}. \end{aligned}\right\} \tag{86}$$

The proportionality of the amplitudes v_r and v_θ to α_n $(= \rho_n/\rho)$ clearly derives from the circumstance that the superfluid owes its motion to mutual friction with the normal fluid.

(*e*) *The origin of the two branches*

It would appear that the two independent branches along which helium II becomes unstable originates in the fact that helium II is a mixture of a viscous and a non-viscous fluid which are coupled via a mutual friction force. More specifically, we may describe the consequence of this peculiar circumstance as follows.

Consider the normal fluid. If there were no mutual friction, it would become unstable for a Taylor number equal to 1708. But the existence of mutual friction implies that when instability sets in, the normal fluid will have to drag along some additional fluid which does not interact with the walls (e.g. $v_z \neq 0$ on the boundaries) and which, on account of its inviscid nature, cannot react to its own shear. Consequently, the onset of instability of the normal fluid will be inhibited. As we have seen, the extent of this inhibition depends on the concentration, α_n, of the normal fluid and on the coupling constant, Γ, through the non-dimensional combination $2\Gamma\,|\,A\,|\,d^2/\nu$.

Turning next to the superfluid, we first observe that for an inviscid fluid we have Rayleigh's criterion for stability, namely,

$$\frac{\partial}{\partial r}(r^2\Omega) > 0, \tag{87}$$

or in our present terminology,

$$A > 0. \tag{88}$$

When $A = 0$, the system will be in a state of neutral stability; and when $A < 0$, it will be definitely unstable. Therefore, in the absence of mutual friction, the superfluid should become unstable for $T > 0$. But again, on account of mutual friction, the onset of the natural Rayleigh instability of the superfluid is also delayed. This accounts for the lower branch and also for the smallness of the lower critical Taylor number (see table 3).

On the basis of these remarks we may consider the higher branch as principally referring to the (Taylor) instability of the normal fluid, and the lower branch as principally referring to the (Rayleigh) instability of the superfluid.

6. The case of isotropic mutual friction

When the mutual friction is isotropic and has the form given by equation (15), the hydrodynamical equations governing the r and the θ components of the velocities are the same as in the case of transverse mutual friction; and these are given by equations (17), (18), (20) and (21). The equations governing the z components of the velocities are different; and equations (19) and (22) are replaced by

$$\frac{\partial u_z}{\partial t} + (\mathbf{u}\,.\,\mathrm{grad})\,u_z = -\frac{\partial \varpi_n}{\partial z} + \nu\nabla^2 u_z - \Gamma\,|\,\boldsymbol{\omega}\,|\,\alpha_s(u_z - v_z) \tag{89}$$

and

$$\frac{\partial v_z}{\partial t} + (\mathbf{v}\,.\,\mathrm{grad})\,v_z = -\frac{\partial \varpi_s}{\partial z} + \Gamma\,|\,\boldsymbol{\omega}\,|\,\alpha_n(u_z - v_z). \tag{90}$$

This difference in the two sets of equations does not affect the existence of the stationary solution (25) and (26); but the perturbation equations governing departures from the stationary state are different. Indeed, by an analysis similar to that described in § 3, we find that in place of equations (44) to (47) we now have

$$\frac{\nu}{\lambda^2}(\mathrm{DD}^* - \lambda^2)\left(\mathrm{DD}^* - \lambda^2 - \frac{p}{\nu}\right)u_r = 2\Omega u_\theta + \frac{K\alpha_s}{\lambda^2}(\mathrm{DD}^* - \lambda^2)(u_r - v_r), \tag{91}$$

$$\nu\left(\mathrm{DD}^* - \lambda^2 - \frac{p}{\nu}\right)u_\theta = 2Au_r + K\alpha_s(u_\theta - v_\theta), \tag{92}$$

$$-\frac{p}{\lambda^2}(\mathrm{DD}^* - \lambda^2)\,v_r = 2\Omega v_\theta - \frac{K\alpha_n}{\lambda^2}(\mathrm{DD}^* - \lambda^2)(u_r - v_r) \tag{93}$$

and

$$-pv_\theta = 2Av_r - K\alpha_n(u_\theta - v_\theta). \tag{94}$$

(a) *The equations governing marginal stability*

On the assumption that the principle of the exchange of stabilities is valid, the equations governing marginal stability can be obtained by putting $p = 0$ in equations (91) to (94). And by a series of transformations and eliminations the resulting equations can be reduced to the pair of equations (cf. equations (58) and (59))

$$\frac{2A}{\lambda^2}(\mathrm{DD}^* - \lambda^2)[\nu(\mathrm{DD}^* - \lambda^2) - K]\,u_r = \left[4\Omega A - \frac{K\alpha_n \nu}{\lambda^2}(\mathrm{DD}^* - \lambda^2)^2\right]u_\theta \tag{95}$$

and
$$\left[4\Omega A - \frac{K\alpha_n \nu}{\lambda^2}(\mathrm{DD}^* - \lambda^2)^2\right]u_r = 2\Omega[\nu(\mathrm{DD}^* - \lambda^2) - K]\,u_\theta; \tag{96}$$

and the boundary conditions are

$$u_r = \mathrm{D}^* u_r = u_\theta = \mathrm{DD}^* u_\theta = 0 \quad \text{for} \quad r = R_1 \text{ and } R_2; \tag{97}$$

and these are the same as before (cf. equation (60)).

(b) *The case when* $(R_2 - R_1) \ll \frac{1}{2}(R_2 + R_1)$ *and the cylinders are rotating in the same direction*

Making the approximations appropriate to this case (cf. § 5) we find that the equations we have to solve are

$$(\mathrm{D}^2 - a^2)(\mathrm{D}^2 - a^2 - C)\,u_r = \frac{2\overline{\Omega} d^2}{\nu} a^2 \left[1 + \frac{C\alpha_n}{Ta^2}(\mathrm{D}^2 - a^2)^2\right]u_\theta \tag{98}$$

and
$$(\mathrm{D}^2 - a^2 - C)\,u_\theta = \frac{2Ad^2}{\nu}\left[1 + \frac{C\alpha_n}{Ta^2}(\mathrm{D}^2 - a^2)^2\right]u_r, \tag{99}$$

together with the boundary conditions

$$u_r = \mathrm{D}u_r = u_\theta = \mathrm{D}^2 u_\theta = 0 \quad \text{for} \quad \zeta = 0 \text{ and } 1. \tag{100}$$

Eliminating u_θ between equations (98) and (99) we obtain

$$(\mathrm{D}^2 - a^2)(\mathrm{D}^2 - a^2 - C)^2 u_r = -Ta^2\left[1 + \frac{C\alpha_n}{Ta^2}(\mathrm{D}^2 - a^2)^2\right]^2 u_r. \tag{101}$$

(c) *The solution satisfying the boundary conditions* (68)

In part II a method is described for solving the characteristic value problem defined by equations (98) to (100). Here, as in § 5, we shall consider the explicit solution which can be written down for the slightly modified (non-physical!) boundary conditions (68). For these latter boundary conditions u_r, u_θ and all their even derivatives vanish for $\zeta = 0$ and 1. The solutions for u_r and u_θ must, therefore, be multiples of $\sin m\pi\zeta$ where m is an integer. Considering the lowest mode ($m = 1$), we obtain from equation (101) the corresponding characteristic equation

$$(\pi^2 + a^2)(\pi^2 + a^2 + C)^2 = Ta^2\left[1 + \frac{C\alpha_n}{Ta^2}(\pi^2 + a^2)^2\right]^2, \tag{102}$$

or, equivalently,
$$T - \frac{1}{a}(\pi^2 + a^2 + C)(\pi^2 + a^2)^{\frac{1}{2}}\sqrt{T} + C\alpha_n\frac{(\pi^2 + a^2)^2}{a^2} = 0. \tag{103}$$

Hence
$$\sqrt{T} = \frac{1}{2a}(\pi^2+a^2+C)(\pi^2+a^2)^{\frac{1}{2}}\left\{1 \pm \left[1-\frac{4C\alpha_n(\pi^2+a^2)}{(\pi^2+a^2+C)^2}\right]^{\frac{1}{2}}\right\}. \quad (104)$$

Thus, as in the case of transverse mutual friction, we have again two branches (see figure 4). The critical Taylor numbers for the two branches derived from

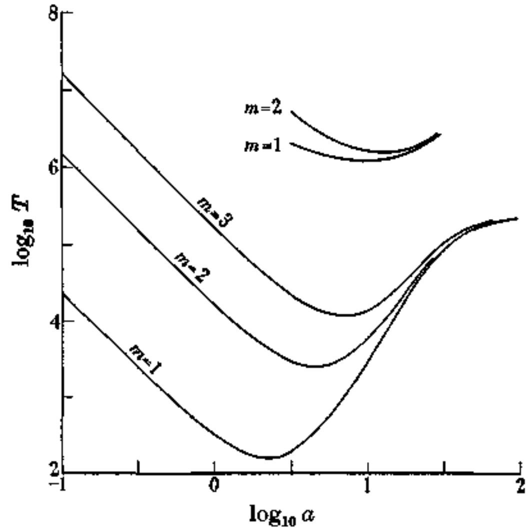


FIGURE 4. The dependence of the Taylor number (T) on the wave number a (in units of $1/d$) for the two branches and various modes for the case of isotropic mutual friction. The curves refer to a value of $C = 1000$ and $\alpha_n = 0{\cdot}5$.

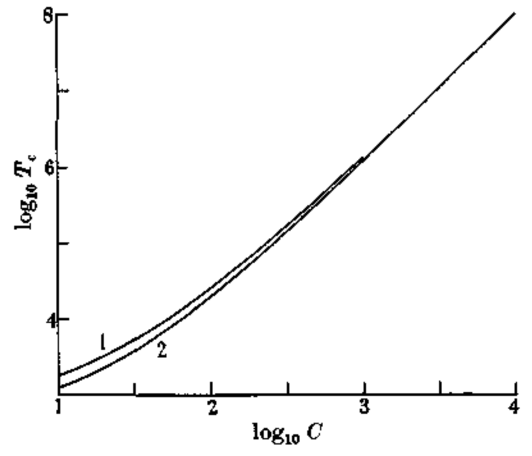


FIGURE 5. The log C-log T_c relation for the case of isotropic mutual friction derived from the solution of the modified boundary value problem. The curves labelled 1 and 2 refer to values of $\alpha_n = 0$ and $0{\cdot}5$, respectively, and represent the instability arising from the higher branch.

equation (104) are given in tables 2 and 3 and illustrated in figure 5. Table 2 also includes the results derived from the correct solution of the characteristic value problem obtained in II; these latter results are illustrated in figure 6.

It will be noticed that apart from differences in detail the solutions obtained for the transverse and the isotropic forms for the mutual friction are remarkably

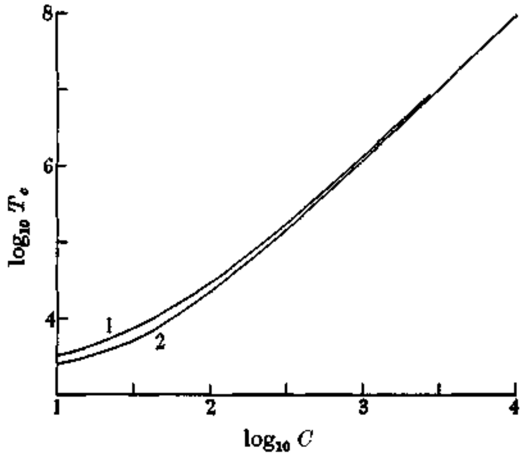


FIGURE 6. The log C-log T_c relation for the case of isotropic mutual friction derived from the solution of the correct boundary value problem. The curves labelled 1 and 2 refer to values of $\alpha_n = 0$ and 0·5, respectively, and represent the instability arising from the higher branch.

TABLE 2. THE CRITICAL TAYLOR NUMBERS APPROPRIATE TO THE HIGHER BRANCH AND FOR THE ISOTROPIC FORM OF THE MUTUAL FRICTION FOR VARIOUS VALUES OF C AND α_n

	the solution for the modified boundary value problem				the solution for the correct boundary value problem			
	$\alpha_n = 0$		$\alpha_n = 0·5$		$\alpha_n = 0$		$\alpha_n = 0·5$	
C	a	T_c	a	T_c	a	T_c	a	T_c
0	2·23	$6·575 \times 10^2$	2·2	$6·575 \times 10^2$	3·13	$1·715 \times 10^3$	3·13	$1·715 \times 10^3$
10	2·78	$1·734 \times 10^3$	2·7	$1·301 \times 10^3$	3·55	$3·184 \times 10^3$	3·4	$2·546 \times 10^3$
20	3·15	$3·158 \times 10^3$	3·2	$2·301 \times 10^3$	3·85	$4·954 \times 10^3$	3·7	$3·676 \times 10^3$
40	3·66	$6·951 \times 10^3$	3·8	$5·196 \times 10^3$	4·27	$9·377 \times 10^3$	4·3	$6·797 \times 10^3$
60	4·03	$1·192 \times 10^4$	4·3	$9·179 \times 10^3$	4·59	$1·494 \times 10^4$	4·7	$1·100 \times 10^4$
80	4·33	$1·801 \times 10^4$	4·8	$1·420 \times 10^4$	4·85	$2·159 \times 10^4$	5·0	$1·624 \times 10^4$
100	4·58	$2·518 \times 10^4$	5·1	$2·021 \times 10^4$	5·1	$2·930 \times 10^4$	5·3	$2·247 \times 10^4$
200	5·46	$7·647 \times 10^4$	6·1	$6·478 \times 10^4$	5·9	$8·320 \times 10^4$	6·4	$6·808 \times 10^4$
300	6·06	$1·524 \times 10^5$	6·9	$1·326 \times 10^5$	6·4	$1·616 \times 10^5$	7·1	$1·369 \times 10^5$
400	6·53	$2·522 \times 10^5$	7·5	$2·232 \times 10^5$	6·9	$2·638 \times 10^5$	7·7	$2·285 \times 10^5$
500	6·91	$3·752 \times 10^5$	7·9	$3·361 \times 10^5$	7·2	$3·892 \times 10^5$	8·1	$3·423 \times 10^5$
600	7·24	$5·058 \times 10^5$	8·4	$4·711 \times 10^5$	7·6	$5·375 \times 10^5$	8·6	$4·782 \times 10^5$
700	7·53	$6·899 \times 10^5$	8·7	$6·280 \times 10^5$	7·8	$7·085 \times 10^5$	8·9	$6·360 \times 10^5$
800	7·80	$8·811 \times 10^5$	9·1	$8·067 \times 10^5$	8·1	$9·020 \times 10^5$	9·2	$8·156 \times 10^5$
900	8·04	$1·095 \times 10^6$	9·3	$1·007 \times 10^6$	8·3	$1·118 \times 10^6$	9·5	$1·017 \times 10^6$
1000	8·26	$1·330 \times 10^6$	9·6	$1·229 \times 10^6$	8·5	$1·356 \times 10^6$	9·8	$1·240 \times 10^6$
2000	9·86	$4·890 \times 10^6$	11·5	$4·621 \times 10^6$	10·1	$4·938 \times 10^6$	11·7	$4·640 \times 10^6$
5000	12·4	$2·837 \times 10^7$	14·7	$2·736 \times 10^7$	12·6	$2·848 \times 10^7$	14·8	$2·741 \times 10^7$
10000	14·8	$1·093 \times 10^8$	17	$1·066 \times 10^8$	15·0	$1·096 \times 10^8$	17·1	$1·067 \times 10^8$

TABLE 3. THE CRITICAL TAYLOR NUMBERS APPROPRIATE TO THE LOWER BRANCH AND FOR THE ISOTROPIC FORM OF MUTUAL FRICTION. THE VALUES ARE DERIVED FROM THE SOLUTION OF THE CORRECT BOUNDARY VALUE PROBLEM

C	500	600	700	800	900	1000
a	3·7	3·8	3·8	3·9	3·9	3·9
T_c	657	698	733	767	797	825

similar. The most important difference is that, for the transverse form, the regions of instability defined by the two branches are clearly separated by the line $T = \alpha_n C^2$, while this is not the case for the isotropic form (cf. figures 1 and 4).

7. Proposed experimental study and comparisons with some existing experiments

An experimental investigation of the phenomenon theoretically examined in this paper can be conveniently carried out by studying the behaviour of helium II between a rotating inner cylinder and a stationary outer cylinder. In devising such experiments, some care should be exercised in the choice of the geometrical parameters of the apparatus such as the radii of the inner and the outer cylinders. For, if a viscometer should be constructed with some assigned values of R_1 and R_2 (satisfying, however, the requirement $(R_2 - R_1) \ll \frac{1}{2}(R_2 + R_1)$) and the experiment should consist in rotating the inner cylinder with varying angular velocities, then in the (C, T) plane we should be restricted to the locus (cf. equations (27), (64) and (65))

$$\frac{T}{C^2} = \frac{\bar{\Omega}}{\Gamma^2 |A|} = \frac{R_2^2 - R_1^2}{2R_1^2 \Gamma^2}. \tag{105}$$

It is important to notice that this relation is independent of the manner in which ν may be defined; i.e. independent of whether ν is defined as the kinematic viscosity with respect to the density of the normal fluid (as is done in this paper) or with respect to the density of helium II (as is more customary).

Hence, in the $(\log C, \log T)$ plane (such as is illustrated in figures 2, 3, 5 and 6) we shall be restricted to a straight line parallel to the asymptote

$$\log T = 2 \log C, \tag{106}$$

to which the various stability curves belonging to the higher branch tend. Consequently, an improper choice of R_1 and R_2 may restrict one to a locus which does not pass through any of the interesting parts of this plane.

It is perhaps worth pointing out in this connexion that the critical curve

$$T = \alpha_n^2 C^2, \tag{107}$$

which (on the transverse form for the mutual friction) limits the lower branch, is, in contrast to the asymptote $T = C^2$ of the upper branch, highly temperature-dependent since it depends directly on $(\rho_n/\rho)^2$.

At the present time, one of us (R. J. D.) is assembling apparatus to study the behaviour of helium II between rotating cylinders. A rotating cylinder viscometer is being constructed with the inner cylinder capable of rotation and the outer cylinder supported by a torsion fibre. Interchangeable inner cylinders are being provided so that a number of different loci in the $(\log C, \log T)$ plane can be traversed. In this manner it is hoped to investigate the different types of instabilities which have been predicted.

We now turn to some experiments which can be interpreted, tentatively, on the basis of the theory developed in this paper. A recent experiment by Kolm &

Herlin (1956) appears to be in this category. In their experiments† a soft iron core was suspended within a rigid coaxial outer cylinder by a magnetic bearing of the type used by Beams and others. The inner cylinder was accelerated electromagnetically and the viscosity was determined by observing the deceleration of the inner cylinder rotating freely in liquid helium. And Kolm & Herlin report an observation which suggests a critical angular velocity (and, therefore, a critical Taylor number). They observed, that, at $2{\cdot}135^\circ$ K, the effective viscosity was constant and equal to $63{\cdot}3\,\mu$P so long as Ω_1 was greater than 120 rad/min; that it decreased almost discontinuously to the value $36{\cdot}6\,\mu$P as Ω_1 decreased below 120 rad/min; and finally, that it remained constant at the lower value as Ω_1 continued to decrease. From the dimensions of their apparatus ($R_1 = 0{\cdot}635$ cm and $R_2 = 0{\cdot}794$ cm) and the coefficient of kinematic viscosity ($= 1{\cdot}34 \times 10^{-4}\,\mathrm{cm^2\,s^{-1}}$ if referred to the density of helium II, and $1{\cdot}61 \times 10^{-4}\,\mathrm{cm^2\,s^{-1}}$ if referred to the density of the normal fluid ($= 0{\cdot}83\rho$) at the temperature ($= 2{\cdot}135^\circ$ K) of the experiments) as measured by Heikkila & Hollis-Hallett (1955), we find that the value of Ω_1 ($= 2\,\mathrm{rad\,s^{-1}}$ leading to $A = -3{\cdot}56$)) corresponds to a Taylor number $T_c = 3{\cdot}49 \times 10^5$. The effective viscosity below this value of T_c (namely, $36{\cdot}6\,\mu$P) is still a factor two higher than the value given by Heikkila & Hollis-Hallett (1955). This difference may be due to some residual turbulence resulting from the instability due to the lower branch.

If we interpret $T_c = 3{\cdot}49 \times 10^5$ as the critical Taylor number for the onset of instability on the higher branch when $\alpha_n = 0{\cdot}83$ (appropriate for the temperature at which the discontinuity was observed) then from the results given in table 1 (the last column) we deduce a value of $C = 575$. From this value of C we derive $\Gamma = C\nu/2Ad^2 = 0{\cdot}52$. This value of Γ should be compared with the value $\Gamma = 0{\cdot}55$ which Hall & Vinen (1956*a*; their coefficient B is related to Γ by $\Gamma = \frac{1}{2}B$) deduce from an entirely unrelated experiment. This agreement between the two determinations may be regarded as a confirmation of the interpretation we have placed on Kolm & Herlin's experiment. On this interpretation the fact that they measured a constant viscosity before T_c was reached can also be understood; for, the locus described by their experiments in the ($\log C$, $\log T$) plane intersects the line of instability at so small an angle that the level of turbulence must have remained approximately constant prior to the intersection.

An examination of another experiment by Wheeler, Blakewood & Lane (1955) is also instructive. In their experiments which demonstrated the attenuation of second sound in rotating helium II, the liquid helium was confined between an outer cylinder of radius $R_2 = 1{\cdot}588$ cm and an inner rotating cylinder of radius $R_1 = 0{\cdot}794$ cm. These dimensions together with the value of Γ as determined by Hall & Vinen imply that the locus described by their experiments in the ($\log C$, $\log T$) plane lies far above the limit of stability set by the higher branch. Under the conditions of their experiments, therefore, helium II must have been highly turbulent. This will account for the fact that Wheeler *et al.* deduce a value of $\Gamma = 4{\cdot}3$ which is highly discordant with Hall & Vinen's determination. One might conclude from this that the attenuation of second sound is sensitive to the prevalence of

† We are indebted to the authors for some additional (unpublished) information regarding their experiments.

turbulence in helium II; this in turn may provide a useful criterion for establishing the different regimes of instability in the $(\log C, \log T)$ plane.

In concluding this paper, we should like to express our thanks to Miss Donna Elbert for carrying out the numerical solutions of the various characteristic value problems; tables 1, 2 and 3 are the results of her efforts. One of us (R.J.D.) is further indebted to the National Science Foundation for supporting in large measure his research.

References

Chandrasekhar, S. 1954 *Mathematika*, **1**, 5.
Chandrasekhar, S. 1957 *Proc. Roy. Soc.* A, **241**, 29 (part II).
Daunt, J. G. & Smith, R. S. 1954 *Rev. Mod. Phys.* **26**, 172.
Feynman, R. P. 1955 *Progress in low temperature physics* (ed. C. J. Gorter), **1**, chap. 2, p. 17. Amsterdam: North Holland Publishing Co.
Hall, H. E. & Vinen, W. F. 1956*a* *Proc. Roy. Soc.* A, **238**, 204.
Hall, H. E. & Vinen, W. F. 1956*b* *Proc. Roy. Soc.* A, **238**, 215.
Heikkila, W. J. & Hollis-Hallett, A. C. 1955 *Canad. J. Phys.* **33**, 420.
Kolm, H. H. & Herlin, M. A. 1956 *Phys. Rev.* **102**, 607.
Taylor, G. I. 1923 *Phil. Trans.* A, **223**, 289.
Wheeler, R. G., Blakewood, C. H. & Lane, C. T. 1955 *Phys. Rev.* **99**, 1667.

PRINTED IN GREAT BRITAIN AT THE UNIVERSITY PRESS, CAMBRIDGE
(BROOKE CRUTCHLEY, UNIVERSITY PRINTER)

Reprinted without change of pagination from the
Proceedings of the Royal Society, A, *volume* 241, pp. 29–36, 1957

The hydrodynamic stability of helium II between rotating cylinders. II

BY S. CHANDRASEKHAR, F.R.S.
Enrico Fermi Institute for Nuclear Studies, University of Chicago

(*Received* 6 *March* 1957)

The characteristic value problems (Chandrasekhar & Donnelly 1957), to which a study of the hydrodynamic stability of helium II between rotating cylinders leads one, are solved here correctly with respect to the boundary conditions which have to be satisfied. The nature and origin of the two types of instabilities which occur in this problem are further examined by considering the growth rate of small perturbations.

1. INTRODUCTION

The stability of the viscous flow of helium II between rotating cylinders has been considered in part I (Chandrasekhar & Donnelly 1957; preceding paper); and in the particular case of the cylinders rotating in the same direction and the difference in the radii of the cylinders being small compared with their mean radius, the problem was reduced to one in characteristic values of certain high order differential equations. In part I the problem was not solved for the correct set of boundary conditions; it was solved for a modified (non-physical!) set of boundary conditions for which an explicit solution could be written down. This latter procedure is sufficient so long as one's principal interest is in the nature of the dependence of the characteristic values on the various parameters of the problem. Nevertheless, it is unsatisfactory to leave the basic mathematical problem unsolved. In this part a method of solving the characteristic value problem with the correct set of boundary conditions will be described. A further problem which will also be considered concerns the character of the transition that takes place as we cross the two stability branches discussed in part I.

2. THE SOLUTION OF THE CHARACTERISTIC VALUE PROBLEM FOR THE CASE OF TRANSVERSE MUTUAL FRICTION

In this case the characteristic value problem to be solved is defined by the equations

$$[(\mathrm{D}^2-a^2)^2+Ca^2]\,u_r = -Ta^2\left[1-\frac{qC}{T}(\mathrm{D}^2-a^2)\right]u_\theta \tag{1}$$

and
$$(\mathrm{D}^2-a^2-C)\,u_\theta = \left[1+\frac{qC}{Ta^2}(\mathrm{D}^2-a^2)^2\right]u_r, \tag{2}$$

and the boundary conditions

$$u_r = \mathrm{D}u_r = u_\theta = \mathrm{D}^2u_\theta = 0 \quad \text{for} \quad \zeta = 0 \text{ and } 1. \tag{3}$$

Equations (1) and (2) follow from part I, equations (62) and (63), by replacing u_θ by $(2Ad^2/\nu)\,u_\theta$; also, q has been written in place of α_n.

The method of solution now to be described is an adaptation of the one which the author (Chandrasekhar 1954) has recently used for the solution of Taylor's original problem.

(a) The method

In view of the boundary conditions on u_θ, we shall expand it in a sine series of the form

$$u_\theta = \sum_{n=1}^{\infty} S_n \sin n\pi\zeta. \tag{4}$$

Having chosen u_θ in this manner we shall next *solve* the equation

$$\begin{aligned}[(\mathrm{D}^2-a^2)^2+Ca^2]\,u_r &= -[Ta^2-qCa^2(\mathrm{D}^2-a^2)]\sum_{n=1}^{\infty} S_n \sin n\pi\zeta \\ &= -\sum_{n=1}^{\infty} S_n[Ta^2+qCa^2(n^2\pi^2+a^2)]\sin n\pi\zeta, \end{aligned} \tag{5}$$

obtained by inserting (4) in (1), and arrange that the solution satisfies the four boundary conditions on u_r; since equation (5) is of the fourth order, there will be just enough constants of integration to do this. With u_r determined in this fashion and u_θ given by (4), equation (2) will lead, as we shall see in detail presently, to an infinite-order determinant which must be zero if all the S_n's are not to vanish identically. It is in this way that we shall obtain the characteristic equation for T.

(b) The characteristic equation

We shall now proceed to obtain the explicit form of the characteristic equation for T.

The solution of equation (5) is, of course, straightforward. The general solution can be written in the form

$$\begin{aligned} u_r = \sum_{m=1}^{\infty} S_m \gamma_m m\pi \Big[& A_1^{(m)} \cosh \beta_1 \zeta + B_1^{(m)} \sinh \beta_1 \zeta \\ & + A_2^{(m)} \cosh \beta_2 \zeta + B_2^{(m)} \sinh \beta_2 \zeta - \frac{1}{m\pi} \sin m\pi\zeta \Big], \end{aligned} \tag{6}$$

where
$$\gamma_m = \frac{a^2[T+qC(m^2\pi^2+a^2)]}{(m^2\pi^2+a^2)^2+Ca^2}, \tag{7}$$

β_1 and β_2 are a pair of complex conjugate roots of the equation

$$(\beta^2-a^2)^2+Ca^2 = 0, \tag{8}$$

and $A_1^{(m)}$, $B_1^{(m)}$, $A_2^{(m)}$ and $B_2^{(m)}$ are constants of integration to be determined by the boundary conditions $u_r = \mathrm{D}u_r = 0$ at $\zeta = 0$ and 1. These latter conditions lead to the equations

$$A_1^{(m)} + A_2^{(m)} = 0, \quad \beta_1 B_1^{(m)} + \beta_2 B_2^{(m)} = 1, \tag{9}$$

$$A_1^{(m)} \cosh \beta_1 + B_1^{(m)} \sinh \beta_2 + A_2^{(m)} \cosh \beta_2 + B_2^{(m)} \sinh \beta_2 = 0, \tag{10}$$

$$A_1^{(m)} \beta_1 \sinh \beta_1 + B_1^{(m)} \beta_1 \cosh \beta_1 + A_2^{(m)} \beta_2 \sinh \beta_2 + B_2^{(m)} \beta_2 \cosh \beta_2 = (-1)^m. \tag{11}$$

On solving these equations we find

$$A_1^{(m)} = \frac{1}{\Delta}\{\beta_2 \sinh\beta_1[\cosh\beta_2 + (-1)^{m+1}] - \beta_1 \sinh\beta_2[\cosh\beta_1 + (-1)^{m+1}]\}$$
$$= -A_2^{(m)}, \tag{12}$$

$$B_1^{(m)} = \frac{1}{\Delta}\{\beta_2(\cosh\beta_2 - \cosh\beta_1)[\cosh\beta_2 + (-1)^{m+1}] - \sinh\beta_2(\beta_2 \sinh\beta_2 - \beta_1 \sinh\beta_1)\}, \tag{13}$$

$$B_2^{(m)} = \frac{1}{\Delta}\{-\beta_1(\cosh\beta_2 - \cosh\beta_1)[\cosh\beta_1 + (-1)^{m+1}] + \sinh\beta_1(\beta_2 \sinh\beta_2 - \beta_1 \sinh\beta_1)\}, \tag{14}$$

where $$\Delta = (\beta_1^2 + \beta_2^2)\sinh\beta_1 \sinh\beta_2 - 2\beta_1\beta_2(\cosh\beta_1 \cosh\beta_2 - 1). \tag{15}$$

Now substituting for u_r and u_θ from equations (4) and (6) in equation (2), we obtain

$$-Ta^2 \sum_{n=1}^{\infty} S_n(n^2\pi^2 + a^2 + C)\sin n\pi\zeta$$
$$= -\sum_{m=1}^{\infty} S_m \gamma_m [Ta^2 + qC(m^2\pi^2 + a^2)^2]\sin m\pi\zeta$$
$$+ a^2(T - qC^2)\sum_{m=1}^{\infty} S_m \gamma_m m\pi [A_1^{(m)}\cosh\beta_1\zeta + A_2^{(m)}\cosh\beta_2\zeta$$
$$+ B_1^{(m)}\sinh\beta_1\zeta + B_2^{(m)}\sinh\beta_2\zeta]. \tag{16}$$

Next, multiplying this equation by $\sin n\pi\zeta$ and integrating over the range of ζ, we obtain

$$\frac{1}{2a^2(T - qC^2)}\{\gamma_n[Ta^2 + qC(n^2\pi^2 + a^2)^2] - Ta^2(n^2\pi^2 + a^2 + C)\}S_n$$
$$= \sum_{m=1}^{\infty}\left\{\frac{1}{n^2\pi^2 + \beta_1^2}[A_1^{(m)}[1 + (-1)^{n+1}\cosh\beta_1] + (-1)^{n+1}B_1^{(m)}\sinh\beta_1]\right.$$
$$\left. + \frac{1}{n^2\pi^2 + \beta_2^2}[A_2^{(m)}[1 + (-1)^{n+1}\cosh\beta_2] + (-1)^{n+1}B_2^{(m)}\sinh\beta_2]\right\}S_m\gamma_m mn\pi^2. \tag{17}$$

Finally, substituting for $A_1^{(m)}$, etc., in accordance with equations (12) to (15) in equation (17) we obtain after some lengthy but elementary reductions

$$\sum_{m=1}^{\infty} S_m\gamma_m\left[\frac{2mn\pi^2a^3(T - qC^2)\sqrt{C}}{(n^2\pi^2 + a^2)^2 + Ca^2}\,\frac{[(-1)^{n+1} + (-1)^{m+1}]\cos\alpha_2 + [1 + (-1)^{m+n}]\cosh\alpha_1}{\alpha_2\sinh\alpha_1 + (-1)^{n+1}\alpha_1\sin\alpha_2}\right.$$
$$\left. - \delta_{mn}\left\{[Ta^2 + qC(n^2\pi^2 + a^2)^2] - T\frac{(n^2\pi^2 + a^2 + C)[(n^2\pi^2 + a^2)^2 + Ca^2]}{T + qC(n^2\pi^2 + a^2)}\right\}\right] = 0; \tag{18}$$

where $$\alpha_1 = [\tfrac{1}{2}(a^4 + Ca^2)^{\frac{1}{2}} + \tfrac{1}{2}a^2]^{\frac{1}{2}} \quad \text{and} \quad \alpha_2 = [\tfrac{1}{2}(a^4 + Ca^2)^{\frac{1}{2}} - \tfrac{1}{2}a^2]^{\frac{1}{2}}. \tag{19}$$

The infinite determinant formed by the coefficients of $S_m\gamma_m$ in equation (18) must clearly vanish; and this provides the required characteristic equation.

(c) The equation for determining the characteristic values belonging to the lowest mode in the first approximation

A method of solving the infinite order characteristic equation which (18) provides for T would be to set the determinant formed by the first n rows and columns of the matrix equal to zero and let n take increasingly large values. In practice the usefulness of the method will depend on how rapidly the process converges. Actually, in this instance by setting the (1, 1) element of the matrix equal to zero we already obtain the required characteristic values within a small percentage of the true value. Therefore, to this accuracy, the characteristic values belonging to the lowest mode are determined by the equation

$$[Ta^2+qC(\pi^2+a^2)^2]-T\frac{(\pi^2+a^2+C)[(\pi^2+a^2)^2+Ca^2]}{T+qC(\pi^2+a^2)}$$
$$=\frac{4a^3\pi^2(T-qC^2)\sqrt{C}}{(\pi^2+a^2)^2+Ca^2}\,\frac{\cosh\alpha_1+\cos\alpha_2}{\alpha_2\sinh\alpha_1+\alpha_1\sin\alpha_2}. \tag{20}$$

Calculations based on this formula have already been included in part I, table 1.

3. The solution of the characteristic value problem for the case of isotropic mutual friction

In this case the equations to be solved are (cf. part I, equations (98) and (99))

$$(\mathrm{D}^2-a^2)(\mathrm{D}^2-a^2-C)\,u_r=-Ta^2\left[1+\frac{qC}{Ta^2}(\mathrm{D}^2-a^2)^2\right]u_\theta \tag{21}$$

and

$$(\mathrm{D}^2-a^2-C)\,u_\theta=\left[1+\frac{qC}{Ta^2}(\mathrm{D}^2-a^2)^2\right]u_r, \tag{22}$$

together with the boundary conditions

$$u_r=\mathrm{D}u_r=u_\theta=\mathrm{D}^2u_\theta=0\quad\text{for}\quad\zeta=0\text{ and }1. \tag{23}$$

This problem can be solved by a procedure very similar to that described in § 2. We shall, therefore, omit all the details and quote only the final determinantal equation. We find

$$\left\|\frac{4\pi^2mnQ_{mn}C}{(m^2\pi^2+a^2)(m^2\pi^2+b^2)}[Ta^2+qC(m^2\pi^2+a^2)^2][Ta^2-qC^2(n^2\pi^2+a^2)]\right.$$
$$\left.-\delta_{mn}\{[Ta^2+qC(n^2\pi^2+a^2)^2]^2-Ta^2(n^2\pi^2+a^2)(n^2\pi^2+b^2)^2\}\right\|=0, \tag{24}$$

where

$$b=\sqrt{(a^2+C)} \tag{25}$$

and

$$Q_{mn}=[b\sinh a\{[(-1)^{n+1}+(-1)^{m+1}]+[1+(-1)^{m+n}]\cosh b\}$$
$$-a\sinh b\{[(-1)^{n+1}+(-1)^{m+1}]+[1+(-1)^{m+n}]\cosh a\}]$$
$$\div 2[(a^2+b^2)\sinh a\,\sinh b-2ab(\cosh a\,\cosh b-1)]. \tag{26}$$

In particular the equation for determining the characteristic values belonging to the lowest mode in the first approximation is

$$[Ta^2+qC(\pi^2+a^2)^2]^2-Ta^2(\pi^2+a^2)(\pi^2+b^2)^2$$
$$=\frac{4\pi^2Q_{11}C}{(\pi^2+a^2)(\pi^2+b^2)}[Ta^2+qC(\pi^2+a^2)^2][Ta^2-qC^2(\pi^2+a^2)]. \tag{27}$$

Again this formula suffices to determine the required characteristic values to an accuracy which is better than 1 %. Calculations based on this formula have already been included in tables 2 and 3 of part I.

4. On the manner of the onset of instabilities in the (a, T) plane

For a proper understanding of the two branches of instability in the (a, T) plane which the analysis of part I has disclosed, it is necesary to go back to the general equations governing the perturbations, i.e. before p which occurs in the time dependent factor, e^{pt} (cf. equation (36) of part I) has been put equal to zero. Thus, we must go back to equations (44) to (47) or (91) to (94) of part I, depending on whether we are dealing with the case of transverse or isotropic mutual friction.

(a) *The case of transverse mutual friction*

Considering equations (44) to (47) of part I for the case when $(R_2-R_1) \ll \frac{1}{2}(R_2+R_1)$ and the two cylinders are rotating in the same direction, and making the approximations appropriate to this case (i.e. not distinguishing between D and D*, and replacing Ω by $\bar{\Omega}$), we have

$$\frac{1}{a^2}(D^2-a^2)\left(D^2-a^2-\frac{pd^2}{\nu}\right)u_r = \frac{2\bar{\Omega}d^2}{\nu}u_\theta - \frac{K\alpha_s d^2}{\nu}(u_r-v_r), \tag{28}$$

$$\left(D^2-a^2-\frac{pd^2}{\nu}\right)u_\theta = \frac{2Ad^2}{\nu}u_r + \frac{K\alpha_s d^2}{\nu}(u_\theta - v_\theta), \tag{29}$$

$$-\frac{pd^2}{a^2\nu}(D^2-a^2)v_r = \frac{2\bar{\Omega}d^2}{\nu}v_\theta + \frac{K\alpha_n d^2}{\nu}(u_r-v_r) \tag{30}$$

and

$$-\frac{pd^2}{\nu}v_\theta = \frac{2Ad^2}{\nu}v_r - \frac{K\alpha_n d^2}{\nu}(u_\theta-v_\theta), \tag{31}$$

where radial distances are now measured from the inner cylinder in units of d $(=R_2-R_1)$ and $\lambda = a/d$.

We shall consider equations (28) to (31) for the case when all the velocity components together with all their even derivatives vanish on the boundaries. This is the 'artificial' case considered in part I; but as we have seen the consideration of this case suffices for an understanding of the nature of the phenomenon we are dealing with. Assuming then that all the velocity components are proportional to $\sin \pi\zeta$ (corresponding to the lowest mode) we find that the equations become

$$\frac{1}{x}(1+x)(1+x+\sigma)u_r - \xi u_\theta + \kappa_s(u_r-v_r) = 0, \tag{32}$$

$$(1+x+\sigma)u_\theta - \eta u_r + \kappa_s(u_\theta - v_\theta) = 0, \tag{33}$$

$$-\frac{1}{x}(1+x)\sigma v_r + \xi v_\theta + \kappa_n(u_r-v_r) = 0 \tag{34}$$

and

$$\sigma v_\theta - \eta v_r - \kappa_n(u_\theta - v_\theta) = 0, \tag{35}$$

where the following abbreviations have been used:

$$x = \frac{a^2}{\pi^2}, \quad \sigma = \frac{pd^2}{\nu\pi^2}, \quad \xi = \frac{2\overline{\Omega}d^2}{\nu\pi^2}, \quad \eta = -\frac{2Ad^2}{\nu\pi^2}, \tag{36}$$

$$\kappa_n = \frac{K\alpha_n}{\nu}\frac{d^2}{\pi^2} = \frac{C\alpha_n}{\pi^2} \quad \text{and} \quad \kappa_s = \frac{K\alpha_s}{\nu}\frac{d^2}{\pi^2} = \frac{C\alpha_s}{\pi^2}. \tag{37}$$

We can rewrite equations (32) to (37) as a single matrix equation in the form

$$\begin{vmatrix} (1+x)(1+x+\sigma)+x\kappa_s & -x\kappa_s & -x\xi & 0 \\ -\eta & 0 & 1+x+\sigma+\kappa_s & -\kappa_s \\ x\kappa_n & -\sigma(1+x)-x\kappa_n & 0 & x\xi \\ 0 & -\eta & -\kappa_n & \sigma+\kappa_n \end{vmatrix} \begin{vmatrix} u_r \\ v_r \\ u_\theta \\ v_\theta \end{vmatrix} = 0. \tag{38}$$

The determinant of the matrix on the left-hand side of this equation must therefore vanish. By adding suitable multiples of the different rows (and columns) to other rows (and columns) we find that the determinant can be reduced to

$$\begin{vmatrix} (1+x)^2 & -[(1+x)(1+x+\sigma)+x\kappa] & 0 & x\tau \\ 0 & 1 & 1+x & -(1+x+\sigma+\kappa) \\ -\sigma(1+x) & -x\kappa_n & x\tau & 0 \\ -1 & 0 & \sigma & \kappa_n \end{vmatrix} = 0, \tag{39}$$

where $$\kappa = \kappa_n + \kappa_s = \frac{Kd^2}{\nu\pi^2} = \frac{C}{\pi^2} \quad \text{and} \quad \tau = \xi\eta = \frac{T}{\pi^4}. \tag{40}$$

Now expanding the determinant (39) by the elements of the first column and rearranging, we finally obtain

$$\begin{aligned} &x(1+x)^2\kappa_n\{\tau+(1+x)\kappa_n+\sigma(1+x+\sigma+\kappa)\} \\ &\quad +\sigma(1+x)\{[(1+x)(1+x+\sigma)+\kappa x][(1+x)\kappa_n+\sigma(1+x+\sigma+\kappa)]-x\sigma\tau\} \\ &\qquad +x\tau\{x[\tau+(1+x)\kappa_n]-(1+x+\sigma+\kappa)[(1+x)(1+x+\sigma)+\kappa x]\} = 0. \end{aligned} \tag{41}$$

Some immediate consequences of equation (41) are the following:

(i) If we put $\sigma = 0$ in equation (41) we find that

$$[\tau+(1+x)\kappa_n][(1+x)^2\kappa_n+x\tau] = \tau(1+x+\kappa)[(1+x)^2+\kappa x]; \tag{42}$$

in virtue of the definitions (37) and (40) this equation is identical with the condition for marginal stability derived in part I (equation (72) for the case $m = 1$) on the principle of the exchange of stabilities.

(ii) To determine the nature of the transition across the two stability branches defined by equation (42), it will suffice to consider terms of order σ only. To this order, the equation is

$$\begin{aligned} &\sigma\{\tau x[(1+x)^2+\kappa x+(1+x)(1+x+\kappa)]-\kappa_n(1+x)^2[(1+x)^2+\kappa x+x(1+x+\kappa)]\} \\ &\quad = x\{[\tau+(1+x)\kappa_n][(1+x)^2\kappa_n+x\tau]-\tau(1+x+\kappa)[(1+x)^2+\kappa x]\}. \end{aligned} \tag{43}$$

In the (x, τ) plane the quantity on the right-hand side of equation (43) is positive in the region above the higher branch; it is negative in the region bounded by the two branches; and it is again positive in the region bounded by the lower branch and the x axis. Also, it can be shown that in the immediate neighbourhood of the higher branch, the coefficient of σ in equation (43) is positive, while in the immediate neighbourhood of the lower branch it is negative. Accordingly, the transition, as we cross *either* branch in the direction of increasing τ, is one from stability to instability.

(iii) On the x axis the equation reduces to

$$[\sigma^2+\sigma(1+x+\kappa)+(1+x)\kappa_n][\sigma^2(1+x)+\sigma\{(1+x)^2+\kappa x\}+x(1+x)\kappa_n] = 0. \quad (44)$$

All four roots of this equation are real and negative. Hence on the x axis there is complete stability.

(iv) Considering next the limiting case $\kappa_n = 0$, we have

$$\sigma^2(1+x)\{[(1+x)(1+x+\sigma)+\kappa x](1+x+\sigma+\kappa)-x\tau\} \\ +x\tau\{x\tau-[(1+x)(1+x+\sigma)+\kappa x](1+x+\sigma+\kappa)\} = 0. \quad (45)$$

This equation can be reduced to the form

$$[\sigma^2(1+x)-x\tau]\{\sigma^2(1+x)+\sigma[(1+x)^2+\kappa x+(1+x)(1+x+\kappa)] \\ +[(1+x)^2+\kappa x](1+x+\kappa)-\tau x\} = 0. \quad (46)$$

One pair of roots is, therefore,

$$\sigma = \pm\left(\frac{\tau x}{1+x}\right)^{\frac{1}{2}}, \quad (47)$$

which would indicate instability over the entire (x, τ) plane. On the other hand, *both* roots of the remaining factor in equation (46) are negative *if*

$$[(1+x)^2+\kappa x](1+x+\kappa) > \tau x; \quad (48)$$

i.e. if we are below the higher stability branch defined by equation (42) (in the case $\kappa_n = 0$); and one of the roots becomes positive when we are above the higher branch. The meaning of these results is the following. In the limit of zero concentration of the normal fluid, the (Rayleigh) instability of the superfluid prevails over the entire (x, τ) plane; but the normal fluid (such of it as there is!) continues to be stable as long as the Taylor number is less than the critical value appropriate to the higher branch.

We have thus shown that the principle of the exchange of stabilities obtains in the limit $\kappa_n = 0$; that all roots of σ are real and negative on the x axis; that instability sets in as we cross either branch; that the upper branch is a continuation of the (Taylor) instability of the normal fluid as inhibited by the superfluid; and finally, that the lower branch represents the (Rayleigh) instability of the superfluid inhibited by the normal fluid.

(b) The case of isotropic mutual friction

For the discussion of this case we must start from equations (91) to (94) of part I. These equations can be treated in an entirely analogous fashion. We find, for example, that equation (42) is replaced by

$$\left[\frac{\tau x}{1+x}+(1+x)\kappa_n\right]^2+2\sigma(1+x)(1+x+\sigma+\kappa)\kappa_n \\ +(1+x+\sigma+\kappa)^2\left(\sigma^2-\frac{\tau x}{1+x}\right)-\frac{\tau x}{1+x}\sigma^2=0; \qquad (49)$$

and this equation leads to the same general conclusions as equation (42).

References

Chandrasekhar, S. 1954 *Mathematika*, **1**, 5.
Chandrasekhar, S. & Donnelly, R. J. 1957 *Proc. Roy. Soc.* A, **241**, 9 (part I).

PRINTED IN GREAT BRITAIN AT THE UNIVERSITY PRESS, CAMBRIDGE
(BROOKE CRUTCHLEY, UNIVERSITY PRINTER)

Rumford Medal Lecture 1957
Thermal Convection

S. Chandrasekhar

I. INTRODUCTION

The simplest example of thermally induced convection arises when a horizontal layer of fluid is heated from below and an adverse temperature gradient is maintained. The adjective "adverse" is used to qualify the prevailing temperature gradient, since, on account of thermal expansion, the fluid at the bottom becomes lighter than the fluid at the top; and this is a top-heavy arrangement which is potentially unstable. Under these circumstances the fluid will try to redistribute itself to redress this weakness in its arrangement. This is how thermal convection originates: It represents the efforts of the fluid to restore to itself some degree of stability. These basic facts concerning thermally induced convection were discovered by Count Rumford when he observed that currents were set up in the bore of a large thermometer which he had been using in an experiment. I am indebted to Professor Sanborn Brown for the following extract from Count Rumford's writings, in which his discovery is announced:

> I saw the whole mass of the liquid in the tube in a most rapid motion running swiftly in two opposite directions, up and down at the same time. The bulb of the thermometer, which is of copper, had been made two years before I found leisure to begin my experiments, and having been left unfilled without being closed with a stopple, some fine particles of dust had found their way into it and these particles which were intimately mixed with the spirits of wine, on their being illuminated by the sun's beam, became perfectly visible . . . and by their motions discovered the violent motions by which the spirits of wine in the tube of the thermometer was agitated. . . . On examining the motion of the spirits of wine with a lens, I found that the ascending current occupied the axis of the tube and that it descended by the sides of the tube. On inclining the tube a little, the rising current moved out of the axis and occupied the side of the tube which was uppermost, while the descending currents occupied the whole of the lower side of it.

However, as everyone has known since King Alfred's time, one cannot always depend on thermally induced circulation to prevent burning at the bottom! The reason is, the natural tendency of the fluid to react instantly to its unstable arrangement is inhibited by its own viscosity, and the more viscous the fluid, the less agile is it to react to its potential instability.

The first quantitative experiments to establish the extent to which viscosity inhibits the onset of instability, and to determine the precise manner in which instability does set in, are those of Bénard at the turn of the century. Bénard worked with very thin layers of liquid (only about 1 mm. deep) standing on a leveled metallic plate which was maintained at a uniform temperature. The upper surface was usually free and, being in contact with air, was at a lower temperature. Various liquids were employed – and some, indeed, which would have been solid under ordinary conditions. Bénard's experiments established the following two fundamental facts: *First,* a certain critical temperature gradient has to be exceeded before instability can set in and, *second,* the motions which ensue on surpassing of the critical temperature gradient have a *cellular pattern.* What actually happens on the onset of instability is that the layer of liquid rapidly resolves itself into a number of cells which after a while become equal and regular and align themselves to form a beautiful hexagonal pattern. Figure 1 is a reproduction of an early photograph of Bénard's; Figure 2, which illustrates the same phenomenon by a different experimental arrangement, is taken from a paper by Schmidt and Milverton.

The correct interpretation of Bénard's experiments was given by Lord Rayleigh in 1916. Rayleigh showed that what decides the stability or otherwise of a fluid heated on the "underside" – as he expressed it – is the numerical value of the nondimensional parameter,

$$R = \frac{g\alpha\beta}{\kappa\nu} d^4 \qquad (1)$$

– now called the Rayleigh number – where g denotes the acceleration due to gravity, d the depth of the layer, $\beta = | dT / dZ |$ the constant adverse temperature gradient which is maintained, and α, κ, and ν are the coefficients of volume expansion, thermometric conductivity, and kinematic viscosity, respectively. Rayleigh showed that instability must set in when R exceeds a certain determinate critical value R_c, and that when $R = R_c$ a stationary pattern of motions will come to prevail.

If β_c denotes the temperature gradient which must be exceeded for the occurrence of convection, then

$$\beta_c = R_c \frac{\kappa\nu}{g\alpha} d^{-4} \; ; \tag{2}$$

and, as we observe, this is higher the higher the viscosity.

In this lecture I want to consider especially the effects of rotation and magnetic field, separately and jointly, on the onset of thermal instability. However, to understand fully the meaning of the predicted effects, it is necessary that we have some appreciation of the manner in which instability under the circumstances envisaged can manifest itself, and how, indeed, one can calculate the critical Rayleigh number for the onset of instability. I shall, therefore, consider these aspects of the problem first.

II. ON THE TWO WAYS IN WHICH INSTABILITY CAN SET IN: AS CONVECTION AND AS OVERSTABILITY

Suppose, then, that an initial *static* state in which a certain adverse temperature gradient is maintained is slightly disturbed. We ask: Will the disturbance gradually die down and the original state be eventually restored? Or, will the disturbance grow in amplitude in such a way that the system progressively departs from the initial state and never reverts to it? If the latter should be the case, then the state we started from is clearly an unstable one. On the other hand, we cannot positively conclude stability if the former obtains, for, in order that we may consider the initial state as stable, it is necessary that not merely a particular disturbance, but all conceivable disturbances to which the state may be subject, be damped. The criterion for instability is, then, that there exist at least *one* mode of disturbance for which the system is unstable; and the criterion for stability is that there exist *no* mode of disturbance for which it is unstable. If all initial states are classified as stable or unstable according to these criteria, then the boundary (in the set theoretical sense) between these two classes of states in the manifold of all states will define a certain *marginal state*. By this definition, a marginal state is a state of *neutral stability*. The isolation and characterization of the marginal state is clearly the prime object of an investigation on stability.

It is not necessary for our present purposes to go into the details of the mathematical processes by which one specifies the marginal state, but it is important to recognize that marginal states can be one of two distinct kinds. The two kinds correspond to the two different ways in which the amplitude of a disturbance can grow or be damped (see figure 3). Thus if $A(t)$ denotes the amplitude of a disturbance, then its dependence on time can be either of the two kinds:

$$A(t) = A_0 e^{qt}, \tag{3}$$

or

$$A(t) = A_0 e^{qt} \cos pt, \tag{4}$$

where p and q are real. In either case, we shall have stability or instability according as q is negative or positive; in the former case (equation 3) the amplitude of the disturbance will be aperiodically damped or amplified; in the latter case (equation 4) the amplitudes of certain characteristic oscillations will be damped or amplified. In both cases the marginal state will be distinguished by $q = 0$, but with this essential difference: In case 3 the marginal state will exhibit a stationary pattern of motions, while in case 4 it will exhibit oscillatory motions with a certain characteristic frequency.

Quite generally, if at the onset of instability a stationary pattern of motions prevails, then one says that the *principle of the exchange of stabilities* is valid and that instability sets in as convection. On the other hand, if instability sets in via a marginal state of purely oscillatory motions, then one says (following Eddington) that one has a case of *overstability*. The use of the term overstability in this connection is not entirely common. So it may be worth recalling Eddington's own definition of that term: "In the usual kinds of *instability* a slight displacement provokes restoring forces tending away from equilibrium; in *overstability* it provokes restoring forces so strong as to overshoot the corresponding position on the other side of equilibrium."

One final general remark: Since stability means stability with respect to all possible disturbances, it is clear that for an investigation of the stability of a system to be complete, it is necessary that the reaction of the system to all possible disturbances be examined. In practice one accomplishes this by expressing an arbitrary disturbance as a superposition of certain basic possible modes and examining the stability of the system with respect to each of these modes. Thus, in the problem of the stability of a layer of fluid heated from below, an arbitrary disturbance is expressed (in accordance with Fourier's theorem) as a superposition of two-dimensional periodic waves, and the stability is then investigated with respect to each of these waves.

III. THE SOLUTION OF THE CLASSICAL BÉNARD PROBLEM

As I have already stated, Rayleigh first gave the correct interpretation of Bénard's experiments. However, in solving the underlying mathematical problem, Rayleigh did not attempt to satisfy the correct (physical) boundary conditions. Later investigations by Jeffreys, Low, and Pellew and Southwell have corrected this deficiency in Rayleigh's original solution, so that we may now consider the problem as solved. I shall briefly describe the outcome of these investigations.

Rayleigh and, more generally, Pellew and Southwell have proved that for this problem the principle of the exchange of stabilities is valid and that, in agreement with the experiments, the onset of instability must manifest itself as a stationary pattern of motions. And the critical Rayleigh number for the onset of instability has been determined as follows:

Considering a horizontal two-dimensional periodic disturbance of an assigned wave number a ($= 2\pi/\lambda$ where λ denotes the wavelength of the disturbance) one asks: What is the lowest Rayleigh number $R(a)$ at which a mode of disturbance with the wave number a, when excited, does not get damped? On solving this problem, one finds that the resulting function $R(a)$ has a single minimum at $a = a_c$ (say) where $R = R_c$. It is clear that R_c specifies the required critical Rayleigh number for the onset of thermal convection, for if $R < R_c$, then all disturbances (expressible, as they are, as superpositions of two-dimensional waves) will be damped; and when $R = R_c$ all disturbances will again be damped *except* for a periodic disturbance with precisely the wave number a_c; this is, therefore, the disturbance which will manifest itself at marginal stability. In this manner the critical Rayleigh number for the onset of instability has been determined for the two cases of interest, namely, when the layer of fluid is confined between two rigid planes and when the layer of fluid is supported by a rigid plane and the top surface is left free. The results of the mathematical analysis are:

$$\begin{array}{lll} \text{Both surfaces rigid} & R_c = 1708, & a_c = 3.13/d\,, \\ \left.\begin{array}{l}\text{Bottom surface rigid and} \\ \text{the top surface free}\end{array}\right\} & R_c = 1100, & a_c = 2.68/d\,. \end{array} \qquad (5)$$

Several experiments have been performed to verify whether or not instability does set in at the predicted Rayleigh numbers. I shall refer to only one such set of experiments, by Schmidt and Milverton, since the principle underlying their experiments is an important one and has provided the basis for other experiments to which I shall presently refer.

In the experiments of Schmidt and Milverton the layer of fluid was confined between two rigid planes which were maintained at constant temperatures, and heat was supplied (by an electrical coil) to the bottom plate at a constant rate. The experiments consisted in determining the difference in temperature, $(T_2 - T_1)$, between the two plates for varying rates of heating of the bottom plate; a measure of the latter is provided by the square of the heating current, C^2. Experiments of this kind generally give a plot similar to the one illustrated in figure 4. It will be seen that this plot shows a distinct break for a particular $T_2 - T_1 = \Delta T_c$ (say); this determines the critical temperature gradient, $\beta = \Delta T_c/d$ for the onset of instability. For, when $T_2 - T_1 < \Delta T_c$, the relation $(T_2 - T_1, C^2)$ is linear, with a constant slope corresponding to what must be (and is verified to be) the conductive temperature gradient; at $T_2 - T_1 = \Delta T_c$, the slope of the $(T_2 - T_1, C^2)$-relation suddenly decreases, indicating that after instability a new mechanism of heat transport – namely, convective heat transport – has begun to be operative. By such experiments Schmidt and Milverton were able to confirm that the critical Rayleigh number for the onset of instability for their experimental arrangement is 1770 ± 140; this is in good agreement with the theoretical value 1708 (cf. equation 5).

IV. THE EFFECT OF ROTATION

We shall now pass on to the consideration of the case when the layer of fluid which is being heated from below is set in rotation with a constant angular velocity Ω about the vertical. The effect of the rotation is to subject the fluid to Coriolis acceleration in addition to that of gravity. And Coriolis acceleration can have a decisive effect on the onset of instability, as can be seen from the following argument:

In the absence of viscosity and temperature gradients, the equations governing the motions of the fluid are:

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u}.\, grad)\boldsymbol{u} = - grad\, P + 2\boldsymbol{u} \times \Omega \qquad (6)$$

and

$$div\, \boldsymbol{u} = 0, \qquad (7)$$

where

$$P = -\tfrac{1}{2}\, |\Omega \times \boldsymbol{r}|^2 + p/\rho \qquad (8)$$

In the foregoing equations $\boldsymbol{u}$ denotes the velocity, p the pressure and ρ the density.

If the state of motions is stationary, and the velocities are sufficiently small for the nonlinear terms in equation 6 to be negligible, then

$$grad\, P = 2\boldsymbol{u} \times \Omega \qquad (9)$$

From this equation it follows that

$$curl\, (\boldsymbol{u} \times \Omega) = 0. \qquad (10)$$

When Ω is assumed to be in the direction of the z-axis, the three components of the single vector equation (10) are

$$\frac{\partial u_x}{\partial z} = 0, \; \frac{\partial u_y}{\partial z} = 0 \text{ and } \frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} = - \frac{\partial u_z}{\partial z} = 0. \qquad (11)$$

Hence, in the absence of viscosity, for sufficiently slow motions, the velocity components cannot depend on z. This result is a special case of a general theorem due to J. Proudman and G I. Taylor that *all slow motions in a rotating inviscid fluid are necessarily two dimensional.* This theorem has an important bearing on the ensuing of thermal convection in a rotating fluid.* For, convection implies that the motions have a three-dimensional character; and this the Taylor-Proudman theorem forbids for an inviscid fluid so long as the nonlinear terms in the equations of motion are neglected. Accordingly, in contrast to the case of a nonrotating fluid, an inviscid fluid in rotation is thermally stable for *all* adverse temperature gradients.

*I am indebted to Dr. Raymond Hide for pointing out to me the relevance of the Taylor-Proudman theorem for these considerations.

Indeed, only in the presence of viscosity can instability arise, for only then can the Taylor-Proudman theorem be violated.

It is clear, then, that the effect of rotation will be to inhibit the onset of convection. More precisely, it follows from a theoretical analysis of this problem that the extent of the inhibition depends on Ω through the nondimensional parameter

$$T = \frac{4\Omega^2}{\nu^2} d^4 ; \qquad (12)$$

this has now come to be called the Taylor number.

A further important difference between the problem with and without rotation is the following: While in the problem without rotation the principle of the exchange of stabilities is always valid, this is no longer the case when Coriolis forces are acting. The discriminating parameter in this connection is the ratio of the kinematic viscosity (ν) to the thermometric conductivity (κ); this is sometimes called the Prandtl number and denoted by

$$\omega = \nu/\kappa . \qquad (13)$$

From a theoretical analysis of the problem it follows that if ω is less than a certain critical value ω^* (say), then for all values of T greater than a certain determinate T^* (depending on ω), the mode of instability which should set in first is overstability and not convection. But if $\omega > \omega^*$, the principle of the exchange of stabilities obtains and instability should set in as ordinary cellular convection. The type of results to be expected under these circumstances is shown in figure 5. The precise value of ω^* depends on the boundary conditions that have to be satisfied on the confining boundaries (such as whether they are rigid or free). For the (nonphysical) case of two free boundaries $\omega^* = 0.677$; the exact value of ω^* for other more realistic boundary conditions has not been determined; but it is known that in all cases it is not very different from unity.

It is remarkable that the Prandtl number should play this decisive role in determining the manner of the onset of thermal instability in rotating fluids; and that overstability should be the rule for $\omega \ll 1$, as is the case with metallic liquids such as mercury.

(*a*) *Theoretical Predictions*

The results of the theoretical calculations on the dependence of R_c on T are shown in figure 6. The cases when instability sets in as a stationary pattern of convection as well as when it sets in via a state of oscillatory motions are both illustrated; in the latter case the curves refer to a value of $\omega = 0.025$ which is appropriate for mercury at room temperatures. The asymptotic behaviors of these dependences for $T \to \infty$ may be noted.

In case the instability sets in as convection, the dependence of the critical Rayleigh number for the onset of instability on the Taylor number T has the asymptotic behavior,

$$R_c \to \text{constant} \times T^{\frac{2}{3}} \qquad (T \to \infty); \tag{14}$$

while the wave number a_c (in units of $1/d$) of the disturbance which manifests itself at marginal stability increases with T according to the law

$$a_c \to \text{constant} \times T^{\frac{1}{6}} \qquad (T \to \infty). \tag{15}$$

These relations apply so long as ω is greater than a certain critical value ω^*. If the Prandtl number is less than this critical value, then for all T greater than a certain determinate value, instability will set in as overstability; and for sufficiently large T, we have the asymptotic relations:

$$\left.\begin{aligned} R_c &\to \text{constant} \times \frac{\omega^{\frac{4}{3}}}{(1+\omega)^{\frac{1}{3}}} T^{\frac{2}{3}}, \\ a_c &\to \text{constant} \times \left(\frac{\omega}{1+\omega}\right)^{\frac{1}{3}} T^{\frac{1}{6}}. \end{aligned}\right\} \quad (T \to \infty). \tag{16}$$

Also, the frequency, p, of the characteristic oscillations at marginal stability has the behavior,

$$p/\Omega \to \text{constant} \times T^{-\frac{1}{6}} \qquad (T \to \infty). \tag{17}$$

(*b*) *Experimental Verifications*

The various predictions described in the preceding paragraph and exhibited in figure 6 have been fully confirmed by some very beautiful experiments of Nakagawa and Frenzen and of Fultz and Nakagawa at the University of Chicago.

First, the experiments of Nakagawa and Frenzen with water (which has a Prandtl number $\omega = 6.0$) fully confirm that in this case, thermal instability does indeed set in as cellular convection (see figure 7) and that the $T^{\frac{2}{3}}$-law is obeyed (see figure 8). On the other hand, the experiments of Fultz and Nakagawa with mercury (which has a Prandtl number $\omega = 0.025$) show that in this instance, in agreement with theoretical prediction, instability does set in as overstability. This is shown in a particularly striking manner by the temperature records (see figures 9 and 10 in which the temperature records obtained in the experiments with water and mercury are contrasted). Further, the experimentally determined (R_c, T)-dependence is in accord with the theoretical relation (see figure 8). And, finally, the frequencies of the characteristic oscillations at marginal stability (as determined from temperature records such as those shown in figure 10) are also in very good agreement with the predicted values (see figure 11).

V. THE EFFECT OF A MAGNETIC FIELD

We now turn to the effect of an impressed magnetic field on thermal instability. We suppose that the fluid under consideration is an electrical conductor (such as mercury) and that an external magnetic field ($\boldsymbol{H}$) is impressed in a direction parallel to gravity, $\boldsymbol{g}$. (The case when the directions of $\boldsymbol{H}$ and $\boldsymbol{g}$ are not parallel has also been treated; but we shall not consider it here.)

On general grounds we may expect that the effect of a magnetic field will also be to inhibit the onset of convection and that the inhibiting effect will be the greater the stronger the magnetic field (H), and the higher the electrical conductivity (σ): for, when the field is strong (or the conductivity high) the lines of magnetic force tend to be glued to the material and this will make motions at right angles to the field difficult. In this latter respect the inhibiting effect of a magnetic field has a different origin from that of rotation: for, while in the absence of viscosity motions parallel to the axis of rotation are forbidden (in accordance with the Taylor-Proudman theorem), in the absence of electrical resistivity motions perpendicular to the field are forbidden.

As to whether instability will set in as convection or as overstability, it can be shown that with liquid metals such as mercury overstability cannot arise. Moreover, when cellular convection sets in, we must in accordance with what we have stated, expect that the cells become progressively elongated as the strength of the magnetic field is increased. And in the limit of infinite electrical conductivity (or, infinite field strength) when the fluid elements are

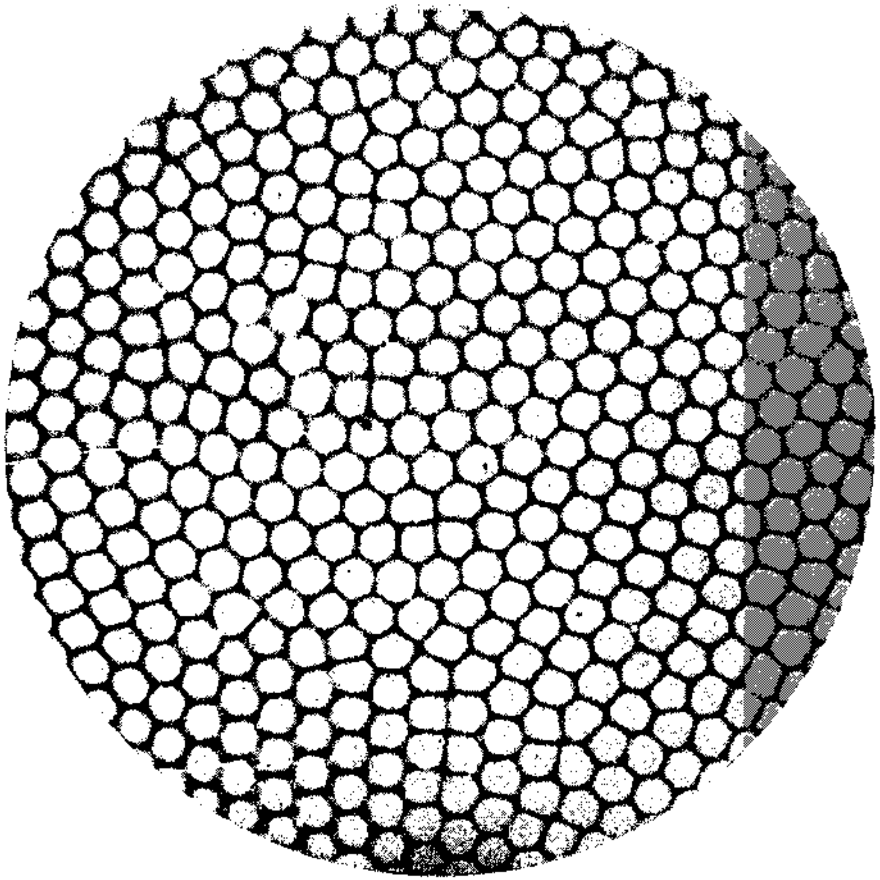


FIGURE 1. Bénard cells in spermaceti. A reproduction of one of Bénard's original photographs.

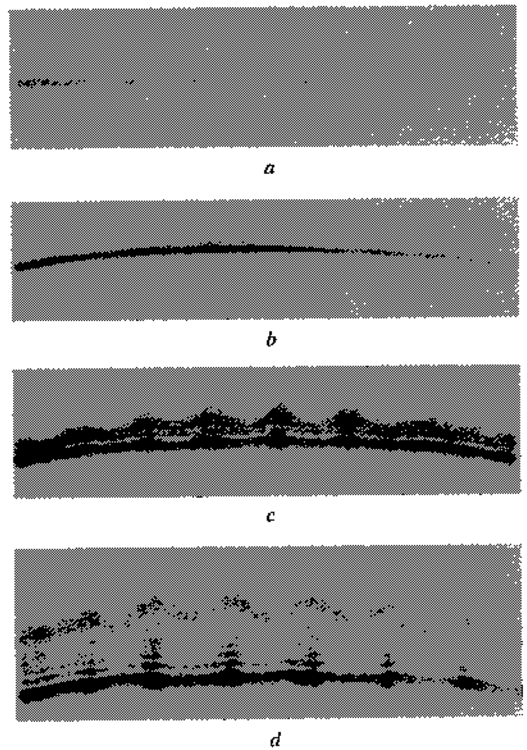


FIGURE 2. The onset of thermal instability as photographed by Schmidt and Milverton using an optical arrangement. (From *Proc. Roy. Soc.*, London, A, 152, 586, 1935.)

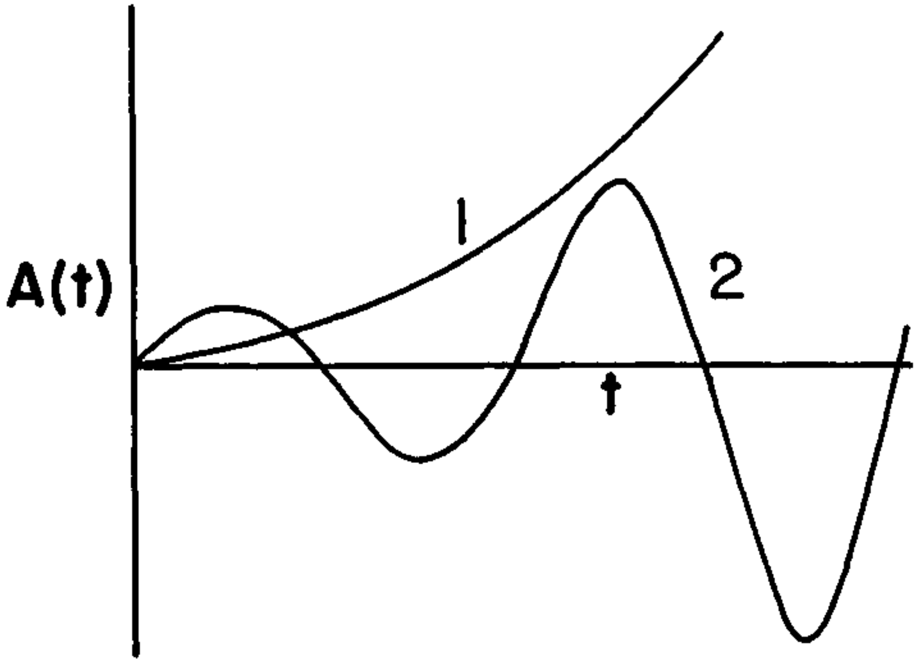


FIGURE 3. Aperiodic instability (curve 1) and overstability (curve 2).

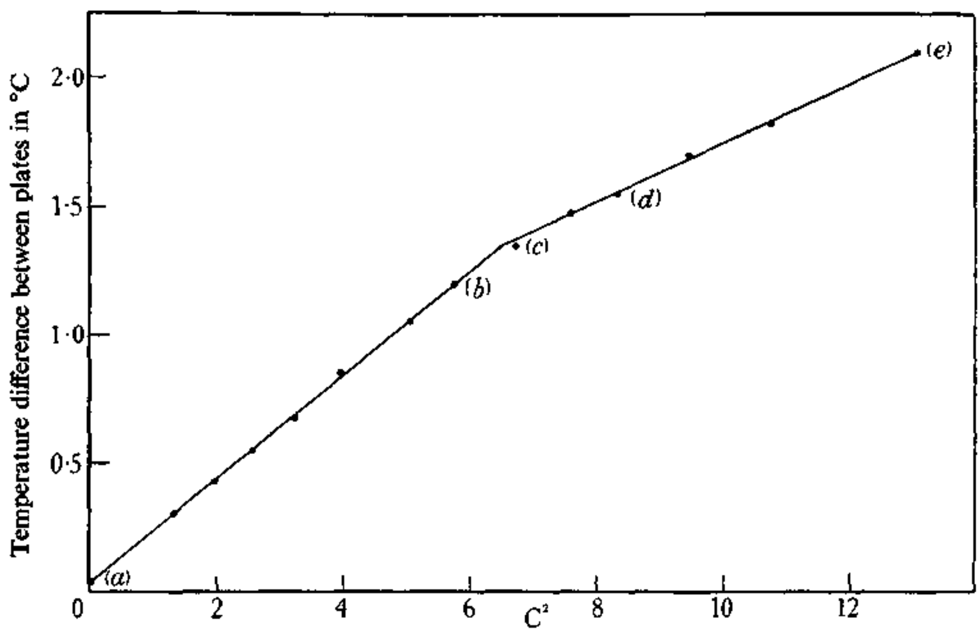


FIGURE 4. Curve showing the discontinuity in the rate of heat transfer in a layer of fluid between two horizontal plates when heated from below. (R. J. Schmidt and S. W. Milverton, *Proc. Roy. Soc.*, London, A, 152, 586, 1935.)

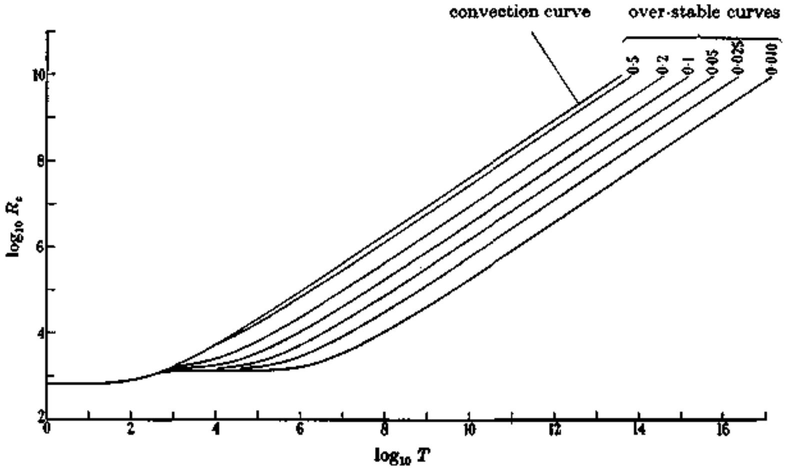


FIGURE 5. The (R_c,T)-relations for a rotating horizontal layer of fluid heated below. The curves have been derived for the case when both bounding surfaces are free. The curve labelled " convection curve " is the (R_c,T)-relation for the onset of ordinary cellular convection. The remaining curves are the corresponding relations for the onset of overstability. The value of ω to which the various curves refer are shown at the top of each curve. It will be seen that for each value of $\omega < 0.677$, the instability sets in as ordinary cellular convection for T less than a certain T^* while it sets in as overstability for $T > T^*$. (From *Proc. Roy. Soc.*, London, A, 231, 198, 1955.)

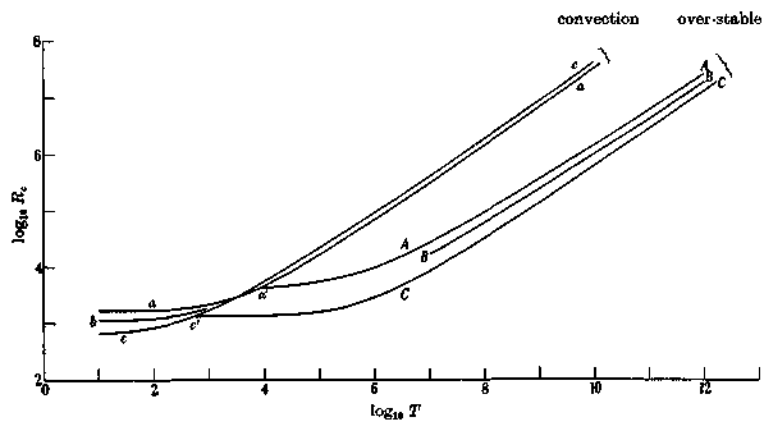


FIGURE 6. The (R_c,T)-relations for the three cases (a) both bounding surfaces rigid, (b) one bounding surface rigid and the other free, and (c) both bounding surfaces free. The curves labelled *aa*, *b*, and *cc* are the relations for the onset of ordinary cellular convection for the three cases, respectively. The curves labelled *a'AA*, *BB*, and *c'CC* are the corresponding relations for the onset of overstability for $\omega = 0.025$. At *a'* (respectively *c'*) we have a change from one type of instability to another as T increases. (From *Proc. Roy. Soc.*, London, A, 231, 198, 1955.)

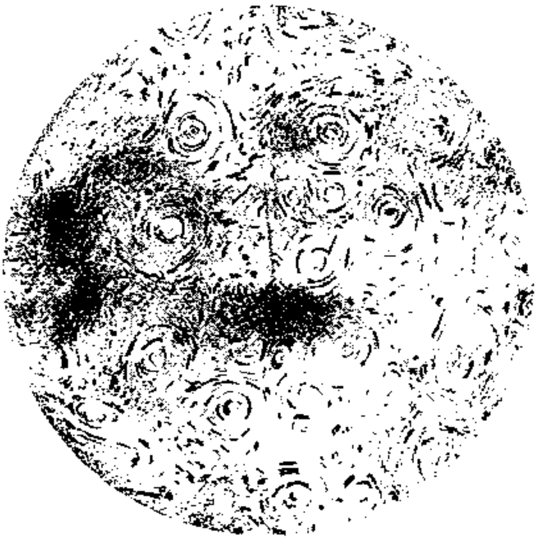


FIGURE 7. Convection cells which appear in water when in rotation and heated from below: depth 18 cm.; difference in temperature 0.7°; rate of rotation 5.0 rpm; Taylor number 1.2×10^9. (Y. Nakagawa and P. Frenzen, *Tellus*, 7, 1, 1955.)

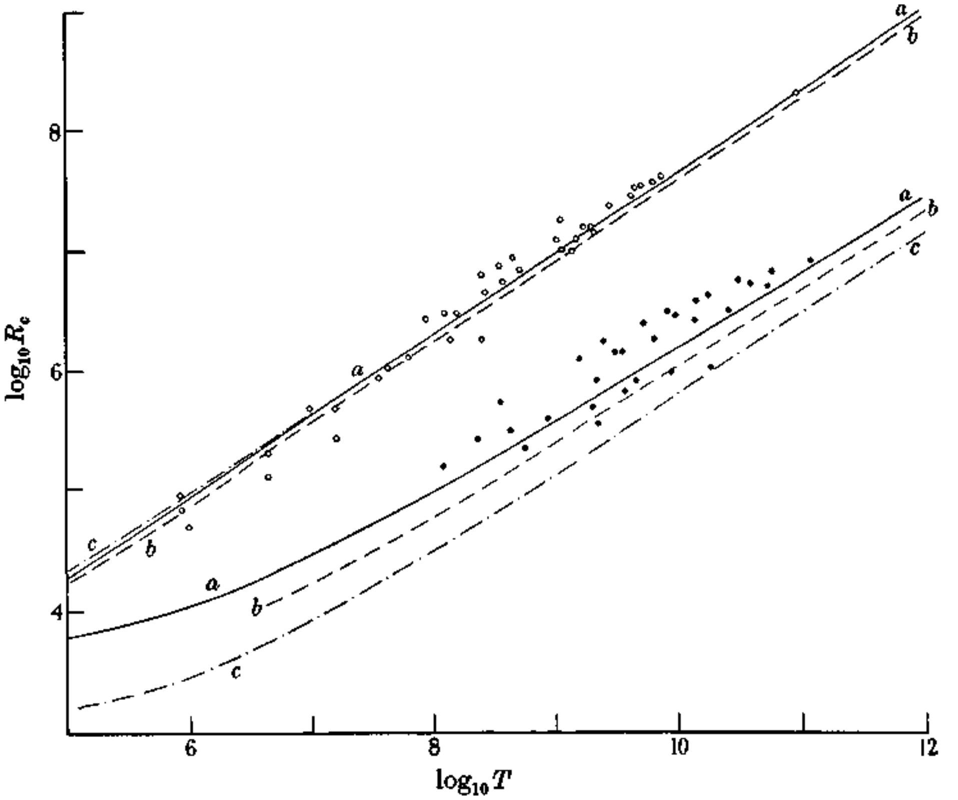


FIGURE 8. A summary of the experimental and the theoretical results. The curves *aa, bb,* and *cc* are the theoretical (R_c, T)-relations derived for the three cases (a) both bounding surfaces rigid, (b) one bounding surface rigid and the other free, and (c) both bounding surfaces free. The upper group of curves are for the onset of instability as ordinary cellular convection. The lower group of curves are for the onset of overstability for a value of the Prandtl number $\omega = 0.025$. The open circles are the experimentally determined points for water ($\omega = 6$). The solid circles are the experimentally determined points for mercury ($\omega = 0.025$). (D. Fultz and Y. Nakagawa, *Proc. Roy. Soc.*, London, A, 231, 211, 1955.)

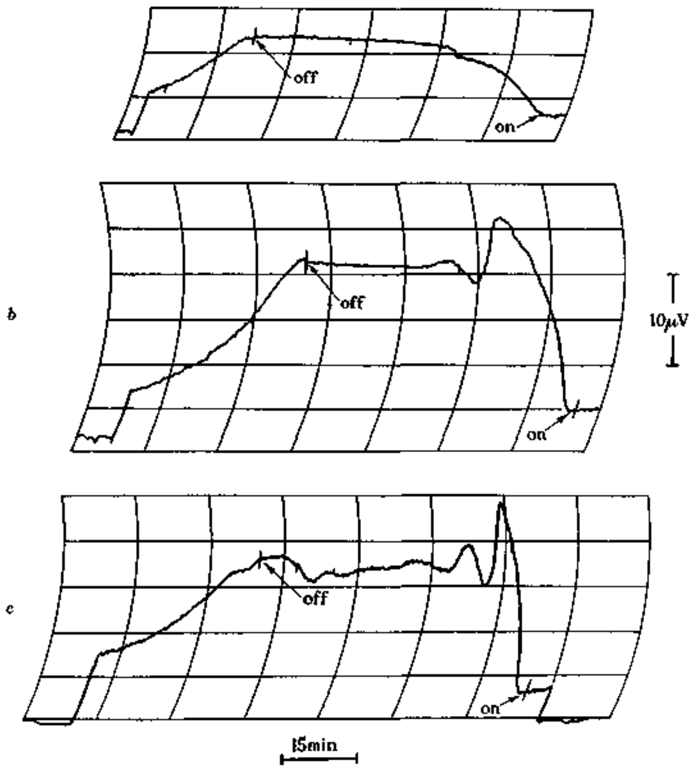


FIGURE 9. A time record of the adverse temperature gradient for water for three different rates of heating; $d = 3$ cm., $\Omega = 10$ rpm. (D. Fultz and Y. Nakagawa, *Proc. Roy. Soc.*, London, A, 231, 211, 1955.)

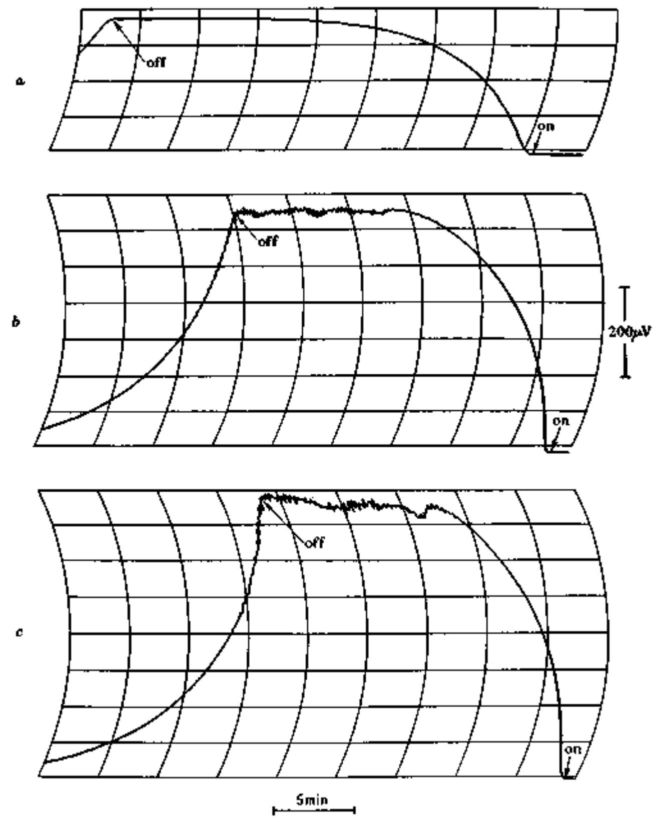


FIGURE 10. A time record of the adverse temperature gradient for mercury for three different rates of heating; $d = 6$ cm., $\Omega = 15$ rpm. (D. Fultz and Y. Nakagawa, *Proc. Roy. Soc.*, London, A, 231, 211, 1955.)

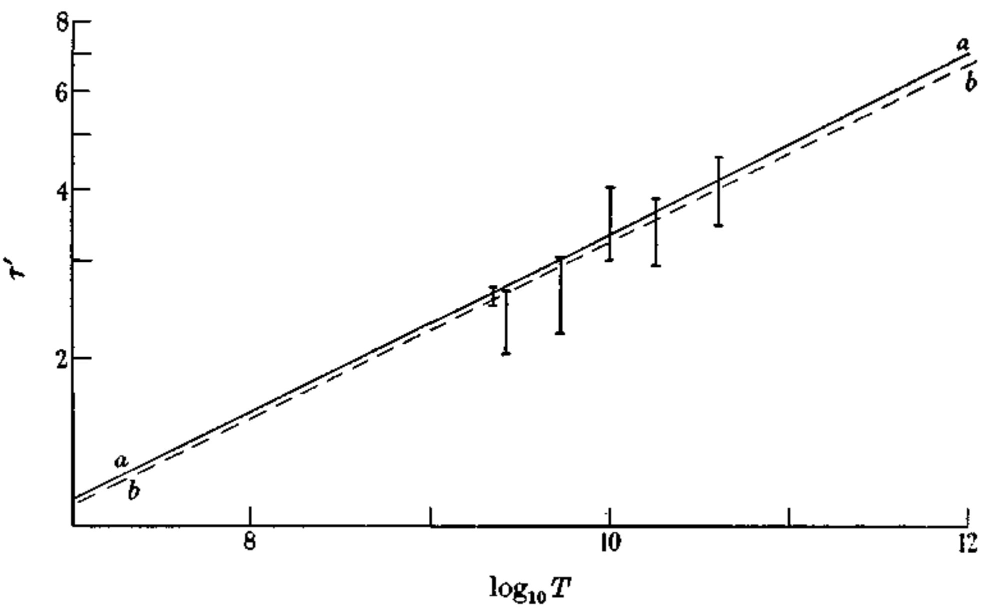


FIGURE 11. A comparison of the observed periods of oscillation at marginal stability with the theoretical periods. The ordinate (τ') gives the period in units of $2\pi/\Omega$. The curves *aa* and *bb* are the theoretical relations derived for $\omega = 0.025$ and for the case of two bounding surfaces rigid (*aa*) and one bounding surface rigid and the other free (*bb*). (D. Fultz and Y. Nakagawa, *Proc. Roy. Soc.*, London, A, 231, 211, 1955.)

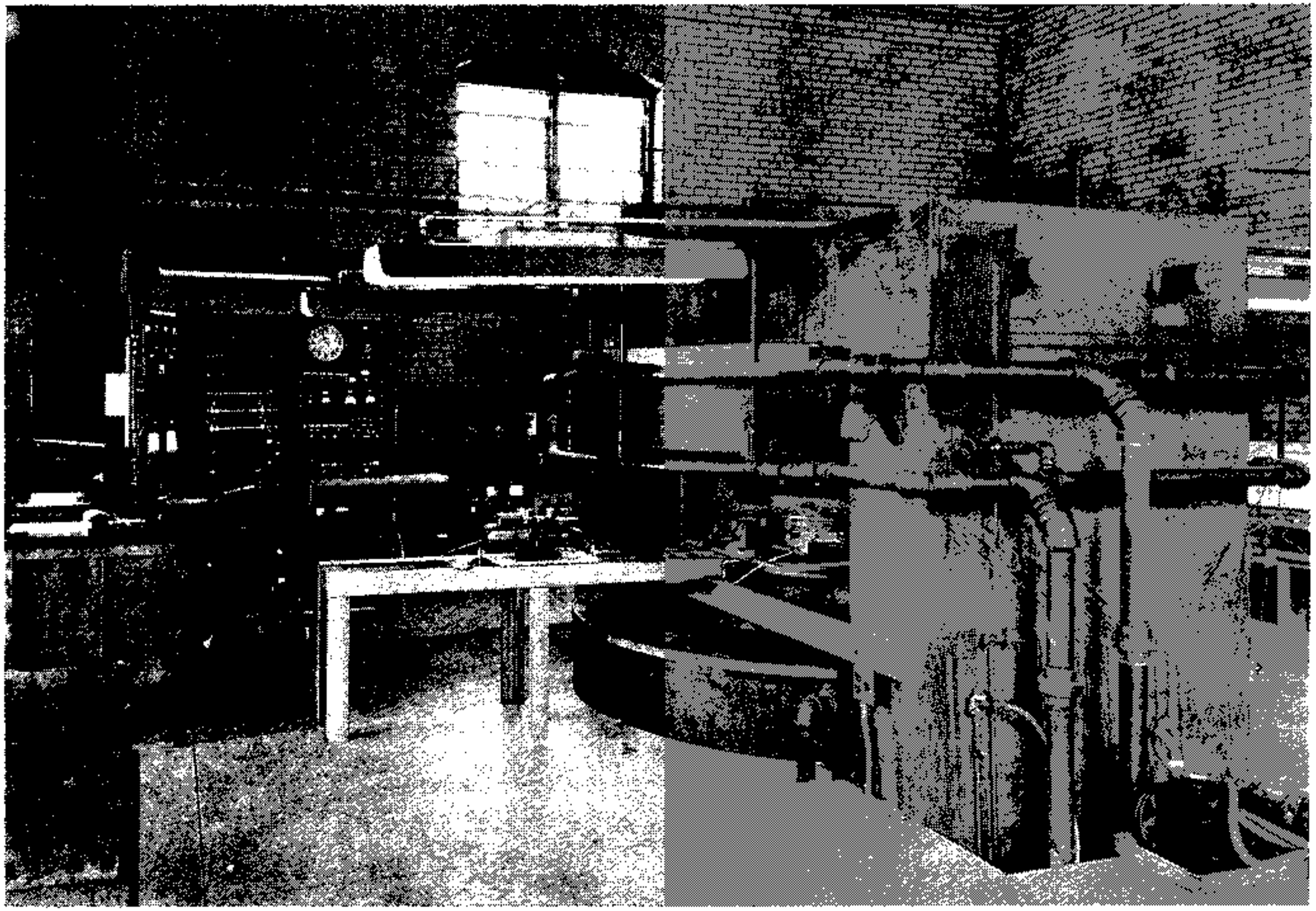


FIGURE 12. The hydromagnetic laboratory at the Enrico Fermi Institute for Nuclear Studies of the University of Chicago.

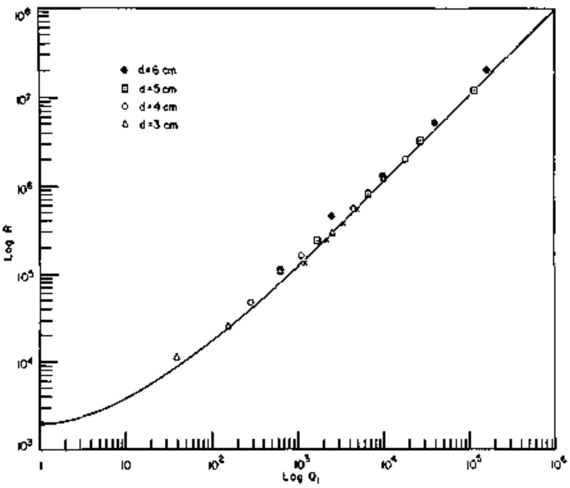


FIGURE 13. A comparison of the experimental and the theoretical results. The theoretical (R_c,Q_1)-relation is shown by the full line curve. The solid circles, squares, open circles, and triangles are the experimentally determined points for $d = 6$, 5, 4 and 3cm., respectively, with the 36½ inch magnet; the triangles represent the results with a smaller magnet with $H = 1500$ gauss and $d = 6$, 5, 4 and 3 cm. (Y. Nakagawa, *Proc. Roy. Soc.*, London, A, in press.)

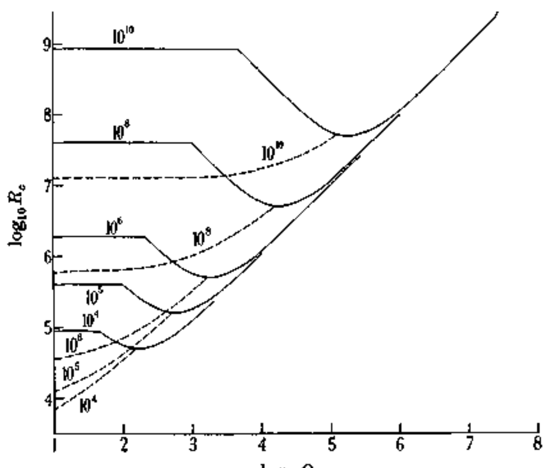


FIGURE 14. The critical Rayleigh number (R_c) for the onset of ordinary cellular convection (solid line) and overstability (for $\omega = 0.025$) (broken line) as a function of Q_1 $(= \mu^2 d^2 H^2 \sigma/\pi^2 \rho\nu)$ for various assigned values of T_1 $(= 4\Omega^2 d^4/\pi^4\nu^2)$. The curves are labelled by the values of T_1 to which they refer. For a given value of T_1, instability will set in as overstability for all values of Q_1 less than that at the point of intersection of the corresponding full line and dashed curves; for all larger values of Q_1, it will set in as ordinary convection. (*Proc. Roy. Soc.*, London, A, 237, 476, 1956.)

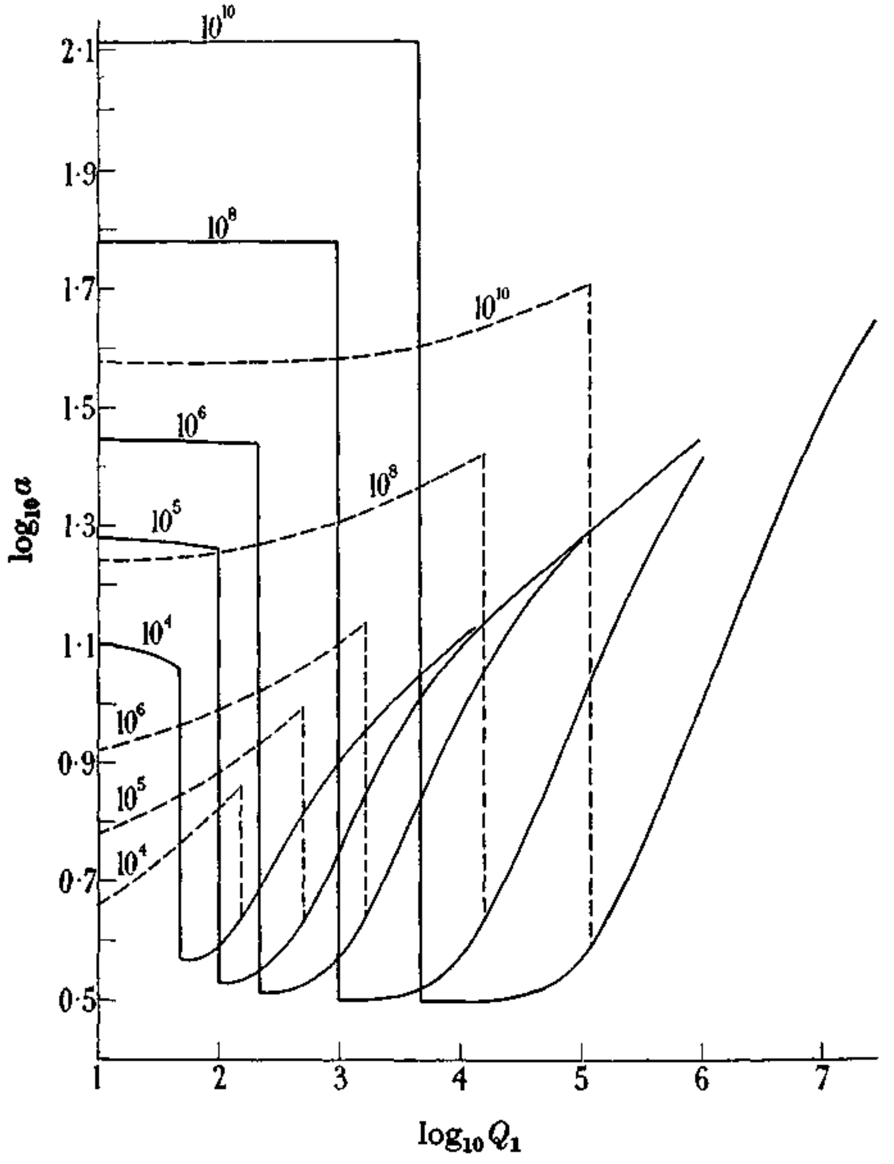


FIGURE 15. The dependence on Q_1 (for various assigned values of T_1) of the wave number a (in the unit $1/d$) of the disturbance at which instability first sets in as convection (solid line) and as overstability (for $\omega = 0.025$) (broken line). It will be observed that a discontinuous change in a occurs when (for increasing Q_1) the manner of instability changes from overstability to cellular convection. (*Proc. Roy. Soc.*, London, A, 237, 476, 1956.)

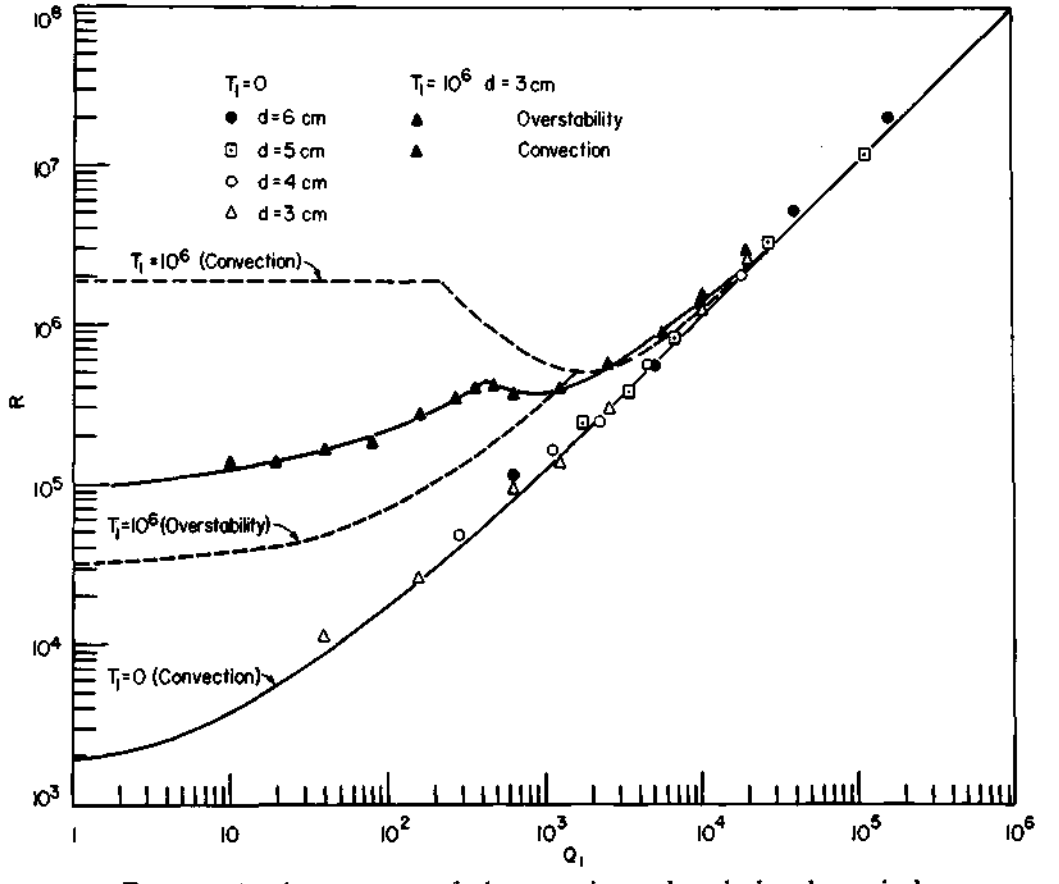


FIGURE 16. A summary of the experimental and the theoretical results. The curves labelled $T_1 = 10^6$ (convection), $T_1 = 10^6$ (overstability) and $T_1 = 0$ (convection) represent the theoretically derived relations. The value of T_1 appropriate for the experimental points represented by the solid and shaded triangles is 7.75×10^5.

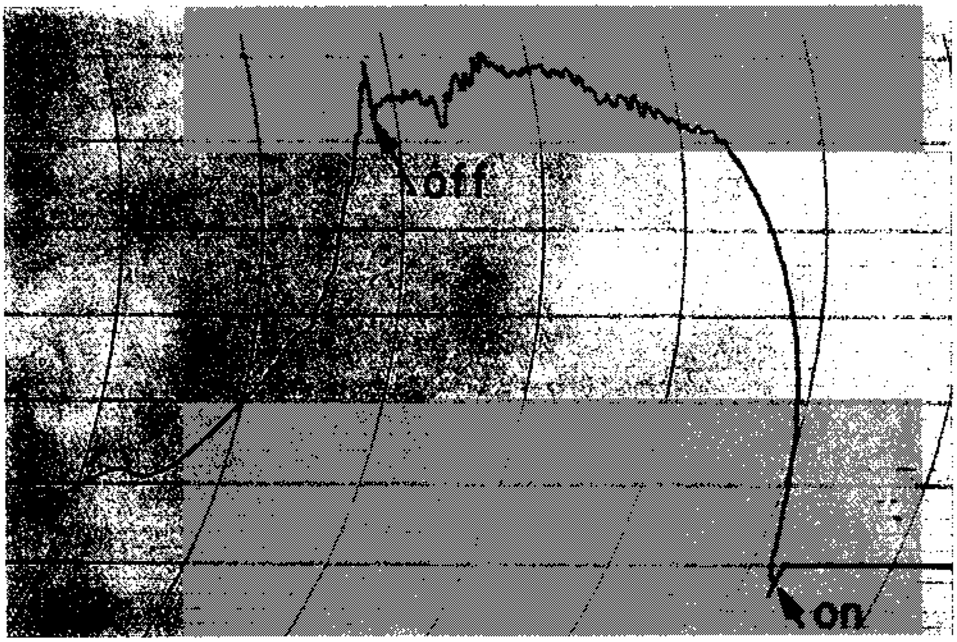


FIGURE 17. A time record of the adverse temperature gradient for mercury: $d = 3$ cm., $\Omega = 5$ rpm, $H = 125$ gauss, $Q_1 = 1.01 \times 10^1$, $T_1 = 7.90 \times 10^5$.

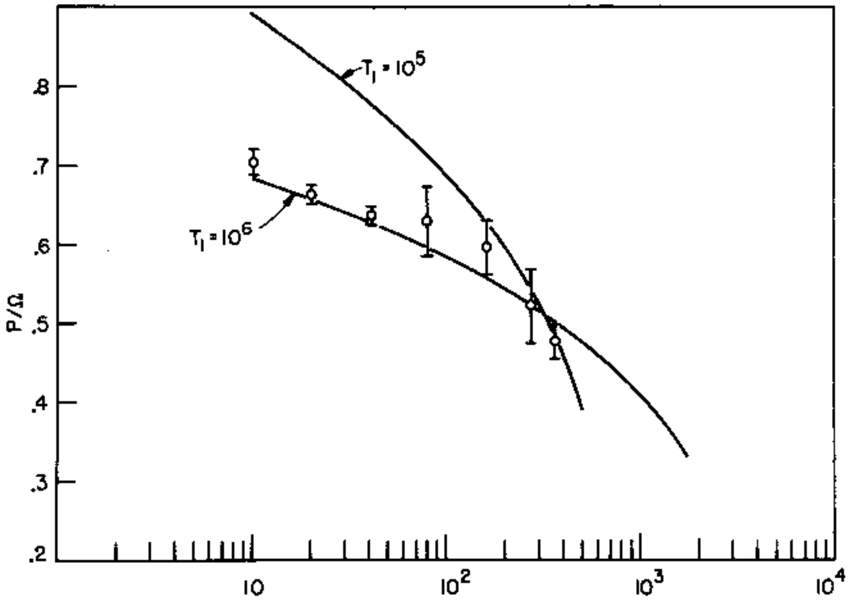


FIGURE 18. A comparison of the observed periods of overstable oscillations with the theoretical values (the full line curves). The value of T_1 appropriate for the experimental results is $(8.05 \pm 0.07) \times 10^5$.

permanently attached to the lines of force and motions parallel to the field are the only ones permissible, convection in the usual sense cannot occur.

(a) Theoretical Predictions

A detailed theoretical treatment of the inhibition of thermal convection by a magnetic field confirms one's general expectations. Precisely, it is found that the critical Rayleigh number, R_c, for the onset of instability depends on the strength of the magnetic field (H) and the electrical conductivity (σ) through the nondimensional parameter

$$Q = \frac{\mu^2 H^2 \sigma}{\rho \nu} d^2, \tag{18}$$

where μ denotes the magnetic permeability (see figure 13). In particular it is found that the dependence of R_c on Q has the asymptotic behavior

$$R_c \to \pi^2 Q \qquad (Q \to \infty), \tag{19}$$

independently of the boundary conditions; and further that the wave number a_c (in units of $1/d$) of the disturbance which manifests itself at marginal stability has the behavior

$$a_c \to \pi^{\frac{2}{3}} (\tfrac{1}{2} Q)^{\frac{1}{6}} \qquad (Q \to \infty), \tag{20}$$

again, independently of the boundary conditions.

If we insert in the asymptotic relation (19) the expressions for R_c and Q given by equations 2 and 18, we obtain

$$\beta_c = \pi^2 \frac{\mu^2 H^2 \sigma \kappa}{g \alpha} d^{-2} \qquad (H^2 \sigma \to \infty): \tag{21}$$

a formula for the critical temperature gradient in which the viscosity no longer enters. The physical meaning of this independence on viscosity is simply that as the strength of the impressed magnetic field is increased, ohmic dissipation – rather than viscous dissipation – becomes the principal factor in arresting incipient convection. Indeed, if we consider an *inviscid* fluid with a finite electrical conductivity, then we shall find that in the presence of an external magnetic field, the critical temperature gradient for the onset of instability is given precisely by equation 21. The presence of a magnetic field, therefore, imparts to the fluid characteristics which we normally associate with viscosity; we may, if we like, even define an effective kinematic viscosity

$$\nu_{eff} = \frac{\mu^2 H^2 \sigma}{\rho} d^2, \tag{22}$$

for motions perpendicular to the field.

(b) Experimental Verifications

Experiments on the inhibition of thermal convection by a magnetic field have been carried out by Nakagawa at the University of Chicago. For these and similar hydromagnetic experiments, the electromagnet of a (discarded!) 36½ inch cyclotron has been reconditioned at the Enrico Fermi Institute for Nuclear Studies at the university (see figure 12). This magnet provides a uniform magnetic field in a cylindrical volume 78 cm. in diameter and 22 cm. in height; and the strength of the field can be varied up to a maximum of 13,000 gauss. By using layers of mercury of depth 3 to 6 cm. and magnetic fields of strength 500 to 8,000 gauss, Nakagawa has determined the dependence of the critical Rayleigh number for the onset of instability on the parameter Q. The results of his experiments together with the theoretically predicted (R_c, Q)-relation are shown in figure 13. It will be seen that the experiments fully confirm the theoretical predictions.

(c) The Magnetic Inhibition of Convection in Sunspots

It is now generally believed that the lower temperatures of the sunspots relative to the surrounding photosphere is due to the absence of a convection zone in the sunspots at depths where convection is operative in the photospheric layers; and that this absence of convection in the sunspots is due to its having been suppressed by the prevailing magnetic fields. This suggestion was originally made by Walén and Biermann.

VI. THE EFFECT OF ROTATION AND MAGNETIC FIELD

We now consider the effect on thermal instability of a magnetic field and rotation acting simultaneously. From the physical interpretation which we have given of the effects which arise when rotation alone or magnetic field alone is present, it is clear that when they are both present, the fluid must be subject to conflicting tendencies.

For, while both have inhibitive effects on the onset of instability, the origin in the two cases is different. As we have explained, the inhibitive effect of rotation arises from the fact that in the absence of viscosity motions parallel to the axis of rotation are forbidden; while the inhibitive effect of a magnetic field arises from the fact that in the absence of electrical resistivity motions perpendicular to the field are forbidden. Consequently, when both rotation and magnetic field are present, motions compatible with the one are incompatible with the other. The fluid has therefore to obey two masters with contrary inclinations; and we may conclude that the fluid is in danger of a nervous breakdown! To express the same thing less anthropomorphically: The behavior of the fluid must be characterized by discontinuities indicating which of the two laws – the conservation of vorticity or the conservation of magnetic flux – is violated at the expense of the other (by virtue of viscous or ohmic dissipation).

An alternative way of looking at the problem is the following: We have seen that in the presence of rotation viscosity facilitates the onset of instability; we have also seen that a magnetic field imparts to the fluid characteristics associated with viscosity. Consequently, even though each acting separately inhibits the onset of convection, they will not necessarily conspire toward the same ends when they act together; indeed, we may expect that under certain circumstances the effect of a magnetic field on a rotating fluid may actually be one of accelerating the onset of convection.

We may look at the problem in still another way: In the absence of viscosity we have the theorem of Helmholtz and Kelvin that the normal flux of the vorticity ($\boldsymbol{\omega} = \text{curl } \boldsymbol{u}$) across any element of surface is an integral of the equations of motion; i.e.

$$\frac{d}{dt}\int_S \boldsymbol{\omega} \, . \, d\boldsymbol{S} = 0 \, , \tag{23}$$

where the integral is over an element of surface S and we follow the element with the motion. At the same time, in the absence of electrical resistivity we have the theorem of Cowling that the normal flux of the magnetic field ($\boldsymbol{H}$) across any element of surface is also an integral of the equations of motion; i.e.

$$\frac{d}{dt}\int_S \boldsymbol{H} . \, d\boldsymbol{S} = 0 \, . \tag{24}$$

It is the existence of these two integrals simultaneously that is at the base of the complex behavior of a real fluid (i.e. one with finite viscosity and finite electrical resistivity) when circulating and in the presence of a magnetic field.

(*a*) *Theoretical Predictions*

A complete theoretical discussion of the thermal instability of a layer of fluid such as mercury, in rotation and in the presence of an impressed magnetic field, is further complicated by the fact that the onset of instability in the presence of rotation is generally as overstability while it is as convection in the presence of a magnetic field. While the general solution of this complex problem taking correctly into account all the boundary conditions has not been carried out, the analogue of the case of two free boundaries (first considered by Rayleigh in his solution of the classical problem) has been fully analyzed and the results of the analysis are summarized in figures 14 and 15.

Considering first the case when instability sets in as convection, we observe that the following sequence of events will occur if we gradually increase the strength of the magnetic field keeping the angular velocity of rotation constant. Suppose for example that T_1 $(= T/\pi^4) = 10^6$ and we gradually increase the magnetic field; i.e., Q_1 $(= Q/\pi^2)$. When $Q_1 = 0$ (and $T_1 = 10^6$) the cells which appear at marginal stability will be elongated ($a/d = 18.9$); and as Q_1 is increased, at first, the critical Rayleigh number and the dimension of the marginal cells will be hardly affected. However, when the magnetic field has increased to a value corresponding to $Q_1 = 100$, cells of two very different sizes will suddenly appear simultaneously: one set, which will be highly elongated and the other set, which will be relatively quite wide; in fact, in this particular case the ratio of the two cell diameters is about 6. As the magnetic field increases beyond this value, the critical Rayleigh number will start decreasing and pass through a minimum; and eventually the effect of the magnetic field will predominate.

The sequence of events we have described will occur if the Prandtl number ($\omega = \nu/\kappa$) exceeds a certain value of the order of one. But if one considers a fluid such as mercury with a low Prandtl number, the occurrence of overstability (when the effects of rotation dominate) alters the situation. The sequence of events we may then observe can be directly read from figures 14 and 15. It is seen for example from figure 14 that the transition from overstability to convection occurs at about the place where the convection curve passes through its minimum; and from figure 15 it is seen that this transition is accompanied by a substantial discontinuity in the wave number of the disturbance which manifests itself at marginal stability. This latter discontinuity is in the sense that the convection cells (for increasing Q_1) suddenly become very much widened; under experimentally realizable conditions this widening can be by a factor as large as 10.

(*b*) *Experimental Verifications*

Again at the Fermi Institute, Nakagawa has carried out experiments on the thermal instability of rotating layers of mercury in a homogeneous external magnetic field. The results of his experiments are presented in figures 16 to 18.

Nakagawa was able to distinguish whether the onset of instability is as overstability or as convection by examining the temperature records. Figure 16 is one of these temperature records; it shows in a very striking manner the onset of overstability. The experimentally determined critical Rayleigh numbers for $T_1 = 7.75 \times 10^5$ and varying values of Q_1 are exhibited in figure 17 together with the theoretically expected relations for $T_1 = 0$ and $T_1 = 10^6$. It is seen that the transition from overstability to convection does occur, as predicted, at a certain critical field strength. However, the quantitative agreement between theory and experiment in this case is not as good as in the others; but this was not to be expected, since, in the solution of the theoretical problem in this case, the correct boundary conditions were not satisfied. Nevertheless, the experiments do confirm the broad features of the predicted variations; the agreement, in particular, between the observed frequencies of the oscillations at marginal stability and the predicted values is very satisfactory.

* * * * *

That the nonlinearity of the hydrodynamical equations leads to a complex and often unexpected behavior of a fluid is well known. If the fluid should in addition be an electrical conductor and magnetic fields are present, then its behavior contains even greater elements for surprise. It is this unexpectedness which makes hydromagnetics a fascinating field for study; but it also calls for a disciplined and a systematic examination.

In our efforts to understand why a fluid behaves in a particular way under particular conditions the study of hydrodynamic and hydromagnetic stability is likely to be useful, for it may disclose the relative importance and sometimes the conflicting tendencies of the different forces and constraints which are kept dormant in the undisturbed state. Thus, the study of thermal stability in the context of the simplest problem already reveals several phases of fluid behavior which, in these connections, go to make for its complexity. A similar examination of other situations may give us further insight. To the critic who rightly points out that in these studies we are very remote from the conditions under which Nature presents her most extravagant displays, let me only say:

> There is a square. There is an oblong. The players take the square and place it upon the oblong. They place it very accurately. They make a perfect dwelling place. The structure is now visible. What was inchoate is here stated. We are not so various or so mean. We have made oblongs and stood them upon squares. This is our triumph. This is our consolation.
>
> —Virginia Woolf

References

The following list, while very incomplete as a bibliography, represents, however, the principal source for the lecture.

On the Classical Problem

1. Rumford, *Complete Works* (Boston: American Academy of Arts and Sciences, 1870) I, 237. et. seq.
2. Bénard, H., 1900, *R., gen. Sci. por. appl.* 12, 1261, 1309.
3. Rayleigh, 1916, *Phil. Mag.* 6, 32, 529.
4. Jeffreys, H., 1926, *Phil. Mag.* 7, 2, 833.
5. Low, A. R., 1929, *Proc. Roy. Soc.* A, 125, 180.
6. Pellew, A., and Southwell, R. V., 1940 *Proc. Roy Soc.* A, 176, 312.
7. Schmidt, R. J., and Milverton, S. W., 1935, *Proc. Roy. Soc.* A, 152, 586.

On the Effect of Rotation

1. Chandrasekhar, S., 1953, *Proc. Roy. Soc.* A, 217, 306.
2. Chandrasekhar, S., and Elbert, D. D., 1955, *Proc. Roy. Soc.* A, 231, 198.
3. Nakagawa, Y., and Frenzen, P., 1955, *Tellus*, 7, 1.
4. Fultz, D., and Nakagawa, Y., 1955, *Proc. Roy. Soc.* A, 231, 211.

On the Effect of a Magnetic Field

1. Chandrasekhar, S., 1952, *Phil. Mag.* 7, 43, 501.
2. Chandrasekhar, S., 1954, *Phil. Mag.* 7, 45, 1177.

3. Thompson, W. B., 1951, *Phil. Mag.* 7, 42, 1417.
4. Nakagawa, Y., 1957, *Proc. Roy. Soc.* A, 240, 108.

On the Effect of Rotation and Magnetic Field

1. Chandrasekhar, S., 1954, *Proc. Roy. Soc.* A, 225, 173.
2. Chandrasekhar, S., 1956, *Proc. Roy. Soc.* A, 237, 476.
3. Nakagawa, Y., 1957, *Proc. Roy. Soc.* A (in press).

THE THERMODYNAMICS OF THERMAL INSTABILITY IN LIQUIDS

by S. Chandrasekhar

The Enrico Fermi Institute for Nuclear Studies
The University of Chicago
Chicago

"*Every natural process involves in greater or less degree friction or conduction of heat. But in the domain of irreversible processes the principle of least action is no longer sufficient; for the principle of the increase of entropy brings into the system of physics a wholly new element foreign to the action principle and which demands special mathematical treatment.*"[1]

Max Planck (1909)

ABSTRACT

In this paper the thermodynamics of thermal instability in liquids is considered and the following theorem is proved: Thermal instability as cellular convection will set in at the minimum (adverse) temperature gradient which is necessary to maintain a balance between the rate of dissipation of energy by all irreversible processes present and the rate of liberation of the thermodynamically available energy by the buoyancy force acting on the fluid. Likewise, the onset of thermal instability will be as overstable oscillations if it is possible (at a lower adverse temperature gradient) to balance in a synchronous manner the periodically varying amounts of kinetic and other forms of energy with similarly varying rates of dissipation and liberation of energy.

I

INTRODUCTION

It is known that when a horizontal layer of liquid is heated from below, instability with ensuing convection will set in if the prevailing adverse temperature gradient (β) exceeds a certain critical value. As Rayleigh [1] first showed, the stability of the liquid under these circumstances depends on the numerical value of the non-dimensional parameter,

$$R = \frac{g \alpha \beta}{\varkappa \nu} d^4, \tag{1}$$

– called the Rayleigh number – where d denotes the depth of the layer, g the acceleration due to gravity, and α, $\varkappa$ and ν are the coefficients of volume expansion, thermometric conductivity and kinematic viscosity, respectively. The condition for instability is that R exceeds a certain determinate critical value, R_c.

During the past few years, the effects of rotation and magnetic field, separately and jointly, on the onset of thermal instability have been investigated both experimentally and theoretically (for a general account of these investigations see Chandrasekhar [2]).

[1] Translated from M. Planck, Leidener Vortrag, 9.12.1909: „Die Einheit des physikalischen Weltbildes".

The theoretical determination of the critical RAYLEIGH number, R_c, for the onset of instability depends, always, on the solution of a characteristic value problem which gives the value of R at which a disturbance periodic in the horizontal plane with a specified wave number, a, first becomes unstable; the required value of R_c is the minimum of R, determined in this fashion, as a function of a. In all cases which have been investigated so far, it has been found that the underlying characteristic value problem can be solved by a variational method based on a formula for R which expresses it as the ratio of two integrals which are, in most instances, positive-definite. For the classical RAYLEIGH problem, it was shown by MALKUS [3] (see also [4]) that the formula which provides the basis for the variational treatment (in case the instability sets in as a stationary pattern of motions) equates, simply, the rate at which energy is dissipated by viscosity to the rate at which the thermodynamically available energy is released by the buoyancy force acting on the fluid. In this paper we shall generalize this principle to include the effects of rotation and magnetic field and allow also for the possibility of convection setting in as oscillations of increasing amplitude (i.e., as overstability).

II

THE CLASSICAL PROBLEM

As a preliminary to the consideration of the more general cases in the following sections, it will be convenient to have a parallel treatment of the classical problem. As we have stated, this problem has, effectively, been considered by MALKUS [3]; however our point of view is slightly different and the presentation will be adapted to the discussion which is to follow.

The equations which govern the perturbation of an initial static state in which the temperature gradient $-\beta$ prevails, are:

$$\frac{\partial u_i}{\partial t} = -\frac{\partial \varpi}{\partial x_i} + \gamma \theta \lambda_i + \nu \nabla^2 u_i \tag{2}$$

and

$$\frac{\partial \theta}{\partial t} = \beta w + \varkappa \nabla^2 \theta, \tag{3}$$

where u_i denotes the component of the velocity $\mathfrak{u}$ in the direction i ($i = 1, 2, 3$) of a rectangular system of reference x_i; θ the perturbation in the temperature; $\vec{\lambda}$ a unit vector in the direction of the vertical (i.e., of the acceleration of gravity $\mathfrak{g}$); $w (= \lambda_j u_j)$ is the vertical component of the velocity; $\gamma = g\,\alpha$; and $\varrho\,\varpi$ (where ϱ is the density) is related to the departures from the static pressure distribution caused by the perturbation.

By taking the curl of equation (2), we can eliminate ϖ and obtain

$$\frac{\partial \omega_i}{\partial t} = \gamma\, \varepsilon_{ijk} \frac{\partial \theta}{\partial x_j} \lambda_k + \nu \nabla^2 \omega_i, \tag{4}$$

where

$$\vec{\omega} = \operatorname{curl} \mathfrak{u} \tag{5}$$

denotes the vorticity. ε_{ijk} is a tensor with the following properties: $\varepsilon_{ijk} = -\varepsilon_{ikj}$, $\varepsilon_{iik} = 0$, $\varepsilon_{ijk} \cdot \varepsilon_{ljk} = \delta_{il}$, where $\delta_{il} = 1$ for $i = l$ and $\delta_{il} = 0$ for $i \neq l$. Taking once again the curl of equation (4) and making use of the solenoidal character of u, we obtain

$$-\frac{\partial}{\partial t} \nabla^2 u_i = \gamma \left(\lambda_j \frac{\partial^2 \theta}{\partial x_j \partial x_i} - \lambda_i \nabla^2 \theta \right) - \nu \nabla^4 u_i. \tag{6}$$

Multiplying equation (4) by λ_i, we get

$$\frac{\partial \zeta}{\partial t} = \nu \nabla^2 \zeta, \tag{7}$$

where

$$\zeta = \lambda_j \omega_j = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \tag{8}$$

is the z-component of the vorticity and u and v are the horizontal components of the velocity in the rectangular directions x and y respectively. Similarly, the z-component of equation (6) is

$$\frac{\partial}{\partial t} \nabla^2 w = \gamma \left(\frac{\partial^2 \theta}{\partial x^2} + \frac{\partial^2 \theta}{\partial y^2} \right) + \nu \nabla^4 w. \tag{9}$$

In treating the perturbation problem characterized by the foregoing equations, we analyze the disturbance into normal modes and consider each mode separately. In the present problem, the normal modes are disturbances, periodic in the (x, y)-plane and of assigned wave numbers. And a disturbance belonging to a particular wave number, a, will be such that all relevant quantities (q) describing it will satisfy the equations

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) q = -a^2 q \quad \text{and} \quad \nabla^2 q = \left(\frac{\partial^2}{\partial z^2} - a^2 \right) q. \tag{10}$$

Considering then a disturbance belonging to a wave number a and assuming that the marginal state is a stationary one, we can, after eliminating θ between equations (3) and (9), reduce the problem to one in characteristic values, for R (as dependent on a). We shall not carry out this reduction as it is available in the literature in several places. But it is important for our purposes to show how, once equations (3) and (9) have been solved, we can complete the solution by determining the velocities in the horizontal plane as well. We can accomplish this, quite generally, as follows:

Resolving the horizontal component of the velocity into a rotational and an irrotational part in the manner

$$u = \frac{\partial \varphi}{\partial x} - \frac{\partial \psi}{\partial y} \quad \text{and} \quad v = \frac{\partial \varphi}{\partial y} + \frac{\partial \psi}{\partial x}, \tag{11}$$

where φ and ψ are two scalar functions, we have

$$-\frac{\partial w}{\partial z} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = \frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} \tag{12}$$

and

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2}. \tag{13}$$

If φ and ψ are FOURIER-analyzed in the same way as all the other quantities, and we are considering a disturbance belonging to a wave number a, then it follows from equations (12) and (13) that

$$\varphi = \frac{1}{a^2}\frac{\partial w}{\partial z} \quad \text{and} \quad \psi = -\frac{1}{a^2}\zeta. \tag{14}$$

Hence,

$$u = \frac{1}{a^2}\left(\frac{\partial^2 w}{\partial x\,\partial z} + \frac{\partial \zeta}{\partial y}\right) \quad \text{and} \quad v = \frac{1}{a^2}\left(\frac{\partial^2 w}{\partial y\,\partial z} - \frac{\partial \zeta}{\partial x}\right). \tag{15}$$

These equations express u and v in terms of the z-components of the velocity and the vorticity; they are of general applicability.

Returning to equations (3), (7) and (9), we shall consider them under conditions of a stationary state. Then from equation (7), it follows directly that

$$\zeta \equiv 0, \tag{16}$$

and that, therefore,

$$u = \frac{1}{a^2}\frac{\partial^2 w}{\partial x\,\partial z} \quad \text{and} \quad v = \frac{1}{a^2}\frac{\partial^2 w}{\partial y\,\partial z}. \tag{17}$$

Consider, now, the average rate of dissipation of energy by viscosity in a unit column of the fluid (i.e., in a vertical column of unit cross section). This is given by[1]

$$\varepsilon_\nu = -\nu\int_0^d \langle u_i \nabla^2 u_i\rangle\, dz, \tag{18}$$

where the angular brackets signify that the quantity enclosed is to be averaged over the entire (x, y)-plane. (In this paper the angular brackets will always be used in this sense.)]

Making use of equations (10) and (17), we have

$$\langle u_i \nabla^2 u_i\rangle = \left\langle \frac{1}{a^4}\frac{\partial^2 w}{\partial x\,\partial z}\left(\frac{\partial^2}{\partial z^2} - a^2\right)\frac{\partial^2 w}{\partial x\,\partial z} + \frac{1}{a^4}\frac{\partial^2 w}{\partial y\,\partial z}\left(\frac{\partial^2}{\partial z^2} - a^2\right)\frac{\partial^2 w}{\partial y\,\partial z} + w\left(\frac{\partial^2}{\partial z^2} - a^2\right)w \right\rangle. \tag{19}$$

Without loss of generality, we may suppose that w is of the form

$$w = W(z)\sin a_x x \sin a_y y, \tag{20}$$

where $W(z)$ is a function of z only, and

$$a_x^2 + a_y^2 = a^2. \tag{21}$$

With w given by equation (20), equation (19) becomes

$$\langle u_i \nabla^2 u_i\rangle = \frac{1}{4}\left\{W(D^2 - a^2)W + \frac{1}{a^2}DW(D^2 - a^2)DW\right\}, \tag{22}$$

where

$$D = \frac{d}{dz}. \tag{23}$$

Inserting this expression for $\langle u_i \nabla^2 u_i\rangle$ in equation (18), we have

$$\varepsilon_\nu = -\frac{\nu}{4a^2}\int_0^d \left\{a^2 W(D^2 - a^2)W + DW(D^2 - a^2)DW\right\} dz. \tag{24}$$

[1] This is per unit mass. We will not repeat this every time: it must be understood.

By an integration by parts, we find that

$$\int_0^d DW(D^2 - a^2)DW\,dz = -\int_0^d D^2W(D^2 - a^2)W\,dz, \tag{25}$$

the integrated part vanishes on account of the boundary conditions[1] on W. Hence

$$\varepsilon_\nu = \frac{\nu}{4a^2}\int_0^d [(D^2 - a^2)W]^2\,dz, \tag{26}$$

which expresses ε_ν as a positive-definite integral.

Consider next, the average rate of liberation of energy, ε_g, in a unit column of the fluid by the buoyancy force acting on it. This is given by

$$\varepsilon_g = \gamma\int_0^d \langle u_i\,\lambda_i\,\theta\rangle\,dz = \gamma\int_0^d \langle\theta\,w\rangle\,dz. \tag{27}$$

On the other hand, according to equation (3), in a steady state,

$$w = -\frac{\varkappa}{\beta}\nabla^2\theta = -\frac{\varkappa}{\beta}\left(\frac{\partial^2}{\partial z^2} - a^2\right)\theta. \tag{28}$$

Substituting for w in accordance with this equation in the expression for ε_g and writing θ in the form (cf. equation (20))

$$\theta = \Theta(z)\sin a_x x\sin a_y y, \tag{29}$$

we obtain

$$\varepsilon_g = -\frac{\varkappa\gamma}{4\beta}\int_0^d \Theta(D^2 - a^2)\Theta\,dz. \tag{30}$$

By an integration by parts this can be brought to the positive-definite form

$$\varepsilon_g = \frac{\varkappa\gamma}{4\beta}\int_0^d [(D\Theta)^2 + a^2\Theta^2]\,dz. \tag{31}$$

(The integrated part again vanishes since $\theta = 0$ on the boundaries.)

If we now demand that

$$\varepsilon_\nu = \varepsilon_g, \tag{32}$$

then

$$\frac{\varkappa\gamma}{\beta}\int_0^d [(D\Theta)^2 + a^2\Theta^2]\,dz = \frac{\nu}{a^2}\int_0^d [(D^2 - a^2)W]^2\,dz. \tag{33}$$

To verify that this is exactly the relation for β which is minimized in Pellew and Southwell's variational treatment of the problem [5], we first observe that according to equation (9), in a steady state,

$$\gamma a^2\Theta = \nu(D^2 - a^2)^2W. \tag{34}$$

[1] These are $W = 0$; and either $DW = 0$ or $D^2W = 0$ depending on whether the bounding surface is rigid or free.

Hence, we may write

$$\Theta = \frac{\nu}{\gamma a^2} F \quad \text{where} \quad F = (D^2 - a^2)^2 W. \tag{35}$$

With Θ expressed in terms of W in this manner, equation (33) becomes

$$\frac{\varkappa \nu^2}{\beta \gamma a^4} \int_0^d [(DF)^2 + a^2 F^2]\, dz = \frac{\nu}{a^2} \int_0^d [(D^2 - a^2) W]^2 dz, \tag{36}$$

or,

$$\frac{\beta \gamma}{\varkappa \nu} = \frac{\int_0^d [(DF)^2 + a^2 F^2]\, dz}{a^2 \int_0^d [(D^2 - a^2) F]^2 dz}. \tag{37}$$

Equation (37) clearly expresses the RAYLEIGH number (for the specified a) as the ratio of two positive definite integrals. And it can now be verified that in the variational treatment of the problem the expression which is minimized is indeed the quantity on the right-hand side of equation (37). Therefore, *the condition for instability is, simply, that the adverse temperature gradient be greater than the minimum necessary to maintain, steadily, the equality between ε_ν and ε_g.*

III

THE EFFECT OF A MAGNETIC FIELD

Suppose now that the liquid which is being investigated is an electrical conductor and that a uniform external magnetic field of intensity H is impressed on it in the direction of $\mathfrak{g}$. (The generalization to the case when the directions of $\mathfrak{H}$ and $\mathfrak{g}$ do not coincide is not difficult, but since it does not add anything essential to the physics of the problem, we shall not consider it here.) Under these circumstances, equation (2) is replaced by

$$\frac{\partial u_i}{\partial t} = -\frac{\partial \varpi}{\partial x_i} + \gamma \theta \lambda_i + \nu \nabla^2 u_i + \frac{\mu H}{4\pi \varrho} \frac{\partial h_i}{\partial z}, \tag{38}$$

where μ denotes the magnetic permeability and h_i is the perturbation in the magnetic field; the latter is governed by the equations

$$\frac{\partial h_i}{\partial t} = H \frac{\partial u_i}{\partial z} + \eta \nabla^2 h_i, \tag{39}$$

and

$$\frac{\partial h_i}{\partial x_i} = 0. \tag{40}$$

In equation (39)

$$\eta = \frac{1}{4\pi \mu \sigma}, \tag{41}$$

where σ is the coefficient of electrical conductivity.

In a stationary state, equation (39) gives

$$\eta \nabla^2 h_i = \eta \left(\frac{\partial^2}{\partial z^2} - a^2\right) h_i = -H \frac{\partial u_i}{\partial z}. \tag{42}$$

It can be shown that in a steady state the presence of a magnetic field will not induce any vorticity in the z-direction; therefore, the horizontal components of the velocity will still be given by equation (17) of the last section.

Since the liquid considered is an electrical conductor, currents will be generated by motions across the lines of force; and these currents will give rise to JOULE heating. Thus, there is in this case an additional source of irreversible dissipation of energy; and we must take this into account.

Let ε_σ denote, then, the average rate of dissipation of energy in a unit column of the fluid by JOULE heating. This is given by

$$\varepsilon_\sigma = -\frac{\mu \eta}{4\pi \varrho} \int_0^d \langle h_i \nabla^2 h_i \rangle \, dz, \tag{43}$$

or, making use of equations (17) and (42), we have

$$\varepsilon_\sigma = \frac{\mu H}{4\pi\varrho} \int_0^d \left\langle h_i \frac{\partial u_i}{dz} \right\rangle dz = \frac{\mu H}{4\pi\varrho} \int_0^d \left\langle \frac{1}{a^2}\left(h_x \frac{\partial^3 w}{\partial x \, \partial z^2} + h_y \frac{\partial^3 w}{\partial y \, \partial z^2}\right) + h_z \frac{\partial w}{\partial z} \right\rangle dz. \tag{44}$$

Since averaging over the horizontal plane is equivalent to integrating over a square with a side equal to the wave length, λ, of the disturbance, we can write, for example,

$$\left\langle h_x \frac{\partial^3 w}{\partial x \, \partial z^2} \right\rangle = C \iint h_x \frac{\partial}{\partial x} \left(\frac{\partial^2 w}{\partial z^2}\right) dx \, dy, \tag{45}$$

where the integration is over a square of area λ^2 and C is an appropriate constant of proportionality. By an integration by parts, we obtain

$$\left\langle h_x \frac{\partial^3 w}{\partial x \, \partial z^2} \right\rangle = -C \iint \frac{\partial h_x}{\partial x} \frac{\partial^2 w}{\partial z^2} \, dx \, dy = -C \left\langle \frac{\partial h_x}{\partial x} \frac{\partial^2 w}{\quad} \right\rangle. \tag{46}$$

The terms in h_x and h_y in the integral defining ε_σ thus combine to give

$$-\frac{1}{a^2} \int_0^d \left\langle \left(\frac{\partial h_x}{\partial x} + \frac{\partial h_y}{\partial y}\right) \frac{\partial^2 w}{\partial z^2} \right\rangle dz = \frac{1}{a^2} \int_0^d \left\langle \frac{\partial h_z}{\partial z} \frac{\partial^2 w}{\partial z^2} \right\rangle dz. \tag{47}$$

Hence,

$$\varepsilon_\sigma = \frac{\mu H}{4\pi\varrho} \int_0^d \left\langle h_z \frac{\partial w}{\partial z} + \frac{1}{a^2} \frac{\partial h_z}{\partial z} \frac{\partial^2 w}{\partial z^2} \right\rangle dz. \tag{48}$$

A further integration by parts (this time with respect to z) of the term in $\partial h_z/\partial z$ n (48), transforms this expression into

$$\varepsilon_\sigma = -\frac{\mu H}{4\pi \varrho a^2} \int_0^d \left\langle \frac{\partial w}{\partial z} \left(\frac{\partial^2}{\partial z^2} - a^2\right) h_z \right\rangle dz. \tag{49}$$

But, according to equation (42),

$$\left(\frac{\partial^2}{\partial z^2} - a^2\right) h_z = -\frac{H}{\eta}\frac{\partial w}{\partial z}. \tag{50}$$

Hence

$$\varepsilon_\sigma = \frac{\mu H^2}{4\pi \varrho a^2 \eta}\int_0^d \left\langle \left(\frac{\partial w}{\partial z}\right)^2 \right\rangle dz = \frac{\mu H^2}{16\pi \varrho a^2 \eta}\int_0^d (D W)^2\, dz. \tag{51}$$

On substituting for η from equation (41), the expression for ε_σ finally becomes

$$\varepsilon_\sigma = \frac{\mu^2 H^2 \sigma}{4 \varrho a^2}\int_0^d (D W)^2\, dz. \tag{52}$$

Combining equation (52) with equations (27) for ε_ν, we obtain, for the total average rate of irreversible dissipation of energy in a unit column of the fluid, the expression

$$\varepsilon_\nu + \varepsilon_\sigma = \frac{\nu}{4a^2}\int_0^d \left\{[(D^2 - a^2) W]^2 + \frac{\mu^2 H^2 \sigma}{\varrho \nu}(D W)^2\right\} dz. \tag{53}$$

The expression for the average rate of liberation of energy, ε_g, by the buoyancy force is unchanged. We have (cf. equation (31)):

$$\varepsilon_g = \frac{\varkappa \gamma}{4\beta}\int_0^d [(D\Theta)^2 + a^2\Theta^2]\, dz. \tag{54}$$

(However, the further equation (35) relating Θ with W is not valid in the present problem.)

By demanding

$$\varepsilon_\sigma + \varepsilon_\nu = \varepsilon_g, \tag{55}$$

we obtain a relation which expresses β as the ratio of two positive-definite integrals; and it can be verified that it is indeed this expression which is minimized in the variational treatment of the problem [6].

IV

THE EFFECT OF ROTATION

We shall now consider a case in which $\zeta \neq 0$ and the motion in the (x, y)-plane is not irrotational. Such a case arises when the fluid is subject to a rotation and we have to allow for the resulting Coriolis acceleration in the equation of motion.

To illustrate the problem in its simplest context, we shall suppose that the axis of rotation coincides with the vertical. Equation (2) is then replaced by

$$\frac{\partial u_i}{\partial t} = -\frac{\partial \varpi}{\partial x_i} + \gamma\theta\lambda_i + \nu\nabla^2 u_i + 2\Omega\varepsilon_{ijk} u_j \lambda_k, \tag{56}$$

where Ω denotes the angular velocity of rotation about $\vec{\lambda}$; further, in place of equations (4), (7) and (9) we now have

$$\frac{\partial \zeta}{\partial t} = \nu \nabla^2 \zeta + 2\Omega \frac{\partial w}{\partial z}, \tag{57}$$

and

$$\frac{\partial}{\partial t}(\nabla^2 w) = \gamma \left(\frac{\partial^2 \theta}{\partial x^2} + \frac{\partial^2 \theta}{\partial y^2}\right) + \nu \nabla^4 w - 2\Omega \frac{\partial \zeta}{\partial z}. \tag{58}$$

In a stationary state, these equations become

$$\nu \nabla^2 \zeta = -2\Omega \frac{\partial w}{\partial z} \tag{59}$$

and

$$\nu \nabla^4 w - 2\Omega \frac{\partial \zeta}{\partial z} = -\gamma \left(\frac{\partial^2 \theta}{\partial x^2} + \frac{\partial^2 \theta}{\partial y^2}\right) = \gamma a^2 \theta. \tag{60}$$

From equation (59) it is apparent that in this problem $\zeta \neq 0$. We must, therefore, use the general solution (15) for the horizontal components of the velocity.

Consider again the average rate of dissipation of energy, ε_ν, by viscosity in a unit column of the fluid. The contribution to ε_ν by the horizontal component, u, of the velocity is

$$-\nu \int_0^d \langle u \nabla^2 u \rangle \, dz = -\frac{\nu}{a^4} \int_0^d \left\langle \left(\frac{\partial^2 w}{\partial x \partial z} + \frac{\partial \zeta}{\partial y}\right) \nabla^2 \left(\frac{\partial^2 w}{\partial x \partial z} + \frac{\partial \zeta}{\partial y}\right) \right\rangle dz. \tag{61}$$

It is clear from equation (59) (and indeed on general grounds) that the phases of the waves associated with w and ζ must agree. Therefore, $\partial^2 w/\partial x \partial z$ and $\partial \zeta/\partial y$ are out of phase and the average of their product must vanish. The cross terms in (61) do not, therefore, contribute and we are left with

$$\begin{aligned} -\nu \int_0^d \langle u \nabla^2 u \rangle \, dz &= -\frac{\nu}{a^4} \int_0^d \left\langle \frac{\partial^2 w}{\partial x \partial z} \nabla^2 \frac{\partial^2 w}{\partial x \partial z} + \frac{\partial \zeta}{\partial y} \nabla^2 \frac{\partial \zeta}{\partial y} \right\rangle dz \\ &= -\frac{\nu}{a^4} \int_0^d \left\{a_x^2 DW(D^2 - a^2) DW + a_y^2 Z(D^2 - a^2) Z\right\} dz, \end{aligned} \tag{62}$$

where we have supposed (in agreement with equations (20) and (59)) that

$$\zeta = Z(z) \sin a_x x \sin a_y y. \tag{63}$$

We have a similar contribution from the terms in v. Thus, all together, we have

$$\varepsilon_\nu = -\frac{\nu}{4a^2} \int_0^d \left\{a^2 W(D^2 - a^2) W + DW(D^2 - a^2) DW + Z(D^2 - a^2) Z\right\} dz. \tag{64}$$

The terms in W clearly combine to give the same expression as that found in II. The terms in Z can also be transformed into a positive-definite form by an integration by parts.[1] Thus, we finally obtain

$$\varepsilon_\nu = \frac{\nu}{4a^2} \int_0^d \left\{[(D^2 - a^2) W]^2 + (DZ)^2 + a^2 Z^2\right\} dz. \tag{65}$$

[1] The integrated part vanishes on account of the boundary conditions on Z. These are: $Z = 0$ on a rigid surface and $DZ = 0$ on a free surface.

The expression for ε_g is unchanged:

$$\varepsilon_g = \frac{\varkappa \gamma}{4 \beta} \int_0^d [(D\Theta)^2 + a^2 \Theta^2]\, dz. \tag{66}$$

By equating these expressions for ε_ν and ε_g we again obtain an expression for β as the ratio of two positive-definite integrals which is identical with the quantity which is minimized in the variational treatment of this problem [7].

It should be remarked that the expression for ε_ν given by equation (65) is entirely general[1]; it is in no way restricted to the particular problem under discussion: for, we have not used in its derivation equation (59) relating w and ζ which distinguishes this problem. (However, in the variational treatment this relation must be used.)

V

THE CASE WHEN OVERSTABILITY OCCURS

So far we have considered the marginal state as a stationary one. But it is known that instability can set in also via a state of oscillatory motions. Indeed, in the case of a layer of liquid heated from below and subject to rotation, instability can arise either, as cellular convection, or as overstable oscillations, depending on the value of the Prandtl number, $\nu/\varkappa$; the theoretical predictions [8] relative to this problem have been fully confirmed by experiments [9]. Examples of overstability also occur when rotation and magnetic field are simultaneously present [10], [11], [12]. The question, therefore, arises: How should one modify, under these circumstances, the principle which equates the rate of irreversible dissipation of energy with the rate of liberation of thermodynamically available energy by the buoyancy force, as a criterion for stability. It would appear natural that we generalize the principle along the following lines:

If the motion is periodic, then the kinetic energy of the fluid, as well as the rate of release of thermodynamic energy by the buoyancy force, will be subject to similar variations. Suppose that all quantities which describe the perturbation (such as w, θ, etc.,) vary with a circular frequency p so that all the amplitudes have a time dependent factor e^{ipt}. Then

$$u_i \frac{\partial u_i}{\partial t} = i p\, u_i^2. \tag{67}$$

We must allow for this change in the kinetic energy (per unit mass) in writing an equation of energy balance. If there are no sources of dissipation besides viscosity, then it would appear that we must equate

$$\varepsilon_\nu + i p \int_0^d \langle u_i^2 \rangle\, dz \tag{68}$$

[1] The assumption that the waves associated with $\partial^2 w/\partial x\, \partial z$ and $\partial \zeta/\partial y$ are out of phase is, however, essential: There are situations in which this need not be the case, e.g. when the directions of the axis of rotation and of $\mathfrak{g}$ do not coincide.

to the rate of release of energy

$$\varepsilon_g = \gamma \int_0^d \langle \theta\, w \rangle\, dz, \tag{69}$$

and allow for the fact that θ and w are no longer stationary.

On first sight, the appearance of the imaginary i, and complex numbers generally, in the equation for the energy balance may strike one as very odd. Its origin must be traced to the fact that the oscillations in the velocity and in the acceleration are out of phase. Consequently, the excess (or defect) of energy dissipated during one phase of the cycle must be exactly compensated by a similar excess (or defect) of energy liberated in a synchronous manner. It is this need for synchronism which determines the period of oscillation, as well as the RAYLEIGH number, as characteristics of the marginal state.[1]

We shall now verify that the thermodynamic principle as reformulated above is in agreement with the variational treatment of the problem considered in IV, for the case of overstability.

When the state is not stationary, the vertical component of the vorticity is necessarily non-vanishing and we must allow for this in the solution for the horizontal components of the velocity. The expression for ε_ν is, therefore, formally the same as that found in IV.

We must now evaluate the additional term in $\langle u_i^2 \rangle$ in (68). Considering the contribution to the integral by the x-component of the velocity, we have

$$\int_0^d \langle u^2 \rangle\, dz = \frac{1}{a^4} \int_0^d \left\langle \left(\frac{\partial^2 w}{\partial x\, \partial z} + \frac{\partial \zeta}{\partial y} \right)^2 \right\rangle dz. \tag{70}$$

Remembering that in the case under consideration the waves in the horizontal plane associated with $\partial^2 w / \partial x\, \partial z$ and $\partial \zeta / \partial y$ are out of phase and that, therefore, the average of their product vanishes, we have

$$\int_0^d \langle u^2 \rangle\, dz = \frac{1}{4 a^4} \int_0^d [a_x^2 (D\, W)^2 + a_y^2 Z^2]\, dz. \tag{71}$$

We have a similar contribution from $\langle v^2 \rangle$. Thus, all together, we have

$$i p \int_0^d \langle u_i^2 \rangle\, dz = \frac{i p}{4 a^2} \int_0^d [a^2 W^2 + (D\, W)^2 + Z^2]\, dz. \tag{72}$$

This must be added to the expression (65) for ε_ν.

Turning to the evaluation of ε_g, we must now use

$$\beta w = -\varkappa \nabla^2 \theta + i p\, \theta, \tag{73}$$

[1] The mathematical counterpart of this is that the determination of R and p is through the solution of a *double* characteristic value problem [8].

instead of (28) which we have used up to the present, to eliminate w. Accordingly we now have

$$\varepsilon_g = -\frac{\varkappa\gamma}{\beta}\int_0^d \left\langle \theta\left\{\left(\frac{\partial^2}{\partial z^2}-a^2\right)\theta - \frac{ip}{\varkappa}\theta\right\}\right\rangle dz = \frac{\varkappa\gamma}{4\beta}\int_0^d\left[(D\Theta)^2+a^2\Theta^2+\frac{ip}{\varkappa}\Theta^2\right]dz. \quad (74)$$

Hence, the equation which, under these circumstances, will determine a marginal state which is oscillatory is:

$$\frac{\nu}{a^2}\int_0^d\left\{[(D^2-a^2)W]^2+(DZ)^2+a^2Z^2+\frac{ip}{\nu}[a^2W^2+(DW)^2+Z^2]\right\}dz$$
$$=\frac{\varkappa\gamma}{\beta}\int_0^d\left\{(D\Theta)^2+a^2\Theta^2+\frac{ip}{\varkappa}\Theta^2\right\}dz; \quad (75)$$

and this is indeed the expression for β which is used in the variational treatment of the problem [8].

VI

CONCLUDING REMARKS

The analysis of the foregoing sections allows us to formulate the following general principle:

Thermal instability as cellular convection will set in at the minimum (adverse) temperature gradient which is necessary to maintain a balance between the rate of dissipation of energy by all irreversible processes present and the rate of liberation of the thermodynamically available energy by the buoyancy force acting on the fluid. Likewise, the onset of thermal instability will be as overstable oscillations if it is possible (at a lower adverse temperature gradient) to balance in a synchronous manner the periodically varying amounts of kinetic and other forms of energy with similarly varying rates of dissipation and liberation of energy.

REFERENCES

[1] Rayleigh, Phil. Mag. [6], *32*, 529, 1916.
[2] Chandrasekhar, S., Daedalus *86*, 323, 1957.
[3] Malkus, W.V.R., Proc. Roy. Soc. A, *225*, 185, 196, 1954.
[4] Jeffreys, H., Quart. Jour. of Mech. and Applied Math. *9*, 1, 1956.
[5] Pellew, A., and R. Southwell, Proc. Roy. Soc. A, *152*, 586, 1940.
[6] Chandrasekhar, S., Phil. Mag. [7], *43*, 501, 1952.
[7] Chandrasekhar, S., Proc. Roy. Soc. A, *217*, 306, 1953.
[8] Chandrasekhar, S., and D.D. Elbert, Proc. Roy. Soc. A, *231*, 198, 1955.
[9] Fultz, D., and Y. Nakagawa, Proc. Roy. Soc. A, *23*, 211, 1955.
[10] Chandrasekhar, S., Proc. Roy. Soc. A, *237*, 476, 1956.
[11] Nakagawa, Y., Proc. Roy. Soc. A, *242*, 81, 1957.
[12] Nakagawa, Y., Proc. Roy. Soc. A, (in press) 1958.

VARIATIONAL METHODS IN HYDRODYNAMICS

BY

S. CHANDRASEKHAR

1. Introduction. Studies in hydrodynamic and hydromagnetic stability have led to characteristic-value problems in differential equations of high order, and it has been possible to solve several of them by variational methods which, at least in the manner of their applications, appear novel. In this paper we shall briefly describe two such examples; a list of further examples will be found in Chandrasekhar [1].

2. First example. In the study of the instability of a layer of fluid heated from below and subject to Coriolis forces resulting from rotation with an angular velocity Ω about the vertical, one is led to the following problem: To solve

$$(D^2 - a^2 - i\sigma)Z = -\frac{2\Omega}{\nu} d\, DW, \tag{1}$$

$$(D^2 - a^2)(D^2 - a^2 - i\sigma)W - \frac{2\Omega}{\nu} d^3\, DZ = F, \tag{2}$$

and

$$(D^2 - a^2 - i\tilde{\omega}\sigma)F = -Ra^2W, \tag{3}$$

together with the boundary conditions

$$W = F = 0 \text{ for } z = \pm\tfrac{1}{2}, \tag{4}$$

and

$$\begin{aligned} &\textit{Either}\ DW = Z = 0 \text{ on } z = \pm\tfrac{1}{2},\\ &\textit{Or}\quad D^2W = DZ = 0 \text{ on } z = \pm\tfrac{1}{2},\\ &\textit{Or}\quad DW = Z = 0 \text{ on } z = +\tfrac{1}{2} \quad \text{and} \quad D^2W = DZ = 0 \text{ on } z = -\tfrac{1}{2}, \end{aligned} \tag{5}$$

where $D = d/dz$, a, Ω, ν, and $\tilde{\omega}$ are assigned constants, and σ is a parameter to be determined by the condition that the characteristic value R is real. The solution of the physical problem requires the minimum (with respect to a^2) of these real characteristic values of R for various assigned values of T $(= 4\Omega^2 d^4/\nu^2)$ and $\tilde{\omega}$.

A consideration of the foregoing *double* characteristic-value problem (double since both σ and R are to be determined) leads to the following variational principle:

From equations (1) to (3) it follows that

$$R = \frac{\int_{-\frac{1}{2}}^{+\frac{1}{2}} [(DF)^2 + (a^2 + i\tilde{\omega}\sigma)F^2]\, dz}{a^2 \int_{-\frac{1}{2}}^{+\frac{1}{2}} \{[(D^2 - a^2)W]^2 + d^2[(DZ)^2 + a^2Z^2] + i\sigma[(DW)^2 + a^2W^2 + d^2Z^2]\}\, dz}. \tag{6}$$

It can now be readily verified that the variation δR in R given by equation (6) due to variations δW and δZ in W and Z compatible only with the boundary conditions on W, Z, and F, is given by

$$(7) \qquad \delta R = -\frac{2}{a^2 I_2} \int_{-\frac{1}{2}}^{+\frac{1}{2}} \delta F[(D^2 - a^2 - i\tilde{\omega}\sigma)F + Ra^2 W]\, dz,$$

where I_2 stands for the integral in the denominator of the expression on the right-hand side of (6). Accordingly, $\delta R \equiv 0$ for all small arbitrary variations δF, provided that

$$(8) \qquad (D^2 - a^2 - i\tilde{\omega}\sigma)F + Ra^2 W = 0,$$

i.e., if the differential equation governing W is satisfied. On this account, formula (6) provides the basis for a variational procedure for solving equations (1) to (3) (for any assigned a^2 and σ) and satisfying the boundary conditions of the problem.

It should be noted that formula (6) does not express R as the ratio of two positive-definite integrals; indeed, for an arbitrarily assigned σ, R will be complex. Nevertheless, it appears that the simplest trial function for F, namely, $F = \cos \pi z$, already leads to surprisingly accurate determinations for the characteristic values provided that, for the chosen form of F, the functions W and Z are determined as solutions of equations (1) and (2). (For the details of the solution and for the comparison with experimental results, see Chandrasekhar and Elbert [2] and Fultz and Nakagawa [3].)

3. Second example. The stability of an incompressible, heavy, viscous fluid of variable density leads to the following characteristic-value problem (Chandrasekhar [4]): To solve

$$(9) \quad D\left\{\left[\rho - \frac{\mu}{n}(D^2 - k^2)\right] Dw - \frac{1}{n}(D\mu)(D^2 + k^2)w\right\}$$
$$= k^2 \left\{-\frac{g}{n^2}(D\rho)w + \left[\rho - \frac{\mu}{n}(D^2 - k^2)\right] w - \frac{2}{n}(D\mu)(Dw)\right\},$$

together with the boundary conditions

$$(10) \qquad w = 0 \text{ for } z = 0 \text{ and } l,$$

and

$$(11) \quad \begin{array}{llll} \textit{Either} & Dw = 0 \text{ for } z = 0 & \text{and} & l, \\ \textit{Or} & D^2w = 0 \text{ for } z = 0 & \text{and} & l, \\ \textit{Or} & Dw = 0 \text{ for } z = 0 & \text{and} & D^2w = 0 \text{ for } z = l, \\ \textit{Or} & D^2w = 0 \text{ for } z = 0 & \text{and} & Dw = 0 \text{ for } z = l, \end{array}$$

where $\rho = \rho(z)$ and $\mu = \mu(z)$ are given functions of z, k is an assigned (real) constant, and n is the characteristic-value parameter. (Note that n can be complex.)

One can deduce from equation (9) that

$$(12)\quad n\int_0^l \rho\left\{w^2+\frac{1}{k^2}(Dw)^2\right\}dz-\frac{g}{n}\int_0^l (D\rho)w^2\,dz$$
$$=-\int_0^l\left\{\mu\left[k^2w^2+2(Dw)^2+\frac{1}{k^2}(D^2w)^2\right]+(D^2\mu)w^2\right\}dz;$$

and again this last equation provides the basis for a convenient variational procedure for determining n. For, considering the effect on n [determined in accordance with equation (12)] of an arbitrary variation δw in w compatible only with the boundary conditions on w, we find that

$$(13)\quad -\frac{1}{2}k^2\left(I_1+\frac{g}{n^2}I_2\right)\frac{\delta n}{n}$$
$$=\int_0^l \delta w\left\{k^2\left[\rho w-\frac{\mu}{n}(D^2-k^2)w-\frac{g}{n^2}(D\rho)w-\frac{2}{n}(D\mu)(Dw)\right]\right.$$
$$\left.-D\left[\rho\,Dw-\frac{\mu}{n}(D^2-k^2)\,Dw-\frac{1}{n}(D\mu)(D^2+k^2)w\right]\right\}dz,$$

where

$$(14)\qquad I_1=\int_0^l \rho\left\{w^2+\frac{1}{k^2}(Dw)^2\right\}dz$$

and

$$(15)\qquad I_2=\int_0^l (D\rho)w^2\,dz.$$

It will be noticed that the variational procedure in this instance involves the solution of a quadratic equation none of whose coefficients are positive-definite. Nevertheless, as Hide [5] has shown, simple trial functions for w satisfying the boundary conditions enable the complicated dependence of n on the various parameters of the problem to be determined.

Bibliography

1. S. Chandrasekhar, *On characteristic value problems in high order differential equations which arise in studies on hydrodynamic and hydromagnetic stability*, Amer. Math. Monthly vol. 61 (1954) pp. 32–45.

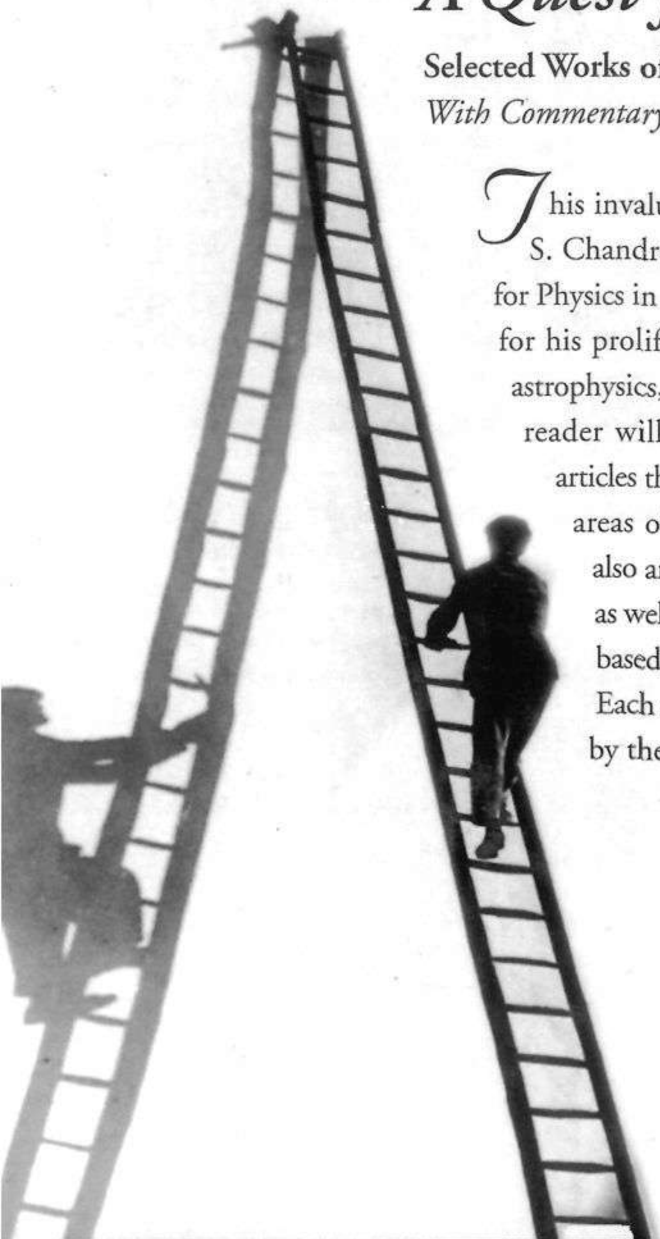
2. ——— and D. Elbert, *The instability of a layer of fluid heated below and subject to Coriolis forces, II*, Proc. Roy. Soc. London Ser. A vol. 231 (1955) pp. 198–210.

3. D. Fultz and Y. Nakagawa, *Experiments on over-stable thermal convection in mercury*, Proc. Roy. Soc. London Ser. A vol. 231 (1955) pp. 211–225.

4. S. Chandrasekhar, *The character of the equilibrium of an incompressible heavy viscous fluid of variable density*, Proc. Cambridge Philos. Soc. vol. 51 (1955) pp. 162–178.

5. R. Hide, *The character of the equilibrium of an incompressible heavy viscous fluid of variable density: an approximate theory*, Proc. Cambridge Philos. Soc. vol. 51 (1955) pp. 179–201.

University of Chicago,
Chicago, Ill.



A Quest for Perspectives

Selected Works of S. Chandrasekhar

With Commentary

This invaluable book presents selected papers of S. Chandrasekhar, co-winner of the Nobel Prize for Physics in 1983 and a scientific giant well known for his prolific and monumental contributions to astrophysics, physics and applied mathematics. The reader will find here most of Chandrasekhar's articles that led to major developments in various areas of physics and astrophysics. There are also articles of a popular and historical nature, as well as some hitherto unpublished material based on Chandrasekhar's talks at conferences. Each section of the book contains annotations by the editor.

P175 hc

ISBN 1-86094-283-0

9 781860 942839

ISBN 1-86094-201-6(set)

9 781860 942013

Imperial College Press

www.icpress.co.uk