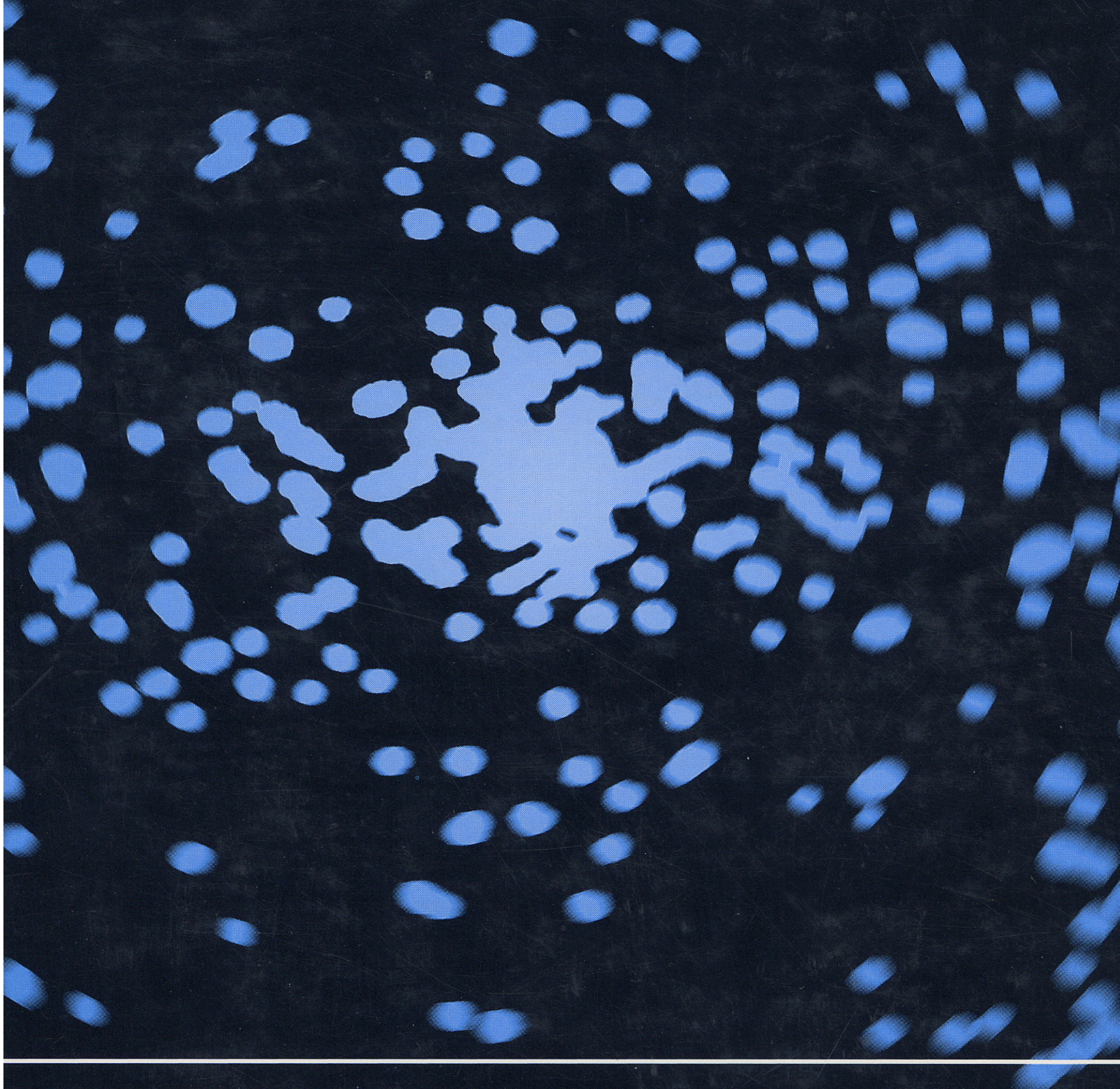
A Life in ASTROPHYSICS

Selected Papers of

Viktor A. Ambartsumian



ALLERTON PRESS, INC.

A Life in ASTROPHYSICS

Selected Papers of
Viktor A. Ambartsumian

A Life in ASTROPHYSICS

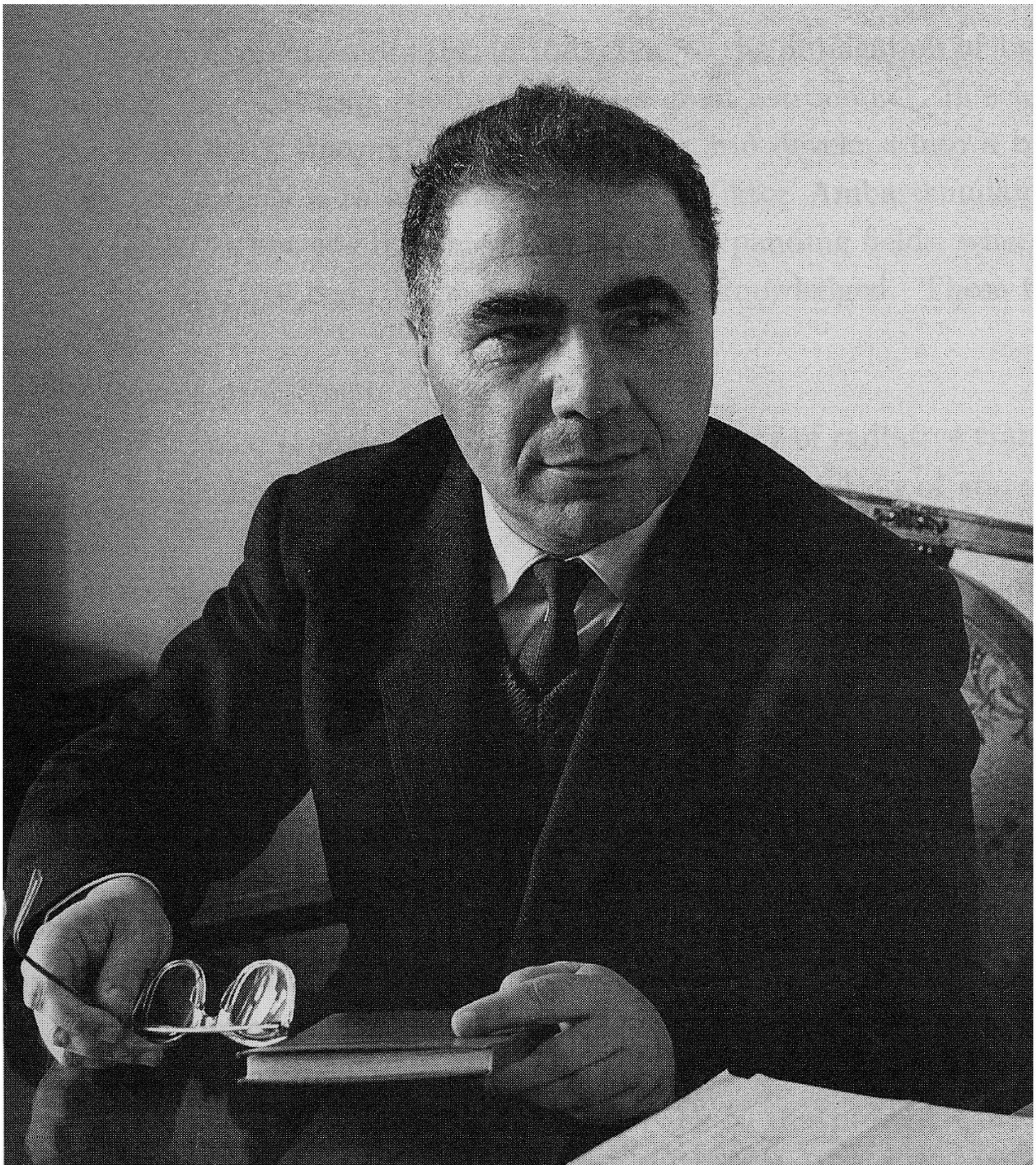
Selected Papers of Viktor A. Ambartsumian

Edited by Rouben V. Ambartsumian;
with an Introduction by Geoffrey Burbidge

Allerton Press, Inc. / New York

Library of Congress Cataloging-in-Publication Data

Ambarts͡umia͡n, V. A. (Viktor Amazaspovich)
A life in astrophysics : selected papers of Viktor Ambartsumian / edited by Rouben V. Ambartsumian ; with an introduction by Geoffrey Burbidge.
p. cm.
ISBN 0-89864-082-2
1. Stars — Evolution. 2. Galaxies — Evolution. 3. Inverse problems (Differential equations) 4. Radiative transfer. 5. Astrophysics.
I. Ambartzumian, R. V. II. Title.
QB806.A43 1998
523.01 — dc21 97-52237
CIP



INTRODUCTION

I am very happy to write this introduction to the publication of a small collection of the most important papers by a great astronomer. In science, to open a new path that expands over decades and develops into a broad avenue of inquiry is a rare accomplishment. Viktor Ambartsumian has been a pioneer in not one but three perpetually expanding fields, where his outstanding creative contributions are widely acknowledged. These fields have been:

(i) Inverse problems in mathematics.

(ii) Invariance principles as applied to the theory of radiative transfer.

(iii) Radically new concepts of the origin and evolution of stars and galaxies.

Key papers in each of these areas are included in this volume.

The landscape of inverse problems can hardly yield to brief description today due to its diversity. However, all ramifications of this topic grow from the same root — the inverse Sturm-Liouville problem, which is a vast and deep mathematical theory in itself. Yet the initial germ for the latter theory was the paper by Ambartsumian which opens the present collection. Reading this book produces the impression that with this, his first publication, young Ambartsumian set the standard for the scientific level of his own work in subsequent years.

Ambartsumian concentrated on the theory of radiative transfer in the period up to and immediately after World War II and published a number of now classical papers, including papers 7–11 of the present collection. Later, the principles of invariance, discovered and elegantly developed by Ambartsumian, became the main problem-solving tools in the writings of S. Chandrasekhar, G. Munch, R. Bellman, and many others who worked in this field.

Ambartsumian established new directions in both theory and observations for over thirty years. The last few papers included in this collection cover the third area which is of special interest in cosmogony and cosmology.

Many years ago, empirical evidence suggested to Ambartsumian that systems of stars grouped into associations often have positive total energy. He subsequently concluded that they were in fact dispersing; he went on to develop a theory of expanding associations which has been generally accepted. Following this he saw much more clearly than others in the 1950s and 1960s that many groups and clusters of galaxies also appear to have positive total energy. He therefore concluded that they too are coming apart, though the common point of view has been that in general such systems are bound by unseen matter. Ambartsumian also applied the same arguments to what appear to be violent outbursts in the nuclei of galaxies. His conclusion was that these explosions were the manifestations of creation events.

In all these ideas he was initially alone. For many it is still not easy to admit the possibility of changing the traditional paradigm. But the pressure of observation does its work, slowly but steadily. Truly, by their impact on cosmogonical thinking the ideas of Ambartsumian started a revolution of Copernican scale.

Viktor Ambartsumian was born in Tbilisi to Armenian parents. He studied at Leningrad University (now, again the University of St. Petersburg) from 1925-1928 where his greatest interests were in mathematics and astronomy. As well as being one of the giants of astronomical research in the 20th century, Viktor Ambartsumian was also a great leader and organizer of science in Armenia, in Russia, and on the international level. He was Director of the Byurakan Observatory for a very long period. He was a member of the Academy of Sciences of the USSR, President of the Armenian Academy of Sciences for a number of years, and President of the International Astronomical Union in the period 1958-1961. In his lifetime, he was honored in many ways.

Viktor Amazaspovich Ambartsumian died on August 12, 1996, at his beloved Byurakan, and sadly did not see this book in print. I hope that this volume will enable his work to reach a wider audience than has so far been possible.

Geoffrey Burbidge
September 1997

CONTENTS

ORIGINAL PUBLICATION INDEX

(The titles are in the language of original publication.)

1. "Über eine Frage der Eigenwerttheorie," *Z. f. Phys.*, vol. 53, pp. 690–695, 1929.
2. "The excitation of the metastable states in the gaseous nebulae," Циркуляры ГАО [Pulkovo Obs. Circ.], no. 6, pp. 10–17, 1933.
3. "On the radiative equilibrium of a planetary nebula," Известия ГАО [Bulletin De L'Observatoire Central à Poulkovo], vol. XIII, no. 114, pp. 1–27, 1933.
4. "On the derivation of the frequency function of space velocities of the stars from the observed radial velocities," *Mon. Not. R. Astr. Soc.*, vol. 96, no. 3, pp. 172–179, 1936.
5. "К статистике двойных звёзд," Астр. журн. [Sov. Astr.], vol. 14, no. 3, pp. 207–219, 1937.
6. "К вопросу о динамике открытых скоплений," Уч. записки ЛГУ [Uch. Zap. LGU], no. 22, pp. 19–22, 1938.
7. "О рассеянии света атмосферами планет," Астр. журн. [Sov. Astr.], vol. 19, pp. 30–41, 1942.
8. "К вопросу о диффузном отражении света мутной средой," ДАН СССР [Dokl. Akad. Nauk SSSR], vol. 38, pp. 257–261, 1943.
9. "On the problem of the diffuse reflection of light," *Journal of Physics*, vol. 8, no. 2, pp. 65–75, 1944.
10. "К теории флуктуаций яркости в Млечном пути," ДАН СССР [Dokl. Akad. Nauk SSSR], vol. 44, no. 6, pp. 244–247, 1944.
11. "О поверхностных яркостях в Галактике," Астр. журн. [Sov. Astr.], vol. 23, pp. 257–268, 1946.

12. “Звёздные ассоциации,” Астр. журн. [Sov. Astr.], vol. 26, pp. 3–9, 1949.
13. “Кратные системы типа трапеции,” Сообщения Бюраканской обсерватории [Communications of Byurakan Obs.], vol. 15, pp. 3-35, 1954.
14. “On the patchy structure of the interstellar absorbing layer,” *Trans. IAU*, vol. 7, pp. 452–455, 1950.
15. “Superassociations in distant galaxies,” *The Galaxy and the Magellanic Clouds, IAU–URSI Symposium, no. 20, Canberra, March 18–28*, 1963, F. J. Kerr and A. W. Rodgers (eds.), Canberra, Austr. Acad. Sci., pp. 122–126, 1964.
16. “Flare stars,” *Inter. Astron. Union, Colloq. no. 15, Bamberg, 31 August–3 Sept.* 1971, in *New Directions and New Frontiers in Variable Star Research*, Bamberg, pp. 98–108, 1971.
17. “Фуоры,” Астрофизика, vol. 7, no. 4, pp. 557–572, 1971.
18. “Instability phenomena in systems of galaxies,” *Astron. Journal*, vol. 66, no. 10, pp. 536–540, 1961.
19. “On the nuclei of galaxies and their activity,” *Proceedings of the Thirteenth Conference on Physics at the University of Brussels, September 1964: The Structure and Evolution of Galaxies*, Interscience Publishers, pp. 1–14, 1965.
20. “Problems of extragalactic research,” Invited Discourse C at the General Assembly of IAU, *Trans. IAU*, vol. XI B, Academic Press, pp. 145–160, 1962.
21. “Galaxies and their nuclei,” *Highlights of Astronomy*, G. Contopoulos (ed.), IAU, pp. 51–66, 1974.
22. Introduction, *Nuclei of Galaxies*, in *Pont. Acad. Scient. Scripta Varia*, 35, pp. 1–12, 1971.
23. Epilogue adapted from “On some trends in the development of astrophysics,” *Ann. Rev. Astron. Astrophy.*, pp. 1–13, 1980.

EDITOR'S INTRODUCTION

"Who on earth is going to read all these papers? They are almost as old as I am." With this question my father would usually open the discussion of the edited versions of his papers as I brought them to his attention one by one. On one occasion I asked him to elaborate. "My doubts are not about the papers," he laughed. "They have not changed, unlike the fashions in science."

For this collection V.A. Ambartsumian selected several papers which have already proven to have a tremendous impact on their respective fields. The eighth paper in this collection, which presents the method of addition of layers, is a good example. In their book *Invariant Imbedding & Radiative Transfer in Slabs of Finite Thickness*, R.E. Bellman, R.E. Kalaba and M.C. Prestrud write that "As a result of this pioneering work, new analytic treatments were made available and simplified computational solutions were obtained. These ideas were further developed and extensively generalized by Chandrasekhar in a series of fundamental papers and in 1950 in his book....Many otherwise intractable problems were tamed, and great advances were made."[1] The invariance principle introduced by this paper has since found widespread application even in fields distant from astrophysics.

In choosing papers for this collection Ambartsumian also gave preference to those which, in his opinion, retain potential for future development. In earlier years he would often advise me to spend more time reading the scientific classics of the past. He firmly shared the old belief that, along with seeds which have since grown and produced scientific fruit, those classical works still contain many other seeds that remain hidden and await their chance for development. I assume that this belief, as applied to his own work, influenced my father's selections. For instance, Ambartsumian's first Solvay lecture, delivered in 1958, concerning the explosion of nuclei of the galaxies, was groundbreaking. That first lecture and the development

[1] American Elsevier Publishing Co., 1963, p. 1.

it started were described by Jerzy Neyman in his article "Reminiscences of a Revolutionary Period in Cosmology."[2] As Neyman recalled, despite initial skepticism, "Ambartsumian's arguments and many-sided documentation made the attending scholars think, and there followed several important international developments...." Neyman concluded that "Evidence in favor of the Ambartsumian Hypothesis is now overwhelming. My hearty congratulations to Professor V.A. Ambartsumian, the Copernican Revolutionary...!"

In preparing his second Solvay report, presented at the 1964 conference in Brussels, Ambartsumian had the benefit of numerous observations made in the intervening six years, which he used to considerably refine and develop the ideas outlined in the earlier lecture. Ultimately, he chose to include in this collection the second, more elaborate report (paper 19), rather than the first, more famous lecture. He could not avoid contemplating the chance of a recurrence of events such as he witnessed several times during his life, when ideas he had planted grew to provide splendid scientific results. Now, after he has passed away, I recall the following episode. During my editorial work on "Flare Stars," I became absorbed by the fascinating statistical problem the paper raises and wrote some mathematical comments, which I presented to my father. His reaction: "Put it in. After all, the purpose of all this is the further development of what has already been done." These comments are presented here as an addendum to paper 16.

The papers in this book are presented chronologically. However, they can be viewed as falling into two main groups. The first is the series which takes the reader successively from individual stars (papers 5, 16 and 17), multiple systems and star clusters (papers 6 and 13), associations and superassociations (papers 12, 15) into the world of galaxies (papers 18-22). The methodological and conceptual interconnection between these papers can be briefly described by H. Alfven's expression: "Science versus Myth." Among the weapons science uses in this struggle, mathematics is not the least important, and the papers of the second group (the remaining papers in this volume), complementary to this series, testify to Viktor Ambartsum-

---

[2] *Problems of Physics and Evolution of the Universe*, Publishing House of the Armenian Academy of Sciences, Yerevan, 1978, pp. 243-250.

ian's supreme mastery of the weapon of mathematics. In this connection, in his paper "Computed Tomography: Some History and Recent Developments," published in *Proceedings of Symposia in Applied Mathematics*, (Volume 27, 1982), Nobel Laureate A. M. Cormack made the following comment pointing to the relation of my father's work to computer tomography: "Ambartsumian gave the first numerical inversion of the Radon transform and it gives the lie to the often made statement that computed tomography would have been impossible without computers." The opening paper in this collection presents the first "inverse problem" ever solved. Now there are journals and series of monographs in mathematical physics which include in their titles that very term, coined by Ambartsumian in his first paper.

The papers are presented here with minor editorial changes and abridgements. Our general purpose in editing was to make the texts more precise. However, more serious abridgements have been made in papers [3], [13] and [17], where my father selected only those sections of the original papers which in his judgement have retained the most value. Among the texts now in this book, those which existed only in Russian have been translated from the three-volume collection of Ambartsumian's papers published in 1960 (Volumes 1 and 2) and 1988 (Volume 3) by the Armenian Academy of Sciences, Yerevan. In cases where an English version existed, it has been used here. However, some minor changes have been made in conformity with the corresponding texts in the Armenian Academy volumes. The original papers and translations were not always uniform in style and format, as well as in form of citation, and much of that variation has been preserved here as well.

The publication of several papers in the collection has been possible only with the consent of their owners, who are acknowledged herein and whom we thank for graciously allowing us to include these articles. Finally, I would like to express the hope that the reader will perceive in this collection depth and originality of the analyses and dynamic development of the topics — qualities that are always in fashion.

Rouben Ambartsumian
Yerevan, January 1998

A PROBLEM IN THE THEORY OF EIGENVALUES

In a certain special case (oscillating string, the natural boundary conditions) the spectrum of eigenvalues determines uniquely the differential equation to which it corresponds (in Schrödinger's theory, the "equation of amplitudes").

In those fields of theoretical physics (wave mechanics, theory of oscillations) where eigenvalue problems arise, the question of the uniqueness of determination of the mechanical system (i.e., of the Hamiltonian) by the set of the eigenvalues of the corresponding linear equation can be important. If a spectrum really completely determines the differential equation, then in principle it would become possible to determine the structure of an atomic system from the frequencies it is radiating or absorbing. This would mean solving a problem which is inverse to the Schrödinger problem. However, an approach to the general problem leads to many difficulties. Therefore, below we consider only a special case.

We prove that among all equations

$$\mu \frac{d^2\varphi}{dx^2} - q(x)\,\varphi + \alpha\varphi = 0,$$

where α is a "parameter" of eigenvalues, μ is a constant, $q(x)$ is a continuous function, for "natural boundary conditions"

$$\varphi'(0) = \varphi'(\pi) = 0$$

only the equation of the oscillating string

$$\mu \frac{d^2\varphi}{dx^2} + \alpha\varphi = 0$$

has the eigenvalues

$$\alpha_n = \kappa n^2.$$

§**1.** We start with the differential equation

$$(py')' - qy - \lambda ry + \alpha y = 0, \tag{1}$$

where λry is a perturbation term, q, r, p, p' are continuous functions of x and $p > 0$.

For the boundary conditions $y'(0) = y'(\pi) = 0$ the differential equation (1) has a countable set of eigenvalues which we can arrange in increasing order:

$$\alpha_1, \alpha_2, \alpha_3, ... \tag{2}$$

These eigenvalues are functions of λ. In this section our purpose is to show that these functions have no singularities on the real axis. It is sufficient to demonstrate that $\alpha_i(\lambda)$ are regular analytical functions of λ in the vicinity of the point $\lambda = 0$. The proof is by observation that the latter statement applies as well to the equation

$$(py')' + (q - \lambda_0 r)y - (\lambda - \lambda_0)ry + \alpha y = 0 \tag{3}$$

in the vicinity of arbitrary real λ_0. But equations (1) and (3) are identical.

First, let us suppose that $\alpha = 0$ is not an eigenvalue of the equation

$$(py')' + qy - \alpha y = 0. \tag{1'}$$

In this case, the differential operator

$$L(y) = (py')' + qy$$

has a Green function $G(x, \xi)$.

Then the power series

$$S(x, \xi, \lambda) = G_1(x, \xi) + \lambda G_2(x, \xi) + \lambda^2 G_3(x, \xi) + ..., \tag{4}$$

in which

$$G_n(x, \xi) =$$
$$\int \cdots \int G(x, t_1)\, r(t_1)\, G(t_1, t_2)\, r(t_2) ... r(t_{n-1})\, G(t_{n-1}, \xi)\, dt_1\, dt_2 ... dt_{n-1}$$

converges inside some disk $|\lambda| < \rho$, since $G(x,\xi)$ and $r(x)$ are bounded. We have

$$S(x,\xi,\lambda) = \frac{k(x,\xi,\lambda)}{r(\xi)},$$

where $k(x,\xi,\lambda)$ is the resolvent of the kernel $G(x,\xi)\,r(x)$. The function $S(x,\xi,\lambda)$ represents for $|\lambda| < \rho$ the Green function of the differential operator

$$L(y) - \lambda\, ry = (py')' - qy - \lambda\, ry.$$

The eigenvalues of equation (1) are the null-points of the Fredholm denominator of the kernel $S(x,\xi,\lambda)$. This means that for $|\lambda| < \rho$ we can determine these eigenvalues from the equation:

$$D(\alpha,\lambda) = 1 - \frac{1}{1!}D_1(\lambda)\,\alpha + \frac{1}{2!}D_2(\lambda)\,\alpha^2 - \frac{1}{3!}D_3(\lambda)\,\alpha^3 + ... = 0, \qquad (5)$$

where

$$D_n(\lambda) = \int\cdots\int \begin{vmatrix} S(x_1,x_1,\lambda) & S(x_1,x_2,\lambda) & \ldots\ldots & S(x_1,x_n,\lambda) \\ S(x_2,x_1,\lambda) & S(x_2,x_2,\lambda) & \ldots\ldots & S(x_2,x_n,\lambda) \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ S(x_n,x_1,\lambda) & S(x_n,x_2,\lambda) & \ldots\ldots & S(x_n,x_n,\lambda) \end{vmatrix} dx_1\, dx_2 ... dx_n.$$

The series (5) is uniformly convergent for $|\lambda| \le \rho - \varepsilon$, where ε is a positive number and for all finite values of α, see [1]. Consequently it is an analytic function of two variables, and the regions of convergence are the whole α-plane and the circle $|\lambda| < \rho$.

If we expand $D(\alpha,\lambda)$ by powers of $\alpha - \alpha_i(0)$ and λ and take into account that $\alpha_i(0)$ is a simple root of the equation $D(\alpha,0) = 0$, we will find that the constant term of the expansion vanishes, but the coefficient of the term $[\alpha - \alpha_i(0)]$ is nonzero. According to the theorem about implicit functions, we can state that within some convergence circle the function $\alpha_i(\lambda)$, i.e., the root of the equation $D(\alpha,\lambda) = 0$ which coincides with $\alpha_i(0)$ at $\lambda = 0$, can be expanded into series by powers of λ. Thus the eigenvalues are analytic functions of the perturbation parameter λ, provided $\alpha = 0$

is not an eigenvalue of equation (1′). The last condition, however, is not essential. Indeed, assuming $\alpha = 0$ is an eigenvalue of (1′), let us denote by k the nearest eigenvalue by module. Then we consider the equation

$$(py')' - \left(q + \frac{k}{2}\right) y + \beta y = 0, \tag{6}$$

for which β is not an eigenvalue. The eigenvalues of the equation

$$(py')' - \left(q + \frac{k}{2}\right) y - \lambda r y + \beta y = 0 \tag{7}$$

are analytic functions of λ in the vicinity of $\lambda = 0$. However, the eigenvalues of (7) and (1) differ merely by a constant $\frac{k}{2}$. Therefore, they are also analytic functions of λ.

We have proved that for any value of λ the eigenvalues of equation (1) are analytic functions of λ.

§**2.** The same reasoning shows that $D(x, \xi, \alpha, \lambda)$ (Fredholm's numerator) is also an analytic function of α and λ in the whole α-plane and in some circle $|\lambda| < \rho$.

We denote the normalized eigenfunctions of equation (1′) by

$$\varphi_1(x, \lambda), \varphi_2(x, \lambda), \dots . \tag{8}$$

It is well known that the products $\varphi_i(x, \lambda)\, \varphi_i(\xi, \lambda)$ are residues of the resolvent

$$\Gamma(x, \xi; \alpha, \lambda) = \frac{D(x, \xi, \alpha, \lambda)}{D(\alpha, \lambda)}$$

at the point $\alpha = \alpha_i(\lambda)$. Thus we have

$$\varphi_i(x, \lambda)\, \varphi_i(\xi, \lambda) = \frac{1}{2\pi i} \int_C \Gamma(x, \xi; \alpha, \lambda)\, d\alpha, \tag{9}$$

where C is a curve on the α-plane which encircles the point $\alpha_i(\lambda)$, but no other $\alpha_j(\lambda)$ $(j \neq i)$.

When λ belongs to the region $|\lambda| < a < \rho$ where a is a positive number to be selected later, each eigenvalue remains in some region B_i. It is easy to see that for sufficiently small a no region B_j has points in common with

B_i $(j \neq i)$. This follows from the following two facts: 1) in the case of limited changes of λ all variations of the eigenvalues $\alpha_i(\lambda)$ are uniformly bounded, and 2) if $\alpha_1(\lambda), ..., \lambda_N(\lambda)$ are the first N eigenvalues that are analytic functions of λ, we can take N so large that $\alpha_{N+1}(\lambda), ...$ for every $|\lambda| < \rho$ remain greater than $\alpha_i(\lambda)$ for the same λ.

Since $\alpha_i(0)$ are all distinct, we can take a so small that for $|\lambda| < a$ all $\alpha_i(\lambda)$ $(i = 1, ..., N)$ are regular and the regions B_i are pairwise disjoint. Now we can choose C in such a way that it encircles B_i, but does not encircle any point of B_j $(j \neq i)$ (see Fig. 1). Formula (9) then shows that for sufficiently small λ, the function $\varphi_i(x, \lambda)\,\varphi_i(\xi, \lambda)$ depends analytically on λ. From this we can conclude that $\varphi_i(x, \lambda)$ is also an analytic function of λ.

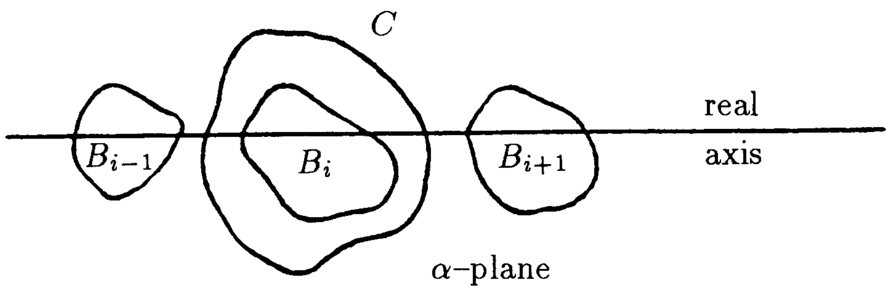


Fig.1.

For further reasoning the expressions of the perturbed eigenvalues are essential. We write here only the first three terms:

$$\alpha_i(\lambda) = \alpha_i(\lambda_0)+(\lambda-\lambda_0)\,\varepsilon_{ii}(\lambda_0)+(\lambda-\lambda_0)^2 \sum_{j=1}^{\infty}{}' \frac{\varepsilon_{ij}(\lambda_0)}{\alpha_i(\lambda_0) - \alpha_j(\lambda_0)} + \cdots, \quad (10)$$

where $\sum'$ denotes summation with the case $i = j$ excluded, and

$$\varepsilon_{ij}(\lambda_0) = \int_0^{\pi} r(x)\varphi_i(x, \lambda_0)\varphi_j(x, \lambda_0)\, dx. \quad (11)$$

§**3.** Let us now assume for a moment that the equation

$$\mu \frac{d^2\varphi}{dx^2} - r(x)\,\varphi + \alpha\varphi = 0 \quad (12)$$

has the same system of eigenvalues as the equation

$$\kappa \frac{d^2\varphi}{dx^2} + \alpha\varphi = 0 \tag{12'}$$

when the boundary conditions are $\varphi'(0) = \varphi'(\pi) = 0$.

Then of necessity, $\mu = \kappa$. This follows from the comparison of the asymptotic expressions for the eigenvalues

$$\alpha_n = n^2\mu + O(1) \quad \text{of (12) and} \quad \alpha_n = n^2\kappa + O(1) \quad \text{for (12')}.$$

Let us write the equations (12), (12′) for the case $\lambda = 0$ and $\lambda = 1$

$$\kappa \frac{d^2\varphi_i(x,0)}{dx^2} + \alpha\varphi_i(x,0) = 0,$$

$$\kappa \frac{d^2\varphi_i(x,1)}{dx^2} - r(x)\,\varphi_i(x,1) + \alpha_i\varphi_i(x,1) = 0.$$

We multiply the first of these equation by $\varphi_i(x,1)$, the second by $\varphi_i(x,0)$, subtract and integrate the result. Then according to the Green formula we obtain

$$\int r(x)\,\varphi_i(x,0)\,\varphi_i(x,1)\,dx = 0. \tag{13}$$

For large values of i we have asymptotic formulas:

$$\varphi_i(x,0) = \sqrt{\frac{2}{\pi}}\cos ix + O\left(\frac{1}{i}\right),$$

$$\varphi_i(x,1) = \sqrt{\frac{2}{\pi}}\cos ix + O\left(\frac{1}{i}\right),$$

yielding the asymptotic expression

$$\varphi_i(x,0)\,\varphi_i(x,1) = \frac{2}{\pi}\cos^2 ix + O\left(\frac{1}{i}\right) = \frac{1}{\pi}\left[1 + \cos 2ix\right] + O\left(\frac{1}{i}\right).$$

Now from

$$\lim_{i\to\infty}\int_0^\pi r(x)\,\cos 2ix\,dx = 0 \quad \text{and} \quad \lim_{i\to\infty}\int_0^\pi r(x)\,O\left(\frac{1}{i}\right)\,dx = 0$$

and from (13) we conclude that

$$\lim_{i\to\infty} \frac{1}{\pi} \int r(x)\,\varphi_i(x,0)\,\varphi_i(x,1)\,dx = \frac{1}{\pi} \int r(x)\,dx = 0.$$

However, since $\varphi_1(x,0) = 1/\sqrt{\pi}$, we can write the expansion of $\alpha_1(\lambda)$ for $\lambda_0 = 0$ according to (10) in the form

$$\alpha_1(\lambda) = -\lambda^2 \sum_{i=2}^{\infty} \frac{\varepsilon_{1j}^2}{\alpha_j(0)} + \dots . \tag{14}$$

From this we conclude that for sufficiently small λ, the value of $\alpha_1(\lambda)$ is negative.

Differentiating (14) we see that the derivative of $\alpha_1(\lambda)$ for sufficiently small positive values of λ is negative. But we have already adopted that $\alpha_1(0) = \alpha_1(1) = 0$. Therefore, $\alpha_1'(\lambda)$ somewhere in the interval $(0,1)$ is positive and changes its sign.

Let $\lambda = \delta$ be the point where $\alpha'(\delta) = 0$. Since $\alpha_1'(0) = 0$ we find that at some point δ_1 the second derivative must vanish.

According to (10), this means

$$\sum_{j=2}^{\infty} \frac{\varepsilon_{1j}^2(\delta_1)}{\alpha_1(\delta_1) - \alpha_j(\delta_1)} = 0.$$

Since all terms here are negative, we obtain

$$\varepsilon_{1j}^2(\delta_1) = 0, \quad \varepsilon_{1j}(\delta_1) = 0 \quad (j \neq 1). \tag{15}$$

However, according to (11) ε_{1j} are the coefficients of expansion of the function $q(x)\,\varphi(x_i,\delta_1)$ by the series of orthogonal functions $\varphi_j(x,\delta_1)$ $(j = 1, 2, \dots)$. The system is complete and, therefore, it follows from (15) that

$$r(x)\,\varphi_1(x,\delta_1) = C\varphi_1(x,\delta_1)$$

or $r(x) = C$.

On the other hand $\int_0^{\pi} r\,dx = 0$ implying $C = 0$. It follows that $r(x) = 0$.

I express my deep gratitude to Professor V. I. Smirnoff for his valuable advice during this work.

REFERENCES

1. R. Courant und D. Hilbert, *Methoden der Mathematischen Physik*, 126, 1924.

21 December 1928 Pulkovo Observatory

THE EXCITATION OF METASTABLE STATES IN GASEOUS NEBULAE

There are two different types of atomic processes which are responsible for the excitation of metastable states in the gaseous nebulae: the fluorescence phenomenon and electronic collisions.

The fluorescence phenomenon. We consider an atom which has three stationary levels 1, 2, 3 with the energies $\varepsilon_1 < \varepsilon_2 < \varepsilon_3$. Let n_1, n_2, n_3 be the number of atoms in cubic centimeters in corresponding levels. The relative values of these numbers in the case of stationary distribution are determined by the radiation field and atomic constants (transition probabilities). The stationarity conditions are:

$$\begin{aligned} &B_{12}\left(n_1 - \frac{g_1}{g_2}\cdot n_2\right)\rho_{12} + B_{13}\left(n_1 - \frac{g_1}{g_3}\cdot n_3\right)\rho_{13} - \\ &n_2\frac{g_1}{g_2}B_{12}\,\sigma_{12} - n_3\frac{g_1}{g_3}B_{13}\,\sigma_{13} = 0, \\ &B_{13}\left(n_1 - \frac{g_1}{g_3}\cdot n_3\right)\rho_{13} + B_{23}\left(n_2 - \frac{g_2}{g_3}\cdot n_3\right)\rho_{23} - \\ &n_3\left\{\frac{g_1}{g_3}B_{13}\,\sigma_{13} + \frac{g_2}{g_3}\sigma_{23}\,B_{23}\right\} = 0, \end{aligned} \tag{1}$$

where B_{ik} is Einstein's probability coefficient corresponding to the transition $i \to k$, g_k is the weight of the k-th level, ρ_{ik} is the density of radiation in the frequency

$$\nu_{ik} = \frac{\varepsilon_k - \varepsilon_i}{h}$$

and

$$\sigma_{ik} = \frac{8\pi\, h\nu_{ik}^3}{c^3}, \tag{2}$$

h, c and π have their usual meaning.

Thus the product

$$A_{ki} = B_{ik}\,\sigma_{ik}\frac{g_i}{g_k} \tag{3}$$

is Einstein's probability coefficient of spontaneous transition $k \to i$.

We write equations (1) to become

$$B_{12}\,\rho_{12} + B_{13}\,\rho_{13} = \frac{g_1}{g_2}\,B_{12}\,(\sigma_{12}+\rho_{12})\,\frac{n_2}{n_1} + \frac{g_1}{g_3}\,B_{13}\,(\sigma_{13}+\rho_{13})\,\frac{n_3}{n_1},$$
$$B_{13}\,\rho_{13} = \tag{4}$$
$$-\,B_{23}\,\rho_{23}\frac{n_2}{n_1} + \left[\frac{g_1}{g_3}\,B_{13}\,(\sigma_{13}+\rho_{13}) + \frac{g_2}{g_3}\,B_{23}\,(\sigma_{23}+\rho_{23})\right]\frac{n_3}{n_1}.$$

Before solving these equations we make some simplifications, corresponding to the physical conditions in nebulae. The radiation density ρ_{ik} may be represented as

$$\rho_{ik} = W\frac{\sigma_{ik}}{\exp\left(\frac{h\nu_{ik}}{kT}\right)-1}. \tag{5}$$

Here T is the surface temperature of the nucleus, and the factor W is defined by the relation:

$$W = \frac{1}{4}\left(\frac{r_*}{r_n}\right)^2, \tag{6}$$

where r_* is the radius of the nucleus and r_n is the distance of the point of the nebula we consider from the nucleus. If W is a small quantity ($W < 10^{-3}$) the densities ρ_{ik} in the brackets of (4) may be neglected, compared with σ_{ik} and we have:

$$B_{12}\,\rho_{12} + B_{13}\,\rho_{13} = \frac{g_1}{g_2}\,B_{12}\,\sigma_{12}\,\frac{n_2}{n_1} + \frac{g_1}{g_3}\,B_{13}\,\sigma_{13}\,\frac{n_3}{n_1},$$
$$B_{13}\,\rho_{13} = -B_{23}\,\rho_{23}\frac{n_2}{n_1} + \left(\frac{g_1}{g_3}\,B_{13}\,\sigma_{13} + \frac{g_2}{g_3}\,B_{23}\,\sigma_{23}\right)\frac{n_3}{n_1}. \tag{7}$$

Solving these equations we obtain:

$$\frac{n_2}{n_1} = \frac{B_{12}\,\rho_{12}\,(g_1 B_{13}\sigma_{13} + g_2 B_{23}\sigma_{23}) + g_2\,B_{13}B_{23}\rho_{13}\sigma_{23}}{\frac{g_1}{g_2}\,B_{12}\,\sigma_{12}\,(g_1 B_{13}\sigma_{13} + g_2 B_{23}\sigma_{23}) + g_1\,B_{13}B_{23}\sigma_{13}\rho_{23}}. \tag{8}$$

We suppose that the second level is a metastable one, i.e., that the quantity B_{12} is smaller than B_{13} and B_{23}. Therefore, the members containing the factor $B_{12}\,\rho_{12}$ may be neglected compared with the term containing $B_{13}\,\rho_{13}$. Expression (8) then becomes:

$$\frac{n_2}{n_1} = \frac{g_2\,B_{13}B_{23}\rho_{13}\sigma_{23}}{\frac{g_1}{g_2}\,B_{12}\,\sigma_{12}\,(g_1 B_{13}\sigma_{13} + g_2 B_{23}\sigma_{23}) + g_1\,B_{13}B_{23}\sigma_{13}\rho_{23}}. \tag{9}$$

We may write

$$\rho_{ik} = W\sigma_{ik}\,\overline{\rho}_{ik} \quad \text{where} \quad \overline{\rho}_{ik} = \frac{1}{\exp\left(\frac{h\nu_{ik}}{kT}\right) - 1}. \tag{10}$$

Then

$$\frac{n_2}{n_1} = \frac{g_2\overline{\rho}_{13}\,W}{\frac{g_1^2}{g_2}\frac{B_{12}}{B_{23}}\frac{\sigma_{12}}{\sigma_{23}} + g_1\frac{B_{12}}{B_{13}}\frac{\sigma_{12}}{\sigma_{13}} + g_1\overline{\rho}_{23}\,W}. \tag{11}$$

Neither the transition $3 \to 1$, nor the transition $3 \to 2$ is forbidden. Therefore, the quantities B_{13} and B_{23} will be of the same order of magnitude, and the ratios $\frac{B_{12}}{B_{13}}$ and $\frac{B_{12}}{B_{23}}$ are small quantities of the same order of magnitude.

Case I. $W < \frac{B_{12}}{B_{13}}, \frac{B_{12}}{B_{23}}$. In this case the last term in the denominator may be neglected ($\overline{\rho}_{23}$ is ordinarily of the order of unity), and we have

$$\frac{n_2}{n_1} = \frac{g_2}{g_1}W\frac{\overline{\rho}_{13}}{\frac{g_1}{g_2}\frac{B_{12}}{B_{23}}\frac{\sigma_{12}}{\sigma_{23}} + \frac{B_{12}}{B_{13}}\frac{\sigma_{12}}{\sigma_{13}}}.$$

In order to estimate the order of magnitude of $\frac{n_2}{n_1}$ we may put: $g_1 = g_2$; $B_{23}\,\sigma_{23} = B_{13}\,\sigma_{13}$. We get

$$\frac{n_2}{n_1} \cong W\frac{B_{13}\,\sigma_{13}}{2\,B_{12}\,\sigma_{12}}\cdot\overline{\rho}_{13}. \tag{12}$$

If on the other hand the second level is not metastable (ordinary level) and B_{12} is of the same order of magnitude as B_{13} and B_{23}, we may neglect the last term in the denominator of (8) and write

$$\left(\frac{n_2}{n_1}\right)_{\text{ord}} = \frac{B_{12}\,\sigma_{12}\,(g_1B_{13}\sigma_{13} + g_2B_{23}\sigma_{23})\;W\overline{\rho}_{12} + g_2\,B_{13}B_{23}\sigma_{13}\sigma_{23}\;W\overline{\rho}_{13}}{\frac{g_1}{g_2}\,B_{12}\,\sigma_{12}\,(g_1B_{13}\sigma_{13} + g_2B_{23}\sigma_{23})}. \tag{13}$$

When estimating the order of magnitude we may put: $g_1 = g_2$; $B_{12}\,\sigma_{12} = B_{13}\,\sigma_{13} = B_{23}\,\sigma_{13}$. Then

$$\left(\frac{n_2}{n_1}\right) = W\overline{\rho}_{12} + \frac{1}{2}\,W\,\overline{\rho}_{13}. \tag{14}$$

The main difference between (12) and (14) is the presence in (12) of a large factor $\frac{B_{13}\,\sigma_{13}}{B_{12}\,\sigma_{12}}$. Therefore, we may assert that in our case the number of atoms in the metastable state is $\frac{B_{13}\,\sigma_{13}}{B_{12}\,\sigma_{12}}$ times larger than in any ordinary excited state.

Case II. $W > \frac{B_{12}}{B_{13}}, \frac{B_{12}}{B_{23}}$. In this case the first two numbers in the denominator of (11) may be neglected. Therefore, we have

$$\frac{n_2}{n_1} = \frac{g_2}{g_1}\frac{\bar{\rho}_{13}}{\bar{\rho}_{23}} = \frac{g_2}{g_1}\frac{\exp\left(\frac{h\nu_{23}}{kT}\right) - 1}{\exp\left(\frac{h\nu_{13}}{kT}\right) - 1}. \tag{15}$$

If $h\,\nu_{13} > k\,T$ we obtain

$$\frac{n_2}{n_1} = \frac{g_2}{g_1}\left(\exp\left(-\frac{h\nu_{12}}{kT}\right) - \exp\left(-\frac{h\nu_{13}}{kT}\right)\right). \tag{16}$$

If at the same time $h\,\nu_{23} > k\,T$ and $h\,\nu_{12} > k\,T$

$$\frac{n_2}{n_1} = \frac{g_2}{g_1}\exp\left(-\frac{h\nu_{12}}{kT}\right). \tag{17}$$

In this case the ratio $\frac{n_2}{n_1}$ is approximately defined by Boltzmann's law. This result was already obtained by Rosseland [1].

The physical meaning of the above is the following: The first two terms in the denominator of (11) correspond to the forbidden transition from the metastable state to the normal state. The last term in the denominator of (11) corresponds to the transitions from the metastable state to the higher states. These transitions are stimulated by the corresponding radiation. In the first case the forbidden transitions are predominant. The forbidden line will appear then in full strength. In the second case the stimulated transitions to the higher levels are predominant, and, if W is sufficiently large, the relative number of forbidden transitions will be very small, and the forbidden line will disappear. Briefly, in the second case the density of radiation will be large enough to make the collisions of the metastable atoms with light-quanta sufficiently frequent.

There may be some doubt as to the possibility of application of our formulas to the gaseous nebulae because in these nebulae photoelectric ionization plays a far more important role than line excitation. But we may treat the ionized atom as an atom in the energy level with very large weight g_3, and the continuous spectrum behind the head of the principal series of the atoms as a very wide spectral line. In fact, the quantity B_{13}, determined from the condition that $n_1 B_{13} \rho_{13}$ is the number of atoms ionized per second, will be of the same order of magnitude as the B-coefficients for the first lines of the principal series. We remark that on account of the large optical thickness of the nebula in ordinary lines of the principal series, the radiation of the nucleus in these lines will be absorbed in the inner layers of the nebula, and, therefore, the first member of (14) vanishes while the second member remains nearly unchanged since the optical thickness in the continuous spectrum is about 10^4 times smaller than in the ordinary lines of the principal series. Hence

$$\left(\frac{n_2}{n_1}\right)_{\text{ord}} = \frac{1}{2} \cdot W \overline{\rho}_{13} = \frac{1}{2} \frac{W}{\exp\left(\frac{h\nu_{13}}{kT}\right) - 1} \tag{18}$$

when the second level is not a metastable one.

The author's observations [2] are in good agreement with this formula. Expression (18) shows that our assertion that in Case I the number of atoms in the metastable state is $\dfrac{B_{13}\,\sigma_{13}}{B_{12}\,\sigma_{12}}$ times larger than in any ordinary excited state must be satisfied more exactly than expected.

Applications to the gaseous nebulae. We have

$$A_{ki} = \frac{g_i}{g_k} \cdot B_{ik}\, \sigma_{ik}.$$

For the first lines of each principal series A_{ki} is of the order 10^8 sec^{-1} if the corresponding transition is not forbidden. Taking $g_i = g_k$ we obtain for these lines

$$B_{ik} = \frac{10^8}{\sigma_{ik}}.$$

As we have mentioned above, the quantity B_{13} corresponding to the bound–free transitions will be of this order of magnitude. In the first case it will be

$$W < \frac{B_{12}}{B_{13}} \quad \text{or} \quad W < \frac{B_{12}\,\sigma_{13}}{10^8},$$

or introducing $B_{12} = \frac{A_{21}}{\sigma_{12}} \cdot \frac{g_2}{g_1}$

$$W < \frac{A_{21}}{10^8} \cdot \frac{\sigma_{13}}{\sigma_{12}} \cdot \frac{g_2}{g_1} = 10^{-8} \frac{g_2}{g_1} \cdot \left(\frac{\nu_{13}}{\nu_{12}}\right)^3 .$$

The quantity $\frac{g_2}{g_1} \cdot \left(\frac{\nu_{13}}{\nu_{12}}\right)^3$ is usually of the order of unity and we find

$$W < 10^{-8} A_{21}.$$

In the planetaries and diffuse nebulae, W is of the order 10^{-14}. Hence

$$\tau_2 = \frac{1}{A_{21}} < 10^6 \text{ sec.},$$

where τ_2 is the mean lifetime of the metastable state.

Thus, if the mean lifetime of the metastable state is shorter than a week, the conditions of Case I are fulfilled. Only when the mean lifetime of the given state is larger than 10^6 sec will the formulas of Case II be applicable. As examples we shall consider the following metastable levels: the states $2\,S$ of H, $2\,S$ of He^+, $2^1\,S$ of parhelium and the state $2^3\,S$ of orthohelium. The first three of these are metastable because the only possible transition of the type

$$2\,S \longrightarrow 1\,S$$

is "forbidden" as a transition between two even states. The last state $2^3\,S$ of orthohelium is metastable because the only possible transition of the type

$$2^3\,S \longrightarrow 1^1\,S$$

is forbidden not only as a transition from one even state to another but also as an intercombination between orthohelium and parhelium levels. The metastability of $2^3\,S$ of He will therefore be of a higher degree than the metastability of the first three levels.

If we suppose that the mean lifetime for the first three types is of the order of 1 sec or 10 sec, i.e., of the same order as the mean lifetime of the levels corresponding to the "nebulium" radiation, the formulas of Case I will be applicable. The ratio $\frac{n_2}{n_1}$ will be for these states 10^8 or 10^9 times larger than the same ratio for ordinary lines.[1]

Only for the level $2^3\,S$ of He may we expect such a long mean lifetime that Case II may occur. A large proportion of He atoms will then be in the state 2^3S and in favorable conditions a considerable optical depth of the nebula in the corresponding series may arise.

Application to the Wolf-Rayet stars. To determine which of our two cases is realized in the gaseous envelope surrounding a Wolf-Rayet star, knowledge of W is required. We have no data about this subject, but it seems that W will be scarcely smaller than 10^{-8}, or perhaps larger. We know, indeed, that during the month after an outburst, the Novae develop many features of the Wolf-Rayet Spectrum. Taking the velocity of the expansion of the gaseous envelope 1000 km/sec and the radius of the star after the ejection of gases 10^6 km, we obtain for W at the end of the month the value $0.5 \cdot 10^{-7}$. For such value of W formula (16) will be applicable to the levels with mean lifetime longer than 10^{-1} sec. Some of the metastable states will have longer mean lifetimes. Such is undoubtedly the state 2^3S of orthohelium. The accumulation of atoms in this state may cause considerable optical depth in the lines of the principal series of orthohelium. The number of atoms in the state 2^3S of He per square centimeter of the surface of the envelope may be estimated in the following manner.

Not all quanta able to ionize the normal He atom emitted by the central star are absorbed by the gaseous envelope, because in the opposite case at the temperatures of the Wolf-Rayet stars the lines of He would be much stronger than the lines of He^+. To explain the observed intensities of He lines, let us suppose that about 1% of the mentioned quanta are absorbed. The optical thickness of the gaseous envelope for the frequencies lying behind the frequency of ionization of normal He then will be about 0.01. If the

[1]Observations have shown that the number of excited atoms in the ordinary excited states of hydrogen is of the order 10^4 per square centimeter of the nebular disk. The number of hydrogen atoms in the state $2\,S$ will therefore be 10^{12} or 10^{13} per square centimeter, and the optical thickness of the nebula in the first two lines of the Balmer series may reach 0.1 or 1.

absorption coefficient per He atom is of the same order as the absorption coefficient behind the head of the Lyman series of H, the number of normal He atoms per square centimeter will therefore be about $2 \cdot 10^{15}$. Applying formula (16) we find that the number of atoms in the state 2^3S of He per square centimeter of the surface of the envelope will be about 10^{14}. Such number of atoms will produce considerable optical thickness of the envelope in the lines of the principal series of orthohelium.

....

Collisional excitation. Now we consider an atom which has only two levels 1 and 2 with the energies ε_1 and ε_2. We suppose that the collisions of the first kind may excite some nebular atoms to the metastable state. The atom may after that pass into a normal state either spontaneously emitting a quantum of the forbidden line or transmitting the energy of the excitation to a free electron. All other types of transitions are neglected. Let $b_{12}\,dt$ be the probability of an inelastic collision which excites the normal atom and a_{21} be the probability of the transition of an excited atom in the normal state by means of a superelastic collision. The condition of stationarity will have the form:

$$b_{12}\,n_1 - (A_{21} + a_{21})\,n_2 = 0. \tag{19}$$

When the velocity distribution of electrons obeys Maxwell's law, we have

$$b_{12} = \frac{g_1}{g_2} \cdot a_{21} \exp\left(-\frac{h\nu_{12}}{kT}\right).$$

Expression (19) becomes

$$\frac{n_2}{n_1} = \frac{g_2}{g_1}\frac{a_{21}}{A_{21} + a_{21}} \cdot \exp\left(-\frac{h\nu_{12}}{kT}\right).$$

If $A_{21} > a_{21}$, i.e., if the density of electrons is low, the ratio $\frac{n_2}{n_1}$ is smaller than

$$\frac{g_2}{g_1} \cdot \exp\left(-\frac{h\nu_{12}}{kT}\right).$$

In this case the spontaneous transitions are predominant and the forbidden lines appear in their full strength. If $A_{21} < a_{21}$, the forbidden lines will be weakened or will disappear altogether.

REFERENCES

1. S. Rosseland, *Astrophysik.*
2. *Zeitschrift für Astrophysik*, 6, 107, 1933.

Yerevan
Astronomical Observatory of the Academy
of Sciences of Armenia

ON THE RADIATIVE EQUILIBRIUM OF A PLANETARY NEBULA[1]

The excellent works of Hubble, Bowen and Zanstra have solved at least qualitatively the problem of the origin of nebular luminosity and of nebular spectra. Carroll and Cillié have made an attempt to compute the relative intensities of the members of the Balmer series of hydrogen in nebular spectra, opening the door to theoretical interpretation of modern spectrophotometric data. However, the dynamics of nebulae as well as the nature and the origin of forces acting in them remain unknown. It is now scarcely possible to answer these questions and to form a complete theory of planetary nebulae.

Selective radiation pressure, owing to specific nebular conditions, plays a very important part in the nebulae. Perhaps radiation pressure is in this case greater than any other force. Computation of radiation pressure is therefore a matter of considerable interest. This computation can be based on the analysis of the field of radiation. An approximate analysis can be done without the solution of other problems connected with planetary nebulae. The purpose of the present paper is to investigate the radiative equilibrium of planetary nebulae. We concentrate on planetary nebulae, though many results may be applied to the diffuse nebulae, as well as to the gaseous envelopes surrounding some stars with emission line spectrum (P Cygni and others). The method applied here was proposed by the author in an earlier paper [1].

---

[1]The first section of a longer paper published in Известия ГАО [*Bulletin De L'Observatoire Central à Poulkovo*], vol. XIII, no. 114, 1933, pp. 1–27. The first version of this section was originally published as "The Radiative Equilibrium of a Planetary Nebula" in *Monthly Notices of the Royal Astronomical Society*, vol. 93, no. 1, 1932 and is used with permission of Blackwell Science Ltd.

The Radiative Equilibrium of a Non-expanding Hydrogen Nebula

We will consider below planetary nebula consisting of hydrogen only. There are some observational data indicating the expansion of planetary nebulae. (Owing to Doppler displacement, the frequencies of a spectral line in different parts of a nebula are different.) This yields an appreciable change in the type of radiative equilibrium.

But in some cases the expansion velocity is so small that the frequency differences of a line for different parts of the nebula are smaller than its Doppler-broadening, caused by the thermal motion of atoms. The behavior of such nebulae is the same as the behavior of a non-expanding nebula, if we confine ourselves to the field of radiation and its interaction with nebular matter.

According to the theory of nebular luminosity developed by Zanstra, all, or at least a considerable part, of the quanta emitted by the central star, which have frequencies greater than ν_0 (the frequency of the limit of the Lyman series), are absorbed by the hydrogen atoms in the nebula. This circumstance requires that the optical thickness of the nebula τ_1 for these frequencies should be larger than unity or at least not much smaller than unity.

Following Zanstra's line of argument, we will show that in the place of each absorbed quantum having a frequency greater than ν_0 there is a certain probability p of re-emission in the same frequency and the probability $1-p$ of re-emission in the line $\mathbf{L}_\alpha$ (first line of the Lyman series of hydrogen). For simplicity, we will call the radiation beyond the head of the Lyman series briefly "ultraviolet radiation" and the corresponding quanta "ultraviolet quanta."

We consider the possible transformations of an ultraviolet quantum, which is emitted from the surface of the central star. It will be absorbed by the nebular envelope, and the absorption will be accompanied by the ionization of a hydrogen atom. After an interval of time the electron that became free will be captured again by a proton. There are two possibilities at this capture: (I) the electron will jump immediately into the deepest level $1S$ (first level); and (II) the electron will jump into one of the excited levels.

In Case I, a new ultraviolet quantum is emitted and the initial state is restored. In Case II, the electron makes a chain of transitions, the last link of which will be a transition into the first level. The dilution of radiation is so great, and the density of matter so low, that interruption of these transitions is very improbable. The last transition into the normal state is accompanied by the emission of a quantum of the Lyman series. There are again two possibilities:

(a) An $\mathbf{L}_\alpha$-quantum is emitted. However, Zanstra's theory requires that the optical thickness of the nebula in ultraviolet light be at least of the order of unity. But the coefficient of line absorption in the Lyman series is some thousand times larger than the absorption coefficient beyond the head of the series. The emitted $\mathbf{L}_\alpha$-quantum, therefore, will be absorbed by a hydrogen atom in the normal state. This atom passes into the second level, and then, owing to the absence of external perturbations during its short lifetime, turns back to the first level, emitting again an $\mathbf{L}_\alpha$-quantum. Thus the $\mathbf{L}_\alpha$-quantum remains unchanged, and we may say that it is merely scattered. These processes may be repeated many times until the quantum reaches the outer boundary of the nebula and escapes.

(b) A quantum of some other line of the Lyman series is emitted. For simplicity we assume that it is an $\mathbf{L}_\beta$-quantum. In this line the optical thickness of the nebula is also very large, and the emitted quantum will be absorbed. This absorption is accompanied by the transition of an atom from the first to the third level. The atom in the third level has two possibilities: it either makes the transition of the type $3 \longrightarrow 2 \longrightarrow 1$, emitting the quanta $\mathbf{H}_\alpha$ and $\mathbf{L}_\alpha$ successively, or it passes immediately into the first level, emitting again $\mathbf{L}_\beta$. In the first case the final product is an $\mathbf{L}_\alpha$-quantum. Its further fate is described above. In the second case the quantum $\mathbf{L}_\beta$ will be absorbed, and thus there exists a finite probability of creation of a quantum $\mathbf{L}_\alpha$. After many absorptions and re-emissions the probability of the survival of an $\mathbf{L}_\beta$-quantum will be very small and the probability of the creation of an $\mathbf{L}_\alpha$-quantum will be practically equal to unity.

In this manner in both cases (a) and (b) the final product is an $\mathbf{L}_\alpha$-quantum. It is easily seen that our considerations may be generalized to the cases where the transitions are accompanied, instead of by the emission of an $\mathbf{L}_\beta$-quantum, by the emission of an $\mathbf{L}_\gamma$, $\mathbf{L}_\delta$, etc. Let p be the probability of Case I and $1-p$ the probability of Case II.

Thus we see indeed that after the absorption of an ultraviolet quantum there is a finite probability p of re-emission of it with the same wavelength[2] and a finite probability $1-p$ of re-emission of quantum $\mathbf{L}_\alpha$. We shall not take into account the intermediate stages in which the absorbed quantum may appear as $\mathbf{L}_\beta$-, $\mathbf{L}_\gamma$-quantum, etc. These will have little influence on the results. The quanta $\mathbf{L}_\alpha$ cannot be transformed and can only be scattered.

In this way the problem is reduced to the study of two superposed fields of radiation: the field of ultraviolet quanta and the $\mathbf{L}_\alpha$-radiation field in the planetary nebulae.

The Field of Ultraviolet Quanta. In this paper we shall use the method of reduction of a spherical problem to a plane problem, developed by Professor Milne. Let k be the absorption coefficient of the ultraviolet radiation per atom. This coefficient depends upon the wavelength. We shall use its mean value. Let, further, n be the number of H atoms in the first level in 1cm^3, r_1 and r_2 the distances of the inner and outer boundary of the nebular ring from the central star. Then the optical depth at the distance r from the central star is

$$\tau = \int_{r_1}^{r_2} nk\, dr. \tag{1}$$

The equations of transfer of the energy of ultraviolet quanta may be written in Milne form

$$\frac{1}{2}\frac{dI(\tau)}{d\tau} = I(\tau) - B(\tau) \tag{2}$$

[2]Owing to the free-free transitions as well as to the inelastic collisions of free electrons the wavelength of the re-emitted quantum may differ considerably from the wavelength of the absorbed quantum. But it always remains shorter than $\lambda_0 = \frac{c}{\nu_0}$, where c is the velocity of light. In this paper we do not distinguish between the ultraviolet quanta of different wavelengths.

$$\frac{1}{2}\frac{dI'(\tau)}{d\tau} = B(\tau) - I'(\tau) \tag{3}$$

if we use an approximation of Schwarzschild-Schuster type.

Here $I(\tau)$ is the average intensity of the diffuse ultraviolet radiation of the nebula in the outward direction at point τ, and $I'(\tau)$ is the average intensity of the same radiation at the same point in the inward direction. The quantity $4\pi\, B(\tau)\, d\tau$ is the amount of energy of ultraviolet quanta emitted in the layer $d\tau$ per second. The same layer absorbs the diffuse ultraviolet radiation from various parts of the nebular ring. The absorbed energy is equal to $2\pi[I(\tau) + I'(\tau)]\, d\tau$. Besides this, the layer absorbs the radiation of the central star. Let πS be the amount of ultraviolet energy falling on each square centimeter of the inner surface of the nebula. At the point τ this amount is reduced to $\pi S \exp(-(\tau_1 - \tau))$, where

$$\tau_1 \equiv \int_{r_1}^{r_2} nk\, dr \tag{4}$$

is the optical thickness of the nebula. From this amount our layer absorbs

$$\pi S \exp(-(\tau_1 - \tau))\, d\tau.$$

Since from the quanta absorbed only the portion p is re-emitted again as ultraviolet quanta, the equation of radiative equilibrium may be written:

$$p\left[I(\tau) + I'(\tau) + \frac{1}{2} \cdot S \exp(-(\tau_1 - \tau))\right] = 2\, B(\tau). \tag{5}$$

Introducing the boundary conditions [2]

$$I'(0) = 0, \quad I(\tau_1) = I'(\tau_1), \tag{6}$$

we take into account the diffuse radiation incident on any portion of the inner face of the nebular envelope and arriving from other portions of the inner face.

From equations (2) and (3) we have

$$\frac{1}{2} \cdot \frac{d(I + I')}{d\tau} = I - I' \tag{7}$$

$$\frac{1}{2} \cdot \frac{d(I - I')}{d\tau} = I + I' - 2B. \tag{8}$$

Differentiating (7) and comparing with (8) we obtain

$$\frac{1}{4} \cdot \frac{d^2(I + I')}{d\tau^2} = I + I' - 2B. \tag{9}$$

Substituting (5) into (9) we find the following equation for $I + I'$:

$$\frac{1}{4} \cdot \frac{d^2(I + I')}{d\tau^2} = (1 - p)(I + I') - \frac{p}{2} \cdot S \exp(-(\tau_1 - \tau)). \tag{10}$$

The general solution of this equation is

$$I + I' = A \exp(-\lambda\tau) + B \exp(\lambda\tau) + \frac{2p}{3 - 4p} S \exp(-(\tau_1 - \tau)), \tag{11}$$

where A and B are constants of integration and $\lambda = 2\sqrt{1 - p}$. Substituting (11) in (5) we find

$$B(\tau) = \frac{p}{a}\left(A \exp(-\lambda\tau) + B \exp(\lambda\tau) + \frac{3}{2(3 - 4p)} \cdot S \exp(-(\tau_1 - \tau))\right). \tag{12}$$

Substituting (11) in (7) we obtain

$$I(\tau) - I'(\tau) = -\frac{\lambda}{2} A \exp(-\lambda\tau) + \frac{\lambda}{2} B \exp(\lambda\tau) + \frac{p}{3 - 4p} \cdot S \exp(-(\tau_1 - \tau)). \tag{13}$$

Adding and subtracting (11) and (13), we find $I(\tau)$ and $I'(\tau)$:

$$I(\tau) = \frac{1}{2}\left(1 - \frac{\lambda}{2}\right) A \exp(-\lambda\tau) + \frac{1}{2}\left(1 + \frac{\lambda}{2}\right) B \exp(\lambda\tau) + \frac{3p}{2(3 - 4p)} \cdot S \exp(-(\tau_1 - \tau)), \tag{14}$$

$$I'(\tau) = \frac{1}{2}\left(1 + \frac{\lambda}{2}\right) A \exp(-\lambda\tau) + \frac{1}{2}\left(1 - \frac{\lambda}{2}\right) B \exp(\lambda\tau) + \frac{p}{2(3 - 4p)} \cdot S \exp(-(\tau_1 - \tau)). \tag{15}$$

The first of conditions (6) may be written according to (15) in the form:

$$A\left(1+\frac{\lambda}{2}\right)+B\left(1-\frac{\lambda}{2}\right)+\frac{p}{3-4p}\cdot S\,\exp\left(-\tau_1\right)=0. \qquad (16)$$

The second of conditions (6) gives

$$\lambda B\exp\left(\lambda\tau_1\right)+\frac{2pS}{3-4p}=\lambda A\cdot\exp\left(-\lambda\tau_1\right). \qquad (17)$$

From equations (16) and (17) we find the following coefficients A and B:

$$A=\frac{\left(1-\frac{\lambda}{2}\right)\exp\left(\tau_1\right)-\frac{\lambda}{2}\exp\left(\lambda\tau_1\right)}{\frac{\lambda}{2}\left[\left(1-\frac{\lambda}{2}\right)\exp\left(-\lambda\tau_1\right)+\left(1+\frac{\lambda}{2}\right)\exp\left(\lambda\tau_1\right)\right]}\cdot\frac{p\cdot S\,\exp\left(-\tau_1\right)}{3-4p}, \qquad (18)$$

$$B=\frac{\left(1+\frac{\lambda}{2}\right)\exp\left(\tau_1\right)+\frac{\lambda}{2}\exp\left(-\lambda\tau_1\right)}{\frac{\lambda}{2}\left[\left(1-\frac{\lambda}{2}\right)\exp\left(-\lambda\tau_1\right)+\left(1+\frac{\lambda}{2}\right)\exp\left(\lambda\tau_1\right)\right]}\cdot\frac{p\cdot S\,\exp\left(-\tau_1\right)}{3-4p}. \qquad (19)$$

These values of A and B in conjunction with (12), (14) and (15) give the solution for the field of the ultraviolet quanta.

For the net flux of the diffuse ultraviolet radiation at the outer boundary of the nebula we obtain:

$$\pi F_u=\pi\left[I(0)-I'(0)\right]=$$
$$\pi\left[1-\frac{2\exp\left(\tau_1\right)-\lambda\sinh\left(\lambda\tau_1\right)}{\left(1-\frac{\lambda}{2}\right)\exp\left(\lambda\tau_1\right)+\left(1+\frac{\lambda}{2}\right)\exp\left(\lambda\tau_1\right)}\right]\cdot\frac{p\cdot S\,\exp\left(-\tau_1\right)}{3-4p}. \qquad (20)$$

For the representation of the solution in numerical form it is necessary to know p and τ_1. The value of p can be calculated from pure physics. Cillié [3] has computed the relative probabilities of the capture of electrons by protons on different hydrogen levels. From his results we have deduced the fraction of captured electrons which pass immediately from a free state to the first level, re-emitting the ultraviolet quanta. This fraction is our p. The value of p depends on the temperature of free electrons. For different temperatures we have:

T	$10,000^\circ$	$20,000^\circ$	$50,000^\circ$
p	0.46	0.49	0.57

Putting $p = 0.5$, we find for large values of τ_1 ($\tau_1 > 3$) the following asymptotic expression for the net flux πF_u at the outer boundary:

$$\pi F_u = 0.7\,\pi S\, e^{-\tau_1}.$$

The net flux of the direct ultraviolet radiation of a star will be simply $\pi S\, e^{-\tau_1}$, while the whole net flux will be $1.7\,\pi S\, e^{-\tau_1}$.

In the absence of the absorbing envelope the net flux from the star is πS. The fraction $1 - 1.7\,e^{-\tau_1}$ of it is converted into other forms of radiation. Owing to the fact that by the splitting of an ultraviolet quantum an $\mathbf{L}_\alpha$-quantum is certainly created, the net flux in the line $\mathbf{L}_\alpha$ at the outer boundary will contain $\dfrac{1 - 1.7\,e^{-\tau_1}}{h\,\nu_c}\mathbf{L}_\alpha$-quanta, where ν_c is the average frequency of the ultraviolet quanta. If τ_1 is large, the flux of the $\mathbf{L}_\alpha$ energy at the outer boundary of nebula is approximately $\dfrac{\nu_\alpha}{\nu_c}\,\pi S$. Here ν_α is the frequency of the line $\mathbf{L}_\alpha$.

The Field of $\mathbf{L}_\alpha$-Radiation. Let κ be the absorption coefficient within the line $\mathbf{L}_\alpha$ per hydrogen atom in the normal state. The optical depth for this line is defined by

$$t = \int_r^{r_2} n\kappa\, dr. \tag{21}$$

The ratio $\dfrac{\kappa}{k} = \omega$ may be assumed constant when the temperature variations within the nebula are neglected. In fact, κ is a function of atomic constants and of the Doppler breadth of the line only. This breadth depends upon the temperature. When $\dfrac{\kappa}{k} = \omega$ is constant, the ratio $\dfrac{t}{\tau}$ is also constant, and we have

$$\frac{t}{\tau} = \frac{\kappa}{k} = \omega. \tag{22}$$

If the temperature of the nebula is of the order $10^3 - 10^4$ degrees, the quantity ω will also be of the order $10^3 - 10^4$. Since we have supposed that the optical thickness τ_1 of the nebula in the ultraviolet region is of the order of unity or larger, the optical thickness in the line $\mathbf{L}_\alpha$,

$$t_1 = \int_{r_1}^{r_2} n\kappa\, dr,$$

will be of the order $10^3 - 10^4$, or larger.

The equation of transfer of the radiation in the line $\mathbf{L}_\alpha$ has the same form as (2) and (3). Let $K(t)$ be the average intensity of the diffuse $\mathbf{L}_\alpha$-radiation of the nebula in the outward direction at point t, and $K'(t)$ be the average intensity of the same radiation at the same point in the inward direction. The equations of transfer are:

$$\frac{1}{2}\frac{dK(t)}{dt} = K(t) - C(t), \tag{23}$$

$$\frac{1}{2}\frac{dK'(t)}{dt} = C(t) - K'(t), \tag{24}$$

where $4\pi C(t)\, dt$ is the amount of energy emitted by the layer dt in the line $\mathbf{L}_\alpha$ per second. This layer absorbs the diffuse $\mathbf{L}_\alpha$-radiation from the other parts of the nebula. The quantity of diffuse radiation absorbed is $2\pi\,[K(t) + K'(t)]\, dt$. The number of $\mathbf{L}_\alpha$-quanta emitted by the central star is negligible, since the number of ultraviolet quanta transformed into $\mathbf{L}_\alpha$-quanta is some thousand times larger.

The number of ultraviolet quanta which are absorbed in the layer dt and are transformed into $\mathbf{L}_\alpha$-quanta is

$$\frac{(1-p)[2\pi(I + I') + \pi\, S \exp\left(-(\tau_1 - \tau)\right)]\, dt}{h\,\nu_c}.$$

Thus the $\mathbf{L}_\alpha$-radiation created in dt according to (5) is

$$\frac{1-p}{p} \cdot \frac{\nu_\alpha}{\nu_c} \cdot 4\pi\, B(\tau)\, d\tau = 4\pi\, \frac{1-p}{p} \cdot \frac{\nu_\alpha}{\nu_c} \cdot B(\tau)\, \frac{dt}{\omega}.$$

Hence the equation of radiative equilibrium is

$$4\pi\, C(t)\, dt = 2\pi\,[K(t) + K'(t)]\, dt + 4\pi\, \frac{1-p}{p} \cdot \frac{\nu_\alpha}{\nu_c} \cdot B(\tau)\, \frac{dt}{\omega},$$

or

$$C(t) = \frac{1}{2}[K(t) + K'(t)] + \frac{\nu_\alpha}{\nu_c} \cdot \frac{1-p}{p\,\omega}\, B(\tau). \tag{25}$$

The boundary conditions are

$$K'(0) = 0, \quad K'(t_1) = K(t_1). \tag{26}$$

From equations (23) and (24) we have

$$\frac{1}{2}\frac{d(K+K')}{dt} = K - K', \tag{27}$$

$$\frac{1}{2}\frac{d(K-K')}{dt} = K + K' - 2C(t). \tag{28}$$

Differentiating (27) and using (28) we find

$$\frac{1}{4}\frac{d^2(K+K')}{dt^2} = K + K' - 2C(t), \tag{29}$$

or according to (25)

$$\frac{1}{4}\frac{d^2(K+K')}{dt^2} = -\frac{2\nu_\alpha}{\nu_c}\cdot\frac{1-p}{p\,\omega}B(\tau). \tag{30}$$

Writing $B(\tau)$ in the form

$$\begin{aligned} B(\tau) &= \frac{p}{2}\left(A\,e^{-\lambda\tau} + Be^{\lambda\tau} + D\exp\left(-(\tau_1-\tau)\right)\right) \\ &= \frac{p}{2}\left(A\,e^{-\frac{\lambda}{\omega}t} + B\,e^{\frac{\lambda}{\omega}t} + D\,e^{-\frac{t_1-t}{\omega}}\right), \end{aligned} \tag{31}$$

where

$$D = \frac{3}{2(3-4p)}\cdot S, \tag{32}$$

we find the following solution of equation (30):

$$K(t) + K'(t) = a + bt - \frac{4\nu_\alpha}{\nu_c}\cdot\frac{1-p}{\lambda^2}\omega\left(A\,e^{-\frac{\lambda}{\omega}t} + B\,e^{\frac{\lambda}{\omega}t} + D\lambda^2\,e^{-\frac{t_1-t}{\omega}}\right),$$

where a and b are constants of integration.

Differentiating this expression and using (27), we find

$$K(t) - K'(t) = \frac{b}{2} - \frac{2\nu_\alpha}{\nu_c}\cdot\frac{1-p}{\lambda}\left(-A\,e^{-\frac{\lambda}{\omega}t} + B\,e^{\frac{\lambda}{\omega}t} + D\lambda\,e^{-\frac{t_1-t}{\omega}}\right).$$

According to the definition of λ,

$$\lambda = 2\sqrt{1-p}.$$

Therefore,

$$K(t) + K'(t) = a + bt - \frac{\nu_\alpha}{\nu_c}\,\omega\,\left(A\,e^{-\frac{\lambda}{\omega}\,t} + B\,e^{\frac{\lambda}{\omega}\,t} + D\lambda^2\,e^{-\frac{t_1 - t}{\omega}}\right), \quad (33)$$

$$K(t) - K'(t) = \frac{b}{2} - \frac{\nu_\alpha}{\nu_c}\cdot\frac{\lambda}{2}\,\left(-A\,e^{-\frac{\lambda}{\omega}\,t} + B\,e^{\frac{\lambda}{\omega}\,t} + D\lambda\,e^{-\frac{t_1 - t}{\omega}}\right). \quad (34)$$

From (33) and (34) we have

$$\begin{aligned} K(t) = \frac{a}{2} + \frac{b}{4} + \frac{b}{2}\,t - \frac{\nu_\alpha}{2\nu_c}\,\omega\,\Bigg[A\left(1 - \frac{\lambda}{2\omega}\right)\,e^{-\frac{\lambda}{\omega}\,t} + \\ B\left(1 + \frac{\lambda}{2\omega}\right)\,e^{\frac{\lambda}{\omega}\,t} + D\lambda^2\left(1 + \frac{1}{2\omega}\right)\,e^{-\frac{t_1 - t}{\omega}}\Bigg], \end{aligned} \quad (35)$$

$$\begin{aligned} K'(t) = \frac{a}{2} - \frac{b}{4} + \frac{b}{2}\,t - \frac{\nu_\alpha}{2\nu_c}\,\omega\,\Bigg[A\left(1 + \frac{\lambda}{2\omega}\right)\,e^{-\frac{\lambda}{\omega}\,t} + \\ B\left(1 - \frac{\lambda}{2\omega}\right)\,e^{\frac{\lambda}{\omega}\,t} + D\lambda^2\left(1 - \frac{1}{2\omega}\right)\,e^{-\frac{t_1 - t}{\omega}}\Bigg]. \end{aligned} \quad (36)$$

The conditions (26) are reduced to

$$\begin{aligned} &\frac{a}{2} - \frac{b}{4} = \\ &\frac{\nu_\alpha}{2\nu_c}\omega\left[A\left(1 + \frac{\lambda}{2\omega}\right) + B\left(1 - \frac{\lambda}{2\omega}\right) + 4D(1-p)\left(1 - \frac{1}{2\omega}\right)e^{-\frac{t_1}{\omega}}\right], \end{aligned} \quad (37)$$

and

$$b = \frac{\nu_\alpha}{\nu_c}\left[-\lambda A\,e^{-\lambda\tau_1} + \lambda B\,e^{\lambda\tau_1} + 4D(1-p)\right]. \quad (38)$$

Substituting (38) in (37) we find:

$$\begin{aligned} &a = \frac{\nu_\alpha}{2\nu_c}\left[-\lambda A\,e^{-\lambda\tau_1} + \lambda B\,e^{\lambda\tau_1} + 4D(1-p)\right] + \\ &\frac{\nu_\alpha}{\nu_c}\omega\left[A\left(1 + \frac{\lambda}{2\omega}\right) + B\left(1 - \frac{\lambda}{2\omega}\right) + 4D(1-p)\left(1 - \frac{1}{2\omega}\right)e^{-\frac{t_1}{\omega}}\right]. \end{aligned} \quad (39)$$

Equations (35) and (36) together with (38) and (39) give the solution for the $\mathbf{L}_\alpha$-field.

The Density of Radiation in the Inner Layers of the Nebula. We have denoted above by πS the amount of energy of the ultraviolet radiation falling from the star on each square centimeter of the inner surface of the nebula. In the absence of re-emission the mean intensity of the ultraviolet radiation in this region will equal $\frac{\pi S}{4\pi} = 0.25\,S$. In the case where re-emission is taken into account, the average intensity of ultraviolet radiation increases and is equal to $\frac{1}{4}\,S + \frac{1}{2}(I_1 + I_2)$. According to (11), (18) and (19),

$$I(\tau_1) + I'(\tau_1) = \frac{2p}{3-4p} \cdot S \left[1 - \frac{1}{\lambda} \cdot \frac{\lambda\, e^{-\tau_1} + \lambda \cosh(\lambda\tau_1) + 2\sinh(\lambda\tau_1)}{2\cosh(\lambda\tau_1) + \lambda\sinh(\lambda\tau_1)}\right]. \tag{40}$$

The expression in brackets remains between 0 and $1 - \frac{1}{\lambda}$, when τ_1 changes between 0 and ∞. Therefore we have

$$\frac{1}{4} \cdot S \leq \frac{1}{4} \cdot S + \frac{1}{2}(I_1 + I_2) \leq \frac{1}{4} \cdot S + \frac{2p\left(1 - \frac{1}{\lambda}\right)}{3-4p} \cdot S.$$

Putting $p = 0.5$, we obtain:

$$\frac{1}{4} \cdot S \leq \frac{1}{4} \cdot S + \frac{1}{2}(I_1 + I_2) < 0.40\,S.$$

Thus the mean intensity of ultraviolet radiation at the inner boundary of nebula is of the same order of magnitude as in the absence of a nebular envelope. It should be doubled if τ_1 is very large.

The state of affairs entirely changes when we consider the $\mathbf{L}_\alpha$-field. Owing to the large optical thickness of the nebula in the $\mathbf{L}_\alpha$-line, and to the fact that all $\mathbf{L}_\alpha$-quanta absorbed are re-emitted in the same frequency, the density of $\mathbf{L}_\alpha$-radiation in the inner layers of the ring is very large.

In order to estimate this density we consider a modification of (38) and (39). In fact, comparing (38) with (17) and (32), we find

$$b = \frac{2\nu_\alpha}{\nu_c} \cdot S, \tag{41}$$

and neglecting in (39) the terms not containing the factor ω

$$a = \frac{\nu_\alpha}{\nu_c}\,\omega \cdot \left[A + B + 4D(1-p)\,e^{-\tau_1}\right]. \tag{42}$$

Substituting the values of A, B and D, we find:

$$a = \frac{\nu_\alpha}{\nu_c}\,\omega \cdot \left[3(1-p)\,e^{-\tau_1} - \frac{1+e^{-\tau_1}\cosh(\lambda\tau_1)}{2\cosh(\lambda\tau_1)+\lambda\sinh(\lambda\tau_1)}\right]\frac{2p}{3-4p}\cdot S. \quad (43)$$

For the brackets in (33) we have

$$A\,e^{-\lambda\tau_1} + B\,e^{\lambda\tau_1} + D\,\lambda^2 = \frac{S}{3-4p}\left[6(1-p) - \frac{2p}{\lambda}\cdot\frac{\lambda\,e^{-\tau_1} + \lambda\,\cosh(\lambda\tau_1) + 2\sinh(\lambda\tau_1)}{2\cosh(\lambda\tau_1)+\lambda\sinh(\lambda\tau_1)}\right]. \quad (44)$$

This expression varies within the limits

$$2S \le A\,e^{-\lambda\tau_1} + B\,e^{\lambda\tau_1} + D\,\lambda^2 < \frac{S}{3-4p}\left[6(1-p) - \frac{2p}{\lambda}\right]$$

or, if $p = 0.5$

$$2S \le A\,e^{-\lambda\tau_1} + B\,e^{\lambda\tau_1} + D\,\lambda^2 \le 2.29\,S,$$

and in the first approximation

$$A\,e^{-\lambda\tau_1} + B\,e^{\lambda\tau_1} + D\,\lambda^2 = 2.15\,S. \quad (45)$$

For the mean intensity of $\mathbf{L}_\alpha$-radiation at the inner boundary, we obtain approximately:

$$\frac{1}{2}\left[K(t_1) + K'(t_1)\right] = \frac{\nu_\alpha}{\nu_c}\,\omega\,S\cdot\left[\tau_1 + \frac{1}{2}f(\tau_1) - 1.07\right],$$

where

$$f(\tau_1) = \left[3(1-p)\cdot e^{-\tau_1} - \frac{1+e^{-\tau_1}\cosh(\lambda\tau_1)}{2\cosh(\lambda\tau_1)+\lambda\sinh(\lambda\tau_1)}\right].$$

If $\tau \ge 2$ we may neglect $\frac{1}{2}f(\tau_1)$ and have approximately

$$\frac{1}{2}\left[K(t) + K'(t)\right] = \frac{\nu_\alpha}{\nu_c}\,\omega\,S\cdot[\tau_1 - 1].$$

We may take $\omega = 10{,}000$ (see [4]). Therefore, if $\tau_1 = 2$ the mean density of the $\mathbf{L}_\alpha$-radiation at the inner boundary will be of the order $10{,}000\,\pi\,S$, where πS is again the energy of the whole ultraviolet radiation falling on each square centimeter of the inner surface of the nebula from the central star. Therefore, the density of L_α-radiation in this example is 40,000 times larger than the density of the whole diluted ultraviolet radiation of the nucleus in the absence of the absorbing shell at the same distance. A rough estimate shows that the ultraviolet radiation of the black body at temperatures of the order $40,000^\circ - 50,000^\circ$ is about $5 \cdot 10^4$ times stronger than the same radiation within the Doppler width of the L_α-line corresponding to the temperature of the nebular matter. Thus the density of L_α-radiation in the inner layers of the nebular envelope will be

$$40,000 \times 5 \cdot 10^4 = 2 \cdot 10^9$$

times larger than the density of the direct L_α-radiation of the central star within the Doppler line-width.

Such a large density of L_α-radiation will produce a large accumulation of atoms in the state $2P$. On the other hand there will also be a large accumulation of hydrogen atoms in the metastable state $2S$. It may happen, therefore, that the optical thickness of the nebula in the lines of the Balmer series will not be very small.

Radiation Pressure in the Outer Parts of the Nebula. The greater part of the ultraviolet radiation of the star is transformed by the nebula into L_α-quanta. Therefore, the flux of radiation emitted by the nebula will consist chiefly of L_α-quanta. For sufficiently large τ_1 each ultraviolet quantum will give rise to an L_α-quantum, and the flow of L_α-radiation from the nebula will be of the order $\frac{\nu_\alpha}{\nu_c} \cdot \pi S$. On the inner surface of the nebula the flux of L_α-radiation will be practically equal to zero, and the flux of ultraviolet radiation will be πS. Now the radiation pressure in a layer of gas is proportional to the absorption coefficient. On the inner surface of the nebula the resulting flux of radiation consists of ultraviolet quanta, for which the absorption coefficient is small. Therefore, the radiation pressure will not be very great. In the outer parts of the nebula, on the contrary, the flux of radiation consists chiefly of L_α-quanta, and the absorption coefficient is about 10^4 times larger than in the case of ultraviolet quanta, while the flux of radiation is of the same order. The radiation pressure, or more exactly, the gradient of radiation pressure, will be here 10^4 times greater than on the inner boundary of the ring. It is physically clear that for large τ_1 the net flux of $\mathbf{L}_\alpha$-radiation πF_α in the outer layer of the ring will be determined by

$$\pi F_\alpha = \left(\frac{r}{r_n}\right)^2 \cdot \frac{2\pi h\nu_\alpha}{c^2} \int_{\nu_0}^{\infty} \frac{\nu^2\,d\nu}{\exp\left(\frac{h\nu}{kT}\right) - 1},$$

where r and r_n are respectively the radius of the central star and the radius of the nebula. The average impulse received by a hydrogen atom in a normal state per second from $\mathbf{L}_\alpha$-quanta will be

$$\frac{\kappa\pi F_\alpha}{c} = \left(\frac{r}{r_n}\right)^2 \cdot \frac{2\kappa\pi h\nu_\alpha}{c^3} \int_{\nu_0}^{\infty} \frac{\nu^2\,d\nu}{\exp\left(\frac{h\nu}{kT}\right) - 1}.$$

The impulse received by each hydrogen atom from the gravitational field of the central star per second is

$$g\left(\frac{r}{r_n}\right)^2 m,$$

where g is the gravitational acceleration on the surface of the central star. However, not only the normal hydrogen atoms but also the free protons are subject to gravitational force. Therefore, the ratio μ of repulsive force R to attractive force G is given by

$$\mu = \frac{R}{G} = \frac{\kappa\,\pi}{mg\left(1 + \frac{n^+}{n_1}\right)} \frac{2h\nu_\alpha}{c^3} \int_{\nu_0}^{\infty} \frac{\nu^2\,d\nu}{\exp\left(\frac{h\nu}{kT}\right) - 1},$$

where n^+ is the number of protons per cubic centimeter, n_1 is the number of normal hydrogen atoms in the same volume. Even in the case when we put

$$\frac{n^+}{n_1} = 500$$

which value is probably too high (see [5]), we obtain for $T = 40,000°$.

$$\mu = \frac{10^{10}}{g}.$$

The value of g for the nuclei of planetary nebulae will be much larger than that for the Sun. But it is very improbable that it may reach 10^{10} cm sec^{-2}. Therefore, we may conclude that if the optical thickness of the nebula in the ultraviolet region is not too small, the radiation pressure will be the dominant factor in the exterior parts of the nebula.

Remark. The essential point was the study of the $\mathbf{L}_\alpha$-field. The second part of the Pulkovo paper was devoted to helium nebulae. It is omitted here because since its appearance 60 years ago more complicated problems have been discussed and solved by other authors (the most important steps have been made by V. V. Sobolev).

REFERENCES

1. Ambartsumian, *M. N.*, 93, 50, 1932.
2. Milne, *Z. f. Astroph.*, 1, 105, 1980.
3. Cillié, *M. N.*, 92, 820, 1932.
4. Ambartsumian, *M. N.*, 93, 58, 1932.
5. Zanstra, *Z. f. Astroph.*, 2, 337, 1931.

Astronomical Observatory
of Leningrad University

ON THE DERIVATION OF THE FREQUENCY FUNCTION OF SPACE VELOCITIES OF THE STARS FROM THE OBSERVED RADIAL VELOCITIES[1]

One of the most important problems of stellar statistics is the derivation of the frequency function of the space velocities of stars of various spectral types and of different absolute magnitudes. The direct solution of this problem requires knowledge of the space velocities of a great number of stars. The derivation of the space velocity of a given star is possible only if three different quantities are measured: the radial velocity, the proper motion and the parallax. These quantities are measurable with different relative degrees of accuracy and are exposed to systematic errors of quite different kinds. For some important groups of stars (for example B-type stars) we have very few reliable individual parallaxes. The number of stars with reliable parallaxes is generally small, and among them the radial velocities are known only for a fraction.

Therefore, several authors have attempted to obtain some knowledge about the distribution law of space velocities from the radial velocities alone. However, in every case some more or less arbitrary form of this law was assumed, and the problem was restricted to finding the numerical values of some parameters entering in this form of distribution law. In the majority of cases these constants are the elements of the velocity ellipsoids.

Owing to the relative uniformity of the catalogues of radial velocities, the results of the statistical investigations based on them are almost free from the influence of systematic errors. It seems desirable, then, to try to solve the problem of derivation of the frequency function of space velocities

[1]Communicated by Sir Arthur Eddington.

Originally published by Blackwell Scientific Publications Ltd. © 1936 in *Monthly Notices of the Royal Astronomical Society*, vol. 96, no. 3, 1936, pp. 172–179. Used here with permission of Blackwell Scientific Publications Ltd.

from the distribution of radial velocities without making any hypothesis about the form of this function.

So far as is known to the author, this problem not only remains unsolved but has not even been discussed in any detail. The purpose of the present paper is to derive a general formula for the frequency function of space velocities from the distribution of radial velocities.

For the frequency function of space velocities we will derive and solve an integral equation. In this equation the observed frequency function of the radial velocities for different parts of the sky enters as the known function.

The Fundamental Assumption. We shall assume that the different elementary volumes of space in our neighborhood have practically identical frequency functions of the space velocities. In reality, for rare types of stars (for example, Cepheids), it is necessary to consider also the distant stars, since the number of stars of such types in our neighborhood is very small. In such cases some corrections for the difference between the frequency functions in various parts of the galaxy are required. These corrections are beyond the scope of the present paper. We assume that radial velocities of a sufficiently large number of near stars situated in different parts of the sky are given, and our aim is to derive the frequency function of the space velocities.

We shall consider first the two-dimensional problem. It is of special interest since some types of stars are strongly concentrated near the galactic plane and the z-components of their velocities are small.

The Two-Dimensional Problem. The stars are distributed over a plane, and we are situated in the same plane. We measure for each star the radial velocity V and its apparent position or azimuth. In the case of stars of high galactic concentration the role of such azimuth is played by the galactic longitude. Let $f(V, \alpha)\, dV\, d\alpha$ be the number of observed stars with azimuths between α and $\alpha + d\alpha$ and with radial velocities between V and $V + dV$. The function $f(V, \alpha)$ can be obtained from the lists of the radial velocities of stars. If, further, $n(\alpha)\, d\alpha$ is the total number of observed stars in the directions between α and $\alpha + d\alpha$, we have

$$n(\alpha) = \int_{-\infty}^{\infty} f(V, \alpha)\, dV.$$

Let $\phi(\xi,\eta)$ be the unknown frequency function of the velocities, i.e., $\phi(\xi,\eta)$ $d\xi\,d\eta$ is the relative number of stars for which the velocity components fall within the limits ξ and $\xi + d\xi$, η and $\eta + d\eta$. We have

$$\int_{-\infty}^{\infty} d\eta \int_{-\infty}^{\infty} \phi(\xi,\eta)\,d\xi = 1.$$

Among the observed stars with azimuths within the interval $(\alpha, \alpha + d\alpha)$, we have $n(\alpha)\,d\alpha\,\phi(\xi,\eta)\,d\xi\,d\eta$ stars, which have velocities within the element $d\xi\,d\eta$ of the "velocity plane" ξ,η.

We may choose the direction of the ξ-axis to have the azimuth $\alpha = 0$. Then it is clear that all stars observed in the azimuth α for which the velocities lie within the strip S of ξ,η plane (see Fig.1) have radial velocities lying between V and $V + dV$. Therefore, from $n(\alpha)\,d\alpha$ stars within the interval $(\alpha, \alpha + d\alpha)$,

$$n(\alpha)\,d\alpha \int_{(S)} \phi(\xi,\eta)\,d\xi\,d\eta$$

stars will have radial velocities between V and $V + dV$. The integration is carried over the strip (S), perpendicular to the direction α and with the width dV.

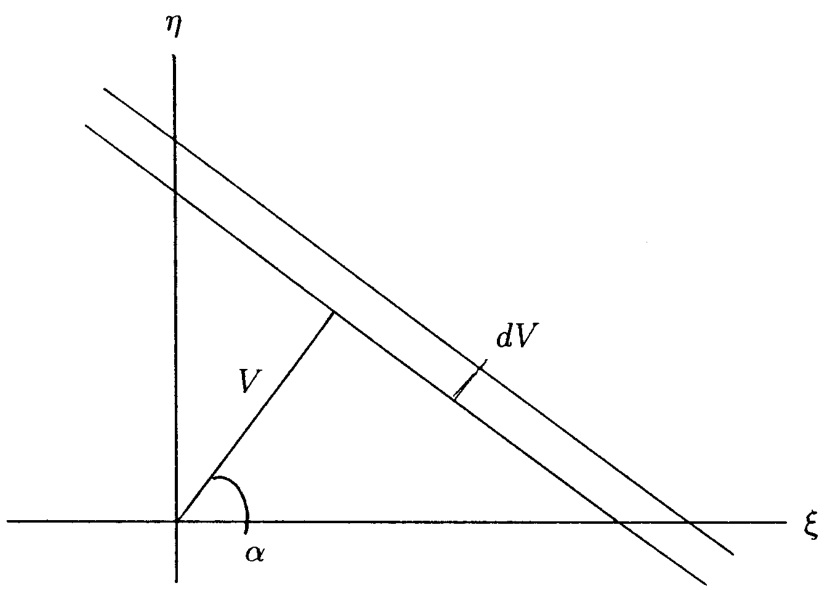


Fig.1

On the other hand, we have denoted the number of such stars as

$$f(V,\alpha)\,dV\,d\alpha.$$

Therefore, we have the equation

$$f(V,\alpha)\,dV = n(\alpha)\int_{(S)}\phi(\xi,\eta)\,d\xi\,d\eta. \tag{1}$$

Let us make a change of variables:

$$\xi' = \xi\cos\alpha + \eta\sin\alpha,$$
$$\eta' = -\xi\sin\alpha + \eta\cos\alpha.$$

It is clear that ξ' within the strip (S) varies between V and $V+dV$, and η' varies between $-\infty$ and ∞. Therefore,

$$\int_{(S)}\phi(\xi,\eta)\,d\xi\,d\eta =$$
$$\int_{V}^{V+dV} d\xi' \int_{-\infty}^{\infty}\phi(\xi'\cos\alpha - \eta'\sin\alpha, \xi'\sin\alpha + \eta'\cos\alpha)\,d\eta'.$$

We divide equation (1) by $n(\alpha)$ and set

$$F(V,\alpha) = \frac{f(V,\alpha)}{n(\alpha)} = \frac{f(V,\alpha)}{\int f(V,\alpha)\,dV}$$

to obtain the equation

$$F(V,\alpha) = \int_{-\infty}^{\infty}\phi(V\cos\alpha - \eta'\sin\alpha, V\sin\alpha + \eta'\cos\alpha)\,d\eta'. \tag{2}$$

The left-hand side of this equation may be obtained from the counts of stars in the radial-velocities lists.

Returning to the old coordinates ξ and η, we may write equation (2) in the form

$$F(V,\alpha) = \int_{(L)}\phi(\xi,\eta)\,ds, \tag{3}$$

where the integration is carried over the straight line (L) whose equation is

$$\xi\cos\alpha + \eta\sin\alpha = V, \tag{4}$$

and ds is an elementary length on this line.

We have come to the following problem:
The value of the integral (3) for every straight line of the ξ, η plane is given as the function of the parameters V and α, defining the straight line. The integrand $\phi(\xi, \eta)$ is to be found.

The solution of this problem is comparatively simple. Let us introduce in both parts of (2) instead of V the expression

$$V = x\cos\alpha + y\sin\alpha + W, \tag{5}$$

where x, y and W are some arbitrary parameters. Then we can rewrite (2) as

$$\begin{aligned}&F(x\cos\alpha + y\sin\alpha + W, \alpha) = \\ &= \int_{-\infty}^{\infty} \phi(x\cos^2\alpha + y\sin\alpha\cos\alpha + W\cos\alpha - \eta'\sin\alpha, \\ &x\cos\alpha\sin\alpha + y\sin^2\alpha + W\sin\alpha + \eta'\cos\alpha)\, d\eta'.\end{aligned}$$

If we introduce the new variable of integration

$$\eta' = U - x\sin\alpha + y\cos\alpha, \tag{6}$$

our equation takes the simple form

$$\begin{aligned}&F(x\cos\alpha + y\sin\alpha + W, \alpha) = \\ &= \int_{-\infty}^{\infty} \phi(x + W\cos\alpha - U\sin\alpha, y + W\sin\alpha + U\cos\alpha)\, dU.\end{aligned} \tag{7}$$

Multiplying both parts by $d\alpha$, integrating between 0 and 2π and changing the order of the integration on the right-hand side, we find

$$\begin{aligned}&\int_0^{2\pi} F(x\cos\alpha + y\sin\alpha + W, \alpha)\, d\alpha = \\ &= \int_{-\infty}^{\infty} dU \int_0^{2\pi} \phi(x + W\cos\alpha - U\sin\alpha, y + W\sin\alpha + U\cos\alpha)\, d\alpha.\end{aligned} \tag{8}$$

Now it is easy to see that the integral

$$\Phi = \int_0^{2\pi} \phi(x + W\cos\alpha - U\sin\alpha, y + W\sin\alpha + U\cos\alpha)\, d\alpha, \tag{9}$$

depends only on x, y and $\sqrt{W^2+U^2}$. In fact, if we introduce in (9)

$$\begin{aligned} W &= G\cos\beta, \\ U &= G\sin\beta, \end{aligned} \tag{10}$$

we obtain

$$\Phi = \int_0^{2\pi} \phi[x + G\cos(\alpha+\beta), y + G\sin(\alpha+\beta)]\, d\alpha,$$

and it is obvious that this integral depends only on x, y and $G = \sqrt{W^2+U^2}$ and is independent of $\beta \equiv \arctan\frac{U}{W}$.

Therefore, we may write simply:

$$\Phi(x,y,G) = \int_0^{2\pi} \phi(x + G\cos\alpha, y + G\sin\alpha)\, d\alpha, \tag{11}$$

and

$$\int_0^{2\pi} F(x\cos\alpha + y\sin\alpha + W, \alpha)\, d\alpha = \int_{-\infty}^{\infty} \Phi(x,y,G)\, dU. \tag{12}$$

However,

$$dU = \frac{G\, dG}{\sqrt{G^2 - W^2}},$$

and we may rewrite (12) in the form

$$\int_0^{2\pi} F(x\cos\alpha + y\sin\alpha + W, \alpha)\, d\alpha = 2\int_W^{\infty} \Phi(x,y,G) \frac{G\, dG}{\sqrt{G^2-W^2}}. \tag{13}$$

This equation is an integral equation of Abel's type for the function

$$\Phi(x,y,G),$$

and its solution is given by

$$\begin{aligned} &\Phi(x,y,G) = \\ &-\frac{1}{\pi}\frac{1}{G}\frac{d}{dG}\int_G^{\infty} \frac{W\, dW}{\sqrt{W^2-G^2}} \int_0^{2\pi} F(x\cos\alpha + y\sin\alpha + W, \alpha)\, d\alpha. \end{aligned} \tag{14}$$

We have, according to (11),

$$\Phi(x,y,0) = 2\pi\phi(x,y), \tag{15}$$

and we may rewrite (14) in the form

$$\phi(x,y) = -\frac{1}{2\pi^2} \lim_{G\to 0} \frac{1}{G}\frac{d}{dG} \int_G^\infty \overline{F}(x,y,W)\,\frac{W\,dW}{\sqrt{W^2-G^2}}, \tag{16}$$

where the function

$$\overline{F}(x,y,W) = \int_0^{2\pi} F(x\cos\alpha + y\sin\alpha + W, \alpha)\,d\alpha \tag{17}$$

may be obtained from the observations. After some algebra we reduce (16) to

$$\phi(x,y) = -\frac{1}{2\pi^2} \int_0^\infty \frac{1}{W}\frac{d\overline{F}(x,y,W)}{dW}\,dW. \tag{18}$$

This equation is the solution of our problem. The numerical calculation of $\overline{F}(x,y,W)$, when $F(V,\alpha)$ is given, may be carried out without difficulty.

We have actually applied our formula to the radial velocities of B-type stars observed in the galactic belt $|b| < 20^\circ$, and the results of this application are in satisfactory agreement with the velocity distribution derived from the direct counts of the known space velocities. The details of this application will be given elsewhere.

The Three-Dimensional Problem. In the case of the three-dimensional problem we may derive from the catalogues the number of stars observed within the given solid angle $d\omega$ in the given direction, having the radial velocity confined within the limits V and $V+dV$. Let us denote this number by $f(V,l,b)\,d\omega$, where l and b are galactic longitude and latitude. If, further, $n(l,b)\,d\omega$ is the total number of observed stars in the same solid angle, we shall have

$$n(l,b) = \int f(V,l,b)\,dV. \tag{19}$$

We have the following relation between the frequency function of the space velocities $\phi(\xi,\eta,\zeta)$ and the observed function $f(V,l,b)$:

$$f(V,l,b)\,dV = n(l,b) \iiint_\Omega \phi(\xi,\eta,\zeta)\,d\xi\,d\eta\,d\zeta, \tag{20}$$

where the integration is extended over the volume (Ω) between two parallel planes in ξ,η,ζ space, which are perpendicular to the direction (l,b) and have the distances V and $V+dV$ from the origin.

Dividing (20) by $n(l, b)$, we bring this equation after some transformations to the form

$$F(V, l, b) = \iint_{\Sigma} \phi(\xi, \eta, \zeta)\, d\sigma, \tag{21}$$

where

$$F(V, l, b) = \frac{f(V, l, b)}{n(l, b)}, \tag{22}$$

and the integration is extended over the plane (Σ), perpendicular to the direction (l, b), at distance V from the origin.

The equation of this plane is:

$$\xi \cos l \cos b + \eta \sin l \cos b + \zeta \sin b = V. \tag{(\Sigma)}$$

In polar coordinates ρ and θ in the plane (Σ) with origin at the point

$$\xi = V \cos l \cos b; \quad \eta = V \sin l \cos b; \quad \zeta = V \sin b,$$

(Σ) is written as

$$\begin{aligned} \xi &= V \cos l \cos b + \rho(\alpha_1 \cos\theta + \beta_1 \sin\theta), \\ \eta &= V \sin l \cos b + \rho(\alpha_2 \cos\theta + \beta_2 \sin\theta), \\ \zeta &= V \sin b + \rho(\alpha_3 \cos\theta + \beta_3 \sin\theta), \end{aligned}$$

where the coefficients $\alpha_1, \alpha_2, \alpha_3$ and $\beta_1, \beta_2, \beta_3$ satisfy the conditions

$$\alpha_1 \cos l \cos b + \alpha_2 \sin l \cos b + \alpha_3 \sin b = 0,$$

$$\beta_1 \cos l \cos b + \beta_2 \sin l \cos b + \beta_3 \sin b = 0,$$

$$\alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3 = 0,$$

$$\alpha_1^2 + \alpha_2^2 + \alpha_3^2 = 1, \qquad \beta_1^2 + \beta_2^2 + \beta_3^2 = 1.$$

Now we can rewrite equation (21) in these coordinates:

$$F(V, l, b) = \int_0^\infty \rho\, d\rho \int_0^{2\pi} \phi(V \cos l \cos b + \rho(\alpha_1 \cos\theta + \beta_1 \sin\theta),$$
$$V \sin l \cos b + \rho(\alpha_2 \cos\theta + \beta_2 \sin\theta), V \sin b + \rho(\alpha_3 \cos\theta + \beta_3 \sin\theta))\, d\theta.$$

Integrating over all directions and changing the order of the integration on the right-hand side, we obtain

$$\int F(V,l,b)\,d\omega = \int \Phi\,\rho\,d\rho, \tag{23}$$

where

$$\Phi = \int d\omega \int_0^{2\pi} \phi(V\cos l\cos b + \rho(\alpha_1\cos\theta + \beta_1\sin\theta),$$
$$V\sin l\cos b + \rho(\alpha_2\cos\theta + \beta_2\sin\theta), V\sin b + \rho(\alpha_3\cos\theta + \beta_3\sin\theta))\,d\theta.$$

In terms of parameters

$$V = G\gamma_1; \qquad \rho\cos\theta = G\gamma_2; \qquad \rho\sin\theta = G\gamma_3; \qquad \gamma_1^2 + \gamma_2^2 + \gamma_3^2 = 1,$$

the integral Φ takes the form

$$\Phi = \int d\theta \int \phi(G(\gamma_1\cos l\cos b + \gamma_2\alpha_1 + \gamma_3\beta_1),$$
$$G(\gamma_1\sin l\cos b + \gamma_2\alpha_2 + \gamma_3\beta_2), G(\gamma_1\sin b + \gamma_2\alpha_3 + \gamma_3\beta_3))\,d\omega,$$

and it may be shown that it depends only on G. We may write

$$\Phi(G) = 2\pi \int \phi(G\cos l\cos b, G\sin l\cos b, G\sin b)\,d\omega; \quad \Phi(0) = 8\pi^2\phi(0,0,0). \tag{24}$$

We now have

$$G^2 = V^2 + \rho^2; \qquad \rho\,d\rho = G\,dG.$$

Therefore,

$$\int F(V,l,b)\,d\omega = \int_V^\infty \Phi\,G\,dG \tag{25}$$

and

$$\Phi = -\frac{1}{V}\frac{d}{dV}\int F(V,l,b)\,d\omega. \tag{26}$$

Comparing (26) with (24) we find

$$\phi(0,0,0) = \frac{1}{8\pi^2}\Phi(0) = -\lim_{V\to 0}\frac{1}{V}\frac{d}{dV}\int F(V,l,b)\,d\omega.$$

Thus we may find $\phi(0,0,0)$. In the same way, after some lengthy algebra we obtain

$$\phi(\xi,\eta,\zeta) = -\frac{1}{8\pi^2}\frac{1}{W}\frac{W}{dW}\int F(\xi\cos l\cos b + \eta\sin l\cos b + \zeta\sin b + W, l, b)\,d\omega.$$

This formula represents the solution of the three-dimensional problem.

Remark. In our method it is assumed that the K-effect is absent. Actually, however, we may determine the K-term from the radial velocities and exclude it.

20 November 1935

Leningrad
Astronomical Observatory of the University

THE STATISTICS OF DOUBLE STARS

It has been indicated by a number of authors that the study of the distribution law of elements of double star orbits, as well as of other statistical interrelations for these objects, can provide interesting results for cosmogony in general and for the age problem of our star system in particular. However, as indicated by the author in the preliminary note [1], wrong conclusions are often drawn from observational data. The aim of this study is to show the erroneous nature of some old conclusions widespread in the literature [2]. We also indicate some new implications from the observational material concerning double stars.

Contrary to Jeans the observed distribution of eccentricities among the double stars with known orbits far from proves the equipartition of energies. Direct consideration of energies of double stars (or large semiaxes of their orbits) proves that equipartition of energy does not occur even among wide pairs. This circumstance, together with the absence of dissociative equilibrium between double and single stars, leads to an upper age bound of 10^{10} years for the ensemble of double stars.

1. Distribution of eccentricities of the double star orbits

The distribution of eccentricities of double star orbits has been discussed often enough. In particular, it was established that among double stars whose orbits have been determined, the number of pairs with eccentricities less than ε is proportional to ε^2.

On the other hand, Jeans has shown that under statistical equilibrium (Boltzman distribution) the same dependence should be observed. From this it is concluded that we already deal with the most probable distribution, and this implies the long time scale. According to Jeans' more careful formulation, given in his response [3] to the author's preliminary note, equipartition exists at least for some parameters.

First of all one should realize that the distribution of eccentricities now at hand can differ to a large degree from the real distribution due to selectivity of observational material.

For the time being we know only the orbits of pairs with comparatively short periods. On the other hand, as observations show, the average eccentricity certainly increases with period. Therefore, the true relative number of pairs with large eccentricities exceeds the relative number of such pairs among the double stars whose orbits have been determined.

To make things clear from a theoretical point of view, we consider the distribution of states of a satellite in the phase space.

For coordinates in the phase space we take three components of a satellite's position and three components of its impulse with respect to the main star. Under statistical equilibrium, the number of satellites dN in the volume element $dx\,dy\,dz\,dp_x\,dp_y\,dp_z$ of the phase space is:

$$dN = C\exp\left(-\frac{E(x,y,z,p_x,p_y,p_z)}{\theta}\right) dx\,dy\,dz\,dp_x\,dp_y\,dp_z, \tag{1}$$

where

$$E = \frac{1}{2m}(p_x^2 + p_y^2 + p_z^2) - \frac{\gamma M m}{\sqrt{x^2+y^2+z^2}} \tag{2}$$

is the energy of the satellite, M and m are masses of the main star and the satellite, γ is the gravitational constant and θ is the Boltzman distribution module.

We will consider a broader class of distributions. Suppose that in the phase space the density is an arbitrary function $f(E)$ of the energy E, rather than having a special form $C\exp\left(-\frac{E}{\theta}\right)$.

Then

$$dN = f(E)\,dx\,dy\,dz\,dp_x\,dp_y\,dp_z.$$

Now let us make canonical transformations in the phase space from the variables x, y, z, p_x, p_y, p_z to the variables L, G, H, l, g, and h of the Delonais lunar theory [4]. The first three of these quantities are expressed in the following way in terms of the usual elements of elliptical motion, which are large semiaxis a, inclination i and eccentricity ε:

$$L = m\sqrt{\gamma M}\,a^{1/2},$$
$$G = m\sqrt{\gamma M}\,a^{1/2}(1-\varepsilon^2)^{1/2},$$
$$H = \sqrt{\gamma M}\,a^{1/2}(1-\varepsilon^2)^{1/2}\cos i.$$

The angular coordinates l, g and h represent the average anomaly, the distance of periastron from the node and the longitude of ascending node, respectively.

It is known that under the canonical transformation of the phase space volumes remain intact (the Jacobian is equal to 1). In other words,

$$dx\,dy\,dz\,dp_x\,dp_y\,dp_z = dL\,dG\,dH\,dl\,dg\,dh\,.$$

On the other hand,

$$E = -\frac{\gamma^2 M^2 m^3}{2L^2} = -\frac{1}{2}\frac{\gamma M m}{a},$$

i.e., the energy depends solely on L. Hence, in our case the density in the phase space also depends on L only, and we can write:

$$dN = f(L)\,dL\,dG\,dH\,dl\,dg\,dh\,.$$

This implies that the number of pairs with L between L and $L + dL$, and $G > G_0$ is equal to

$$8\pi^3 f(L)\,dL \int_{G_0}^{L} dG \int_0^G dH = 4\pi^3 f(L)(L^2 - G_0^2)\,dL,$$

since l, g and h vary independently within $(0, 2\pi)$.

We have

$$L^2 - G_0^2 = m^2\gamma^2 M^2 a\varepsilon_0^2 = L^2\varepsilon_0^2,$$

where ε_0 is the eccentricity, corresponding to an orbit L, G_0.

Thus the number of stars with $\varepsilon < \varepsilon_0$ (i.e., $G > G_0$) and L between L and $L + dL$ is

$$4\pi^3 f(L) L^2 \varepsilon_0^2\,dL,$$

which implies that the number of orbits for which $\varepsilon < \varepsilon_0$ is

$$N(\varepsilon_0) = 4\pi^3\varepsilon_0^2 \int f(L) L^2\,dL. \tag{3}$$

We have the following theorem:

If the density in the phase space is a function of L, i.e., of total energy and solely of this quantity, then for arbitrary density function f(L) the number of stars with eccentricities less than ε_0 *is proportional to* ε_0^2.

It follows that even if we assume that the observed $N(\varepsilon_0)$ is proportional to ε_0^2 (we have our doubts about this because of the selectivity of the material) this does not imply that the phase density is proportional to $\exp(-E/\theta)$, or in other words, that equipartition of energy holds. Actually for any distribution of energy under the single condition that the phase density does not depend on other elements, we have $N(\varepsilon) \sim \varepsilon_0^2$.

Thus, even if we assume that in fact $N(\varepsilon) \sim \varepsilon_0^2$, we cannot conclude about equipartition of energy, let alone the lifetime of a star system.

However, we note the following circumstance. According to the above, in the case where phase density depends on L only (equivalently on E or on the large semiaxes), for each interval dL the number of orbits with eccentricities less than ε is also proportional to ε_0^2. In other words, the number of orbits with eccentricities between ε and $\varepsilon + d\varepsilon$ should be proportional to $\varepsilon\, d\varepsilon$ regardless of a. Therefore, the average eccentricity for each interval of values of large semiaxes equals

$$\bar{\varepsilon} = \frac{\int_0^1 \varepsilon^2\, d\varepsilon}{\int_0^1 \varepsilon\, d\varepsilon} = \frac{2}{3} \tag{4}$$

independent of a.

The observational material contradicts this, as demonstrated by the following table, obtained by Aitken [5].

Table 1.

$\overline{P}$	$\bar{\varepsilon}$	n
16.8 years	0.43	14
37.1	0.40	24
73.0	0.53	24
138.0	0.57	23
200.0	0.62	18

The table contains average eccentricities for stars grouped according to periods.

The first column gives the average period for each group while the last — the number of stars in a group. If we add to this the statistical result of Russel, which says that for stars with periods about 5,000 years the average eccentricity is 0.76, we should conclude that ε depends on P. It is known that $P \sim L^3$. Therefore, $\bar{\varepsilon}$ depends on L. It becomes evident that the main assumption made above must be false and the phase density does not depend on large semiaxes alone. This means that the phase density is by no means proportional to $\exp(-E/\theta)$. Even the assumption that it depends on energy alone is false. There are indications that the given dependence of $\bar{\varepsilon}$ on P is subject to strong observational selection [10, 11, 12].

Possibly for distant components ($P >$ 100 years) the variation of $\bar{\varepsilon}$ is small and phase density depends only on E. It would be interesting to study the dependence of phase density on E based upon observations and to measure its deviation from Boltzman dependence.

2. Derivation of the phase density from observational data

In this section we assume that the phase density depends solely on E. We will try to obtain the form of this dependence from the empirical material. We saw that at least for smaller values of L the phase density probably depends on other elements too. Hence our result should be treated rather as an average with respect to other elements. Even in this form our conclusion has some value; in particular, for distant components our assumption is probably valid. Let the phase density again be $f(L)$. This means that in the volume element $dx\,dy\,dz\,dp_x\,dp_y\,dp_z$ the number of stars is

$$f\left(\gamma M m \cdot \sqrt{\frac{m}{\frac{2Mm\gamma}{r} - \frac{p^2}{m}}}\right) dx\,dy\,dz\,dp_x\,dp_y\,dp_z,$$

where

$$r = \sqrt{x^2 + y^2 + z^2}, \quad p = \sqrt{p_x^2 + p_y^2 + p_z^2}.$$

Therefore, the distribution density in the space is

$$\rho = \iiint f\left(\gamma Mm \cdot \sqrt{\frac{m}{\frac{2Mm\gamma}{r} - \frac{p^2}{m}}}\right) dp_x\, dp_y\, dp_z =$$

$$= 4\pi \int_0^{m\sqrt{\frac{2M\gamma}{r}}} f\left(\gamma Mm \cdot \sqrt{\frac{m}{\frac{2Mm\gamma}{r} - \frac{p^2}{m}}}\right) p^2\, dp.$$

By the spatial distribution of satellites we mean the distribution obtained after bringing the main stars of the pairs to a single point by parallel shifts. The upper bound in the last integral is obtained from the condition that our satellites move along elliptical orbits, i.e., our system has negative total energy.

Now a change of variable

$$L = \gamma Mm \cdot \sqrt{\frac{m}{\frac{2Mm\gamma}{r} - \frac{p^2}{m}}}$$

in the last integral yields

$$\rho(r) = \int_{m\sqrt{\frac{\gamma M}{2}}\, r^{1/2}}^{\infty} f(L)\sqrt{\frac{2Mm^2\gamma}{r} - \frac{\gamma^2 M^2 m^4}{L^2}} \cdot \frac{4\pi^2\gamma^2 M^2 m^4\, dL}{L^3}.$$

Denote

$$K = m\sqrt{\frac{\gamma M}{2}}\, r^{1/2},$$

then

$$\rho(r) = 4\pi^2\gamma^3 M^3 m^6 \int_K^{\infty} f(L)\sqrt{\frac{1}{K^2} - \frac{1}{L^2}}\, \frac{dL}{L^3},$$

or

$$\rho(K) = C\int_K^{\infty} f(L)\sqrt{L^2 - K^2}\, \frac{dL}{KL^4}. \qquad (5)$$

Using this integral equation we can find the phase density $f(L)$ in terms of a given ρ. Epic in [6] has shown that the observational material at hand, after correction for observational selectivity, gives

$$\rho \sim \frac{1}{r^3} \tag{6}$$

or

$$\rho \sim \frac{1}{K^6}.$$

It is evident that if ρ has this special form, then the function

$$f(L) \sim \frac{1}{L^3} \tag{7}$$

satisfies equation (5).

Let us compare this "observed" density in the phase space with that under statistical equilibrium where we have

$$f(L) = C \exp\left(-\frac{E}{\theta}\right) = C \exp\left(\frac{\gamma^2 M^2 m^3}{rL^2\theta}\right). \tag{8}$$

It remains to determine the value of θ. If the double stars ensemble reached statistical equilibrium as a result of interaction with other stars, then θ by order should equal two-thirds of the average kinetic energy of translational movement of surrounding stars. Suppose that the average speed of translational movement of stars is 25 km/sec by order; then already for $a > 20$ AU the exponent on the right-hand side of (8) becomes much less then 1. Therefore, for large values of L (as well as of a)

$$f(L) = \text{const} \tag{9}$$

is a satisfactory approximation.

Meanwhile the result obtained by Epic refers only to distant components. Hence (7) refers to larger values of L.

We see that the "observed" phase density (7) is quite different from that of statistical equilibrium (9).

One can show that "the distribution in space" in the case of statistical equilibrium also differs strongly from the one observed. Indeed from (9) and (5) it follows that under statistical equilibrium

$$\rho \sim \frac{1}{r^{3/2}} \tag{10}$$

in contradiction of the observed distribution (6). The difference between (10) and (6) is so great that it leaves no doubt that in fact (10) is unsatisfactory even as a rough approximation. The formula (3) was established by Epic for distant components, up to 10,000 AU. We conclude that even for such distant components the encounters have not yet led to statistical equilibrium (that is to the most probable distribution) in the sizes of great semiaxes, i.e., the energies. We will see that this strongly reduces the upper bound for the life span of a star system.

3. Testing Epic's law of inverse cubes by new observational material

In this section we consider a very simple testing method for law (6), obtained by Epic for the distribution of components in space. We will see that this new method of analysis confirms the approximate correctness of formula (6).

As a matter of fact, if components are distributed around central stars by the law $1/r^n$, where n is arbitrary, then the projection of the distribution density onto the celestial sphere will be $1/r^{n-1}$.

Assume we have double stars aggregate governed by this distribution within a volume. For any shifts in the volume, the distribution of apparent distances obviously will continue to satisfy the law $1/r^{n-1}$. A summation of this distribution for volume elements both in one direction and in various directions also leads to $1/r^{n-1}$. Therefore, for an arbitrary large part of the sky, for apparent (projected) distances we obtain the same law.

In particular under alternative assumptions

$$\rho \sim \frac{1}{r^3} \quad \text{and} \quad \rho \sim \frac{1}{r^{3/2}}$$

for distribution densities in projections we obtain

$$\rho \sim \frac{1}{r^2} \quad \text{and} \quad \rho \sim \frac{1}{r^{1/2}},$$

which implies that the number of stars with apparent distances between r_2 and r_1 is proportional to

$$\ln \frac{r_2}{r_1} \quad \text{and} \quad r_2^{3/2} - r_1^{3/2}. \tag{11}$$

From Aitken's catalogue [7] we took all stars with apparent magnitude up to 9.0 lying in the Northern Hemisphere (4640 stars altogether). Within this limit Aitken's catalogue seems to be sufficiently homogeneous, since all stars with magnitudes up to 9.0 were tested by Aitken at Lick. The following table gives the number of pairs with distances ranging from 0.5″ to 8″. The second and third lines give numbers proportional to $\ln \frac{r_2}{r_1}$ and to $r_2^{3/2} - r_1^{3/2}$ respectively. The proportionality coefficient C was chosen to have the same total number in each line.

Interval	0.5 − 1″	1 − 2″	2 − 4″	4 − 8″	Total
The observed number of pairs	883	1160	1283	1314	4640
$C \ln \dfrac{r_2}{r_1}$	1160	1160	1160	1160	4640
$C\left(r_2^{3/2} - r_1^{3/2}\right)$	136	382	1080	3040	4638

It is seen from this comparison that $C \ln \dfrac{r_2}{r_1}$ indeed yields an approximation (with about 10% precision) to the observed numbers, while $C\left(r_2^{3/2} - r_1^{3/2}\right)$ cannot be justified.

The deviations from $C \ln \dfrac{r_2}{r_1}$ will definitely become smaller if optical pairs are excluded.

Thus Epic's law $\rho \sim \dfrac{1}{r^3}$ is confirmed in the first approximation. Once again this shows that the energies of stellar pairs are not distributed by the Boltzman law.

In his answer to the author's preliminary note, Jeans acknowledged [3] that equipartition in energies does not exist, but he added that "in certain respects there is a tolerably good approximation to equipartition." However, in view of the above corollaries from Epic's law, it is hardly possible to speak about any approximation to equipartition at all.

4. Relaxation time for an ensemble of double stars

Let us consider the time required for an ensemble of double stars in a star system to reach statistical equilibrium with surrounding stars. In statistical equilibrium we usually have two processes acting in opposite directions: on one hand, the destruction of physical pairs as a consequence of interaction with stars of the field and, on the other hand, formation of pairs when three initially independent stars come together. In the latter case the third body carries away the excess energy. Below we show that because of the absence of statistical equilibrium in our star system, complete mutual compensation of these processes does not occur: the number of pairs formed is negligible in comparison with the number of pairs destroyed.

Along with the destruction of pairs by approaching stars, small energy variations may accumulate to cause destruction. These processes lead to a statistical equilibrium in the sense of Boltzman distribution.

It will be evident that the average time of destruction of a star pair, which we calculate below, is quite enough to reach Boltzman distribution.

Boltzman distribution is reached by means of energy variations smaller than those needed for destruction. Therefore, the time required for this is not greater than the average lifetime of a pair. Thus the average time necessary for destruction of a pair gives the order of "relaxation time" of a double star system. Our computations will refer to "distant pairs" with a distance between components greater than 100 AU and thousands of AU on average, by order.

Passage of a third star near a stellar pair can be of two types:

1) the minimal distance of the passing body from the center of gravity of the pair is large, compared with the great semiaxis of the orbit;

2) the minimal distance of the third body from one of the components is smaller than the large semiaxis of the pair.

The corresponding types we call "distant" and "close" passages. Passages of intermediate type also may occur, but we will not dwell on them since their role is secondary.

Considering passages of Coulomb particles near an atom, Bohr has shown [8, 9] that the role of distant passages is negligibly small compared with the role of close passages. Therefore, we will consider only close passages. The contribution of distant passages somewhat shortens the relaxation time but leaves its order intact.

For pairs of the type we now consider, velocities of orbital motions around the gravity center have the order of one or at most 2 or 3 km/sec. Meanwhile, the relative velocities in the star system are about 30 km/sec. Therefore, in the coordinate system attached to the gravity center, the satellite can practically be considered motionless.

Due to the velocity ratio mentioned a closely passing star will exert most of its influence on the satellite during a small fraction of the rotation period of a pair.

The satellite will acquire extra kinetic energy while its potential energy will remain unchanged. As a result, we will have either growth of the greater semiaxis or complete destruction of the pair. A contrary pattern of interaction requires lower kinetic energy of the passing star compared with that of the satellite. However, the probability of such an event is too low.

The above remarks imply that to influence the satellite, the central star needs a time period much greater than the duration of an encounter.

In this way we come to the problem of evaluation of the change in kinetic energy of a satellite under the influence of a passing star, in the coordinate system attached to the pair's center of masses.

A simple calculation gives the energy increase due to a distant passage

$$\Delta E = \frac{mv^2}{2} \frac{1}{1 + \frac{p^2 v^4}{4m^2\gamma^2}}, \tag{12}$$

assuming that the masses of the passing star and the satellite are equal. Here p is the encounter parameter, i.e., the distance of the satellite from the line along which the passing star moved before the encounter. The number of encounters for which p lies between p and $p + dp$, and the velocity of the passing star between v and $v + dv$, and that occur within time interval dt equals

$$2\pi p\, dp\, v\, dt\, dn,$$

where dn is the number of stars within a unit volume possessing velocities between v and $v + dv$. Therefore, the energy increase during time t will be

$$\pi t \int mv^3\, dn \int \frac{p\, dp}{1 + \dfrac{p^2 v^4}{4m^2\gamma^2}}.$$

Integration in p is over values corresponding to a close encounter, i.e., over $p < a$, a is the great semiaxis.

Therefore,

$$\Delta E = 2\pi t m^3 \gamma^2 \int \frac{\log\left(1 + \frac{a^2 v^4}{4m^2\gamma^2}\right)}{v}\, dn,$$

or

$$\Delta E = 2\pi t m^3 \gamma^2 \frac{n}{\overline{v}} \log\left(1 + \frac{a^2 \overline{v}^4}{4m^2\gamma^2}\right), \tag{13}$$

where n is the total number of stars in the unit volume (the stellar density), $\overline{v}$ is the mean velocity. The time required for ΔE to reach the total energy of the system, which is $-\frac{\gamma m^2}{2a}$, is

$$t = \frac{\overline{v}}{4\pi m \gamma a n \log\left(1 + \frac{a^2 \overline{v}^4}{4m^2\gamma^2}\right)}. \tag{14}$$

This we can (and do) consider to be the relaxation time. Here a is some mean value over this period which is close to the initial value a_0, since for most of the interaction time the values of a remain less than a_0.

Let us substitute in (14) $\overline{v} = 3 \cdot 10^6$ cm/sec and $m =$ the mass of the Sun. The observed values of A can reach $\frac{1}{20}$ parsec, while $n = 0.1(\text{parsec})^{-3}$. We get the value $t = 5 \cdot 10^9$ years. For smaller values of a we get values of the order 10^{10} and 10^{11} years.

Thus, for double stars with distances between components less than 10,000 AU, the Boltzman distribution is reached only after a period of 10^{10} years. The Epic distribution $\left(\rho \sim \frac{1}{r^3}\right)$ was derived for pairs from just this class. Hence for them the Boltzman distribution is not valid. We conclude that the age of these pairs cannot exceed 10^{10} years. In other words, the distribution of great axes of double star orbits favors rather definitely the "shorter time scale."

Our previous note mentioned this circumstance, albeit the above calculation of the relaxation time was not presented there. This gave Jeans a chance to write: "I cannot see that Prof. Ambartsumian's remarks in any way challenge this position, so that it seems to me the observational data he mentions are not opposed to the long time scale of 10^{13} years, but only to an infinitely long time scale."

Meanwhile we have seen that simple calculations indicate that the observational data in question are in contradiction not only to the time scale of 10^{13} years, but even to the time scale of 10^{10} years, i.e., they completely favor the shorter time scale.

5. Dissociative equilibrium for double stars

Additional important data confirming that the encounters have not by now created a statistical equilibrium for pairs with distances about 10^4 AU by order are deviations in the number of such pairs from what is expected under dissociative equilibrium.

We denote by δn_D the number of pairs whose satellites lie within element $\delta\Gamma$ of the phase space which we considered above. According to standard rules of the kinetic theory of gases, by a dissociative equilibrium we have

$$\frac{\delta n_D}{n^2} = \frac{\delta\Gamma}{(\pi m\theta)^{3/2}} \exp\left(-\frac{E}{\theta}\right). \tag{15}$$

Here E is again the internal energy of a pair with a satellite in $\delta\Gamma$, θ is the module of the Boltzman distribution for motion of the stars, n is the number of single stars in a unit volume. If we choose $\delta\Gamma$ in that part of the phase space where $a > 100$ AU, then the factor $\exp(-E/\theta)$ can be replaced by 1. We will have

$$\frac{\delta n_D}{n^2} = \frac{\delta\Gamma}{(\pi m\theta)^{3/2}}. \tag{16}$$

By summation this formula extends to volumes $\delta\Gamma$ which are no longer infinitesimal. The only restriction is that $\delta\Gamma$ should not include parts where E/θ is no longer small as compared with 1.

We can (and do) take that part of the phase space where $a_1 < a < a_2$ for some bounds a_1, a_2. The corresponding bounds for L will be

$$L_1 = m\sqrt{\gamma M}a_1^{1/2} \quad \text{and} \quad L_2 = m\sqrt{\gamma M}a_2^{1/2}.$$

The corresponding phase volume is found to be

$$\delta\Gamma = 8\pi^3 \int_{L_1}^{L_2} M \int_0^L dG \int_0^G dH = \frac{4\pi^3}{3}\left(L_2^3 - L_1^3\right) = \\ = \frac{4\pi^3}{3} m^3 (\gamma M)^{3/2} \left(a_2^{3/2} - a_1^{3/2}\right). \tag{17}$$

Substituting (17) into (16) and putting $M = m$ there we obtain

$$\frac{\delta n_D}{n^2} = \frac{4}{3} m^3 \left(\frac{\pi\gamma}{\theta}\right)^{3/2} \left(a_2^{3/2} - a_1^{3/2}\right). \tag{18}$$

For $a_1 = 10^2$ AU, $a_2 = 10^4$ AU and for the same numerical values of constants used above, we get

$$\frac{\delta n_D}{n} = 10^{-8}.$$

This is the fraction of double stars which under dissociative equilibrium will have a satellite with a between 100 and 10^4 AU. In reality, at least one of every several dozen has this property, i.e., the event in question occurs millions of times more often than it would under dissociative equilibrium.

This perhaps is the most striking evidence indicating that our Galaxy is very far from a statistical equilibrium state and, in conjunction with the results of the previous section, speaks for the validity of the shorter time scale of 10^{10} years.

Because at present we have in our Galaxy an excess number of double stars as compared with the equilibrium state, dissolutions occur considerably (perhaps a million times) more often than the creation of pairs.

The result of this section we roughly formulate as follows.

The existence of such pairs as α and Proxima Centauri or Washington 5583–5584 proves the validity of the shorter time scale. In fact, satellite-stars having a about 10^4 AU are so numerous that even the star closest to us possesses such a satellite.

Conclusion

Until now the opinion was widespread, mainly due to Jeans, that the statistics of double stars speaks in favor of a longer time scale. While new facts from other domains of astronomy kept confirming the shorter scale, double stars remained the main argument for a longer evolutionary scale. In the present paper the latter argument is shown to be an illusion. A proper treatment of statistical data very definitely points to the shorter time scale.

REFERENCES

1. *Nature*, 137, 537, 1936.
2. *Nature*, 136, 432, 1935.
3. *Nature*, 137, 537, 1936.
4. Frank und Mises, "Die Differential und Integralgleichungen der Mechanik und Physik," Zweiter Teil, SS, 158–162, Braunschweig, 1927.
5. Aitken, *The Binary Stars*, 1935.
6. *Tartu Observatory Publications*, 25, 1924.
7. Aitken, *The New General Catalogue of Double Stars*, Washington, 1933.
8. Bohr, *Phil. Mag.*, 25, 10, 1913.
9. Bohr, *Phil. Mag.*, 30, 581, 1915.
10. Seyfert, *H. B.*, 896, 1934.
11. Barbier, *C. R.*, 199, 930, 1934.
12. Finsen, *M. N.*, 96, 862, 1936.

4 December 1936

Astronomical Observatory,
Leningrad University

ON THE DYNAMICS OF OPEN CLUSTERS[1]

It has been pointed out in the literature that, due to several causes, open star clusters dissipate with time. For instance, Rosseland showed that when external stars move through a cluster, they cause perturbation of the motion of the stars in the cluster and could transfer enough momentum to individual stars to cause their escape from the cluster's gravitational field. In this way the cluster will lose stars gradually, i.e., it will dissipate. According to Rosseland the time needed for the star cluster to dissipate following the outlined mechanism is 10^{10} years. However, as pointed out by the present author in the supplement to the Russian edition of Rosseland's book, there is another factor that makes the life of an open cluster even shorter: the stars in the cluster may have close encounters with each other. As a result, they exchange kinetic energy and gradually tend toward the most probable distribution, i.e., a Maxwell-Boltzmann distribution. And this, as we shall see shortly, also causes the dissipation of the cluster.

The relaxation time, i.e., the time in which the encounters of the stars in a cluster will lead to statistical equilibrium, is given approximately by the formula:

$$\tau = \frac{3\sqrt{2}}{32\pi n} \cdot \frac{v^3}{G^2 m^2 \log\left(\frac{\rho}{\rho_0}\right)}, \tag{1}$$

where n is the number of stars per unit volume, m the stellar mass, G the gravitational constant, v the average stellar velocity in the cluster, ρ the radius of the cluster, and ρ_0 the distance at which the potential energy of two stars is equal to the average kinetic energy of stars in the cluster, i.e.,

$$\rho_0 = \frac{2Gm}{v^2}. \tag{2}$$

[1]English translation originally published by Kluwer Academic Publishers in *Dynamics of Star Clusters*, Jeremy Goodman and Piet Hut (eds.), 1985, pp. 521-524, IAU © 1985. Used here with permission of the International Astronomical Union.

Formula (1) was derived for the case of stars with equal masses.

The average velocity v enters formula (1) both explicitly and through ρ_0. To find v we assume that for shorter time intervals the cluster is stationary. We can do this since the time necessary to change the distribution law for the stars in the cluster, considered as a system in phase space, is large compared with the time necessary for a star to cross the cluster. In the case of a stationary system consisting of particles interacting according to Newton's law, using the virial theorem we can write

$$U = 2T, \tag{3}$$

where U is the absolute value of the potential energy of the system, and T is its kinetic energy.

The exact formula for U is:

$$U = \frac{1}{2} \sum_{i \neq k} \frac{Gm^2}{r_{ik}},$$

where we assume again that all stellar masses are equal, and r_{ik} denotes the distance between the i-th and the k-th star. We replace all of the r_{ik}'s with their mean harmonic value which is apparently close to the radius of the cluster ρ. Then approximately

$$U = \frac{1}{2} \frac{GN(N-1)\,m^2}{\rho},$$

where N is the total number of stars in the cluster. For $N \gg 1$,

$$U = \frac{1}{2} \frac{GN^2\,m^2}{\rho}.$$

On the other hand

$$2T = Nmv^2.$$

Therefore, the virial theorem assumes the form:

$$v^2 = \frac{GNm}{2\rho}. \tag{4}$$

Comparing (4) with (2), we find that

$$\log\left(\frac{\rho}{\rho_0}\right) = \log\left(\frac{N}{4}\right). \tag{5}$$

Substituting (4) and (5) in (1) and taking into account that

$$n = \frac{N}{\frac{4}{3}\pi\rho^3},$$

we find

$$\tau = \frac{2}{16\log\left(\frac{N}{4}\right)} \cdot \sqrt{\frac{N\rho^3}{Gm}}. \tag{6}$$

Assuming that for a typical cluster $N = 400$, $\rho = 2$ parsecs, $m = 2 \times 10^{33}$ g, we find the relaxation time to be $\tau \approx 4 \times 10^7$ years.

A result of the evolution of the distribution towards a Maxwell-Boltzmann distribution is the presence of stars with kinetic energy larger than the escape energy for the cluster. Such stars tend to leave the cluster. The whole question is, what is the percentage of such stars in a cluster with the Boltzman distribution? If this percentage is small, then the dissipation of the cluster as a result of this process will be very slow. It is apparent that the ratio of the number of stars which can escape in a relaxation time τ to the total number of stars in the cluster is equal to

$$P = \frac{\int_{\varepsilon_0}^{\infty} \exp\left(-\frac{\varepsilon}{\theta}\right)\sqrt{\varepsilon}\,d\varepsilon}{\int_0^{\infty} \exp\left(-\frac{\varepsilon}{\theta}\right)\sqrt{\varepsilon}\,d\varepsilon}, \tag{7}$$

where ε_0 is the escape energy, and θ is equal to two thirds of the average kinetic energy, i.e.,

$$\theta = \frac{2}{3}\frac{T}{N} = \frac{1}{3}\frac{U}{N}. \tag{8}$$

On the other hand, the mean value of ε_0, i.e., the escape energy, is equal to

$$\varepsilon_0 = \overline{\sum_k \frac{Gm^2}{r_{ik}}} = \frac{2U}{N}, \tag{9}$$

where the line over the sum indicates an average over i. Comparing (8) with (9) we find:

$$\varepsilon_0 = 6\theta.$$

Substituting in (7), we obtain the approximation:

$$P \approx \frac{\exp\left(-\frac{\varepsilon_0}{\theta}\right)\theta\sqrt{\varepsilon_0}}{\frac{1}{2}\sqrt{\pi}\,\theta^{3/2}} = 2\,e^{-6}\sqrt{\frac{6}{\pi}},$$

i.e., one hundredth of the total number of stars should escape a cluster in a relaxation time. Therefore, the dissipation time of the cluster should be of the order of several billion years.

This result was obtained for a cluster consisting of stars with equal masses. Therefore, these numbers are applicable only to stars of the cluster with masses close to the average mass of the stars in the cluster. For stars with masses two to three times less than the average, the escape time will be of the order of a few hundred million years. Remarkably, the open clusters are known to possess rather few dwarfs.

Let us assume for a moment that the open clusters we observe are different stages in the evolution of one and the same cluster. Since the stars escaping from the cluster carry away positive kinetic energy, the total cluster energy

$$H = T - U \tag{10}$$

should decrease with the transition from richer to poorer clusters.

If we substitute (3) in (10), we find

$$H = \frac{1}{2}U. \tag{11}$$

Therefore, under the above assumption, U should increase. The data in the article by Orlova show that no increase of U with the decrease of N is observed.

Another possible hypothesis is that all clusters were formed approximately at the same epoch (perhaps even at the epoch of the formation of the galaxy itself). Then the evolution of rich clusters with large diameters should be slower. Among other things, these rich and large clusters should contain a higher percentage of dwarfs. It seems to the author that this conclusion is supported by observation. That is, the clusters h and χ Persei are both rich and contain a high percentage of dwarfs. On the other hand, a number of poor clusters have hardly any dwarfs. Hence, it becomes clear that to make further conclusions it is of great interest to determine not only the luminosity function for different clusters, but also the total energy H. According to (11), H can be determined from the absolute value of the potential energy.

Leningrad University

THE SCATTERING OF LIGHT BY PLANETARY ATMOSPHERES

The problem of scattering and absorption of light by planetary atmospheres has been the subject of many theoretical studies. However, because of mathematical difficulties, no satisfactory solution adequate to the real physical conditions has been found.

It is known that the problem of distribution of brightness on the planetary disk in some approximation is equivalent to the problem of diffuse reflection from a turbid plane-parallel layer. In turn, the problem of diffuse reflection from a turbid medium requires consideration of the multiple elementary processes of scattering.

In an earlier paper [1] the author has given an exact solution for the case where the medium stretches to infinity in two opposite directions and for any scattering indicatrix. That solution became a basis for treating the diffuse reflection problem for a layer which stretches to infinity only in one direction by means of successive approximations.

In the present paper a new and more exact solution is given to the problem by reducing it to an easily numerically solvable functional equation. One of the advantages of this method is that the unknown function can be found without intermediate calculation of different functions describing the radiation field in the inner layers of the medium.

The method is not restricted to the case of the spherical indicatrix of scattering. However, in order to simplify the presentation, we treat here the case of the spherical indicatrix only, postponing the treatment of the general case to another occasion. We show that at the same time, the method yields a solution to the problem of distribution of brightness over the solar disk.

1. Integral equation of the problem of diffuse reflection

Let us consider plane-parallel layers of matter which is able both to scatter and to absorb passing radiation. Let this matter fill a halfspace (one-sided infinity). Assume that the ratio of the absorption coefficient to the extinction coefficient (the latter is the sum of absorption and extinction coefficients) is a constant to be denoted by λ. We suppose that on the boundary plane the radiation falls under angle θ_0 to the normal. Let the density of radiation flow (the energy passing through the unit area perpendicular to the flow) be πS.

The integral equation of the theory of scattering in the case of a spherical indicatrix is known to have the form

$$B(\tau) = \frac{\lambda}{4} S \exp(-\tau \sec\theta_0) + \frac{\lambda}{2} \int_0^\infty \mathrm{Ei}\,|\tau - t|\, B(t)\, dt, \tag{1}$$

where $B(\tau) = \eta/\alpha$ is the ratio of coefficients of radiation and extinction and τ is the optical depth. If the solution of (1) is known, the intensity of light diffusely reflected in the direction making an angle θ with the normal can be found from the formula

$$I(\theta_1) = \int_0^\infty \exp(-\tau \sec\theta_0)\, B(\tau) \sec\theta_1\, d\tau. \tag{2}$$

The usual method consists of finding a solution of (1) and substituting $B(\tau)$ into (2). In this way one can find the intensity I as a function of the angles of incidence θ_0 and reflection θ_1. From the linearity of the problem it is clear that both $B(\tau)$ and I will be proportional to S. Let us denote

$$\frac{I}{S} = r(\theta_1, \theta_0).$$

This quantity will depend on θ_1 and θ_0 but not on S.

The problem of diffuse reflection requires us to find the function $r(\theta_1, \theta_0)$. Below we derive an equation directly for $r(\theta_1, \theta_0)$ and find its solution.

2. The functional equation determining the function $r(\theta_1, \theta_0)$

Let us set in (1) $\xi = \sec\theta_0$ and $\tau = \sigma + a$, $t = s + a$:

$$B(\sigma + a) = \frac{\lambda}{4} \exp(-\xi(\sigma + a)) + \frac{\lambda}{2} \int_{-a}^\infty \mathrm{Ei}\,|\sigma - s|\, B(s + a)\, ds, \tag{3}$$

and take for a moment $S = 1$. Differentiating this equation by a we have

$$B'(\sigma + a, \xi) - \frac{\lambda}{2} \int_{-a}^{\infty} \mathrm{Ei}\,|\sigma - s|\, B'(s + a, \xi)\, ds =$$
$$= -\frac{\lambda}{4}\xi \exp\left(-\xi(\sigma + a)\right) + \frac{\lambda}{2}\mathrm{Ei}(\sigma + a)\, B(0, \xi),$$

where we show explicitly the dependence of B on the parameter $\xi = \sec\theta_0$. Putting $a = 0$ we obtain

$$\begin{aligned} &B'(\sigma, \xi) - \frac{\lambda}{2} \int_{0}^{\infty} \mathrm{Ei}\,|\sigma - s|\, B'(s, \xi)\, ds = \\ &- \frac{\lambda}{4}\xi \exp(-\xi\sigma) + \frac{\lambda}{2}\mathrm{Ei}(\sigma)\, B(0, \xi). \end{aligned} \tag{4}$$

But

$$\mathrm{Ei}(\sigma) = \int_{1}^{\infty} \exp(-\sigma\xi)\, \frac{d\zeta}{\zeta},$$

and we conclude that the right-hand side of (4) is a superposition of terms of the type $\exp(-\sigma\xi)$. The same is true for the right-hand side of (1). Owing to the linearity, we can write the solution of (4) as a superposition of solutions of the equations of type (1):

$$B'(\sigma, \xi) = -\xi B(\sigma, \xi) + 2B(0, \xi) \int_{1}^{\infty} B(\sigma, \zeta)\, \frac{d\zeta}{\zeta}. \tag{5}$$

Multiplying (5) by $\exp(-\eta\sigma)$ and integrating over σ, we find

$$\begin{aligned} \int_{0}^{\infty} \exp(-\eta\sigma) B'(\sigma, \xi)\, d\sigma &= -\xi \int_{0}^{\infty} \exp(-\eta\sigma)\, B(\sigma, \xi)\, d\sigma + \\ &+ 2B(0, \xi) \int_{1}^{\infty} \frac{d\zeta}{\zeta} \int_{0}^{\infty} B(\sigma, \zeta) \exp(-\eta\sigma)\, d\sigma. \end{aligned} \tag{6}$$

Integrating in the left-hand side we can eliminate the derivative of B

$$\begin{aligned} \eta \int_{0}^{\infty} \exp(-\eta\sigma) B(\sigma, \xi)\, d\sigma - B(0, \xi) &= -\xi \int_{0}^{\infty} \exp(-\eta\sigma) B(\sigma, \xi)\, d\sigma + \\ &+ 2B(0, \xi) \int_{1}^{\infty} \frac{d\zeta}{\zeta} \int_{0}^{\infty} B(\sigma, \zeta) \exp(-\eta\sigma)\, d\sigma. \end{aligned} \tag{7}$$

However,

$$\eta \int_{0}^{\infty} \exp(-\eta\sigma)\, B(\sigma, \xi)\, d\sigma = I(\eta) = r(\eta, \xi),$$

since we have adopted $s = 1$.

Therefore, equation (7) can be rewritten in the form

$$\frac{\xi+\eta}{\eta} r(\eta,\xi) = B(0,\xi)\left[1+\frac{2}{\eta}\int_1^\infty r(\eta,\zeta)\,\frac{d\zeta}{\zeta}\right].$$

For the function

$$R(\eta,\xi) = \frac{r(\eta,\xi)}{\eta}$$

we now find

$$(\xi+\eta)R(\eta,\xi) = B(0,\xi)\left[1+2\int_1^\infty R(\eta,\xi)\,\frac{d\zeta}{\zeta}\right]. \tag{8}$$

On the other hand, from (1) we have, for $\tau = 0$

$$B(0,\xi) = \frac{\lambda}{4}\left[1+2\int_0^\infty \mathrm{Ei}(t)\,B(t,\xi)\,dt\right],$$

or using the integral expression of $\mathrm{Ei}(t)$

$$B(0,\xi) = \frac{\lambda}{4}\left[1+\frac{2}{\xi}\int_1^\infty r(\xi,\zeta)\,\frac{d\zeta}{\zeta}\right]. \tag{9}$$

Taking into account that $r(\xi,\zeta) = \xi R(\xi,\zeta)$, we obtain

$$B(0,\xi) = \frac{\lambda}{4}\left[1+2\int_1^\infty R(\xi,\zeta)\,\frac{d\zeta}{\zeta}\right]. \tag{9a}$$

Substituting (9a) in (8) we find

$$(\xi+\eta)R(\eta,\xi) = \frac{4}{\lambda}B(0,\xi)\;B(0,\eta),$$

$$r(\eta,\xi) = \frac{4}{\lambda}\frac{\eta}{\eta+\xi}B(0,\xi)\;B(0,\eta). \tag{10}$$

Thus we conclude that in the case of a spherical indicatrix of scattering the function $r(\eta,\xi)$ of diffuse reflection is represented by the product of two identical functions each of which depends on one variable, multiplied by the ratio $\dfrac{\eta}{\eta+\xi}$. We recall in this connection that Minnaert [2] has recently noted that the ratio $\frac{r(\eta,\xi)}{\eta}$ is always a symmetrical function of η and ξ independent of the form of the scattering indicatrix.

In a paper of V. A. Fock (now at press [3]) in which the exact solution of (1) was obtained, it was shown that the ratio $\frac{\eta+\xi}{\eta}\, r(\eta,\xi)$ is a product of some function of η and of the same function of ξ, and this function was expressed as an integral depending on a parameter. Our aim now is to find a functional equation for $B(0,\xi)$.

Substituting (10) into the right-hand side of (9a) we find

$$B(0,\xi) = \frac{\lambda}{4}\left[1 + \frac{8}{\lambda} B(0,\xi) \int_1^\infty \frac{B(0,\zeta)}{\xi+\zeta}\,\frac{d\zeta}{\zeta}\right].$$

Instead of $\xi = \sec\theta_0$ we consider $x = 1/\xi = \cos\theta_0$ to be an argument and denote

$$\frac{2}{\sqrt{\lambda}} B(0,\xi) = \frac{2}{\sqrt{\lambda}} B\left(0,\frac{1}{x}\right) = \varphi(x). \tag{11}$$

Then we find a functional equation for $\varphi(x)$

$$\varphi(x) = \frac{\sqrt{\lambda}}{2}\left[1 + 2\varphi(x)\, x \int_0^1 \frac{\varphi(z)}{x+z}\, dz\right]. \tag{12}$$

Now we can represent $r(\eta,\xi)$ as a function of y and x

$$r(y,x) = \frac{x}{x+y}\,\varphi(x)\,\varphi(y). \tag{13}$$

Thus the solution of the functional equation (12) will give us immediately the function $r(y,x)$ of diffuse reflection. In the next paragraph we present this solution for different values of λ.

The advantage of this approach is that in this way we avoid consideration of the functions which describe the radiation field inside the medium.

Of course the proposed method of reduction of the integral equation (1) to a functional equation by means of Laplace transformation can be extended to other integral equations which have kernels depending only on the difference $\tau - t$.

3. Solution of the functional equation for $\varphi(x)$

Instead of $\varphi(x)$ we consider a function

$$\psi(x) = \frac{2}{\sqrt{\lambda}}\varphi(x) \tag{14}$$

which obviously satisfies the equation

$$\psi(x) = 1 + \frac{\lambda}{2}\, x\psi(x) \int_0^1 \frac{\psi(z)}{x+z}\, dz. \tag{15}$$

For $\lambda < 1$ the numerical solution of this equation can be obtained by successive approximation. We begin by taking in the right side of (15) the approximation of the zero order $\psi_0(x) = 1$. As the first approximation we obtain

$$\psi_1(x) = 1 + \frac{\lambda}{2}\, x \ln \frac{1+x}{x},$$

and so on.

Table 1. Values of $\varphi(x) = \frac{\sqrt{\lambda}}{2}\psi(x)$ for different λ.

$\lambda \backslash x$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.519	0.360	0.387	0.403	0.415	0.424	0.431	0.438	0.443	0.448	0.452	0.455
0.612	0.391	0.428	0.449	0.466	0.479	0.490	0.500	0.508	0.516	0.522	0.528
0.728	0.426	0.477	0.508	0.532	0.552	0.570	0.585	0.598	0.609	0.620	0.629
0.806	0.449	0.512	0.553	0.586	0.613	0.637	0.658	0.677	0.694	0.708	0.722
0.865	0.465	0.540	0.590	0.631	0.666	0.697	0.724	0.749	0.771	0.790	0.809
0.910	0.477	0.563	0.621	0.672	0.715	0.754	0.789	0.821	0.851	0.878	0.903
0.944	0.486	0.580	0.647	0.706	0.759	0.805	0.850	0.889	0.926	0.962	0.993
0.969	0.492	0.594	0.669	0.736	0.798	0.852	0.906	0.955	1.001	1.045	1.086
0.986	0.497	0.607	0.691	0.767	0.837	0.903	0.967	1.026	1.081	1.137	1.188
0.997	0.499	0.617	0.710	0.797	0.878	0.958	1.083	1.108	1.179	1.250	1.319
1.000	0.500	0.624	0.725	0.821	0.915	1.007	1.097	1.188	1.276	1.365	1.455

All our numerical integrations have been carried out using the Simpson formula.

Actually the process of successive approximation will converge much more rapidly if we begin not with $\psi_0(x) = 1$ but rather with a function which is more or less near the exact solution.

For example, from (15) it is clear that $\psi(0) = 1$. This suggests taking for $\psi_0(x)$ a linear function

$$\psi_0(x) = 1 + ax,$$

where the constant a can be determined from the condition that the integral of $\psi_0(x)$ over the whole interval is equal to the same integral of the exact solution $\psi(x)$, i.e.,

$$1 + \frac{a}{2} = \int_0^1 \psi(x)\,dx. \tag{16}$$

But we are able to find the exact value of the integral on the right-hand side of (16) in the following way. Let us integrate both sides of (15)

$$\int_0^1 \psi(x)\,dx = 1 + \frac{\lambda}{2}\int_0^1\int_0^1 \frac{\psi(x)\psi(z)\,z\,dz\,dx}{x+z}.$$

Taking into account that the integral on the right-hand side is equal to

$$\frac{1}{2}\int_0^1\int_0^1 \psi(x)\psi(z)\,dz\,dx = \frac{1}{2}\left[\int_0^1 \psi(x)\,dx\right]^2,$$

we obtain a quadratic equation which yields

$$\int_0^1 \psi(x)\,dx = \frac{2}{\lambda}\left(1 - \sqrt{1-\lambda}\right). \tag{17}$$

With the same aim of improving the initial approximation we can use the processes of interpolation and extrapolation as soon as approximate solutions for two different values of λ are calculated. With moderate effort it is possible to find such a $\psi_0(z)$ that ψ_1 differs from ψ_0 by not more than two or three units of the third decimal place.

In the case $\lambda > 0.95$ it is better to modify the process in the following way. Under the integral sign in (15) let us substitute

$$\frac{x}{x+z} = 1 - \frac{z}{x+z}.$$

Then

$$\psi(x) = 1 + \frac{\lambda}{2}\,\psi(x) \int_0^1 \psi(z)\,dz - \frac{\lambda}{2}\,\psi(x) \int_0^1 \frac{\psi(z)\,z\,dz}{x+z},$$

or using (17)

$$\sqrt{1-\lambda}\,\psi(x) = 1 - \frac{\lambda}{2}\psi(x) \int_0^1 \frac{\psi(z)\,z\,dz}{x+z},$$

from which

$$\psi(x) = \left[\sqrt{1-\lambda} + \frac{\lambda}{2} \int_0^1 \frac{\psi(z)\,z\,dz}{x+z}\right]^{-1}. \tag{18}$$

Now putting in the right-hand side $\psi(z) = \psi_0(z)$ we calculate $\psi_1(z)$. Then in the right-hand side of (18) we put $\psi(z) = \frac{1}{2}\left[\psi_0(z) + \psi_1(z)\right]$ and obtain some function $\psi_2(z)$. Then we again form the mean, etc. When $\lambda > 0.95$ the process is rapidly converging.

Let us note that at $\lambda = 1$ equation (18) becomes

$$\psi(x) \int_0^1 \frac{\psi(z)\,z\,dz}{x+z} = 2,$$

or

$$\frac{2}{x} \int_0^1 r(z,x)\,z\,dz = 1.$$

Multiplying both sides by $\pi\,Sx$, we find that this equation represents the condition of equality of incident and reflected flows, i.e., that for $\lambda = 1$ the albedo for every angle of incidence is equal to 1.

4. The distribution of brightness over the planetary disk

The results obtained enable us to determine the distribution of brightness over the disk in different phases. The simplest result we have is for the case when the planet is in opposition. In such case $\theta_1 = \theta_0$ and $y = x$. Therefore, according to (13)

$$r(x,x) = \frac{1}{2}\left[\varphi(x)\right]^2, \tag{13$'$}$$

where x is the cosine of angular distance from the center of the disk. For the intensity we have

$$I = \frac{1}{2}\left[\varphi(x)\right]^2 S. \tag{19}$$

Consider an absolutely white surface which scatters light according to Lambert's law, situated at the distance of the planet from the Sun, perpendicular to solar radiation (we use the language of visual photometry). It is evident that $\frac{1}{2}[\varphi(x)]^2$ is the ratio of brightness at point x of the planetary disk to the brightness S of such white surface.

The maximal contrast, i.e., the maximal ratio of brightness at the center to that at the edge of the disk, occurs at $\lambda = 1$ (the case of pure scattering). When λ tends to zero the planetary disk is becoming homogeneously bright.

Comparing our results with observations it is important to keep in mind that they are applicable only to gaseous envelopes of great optical thickness such as the atmospheres of Jupiter, Saturn and Venus. One should also not forget that we have assumed that the scattering indicatrix is spherical. At the same time we assumed that in all regions of the atmosphere the optical properties are identical, i.e., we neglected the possible presence of local details.

It is quite possible and even probable that the scattering indicatrix is not spherical. The theoretical calculations for nonspherical indicatrices we will give in another paper. With all these reservations, we still have made a comparison with observations in the case of Jupiter, using the absolute measurements of brightness published by V. V. Sharonov [4].

Of greatest importance are comparisons with absolute values. It is possible to find some λ for which the theoretical ratio center/edge has the observed value if this ratio is between 1.0 and 8.0. Comparison with absolute measurements has the advantage that from the observed brightness at the center of the disk ($x = 1.0$) one can determine λ and then find the contrast. This is a more severe test of the theory. For this reason we have taken the observations of Sharonov. Since for the center of the disk the observations give $r = 0.590$, we concluded that for the atmosphere of Jupiter $\lambda = 0.969$. It was found that by $\lambda = 0.969$ the theoretical curve $\frac{1}{2}\varphi^2$ represents sufficiently well the distribution of brightness along the equatorial diameter. Only at the edge is the discrepancy greater than 11%. Since the precision of measurements decreases at the edges, we can consider the agreement sufficiently good.

Table 2. Theoretical distribution of brightness over the planetary disk in the case of opposition of the planet for different λ, according to (13′).

$\lambda\backslash x$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.519	0.065	0.075	0.081	0.086	0.090	0.093	0.096	0.098	0.100	0.102	0.104
0.612	0.076	0.092	0.101	0.108	0.114	0.120	0.125	0.129	0.133	0.137	0.140
0.728	0.091	0.112	0.129	0.141	0.152	0.162	0.171	0.179	0.186	0.192	0.198
0.806	0.101	0.131	0.153	0.172	0.188	0.203	0.216	0.229	0.241	0.251	0.261
0.865	0.108	0.146	0.174	0.199	0.222	0.243	0.262	0.280	0.297	0.312	0.327
0.910	0.114	0.158	0.193	0.226	0.256	0.284	0.311	0.337	0.362	0.385	0.408
0.944	0.118	0.168	0.209	0.249	0.288	0.324	0.361	0.395	0.429	0.463	0.493
0.969	0.121	0.176	0.224	0.271	0.318	0.363	0.410	0.456	0.501	0.546	0.590
0.986	0.124	0.184	0.239	0.294	0.350	0.408	0.468	0.526	0.584	0.646	0.706
0.997	0.125	0.190	0.252	0.318	0.385	0.459	0.534	0.614	0.695	0.781	0.870
1.000	0.125	0.195	0.263	0.337	0.419	0.507	0.602	0.705	0.814	0.931	1.058

5. The theoretical albedo

The auxiliary functions we have introduced allow us to determine the theoretical value of the albedo, i.e., the ratio of the flux reflected by the atmosphere to the incident flux. Generally this theoretical albedo depends on the angle between the incident flux and the direction normal to the layer.

For the flux scattered from the unit surface of the scattering layer we have

$$H = \int I \cos\theta_1 \, d\omega_1 = S \int r(\theta_1, \theta_0) \cos\theta_1 \, d\omega_1,$$

where $d\omega_1$ is an element of the solid angle. For the flux of the incident radiation we have

$$F = \pi \, S \cos\theta_0.$$

For the albedo we find:

$$A = \frac{H}{F} = \frac{2}{\cos\theta_0} \int r(\theta_1, \theta_0) \cos\theta_1 \sin\theta_1 \, d\theta_1,$$

and since

$$r(\theta_1, \theta_0) = R(\theta_1, \theta_0) \sec\theta_1,$$

we conclude that

$$A = \frac{2}{\cos\theta_0} \int R(\theta_1, \theta_0) \sin\theta_1 \, d\theta_1,$$

or

$$A = \frac{2}{y} \int_0^1 \frac{xy\varphi(x)\varphi(y)}{x+y} \, dx,$$

or

$$A = 2\varphi(y) \int_0^1 \varphi(x) \, dx - 2\varphi(y) \int_0^1 \frac{y\varphi(x) \, dx}{x+y}.$$

By means of (12), (14) and (17) we transform this equation to

$$A = 1 - 2\sqrt{\frac{1}{\lambda} - 1}\, \varphi(y). \tag{20}$$

This is the final expression for albedo. Table 3 contains the values computed according to (20) for A dependent on y and λ. We see that for the skew rays the albedo is smaller than when the incident flow is normal. This difference is greater in the case of smaller λ, i.e., for the media with a smaller albedo.

6. The distribution of brightness over the solar disk and similar problems

Let us again consider a plane-parallel absorbing and scattering layer. Suppose that the layer has a finite optical thickness τ_0 and that the sources of radiation lie behind the layer. In this situation we have some flux of radiation which has penetrated through the layer after scatterings. Let us increase the optical thickness of the layer, maintaining everywhere its optical parameters, including the constant ratio of the scattering coefficient to the absorption coefficient. At the same time let us increase the radiation power of illuminating sources behind the layer in such proportion that the flux of penetrating radiation remains constant. In the limit we will have a half-space medium and sources situated infinitely deep. Posing in this way the problem of radiation field, we arrive at the famous integral equation

$$\overline{B}(\tau) = \frac{\lambda}{2} \int_0^\infty \mathrm{Ei}|\tau - t| \, \overline{B}(t) \, dt. \tag{21}$$

Table 3. Values of $A = 1 - 2\sqrt{\frac{1}{\lambda} - 1}\,\varphi(y)$ for different λ.

$\lambda \backslash x$	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.519	0.307	0.255	0.224	0.201	0.183	0.170	0.156	0.146	0.137	0.130	0.124
0.612	0.376	0.318	0.274	0.257	0.236	0.218	0.201	0.190	0.177	0.168	0.158
0.728	0.478	0.414	0.378	0.348	0.324	0.302	0.284	0.257	0.254	0.240	0.230
0.806	0.560	0.498	0.458	0.426	0.399	0.376	0.355	0.336	0.320	0.306	0.292
0.865	0.633	0.573	0.535	0.502	0.474	0.450	0.428	0.408	0.391	0.376	0.361
0.910	0.700	0.646	0.610	0.573	0.551	0.526	0.505	0.484	0.466	0.448	0.433
0.944	0.763	0.717	0.685	0.656	0.680	0.608	0.586	0.567	0.549	0.531	0.516
0.969	0.825	0.788	0.762	0.738	0.716	0.697	0.677	0.660	0.644	0.628	0.614
0.986	0.883	0.857	0.837	0.819	0.802	0.787	0.772	0.758	0.745	0.732	0.720
0.997	0.949	0.936	0.926	0.917	0.909	0.900	0.893	0.885	0.877	0.870	0.863
1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

In particular, for $\lambda = 1$ we obtain an infinite purely scattering layer and the mathematical problem of E. Milne. Thus our problem for $\lambda = 1$ is equivalent to the problem of radiative equilibrium of the photosphere, although the physical picture can be quite different. Therefore, the angular dependence obtained from (21) yields simultaneously the distribution of brightness over the solar disk.

The intensity of the photospheric radiation in a given direction is determined by the integral

$$r(\eta) = \int_0^\infty \exp(-\eta t)\, \eta \overline{B}(t)\, dt, \tag{22}$$

where η is the secants of the angle between the direction of radiation and the normal to the layers. Now $r(\eta)$ may be expressed through the function φ, which was introduced above.

For the derivative $\overline{B}'(\tau)$ we obtain from (21):

$$\overline{B}'(\tau) = \frac{\lambda}{2} \int_0^\infty \mathrm{Ei}|\tau - t|\, \overline{B}'(t)\, dt = \frac{\lambda}{2} \mathrm{Ei}(\tau) \overline{B}(0). \tag{23}$$

Since

$$\mathrm{Ei}(\tau) = \int_1^\infty \exp(-\tau\zeta)\, \frac{d\zeta}{\zeta}, \tag{24}$$

we can obtain the solution of (23) as a superposition of solutions of the equations of type (23). This leads to

$$\overline{B}'(\tau) = 2\overline{B}(0) \int_1^\infty B(\tau, \zeta)\, \frac{d\zeta}{\zeta} + \mu \overline{B}(\tau), \tag{25}$$

where μ is a constant, which must be chosen in a way to reduce the right-hand side of (25) to $\overline{B}(\tau)$. Multiplying both sides of (25) by $\exp(-\tau\eta)$ and integrating we find

$$\int_0^\infty \exp(-\eta\tau)\, B'(\tau)\, d\tau = 2\frac{B(0)}{\eta} \int_1^\infty r(\eta, \zeta)\, \frac{d\zeta}{\zeta} + \frac{\mu}{\eta}\, r(\eta),$$

where $r(\eta, \zeta)$ is the function introduced in earlier paragraphs. Integrating by parts in the left-hand side of this equation, we find

$$\eta \int_0^\infty \exp(-\eta\tau)\, B(\tau)\, d\tau = \overline{B}(0) \left[1 + \frac{2}{\eta} \int_1^\infty r(\eta, \zeta)\, \frac{d\zeta}{\zeta}\right] + \frac{\mu}{\eta}\, r(\eta),$$

or

$$r(\eta)\left(1 - \frac{\mu}{\eta}\right) = \overline{B}(0) \left[1 + 2\int_1^\infty R(\eta, \zeta)\, \frac{d\zeta}{\zeta}\right]. \tag{26}$$

Substituting here instead of $R(\eta, \zeta)$ its expression, and writing y instead of η^{-1} and z instead of ζ^{-1}, we obtain

$$r(y) = \frac{\overline{B}(0)}{1 - \mu y} \left[1 + 2y\varphi(y) \int_0^1 \frac{\varphi(z)\, dz}{z + y}\right], \tag{27}$$

with $\varphi(z)$ as defined in section 2 (see also Table 2). Taking into account (12) we can rewrite (27) in the form

$$r(y) = \frac{2}{\sqrt{\lambda}}\, \frac{\overline{B}(0)\, \varphi(y)}{1 - \mu y}. \tag{28}$$

We see that the intensity of radiation which leaves the medium under the angle $\arccos \varphi(y)$ is proportional to

$$\frac{\varphi(y)}{1 - \mu y}.$$

Let us now determine the parameter μ. From equation (21) we have

$$\overline{B}(0) = \frac{\lambda}{2} \int_0^\infty \mathrm{Ei}\, t\, B(t)\, dt,$$

or

$$\overline{B}(0) = \frac{\lambda}{2}\int_0^1 \frac{d\zeta}{\zeta}\int_0^\infty \exp\,(-t\zeta)\overline{B}(t)\,dt = \frac{\lambda}{2}\int_1^\infty \frac{r(\zeta)}{\zeta^2}\,d\zeta = \frac{\lambda}{2}\int_0^1 r(y)\,dy.$$

Taking for $r(y)$ expression (28) we find an equation from which μ can be determined as a function of λ:

$$\sqrt{\lambda}\int_0^1 \frac{\varphi(y)\,dy}{1-\mu y} = 1. \tag{29}$$

Let us now show that μ is the solution of

$$\lambda = \frac{2\mu}{\ln\frac{1+\mu}{1-\mu}}. \tag{30}$$

To prove this we consider the bounded solution of the equation

$$\frac{\lambda}{2}\int_1^\infty \exp\,(-\tau\xi)\frac{d\xi}{\xi(\xi-\mu)} = C(\tau) - \frac{\lambda}{2}\int_0^\infty \mathrm{Ei}|\tau - t|C(t)\,dt, \tag{31}$$

where $\mu > 0$ satisfies (30). The bounded solution of (31) can be expressed as superposition of solutions of (3). Hence

$$C(\tau) = 2\int_1^\infty \frac{B(\tau,\xi)\,d\xi}{\xi(\xi-\mu)}. \tag{32}$$

Multiplying this equation by $\exp\,(-\eta\tau)$ and integrating, we find

$$\int_0^\infty C(\tau)\exp\,(-\eta\tau)\,d\tau = 2\int_1^\infty \frac{R(\eta,\xi)\,d\xi}{\xi(\xi-\mu)}. \tag{33}$$

On the other hand, we can write explicitly the only bounded solution of (31) which is

$$C(\tau) = \exp\,(-\mu\tau). \tag{34}$$

Therefore, we rewrite (33) in the form

$$\frac{1}{\mu+\eta} = 2\int_1^\infty \frac{\varphi(y)\,\varphi(\xi)\,d\xi}{\xi(\xi+\eta)(\xi-\mu)}.$$

Substituting again $\eta^{-1} = y$ and $\xi^{-1} = x$, we find

$$\frac{1}{1+\mu y} = 2\varphi(y)\int_0^1 \frac{x\,\varphi(x)\,dx}{(x+y)(1-\mu x)}, \tag{35}$$

or integrating both sides of (35)

$$\frac{1}{\mu}\ln(1+\mu) = 2\int_0^1 \frac{x\varphi(x)\,dx}{1-\mu x}\int_0^1 \frac{\varphi(y)\,dy}{x+y}.$$

By virtue of the functional equation (12)

$$\frac{1}{\mu}\ln(1+\mu) = \int_0^1 \frac{\frac{2}{\sqrt{\lambda}}\varphi(x)-1}{1-\mu x}\,dx.$$

This yields

$$\frac{1}{\mu}\ln\frac{1+\mu}{1-\mu} = \frac{2}{\sqrt{\lambda}}\int_0^1 \frac{\varphi(x)\,dx}{1-\mu x},$$

or on the basis of (30)

$$\sqrt{\lambda} = \int_0^1 \frac{\varphi(x)\,dx}{1-\mu x}.$$

This means that the root of equation (29) actually is the quantity μ determined by (25). In the special case of $\lambda = 1$, taking into account (28) we have

$$r(y) = A\,\varphi(y),$$

where A is a constant. Since the mathematical problem of distribution of brightness over the solar disk is equivalent to the problem considered in this paragraph, we can conclude that $\varphi(y)$ is the function representing the distribution of brightness over the solar disk for $\lambda = 1$.

7. Conclusions

Our method of reduction of an integral equation to a functional equation is applicable not only to the equation of scattering theory but also to more general equations with kernels depending on the difference of variables. The whole argumentation remains the same.

In another paper we intend to apply our method to the case of a nonspherical indicatrix of scattering.

REFERENCES

1. Ambartsumian, V. A., *Trudi AO Leningrad University*, 11, 64, 1941.

2. Minnaert, *Ap. J.*, 93, 403, 1941.
3. Fock, V. A., (editor), *Matem. Sbornik*, 14 (56), 1–2, 1944.
4. Sharonov, V. V., *Circ. Pulk. Obs.*, 48, no. 30, 1939.

4 December 1941–January 1942

Astronomical Observatory,
Leningrad University

THE PROBLEM OF DIFFUSE REFLECTION OF LIGHT BY A TURBID MEDIUM

(Presented by Academician S. I. Vavilov, January 6, 1943)

The problem of diffuse reflection of light by a turbid medium, i.e., by a medium which both scatters and absorbs light, has been the subject of much investigation. The usual approach to this physical problem is based on analysis of the radiation field within the medium and subsequent calculation using the data obtained on the intensities of diffusely reflected radiation. However, this approach has not as yet produced a complete solution to the problem.

In this note we propose another approach to the problem of diffuse reflection of light. It appears that this new method is much more effective than the old one, which led to an integral equation. The new method avoids any calculations connected with evaluation of quantities describing the radiation field inside the medium. Now we consider only what happens near its surface.

We take a medium, which consists of plane-parallel layers, each element of which has both scattering and absorbing ability. Let us suppose that the ratio $\frac{\lambda}{1-\lambda}$ of the scattering coefficient to the absorption coefficient is constant throughout the medium. Let us also assume that the scattering indicatrix has everywhere a spherical form. This means that the light scattered by every element of volume is distributed uniformly in all directions.

Finally, let us suppose that the medium is bounded on one side by a plane A and that on the other side it stretches to infinity. This means that the optical depth is infinite. At the end of this paper we will also consider the case of finite optical thickness.

Let us suppose that the plane A is illuminated by a bundle of parallel rays. By ξ we denote the cosine between the direction of the rays and the inner normal to A. Let the flux of this radiation through a unit of area perpendicular to the bundle be πS. Owing to scattering processes (multiple in the general case), certain intensities of diffuse light will flow from the medium in different directions. Of course, the intensity I of the outgoing radiation will depend on the angle between the direction of the outgoing radiation and the direction normal to A. Let the cosine of that angle be η. Thus I will depend on ξ and η, being at the same time proportional to S:

$$I(\eta, \xi) = r(\eta, \xi)\, S.$$

Our aim is to find the function $r(\eta, \xi)$.

Assume we attach to our infinite medium an additional layer of small optical thickness $\Delta\tau$ consisting of matter with the same optical properties. We think that the layer $\Delta\tau$ is bounded by two parallel planes, A and A'; thus A' forms the boundary of the new medium. The new composite medium will have the same capability of diffuse reflection as the initial infinite medium, characterized by the same function $r(\eta, \xi)$.

We will use this property of invariance with respect to addition of the layer $\Delta\tau$ to derive an equation for the function $r(\eta, \xi)$. Below we will neglect quantities of the order of $\Delta\tau^2$ and higher.

As a result of addition of the layer $\Delta\tau$, now the light first penetrates the new boundary A' of the medium. At the old boundary A, instead of the quantity πS, the quantity $\pi S\left(1 - \frac{\Delta\tau}{\xi}\right)$ of direct radiation now will fall. Therefore, A will reflect the intensity $r(\eta, \xi)\, S\left(1 - \frac{\Delta\tau}{\xi}\right)$. However, during passage of the layer $\Delta\tau$ it will diminish $\left(1 - \frac{\Delta\tau}{\eta}\right)$ times. The corresponding contribution to I will be

$$r(\eta, \xi)\left(1 - \frac{\Delta\tau}{\xi}\right)\left(1 - \frac{\Delta\tau}{\eta}\right) S.$$

However, the layer $\Delta\tau$ will scatter an additional intensity in the direction η. This intensity consists of four parts:

1) The layer $\Delta\tau$ scatters a part of the direct beam in the direction of η. This part is equal to

$$\frac{\lambda}{4}\frac{\Delta\tau}{\eta} S.$$

2) A part of the radiation is scattered by the layer $\Delta\tau$ towards A and is partly reflected from A. This gives an additional intensity

$$\frac{\lambda}{2}\Delta\tau\, S \int_0^1 r(\eta,\zeta)\,\frac{d\zeta}{\zeta}.$$

3) The layer $\Delta\tau$ scatters the radiation reflected from A. The corresponding intensity is

$$\frac{\lambda}{2}\frac{\Delta\tau}{\eta}\, S \int_0^1 r(\zeta,\xi)\, d\zeta.$$

4) A part of the radiation reflected from A is scattered back by the layer $\Delta\tau$ and again is partly reflected from A. As a result the intensity obtains an additional term

$$\lambda\Delta\tau S \int_0^1 r(\zeta,\xi)\, d\zeta \int_0^1 r(\eta,\zeta')\,\frac{d\zeta'}{\zeta'}.$$

In spite of all these increases and decreases, the intensity reflected from A' remains equal to $r(\eta,\xi)\, S$. Therefore,

$$r(\eta,\xi) = r(\eta,\xi)\left(1-\frac{\Delta\tau}{\xi}-\frac{\Delta\tau}{\eta}\right)+\frac{\lambda}{4}\frac{\Delta\tau}{\eta}+\frac{\lambda}{2}\frac{\Delta\tau}{\eta}\int_0^1 r(\zeta,\xi)\, d\zeta +$$
$$+\frac{\lambda}{2}\Delta\tau \int_0^1 r(\eta,\zeta)\,\frac{d\zeta}{\zeta}+\lambda\Delta\tau\int_0^1 r(\zeta,\xi)\, d\zeta \int_0^1 r(\eta,\zeta')\,\frac{d\zeta'}{\zeta'}$$

or

$$\left(\frac{1}{\eta}+\frac{1}{\xi}\right) r(\eta,\xi) = \frac{\lambda}{4}\left[\frac{1}{\eta}+2\int_0^1 r(\eta,\zeta)\,\frac{d\zeta}{\zeta}+\frac{2}{\eta}\int_0^1 r(\zeta,\xi)\, d\zeta + \right.$$
$$\left. +\, 4\int_0^1 r(\eta,\zeta')\,\frac{d\zeta'}{\zeta'}\int_0^1 r(\zeta,\xi)\, d\zeta\right].$$

We introduce the function $R(\eta,\xi)$, defined by

$$R(\eta,\xi) = \frac{4\eta}{\lambda}\, r(\eta,\xi). \tag{1}$$

Then we obtain for $R(\eta,\xi)$ a functional equation

$$\begin{aligned}\left(\frac{1}{\eta}+\frac{1}{\xi}\right) R(\eta,\xi) &= 1+\frac{\lambda}{2}\int_0^1 R(\eta,\zeta)\,\frac{d\zeta}{\zeta}+\\ &+\frac{\lambda}{2}\int_0^1 R(\zeta,\xi)\,\frac{d\zeta}{\zeta}+\frac{\lambda^2}{4}\int_0^1 R(\eta,\zeta')\,\frac{d\zeta'}{\zeta'}\int_0^1 R(\zeta,\xi)\,\frac{d\zeta}{\zeta}.\end{aligned} \tag{2}$$

Evidently if function $R(\eta, \xi)$ satisfies this equation so then does the function $R(\xi, \eta)$. But since our physical problem should have only one solution, it is natural to try to find a symmetric solution, i.e., to assume that

$$R(\eta, \xi) = R(\xi, \eta). \tag{3}$$

But in this case the right-hand side of (2) can be presented as a product

$$\left(\frac{1}{\eta} + \frac{1}{\xi}\right) R(\eta, \xi) = \left[1 + \frac{\lambda}{2} \int_0^1 R(\eta, \zeta) \frac{d\zeta}{\zeta}\right] \left[1 + \frac{\lambda}{2} \int_0^1 R(\xi, \zeta) \frac{d\zeta}{\zeta}\right]. \tag{4}$$

Let us denote

$$\varphi(\eta) = 1 + \frac{\lambda}{2} \int_0^1 R(\eta, \zeta) \frac{d\zeta}{\zeta}. \tag{5}$$

Then (4) implies the following structure of $R(\eta, \xi)$:

$$R(\eta, \xi) = \frac{\varphi(\eta)\, \varphi(\xi)}{\frac{1}{\eta} + \frac{1}{\xi}}. \tag{6}$$

Substituting (6) into (5), we obtain the equation for $\varphi(\eta)$

$$\varphi(\eta) = 1 + \frac{\lambda}{2} \eta\, \varphi(\eta) \int_0^1 \frac{\varphi(\xi)\, d\xi}{\eta + \xi}. \tag{7}$$

Thus the function $R(\eta, \xi)$ which determines the law of diffuse reflection is expressed in the form (6), where $\varphi(\eta)$ is the solution of the functional equation (7).

Since in our physical problem we always have $\lambda \leq 1$, equation (7) is easily solved numerically by means of successive approximation, beginning with zero approximation $\varphi_0(\eta) = 1$. In this way the problem of diffuse reflection from an infinitely deep medium can be completely solved.

One can also deduce the functional equations (2) and (7) not from the physical considerations as above but rather from the well-known integral equation of the scattering theory. This was shown in our earlier paper (*Astr. Journ.*, vol. 19, no. 5, p. 30, 1942).

The above method can be applied not only to media of infinite depth but also to layers of finite optical thickness τ, enclosed between two parallel planes A and B. In this case, however, we consider not only the function of diffuse reflection $r(\eta, \xi)$ but also a function $s(\eta, \xi)$ which describes the diffuse transparency, i.e., the part of light which enters via A and leaves the medium via B.

To use invariance, we add a thin layer $\Delta\tau$ on side A and subtract the same layer on side B.

Then it is possible to present the functions $r(\eta, \xi)$ and $s(\eta, \xi)$ by means of two auxiliary functions $\varphi(\eta)$ and $\psi(\eta)$, each of which depends on only one variable:

$$r(\eta, \xi) = \frac{\lambda}{4}\xi\,\frac{\varphi(\eta)\varphi(\xi) - \psi(\eta)\psi(\xi)}{\eta + \xi}$$

$$s(\eta, \xi) = \frac{\lambda}{4}\xi\,\frac{\psi(\eta)\varphi(\xi) - \varphi(\eta)\psi(\xi)}{\eta - \xi}.$$

The functions $\varphi(\eta)$ and $\psi(\eta)$ must then satisfy the equations

$$\varphi(\eta) = 1 + \frac{\lambda}{2}\,\eta\varphi(\eta)\int_0^1 \frac{\varphi(\xi)\,d\xi}{\eta + \xi} - \frac{\lambda}{2}\eta\psi(\eta)\int_0^1 \frac{\psi(\xi)\,d\xi}{\eta + \xi}$$

$$\psi(\eta) = \exp\left(-\tau/\eta\right) + \frac{\lambda}{2}\eta\int_0^1 \frac{\psi(\eta)\varphi(\xi) - \varphi(\eta)\psi(\xi)}{\eta - \xi}\,d\xi.$$

The numerical solution of this system of two equations can also be found by successive approximations.

9 December 1942

Leningrad University,
Elabuga Branch

ON THE PROBLEM OF THE DIFFUSE REFLECTION OF LIGHT[1]

The problem of the diffuse reflection of light by a scattering medium consisting of plane-parallel layers is considered for multiple scattering.

The problem of the diffuse reflection of light by a medium, every volume element of which both absorbs and reflects (turbid medium), has remained unsolved, even for the simplest case of a medium consisting of plane-parallel layers and a parallel beam of rays falling on the boundary of the medium. The aim of the present paper is to show that in this case the problem admits a solution by rather simple means.

Consider a medium consisting of plane-parallel layers bounded on one side by a plane A and extending on the other side to infinity. A beam of parallel rays falls on the surface A and penetrates into the depths of the medium, undergoing absorption and diffusion. Denote by θ_0 the angle formed by the direction of the rays with the internal normal. Let φ_0 be the azimuth of the incident rays, computed from some given direction on the surface A.

The usual method of studying the problem consisted in considering the equation of flux:

$$\cos\theta \frac{\partial I(\tau,\theta,\varphi)}{\partial\tau} = I(\tau,\theta,\varphi) - B(\tau,\theta,\varphi) \qquad (1)$$

and the generalized steady state equation for flux:

$$B(\tau,\theta,\varphi) = \\ = \frac{\lambda}{4\pi}\iint x(\cos\gamma)\, I(\tau,\theta',\varphi')\, d\omega' + \frac{\lambda}{4}\, S\, \exp\left(-\tau\sec\theta_0\right) x(\cos\gamma_1), \qquad (2)$$

which expresses the fact that the radiation of unit volume is made up of the energy of rays passing through the volume element in different directions, which is scattered by it, and of the scattered energy of the original beam attenuated $\exp(-\tau \sec\theta_0)$ times on its path to the given unit volume element. In these equations $I(\tau, \theta, \varphi)$ is the intensity of the diffused radiation at the optical depth τ in the direction forming an angle θ with the external normal and having an azimuth φ. The optical depth τ is determined for the usual linear depth z computed from the boundary A in terms of the volume coefficient of extinction of light $\alpha(z)$ by means of the formula:

$$\tau = \int_0^z \alpha(z)\, dz.$$

$B(\tau, \theta, \varphi)$ denotes the escaped radiation defined in terms of the coefficient of radiation $\eta(\tau, \theta, \varphi)$ at a depth τ in the direction θ, φ and α:

$$B(\tau, \theta, \varphi) = \frac{\eta(\tau, \theta, \varphi)}{\alpha}.$$

λ is the ratio of the coefficient of pure scattering to the sum of the coefficients of absorption and pure scattering; πS is the flow of external radiation falling on a unit surface perpendicular to it. And finally, $x(\cos\gamma)$ is a function called the scattering indicatrix, which gives the relative distribution of radiation scattered by a volume element from a direction θ, φ in a direction θ', φ' depending on the scattering angle γ between these two directions; the latter is defined by the formula:

$$\cos\gamma = \cos\theta \cos\theta' + \sin\theta \sin\theta' \cos(\varphi - \varphi'). \tag{3}$$

Instead of $x(\cos\gamma)$ we shall also write

$$x(\theta, \varphi; \theta', \varphi') = x(\cos\gamma).$$

As for the angle γ_1 between the direction θ, φ and the direction of the incident radiation,

$$\cos\gamma_1 = -\cos\theta \cos\theta_0 + \sin\theta \sin\theta_0 \cos(\varphi - \varphi_0).$$

Equations (1) and (2) are usually solved approximately by averaging more or less roughly over the angles, or the system (1) and (2) is solved by reduction to one integral equation for an unknown function $B(\tau, \theta, \varphi)$ from which the intensity of radiation emerging from the medium is obtained by integrating. In our approach, the problem is reduced to a functional equation, which we shall proceed to solve.

Derivation of functional equation

The quantities I and B entering (1) depend not only on the arguments θ and φ, but also on parameters θ_0, φ_0 characterizing the direction of external radiation. In particular, the value of I for $\tau = 0$, i.e., the intensity of the diffuse radiation $I(0, \theta, \varphi)$ emerging from the boundary, which we call the diffusely reflected radiation, will also depend on parameters θ_0, φ_0. We now denote the intensity of this radiation by

$$I(0, \theta, \varphi) = r(\theta, \varphi; \theta_0, \varphi_0)\, S,$$

since due to the linear character of the problem it is proportional to the incident flow πS. The function $r(\theta, \varphi; \theta_0, \varphi_0)$ characterizes the diffuse reflection power of our medium. If the flux incident on the medium is not a beam of parallel rays, but radiation is coming from various directions θ_0, φ_0 where θ_0, as before, is the angle formed by the radiation with the internal normal, then the intensity of the diffusely reflected light $I_2(\theta, \varphi)$ will be equal to

$$I_2(\theta, \varphi) = \frac{1}{\pi} \int_0^{2\pi} \int_0^{\frac{\pi}{2}} r(\theta, \varphi; \theta_0, \varphi_0)\, I_1(\theta_0, \varphi_0) \sin\theta_0 \, d\theta_0 \, d\varphi_0. \qquad (4)$$

We shall look for the function r which we call the reflection function.

Let us return to the case of a parallel beam of rays and construct a plane A' at an optical depth $d\tau$ from the boundary plane A. At the plane A' we have two kinds of radiation penetrating the interior: 1) incident radiation πS, attenuated to the value $\pi S\,(1 - \sec\theta_0\, d\tau)$ and 2) radiation diffused from the above layer of optical depth $d\tau$. Its intensity, as evident from the equation of flux, equals

$$-B(0, \theta, \varphi)\ \sec\theta\, d\tau,$$

where $\sec\theta < 0$ since for flux moving towards the interior of the medium the angle with the external normal $\theta > \frac{\pi}{2}$. If instead of θ we introduce the angle formed with the internal normal $\theta' = \pi - \theta$, then the same intensity is given by

$$B(0, \pi - \theta', \varphi) \sec\theta' \, d\tau.$$

The part of the medium under A' reflects both radiations diffusely; the function r characterizing reflection remains the same, since the removal of a layer $d\tau$ from a medium of infinite optical thickness does not affect the diffuse reflection power of the medium. This invariance of the diffuse reflection power of a medium with respect to the removal or addition of a layer comprises the point of departure in our method.

Utilizing the definition of the function r we find that A' should reflect in the direction θ, φ with an intensity

$$\begin{aligned} &S\, r(\theta, \varphi; \theta_0, \varphi_0)(1 - \sec\theta_0 \, d\tau) + \\ &+ \frac{d\tau}{\pi} \iint r(\theta, \varphi; \theta', \varphi')\, B(0, \pi - \theta', \varphi') \sec\theta' \sin\theta' \, d\theta' \, d\varphi'. \end{aligned}$$

On the other hand, on the basis of the same equation of flux, we can write down the intensity of radiation moving outwards from the plane A'. Indeed, for $\tau = 0$ the intensity of radiation directed outwards equals $r(\theta, \varphi; \theta_0, \varphi_0)\, S$. Hence, according to the equation of flux, at a depth $d\tau$ it will be

$$S\, r(\theta, \varphi; \theta_0, \varphi_0)(1 + \sec\theta \, d\tau) - B(0, \theta, \varphi) \sec\theta \, d\tau.$$

Equating these two expressions for the intensity of the outward flux from the plane A', we obtain:

$$\begin{aligned} &(\sec\theta + \sec\theta_0) r(\theta, \varphi; \theta_0, \varphi_0)\, S = \\ &= B(0, \theta, \varphi) \sec\theta + \frac{1}{\pi} \iint r(\theta, \varphi; \theta', \varphi')\, B(0, \pi - \theta', \varphi') \tan\theta' \, d\theta' \, d\varphi'. \end{aligned} \tag{5}$$

On the other hand, putting $\tau = 0$ in (2), we find for $B(0, \theta, \varphi)$

$$B(0, \theta, \varphi) = \frac{\lambda}{4} \cdot S\, x(\cos\gamma_1) + \frac{\lambda S}{4\pi} \iint x(\cos\gamma)\, I(0, \theta', \varphi') \sin\theta' \, d\theta' \, d\varphi'$$

or, since

$$I(0, \theta', \varphi') = S\, r(\theta', \varphi'; \theta_0, \varphi_0),$$

we have

$$B(0,\theta,\varphi) = \frac{\lambda}{4}\cdot S\,x(\cos\gamma_1) + \\ + \frac{\lambda S}{4\pi}\int_0^{2\pi}\int_0^{\frac{\pi}{2}} x(\theta,\varphi;\theta',\varphi')\,r(\theta',\varphi';\theta_0,\varphi_0)\sin\theta'\,d\theta'\,d\varphi'. \tag{6}$$

Inserting (6) in (5) we find

$$(\sec\theta + \sec\theta_0) r(\theta,\varphi;\theta_0,\varphi_0) = \frac{\lambda}{4}\cdot x(\cos\gamma_1)\sec\theta + \\ + \frac{\lambda}{4\pi}\sec\theta\int_0^{2\pi}\int_0^{\frac{\pi}{2}} x(\theta,\varphi;\theta',\varphi')\,r(\theta',\varphi';\theta_0,\varphi_0)\sin\theta'\,d\theta'\,d\varphi' + \\ + \frac{\lambda}{4\pi}\int_0^{2\pi}\int_0^{\frac{\pi}{2}} r(\theta,\varphi;\theta',\varphi')\,x(\theta',\varphi';\theta_0,\varphi_0)\,\tan\theta'\,d\theta'\,d\varphi' + \\ + \frac{\lambda^2}{4\pi^2}\int_0^{2\pi}\int_0^{\frac{\pi}{2}}\int_0^{2\pi}\int_0^{\frac{\pi}{2}} r(\theta,\varphi;\theta',\varphi')\,x(\pi-\theta',\varphi';\theta'',\varphi'')\times \\ \times r(\theta'',\varphi'';\theta_0,\varphi_0)\tan\theta'\sin\theta''\,d\theta'\,d\varphi'\,d\theta''\,d\varphi''. \tag{7}$$

This is the main functional equation for the function $r(\theta,\varphi;\theta_0,\varphi_0)$ characterizing diffuse reflection. For convenience in writing we shall take the argument of the functions r and x to be $\eta = \cos\theta$. We shall write $\cos\theta' = \eta'$, $\cos\theta'' = \eta''$ also. In this case our functional equation can be rewritten in the form

$$\left(\frac{1}{\eta} + \frac{1}{\eta_0}\right) r(\eta,\varphi;\eta_0,\varphi_0) = \frac{\lambda}{4\eta}\cdot x(\eta,\varphi;-\eta_0,\varphi_0) + \\ + \frac{\lambda}{4\pi\eta}\int_0^{2\pi}\int_0^1 x(\eta,\varphi;\eta',\varphi')\,r(\eta',\varphi';\eta_0,\varphi_0)\,d\eta'\,d\varphi' + \\ + \frac{\lambda}{4\pi}\int_0^{2\pi}\int_0^1 r(\eta,\varphi;\eta',\varphi')\,x(\eta',\varphi';\eta_0,\varphi_0)\,\frac{d\eta'}{\eta'}\,d\varphi' + \\ + \frac{\lambda}{4\pi^2}\int_0^{2\pi}\int_0^1\int_0^{2\pi}\int_0^1 r(\eta,\varphi;\eta',\varphi')\,x(-\eta',\varphi';\eta'',\varphi'')\times \\ \times r(\eta'',\varphi'';\eta_0,\varphi_0)\,\frac{d\eta'}{\eta'}\,d\varphi'\,d\eta''\,d\varphi''. \tag{8}$$

This equation becomes more symmetrical if we insert

$$r(\eta,\varphi;\eta_0,\varphi_0) = \frac{\lambda}{4\eta} R(\eta,\varphi;\eta_0,\varphi_0). \tag{9}$$

Indeed, we shall then have:

$$\left(\frac{1}{\eta}+\frac{1}{\eta_0}\right) R(\eta,\varphi;\eta_0,\varphi_0) = x(\eta,\varphi;-\eta_0,\varphi_0)+$$
$$+\frac{\lambda}{4\pi}\int_0^{2\pi}\int_0^1 x(\eta,\varphi;\eta',\varphi')\,R(\eta',\varphi';\eta_0,\varphi_0)\,\frac{d\eta'}{\eta'}\,d\varphi'+$$
$$+\frac{\lambda}{4\pi}\int_0^{2\pi}\int_0^1 R(\eta,\varphi;\eta',\varphi')\,x(\eta',\varphi';\eta_0,\varphi_0)\,\frac{d\eta'}{\eta'}\,d\varphi'+ \tag{10}$$
$$+\frac{\lambda^2}{16\pi^2}\int_0^{2\pi}\int_0^1\int_0^{2\pi}\int_0^1 R(\eta,\varphi;\eta',\varphi')\,x(-\eta',\varphi';\eta'',\varphi'')\times$$
$$\times\, R(\eta'',\varphi'';\eta_0,\varphi_0)\,\frac{d\eta'}{\eta'}\,d\varphi'\frac{d\eta''}{\eta''}\,d\varphi''.$$

This equation possesses the following property: if it is satisfied by some function $R(\eta,\varphi;\eta_0,\varphi_0)$, then it is also satisfied by the function $R(\eta_0,\varphi_0;\eta,\varphi)$ (arguments reversed). On the other hand, since our physical problem has but one solution, equation (10) also has but one regular solution. Hence,

$$R(\eta,\varphi;\eta_0,\varphi_0) = R(\eta_0,\varphi_0;\eta,\varphi), \tag{11}$$

i.e., the function R is symmetrical. This is in complete agreement with the symmetry of the expression

$$\eta\, r(\eta,\varphi;\eta_0,\varphi_0) = \frac{\lambda}{4}\, R(\eta,\varphi;\eta_0,\varphi_0)$$

derived from physical considerations in [1].

Expression of the indicatrix by means of Legendre polynomials

In several theoretical approaches (Rayleigh's formula), the scattering indicatrix is represented in the form of a finite sum of Legendre polynomials. In the general case it can be developed in a series by Legendre polynomials. If we assume only $n+1$ terms, then

$$x(\cos\gamma) = \sum_{i=0}^{n} x_i\, P_i(\cos\gamma). \tag{12}$$

Since the function $x(\cos\gamma)$ gives the relative distribution of the directions of scattered light in an elementary act of scattering, it satisfies the normalization condition

$$\frac{1}{4\pi}\iint x(\cos\gamma)\,d\omega = 1$$

or

$$\frac{1}{2}\int_0^{\pi} x(\cos\gamma)\sin\gamma\,d\gamma = \frac{1}{2}\int_{-1}^{+1} x(\eta)\,d\eta = 1. \tag{13}$$

Condition (13) gives the value $x_0 = 1$ for the first coefficient in the series (12).

In the case of a spherical scattering indicatrix we have simply

$$x(\cos\gamma) = 1,$$

i.e., only one term in the development.

In the case of the Rayleigh scattering indicatrix ,

$$x(\cos\gamma) = \frac{3}{4}(1+\cos^2\gamma) =$$
$$1 + \frac{1}{2}\left(\frac{3}{2}\cos^2\gamma - \frac{1}{2}\right) = P_0(\cos\gamma) + \frac{1}{2}P_2(\cos\gamma),$$

i.e., in this case $x_0 = 1$ and $x_2 = 1/2$, while all the other coefficients are equal to zero.

Of considerable interest is the group of elongated scattering indicatrices of the type

$$x(\eta) = 1 + x_1\eta,$$

where the quantity x_1 characterizes the degree of elongation of the indicatrix in the direction of the incident ray.

In what follows we shall utilize the development (12) to solve the main functional equation (10), assuming that n can be both finite and infinite.

Solution of the main functional equation

According to the well-known formula for the addition of spherical functions:

$$P_i(\cos\gamma) = P_i\left(\cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\varphi-\varphi')\right) = P_i(\cos\theta)P_i(\cos\theta') +$$
$$+\,2\sum_{m=1}^{i}\frac{(i-m)!}{(i+m)!}P_i^m(\cos\theta)P_i^m(\cos\theta')\cos[m(\varphi-\varphi')].$$

Hence the scattering indicatrix $x(\cos\gamma)$ can be written in the form

$$x(\cos\gamma) = \sum_{m=0}^{\infty}\sum_{i=m}^{\infty} c_{im}\, P_i^m(\cos\theta)\, P_i^m(\cos\theta')\, \cos[m(\varphi-\varphi')] = \\ = \sum_{m=0}^{\infty}\sum_{i=m}^{\infty} c_{im}\, P_i^m(\eta)\, P_i^m(\eta')\, \cos[m(\varphi-\varphi')], \tag{14}$$

where

$$c_{im} = (2-\delta_{0m})\, x_i \frac{(i-m)!}{(i+m)!} \qquad (m = 0, 1, 2, \ldots), \tag{14'}$$

and

$$\delta_{00} = 1 \quad \text{and} \quad \delta_{0m} = 0 \quad \text{if} \quad m \neq 0$$

or in the form

$$x(\cos\gamma) = \sum_{m=0}^{\infty} q_m(\eta,\eta')\cos[m(\varphi-\varphi')], \tag{15}$$

where

$$q_m(\eta,\eta') = \sum_{i=m}^{\infty} c_{im}\, P_i^m(\eta)\, P_i^m(\eta') \tag{15'}$$

are symmetrical functions of their arguments.

The reflection function r, and hence R, depend on the difference between the azimuths $\varphi-\varphi_0$ of the incident and reflected light.

Also R must be an even function of $\varphi-\varphi_0$, since it is invariant with respect to a change in the zero direction of the azimuths. In view of this, in the Fourier development of the function R will have the form

$$R(\eta,\varphi;\eta_0,\varphi_0) = \sum_{m=0}^{\infty} f_m(\eta,\eta_0)\cos[m(\varphi-\varphi_0)]. \tag{16}$$

Our task consists in finding the functions $f_m(\eta,\eta_0)$, i.e., the Fourier coefficients of the function R. We insert (16) and (15) in (10). This gives

$$\left(\frac{1}{\eta}+\frac{1}{\eta_0}\right)\sum_{m=0}^{\infty} f_m(\eta,\eta_0)\cos[m(\varphi-\varphi_0)] =$$
$$\sum_{m=0}^{\infty} q_m(\eta,-\eta_0)\cos[m(\varphi-\varphi_0)]+$$
$$+\frac{\lambda}{2}\sum_{m=0}^{\infty}\frac{\cos[m(\varphi-\varphi_0)]}{2-\delta_{0m}}\int_0^1 q_m(\eta,\eta')\,f_m(\eta',\eta_0)\frac{d\eta'}{\eta'}+$$
$$+\frac{\lambda}{2}\sum_{m=0}^{\infty}\frac{\cos[m(\varphi-\varphi_0)]}{2-\delta_{0m}}\int_0^1 q_m(\eta',\eta_0)\,f_m(\eta,\eta')\frac{d\eta'}{\eta'}+$$
$$+\frac{\lambda^2}{4}\sum_{m=0}^{\infty}\frac{\cos[m(\varphi-\varphi_0)]}{(2-\delta_{0m})^2}\int_0^1\int_0^1 f_m(\eta,\eta')q_m(-\eta',\eta'')f_m(\eta'',\eta_0)\frac{d\eta'}{\eta'}\frac{d\eta''}{\eta''}.$$

Equating the coefficients of $\cos[m(\varphi-\varphi_0)]$ on both sides we find

$$\left(\frac{1}{\eta}+\frac{1}{\eta_0}\right) f_m(\eta,\eta_0) = q_m(\eta-\eta_0)+$$
$$+\frac{\lambda}{2(2-\delta_{0m})}\int_0^1 [q_m(\eta,\eta')\,f_m(\eta',\eta_0)+q_m(\eta',\eta_0)\,f_m(\eta,\eta')]\,\frac{d\eta'}{\eta'}+ \qquad (17)$$
$$+\frac{\lambda^2}{4(2-\delta_{0m})^2}\int_0^1\int_0^1 f_m(\eta,\eta')\,q_m(-\eta',\eta'')\,f_m(\eta'',\eta_0)\frac{d\eta'}{\eta'}\,\frac{d\eta''}{\eta''}.$$

We thus have obtained functional equations for the Fourier coefficients $f_m(\eta,\eta')$. Each $f_m(\eta,\eta_0)$ can be found separately.

It is evident from formulas (14) and (15) that if the development of the scattering indicatrix by Legendre polynomials contains only a finite number of terms, n being the degree of the highest term in the development, then for $m > n$ all the $q_m(\eta,\eta_0)$ disappear. Equations (17) show that in this case for $m > n$ the f_m also disappear.

In the Fourier development (16) of the reflection function the number of terms is equal to the degree of the highest term in the development (12) of the scattering indicatrix by Legendre polynomials.

Let us now investigate equations (17) for the functions $f_m(\eta,\eta_0)$. To this end we insert (15) in (17), which yields

$$\left(\frac{1}{\eta}+\frac{1}{\eta_0}\right) f_m(\eta,\eta_0) = \sum_{i=m}^{\infty}(-1)^{i+m} c_{im} P_i^m(\eta)\, P_i^m(\eta_0) +$$

$$+\frac{\lambda}{2(2-\delta_{0m})}\sum_{i=m}^{\infty} c_{im} P_i^m(\eta)\int_0^1 P_i^m(\eta')\, f_m(\eta',\eta_0)\frac{d\eta'}{\eta'}+$$

$$+\frac{\lambda}{2(2-\delta_{0m})}\sum_{i=m}^{\infty} c_{im} P_i^m(\eta_0)\int_0^1 P_i^m(\eta')\, f_m(\eta,\eta')\frac{d\eta'}{\eta'}+$$

$$+\sum_{i=m}^{\infty}\frac{\lambda^2(-1)^{i+m}c_{im}}{4(2-\delta_{0m})^2}\int_0^1 P_i^m(\eta')f_m(\eta,\eta')\frac{d\eta'}{\eta'}\int_0^1 P_i^m(\eta'')f_m(\eta'',\eta_0)\frac{d\eta''}{\eta''}.$$

We see that the right-hand side can be represented as a sum of products:

$$\left(\frac{1}{\eta}+\frac{1}{\eta_0}\right) f_m(\eta,\eta_0) =$$

$$= \sum_{i=m}^{\infty}(-1)^{i+m} c_{im}\left[P_i^m(\eta)+\frac{(-1)^{i+m}\lambda}{2(2-\delta_{0m})}\int_0^1 f_m(\eta,\eta')P_i^m(\eta')\frac{d\eta'}{\eta'}\right]\times \tag{18}$$

$$\times\left[P_i^m(\eta_0)+\frac{(-1)^{i+m}\lambda}{2(2-\delta_{0m})}\int_0^1 f_m(\eta',\eta_0)\, P_i^m(\eta')\,\frac{d\eta'}{\eta'}\right].$$

In view of the symmetry of the function $f_m(\eta,\eta_0)$ the two factors in parentheses, entering each term of the sum in formula (18), represent the same functions, one depending on η, the other on η_0. In other words, (18) can be rewritten

$$\left(\frac{1}{\eta}+\frac{1}{\eta_0}\right) f_m(\eta,\eta_0) = \sum_{i=m}^{\infty}(-1)^{i+m} c_{im}\, \varphi_i^m(\eta)\, \varphi_i^m(\eta_0), \tag{19}$$

where

$$\varphi_i^m(\eta) = P_i^m(\eta)+\frac{(-1)^{i+m}\lambda}{2(2-\delta_{0m})}\int_0^1 f_m(\eta,\eta')\, P_i^m(\eta')\,\frac{d\eta'}{\eta'}. \tag{20}$$

It follows from (19) that the function $f_m(\eta,\eta_0)$ has the structure

$$f_m(\eta,\eta_0) = \sum_{i=m}^{\infty}(-1)^{i+m}\frac{c_{im}\varphi_i^m(\eta)\cdot\varphi_i^m(\eta_0)}{\frac{1}{\eta}+\frac{1}{\eta_0}}. \tag{21}$$

The equations for the auxiliary functions $\varphi_i^m(\eta)$ are obtained by inserting (21) into the right-hand side of (20):

$$\varphi_i^m(\eta) = P_i^m(\eta) + \\ + \frac{(-1)^{i+m}\lambda}{2(2-\delta_{0m})} \int_0^1 \sum_{k=m}^{\infty} \frac{(-1)^{k+m} c_{km}\, \varphi_k^m(\eta)\, \varphi_k^m(\eta')}{\frac{1}{\eta}+\frac{1}{\eta'}} P_i^m(\eta') \cdot \frac{d\eta'}{\eta'}$$

or (using (14′))

$$\varphi_i^m(\eta) = P_i^m(\eta) + \\ + \frac{\lambda}{2} \sum_{k=m}^{\infty} (-1)^{i+k}\, \eta\, \frac{(k-m)!}{(k+m)!} \int_0^1 \frac{x_k\, \varphi_k^m(\eta)\, \varphi_k^m(\eta')}{\eta+\eta'} P_i^m(\eta')\, d\eta'. \qquad (22)$$

Putting $i = m, m+1, \ldots$, we derive from (22) a system of functional equations for the unknown functions $\varphi_m^m(r), \varphi_{m+1}^m(r), \ldots$ For each m the number of unknown functions is equal to the number of equations. If n is the degree of the highest Legendre polynomial in (12), then the total number of unknown functions as well as the number of functional equations is $\frac{(n+1)(n+2)}{2}$. However, the entire group of equations divides into a number of subsystems, corresponding to various m. The subsystem for the function φ_i^m, where m has a fixed value, can be solved independently of the other subsystems.

We summarize: the unknown function R of four variables by means of formulas (16) and (21) is expressed via functions of one variable $\varphi_i^m(\eta)$, the latter functions themselves being determined by the system of functional equations (22).

This form of representing R is especially convenient in the case of a finite development (12) of the scattering indicatrix, as seen from the particular examples which follow.

Spherical scattering indicatrix

In this case

$$x(\cos\gamma) = 1 \qquad (23)$$

and all the x_i's are zero for $i > 0$. The highest polynomial in the development of the scattering indicatrix is of zero degree. Therefore, only one term is left in the development (16), corresponding to $m = 0$

$$R = f_0(\eta, \eta_0). \tag{24}$$

Formula (21) is reduced in this case to

$$f_0(\eta, \eta_0) = \frac{\varphi_0^0(\eta)\,\varphi_0^0(\eta_0)}{\frac{1}{\eta} + \frac{1}{\eta_0}}, \tag{25}$$

and the unique auxiliary function $\varphi_0^0(\eta)$ is determined from the one functional equation

$$\varphi_0^0(\eta) = 1 + \frac{\lambda}{2}\eta\varphi_0^0(\eta)\int_0^1 \frac{\varphi_0^0(\eta')\,d\eta'}{\eta + \eta'}, \tag{26}$$

to which system (22) is reduced. For each λ this system is easily solved numerically by the method of successive approximations.

The case of spherical scattering indicatrix has been discussed by us in [2] where tables for different values of λ obtained are given.

Elongated scattering indicatrix

Consider an indicatrix of the type

$$x(\cos\gamma) = 1 + x_1 \cos\gamma. \tag{27}$$

If $x_1 > 0$ such an indicatrix is elongated in the direction $\gamma = 0$, whereas in the case of $x_1 < 0$ it is elongated in the direction $\gamma = \pi$.

In these cases the degree of the highest Legendre polynomials in the development of the indicatrix is one. Hence (16) reduces to

$$R(\eta, \varphi; \eta_0, \varphi_0) = f_0(\eta, \eta_0) + f_1(\eta, \eta_0)\cos(\varphi - \varphi_0). \tag{28}$$

As for f_0 and f_1, according to (21), they have the following structure:

$$\begin{aligned} f_0(\eta, \eta_0) &= \frac{\varphi_0^0(\eta)\,\varphi_0^0(\eta_0) - x_1\varphi_1^0(\eta)\,\varphi_1^0(\eta_0)}{\frac{1}{\eta} + \frac{1}{\eta_0}}, \\ f_1(\eta, \eta_0) &= x_1\,\frac{\varphi_1^1(\eta)\,\varphi_1^1(\eta_0)}{\frac{1}{\eta} + \frac{1}{\eta_0}}. \end{aligned} \tag{29}$$

The auxiliary functions φ_0^0 and φ_1^0 should be determined, according to (22), from the subsystems

$$\varphi_0^0(\eta) = 1 + \frac{\lambda}{2}\,\eta\,\varphi_0^0(\eta)\int_0^1 \frac{\varphi_0^0(\eta')\,d\eta'}{\eta+\eta'} - \frac{\lambda}{2}\,x_1\,\eta\,\varphi_1^0(\eta)\int_0^1 \frac{\varphi_1^0(\eta')\,d\eta'}{\eta+\eta'},$$

$$\varphi_1^0(\eta) = \tag{30}$$

$$= \eta - \frac{\lambda}{2}\,\eta\,\varphi_0^0(\eta)\int_0^1 \frac{\varphi_0^0(\eta')\,\eta'\,d\eta'}{\eta+\eta'} + \frac{\lambda}{2}\,x_1\,\eta\,\varphi_1^0(\eta)\int_0^1 \frac{\varphi_1^0(\eta')\,\eta'\,d\eta'}{\eta+\eta'},$$

and the auxiliary function $\varphi_1^1(\eta)$ from the equation

$$\varphi_1^1(\eta) = \sqrt{1-\eta^2} + \frac{\lambda}{4}\,x_1\,\eta\,\varphi_1^1(\eta)\int_0^1 \frac{\varphi_1^1(\eta')\sqrt{1-(\eta')^2}}{\eta+\eta'}\,d\eta'. \tag{31}$$

The last equation can be solved numerically by successive approximations. By multiplying by $\sqrt{1-\eta^2}$ and substituting $\varphi_1^1(\eta)\sqrt{1-\eta^2} = \psi(\eta)$ we bring it to the convenient form

$$\psi(\eta) = 1 - \eta^2 + \frac{\lambda\,x_1}{4}\eta\,\psi(\eta)\int_0^1 \frac{\psi(\eta')\,d\eta'}{\eta+\eta'}. \tag{32}$$

A simple relation exists between the functions φ_0^0 and φ_0^1 given by (30). To establish it we transform the second equation of this system by substituting

$$\frac{\eta'}{\eta+\eta'} = 1 - \frac{\eta}{\eta+\eta'}.$$

Hence

$$\varphi_1^1(\eta) = 1 - \frac{\lambda}{2}\cdot\eta\,\varphi_0^0(\eta)\int_0^1 \varphi_0^0(\eta')\,d\eta' + \frac{\lambda\,x_1}{2}\cdot\eta\,\varphi_1^0(\eta)\int_0^1 \varphi_1^0(\eta')\,d\eta' +$$
$$+ \frac{\lambda}{2}\eta^2\varphi_0^0(\eta)\int_0^1 \frac{\varphi_0^0(\eta')\,d\eta'}{\eta+\eta'} - \frac{\lambda\,x_1}{2}\eta^2\varphi_1^0(\eta)\int_0^1 \frac{\varphi_1^0(\eta')\,d\eta'}{\eta+\eta'}.$$

On the basis of the first equation in (30), the last two terms here can be replaced by

$$\eta\left(\varphi_0^0(\eta) - 1\right).$$

Therefore, in terms of constants:

$$\alpha = \int_0^1 \varphi_0^0(\eta)\,d\eta, \qquad \beta = \int_0^1 \varphi_1^0(\eta)\,d\eta, \tag{33}$$

we have

$$\varphi_1^0(\eta) = \frac{\left(1 - \frac{\lambda}{2} \cdot \alpha\right) \eta \, \varphi_0^0(\eta)}{1 - \frac{\lambda x_1}{2} \cdot \beta\eta}. \tag{34}$$

Table 1. Values of $\varphi_0^0(\eta)$.
Scattering indicatrix $x(\cos\gamma) = 1 + \cos\gamma$

$\eta\backslash\lambda$	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.1	1.054	1.070	1.088	1.109	1.134	1.166	1.248
0.2	1.080	1.106	1.136	1.171	1.215	1.276	1.450
0.3	1.097	1.130	1.168	1.216	1.276	1.365	1.642
0.4	1.108	1.146	1.192	1.249	1.324	1.439	1.829
0.5	1.115	1.157	1.208	1.274	1.382	1.502	2.013
0.6	1.119	1.163	1.219	1.291	1.391	1.535	2.194
0.7	1.120	1.164	1.226	1.304	1.414	1.599	2.375
0.8	1.120	1.167	1.236	1.312	1.431	1.637	2.552
0.9	1.118	1.166	1.230	1.317	1.443	1.669	2.730
1.0	1.115	1.164	1.228	1.318	1.452	1.696	2.909

Table 2. Values of $\varphi_1^0(\eta)$.
Scattering indicatrix $x(\cos\gamma) = 1 + \cos\gamma$

$\eta\backslash\lambda$	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.1	0.083	0.077	0.071	0.064	0.054	0.041	0.000
0.2	0.172	0.162	0.150	0.136	0.118	0.090	0.000
0.3	0.264	0.251	0.235	0.215	0.189	0.177	0.000
0.4	0.359	0.344	0.324	0.299	0.265	0.210	0.000
0.5	0.456	0.439	0.417	0.387	0.346	0.277	0.000
0.6	0.555	0.536	0.512	0.478	0.431	0.349	0.000
0.7	0.654	0.635	0.609	0.572	0.519	0.425	0.000
0.8	0.755	0.734	0.708	0.669	0.610	0.505	0.000
0.9	0.856	0.836	0.808	0.767	0.704	0.588	0.000
1.0	0.959	0.939	0.910	0.867	0.800	0.674	0.000

Thus $\varphi_1^0(\eta)$ is expressed through $\varphi_0^0(\eta)$, but two constants α and β enter, which are determined from (33). The first equation in (30) in conjunction with (34) and (33) completely determines the functions φ_0^0 and

φ_1^0. The numerical values of the functions φ_0^0, φ_1^0 and φ_1^1 obtained by the method of successive approximations for different λ and for $x_1 = 1$ are given in Tables 1, 2, and 3. Our solutions are accurate up to the third decimal.

Table 3. Values of $\varphi_1^1(\eta)$.
Scattering indicatrix $x(\cos\gamma) = 1 + \cos\gamma$

$\eta\backslash\lambda$	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.1	1.016	1.021	1.027	1.032	1.038	1.044	1.051
0.2	1.010	1.018	1.026	1.034	1.043	1.053	1.062
0.3	0.988	0.998	1.007	1.018	1.028	1.040	1.050
0.4	0.954	0.964	0.975	0.986	0.997	1.009	1.022
0.5	0.903	0.915	0.926	0.938	0.951	0.982	0.976
0.6	0.836	0.647	0.859	0.870	0.883	0.895	0.908
0.7	0.749	0.759	0.770	0.780	0.790	0.804	0.815
0.8	0.632	0.640	0.648	0.658	0.668	0.679	0.688
0.9	0.459	0.466	0.4730	0.479	0.486	0.494	0.502
1.0	0.000	0.000	0.000	0.000	0.000	0.000	0.000

For an indicatrix of the type considered, special attention is attached to the case $\lambda = 1$ (i.e., pure scattering, in the absence of absorption) with arbitrary x_1. In this case the system (30) is satisfied if we put φ_1^0 and φ_0^0 equal to the solution of (26) for a spherical scattering indicatrix.

If we put $\lambda = 1$ and $\varphi_1^0 = 0$ in (30), then our two equations are reduced to the following:

$$\varphi_0^0(\eta) = 1 + \frac{1}{2}\cdot\eta\,\varphi_0^0(\eta)\int_0^1 \frac{\varphi_0^0(\eta')\,d\eta'}{\eta+\eta'}, \tag{35}$$

$$1 = \frac{1}{2}\cdot\varphi_0^0(\eta)\int_0^1 \frac{\varphi_0^0(\eta')\,\eta'\,d\eta'}{\eta+\eta'}. \tag{36}$$

The first of these is identical to (26). Let us prove that the second is also equivalent to (26).

Under the integral sign in the second equation we substitute

$$\frac{\eta'}{\eta+\eta'} = 1 - \frac{\eta}{\eta+\eta'}$$

and obtain

$$1 = \frac{1}{2}\varphi_0^0(\eta) \int_0^1 \varphi_0^0(\eta)\, d\eta' - \frac{1}{2}\eta\varphi_0^0(\eta) \int_0^1 \frac{\varphi_0^0(\eta')\, d\eta'}{\eta + \eta'}. \tag{37}$$

On the other hand, integrating (36) with respect to η we obtain:

$$1 = \frac{1}{2} \int_0^1 \int_0^1 \frac{\varphi_0^0(\eta)\, \varphi_0^0(\eta')\, \eta'\, d\eta'\, d\eta}{\eta + \eta'} =$$
$$= \frac{1}{4} \int_0^1 \int_0^1 \varphi_0^0(\eta)\, \varphi_0^0(\eta')\, d\eta\, d\eta' = \frac{1}{4} \left[\int_0^1 \varphi_0^0(\eta)\, d\eta \right]^2,$$

i.e.,

$$\int_0^1 \varphi_0^0(\eta)\, d\eta = 2.$$

Substituting this value in (37) we see that (36) reduces to (26). Thus, for $\lambda = 1$ both equations (30) are satisfied if $\varphi_0^0(\eta)$ is taken as equal to the solution of (26), and φ_1^0 equal to zero.

Therefore, on the basis of (28) and (29) we can write for the function of reflection R:

$$R = \frac{\varphi_0^0(\eta)\, \varphi_0^0(\eta_0)}{\frac{1}{\eta} + \frac{1}{\eta_0}} + x_1 \cdot \frac{\varphi_1^1(\eta)\, \varphi_1^1(\eta_0)}{\frac{1}{\eta} + \frac{1}{\eta_0}} \cdot \cos(\varphi - \varphi_0).$$

After averaging over the azimuth, the second term disappears and we come to the following remarkable result.

In the case of pure scattering the reflection function for an elongated indicatrix $x(\cos\gamma) = 1 + x_1 \cos\gamma$, averaged over the azimuth, is exactly equal to the reflection function (25) for a spherical scattering indicatrix. For $\lambda \neq 1$ this rule no longer holds.

Remark on Lambert's law

In photometry in addition to the reflection function $r(\eta, \varphi; \eta_0, \varphi_0)$, use is often made of the coefficient of brightness

$$\rho = \frac{r}{\eta} = \frac{\lambda}{4\eta\eta_0} R.$$

According to Lambert's empirical law for certain material in which λ is close to unity, ρ is constant. In the case of a spherical scattering indicatrix it follows from (25) that

$$\rho = \frac{\lambda}{4} \frac{\varphi_0^0(\eta)\, \varphi_0^0(\eta_0)}{\eta + \eta'}.$$

For $\lambda = 1$, using the values of $\varphi_0^0(\eta)$, we present in Table 4 the values of the coefficient of brightness for various pairs of values of η and η_0. We see that for angles θ and θ_0, not too close to 90°, the quantity ρ is almost constant and has values close to unity. On the basis of the results of the preceding paragraph, we can assert that for a spherical indicatrix of the type $1 + x_1 \cos\gamma$ the mean (with respect to the azimuth) coefficient of brightness is almost constant and close to unity. For an arbitrary scattering indicatrix and for $\lambda = 1$, the coefficient of brightness, averaged over the azimuth, for not too large angles with the normal, is almost constant and close to unity.

Table 4. Spherical scattering indicatrix.
Coefficients of brightness for $\lambda = 1.0$

$\eta \backslash \eta$	−0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.0	—	3.12	1.810	1.370	1.140	1.010	0.913	0.849	0.798	0.758	0.728
0.1	3.12	1.95	1.510	1.280	1.140	1.050	0.977	0.926	0.884	0.862	0.825
0.2	1.81	1.51	1.320	1.190	1.110	1.040	0.994	0.957	0.925	0.900	0.878
0.3	1.37	1.28	1.190	1.130	1.070	1.030	1.000	0.975	0.953	0.946	0.918
0.4	1.14	1.14	1.110	1.070	1.050	1.020	1.000	0.987	0.972	0.961	0.951
0.5	1.01	1.05	1.040	1.030	1.020	1.010	1.000	0.997	0.988	0.981	0.977
0.6	0.913	0.977	0.994	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.998
0.7	0.849	0.926	0.957	0.975	0.987	0.997	1.000	1.010	1.010	1.010	1.020
0.8	0.798	0.884	0.925	0.953	0.972	0.988	1.000	1.010	1.020	1.020	1.030
0.9	0.759	0.852	0.900	0.946	0.961	0.981	0.998	1.010	1.020	1.030	1.040
1.0	0.728	0.825	0.878	0.918	0.951	0.977	0.998	1.020	1.030	1.040	1.060

Hence, the theory brings us to the following conclusions and limitations in the applicability of Lambert's law:

1) It is applicable only to purely scattering media.

2) It refers only to a coefficient of brightness averaged over the azimuth and does not take into account the dependence on the azimuth.

3) Even with these limitations it is valid only for not too large angles of incidence and reflection (up to 70°).

This method of solving the classical problem of the scattering of light can be easily generalized for the case of layers of finite optical thickness.

REFERENCES

1. Minnaert, Astrophys. Journ., 93, 403 (1941).
2. Ambartsumian, Astronom. Journ., 19, No. 1 (1942).

30 September 1943 Leningrad State University

THE THEORY OF FLUCTUATIONS OF THE SURFACE BRIGHTNESS IN THE MILKY WAY

At present we can take as established that the absorbing layer in our Galaxy has a patchy structure, i.e., it consists of a large number of absorbing clouds whose cumulative effect is responsible for the general cosmic absorption. The absorbing clouds which are nearer to us and have higher absorbing power appear to us as *dark nebulae.* A large number of papers devoted to the study of dark nebulae has been published. However, a statistical study of the whole complex of absorbing clouds, which sometimes have low absorbing power, is still missing.

In an earlier paper [1], the author has shown that the fluctuations in the number of extragalactic nebulae are at least partly caused by the patchy structure of the galactic absorbing layer. In [1] some mean value of optical thickness of a cloud was deduced. It turned out to be equal to $0^m.27$.

We can adopt as a working hypothesis that the fluctuations of surface brightness along the Milky Way are also caused by absorbing clouds. Apparently such an assumption can be considered as the first approximation to the real situation. Then the question arises about the distribution law of the brightness fluctuations. In the following paragraphs we derive a differential equation which determines the distribution function. From the latter we find the values of moments of all orders.

1. Let us suppose that the equatorial plane of the Galaxy is homogeneously filled by stars and absorbing clouds up to an *infinite* distance.

This model assumption is not as bad as it seems, since absorption at very large distances is almost complete, and distant stars and clouds do not influence the observed brightnesses. At the same time we will assume that during the passage of light through each cloud the same fraction of the intensity is absorbed. The transparency of a cloud we will denote by q.

If the number of clouds in a certain direction at distances less than s is $n(s)$ (this is, of course, a random function), then the light of a star placed at distance s will diminish by the factor $q^{n(s)}$. Assume that an element dV

of galactic space emits total energy $4\pi\eta dV$. Then the observed brightness in any direction within the galactic plane will be

$$\int_0^\infty q^{n(s)}\,\eta\,ds.$$

We consider the distribution function of values of this integral:

$$f(I) = P\left(\int_0^\infty q^{n(s)}\,\eta\,ds > I\right).$$

For small values of $a > 0$ we can write

$$f(I) = P\left(\int_0^a q^{n(s)}\,\eta\,ds + q^{n(a)}\int_a^\infty q^{n(s)-n(a)}\eta\,ds > I\right). \tag{1}$$

Since a is small, the number $n(a)$ can take on only two values, $n(a) = 0$ and $n(a) = 1$. The probability of 0 is $1 - \nu a$ and of 1 is νa. Here ν is the mean number of clouds per unit length of light path. The probabilities of other values of $n(s)$ are small numbers of higher orders and can be neglected.

Correspondingly, the integral $\int_0^a q^{n(s)}\,\eta\,ds$ can take on either a value ηa or $\eta\theta a$, where $0 < \theta < 1$. Therefore,

$$\begin{aligned} f(I) =& (1-\nu a)\,P\left(\int_a^\infty q^{n(s)-n(a)}\eta\,ds > I - a\eta\right) + \\ & + \nu a\,P\left(\int_0^\infty q^{n(s)-n(a)}\eta\,ds > \frac{I-\eta\theta a}{q}\right). \end{aligned} \tag{2}$$

But owing to homogeneous distribution of clouds in space

$$P\left(\int_a^\infty q^{n(s)-n(a)}\eta\,ds > I\right) = P\left(\int_0^\infty q^{n(s)}\,\eta\,ds > I\right),$$

and equation (2) can be rewritten in the form

$$f(I) = (1-\nu a)\,f(I - a\eta) + \nu a\,f\left(\frac{I-\eta\theta a}{q}\right). \tag{3}$$

Up to the terms of the second order in a, this is equivalent to

$$f(I) = f(I) - \nu a\,f(I) - a\eta\,f'(I) + \nu a\,f\left(\frac{I}{q}\right).$$

From this

$$f(I) + \frac{\eta}{\nu} f'(I) = f\left(\frac{I}{q}\right) \tag{4}$$

or using a new variable $u = I\frac{\nu}{\eta}$

$$f(u) + f'(u) = f\left(\frac{u}{q}\right). \tag{5}$$

By differentiating, for the density $g(u) = f'(u)$ we obtain the equation

$$g(u) + g'(u) = \frac{1}{q} g\left(\frac{u}{q}\right). \tag{6}$$

From this functional equation it is possible to find the mean values of all powers of brightness.

From (5) we can see that

$$f'(0) = g(0) = 0.$$

Now multiplying (6) by u and integrating, we find

$$\overline{u} + \int_0^\infty g'(u)\, u\, du = q\overline{u},$$

where $\overline{u}$ is the mean brightness.

Integrating by parts and taking into account that

$$\int_0^\infty g(u)\, du = 1,$$

we find

$$\overline{u} = \frac{1}{1-q}. \tag{7}$$

Multiplying (6) by u^2 and integrating, we get

$$\overline{u^2}\,(1-q^2) = -\int_0^\infty g'(u)\, u^2\, du = 2\overline{u},$$

i.e., the mean value of the squared brightness equals

$$\overline{u^2} = \frac{2\overline{u}}{1-q^2} = \frac{2}{(1-q)^2(1+q)}.$$

For the relative square deviation, we obtain

$$\frac{\overline{(u-\overline{u})^2}}{\overline{u}^2} = \frac{\overline{u^2}}{\overline{u}^2} - 1 = \frac{1-q}{1+q}, \tag{8}$$

i.e., it is completely determined by the transparency of one cloud.

2. Now we abandon the assumption that all absorbing clouds have the same optical thickness. Rather let them have random optical thicknesses. However, we assume that in different parts of space the distribution of transparency q remains the same. Let the probability of having the value of transparency between q and $q + dq$ be $d\varphi(q)$.

In this case for the distribution function f of brightness, we obtain the following generalization of (5)

$$f(u) + f'(u) = \int_0^1 f(u/q)\, d\varphi(q). \tag{9}$$

For the probability density $g(u) = -f'(u)$, we get

$$g(u) + g'(u) = \int_0^1 g\left(\frac{u}{q}\right) \frac{d\varphi(q)}{q}. \tag{10}$$

The moments $\overline{u}$ and $\overline{u^2}$ are found to be

$$\overline{u} = \frac{1}{1 - \int_0^1 q\, d\varphi(q)} = \frac{1}{1 - \overline{q}}; \qquad \overline{u^2} = \frac{2\overline{u}}{1 - \int_0^1 q^2\, d\varphi(q)} = \frac{2\overline{u}}{1 - \overline{q^2}}, \tag{11}$$

where $\overline{q}$ and $\overline{q^2}$ are the mean values

$$\overline{q} = \int_0^1 q\, d\varphi(q); \qquad \overline{q^2} = \int_0^1 q^2\, d\varphi(q).$$

Since $\overline{q^2} > \overline{q}^2$, we have

$$\overline{u^2} > \frac{2\overline{u}}{1 - \overline{q}^2}.$$

It turns out that given $\overline{q}$, the relative mean square deviation

$$\frac{\overline{u^2} - \overline{u}^2}{\overline{u}^2} = \frac{\int_0^1 (1-q)^2\, d\varphi(q)}{\int_0^1 (1-q^2)\, d\varphi(q)}$$

is minimal in the case of clouds of identical transparency equal to $\overline{q}$.

For a number of simpler laws $\varphi(q)$, Laplace transforms of the solutions of equation (10) can be derived. In this way we find that when $\varphi(q) = q^k$, equation (10) has a solution $g(u) = u^k\, e^{-u}(k!)^{-1}$.

REFERENCES

1. Ambartsumian, V. A., *Bull. Abast. Obs.*, 4, 17 (1940).

11 March 1944

Yerevan
Astronomical Observatory of the Academy
of Sciences of Armenia

SURFACE BRIGHTNESSES IN OUR GALAXY

§**1.** ***Integral surface brightnesses.*** Stellar counts are one of the main tools in any study of the structure of our Galaxy. However, the use of stellar counts is connected with certain difficulties both of a principal and practical nature. In particular, it requires knowledge of the luminosity function and its changes in space.

However, it is possible to obtain *some* data on the structure of the Galaxy avoiding detailed analysis of stellar counts. Sometimes it is sufficient to study the *surface brightness* produced by stars, i.e., the sums of apparent brightnesses of stars, seen in a unit solid angle (for example per square degree). We shall use below the concept of "integral surface brightness" in this particular sense. This means that other sources of brightness, such as diffuse illumination of the night sky or zodiacal light, are excluded. We also exclude "galactic light," i.e., the light scattered by interstellar dust and by diffuse nebulae, as well as all extragalactic sources.

Unfortunately, at the moment the only source of data concerning total surface brightnesses is the integration of counts of stars of different magnitudes according to the formula

$$I = \int_{-\infty}^{+\infty} A(m) 10^{0.4(m_0 - m)}\, dm, \qquad (1)$$

where the unit in which I is measured is one star of magnitude m_0 and $A(m)$ is the number of stars per square degree with magnitudes between $m - \frac{1}{2}$ and $m + \frac{1}{2}$. Of course, the quantity I is a function of celestial coordinates.

An attempt to determine the quantity I from observed brightness of the sky was made in a recent paper by Fesenkov [1]. Further accumulation and improvement of observations of this type is extremely desirable.

Calculating I following (1), we encounter difficulties of two types. Sometimes faint stars (weaker than 20.0) which avoided the count make a substantial contribution to the integral (1). This contribution should be calculated by means of extrapolation. On the other hand, for the brightest stars the values of $A(m)$ become indefinite. Because of this, I is strongly dependent on the occasional presence of one or two brighter stars in the field under consideration.

Nevertheless the counts of Van Rhijn [2] together with the data on individual brightest stars enable us to determine approximate values of I for different directions in the sky.

§**2.** ***Partial surface brightnesses.*** Along with the total surface brightnesses we can consider the *partial surface brightnesses* caused by the radiation of stars of a certain physical type. Thus, we can speak about the partial surface brightness generated by F-type or G-type stars.

The partial surface brightnesses are calculable by means of the same formula (1), where $A(m)$ now represents the numbers of stars of the given type.

The aforementioned difficulties in calculation of total surface brightnesses are also present for the partial surface brightnesses, especially due to a scarcity of data on the spectral types of faint stars. The material for calculation of partial surface brightnesses for different spectral classes is supplied by spectral catalogues like HDC, BSD and HDE. Of course, if the lists of stars of different types are not overlapping, then the sum of the partial surface brightnesses will be equal to the total surface brightness.

Some astronomers have recently introduced the concept of a subsystem of stars of the same physical type (Lindblad [3], Kukarkin [4]). Evidently the total surface brightness is the sum of surface brightnesses calculated for such subsystems.

§**3.** ***The emission coefficients.*** Calculating the surface brightnesses, we have to consider the total emission coefficient as well as the partial emission coefficients. The total emission coefficient η is defined as the total brightness of all stars in a cubic parsec expressed in units equal to the radiation of a star of some standard absolute magnitude M_0.

If $\Phi(M)$ is the luminosity function normalized in such a way that

$$\int \Phi(M)\, dM = n, \tag{2}$$

where n is the total number of stars in a cubic parsec, then

$$\eta = \int \Phi(M) 10^{0.4(M_0 - M)}\, dM. \tag{3}$$

In the same way the partial emission coefficients can be defined for stars belonging to different physical types. Passage from numbers of stars in a unit volume or unit solid angle to the emission coefficient and the surface brightness corresponds to passage from a "microscopic" to a "macroscopic" point of view.

The macroscopic approach has in fact long been applied to the study of external galaxies, for example, to the study of the ellipticals (in particular in the work of Oort [5] on ellipticals NGC 3115 and 4494). In galactic astronomy it was also used in a few cases, for example, in the paper of Henyey and Greenstein [6] on the light scattered by cosmic dust and in the author's studies [7,8] on fluctuations of surface brightness of the Milky Way.

In particular, Henyey and Greenstein obtained formula (9) of this paper and did a comparison similar to that presented in Table 1 below. However, they missed the opportunity to make inferences about the relative distribution of dark and bright matter.

Let us note that the value of the total emission coefficient in the vicinity of the Sun can be successfully deduced from the Van Rhijn luminosity function and equals 4.8 stars of absolute magnitude 10.0 in a photographic system and 6.1 stars of the same absolute magnitude in the visual part of the spectrum. This means that the color index of the mean radiation in the vicinity of the Sun is about $+0^m.26$.

§4. ***The coefficient of absorption and the optical distance from the plane of the Galaxy.*** In the following discussion we digress from the patchy structure of the absorbing medium in the Galaxy and consider a macroscopic coefficient of absorption α assumed to be a continuous function of coordinates. For any point in the Galaxy we form an integral over values of α along the perpendicular to the equatorial plane of the Galaxy:

$$\tau = \int_0^z \alpha\, dz, \tag{4}$$

where z is the distance of the point from the equatorial plane. We call τ the optical distance from the equatorial plane. When z is increasing indefinitely, τ is approaching τ_0, half the total optical thickness of the absorbing layer:

$$\tau_0 = \int_0^\infty \alpha\, dz. \tag{5}$$

Of course, τ is a function of x, y and z, and τ_0 is a function of x and y only. In the approximation of plane-parallel layers (in the vicinity of the Sun), we consider τ_0 to be a constant. In the same approximation η is a function of z and, therefore, can be considered a function of τ.

The value of τ_0 for the photographic part of the spectrum is $0^m.25$ according to Hubble [9] and is $0^m.34$ according to Parenago [10].

§5. ***The total photographic brightness in the case of plane-parallel layers.*** As we know from the work of Vashakidze [11] and Oort [12], the model of plane-parallel layers is in satisfactory agreement with the observed distribution of stars by apparent magnitudes for high galactic latitudes. Although deeper study reveals some deviations from such a model, we will use it as a satisfactory approximation. In this model both η and α are functions of τ and so is their ratio

$$\frac{\eta}{\alpha} = B(\tau).$$

The surface brightness at the galactic latitude b in this model will be

$$I(b) = \int_0^{\tau_0} \exp\left(-\frac{\tau}{\sin b}\right) B(\tau) \frac{d\tau}{\sin b}. \tag{6}$$

This formula corresponds to the case where the stars are immersed in an absorbing layer. Assume a fraction of the stars lies far enough from the equatorial plane to fall out of the absorbing layer. Then instead of (6) we will have

$$I(b) = \int_0^{\tau_0} \exp\left(-\frac{\tau}{\sin b}\right) B(\tau) \frac{d\tau}{\sin b} + C \exp\left(-\frac{\tau_0}{\sin b}\right) \frac{1}{\sin b}. \tag{7}$$

The second term in this expression corresponds to the radiation of stars outside the absorbing layer.

Let us dwell on the case of formula (6), i.e., suppose that all stars are immersed in the absorbing layer. This will be true in particular if τ/α tends to zero as τ increases.

Let us accept that

$$B(\tau) = B_0 = \text{const} \quad \text{for} \quad \tau < \tau_1 < \tau_0 \tag{8a}$$

and

$$B(\tau) = 0 \quad \text{for} \quad \tau_1 < \tau < \tau_0, \tag{8b}$$

where τ_1 is a constant. This means that we assume that the ratio emission/absorption is constant within some distance from the galactic plane and that at larger distances the emission coefficient is zero.

Then from (6) we have

$$I(b) = B_0 \left(1 - \exp\left(-\tau_1 \operatorname{cosec} b\right)\right). \tag{9}$$

It is clear that B_0 is the brightness of the galactic equator and can be determined from observations.

The question arises: Is it possible to find a value of τ_1 yielding agreement between the surface brightnesses calculated from (9) and values calculated on the basis of Van Rhijn data on the mean values of A_n?

The second column of Table 1 gives some values of $I(b)$, calculated according to (9), assuming that $B_0 = 220$ stars of the tenth magnitude (photographic) from a square degree and $\tau_1 = 0.12$. In the third column of the same table the values of I obtained via real counts are given.

Table 1.

b	I_c	I_{obs}
0°	220	220
10°	108	104
20°	65	66
30°	46	47
40°	37	37
90°	26	25

Comparing the two columns, we see that our model is in approximate agreement with the average data taken from observations. Therefore, the stars outside the absorbing layer do not influence the distribution of brightness in question. Thus the second term in (7) can be neglected. The optical thickness in (9) turns out to be $\tau_1 = 0.12$. This is half the optical thickness of the layer in which the stars are immersed.

On the other hand, since the minimal value of τ_0 is 0.23 or 0.25 (if expressed in stellar magnitudes), no doubt remains that a considerable part of the absorbing layer belongs to the region where stars are rare. In other words, η/α tends to zero as the distance from the equatorial plane decreases.

Other astronomers also came to the same conclusion, for example, Aller and Trumpler [13]. Oort [14] noticed that selectively absorbing matter tends to concentrate near the equatorial plane of the Galaxy, while matter causing neutral absorption is more dispersed.

§**6.** ***More exact interpretation.*** After this confirmation of applicability of the model of plane-parallel layers, we concentrate on the interpretation of numerical values of B_0 and τ_1.

Let us note that for smaller values of τ_1, formula (9) for the high galactic latitudes reduces to

$$I(b) = B_0 \tau_1 \operatorname{cosec} b, \tag{10}$$

yielding

$$I\left(\frac{\pi}{2}\right) = B_0 \tau_1; \quad I(0) = B_0. \tag{11}$$

On the other hand, for $b = \frac{\pi}{2}$ and small τ_1, formula (6) yields

$$I\left(\frac{\pi}{2}\right) = \int_0^{\tau_1} B(\tau)\, d\tau, \tag{12}$$

and for $b = 0$

$$I(0) = B(0). \tag{13}$$

From (12) and (13) we get

$$\frac{I\left(\frac{\pi}{2}\right)}{I(0)} = \frac{\int_0^{\tau_1} B(\tau)\, d\tau}{B(0)},$$

and from (10) and (11)

$$\tau_1 = \frac{I\left(\frac{\pi}{2}\right)}{I(0)}. \tag{14}$$

Comparing with the preceding formulas, we obtain

$$\tau_1 = \frac{\int_0^{\tau_1} B(\tau)\, d\tau}{B(0)}, \tag{15}$$

and taking into account that

$$\int_0^{\tau_1} B(\tau)\, d\tau = \int_0^{\infty} \eta\, dz \quad \text{and} \quad B(0) = \frac{\eta(0)}{\alpha(0)}, \tag{16}$$

we find

$$\tau_1 = \alpha(0) \frac{\int_0^{\infty} \eta\, dz}{\eta(0)} = \alpha(0)\, z_1. \tag{17}$$

Thus τ_1 represents the optical thickness of a layer, whose coefficient of absorption equals that of the galactic plane and which has linear thickness

$$z_1 = \frac{\int_0^{\infty} \eta\, dz}{\eta(0)}. \tag{18}$$

In its turn this quantity equals half the perpendicular linear thickness of a homogeneous emitting Galaxy in the region around the Sun.

From (17) one can determine z_1, if the value of $\alpha(0)$ is known.

§7. ***Determination of*** $\alpha(0)$. From (13) and (16) we have immediately

$$I(0) = \frac{\eta(0)}{\alpha(0)}. \tag{19}$$

From (19) we can find $\alpha(0)$ since $I(0)$ is known from observations, and $\eta(0)$ can be found from (3) because the luminosity function in the vicinity of the Sun is also known.

We have already stated that $I(0)$ is equal to 220 stars of the tenth apparent magnitude per square degree. Multiplying by 33, we express the same quantity in the stars of the tenth absolute magnitude per square parsec. We obtain $7.3 \cdot 10^3$ stars of the tenth absolute photographic magnitude from a square parsec. Using the luminosity function of Van Rhijn, we find that $\eta(0)$ is equal to 0.048 stars of tenth absolute photographic magnitude from one cubic parsec. This amounts to $\alpha = 0.66$ per kiloparsec.

But this is the value of the absorption coefficient determined in the usual way. The same coefficient expressed in stellar magnitudes is 0.72 per kiloparsec.

Unfortunately, we have no idea of the accuracy of this value. This is connected in the first place with errors in determination of $\eta(0)$. The assumption of plane-parallel layers is another source of error. It is possible that the value of $\alpha(0)$ also requires some correction, particularly if the Sun lies in a relatively less populated region of the galactic equatorial plane.

Because for the calculation of $I(0)$ we used the numbers of stars averaged over longitudes and since $I(0) = \frac{\eta(0)}{\alpha(0)}$, our value of $\alpha(0)$ is a harmonic mean. There can be considerable fluctuations in the real value of $\alpha(0)$ depending on the direction (longitude). Because of great fluctuations, the harmonic mean will essentially be smaller than the arithmetic mean. Thus the arithmetic average must be larger than $0^m.72$ per kiloparsec. In a preliminary estimate, the difference can reach $0^m.10$. Therefore, we evaluate the mean value of $\alpha(0)$ to be

$$\alpha(0) = 0.82 \frac{\text{stellar magnitude}}{\text{kiloparsec}}.$$

All these estimates should be considered preliminary.

§8. ***Determination of*** z_1**.** From (17) we have

$$z_1 = \frac{\tau_1}{\alpha(0)}.$$

Using the values of τ_1 and $\alpha(0)$ obtained above, we get

$$z_1 = 160 \text{ parsec}.$$

This means that the total thickness of the luminous Galaxy in the vicinity of the Sun must be 320 parsec.

§9. ***Determination of*** τ_1 ***for different types of stars.*** Using partial surface brightnesses for different classes of stars, we can find the corresponding values of τ_1 as we did above for the ensemble of all stars.

However owing to a scarcity of data on the spectral type or other parameters of faint stars, we are able to obtain more or less real data on the partial surface brightnesses only for such classes for which the surface brightness is determined mainly by stars of higher apparent luminosity.

To such classes belong, for example, B stars and Cepheids. We determined τ_1 for them in the following way. For higher galactic latitudes, we have from (9)

$$I(b) = B_0 \tau_1 \operatorname{cosec} b$$

and for the galactic equator

$$I(0) = B_0.$$

Therefore,

$$\tau_1 = \frac{I(b)}{I(0)} \sin b \tag{20}$$

and in particular

$$\tau_1 = \frac{I(\frac{\pi}{2})}{I(0)}. \tag{20'}$$

Unfortunately, for the classes of stars under consideration the number of stars in the high galactic latitudes is small. Therefore, it is necessary to take a sufficiently large circumpolar zone (for example, from ±30° to ±90°). In such a zone the latitude b changes considerably. Therefore, we have to integrate over some solid angle ω. We write

$$\tau_1 = \frac{1}{\omega I(0)} \int I(b) \sin b \, d\omega. \tag{21}$$

Since the brightness under consideration consists of the contributions of separate stars, the integral reduces to a sum

$$\tau_1 = \frac{1}{\omega I(0)} \sum i \sin b, \tag{22}$$

where i is the brightness of an individual star, the summation is over all stars of the given class in the region ω.

Comparing (20′) and (22), we find

$$I\left(\frac{\pi}{2}\right) = \frac{1}{\omega} \sum i \sin b. \tag{23}$$

From (23), $I(\frac{\pi}{2})$ can be found even in the case where there are (almost) no stars of the class in question in the small circumpolar region.

As regards $I(0)$, this quantity can be determined by summation of brightnesses of stars in some narrow equatorial region, for example, between $\pm 5^\circ$ of galactic latitude.

Calculations based on some published lists of photographic brightnesses gave the following results.

a) The average partial visual surface brightness for stars O-B_2 near $b = 0^\circ$ is 2.7 stars of the tenth magnitude per square degree.

b) The average partial surface brightness of the totality of stars O-B_2 for $b = 90^\circ$ is 0.23 stars of the tenth magnitude per square degree.

These two conclusions are based on the data of HDC, which we consider complete, since the B stars fainter than $8^m.0$ make no essential contribution to the surface brightness.

From (22) and from numerical data cited above, for the stars O-B_2 we have

$$\tau_1 = 0.09.$$

However, it is necessary to indicate a possible source of error in such estimates. Since the number of stars in question in high galactic latitudes is small, large relative fluctuations in the estimate of $I(\frac{\pi}{2})$ according to (23) are possible. In other words, the occasional presence or absence near the Sun of a single star of the class in question can lead to quite different values of surface brightness.

Thus in our case the star α Virginis adds to (23) almost as much as all 26 remaining stars of O-B_2 type, for which $|b| > 20^\circ$. Therefore, we decided to adopt for O-B_2 stars the value $\tau_1 = 0.07$.

c) The average partial photographic surface brightness due to Cepheids (classical) at $b = 0^\circ$ is 0.14 stars of the tenth magnitude per square degree.

d) The average partial photographic surface brightness due to classical Cepheids at $b = 90^\circ$ is 0.006 stars of the tenth magnitude per square degree.

From the last two estimates we find $\tau_1 = 0.04$.

Unfortunately, the star δ Cephei is responsible for a considerable part of the brightness for the equatorial zone. If we disregard this star, we obtain for this zone $\tau_1 = 0.07$ and the mean value becomes $\tau_1 = 0.06$.

Thus we can adopt

for O-B$_2$ $\tau_1 = 0.07$
for Cepheids $\tau_1 = 0.06$.

This result suggests a strong concentration of stars of both types near the galactic plane within the absorbing layer.

§**10.** ***The planetary nebulae.*** The study of the distribution of planetaries in the sky brings rather unexpected results. We have calculated the sum of the photographic brightnesses for different galactic belts. From these we deduced the average surface brightnesses now in the second column of Table 2.

Table 2.

b	I_{obs}	I_{c}
0°	44	40
7°	22	25
12°	46	30
20°	15	29
32°	15	25
60°	21	18

Compiling this table we used data from the list of planetary nebulae from the book by Vorontsov-Veljaminov *Novae and Planetary Nebulae*, (Academy of Sciences, USSR, Moscow, 1948 [in Russian]). It is clear that the presence of a secondary minimum in $I(b)$ at $b = 7°$ contradicts formula (9). Therefore, in formula (7) we assume that $B(\tau) = B(0)$ in the interval $0 \leq \tau \leq 0.04$, otherwise $B(\tau) = 0$. On the other hand, we have taken $C \neq 0$ due to a considerable number of planetaries outside the absorbing layer.

This corresponds to the idea that there are two groups of planetaries: one inside the $2r_1$ layer, while the other is widely dispersed with the majority falling outside the absorbing layer.

For this case

$$I = B_0 \left(1 - \exp - \left(\frac{\tau_1}{\sin b}\right)\right) + C \exp\left(-\frac{\tau_1}{\sin b}\right). \qquad (24)$$

Table 2 contains values of I calculated for optimal values of constants:

$B_0 = 40$ stars of 20-th magnitude per square degree.

$C = 20$ stars of 20-th magnitude per square degree.

$\tau_1 = 0.04$; $\tau_0 = 0.3$.

We see that in spite of the optimal choice of 3 constants the results are not as good as in the case of all stars (Table 1).

Still, some of the discrepancy can be due to the small number of objects counted.

§**11.** ***Color of the Galaxy.*** Imagine an observer situated outside our Galaxy far in the direction of its axis of rotation. What surface brightness would he observe in the region of the Sun? The answer is well known to be the doubled surface brightness which we observe in the direction of the galactic pole, the interstellar absorption neglected. Since according to the existing data (Seyfert, [15]) the color of spiral arms is fundamentally different from the color of other parts of the same stellar system, this indicator will be essential for deciding in which part of the Galaxy our Sun is situated.

The color of the full stellar radiation from the poles of the Galaxy can be deduced from statistics of colors of stars in those regions. However, the data in this respect which exist (Malmquist, [16]) are incomplete and inexact.

Still, if the data available allow us to estimate merely the lower bound of the color index, this will be of some interest. In particular, if we can show that for the poles this index exceeds 0.50 then the problem will be solved since in the spiral arms $c < 0.50$. We will see that such an estimate is possible.

Let us write i_k for the partial visual surface brightness of the galactic pole, produced by stars of some k-th spectral type. The photographic surface brightness coming from these stars will be $i_k \cdot 10^{-0.4c_k}$, where c_k is the color index of stars of this type. Evidently the color index of the total radiation will be

$$c = -2.5 \log \frac{\sum i_k \cdot 10^{-0.4c_k}}{\sum i_k}. \tag{25}$$

To determine a lower bound we consider two possibilities: either 1) $c \geq 0.^m6$ or 2) $c < 0.^m6$. If the first assumption is true, we already have a lower bound for the color index equal to 0.6. In the case of the second assumption the observational data yield another estimate of the lower bound. Assume we have some quantities i'_k, which are equal to i_k for spectral types B, A and F and are smaller than i_k for the types G, K, M. Evidently

$$\frac{\sum i'_k \cdot 10^{-0.4c_k}}{\sum i'_k} > \frac{\sum i_k \cdot 10^{-0.4c_k}}{\sum i_k},$$

since the weighted mean of the quantities $10^{-0.4c_k}$ determined by the left-hand side of this equation is obtained by replacing the weights i_k for later spectral types by smaller weights. Therefore,

$$c_1 = -2.5 \log \frac{\sum i'_k \cdot 10^{-0.4c_k}}{\sum i'_k} < -2.5 \log \frac{\sum i_k \cdot 10^{-0.4c_k}}{\sum i_k} = c \tag{26}$$

will be the desired lower bound for c.

Using the data of HDC, we find with sufficient accuracy the values of i_k for types B, A, and F since for determination of total brightnesses of stars of these types in polar directions, this catalogue is sufficiently good. One can make some small corrections using the data of BSD.

For this purpose we have used the statistics of HDC stars compiled by Charlier [17] and counts of stars in BSD given in the introduction to the second volume of this catalogue.

As regards G, K, M spectra, HDC and even BSD can hardly be considered complete. Therefore, instead of i_k we will find some quantities i'_k which are smaller than i_k.

In this way we have obtained the following values of i'_k for different spectral types:

S_p	B	A	F	G	K	M
i'_k	1	4	5	12	7	0

in stars of the tenth magnitude per square degree.

From this using (26) we obtained $c_1 = 0^m.55$. Thus, in the case of the first assumption the lower bound of c is 0.6 and in the second case it is 0.55, i.e., we conclude that

$$c > 0^m.55.$$

This suggests (albeit indirectly) that the Sun is situated somewhere between the spiral arms of the Galaxy. Calculating the value of c, we have used the normal colors of the stars of each spectral type since the selective absorption in the direction of poles requires rather small corrections.

§**12.** ***Comparison of partial surface brightnesses for different galaxies.*** It is interesting to compare the partial surface brightnesses caused by stars of a certain physical type in outer galaxies with the same parameter in our Galaxy, when they are observed from outside under the same angle to the axis of rotation. A comparison for one particular case has been published by the author in [18].

§**13.** ***Conclusions.*** The study of surface brightnesses in our Galaxy leads to a number of interesting results. The main conclusions are: a) the stronger concentration of stars and lower concentration of absorbing matter near the Galactic plane and b) the yellow color of the region of our Galaxy in which the Sun is situated, if observed from outside the Galaxy.

All conclusions in this paper are of a preliminary nature. Our aim was to attract attention to a new direction of study, able to produce convincing results by rather simple methods.

REFERENCES

1. Fesenkov, *Astron. Journal*, 22, 271, 1945.
2. Van Rhijn, *Groningen Publ.*, no. 43, 1929.
3. Lindblad, *Handbuch der Astrophysik*, Bd. 5, 1933.
4. Kukarkin, Manuscript (unpublished).
5. Oort, *Ap. J.*, 91, 273, 1940.
6. Henyey and Greenstein, *Ap. J.*, 93, 70, 1941.
7. Ambartsumian, *Dokl. Akad. Nauk Arm. SSR*, 44, 244, 1944.
8. Ambartsumian, *Dokl. Akad. Nauk Arm. SSR*, 1, 9, 1944.
9. Hubble, *Mount Wilson Contr.*, no. 485, 1934.
10. Parenago, *Astron. Journal*, 17, 1, 1940; 22, 129, 1945.
11. Vashakidze, *Bull. Abast. Obs.*, 1, 87, 1937.
12. Oort, *Bull. Astr. Inst.*, Netherlands, 8, 245, 1938.
13. Aller and Trumpler, *PASP*, 50, 275, 1938.
14. Oort, *Annales d'Astrophysique*, 1, 71, 1938.

15. Seyfert, *Ap. J.*, 91, 528, 1940.
16. Malmquist, *Lund Med.*, no. 37, 1927.
17. Charlier, *Lund Med.*, no. 34, 1926.
18. Ambartsumian, *Dokl. Akad. Nauk Arm. SSR*, 3, 39, 1945.

Yerevan
Astronomical Observatory,
Academy of Sciences of Armenia

STELLAR ASSOCIATIONS

Until recently only two types of "small" stellar systems were studied within our Galaxy: open and globular clusters. The orders of magnitude of gravitional interaction energy for open clusters and for many visual double stars are the same. This speaks for their kinship. However, the author recently established [1] that along with open and globular clusters, we have within our Galaxy still another type of stellar system: the *stellar associations.* They are of outstanding interest from the viewpoint of stellar evolution. In the present paper we consider some examples of stellar associations and discuss their types and properties. An association is a system of stars which have a common origin and are of much lower spatial concentration than that of the general stellar field of the Galaxy where the associations are imbedded. The most important examples of associations are the groups of variable dwarf stars of T Tauri type and the groups of supergiants of O and B types.

Associations consisting of supergiants often have in their center an ordinary star cluster, the associations nucleus.

By all indications, the giant clusters in the Large Magellanic Nebulae are stellar associations.

Examples of stellar associations.

1) A group of variable stars of T Tauri type and the associated stars in Taurus and Auriga.

It is well known that the stars of T Tauri type are restricted to certain parts of the sky. In particular, eight such stars form an isolated group in the constellations of Taurus and Auriga within a solid angle of $12^\circ \times 12^\circ$.

The distance order of 100 parsecs means that the diameter of the group is of the order of 25 parsecs. Joy has discovered within the same domain a series of dwarf stars whose spectra possess bright lines. This indicates their probable interrelation with the stars of T Tauri type. At this point we stress two important facts:

a) the compactness of this group of stars cannot be explained as accidental. Obviously we deal with a single system.

b) the density of this system of stars is so low that it could never be identified by direct observation as a cluster, even if it were situated several times nearer to us.

The discovery of this association became possible solely because its stars belong to a definite class of variable stars. An important characteristic of this system is its low spatial density.

Even if the system of 7–8 stars of T Tauri type is complemented by 40 dwarfs with bright lines, we obtain a spatial density which is much lower than that of the Galactic stellar field where the association is imbedded. This circumstance remains true even after we attach to the system some dwarfs lacking bright lines (they are several times greater in number than dwarfs possessing bright lines).

According to well-known dynamic criteria this means that the system we consider is unstable and will be destroyed under the tidal influence of the general gravitational field of the Galaxy. We are compelled to think that the system consists of stars diverging in space.

Possible relations between the stars of the system and both luminous and dark interstellar matter also deserve attention.

2) The Kukarkin and Parenago [2] catalogue of variable stars in a small region centered at $\alpha = 18^h 40 : \delta = 9°0'$ indicates eight stars of T Tauri type (or by the catalogue terminology — of RW Auriga type). Three of them have a question mark, which refers to the type of variability. We present a list of these stars.

Name	α	δ	Stellar magnitude	Type
*V*637 Ophiuchus	$18^h 31^m 59^s$	$+9°51'.1$	$13^m.6 - 16^m.0$	T Taurus
*V*681 Ophiuchus	32 34	$+9°06'.1$	$14^m.2 - 15^m.4$	T Taurus ?
*V*643 Ophiuchus	32 34	$+6°16'.7$	$13^m.6 - 15^m.6$	T Taurus ?
*V*645 Ophiuchus	33 02	$+11°47'.1$	$14^m.6 - 15^m.6$	T Taurus ?
*V*476 Aquila	43 57	$+7°02'.3$	$13^m.3 - 14^m.2$	T Taurus
*V*480 Aquila	45 43	$+7°00'.7$	$14^m.0 - 16^m.0$	T Taurus
*V*489 Aquila	50 55	$+11°56'.3$	$13^m.1 - 14^m.8$	T Taurus
*V*490 Aquila	53 49	$+12°51'.1$	$14^m.1 - 15^m.5$	T Taurus

The stars are situated in a small, $6^\circ \times 7^\circ$ region, not far from the Galactic equator. Even if we neglect three stars, whose types need confirmation, the remaining five stars of T Tauri type still produce a concentration which cannot be accidental. We deal with members of a system of stars. Average maximal brightness of the variables for this system is $3 - 4^m$ less than that for the association in Taurus. Apparently, this is evidence of greater distance to the association in Aquila and Ophiuchus.

3) A group of stars of type O and B, as well as of red supergiants around the double open cluster χ and h of Perseus. This system was studied by Bidelman [3]. Observations leave no doubt about the existence of a group of supergiants of early and late types, which surrounds the χ and h Perseus clusters. The double cluster is the nucleus of this association.

The system diameter is 170 parsecs by order, while the diameter of both χ and h Perseus clusters is 10 parsecs by order (or 7 parsecs by Osterhoff). A characteristic feature of the system is the presence of a number of B-type stars with bright lines. In particular, there are at least five stars of P Cygni type (HD 12953, 13841, 14134, 14143 and 14818). Even if we accept that the association as a whole contains tens of thousands of stars, its mean density will still be less than that of the Galactic field. Doubtless, the stars of the association diverge in space. It should be noted, however, that the nuclei χ and h of Perseus, which are ordinary open clusters, are perhaps stable and their decay should necessarily follow the patterns common for open clusters.

4) The open cluster NGC 6231 is surrounded by a group of supergiants of O and B types. The study of radial velocities by Struve [4] shows that all these supergiants, together with the cluster, form a single stellar association. Its distance from us is about 1000 parsecs. The association's diameter exceeds the cluster's diameter by almost five times and is about 30 parsecs. Remarkably, the association includes two Wolf-Rayet stars and two stars of P Cygni type.

It is self-evident that the possibility of accidental concentration of these stars around the cluster is out of the question. In this case we again have to accept that the average density of the association is low in comparison with the density of the galactic field. The association is unstable, although the nucleus (NGC 6231 open cluster) probably is stable.

5) NGC 1910 system in the Large Magellanic Cloud is of peculiar interest. It is a large group of supergiants of early types where some stars are P Cygni-type, including the famous S Doradus. The diameter of this system is about 70 parsecs, i.e., many times larger than the sizes of common Galactic clusters.

6) The stellar association in Captein area SA 8 (around $\alpha = 1^h 00^m, \delta = +60° 10'$). The association is a group of weak stars of O and B type, occupying a region 2.5 degrees in diameter. The association includes a Wolf-Rayet star and two stars of B type with bright lines. Apparently at least 23 members of this association belong to the BO type. We note that the association is situated in a region poor in bright stars of B type (brighter than $8^m.0$). Judging from visible stellar magnitudes of stars of early types, this association is situated at a distance not less than 2000 parsecs. This suggests a diameter of about 100 parsecs. This extremely interesting distant association was discovered at the Byurakan Observatory in 1948 using the Bergedorf catalogue. The association's nucleus is the open cluster NGC 381, $7'$ or not less than 4 parsecs in diameter.

Basic characteristics of stellar associations. The above facts suggest the following general conclusions about stellar associations:

1) The associations are systems with small average densities compared with the density of the Galactic field. However, if we take partial concentrations of stars of separate spectral types, then associations are sharply distinguishable, owing to an abundance in them of stars of comparatively rare types. In some cases we deal with supergiants of O and B types, in others — with stars of T Tauri type. Because of their low density, the associations cannot be stationary in the sense of stellar dynamics. Unlike globular and open clusters, associations are nonsteady systems. Obviously members of associations diverge in space, and will eventually dissolve among the field stars.

2) Associations always contain stars which continuously emit matter. In three of the above six examples we find stars of P Cygni type. In examples 4) and 6) above, stars of Wolf-Rayet type are present. In the first two examples variable stars of T Tauri type are present, whose spectral bright lines possess absorption components on the violet side, i.e., they show the same peculiarity as the bright lines in the spectra of stars of P

Cygni type. A natural conclusion is that these stars also continuously emit matter.

3) In some cases the associations have nuclei looking like open stellar clusters.

Stellar associations in Large Magellanic Cloud. The Large Magellanic Cloud is very rich with open clusters. Remarkably, clusters of the Large Cloud in some cases have rather large sizes (several dozen parsecs) [5]. The most striking is the example of NGC 1910. The curve representing distribution of the open clusters in the Large Cloud according to their diameters has a minimum which divides all open clusters into two groups: a) clusters with diameters greater than twenty parsecs and b) clusters with diameters less than twenty parsecs. This alone makes us suspect that we are dealing with objects of two different types and scales. The presence of P Cygni stars in some clusters of the first group suggests that those are objects of the stellar associations type existing within our Galaxy, while the objects of another group are common open clusters.

The following consideration reduces this hypothesis to almost a certainty. Suppose we observe our Galaxy from some external system, e.g., from the Large Magellanic Cloud. Then the association around χ and h Perseus will contrast sharply with the surrounding background, due to the existence of a great number of supergiants in the association. Observing the same system from inside the Galaxy we face the fact that the stars of low luminosity, which are at far shorter distances than the association, are projected upon the latter. Both the members of the association and the projected stars will have a visible magnitude of the same order and, therefore, the former will be lost among the latter.

To an observer in the Large Magellanic Cloud the association would appear as a cluster of supergiants having a diameter of 170 parsecs. The χ and h Perseus clusters would appear to him as mere condensations in this magnificent system. On the other hand, the system NGC 1910 if transferred from the Large Cloud into the Galaxy and placed at the locus of χ and h Perseus would be observed as a typical association, i.e., it would not form a visible condensation of stars. Only a separate study of stars of early spectral types could betray its existence. Thus it seems that all giant systems in the Large Magellanic Cloud (about 15 in number) are in fact stellar associations, whose characteristics were described in the previous section.

Kinematics of stellar associations. The forces of interaction between the stars in an association are smaller than the tidal action of the general force field of the Galaxy. Therefore, at least for peripheral members of associations, the interaction forces can be neglected.

Considering the dynamics of stars of an association under the Galactic force field, it should be noted that the differential effect of Galactic rotation implies growth of distances between members of the association.

For a pair of stars distance r apart, the growth rate of r due to Galactic rotation is expressed via the well-known Oort coefficient A as follows:

$$\frac{dr}{dt} = Ar \sin 2(l - l_0).$$

Accordingly, for the radius R of the system at given galactic longitude l, we have

$$\frac{dR}{dt} = AR \sin 2(l - l_0).$$

For the radii R_1, R_2 at two epochs t_1, t_2 this implies

$$\ln R_2/R_1 = A(t_2 - t_1) \sin 2(l - l_0).$$

This formula says that for $l - l_0 = 45°$, the distance will be doubled after a period of $4 \cdot 10^7$ years.

Our derivation was based on the assumption that all stars in the association follow circular orbits around the Galactic center. Real orbits can of course be different. However, if we exclude too high relative speeds within an association, the radius duplication period will always have this order of magnitude.

Other possible causes of expansion can only support our conclusion that *each individual association came into existence rather recently and that it consists of stars which diverge from some primary volume where they have originated.*

However, the differential effect of Galactic rotation can produce expansion solely within the Galactic plane. If this were the only cause of expansion, then the associations would very soon acquire highly flattened shapes.

It should be stated that the rate of possible expansion of an association due to differences in the periods of oscillatory movements of its members along the Z axis should be much lower. The reason for this lies in the asymptotic independence of the oscillation periods and the amplitudes, for smaller values of amplitudes. Recall that for a star at the height z above the Galactic plane

$$\frac{d^2 z}{dt^2} = -2\pi G \int_{-z}^{z} \rho(z)\, dz,$$

where $\rho(u)$ is the Galactic density at height u.

For smaller values of z this reduces to

$$\frac{d^2 z}{dt^2} = -2\pi G \rho(0)\, z,$$

i.e., we have harmonic oscillations with period and amplitude mutually independent.

Since the associations which we observe are situated at low Galactic latitudes, their stars necessarily have almost equal oscillation periods in the z coordinate.

Therefore, the effect we consider is much less than the effect of differential rotation. Meanwhile observations do not show any considerable flatness in systems of the above examples. This compels us to think that there must be another cause of expansion, which prevails over Galactic rotation. We can suppose that the stars of the association left the initial volume where they formed, with certain velocities in different directions.

These initial velocities had to be no less than 1 km/sec. Otherwise, the effect of differential rotation should be evident for an association, whose sizes are several dozen parsecs. However, they should be less then 10 km/sec, for greater initial velocities would have been reflected in the distribution of radial velocities in the present epoch, for instance, in the association around NGC 6231.

If the distance from the center of an association increases at a rate of 5 km/sec, then the differential effect of Galactic rotation will fail to dominate until the sizes of the association reach several hundred parsecs. But such sizes would mean complete dissolution of the association among the field stars, i.e., the end of the association.

Consequently, flattening of the associations would not be observed. Therefore, the expansion velocities of 5 km/sec by order are most likely. We come to the conclusion that T Tauri-type stars in the association in Taurus-Auriga were expelled from a primary volume several million years ago, and the stars of the association around χ Perseus 10–20 million years ago, etc.

Expansion of an association begins without delay after the birth of its star members, since the assumption that the system spent considerable time in a stationary state before the expansion began contradicts stellar dynamics. This implies that the age of stars in the associations is measured by millions or at most by tens of millions of years.

This estimate is in good agreement with the fact that in associations we find stars of P Cygni, Wolf-Rayet or T Tauri type. A star cannot remain in P Cygni state more than one or two million years, as determined by the high emission rate of matter. On the other hand, P Cygni stars possess not only maximal luminosities among known stars, but probably maximal masses as well. If other states, which correspond to greater or equal masses, exist, then their lifespan should be rather short, for such masses are extremely rare. But P Cygni-type stars could not develop from stars possessing lesser masses. Consequently, they should be ranked among the youngest stars.

Number of stellar associations in the Galaxy. At present it is difficult to give a definite answer to the question concerning the number of stellar associations in the Galaxy. Associations containing supergiants of early type can be discovered at rather great distances (about 2000–3000 parsecs). Therefore, a considerable fraction of them can be observed. Such observable associations are probably several dozen in number. This means that the total number of such associations in the Galaxy is of the order of one hundred. As for associations consisting of T Tauri-type stars and other dwarfs with bright spectral lines, we know, at present, of only two.

However, both lie at rather short distances. In a ball of 100 parsecs radius there is one such association. This suggests that in the Galaxy they number in the thousands.

Assume that this number is 10,000. Also take into account that these associations remain detectable for a period of several million years. Then we have to conclude that in order to keep their present level, at least one association consisting of T Tauri-type stars should be generated in 1000 years on the average.

The question of formation of stars. Some astronomers have been putting forward a hypothesis that all stars in the Galaxy were born simultaneously or almost simultaneously several billion years ago, i.e., together with the formation of our Galaxy. The above facts cause this hypothesis to collapse. The birth of stellar associations and formation of stars within the latter from some other form of matter go on continuously almost before our eyes. The number of associations consisting of T Tauri-type stars which emerge during the lifetime of an association is of the order of ten thousand. For the time being, we do not know the mean number of stars born within an association, for we can identify only the brightest members. However, it seems reasonable that this number equals at least several hundred.

This means that at least a billion stars in our Galaxy were formed as a result of development of stellar associations from some other objects which remain unknown to us.

Other possible types of associations. It is highly probable that the system of B and O-type stars in Orion together with Trapezium make up a single giant association whose diameter exceeds 100 parsecs. The stars of Trapezium and the connected open star cluster apparently form the nucleus of this association. The presence of a giant diffuse nebula makes this system especially interesting and deserving of thorough study. The moving cluster of Ursus Major is a system of 32 members more than 200 parsecs in diameter.

The nucleus of this system is a subgroup of 11 stars nine parsecs in diameter. However, the system lacks direct indications of young age of the entering stars. This small system is probably a remnant of a formerly rich association. The Sun is situated in the interior of this system, but fails to be a member.

Conclusions. This article establishes the existence in the Galaxy of a great number of stellar associations, i.e., stellar systems of low density, unstable and dispersing in the Galactic space. The great role of stellar associations in the development of stars is evident. Therefore, they deserve the most thorough study.

REFERENCES

1. Ambartsumian, V. A., *Evolution of Stars and Astrophysics* [in Russian], pp. 12–16, 26–27, Yerevan, 1947.
2. Kukarkin, B. V. and P. P. Parenago, *General Catalogue of Variable Stars* [in Russian], Moscow, 1948.
3. Bidelman, *Ap. J.*, 99, 61, 1943.
4. Struve, *Ap. J.*, vol. 100, 189, 1944.
5. Shapley, H., *Galaxies*, p. 83, 1945.

Astronomical Observatory of the Academy
of Sciences of Armenia

MULTIPLE SYSTEMS OF TRAPEZIUM TYPE[1]

Introductory remarks. After the instability of O-associations was established in 1948, it was decided to pay attention to the pecularities of stars in these associations. In the beginning of 1949, B. Markarian and the author noticed that in the central part of the Cygnus association within a small cluster IC 4996 exists a multiple star ADS 13626 which differs from the majority of multiple stars. This recalled the presence within the Orion association of the Trapezium of Orion. An interesting photograph of that system can be found in the *Byurakan Atlas of Open Clusters.* In the paper by Markarian and the author on the association around P Cygni [1] the concept of Trapezium-type multiples was introduced. In 1950 and 1951 important papers [2,3] of Markarian were published in which the major role of Trapezium-type systems within the O-clusters was disclosed. They were followed by two papers [4,5] by the author on Trapezium-type systems. In 1952 P. P. Parenago [6] published his report based on the treatment of all observations of the Trapezium of Orion, in which conclusions were rather in favor of the positive energy of this system. In 1954 Sharpless [7] confirmed close interconnections between systems of Trapezium type and O-clusters. He also examined connections with diffuse nebulae and found a number of new systems of Trapezium type.

At this stage further measurements of relative positions of components of these systems as well as determination of their spectral types are important.

The terms "Trapezium" or "systems of Trapezium type" are now generally accepted, although among such systems we find triples, five-member groups and groups with even more components. However, the definition we give below seems quite clear in spite of some uncertainty of the boundary

[1]Editor's note: The original version of this paper consisted of 10 sections of which §§ 1, 5, 6, 8 and 9 have been selected by V. A. Ambartsumian for inclusion in the present collection. The numeration of the sections has been changed accordingly.

between the Trapezia and multiples of ordinary type. In this paper we will review the present state of study of such systems.

§1. General considerations

The majority of multiple systems known to us have the following property: in them it is impossible to find three components a, b, c such that the distances ab, ac and bc have the same order of magnitude. Multiple systems having such property we will call "systems of ordinary type" or ordinary systems.

A good example of an ordinary system is ε Lyrae. It consist of two pairs. The distance between these pairs is about $208''$, while the distance between two components of ε^1 Lyrae is $3''.1$, and between the components of ε^2 Lyrae is $2''.3$. Clearly, in any triple from this system, one of the distances will be 10^2 times less than the two remaining distances.

If in a multiple star it is possible to find three components for which all three distances are of the same order of magnitude, then we call it a system of Trapezium type. Remarkably enough, in θ' Orionis all six mutual distances are of the same order of magnitude.

Evidently our class of multiples includes systems which do not even approximately resemble trapezia considered in geometry.

Of course, in compiling a list of Trapezium-type multiple stars it is neccessary to define exactly which distances are considered to be of the "same order." Let us assume that two distances are of the same order if their ratio is between $\frac{1}{3}$ and 3.

It is well known that among triple stars a great preponderance of the ordinary-type triads is observed. However, the picture contains many interesting details.

Among the stars which are nearer to us than 10.5 parsec there are seven triples. In all seven cases the principal component belongs to the main sequence. Table 1 contains the values of logarithms of great semiaxes of orbits of the furthest and nearest satellites (expressed in AU) as well as the ratio κ of great axes. In the last column the spectral types of components (when they are known) are given.

From this table it is seen that 1) the smallest value of κ is 16, i.e., among the nearest stars there is no system of Trapezium type, and 2) the mean value of $\log \kappa$ is 2.27, i.e., the geometric mean of all values of κ is about 200.

Table 1.

Star	$\log a_1$	$\log a_2$	κ	Spectra
40 Eri	1.53	2.72	16	K1, wA, M6
α Cen	1.37	4.12	563	G0, K5, M5e
$-8°$ 4352	0.11	2.80	490	M2, M5
36 Oph	1.50	3.72	166	K2, K1, K6
HR 6426	1.10	2.44	22	K3, K4, M2
μ Her	1.07	2.59	33	G7, M4
{ $-32°$ 16135 { $-31°$ 17815	1.31	4.52	1600	M5, M5, M1

Such large values of the ratio κ allow us to reduce, in the first approximation, the motions in the triple system to the simple Keplerian motions by elliptical orbits. This conclusion for multiples situated in the solar neighborhood is confirmed by an example of the quadruple ξUMa which is also inside the same sphere of 10.5 parsec radius. It is a visual pair with the semiaxis of 18 AU, of which both components are spectral doubles with great axes 1.5 and 0.04 AU, respectively. In this case, too, the system is very far from being a Trapezium. This means that the motions again are reducible approximately to simple elliptical motions.

The picture changes only slightly when, instead of the nearest stars, we consider the stars which have high apparent brightness in the Aitken catalogue. Of all stars brighter than $4^m.0$ apparent magnitude and to the north of $\delta = -30°$, only 15 have two or more physical visual components. They are distributed according to their spectra as follows:

O – B2	B3 – B9	A	F – G	K – M
4	3	3	3	2

Of these 15 systems only 2 (ζ Persei and σ Orionis) are systems of Trapezium type. Thus even in this case ordinary systems prevail. It should be taken into account that the above 15 systems include only multiples in which at least three components can be visually resolved. However, to the north of $\delta = -30^\circ$ there are still more than 16 visual doubles brighter than $4^m.0$ in which at least one component is a spectral double. All these are multiples of ordinary type. Together we have 31 multiples brighter than $4^m.0$. They are distributed by spectral classes as follows:

O – B2	B3 – B9	A	F – G	K – M
8	3	5	8	4

Roughly speaking, this distribution by spectral intervals is uniform. But it is significant that both Trapezium-type systems just mentioned belong to the same interval O – B2. Are this fact and the fact of the absence of stars of O – B2 type within the distance 10.5 parsec around the Sun interrelated?

Let us take the stars brighter than $5^m.5$ to the north of -30°. Now we have 18 stars which are main components of Trapezium-type systems. They are distributed by spectral types in the following way:

O – B2	B3 – B9	A	F – G	K – M
5	6	1	3	3

We again have a strong prevalence of O – B stars. This prevalence will become stronger if we exclude cases where the satellite is very faint or very far from the brightest star, since in such cases the probability of an optical satellite is rather high. To possibly avoid such cases, let us introduce some limits for the distances from principal stars. For example, for satellites of different visible magnitudes we can use the following distance bounds:

m	d
11.5 – 12.5	10″
10.5 – 11.5	30
9.5 – 10.5	50
8.5 – 9.5	80

The cases in which the components are fainter than $12^m.5$ are all excluded. Then only 11 stars remain on our list (see Table 2). In the last column we present the apparent magnitudes of the components whose relative positions provide the basis for describing the system as a Trapezium. A question arises: what are the probabilities that the objects in this list are optical Trapezia? To answer this question let us assume that the optical component is the faintest in the group (which is the most probable case).

Let us take into account that the mean stellar magnitude of the faintest components of our systems is $10^m.1$ and that the number of stars brighter than 10.1 in the equatorial galactic zone is about 10 per square degree. Transformation of a double star into a triple of Trapezium type occurs if a background star projects within a circle around the double star whose radius is of the order of $50''$. An elementary calculation yields the probability of transformation of the optical binary into a Trapezium-type configuration: it is of the order of $\frac{1}{180}$.

However, a system is listed as a Trapezium only if the relative positions of its components satisfy certain additional conditions. Thus the optical (projected) component must not be too near the other two or too far from them. The probability that a randomly projected star will satisfy these conditions must be about $\frac{1}{2}$. Therefore, the probability that any given double star will be transformed by means of projection into a configuration of three stars satisfying Trapezium conditions with values of parameters as in Table 2 will be less than $\frac{1}{2} \cdot \frac{1}{180} = \frac{1}{360}$.

To the north of $\delta = -30°$ there are 444 double stars with the principal star brighter than $5^m.5$. Therefore, the expected number of Trapezium configurations formed owing to projection with the principal star brighter than $5^m.5$ will be 1.2. We can expect that the number of such optical systems in Table 2 is 1 or 2.

The distribution of the stars of Table 2 according to spectral classes is:

O – B2	B3 – B9	A	F – G	K – M
4	5	1	1	0

Thus of 11 systems of Trapezium type, 9 have O – B stars as the main component, i.e., the early spectral types are strongly prevalent. Possibly, the two systems of spectral interval A and F – G are optical triples. It is also possible that they are in fact ordinary systems which acquired the false appearance of a Trapezium via projection.

Table 2.

Star	ADS	HD	Spectra	m_1	m_2	m_3
ζ Per	2843	24398	B1 Ib	2.9	9.3	11.1
+14°796	3579	31764	B8	5.2	6.7	9.0
14 Aur	3824	33959	A2	5.2	7.2	11.0
θ^1 Ori	4186 – 8	37022	O7	5.4	6.8	6.8
σ Ori	4241	37468	O9.5V	4.0	7.5	10.3
30 CMa	5977	57061	O9 III	5.0	10.5	11.2
P Pup	6205	60863	B8	5.2	9.3	10.0
ζ Mon	6617	67954	G0	5.0	8.5	10.7
−21°4908	11169	166937	B8p	4.0	9.5	9.5
59 Cyg	14526	200120	B3ne	4.7	9.0	11.5
+34°4371	14831	202904	B3ne	4.6	10.2	10.2

Of the systems in Table 2, all four that belong to the O – B2 spectral interval are members of an O-association. In particular, ζ Persei is a member of Perseus II, the stars θ' and σ Orionis are members of the association in Orion, while the system 30 Canis Majoris belongs to the cluster NGC 2362 which is a nucleus of a group of hot giants.

Of stars of the later spectral classes, the star μ Sagittarius ADS 11169 (which according to more refined classification is of type cB8e) is apparently included in the association Sagittarius I. Other stars of B3 – B9 type apparently are not members of O-associations.

We conclude there exists a close interconnection between the Trapezium-type systems, the principal stars of which belong to the spectral interval O – B2, and the O-associations.

......

§2. Instability of Trapezium-type systems

The fact that the overwhelming majority of multiple systems are of ordinary (and not Trapezium) type has for a long time attracted the attention of astronomers. In such systems the motions are approximately Kepler type. Evidently such motions can be maintained for a very long time and these systems are stable. In this respect multiples of ordinary type are in sharp contrast with the open clusters where due to the exchange of kinetic energies between members of the cluster some members may acquire energies sufficient to escape. This leads to disintegration of clusters. In multiples of ordinary type such energy exchanges are rather rare and the system remains stable for a very long time.

The situation is quite different in the Trapezium-type systems. Here the motions are similar to those within clusters. Therefore, there are real possibilities for members to escape. But because of the small number of members the lifetime here must be shorter. The small dimensions of these systems also imply a shorter lifetime.

The formula for the disintegration time of a cluster, when applied to Trapezium-type systems, leads to time scales of the order of $2 \cdot 10^6$ years. In many cases this means that a Trapezium-type system can disintegrate during a period in which each star makes only a few crossings of the system.

This conclusion is physically clear. Even for a small number of crossings, a component has considerable opportunity to approach one of the remaining components at a distance where the energy of their interaction may exceed the value of the escape energy. The chances of expulsion of a star from a system are considerable.

Arguments of this type lead to the conclusion that the lifetimes of Trapezium-type systems are of the order of one or two million years. Thus their components are extremely young stars.

We do not exclude the theoretical possibility of some periodic or quasi-periodic motions in the Trapezium systems. However, such motions require very special initial conditions and, therefore, have an exceedingly low probability.

§3. The sign of the energy of Trapezium-type systems

We know that in the Galaxy alone there are billions of double and multiple stars of which the prevailing majority are stable systems. This means that the values of energy in the prevailing majority of cases should be negative. Although this conclusion is based on purely statistical considerations, it is confirmed by direct determination of orbits for many of them. Nevertheless, it is possible that some pairs or multiples (a minority) have positive energies. Some of the Trapezium-type systems can be considered possible carriers of positive energy. This would mean that they are recently formed groups and are now in the process of disintegration.

It is interesting that in stellar systems which are related to Trapezium-type multiples such as O-associations and O-clusters, the sign of the total energy is at least sometimes positive.

If among Trapezia there are groups with positive energy, it is easy to calculate the time they needed to reach their present sizes. In this way we discover that some Trapezium systems have an age of the order of 10^5 years and in any case less than 10^6 years.

According to Parenago, our conclusion regarding the positive sign of total energy of at least some Trapezia is confirmed by data on movements in the Trapezium of Orion itself (θ' of Orion). For the majority of Trapezia, the time during which they were observed is not sufficient to determine conclusively the sign of their total energy. Unfortunately, all such systems, being comparatively wide groups, have not been sufficiently observed.

......

§4. Some very wide systems

In some photographs of regions around O-associations we find wide groups which are the result of the expansion of Trapezium-type groups. Let us take, for example, the region in Cygnus around NGC 6871. On the photographs of that region there are at least five groups which are of this type, even after we exclude similar groups in NGC 6871 itself. The spectral types and stellar magnitudes of stars in this region were determined by Enier [8]. In all these systems the brightest star belongs to the B type. The catalogue numbers and coordinates of these stars according to Enier are given in Table 3. It happens that all are of B type. In the fourth column the numbers of components (multiplicity) are given; in the seventh, the greatest distance between the components expressed in seconds of arc. In the next column the same distance is expressed in AU; the distance of the system from us is taken to be 1500 parsec. These systems are similar geometrically (especially systems A 34°140 and A 35°190) to the Trapezium in Orion. However, in linear units, they are about ten times wider.

The fact that the brightest stars of these groups belong not to the classes O or BO, but to the later subtypes of B speaks in favor of their greater age as compared with the stars in our List of the Trapezium system. This is especially true as regards A 34°140 and A 35°190, which resemble the Trapezium of Orion itself but are wider by one order of magnitude.

Table 3.

Star	m_{pg}	S_p	n	α (1950)	δ (1950)	d_{max}	D_{max}	Note
A 34°140	10.00	B5	4	$20^h03^m.8$	31°10′	52	78.000	
A 35°190	9.56	B2	4	05.0	35 09	93	140.000	
A 35°240	8.22	B2	5	07.1	35 22	55	83.000	BD + 35°3987
A 35°254		B2	4	07.4	35 20	41	61.000	
A 35°283	8.67	B3	6	08.0	35 43	78	117.000	BD + 35°4004

An expansion which is a result of the interaction of stars (as in open clusters) is necessarily accompanied by a change in the shape of the group. Since here the shape remains similar to that of the Trapezium of Orion itself, this implies rather an expansion resulting from the positive total energy of the group.

§5. The narrow systems of Trapezium type

To understand the evolutionary role of the Trapezium systems, it is important to study the most narrow cases. Special attention must be paid to the visually single O or BO stars which sometimes consist of several components. This question was treated in Sharpless' paper [7]. It seems that such narrow systems can persist no more than a few hundred years. Therefore, the probability of finding such systems among visually single stars must be very low. Nevertheless, it is worthwhile to prove the absence of such single stars. At the same time it is necessary to try to find narrow visual multiples of Trapezium type. If the distances of such multiples are of the order of a thousand parsecs, this would mean that we try to find groups with a size less than 4″. ADS 719, ADS 6033, ADS 11344, ADS 364, ADS 4164 and ADS 14010 are such systems. The brightest star of the first of these systems is type O6. The brightest star of the second system is known as variable VY CMa and belongs to the class Ma. The greatest distance between its components is 2″.9. Since the variable giants of late type are in a way related to the early-type stars, it is possible that here we are dealing with a real Trapezium. Thus it is possible that we have here additional evidence of the close relationship between blue and red giants. Unfortunately, data on components of other systems of these groups are absent. We would like to stress the importance of spectral studies of components of these interesting multiples. In any case the scarcity of these systems is in complete accordance with the concept of their instability.

REFERENCES

1. Ambartsumian and Markarian, *Soobshch. Byurakan Obs.*, II, 1949.
2. Markarian, *Soobshch. Byurakan Obs.*, V, 1950.
3. Markarian, *Soobshch. Byurakan Obs.*, IX, 1951.
4. Ambartsumian, *Dokl. Akad. Nauk Arm. SSSR*, 13, 129, 1951.
5. Ambartsumian, *Dokl. Akad. Nauk Arm. SSSR*, 13, 97, 1951.
6. Parenago, *Transactions of the Second Conference on Cosmology Questions* [in Russian], p. 345, Moscow, 1953.
7. Sharpless, *Ap. J.*, 119, 334, 1954.
8. Enier, *Ap. J.*, 118, 77, 1953.

August 1954

Byurakan
Astronomical Observatory of the Academy
of Sciences of Armenia

ON THE PATCHY STRUCTURE OF THE INTERSTELLAR ABSORBING LAYER[1]

More than ten years ago it became clear that the interstellar absorbing medium has a very irregular, patchy structure. We in the Soviet Union have studied in more detail the structure of the system of absorbing clouds. Some numerical parameters describing this variety of clouds were obtained.

As a rule, each of the luminous diffuse nebulae is illuminated by some high-luminosity star. Are the stars and the corresponding nebulae connected genetically and dynamically or is this connection due to the occasional meeting of a star and a nebula during their galactic motion?

In order to answer these questions we decided to check by observations a conclusion that follows from the hypothesis of a chance meeting.

It is clear that each star can illuminate only a sphere of definite radius around itself. A cloud situated within this sphere will attain sufficient illumination to be observed as a bright diffuse nebula. Evidently, the radius of such a sphere will be proportional to the square root of the luminosity of the star.

Let us take a certain volume V of the galactic space. Stars of different spectral types and luminosities will appear in this volume. We consider the spheres illuminated by these stars. Because the luminosity function and the star density for each spectral class

$$\mathrm{O, B\,0, B\,1, B\,2-9, A, F, G, K, M}$$

are known, we can immediately obtain the total volume illuminated by stars of these spectral types. In the case of a chance connection between the nebulae and the illuminating stars, the probability for any cloud to be

[1]Originally published by the International Astronomical Union © 1950 in *Transactions of the International Astronomical Union*, vol. 7, 1950, pp. 452-455. Used here with permission of the International Astronomical Union.

illuminated by a star of some type, let us say A-type, will be equal to the total sum of the volume illuminated by all A-type stars within the volume, divided by the whole volume V. We are able to compute all these probabilities. If the hypothesis of the chance connection is true, then the numbers of nebulae illuminated by stars of different types will be proportional to the corresponding probabilities.

The comparison of computed probabilities with the observed numbers of diffuse nebulae illuminated by stars of different types showed very close proportionality. We may conclude, therefore, that the hypothesis of a chance meeting should be accepted.

These considerations lead also to another important consequence. It is easy to show that stars of all types together illuminate only about 1/2000 of the volume of interstellar space. This means that a nebula has a probability of about 1/2000 of being illuminated. It follows that the number of all clouds in the Galaxy is about 2000 times larger than the number of bright diffuse nebulae. Following this line of argument we have established that the number of cosmic clouds per ps^3 is about 1/10,000,

$$n \approx \frac{1}{10,000}.$$

If σ is the mean cross-section of a nebula, the mean number of clouds crossed by a ray along the path l will be equal to $ln\sigma$.

If, further, ε_0 is the mean optical thickness of a cloud expressed in stellar magnitudes, the total absorption caused by these clouds will be

$$\Delta m = ln\sigma\varepsilon_0 = al,$$

where a is the mean absorption per parsec.

We know from the general data on cosmic absorption the value of a (photographic or visual) and the order of magnitude of the cross-section σ. Therefore, if we assume that the whole interstellar absorption is caused by our system of clouds (or opaque nebulae) we may derive the value of ε_0.

The first and very rough determination of ε_0 showed that it is of the order of $0^m.2$ or $0^m.3$ in the photographic region. It was clear that this value is not in contradiction to our ideas about the mean transparency of diffuse nebulae.

It was concluded that the absorbing layer consists of a large number of discrete clouds, which are small compared with the distances between them. But it was desirable to have another, independent and more accurate method for the determination of ε_0.

The counts of extragalactic nebulae made by Prof. Shapley and Dr. Hubble have demonstrated that the numbers of nebulae brighter than a certain magnitude per square degree show considerable fluctuation. We have shown that even for a given galactic latitude these fluctuations far exceed chance fluctuations according to Poisson's Law. It seems at first glance that this may be attributed to the clustering tendency, which of course exists and the importance of which was emphasized by Shapley.

However, the clustering tendency alone is not able to account for the main part of the fluctuations. This is particularly clear from the following evidence: when we divide the whole sky into galactic latitude zones and determine the fluctuations in each of these zones separately, the relative magnitude of these fluctuations increases with a decrease in the galactic latitude of the zone.

It is evident, however, that chance fluctuations in the numbers of the galactic absorbing clouds in the path of light coming from different extragalactic nebulae will cause additional fluctuations in visible nebulae numbers.

It remains to investigate theoretically how these fluctuations depend on the galactic latitude b. With this aim in view, let us compute

$$\overline{(N_m - \overline{N_m})^2} = \overline{N_m^2} - \overline{N_m}^2,$$

where N_m is the number of nebulae brighter than some m, per square degree. This number N_m in the absence of absorption must be

$$N_m = N_0 \cdot 10^{0.6m}.$$

The transparency of a cloud is $q = 10^{-0.4\varepsilon_0}$. Therefore, n clouds in the line of sight diminish the brightness of nebulae q^n times. The observable number of nebulae should therefore be

$$N_m = N_0 \cdot 10^{0.6(m - n\varepsilon_0)} = N_0\, 10^{0.6m}\, q^{\frac{3}{2}n}.$$

The problem of the computation of $\overline{N_m}$ was reduced thus to the computation of $\overline{q^{\frac{3}{2}n}}$. At the same time, the computation of $\overline{N_m^2}$ was reduced to the computation of $\overline{q^3}$. Using Poisson's Law for the probability of n we have, after some algebra:

$$\overline{N_m} = N_0 \cdot 10^{0.6m} \exp\left(-n_b(1 - q^{\frac{3}{2}})\right).$$

Here n_b is the mean number of absorbing clouds crossed by the line of sight at latitude b.

In the case of plane-parallel layers of clouds we have

$$n_b = n_{\frac{\pi}{2}} \operatorname{cosec} b.$$

In the same way

$$\overline{N_m^2} = N_0^2 10^{1.2m} e^{-n_b(1-q^3)}$$

and

$$\frac{\overline{(N_m - \overline{N_m})^2}}{\overline{N_m}^2} = \exp\left(n_b(1 - q^{\frac{3}{2}})^2\right) - 1 = \exp\left(n_{\frac{\pi}{2}} \operatorname{cosec}\left(b(1 - q^{\frac{3}{2}})^2\right)\right) - 1. \tag{1}$$

On the other hand, we have

$$\tau_{\frac{\pi}{2}} = n_{\frac{\pi}{2}} \varepsilon_0, \tag{2}$$

where $\tau_{\frac{\pi}{2}}$ is half the optical thickness of the galactic absorption layer in the direction perpendicular to the galactic plane. According to the last determination of Parenago $\tau_{\frac{\pi}{2}} = 0^m.32$.

Equations (1) and (2) determine $n_{\frac{\pi}{2}}$ and ε_0. Using the counts of Shapley and Hubble we have computed the values of

$$\frac{\overline{(N_m - \overline{N_m})^2}}{\overline{N_m}^2}$$

for different latitudes and obtained a value of ε_0 of the order of $0^m.25$.

It is necessary to introduce a correction accounting for the dispersion of the limiting magnitude of different plates and for other observational conditions. This correction is somewhat indefinite. However, it is still certain that ε_0 is confined between

$$0^m.20 < \varepsilon_0 < 0^m.30.$$

It is clear from (2) that $n_{\frac{\pi}{2}}$ will be of the order of unity.

Prof. Kukarkin has determined the dispersion of photoelectric color excesses of extragalactic nebulae in different latitudes and obtained the mean color excess of a single cloud equal to $0^m.05$. Multiplying this value by the corresponding factor, he has found for the optical thickness of a cloud the approximate value

$$\varepsilon_0 = 0^m.27.$$

Dr. Markarian (Byurakan Observatory) has determined the value of ε_0 from comparison of the observed fluctuations in the numbers of stars counted in low latitudes with the theory of fluctuations of star numbers in these latitudes based on the model of a layer of clouds as described above. His theory contains too much algebra to be discussed here. He obtained $\varepsilon_0 \approx 0^m.25$.

We may conclude that ε_0 is really of this order, although we do not exclude that the value of ε_0 may vary in different regions of our Galaxy.

The theory of fluctuations in the total brightness of stars contained in a square degree in the galactic equator may take a very simple and elegant form. The distribution function of this quantity, which is nothing else than the intensity of the stellar component of Milky Way radiation, satisfies a certain functional equation.

In the derivation of this equation I have used an invariance principle similar to that in the theory of diffuse reflection from plane-parallel layers. It is based on the fact that the distribution function remains unchanged when the observer displaces himself by a distance Δr along the line of sight.

This principle gives the following functional equation:

$$\Phi'(J) + \Phi(J) = \frac{1}{q}\Phi\left(\frac{J}{q}\right) \tag{3}$$

for the distribution function, when J is measured in some convenient unit.

From equation (3) is found the mean-square deviation of intensity

$$\frac{\overline{(J-\bar{J})^2}}{\bar{J}^2} = \frac{1-q}{1+q}. \tag{4}$$

Unfortunately, we do not have at our disposal a sufficient number of determinations of J at different points of the Milky Way to check the validity of (3) and (4).

Some years ago, Academician Shain called attention to the too weak correlation of the color excesses between B and O stars as determined by Stebbins and his coworkers, and the brightnesses of the corresponding regions of the Milky Way.

The theory of the cloud structure of the absorbing layer explains this phenomenon. The stars of Stebbins's list have a mean distance above 1000 parsecs, while the fluctuations of the Milky Way brightness are caused chiefly by clouds at distances of 200–500 parsecs. Therefore, the sets of clouds responsible for these two phenomena are quite different and the correlation must be weak.

The problem under consideration is also connected with gas clouds. Under the assumption that interstellar gas also has such a patchy structure, Dr. Melnikov has analyzed the curve of growth for the interstellar lines and has obtained dispersion velocities of the gaseous clouds of the order of 8km/sec.

Dr. Lyman Spitzer has told me that the observational data of W. S. Adams on the splitting of the interstellar lines into several components indicate that the number of gas clouds crossed by the line of sight equals the number of dust clouds computed from our theory. He identifies therefore the two systems of clouds. It follows from this identification that the clouds of interstellar gas should be comparatively small in size (about 8–10 parsecs in diameter). It is very important to confirm this conclusion observationally.

In connection with the problem considered in this brief report, a large amount of observational work is being done at the Abastumany Observatory under Dr. Kharadze. The color indices of many thousands of stars in the selected areas are determined as well as the colors of many extragalactic nebulae.

However, the photoelectric data on the color indices of the external galaxies should be very important for the further study of the structure of the galactic absorbing layer.

Byurakan Observatory,
near Yerevan, Armenia

SUPERASSOCIATIONS IN DISTANT GALAXIES[1]

The Large Magellanic Cloud contains in addition to a considerable number of ordinary O-associations a certain number of larger objects which, however, are similar in nature to the associations. These objects were named "constellations" by Shapley. But the large complex 30 Doradus surpasses notably all these objects both in diameter and in absolute brightness. The latter is of the order of $-15^m.0$ while the diameter is of the order of 600 pc. If we take the average absolute brightness of associations in our Galaxy as equal to $-10^m.0$ then it turns out that 30 Doradus is 100 times more luminous than the ordinary associations. Photographic images of more distant galaxies reveal that sometimes complexes occur in them of the same order of luminosity and dimensions as 30 Doradus. Therefore, it seems to us useful to regard these complexes as a special class of objects and call them *superassociations*.

The frequency of occurrence of superassociations within the galaxies is being investigated at the Byurakan Observatory. On plates taken by means of the 21-inch Schmidt reflector the superassociations are almost star-like if the distance of the corresponding galaxy is over 15 million pc. When exposures are of shorter duration (a few seconds) the general background of the corresponding galaxy does not hinder photometric evaluation, and the images of superassociations can be compared with the focal images of the surrounding stars or with those located in standard areas. In determining the stellar magnitude in this way, the error for closer galaxies may, however, attain $0^m.5$.

---

[1]Originally published by the International Astronomical Union © 1964 in *The Galaxy and the Magellanic Clouds, IAU-URSI Symposium, No. 20, Canberra, March 18-28, 1963*, F.J. Kerr and A.W. Rodgers (eds.), Canberra, Austr. Acad. Sci., 1964, pp. 122-126. Used here with permission of the International Astronomical Union.

We quote here only the preliminary results of the review based on the study of 55 galaxies, mostly belonging to the Sc type. These galaxies (with the exception of two) have been selected from the Shapley-Ames catalogue at random, if we disregard the fact that Sc galaxies were preferably observed while E-type galaxies have not been observed at all.

In Table 1 the NGC numbers of 14 galaxies are listed, in which superassociations have been found. Concentrations exceeding $-13^m.5$ in absolute magnitude have been considered as superassociations. Naturally we could not have regarded as superassociations the long sections of the bright parts of the spiral arms.

Table 1. The Galaxies with Superassociations.

NGC	Type	N_{sa}	$\overline{M}_{sa}$	M_g	M_n
1087	Sc	8	$-14 \cdot 9$	$-20 \cdot 6$	$-15 \cdot 9$
1961	Sb	3	$-15 \cdot 8$	$-21 \cdot 5$	$-17 \cdot 1$
2276	Sc	4	$-15 \cdot 2$	$-20 \cdot 7$	$-15 \cdot 9$
3991	Haro	2	$-17 \cdot 2$	–	no
3995	Sc	4	$-14 \cdot 8$	$-20 \cdot 3$	$-17 \cdot 0$
4303	Sc	4	$-14 \cdot 6$	$-21 \cdot 6$	$-17 \cdot 6$
4496	SBc	1	$-14 \cdot 9$	$-19 \cdot 8$	no
4559	Sc0	2	$-13 \cdot 6$	$-19 \cdot 6$	$-14 \cdot 1$
5676	Sc	2	$-15 \cdot 5$	$-20 \cdot 8$	$-16 \cdot 0$
5678	Sc	4	$-17 \cdot 5$	$-20 \cdot 5$	$-15 \cdot 8$
6217	Sc	4	$-14 \cdot 5$	$-19 \cdot 8$	$-16 \cdot 6$
6412	Sc	1	$-15 \cdot 5$	$-19 \cdot 4$	$-15 \cdot 8$
6643	Sb	3	$-14 \cdot 8$	$-20 \cdot 0$	$-15 \cdot 4$
7448	Sc	3	$-15 \cdot 0$	$-20 \cdot 8$	$-16 \cdot 2$

The second column of the table reproduces the galactic types; the third column, the number of superassociations in each galaxy; the fourth column, the mean absolute magnitudes of the superassociations; the fifth, the absolute magnitudes of the galaxies (Sandage's scale of distances assumed); and the sixth, for comparison, the absolute magnitude of the nucleus of the galaxy as determined from the same plates. The data of this table make it clear that superassociations are found particularly often in supergiant galaxies with an absolute magnitude of $-20^m.5$ and above. Of the observed 35 galaxies of the Sc type, eight containing superassociation have an average absolute magnitude of $-20^m.6$, whereas the absolute magnitude of Sc galaxies without superassociations is equal to $-18^m.8$. As for the Sb galaxies, the number of observed objects is small. It is, therefore, hard to speak of any existing difference. However, there is no doubt that Sb galaxies containing superassociations are at the same time systems of high luminosity.

In Table 2 the numbers (N) of all observed galaxies of the Sc type are quoted for three different intervals of absolute magnitude, together with N_{sa}, the numbers of galaxies containing superassociations for the same intervals.

The last column contains the average numbers ν of superassociations per galaxy of the given class of luminosity. This Table shows more convincingly that superassociations are encountered almost exclusively in supergiant galaxies.

Table 2. Frequency of Superassociations in SC Galaxies.

Interval of M	N	N_{sa}	ν
$M > -20 \cdot 0$	21	3	$0 \cdot 3$
$-20 \cdot 5 < M < -20 \cdot 0$	6	1	$0 \cdot 7$
$M < -20 \cdot 5$	8	6	$3 \cdot 1$

Apparently the picture is somewhat different in irregular galaxies. We have not as yet treated the problem in detail, but the NGC 275 Haro-type galaxy, involving at least five superassociations and of an absolute magnitude $-19^m.0$, testifies to the fact that in irregular galaxies the situation is different. This is also attested by the example of the Large Magellanic Cloud.

As became evident from the work of Shapley and Paraskevopoulos [3], most of the luminosity of the 30 Doradus complex is contained in the nebula. But 30 Doradus also contains hundreds of blue supergiant stars with the richest cluster of supergiants located in the center of this complex. At the same time, the lifetime of the superassociations as a whole must considerably exceed the mean lifetime of ordinary associations. This follows from the fact that the lower limit of the duration of life of any such complex must be the value D/v, where D is the diameter of the complex while v indicates the mean relative velocity traced in it. Making $D = 600$ pc and $v = 10$ km/sec, we obtain for the lifetime a lower limit of 6×10^7 years. This is nearly one order of magnitude more than the age of ordinary O-associations and that of hot supergiants. It must, therefore, be assumed that the blue supergiants observed in the superassociations represent only one of the numerous generations of these objects. Many thousands of supergiants appear, presumably, during the lifetime of the superassociations, which thereupon turn into other stars. If we take into account the fact that the T Tauri-type stars also usually originate in similar complexes and in far greater numbers, then we must believe that hundreds of thousands of stars come into being there.

In 1939 in the book *Theoretical Astrophysics* [1] we made the initial estimation of the mass of the nebula 30 Doradus at $2 \times 10^6\ M_\odot$, while at the Observatory of Mount Stromlo, Johnson [2] reestimated, four years ago, the gaseous mass of this nebula. It turned out to be $5 \times 10^6\ M_\odot$. Consequently, hundreds of thousands of stars are likely to be genetically bound with this nebula.

In conclusion, I take the liberty of making one more remark of a cosmogonic nature. If one realizes that superassociations are formed out of gaseous masses distributed within the given galaxy, one must believe that in galaxies of a great total mass the tidal forces should hinder the formation of such large complexes. On the contrary, in galaxies of a small mass no such hindrance will exist. For this reason we should not expect to meet superassociations in supergiant galaxies. In fact, we note the reverse picture. This apparently contradicts the assumption as to the possibility of the formation of superassociations from initial matter diffused all over the galaxy.

REFERENCES

1. Ambartsumian, V. A. (1939), *Theoretical Astrophysics* [in Russian], Moscow, Leningrad.
2. Johnson H. M. (1959), *P. A. S. P.*, 71: 425 – 434.

3. Shapley H. and Paraskevopoulos, J. S. (1937), *Ap. J.*, 86: 340 – 342.

Byurakan
Astronomical Observatory of the Academy
of Sciences of Armenia

FLARE STARS[1]

Introduction

Historically the first flares of brightness in some red dwarf flare stars were observed as early as in the first quarter of our century [1]. However, those stars attracted attention much later, after the discovery and study by W. Luyten [2] and A. Joy and M. Humason [3] of the flares of a prototype of this class of variable stars, i.e., UV Ceti at the close of the 1940s.

A considerable number of flare stars has been discovered over a comparatively short period in the circumsolar volume and in stellar aggregates — associations and clusters. At present the number of discovered flare stars is well over 600. In the Pleiades cluster and the Orion association alone over 500 flare stars have been found [4-6] and about 40 in the vicinity of the Sun [7]. As a result of the growing interest in flare stars, international campaigns were sponsored in recent years to make regular photographic observations of flare stars in some stellar associations and clusters and photoelectric observations of particular flare stars around the Sun in order to study them at great length.

The campaign, the active participants of which are the observatories of Armagh, Asiago, Boyden, Byurakan, Cerro Tololo, the Crimea, Catania, Tokyo, Tonanzintla and others, has substantially amplified our notions of flare stars.

It turned out that the flare phenomenon is more akin to explosions than could be expected from the first observations. Photoelectric observations with high resolving power in time have shown [8] that sometimes it takes

[1] Jointly with L. V. Mirzoyan.

Originally published by the International Astronomical Union © 1971 in *Inter. Astron. Union, Colloq. No. 15, Bamberg, 31 August-3 September, 1971*, in *New Directions and New Frontiers in Variable Star Research*, Bamberg, 1971, pp. 98-108. Used here with permission of the International Astronomical Union.

the star only several seconds to go through a period of increasing brightness while a decrease in brightness may continue over a long period.

In the present paper an attempt is made to consider several sets of data on flare stars from the standpoint of those ideas on the evolution of stars that are advanced at Byurakan (see, for instance, [9]).

The existence of close relations between flare stars and stars of the T Tauri type is an established fact nowadays. More proof in support of such relations has been produced by one of the authors of the present paper [10, 11] as well as by G. Haro [12, 13].

It was proved at Byurakan as early as in 1953 [10, 11] that the unusual features of radiation of T Tauri and UV Ceti-type stars, and the frequent appearance in their spectra of excessive shortwave radiation are signs typical of the earliest stages in the development of low luminosity stars.

Soon the discovery by Haro and his associates [12] of flare stars in stellar aggregates (associations and clusters) corroborated this point of view spanning a bridge between the stars of the T Tauri and UV Ceti types. Then the most typical flares were observed in certain stars of the T Tauri type. It became evident, in the light of those discoveries, that all the above stars form a wide class of comparatively young nonstable objects.

G. Haro [12, 13] was the first to appreciate the tremendous significance of flare stars in the picture of the earliest stages of evolution of dwarf stars. Relying on observational data concerning flare stars in stellar clusters and associations, he came to the major conclusion that the earlier stage in the evolution of dwarf stars, i.e., the stage of T Tauri (or RW Aurigae) is followed, roughly speaking, by a stage characterized by the ability to produce flares of considerable power from time to time.

An appreciable contribution to the discovery and study of flare stars has been rendered by L. Rosino and his associates [14, 15] who confirmed many of Haro's results. This line was further developed in investigations conducted at the Byurakan Observatory [4, 16, 17].

The abundance of flare stars in stellar aggregates

The distribution of flare stars in the Galaxy, at least in the region available for the observation of stars of low luminosity, is quite inhomogeneous. The investigations of Haro, Rosino and their associates [12 – 15, 18, 19] have established the fact that like stars of the RW Aurigae type, flare stars tend to form groups located in stellar associations and in comparatively young clusters. In the general stellar field, at least among the dwarfs of G, K and earlier M types, flare stars are quite few in number. There are some indications, although of no decisive value, to the effect that classical flare stars of the UV Ceti type around the Sun likewise form a physical system [1, 12, 20, 21]. Some definite evidence favoring this view has been obtained in a recent paper of M. A. Arakelian [21].

When the number of flare stars discovered in some aggregates (Pleiades, Orion) reached several dozen, the question of their total number in those systems came to the fore. A solution of this problem seemed feasible by comparing the number of those stars of the system, for which one single flare was observed, with the number of stars for which recurrent flares were observed. Of course, in this case certain assumptions are made concerning the distribution of flares of a given star in time (for instance, on Poisson distribution of the flare moments).

The statistical estimate of the total number of flare stars in the Pleiades cluster being unexpectedly great (of the order 300), a conclusion was drawn in [16] that in this comparatively young system (the age is of the order of 2.10^7 [22]) all or nearly all stars fainter than visual magnitude 13.3 are flare stars. This first and naturally quite rough statistical estimation of the total number of flare stars in the Pleiades was based on data of 60 flare stars known by 1968. Subsequent investigations [4, 17], based on richer observational data made this conclusion more substantial and precise. For the lower limit of the total number of flare stars in the Pleiades, an estimate of the order of 700 was obtained. Evidence favoring great, yet slow changes in the activity of some flare stars in the Pleiades [4] is conducive to the conclusion that the total number of flare stars is considerably larger than the estimations made.

Owing to the intense regular observations of the Pleiades region made largely at Asiago, Byurakan and Tonanzintla, the number of known flare stars in this region of the sky has thus far exceeded 207 [4], providing splendid proof of the conclusion on the abundance of flare stars in this system.

Flare stars in profusion are also observed in the Orion association. The total number of flare stars in this system is estimated to be of the order of 1000 [17].

These results, testifying to the fact that the number of flare stars is comparable to the total number of stars of low luminosities in those systems, show unequivocally that the flare star stage is a natural stage in the life of stars.

Relation between the stages of RW Aurigae and the flare star

The existence of nonstable stars, possessing simultaneously the properties of the T Tauri-type star and those of the flare star, attests that those stages in the evolution of dwarf stars sometimes mutually overlap in time. Haro [23] pictures their gradual transition in the form of the following evolutionary sequence:

1. T Tauri-type stars in which the flares superimpose with irregular changes (examples: DF Tau, YZ Ori, BW Ori).

2. Dwarfs of the later type in which spectral characteristics of the T Tauri-type star still occur, although quite reduced, and the flares remain as most remarkable changes (examples: V 389 Ori, V 390 Ori).

3. "Pure" flare stars, in which properties of the T Tauri-type virtually disappear, at least during prolonged periods of a more or less constant minimum (examples: EY Tau, FF Tau, FH Tau, V 386 Ori, V 498 Ori).

To determine the evolutionary relation between the stage of RW Aurigae and the stage of flare star, one of the authors [24] estimated the total number of RW Aurigae-type variable stars in the Orion association, which have experienced flares, based on a sample of this class of stars in this system. It was shown that only one-fourth of all the RW Aurigae-type stars in the Orion system experience flares with an amplitude exceeding $0^m.5$. In view of the fact that the RW Aurigae stage is much younger than the flare star stage the conclusion was drawn to the effect that flare activity starts only shortly before the end of RW activity. However, a review of this concept might be necessitated in the light of the possible recurrence of flare activity of the RW Aurigae-type stars [4].

Spectra of flare stars

The spectra of all the flare stars in the intervals between the flares belong to the later K and M classes. The earliest spectral class of flare stars in stellar systems correlates with their age. In the quite young systems of Orion and NGC 2264, the earliest class is K 0; in the older system of the Pleiades, it is K 3 and attains as much as M 0 – M 6 in the rather old systems of Hyades, Praesepe and Coma as well as among flare stars in the vicinity of the Sun [25, 26].

In general the spectra of flare stars outside the flares differ from the spectra of normal dwarf stars by the presence of emission lines of different intensities. In comparatively low flare activity only the lines Ca II are observed in the emission, whereas in higher activity the Balmer lines of the hydrogen series are observed as well.

However, the spectra undergo radical changes during a flare. During a flare, the spectral characteristics of flare stars almost completely coincide with the specific features observed in T Tauri-type stars: apart from a bright intense line spectrum a strong continuous emission is present, especially in the ultraviolet.

Thus the ability of a star to produce flare correlates with the presence in its spectrum of emission lines in the time of minimum brightness, testifying to their chromosphere activity. It was also made clear that the intensity of emission lines is reduced as the corresponding system advances in age (Orion-Pleiades-Hyades) [26].

H-R diagram of flare stars

On the Hertzsprung-Russell diagram flare stars fall in a region that coincides in some measure with that taken up by T Tauri stars.

On the diagram (V, B–V), compiled for the Orion association, flare stars are found with approximately 13.5 visual magnitude [18] and fainter. All the bright flare stars are located above the main sequence when deflections from the latter do not exceed 1^m in the blue region of colors and attain 4^m in the red region. Flare stars fainter than 16^m are uniformly distributed on either side of the main sequence and occupy an area of up to several magnitudes on either side [18].

The basic difference between the diagrams of flare stars in the Pleiades and in Orion lies in the fact that in the former deflections from the main sequence are considerably less, while the brighter flare stars of the cluster are absolutely fainter and have a later type spectrum than in Orion. This difference is more pronounced in the case of the clusters Hyades, Praesepe and Coma [18].

The presence of an appreciable number of stars *below* the main sequence forms a characteristic and very important feature of the diagram (V, V–B) of flare stars, particularly of the Orion association. In this sense it recalls the diagram drawn up by P. P. Parenago [27] for stars of the cluster of the Orion nebula in the region of low luminosites.

Although considerable errors in determining the magnitudes and colors of faint stars might have had their possible influence on the diagram, Haro believes [18] that both in Orion and in NGC 2264 faint flare stars exist that certainly are located much below the main sequence.

The existence of stars located on the Hertzsprung-Russell diagram in the region below the main sequence is confirmed by spectroscopic observations of the Pleiades and Hyades stars, made by G. Herbig [28], as well as by multicolor photographic observations of flare stars in Orion, made by A. Andrews [29]. It was made clear, for instance, that in this system out of 19 flare stars possessing large ultraviolet and blue excesses, 14 lie below the main sequence [29] on the diagram (V, B–V). This question needs further elaboration in view of its significance for stellar evolution.

Amplitudes of flares

The investigations of certain flare stars close to us (UV Ceti and others) have made it possible to determine the approximate law of the distribution of the amplitudes of flares. It turned out that flare frequency increases as the amplitude decreases. However, this growth takes place in such a slow way that the mean total energy, radiated by all the flares in photographic rays, appears to be considerably less than the normal radiation of the star over the same period [30, 31].

The largest flares have been observed in photographic rays in some stars in the Pleiades. Of the photographic amplitudes observed in the Pleiades the largest equals about 7^m [32]. In some cases amplitudes equal to 7^m and in one case in Orion an amplitude exceeding 8^m [33] have been observed in U-rays. In fact, photographic observations of flares in associations and clusters are of low resolving power in time. Thus in Tonanzintla, where observations in U-rays are effected, 15-minute exposures are practiced. Meanwhile, the duration of the maximum itself should be much less. Therefore, the observed values of the amplitudes need correction by $2 - 3^m$. In other words, the true maximum amplitudes of flares in U-rays possibly attain 10^m.

The available photographic observations of flares cover only flares with amplitudes $\geq 0^m.5$. Meanwhile, as noted above, flares of great amplitude (power) are far rarer than flares of small amplitudes [31, 34]. On the other hand, observations of flares are relatively complete only for bright stars. In the case of flare stars fainter than the limiting magnitude of the telescope in the minimum, we are deprived of considerably larger amplitudes as well.

Therefore, the observed distribution of amplitudes of flares in stellar aggregates is quite distorted. In particular, a sharp increase of the mean amplitude in the transition to faint stars, discovered for the Pleiades [17], is, no doubt, due appreciably to the selection of observations.

Despite the paucity of observational material, the data available suggest a direct correlation between the amplitude and the duration of flares (minimum-maximum-minimum): with an increase of the amplitude of the flare its average duration increases [35, 36].

It should be added that in most cases the amplitude of a flare grows toward the ultraviolet.

Frequencies of flares

The statistics of flares, observed in the brightest stars of the UV Ceti type, show [31] that the sequence of flares in an individual star is adequately expressed by Poisson's law. The average frequency of flares is, in general, different for various flare stars.

For instance, according to a statistical study [31], based on data of photoelectric observations of the UV Ceti and YZ Canis Minoris, the average frequencies of flares with amplitudes $\geq 0^m.15$ in those stars are equal to 0.7134 hours^{-1} and 0.2274 hours^{-1}, respectively. The average frequencies of flares in flare stars within a stellar aggregate also differ markedly. Thus, the flares observed with an amplitude $\geq 0.^m5$ in the Pleiades are expressed adequately by the superposition of Poisson distributions with two different frequencies differing by more than one order of magnitude [4, 17]. However, observational data evidences changes in the frequencies of flares in flare stars in the course of time. According to [17], the frequency of flares increases on the eve of the cessation of their flare activity. Haro suggests [18, 25] that the frequency of the flares is, on the average, greater in older systems.

Energies of flares

In accordance with Kunkel's estimation [30], the activity of flare stars in the vicinity of the Sun (UV Ceti-type stars) has an upper limit: the averaged in time energy liberated during flares does not go beyond 1% of the energy radiated by the photosphere of the star. If this conclusion, corroborated in the investigation of V. S. Oskanian and V. Yu. Terebizh [31], is valid, one should assume that the flares of flare stars, despite their high intensity, are insignificant in the energy balance of the star because of their short duration and low frequency.

Calculation indicates that continuous and irregular variations in the brightness of the T Tauri stars are considerably more efficient, as far as energy is concerned. The energies, accounting for the variations in the brightness of the T Tauri stars, are comparable with the energy of their total radiation.

It should be noted that in both cases the matter concerns those energies that have manifested themselves in the form of optical radiation. In the meantime, we already know from solar flares that a considerable part of the energy is spent on the formation of particles of high energy (protons of cosmic rays) emitted into surrounding space. If we adopt the interpretation of flares as explosions, going on above the photospheric layers, then we have to admit (see further) that at least in some cases the optical energy of flares forms only a small part of the total energy of the corresponding explosion. The total energy released during flares can apparently exceed the energy contained in optical flares, at least by one or two orders.

Proper colors of flares

The color of the excessive radiation due to flares can be derived from the observed colors of the total radiation of the flare star in the maximum and minimum brightness [37]. Observations show [38] that most of the flare stars in the minimum have the normal colors of B–V, corresponding to their absorption spectra.

The proper colors of flares, B–V and U–B, determined for a number of flare stars in the vicinity of the Sun and also for the member of the Pleiades H II 1306 [34, 37], differ only slightly from each other and are quite at variance with the colors of blackbody radiation. This fact can apparently be regarded as a proof of the general nonthermal nature of the excessive radiation, accounting for the flare, depending only loosely on the physical parameters of the star, and especially on its effective temperature.

It should be pointed out that the proper colors of flares change considerably as the flare intensifies or dies away, remaining nearly all the time on the two–color diagram (U–B, B–V) above the curve representing the colors of blackbody radiation for various temperatures [34].

Physical nature of flares

The problem of the nature of flares is of particular interest in matters of stellar evolution. It is closely related to the problem of the sources of the energy of flares.

According to a hypothesis advanced and substantiated by one of the authors of the present paper [10, 11], the continuous emission usually present in spectra of the T Tauri-type stars and appearing in the spectra of flare stars only during the flare is of nonthermal nature. We ascribe this fact to the comparatively young age of the above stars and explain it in terms of the ejection into the outer layers of the star of certain portions of pre-stellar matter and the release of energy they have brought into those layers. The process of flare, that can often occur high above the photospheric layers, results from the decay of this matter, and elementary processes must take place similar to those that are generally observed in nuclear decay.

The first part of this hypothesis (the Byurakan hypothesis) is supported by later investigations. Thus, for instance, in the works of M. A. Arakelian [39] and one of the present authors [37, 40], certain evidence, based on the analysis of observational data concerning the flares, was obtained in favor of the nonthermal and nonsynchrotron nature of continuous emission, due to the flares. This concept is apparently supported by the fact that optical flares of the UV Ceti-type stars, at least the intense ones, are accompanied by radio flares, also of a nonthermal nature [41].

Numerous attempts to explain the continuous emission by the known mechanisms of radiation failed [40, 42]. Scrutinizing and rejecting most of the existing interpretations of flares, R. E. Gershberg [42], arrives at the conclusion that the totality of optical observations is in line with the nebular hypothesis.

The nebular hypothesis, elaborated by Gershberg [43, 44] and also by Kunkel [34], assumes that the radiation of flare stars during the flares is a superposition of the radiation of the star and of a hot, ionized, and rapidly emitting gas mass ejected by the star. However, the nebular hypothesis also encounters serious difficulties [37] when it comes to an interpretation of the proper colors of flares, i.e., continuous emission. To overcome those difficulties, Kunkel [34] has considered another component of radiation, presenting the total radiation of the star during the flare as the combined radiation of cold star, hot gas envelope and heated hot spot on the surface of the star. The rapid rate of flares is thus far an insurmountable obstacle for the nebular model. The latest observations of S. Cristaldi and M. Rodono [8], made with high resolving power in time, show that the duration of a flare (minimum–maximum–minimum) can be less than 15 sec.

As to the second part of the Byurakan hypothesis [10, 11], it should be noted that in all the interpretations of continuous emission by known mechanisms of radiation, the question of the origin of energy remains open. The presence or appearance of unknown sources of energy in the star, heavily concentrated and localized not far from its surface, is always presumed. This is virtually the initial assumption of the Byurakan hypothesis (see, for instance, [40]).

It should be added that estimations of the energies accounting for the flare and for irregular variations of the T Tauri-type stars, presumably testify to the fact that the activity, produced by the nonstability of a young star, dies away with age. Presumably the complete extinction of flare activity begins when the supplies of energy are consumed.

"Fast" and "slow" flares

As noted above, the duration of a flare (minimum-maximum-minimum) depends, on the whole, on the amplitude. However, following Haro [19, 23], two essentially different types of flares can be distinguished: "fast" and "slow," which are, respectively, shorter and longer than 30 minutes. Because of the low resolving power in time of photographic observations, the threshold of 30 minutes is determined by the conditions of the observations of Haro and his associates [19, 23] who used 15-minute exposures in the U-rays.

It is highly important that in such a division into "fast" and "slow" flares the type of flare is independent of the amplitude: "fast" flares with great amplitudes and "slow" ones with small amplitudes have been observed. For instance, in the flare star FSO 7 (the flare star of Orion No. 7) a flare was recorded with an amplitude $7^m.7$ in the ultraviolet for which the maximum was attained in 23 minutes [33].

Most flares are "fast." "Slow" flares have so far been observed only in the following seven flare stars of Orion:

No.	V	B–V	References
66	$15^m.0$	$1^m.35$	[23, 29]
92	$15^m.86$	$1^m.31$	[23, 29]
149	$16^m.61$	$1^m.14$	[23, 29]
153	$15^m.56$	$1^m.32$	[23, 29]
177	$16^m.7$	$1^m.5 - 1^m.8$	[33]
229	17.6 U	–	[5]
239	19.5 U	–	[5]

In accordance with Haro's observations [19, 23], the "fast" and "slow" flares differ sharply in the nature of changes observed in the spectra. The "fast" flares are characterized by sharp changes in the radiation in U- and B-rays and strong emission lines, while in V-rays, particularly in the red region, the changes are either small or are completely lacking. In the case of "slow" flares, an intensification of the continuum in the red region of the spectrum is noticed, accompanied by low intensity of the emission lines, notably H_a.

An interpretation of the difference of "slow" and "fast" flares and the phenomenon of fuors

The differences noted in the "slow" and "fast" flares can be interpreted within the Byurakan hypothesis [10, 11] if we assume that the phenomena producing the flare take place in various layers of the star. When the excitation energy is released high above the photospheric layers, in the chromosphere or in the corona, a sudden and very rapid increase of the nonthermal shortwave (ultraviolet and blue) continuous emission occurs that causes the flare of brightness of the star. At the same time, an intense emission line spectrum appears. When the energy is released in deeper, subphotospheric layers, an increase of thermal radiation is observed in the visible region of the spectrum, together with relatively weak emission lines. The duration of the flare from the minimum to the maximum should in the latter case be much longer.

Such an interpretation of the difference between "fast" and "slow" flares suggests that the "slow" flare stars can at times experience "fast" flares too. In this connection it should be noted that out of seven "slow" flare stars in Orion, for three (FSO 66, 149, 153) "fast" flares have already been observed, while the star FSO 177 shows irregular variations of brightness outside of flares [33].

On the Hertzsprung-Russell diagram most stars which experience "slow" flares (four of the five with known colors) fall in the region above the main sequence, i.e., possess apparently more extensive photospheres than normal stars. It can be assumed that it is precisely for this reason that the probability of energy release under the photosphere ("slow" flares) is greater for them than for stars located below the main sequence.

It should be added that until recently "slow" flares have been observed only in the Orion association. Haro even concluded [23] that such flares do not occur at all in the Pleiades stars. However, E. S. Parsamian [45] has discovered a truly "slow" flare in the star FSP 103 (the flare star in the Pleiades). This fact supports the above interpretation of the differences in "fast" and "slow" flares.

The energy of the "slow" flare of the star FSO 177 was at least several dozen times larger than the maximum energy of "fast" flares in stars of the same magnitude in Orion. This fact is by now the most sound corroboration of the Byurakan interpretation. If we assume the transition of nearly all of the explosion energy into the optical region during the "slow" flare of the FSO 177, we come inevitably to the conclusion that in fast flares only about 1% of the energy is emitted into the optical part of the spectrum.

......

Luminosities and masses of flare stars

Observations of flare stars in clusters and associations indicate a great variety in their luminosities. According to Haro [18], in the Orion association the visual absolute magnitude of flare stars varies in the interval of values from $+4^m$ to $+13^m$, i.e., their luminosities can differ by four orders. The brightest flare stars in the Pleiades have an absolute magnitude of $\sim +6^m$ [16].

Thus, the flare stars in associations and clusters have in most cases luminosities that considerably surpass the luminosities of flare stars in the vicinity of the Sun.

This and the spectral classes of the UV Ceti-type stars presumably attest that the latter are rather old objects: in this group the flare activity in stars of higher luminosity has long been extinct.

The masses were determined for only a small number of the closest flare stars of the UV Ceti type. They are on the average of the order of 0.1 of the solar mass. For most flare stars the masses can be estimated roughly by means of mass-luminosity ratio, the application of which is not sufficient in this case. In view of the fact, however, that the luminosities of flare stars in stellar aggregates exceed to a great extent the luminosities of the UV Ceti-type stars, we maintain that their masses must be larger, on the average, by half an order; the great dispersion of luminosities indicates the great variety of masses in those stars.

Unusual distribution of flare stars in the Pleiades

A study of the distribution of flare stars in parent systems is of significance in the problem of stellar evolution. Such a study, carried out for Pleiades by M. A. Mnatsakanian and one of the authors of [46], has revealed that flare stars are nearly completely lacking in the central region of the system. The radius of this cavity is equal to 1.4 parsec. The partial density of flare stars reaches maximum at a distance of 1.5 parsec from the center of the system; then it diminishes more rapidly than $\sim r^{-2}$. These results have been obtained under the assumption that all known flare stars in this region belong to the Pleiades system.

The existence of a cavity in the distribution of flare stars in the central region of the Pleiades can hardly be explained by the influence of the absorbing matter.

A plausible explanation is that stars originating in the central part of the cluster move from the center of the system in the course of their aging.

Conclusion

Flare stars form one of the initial stages in the evolution of dwarf stars. In this connection a detailed and profound study of flare stars is of paramount importance for the problem of stellar evolution.

The conclusion of Haro [18] can be quoted to the effect that there is a considerable number of young stars located on the Hertzsprung-Russsell diagram *below* the main sequence.

. . . .

REFERENCES

1. V. S. Oskanian, 1964, *Publ. Obs. Beograd*, No. 10.
2. W. J. Luyten, 1949, *Ap. J.*, 109, 532.
3. A. H. Joy, M. L. Humason, 1949, *PASP*, 61, 133.
4. V. A. Ambartsumian, L. V. Mirzoyan, E. S. Parsamian, H. S. Chavushian, L. K. Erastova, 1971, *Preprint Byurakan Obs.*, No. 2, 1971; *Astrofizika*, 7, in press.
5. G. Haro, E. Chavira, 1969, *Bol. Obs. Tonanzintla*, 5, No. 32, 59.
6. L. Rosino, L. Pigatto, 1969, *Contr. Obs. Asiago*, No. 231, 13.
7. V. S. Oskanian, Private communication.
8. S. Cristaldi, M. Rodono, 1971, *IBVS*, No. 525, No. 526.
9. V. A. Ambartsumian, L. V. Mirzoyan, G. S. Sahakian, S. K. Vsekhsviatski, V. V. Kazutinski, 1969, *Problems of Modern Cosmology*, Nauka, Moscow [in Russian].
10. V. A. Ambartsumian, 1954, *Soobshch. Byurakan Obs.*, 13.
11. V. A. Ambartsumian, 1957, *Non-Stable Stars, IAU Symposium No. 3*, ed. G. H. Herbig, University Press, Cambridge, p. 177.
12. G. Haro, 1957, *Non-Stable Stars, IAU Symposium No. 3*, ed. G. H. Herbig, University Press, Cambridge, p. 26.

13. G. Haro, 1962, *Symposium on Stellar Evolution*, ed. G. Sahade, Astr. Obs. Nat. Univ. La Plata, La Plata, p. 37.
14. L. Rosino et al., 1956, *Contr. Obs. Asiago*, No. 69; 1962, No. 125; 1964, No. 127; 1966, No. 189.
15. L. Rosino, 1969, *Low-Luminosity Stars*, ed. Sh. S. Kumar, Gordon & Breach Science Publishers, N. Y., London, Paris, p. 181.
16. V. A. Ambartsumian, 1969, *Stars, Nebulae, Galaxies*, Acad. Sci. Armenian SSR, Yerevan, p. 183 [in Russian].
17. V. A. Ambartsumian, L. V. Mirzoyan, E. S. Parsamian, H. S. Chavushian, L. K. Erastova, 1969, *Preprint Byurakan Obs.*, No. 1; Astrofizika, 1970, 6, 3.
18. G. Haro, E. Chavira, 1964, *Vistas in Astronomy*, Vol. 8, ed. A. Beer & K. Aa. Strand, Pergamon Press, London, p. 89.
19. G. Haro, 1968, *Stars and Stellar Systems*, Vol. 7, ed. B. M. Middlehurst & L. H. Aller, University Press, Chicago, p. 141.
20. V. A. Ambartsumian, 1957, *Non-Stable Stars*, ed. M. A. Arakelian, Acad. Sci. Armenian SSR, Yerevan, p. 9 [in Russian].
21. M. A. Arakelian, 1968, *Non-Periodic Phenomena in Variable Stars*, ed. L. Detre, Academic Press, Budapest, p. 161.
22. A. R. Sandage, 1957, *Proc. Vatican Conference on Stellar Populations*, p. 41.
23. G. Haro, 1964, *The Galaxy and the Magellanic Clouds, IAU-URSI Symposium No. 20*, Australian Acad. Sci., Canberra, p. 30.
24. V. A. Ambartsumian, 1970, *Astrofizika*, 6, 31.
25. G. Haro, E. Chavira, 1969, *Bol. Obs. Tonanzintla*, 5, No. 31, 23.
26. G. Haro, E. Chavira, 1970, *Bol. Obs. Tonanzintla*, 5, No. 34, 181.
27. P. P. Parenago, 1954, *Trudy Astr. Inst. Sternberga*, 25.
28. G. H. Herbig, 1962, *Ap. J.*, 135, 736.
29. A. D. Andrews, 1970, *Bol. Obs. Tonanzintla*, 5, No. 34, 195.
30. W. E. Kunkel, 1970, *Communication at the General Assembly of the IAU*, Brighton.
31. V. S. Oskanian, V. Yu. Terebizh, 1971, *Astrofizika*, 7, 83, 281.
32. E. S. Parsamian, Private communication.
33. G. Haro, E. S. Parsamian, 1969, *Bol. Obs. Tonanzintla*, 5, No. 31, 45.

34. W. E. Kunkel, 1967, *An Optical Study of Stellar Flares*, The University of Texas, Austin.
35. L. Rosino, 1958, *Mem. Soc. Roy. Sci. Liege*, IVe serie, 20, 285.
36. L. V. Mirzoyan, H. S. Chavushian, 1970, *Soobshch. Byurakan Obs.*, 42, 17.
37. L. V. Mirzoyan, 1966, *Astrofizika*, 2, 121.
38. H. L. Johnson et al., 1962, *Ap. J.*, 128, 31, 1958; 136, 75, 1962.
39. M. A. Arakelian, 1959, *Doklady Akad. Armyan. SSR*, 29, 35, 167.
40. L. V. Mirzoyan, 1967, *Some Problems of Kinematics and Physics of Young Stars*, Main Astr. Obs. Acad. Sci. USSR, Poulkovo (in Russian).
41. B. Lowell, 1964, *Observatory*, 84, No. 938, 18.
42. R. E. Gershberg, 1970, *Flares of Red Dwarf Stars*, Nauka, Moscow (in Russian).
43. R. E. Gershberg, 1964, *Izv. Krym. Astrofiz. Obs.*, 32, 133; 1965, 33, 206.
44. R. E. Gershberg, 1967, *Astrofizika*, 3, 127.
45. E. S. Parsamian, 1971, *Astrofizika*, 7, in press.
46. L. V. Mirzoyan and M. A. Mnatsakanian, 1971, *IBVS*, No. 528.

Byurakan, Armenia
Astronomical Observatory of the Academy
of Sciences of Armenia

Editor's Addendum: Frequencies of Flares in a Stellar Aggregate

This brief comment concerns the problem of statistical determination of the number of "potential" flare stars in a stellar aggregate which showed no flashes during the observation period.

The mathematical model proposed by V. A. Ambartsumian in [1] makes it possible to draw conclusions about this number. The purpose of this note is to point out the remarkable robustness of the statistical estimate proposed by Ambartsumian, which was not mentioned in [1]. Axiomatically, the model is as follows:

a) Within the aggregate of stars there is an unknown number N of flare stars.

b) To each flare star a value of parameter $\lambda > 0$ is assigned. The values of λ for different flare stars are independent, identically distributed random variables. Their common probability density function $f(\lambda)$ is unknown.

c) Conditional upon its value of λ, each flare star produces flashes at moments of stationary Poisson process of rate $\lambda > 0$.

Recall that in a stationary Poisson process of flashes of rate λ the number of flashes in non-overlapping time intervals is independent and that the probability of k flashes in a time interval of length τ is

$$\frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau}, \quad k = 0, 1, 2, \ldots .$$

Let $p_k(\tau)$ be the probability that for a typical flare star the number of flashes observed during the time interval will be k. Clearly

$$p_k(\tau) = \int \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau} f(\lambda)\, d\lambda.$$

The product

$$N_k(\tau) = N \cdot p_k(\tau)$$

equals the mean number of flare stars which produce exactly k flashes during the observation period. It is natural to take the random quantity

n_k = the number of stars which produced k flashes during the observation period $(0, \tau)$.

as an estimate of $N \cdot p_k(\tau)$.

This we express by writing

$$n_k \approx N \cdot p_k(\tau). \tag{1}$$

This estimate becomes reasonable for larger values of τ. The problem is now formulated: *find an estimate for $N_0(\tau)$ in terms of the numbers n_k, $k > 0$.*

We will consider the values

$$n_1 = 123, \quad n_2 = 16, \quad n_3 = 2, \tag{2}$$

which have been observed in Pleiades, see [1].

An example of an estimate for $N_0(\tau)$ can be obtained in the following way:

We write an identity

$$N_0(\tau) = N \cdot p_0(\tau) = \frac{(N \cdot p_1(\tau))^2}{N \cdot p_2(\tau)} c(\tau) = \frac{(N_1(\tau))^2}{N_2(\tau)} c(\tau),$$

where

$$c(\tau) = \frac{p_0(\tau)\, p_2(\tau)}{(p_1(\tau))^2}.$$

Replacing $N_1(\tau)$ and $N_2(\tau)$ by the observed quantities n_1 and n_2, we obtain

$$N_0(\tau) \approx \frac{(n_1)^2}{n_2} c(\tau). \tag{3}$$

The presence of the factor $c(\tau)$ which depends on the unknown density $f(\lambda)$ bars (3) from direct application. However, $c(\tau)$ can be calculated for special choices of $f(\lambda)$. Thus, it was noted in [1] that

$$c(\tau) \equiv \frac{1}{2}, \quad \text{if} \quad f \quad \text{is} \quad \delta\text{–type, i.e.,} \quad f(\lambda) = \delta_{\lambda_0}(\lambda), \text{ for some } \lambda_0 > 0, \tag{4}$$

for any choice of $\lambda_0 > 0$, as well as

$$c(\tau) \equiv 1 \quad \text{if} \quad f(\lambda) = \alpha\, e^{-\alpha\lambda} \tag{5}$$

for any choice of $\alpha > 0$.

The estimate of $N_0(\tau)$ proposed in [1] was based on (3) and assumption (4). Ambartsumian wrote a Schwartz inequality

$$\left(\int \lambda e^{-\lambda\tau} f(\lambda)\, d\lambda\right)^2 \leq \int \lambda^2 e^{-\lambda\tau} f(\lambda)\, d\lambda \int e^{-\lambda\tau} f(\lambda)\, d\lambda,$$

from which upon multiplication by $N^2\tau^2$ he obtained

$$N_0(\tau) \geq \frac{(N_1(\tau))^2}{2\,N_2(\tau)}, \tag{6}$$

i.e., $(n_1)^2/(2n_2)^{-1}$ becomes an estimate of the lower bound for the quantity in question.

It is natural to try to extend these ideas by considering a family of *double ratio estimates*

$$N_0(\tau) \approx \frac{n_k(\tau)\, n_l(\tau)}{n_m(\tau)}\, c_{k,l,m}(\tau), \tag{7}$$

where k, l, m are nonnegative integers and

$$c_{k,l,m} = \frac{p_0(\tau)\, p_m(\tau)}{p_k(\tau)\, p_l(\tau)}. \tag{8}$$

In all cases the functions $c(\tau)$ depend on the unknown density $f(\lambda)$.

Our point is that even if we assume that $c(\tau)$ is known explicitly (as in cases (4) and (5)), still all these estimates are reasonable only for values of τ large enough ($\tau \approx \infty$), so that the law of large numbers becomes valid.

Then, since $\tau \approx \infty$ is a necessary condition, it is natural to replace $c(\tau)$ in these formulae by the corresponding *asymptotic decompositions*, as $\tau \to \infty$. These decompositions may depend solely on a few parameters determining with chosen precision the behavior of the corresponding $c(\tau)$ at $\tau = 0$.

We hope to obtain in this way a simpler problem involving several unknown parameters instead of an unknown function $c(\tau)$.

Assume that at the point $\lambda = 0$ our f has a Taylor approximation

$$f(\lambda) = f(0) + f'(0)\,\lambda + O(\lambda^2) \tag{9}$$

with $f(0) > 0$.

Then by an easy analysis we find the first three terms of the asymptotic expansion of $p_s(\tau)$:

$$p_s(\tau) \approx \frac{1}{\tau} f(0) + \frac{s+1}{\tau^2} f'(0) + O\left(\frac{1}{\tau^2}\right).$$

Substituting these terms into (8) yields, after simplification,

$$c(\tau) \approx \frac{\left[f(0)\tau^2 + f'(0)\tau\right]\left[f(0)\tau^2 + (m+1)\tau f'(0)\right]}{\left[f(0)\tau^2 + (k+1)\tau f'(0)\right]\left[f(0)\tau^2 + (l+1)\tau f'(0)\right]},$$

or asymptotically

$$c(\tau) = 1 + (m - k - l)\, b + O\left(\frac{1}{\tau^2}\right), \tag{10}$$

where

$$b = \frac{f'(0)}{\tau f(0)}.$$

The most attractive are the *robust* cases where

$$k + l = m, \tag{11}$$

i.e., where the above reduces to

$$c(\tau) = 1 + O\left(\frac{1}{\tau^2}\right).$$

The Ambartsumian case $k = 1$, $l = 1$, $m = 2$ is robust. Taking $c(\tau) \equiv 1$, we obtain from (7) and (2)

$$N_0(\tau) \approx \frac{(n_1)^2}{n_2} \approx 946. \tag{12}$$

In another robust case $k = 1$, $l = 2$, $m = 3$, taking $c(\tau) \equiv 1$, we obtain from (7) and (2)

$$N_0(\tau) \approx \frac{n_1\, n_2}{n_3} = 984. \tag{13}$$

The two numbers agree with 4% *precision.*

Although the estimate (13) becomes unstable for smaller values of n_3, this result confirms (12) and, what is more important, the special role of the ratio $(n_1)^2/n_2$ considered in [1].

REFERENCES

1. V. A. Ambartsumian, L. V. Mirzoyan, E. S. Parsamian, H. S. Chavushian, and L. K. Erastova, 1969, *Preprint Byurakan Obs.*, No. 1; *Astrofizika*, 1970, 6, 1969, 3.

FUORS[1]

The characteristic trait of FU Orion-type stars (fuors) is that they suddenly increase their luminosity in some spectral region by over a hundredfold within a short time span and retain enhanced luminosity for many years. An explanation is offered for this phenomenon, based on the assumption of the presence, prior to the rise in the light curve in the region adjacent to the star, of sources of corpuscular radiation. As indicated by observations, at the time of the rise in brightness, an envelope encompassing those external sources is developed. For that reason, following the rise in brightness, almost all energy from those sources is emitted in the form of thermal radiation flux.

If some rapidly unfolding phenomenon occurs, say, only once during the lifetime of a star, that phenomenon will be observed extremely infrequently among the stars surrounding us. However, it may mark a regular and even major stage in the evolution of all stars or, say, of stars having masses within a certain specified range.

Here we wish to draw attention to a group of phenomena which is observed with extreme rarity and which may shed some light on the problems of stellar evolution.

1. **FU Orionis Stars**. At the close of the past year, we received from the Swedish astronomer Gunnar Wellin a preprint of his short report on a star LkH_α 190 located in the North America nebula (NGC 7000) among a group of stars featuring bright lines (type T Tauri, etc.) with the magnitude $m_{pg} = 16.0$ indicated by Herbig in 1957, and varying its brightness only

[1]English translation originally published by Plenum Publishing Corp. © 1971 in *Soviet Astrophysics*, vol. 7, no. 4, 1971. Used here with permission of the Plenum Publishing Corp. The original version of this paper consisted of six sections, of which §§4 and 6 have been omitted from the present collection. The numbering of the sections has been changed correspondingly.

slowly over the ensuing years, with $m_{pg} = 10.0$ reported in 1970 [1]. Comparison of various negatives indicates that an abrupt increase in brightness occurred in late 1969. Since that time the brightness of the star varied slightly. According to photoelectric measurements taken by Grigoryan in mid-July 1970 (at Byurakan), its brightness was $m_V = 10.8$. We are dealing in this case with an abrupt rise in brightness similar to one witnessed in 1936 in FU Ori. Prior to the increase in brightness, the magnitude of the star FU Ori was 16, but it became brighter than a 10th magnitude star in 1936 following a flare; later, after attenuating slowly, it eventually reached the magnitude $m_{pg} = 10.5$; since then the brightness has hardly varied at all. In both instances the star shifted abruptly from one state, in which its brightness fluctuated slightly about some low level, to another state where it was approximately a hundredfold greater.

Because the higher level of brightness lasts for a period of at least decades we prefer not to use the term "flare stars" in this context. We will call these objects "fuors".

Note that P Cyg furnishes an example of essentially a similar phenomenon. Prior to the brightness rise this star was not visible to the unaided eye. Its magnitude is now $m_{pg} = 4.8$.

In the case of FU Ori, we know the approximate brightness prior to the rise in luminosity ($m = 16.0$). There is only one case (LkH_α190) where we know the spectrum prior to the rise in brightness: it corresponded to a T Tauri-type late dwarf [2]. Unfortunately, owing to the small dispersion, Herbig was unable to give a determination of spectral type based on absorption lines. There are also other objects, in which an appreciable enhancement of brightness was observed within a short period of time, resulting in a more or less stationary state. We cannot exclude the possibility that these objects also belong to the group of fuor-stars.

It is an essential point that both FU Ori, and especially LkH_α190, demonstrate, after the rise in brightness, spectral features typical of stars of relatively high luminosity. In particular, in the star LkH_α190 the observed H_α emission line has an absorption component on the shortwave side which is displaced 420 km/sec. In other words, that star now features a continuous outflow of matter similar to that established in the case of P Cyg. Such an outflow of matter must result in an extended shell around the star. In addition, the atmospheres of LkH_α190 and FU Ori are rich in lithium, which is typical of young stars. Finally, the three stars belong to stellar associations.

Within the usual concepts regarding stellar evolution, the shift of a star from one level of more or less stationary luminosity to another level many times higher must be accounted for in terms of the total output of the energy sources present in the star. But it would be difficult to imagine that the internal structure of the star could vary to such an extent within a mere few months that the total output of the energy sources would increase more than a hundredfold. We therefore have to find some other explanation.

The gist of the explanation which we propose is that there exist intense and constantly active energy sources in the space above the photosphere in some or all of the T Tauri-type stars, somewhere in the region of the corona or even higher. A portion of this energy is released in the form of nonthermal continuous emission in blue, violet and ultraviolet light. In some T Tauri-type stars, this emission is so intense that it is observed directly in the form of an "ultraviolet excess" in the spectrum of the corresponding star (e.g., XX Ori, NS Ori, NX Monocerotis). The maximum of this nonthermal emission is found in the far ultraviolet and is not observable from the earth. In many cases, the "tail" of that excess, extending out from the near ultraviolet observable from the earth, is so faint that it is not perceptible against the background of the star's thermal emission in the same wavelengths. Nevertheless, the existence of the excess in the far ultraviolet can be ascertained with reasonable confidence from the existence of the emission lines in the spectra of those stars. We need not dwell here on which portion of the energy released by the nonthermal sources is converted to electromagnetic radiation and which portion is released in the form of the kinetic energy of corpuscular matter ejected into the surrounding space. However, if we assume that the energy released is the result of primary processes of nuclear-type decay, then the conversion factor of that energy into photographic light observable from the earth's surface must be very small in those cases where the energy is released in the rarefied interstellar space. It is probably less than 0.01. All of the remaining energy in the supposed decay process must be released either in the form of the kinetic energy of the particles emitted or else in the form of short-wavelength electromagnetic radiation which is filtered out by our atmosphere.

On the other hand, if a shell which is opaque not only to shortwave radiation but also to high energy particles develops around the star for any reason whatever, then all of the energy due to sources located within the shell will be released in the form of thermal radiation of the shell. When shell temperatures are of the order of $10,000^\circ$, the conversion factor into photographic light will be close to unity.

In other words, the formation of an opaque shell must lead, under those conditions, to intensified conversion of the energy released by the presumed sources into photographic light, by a factor of more than a hundred.

Hence, we assume that we are not dealing here with an increase in the intensity of the sources of energy, but rather with an increase brought about by the development of the shell as the factor of conversion to photographic light of the energy released by the presumed sources.

2. **The Concept of Calorimetric Stellar Magnitudes**. The concept of visible and absolute bolometric magnitudes and bolometric corrections to the visual or photographic magnitudes have proved quite useful in discussions of stellar luminosity. In relation to bodies which emit appreciable corpuscular radiation compared with the amount of kinetic energy carried off, we introduce a system of stellar magnitudes characterizing energy emitted in unit time, including both the total energy of the electromagnetic radiation and the kinetic energy of the corpuscles emitted. It is convenient to term this system the calorimetric system of stellar magnitudes. A natural definition of such stellar magnitudes is provided by the formula

$$m_{kal} = m_{bol} - 2.5 \log \frac{L_k + L}{L}, \tag{1}$$

where L is the luminosity at electromagnetic wavelengths, and L_k is the total kinetic energy carried off in unit time by the particles emitted.

Let us now define the "calorimetric correction"

$$\delta' = m_{kal} - m'_{pg} \tag{2}$$

for a fuor prior to the rise in brightness. Subsequently, all stellar magnitudes referring to the stage preceding the rise in brightness will be designated by a single prime, while those referring to the stage following the rise in brightness will be designated by double primes. The gist of our hypothesis can be expressed in terms of the equation

$$m_{kal} \approx m''_{bol}. \tag{3}$$

Upon comparing (3) and (2), we can state

$$\delta' \approx (m''_{bol} - m''_{pg}) + (m''_{pg} - m'_{pg}). \tag{4}$$

The first term in the right-hand member of (4) is the bolometric correction to the photographic magnitude after the flare. Since a fuor emits normal thermal radiation after the rise in brightness, this correction can be calculated on the basis of the effective temperature. For $T = 10,000°$, we have the value -0.4. The correction is probably good for both FU Ori and for LkH_α 190. As for the second term in (4), it represents the observed rise in brightness, which is 5 magnitudes in both instances. Hence, we have

$$\delta' = -5.4.$$

But the resulting calorimetric correction for an object which has yet to undergo a rise in brightness (hence, a prefuor) consists of two parts: the thermal emission (t) of the star and the nonthermal emission (nt) emanating from the source(s) located above the photosphere. Evidently, the factor of conversion of the energy released by those sources into radiation in the photographic portion of the spectrum is determined preponderantly by the second part. We therefore have to look separately for the calorimetric correction to the nonthermal radiation of a prefuor. We denote this correction as δ. We now have

$$\delta = m'^{nt}_{kal} - m'^{nt}_{pg}. \tag{5}$$

Let us now use the equation

$$10^{-0.4\, m'^{nt}_{kal}} = 10^{-0.4\, m_{kal}} - 10^{-0.4\, m'^{t}_{bol}},$$

where m'^{t}_{bol} denotes the bolometric thermal radiation of the prefuor. The equation means that the calorimetric luminosity of the prefuor is on the whole the sum of the intensity of thermal radiation by the star and the calorimetric luminosity of the nonthermal sources (producing both corpuscular radiation and electromagnetic radiation). Taking (3) and (5) into account, we readily find

$$\delta = m'^{nt}_{kal} - m'^{nt}_{pg} = \\ (m''_{bol} - m''_{pg}) + (m''_{pg} - m'^{nt}_{pg}) - 2.5 \log \left[1 - 10^{-0.4\,(m'^{t}_{bol} - m''_{bol})}\right]. \tag{6}$$

In essence, however, the last term is very small, so that we can resort to the formula

$$m'^{nt}_{kal} - m'^{nt}_{pg} \approx (m''_{bol} - m''_{pg}) + (m''_{pg} - m'^{nt}_{pg}). \tag{7}$$

Unfortunately, we cannot determine the magnitude of the nonthermal component in photographic light from spectral observations of the prefuor LkH_α190 by Herbig. But since this component is weak, we surmise that it accounts for not more than 15% in photographic light. That would mean $m'^{nt}_{pg} \approx 18.0$. On the other hand, the great intensity of the emission lines of the Balmer series in a prefuor spectrum argues in favor of a rather large excess in the far ultraviolet. That supports the assumption that the excess cannot be much less than the indicated 15% in the near ultraviolet. We can therefore settle for a rough assignment $m'^{nt}_{pg} \approx 18.0$. Consequently, (7) yields, for LkH_α 190:

$$\delta = m'^{nt}_{kal} - m'^{nt}_{pg} \approx -7.4.$$

As we readily see from the tabulated bolometric corrections and color indices of Planck radiation, the highest value of the differences $m_{bol} - m_{pg}$ is attained at $T = 8,000°$ and is -0.2 [3]. Consequently, the value we obtained indicates that, in a prefuor, the factor for conversion of energy into photographic light is at least 700 times less than in normal thermal radiation by F-type stars, where it is at a maximum. All of that signifies that the rise in brightness experienced by a fuor is due to an at least several hundredfold increase in the conversion factor.

3. Slow Flares and Fast Flares in Flare Stars. In the present section, it is our intention to go into somewhat further detail than was done in 1954 [8] on flares occurring in several late UV Cet-type dwarfs in the

vicinity of the Sun, and in broader groups of dwarfs present in associations (Orion, NGC 2264, NGC 7023) and in young clusters (Pleiades).

The essence of the concept which we proposed at that time was that each flare is the result of liberation of some amount of energy which is heavily concentrated prior to the flare in some portion of the "prestellar material." We deliberately eschewed constructing hypotheses on the nature of that superdense prestellar material. Certain masses of that matter capable of existing in a stable state for a protracted period were in question, masses capable of being carried out into the space surrounding the star (possibly into the coronal layers, or even further, out to distances in excess of several stellar radii), and susceptible to almost instantaneous decay at those distances.

The fact that the phenomenon observed takes place as a rule above the star's photosphere follows from the particular energy distribution in the continuous spectrum of the flare (large ultraviolet excess). Here there are no significant quantities of absorbing matter or conditions for thermalization of the emission spectrum. The fact that we are dealing here with an explosion, rather than a quiet expansion of the mass of hot gas ejected from the star, as proposed by several authors, is confirmed by photoelectric observations with a high time resolution. They show that the increase in brightness is often measured literally in seconds.

We have pointed out that in addition to the cases where the energy is released above the photosphere layers we can conceive of cases where the energy is liberated underneath the photosphere layers. The latter cases can be separated into two groups:

1) Release of energy occurring deep in the inner layers of the star, with the energy working its way to the surface over the course of many months or years. In that case, the very process of energy release will become extended at least by a matter of weeks or months. That means that we shall not be observing any separate and distinct flares, but only their overall averaged result, which reduces to some small increase in the brightness of the star.

2) The release of energy takes place directly beneath the photosphere layers, at a depth from which the energy makes its way (via diffusion of the radiation or an ionization wave) to the surface in the course of several hours. The observed flare process then should last several hours. The

process of the rise in brightness of the star must proceed at a much slower pace than in those cases where the liberated energy advances above the star's surface, and the color of the additional radiation must be a function of the amplitude of the brightness. The lower that amplitude, the lower must be the color temperature of the additional radiation.

Professor Haro, in his first observations of "slow flares," which differ radically in their nature from "fast flares," fully confirmed the existence of two classes of flares in flare stars in the constellation Orion, and the recent discovery by Parsamian [7] of a slow flare in the Pleiades showed that slow flares may occur in aggregates older than the Orion association.

We would now like to focus attention on some quantitative data derived from observations which proved to be in close agreement with our hypothesis on the nature of fast and slow flares.

The problem is that, if the flares are the result of disintegration of dense matter, i.e., of some mass of nuclear density, into an assemblage of particles, then conversion of the decay energy into the optical radiation at the frequencies we observe will be very small in the vacuum. Most of the decay energy will become converted either into the kinetic energy of the particles formed (as occurs, for example, in β decay) or into electromagnetic radiation such as γ photons, X-ray photons, or radiation in the far ultraviolet.

An entirely different state of affairs prevails when decay takes place underneath the photosphere layers. In that case, all of the decay energy, except perhaps for neutrino energy, will become converted into the thermal energy of the star's radiation. In other words, the total flare energy in the form of optical light must be many times greater in those cases than in the case of fast flares. This ratio depends on the concrete mechanism acting in the flare process. One of the possible concretizations to consider is the mechanism proposed by Gurzadyan, wherein anti-Compton scattering of quanta of the star's thermal radiation takes place on the electrons (or positrons) released in the decay process. That mechanism would imply a value of the conversion factor below 0.01. Then the energy at optical wavelengths must be a hundred times greater in slow flares than the energy at optical wavelengths released in fast flares.

Observations show that: 1) slow flares are observed many times less frequently than fast flares; 2) the amplitudes observed in slow flares are not smaller than those observed in fast flares; while the largest amplitudes of fast flares in photographic light attain a magnitude of 5 in the Orion stellar association, one of the slow flares observed in Orion (at the star VZO 177) by Haro had an amplitude of 8.4 in that range of light; 3) the color of the emission of slow flares is redder than that of fast flares.

The first of the circumstances enumerated here seems to be due to the fact that the energy must be released in a layer of relatively small linear thickness (possibly of the order of a hundred kilometers) beneath the photosphere (case 2) in order for a slow flare of some appreciable amplitude to be observed. For instance, if we assume that the decay of masses ejected outward takes place more or less spontaneously, then the decay probability in any layer must be proportional to the residence time in that layer, i.e., must be proportional to the thickness of that layer. In that light, the infrequency of slow flares presents no difficulty in understanding.

The second of these circumstances is a direct indication that the observed total energy of the optical radiation emitted in slow flares is several dozen times greater than the total energy observed in the optical range in fast flares, since the duration of the slow flare is dozens of times longer when the radiation intensity is of that order of magnitude. Consequently, the observed relationship between slow and fast optical flares is in complete accord with the concept of different values of the conversion factor in those two instances which we developed above.

In sum, the available data on differences between slow flares and fast flares confirm the hypothesis to the effect that flares are associated with high-energy decay processes.

. . . . . .

4. **Duration of the Postfuor Stage**. We can safely infer from the available data on fuors that the process by which the brightness rise takes place is associated with the following transition:

$$\text{T Tauri-type stars with UV excess (prefuor)} \longmapsto$$
$$\text{Fuor phenomenon (brightness rise, shell formation)} \longmapsto$$
$$\text{A-type stars with P Cygni characteristic (postfuor).}$$

Even though postfuors probably represent a group of objects which is fairly homogeneous in many physical properties, it should still be pointed out that the absolute magnitudes are close to zero in the two cases in question (FU Ori and LkH 190), whereas the absolute photographic magnitudes are of the order of -7.0 in P Cygni itself, and also in other P Cygni-type objects found in O-type associations. It should be acknowledged that we are not yet in a position to state from which objects P Cyg-type supergiants originate, but their frequent presence in O-associations suggests that here too the initial phase was a T Tauri-type star with energy sources of enormous intensity.

Once the frequency of occurrence of fuors is known, as well as the number of objects with a P Cyg spectral characteristic, we can estimate roughly the duration of the postfuor stage, or, more precisely, that portion of it where a continuous outflow of matter is present.

Only two typical fuors have been observed in the last 50 years (i.e., since 1920) that in postfuor stage remained brighter than $11^m.0$. Of course, there may have been cases where the fuor phenomenon escaped notice. But it must be assumed that if two plates taken at epochs separated by twenty years or so were ever compared for a given region of the sky, a postfuor brighter than $11^m.0$ that flared in the time period between the two plates would be detectable with a probability close to unity.

Even though such comparisons have been carried out so frequently at observatories that they undoubtedly encompass most of the Northern Hemisphere, the time intervals between two plates are still not very long. Even if we assume that the value of the time interval Δt, averaged over that part of the sky for which comparisons have been carried out, was 20 years, then we could detect only 40% of the stars that fuorized in the last 50 years and remained brighter than $11^m.0$. Then the total number of stars fuorized during this half-century in the Northern Hemisphere will be of the order of 5; in other words, a single star fuorized in one decade on the average. On the other hand, if T is the average duration, expressed in years of that postfuor phase when the P Cygni characteristic in the spectrum is still detectable and, moreover, the brightness shows no appreciable decrease, then we must have

$$N_p = 0.1 \cdot T$$

for the total number N_p of stars brighter than $11^m.0$ with a P-characteristic.

Unfortunately, the available data are not sufficient to estimate N_p. However, not more than 10 or so stars with a P-characteristic are known [5] among the stars in the HD catalogue of the northern sky. Herbig [6] has made a detailed investigation of the spectra of stars associated with cometary nebulae, but could detect only 4 stars with a P-characteristic, and one of those was, however, of magnitude $13^m.0$. Nevertheless, it can be safely assumed that a more detailed study of the spectra of most of the HD stars, particularly in the region of the H_α line, will result in doubling or even tripling the number of objects discerned with a P-characteristic. In addition, the HD catalogue contains only a small portion (about a third) of the stars brighter than $11^m.0$. Consequently, a crude assumption would be $N_p \approx 60$ in the northern sky. That would mean that the duration of the postfuor stage we are interested in here must be of the order of 600 years. By the way, we have to be careful in our conclusions, since we do not know the true frequency of occurrence of fuors becoming transformed into P Cyg-type supergiants, not even roughly. The lifetime of these supergiants can be much greater than in postfuors of lower luminosity.

There is no doubt, however, that the phase of the P Cyg-type spectrum in postfuors of low luminosity is not very protracted. Our calculations were very rough, but we can still state that the postfuor stage in those instances measures by order in thousand of years. But we then are confronted with the question of what happens after that stage has gone to completion; i.e., we have to deal with what we might call the postpostfuors, and whether the star resumes its initial brightness, i.e., reverts to the brightness of a prefuor or, on the contrary, retains its enhanced brightness. It would be difficult to venture an answer to that question at the present time. The only point that is certain is that cases of an abrupt fall-off in brightness (the antifuor case) have not yet been discovered to date. We are then left with two possibilities: retention of the level of brightness achieved or a gradual elimination of the shell with a fall-off in brightness for decades or centuries. But if the first possibility holds, then as many as tens of millions of postfuors exhibiting visible brightness greater than the 11th magnitude would have accumulated over the course of, say, hundreds of millions of years. But there is in reality no such quantity of stars brighter than 11th magnitude. We therefore have to conclude that the brightness of the star must decline again within a short time.

......

The author is deeply grateful to Dr. H. Wellin for sending information on the fuor in Cygnus, to Professor G. Haro for communicating in writing that he had arrived at the concept of a new class of objects independently of the present author and for stressing the possible relationship to stellar flares, and also to É. Parsamian for shedding additional light on the topics pertinent to slow flares. The author is also indebted to the late K. Grigoryan for an oral communication on estimates of the brightness of the fuor in Cygnus.

REFERENCES

1. G. H. Herbig, *IBVS*, No. 543 (1971).
2. G. H. Herbig, *Ap. J.*, 128, 259 (1958).
3. C. W. Allen, *Astrophysical Quantities*, Athlone Press, London (1955).
4. G. Haro and É. Parsamian, *Bol. Obs. Tonatzintla*, 5, No. 31, 45 (1969).
5. P. Merrill and C. Borwell, *Ap. J.*, 78, 87 (1933).
6. G. H. Herbig, *Contr. Lick. Obs., Ser. II*, No. 99, (1960).
7. É. S. Parsamian, *Astrofizika*, 7, 500, (1971).
8. V. A. Ambartsumian, *Soobshch. Byurakan. Obs.*, No. 13, (1954).

3 August 1971

Byurakan
Astronomical Observatory,
Academy of Sciences of Armenia

INSTABILITY PHENOMENA IN SYSTEMS OF GALAXIES[1]

The dynamics of clusters and groups of galaxies essentially depends on their masses. Unfortunately, the data available on the masses of galaxies are poor.

Observations of the orbital motions in double stars have been used to determine stellar masses. A great number of already determined orbits of visual and spectral doubles form a sound basis for our knowledge of stellar masses. In contrast to this, we are in no position to determine the orbits in double galaxies, and attempts to make statistical use of the differences of radial velocities in double galaxies meet with considerable difficulties in introducing this or that hypothesis on the nature of motion (elliptical or hyperbolic). What remains is to determine the masses of galaxies from measurements of rotation and of the internal motions within a given galaxy.

Unfortunately, the data obtained in this way accumulate very slowly. Thus the value commonly accepted at present for the mass of our Galaxy might very likely turn out to be incorrect by a factor of 2. The estimate of the mass of the Large Magellanic Cloud is very uncertain. Our knowledge of the masses of both giant and dwarf elliptical galaxies is quite limited.

Nonetheless, the data available have made it possible to draw the following valuable conclusions as to the values of the ratio $f = M/L$.

(a) The ratio $f = M/L$ decreases at least 10 times when passing from the elliptical galaxies of high luminosity to the spirals and further to the irregular galaxies.

(b) The ratio f does not increase but, most probably, decreases when passing from supergiant elliptical galaxies through giants to the dwarf systems like those in Fornax and Sculptor.

[1]Originally published by the American Astronomical Society © 1961 in *The Astronomical Journal*, vol. 66, 1961, pp. 536-540. Used here with permission of the American Astronomical Society.

As a result, the ratio of the masses, say, of the supergiant and the dwarf galaxies, turns out to be larger than the ratio of their luminosities.

For instance, the supergiant elliptical galaxy NGC 4889 in the Coma cluster surpasses in luminosity the dwarf system discovered in Capricorn by Zwicky, by nearly one million times and the ratio of the masses is probably much larger.

Quite different is the picture in the case of stars where luminosity increases in proportion to a rather high power of the mass. This accounts for the fact that although the luminosity of stars can vary by hundreds of millions of times, their masses vary at most by about a thousand times and most of the stars have masses differing from the average by not more than a few times.

As a consequence, the gravity field in any stellar cluster is determined almost equally by bright as well as by faint members of the cluster. With clusters and groups of galaxies the situation is different. Here the dwarf galaxies have almost no influence upon the structure of the gravity field, which is determined chiefly by a small number of supergiants and partly by the giant galaxies.

Zwicky (1957) has rendered great service proving the monotonic increase in the number of galaxies with decrease of luminosity (the monotonic form of the luminosity function). Probably, such an increase takes place in the majority of clusters and groups of galaxies. However, even a relatively large number of low-luminosity galaxies have a negligible influence upon the structure of the field of gravitation inside and outside the cluster. Suffice it to say that in the Local Group the total mass as well as the gravity field are determined mainly by two members — M31 and our Galaxy.

This circumstance can simplify a number of problems on the dynamics of clusters of galaxies, for we can consider only a small number of its massive members.

We observe multiple galaxies in great numbers. The question of the type of configuration of these systems can be raised as was the case with multiple stars. Division of all configurations into two types seems suitable: *common configurations* and *configurations of the Trapezium type.* The latter include multiple systems in which at least three members can be found having mutual distances of the same order of magnitude. These configurations cannot be stable and they disintegrate within a period of the order of several revolutions in the system.

The observations indicate that in *real stellar systems of the Trapezium type* one of the components belongs to O or B spectral types. Such stars are of recent formation and for them the number of revolutions in the system is expected to be small. However, observations show that a few multiple stars of later spectral classes also possess Trapezium-like configurations. Of course, the configurations we observe are projections of true space configurations. Therefore, even if there are *no* real Trapezium configurations of later-type stars, a small percentage ($\sim$ 8%) of *apparent* configurations of the Trapezium type will be due to projection. This is almost precisely the percentage of Trapezium configurations observed in cases where the components of multiple stars do not belong to the O and B spectral types. In other words, there are no or almost no *real* configurations of the Trapezium type among the late-type multiple stars.

Quite the reverse is the case with multiple galaxies. As we have previously pointed out in (Ambartsumian 1956), out of the 132 multiple galaxies in Holmberg's catalogue of the double and multiple galaxies (Holmberg 1937), 87 have configurations which should, doubtless, be classed as Trapezium type. Thus, systems of galaxies of Trapezium type are markedly dominant, and most of the multiple galaxies are of recent formation; i.e., their components could have made *but a few revolutions* from the moment of formation of the system.

Two remarks are to be made in this connection: (1) The period of revolution in the multiple and double galaxies should be of the order of 10^9 years. Therefore, the age of the multiple systems we have observed could possibly be 5×10^9 years or more. In the sense of instability of multiple systems, the galaxies are probably young, although their age can in some cases be three or four times more than 5×10^9 years. (2) The instability of Trapezium configurations has not yet received a clear-cut mathematical treatment. However, simple considerations make it evident that the above-mentioned time of dissolution (several periods of revolution) holds true only for cases where the masses of all three components on which the Trapezium configuration is based are of the same order of magnitude; otherwise the system can exist considerably longer. Furthermore, the components must be of comparable luminosity. A substantial number of the observed multiple galaxies in fact meets this requirement. In particular, in such systems as Stephan's Quintet and Seyfert's Sextet the differences in stellar magnitudes are comparatively small. By contrast, Trapezium configurations where one of the components is much brighter than the others (for example, the system M31, M32 and NGC 205) are, presumably, much more stable.

On the other hand, there are cases where in a cluster of galaxies there are three or four members which are noticeably brighter than all the rest (and therefore contain the greater part of the mass) and together form a Trapezium configuration. Considering only the interaction of these brighter galaxies, it can be asserted that such systems should be unstable. For example, the four galaxies NGC 3681, 3684, 3686, and 3691 form a typical multiple system of the Trapezium type; at least a dozen other much fainter galaxies are included in this system, but the system is evidently unstable. The galaxies NGC 7383–7390 belong to a small cluster containing six bright members and more than a dozen faint components. The bright members constitute a Trapezium-type system. Finally, the three galaxies, NGC 3613, 3619 and 3625, form a small group containing at least eight fainter objects. In this case there are fainter objects of considerable angular diameters and low surface luminosity. Again we have an unstable group, although this group is not, apparently, a cluster of its own, but forms a condensation in the Ursa Major Cloud.

The arguments put forth in my previous papers speak in favor of the *joint formation* of the members of each cluster or physical group of galaxies. We refrain from repeating these arguments, but because we still encounter published assertions as to the possibility of formation of groups and clusters of galaxies from independent members of the general metagalactic field, a new argument will be advanced. This is based on the existence of systems consisting of a few very bright galaxies and a larger number of faint ones. In principle, it is possible to understand the dynamical formation of one physical pair by the accidental encounter of three galaxies. Generally speaking, this pair in the course of time can capture other galaxies too. However, the exchange of large amounts of energy must take place between the interacting galaxies, and to achieve this end the interacting galaxies must have masses of about the same order. Let us assume that a multiple system of three or more massive galaxies has come into being in this manner (although this can be proved to be highly improbable). No galaxy of an essentially small mass (say by two orders less) can ever be captured by such a group, because the exchange of kinetic energies in the case of large mass ratio is always negligibly small. Thus, the mechanism of capture meets with new difficulties in any attempt to account for the existence of galaxies of small mass in groups and clusters. This difficulty applies to all three of the above examples of multiple systems of bright galaxies with an additional number of fainter members, and also to the case of a *pair* of bright galaxies NGC 521, 533, which has a number of very faint companions.

The great difference in distribution of bright and faint members is most apparent in the large spherical clusters of galaxies. The bright members are densely concentrated, while the faint ones occur relatively more frequently on the periphery. This phenomenon was given special consideration by Zwicky, who showed that the Coma cluster is of very large dimensions if we judge from the distribution of galaxies of low luminosity. However, the case is completely different with irregular clusters. According to Reaves (1956), faint galaxies of low surface luminosity in the Virgo cluster reveal approximately the same distribution and, consequently, the same degree of concentration as bright galaxies.

Such a picture occurs also in the case of the objects of low surface brightness and small density gradient in the well-known cluster in Fornax. As Hodge points out (1959), the search for similar objects in regions neighboring the cluster has led to negative results. Finally, in the above-mentioned case of NGC 3613, 3619 and 3625, galaxies of low surface brightness and low luminosity do not spread far beyond the bounds determined by the group of bright galaxies.

These examples bear testimony to the fact that equipartition of energies between the bright and the faint members of the irregular clusters is out of the question, and that the phenomena of instability are expressed in irregular clusters much more sharply than in the spherical ones.

The existence of a great number of Trapezium configurations signifies that many of the multiple galaxies are unstable formations. If this is the case, we have no right to assume *a priori* that the multiple galaxies should be negative-energy systems. In the case of simple double stars (we exclude

O and B stars) it could be asserted without any knowledge of their orbits that, for the most part, they have negative total energies. In fact, if the majority of the multiple stars had positive total energy then the time of disintegration would be only a few tens of thousands of years, and within such a period most of the multiple stars would be replaced by stars of a new generation. In other words, positive total energy would lead to an erroneous conclusion about the rate at which stars are formed in the Galaxy.

Yet in the case of multiple *galaxies* the assumption of positive total energy for most of them does *not* lead to a similar erroneous deduction. The age of the component galaxies derived in this way is but a few times less than that accepted for our Galaxy. Therefore, we infer that the sign of the energy of the multiple galaxies, groups and clusters of galaxies should be determined in each case, relying upon observational data.

The reasons put forth above support the view that the *a priori* assumption of positive total energy in a number of systems of galaxies cannot be regarded as more audacious than, say, the assumption that almost all such systems possess negative energy. Nevertheless, let us consider the facts: The data show that for a number of multiple systems the negative total energy assumption implies that the ratio $f = M/L$ must be about one order of magnitude larger than follows from other data. Thus it is pointed out by Kalloghlian that the multiple system comprising NGC 68, 69, 71, 72 and one anonymous galaxy, if it has negative total energy, would lead to a value of f greater than 300. For the double galaxy NGC 7385–7386, he finds f greater than 600.

The sign of total energy of Stephan's Quintet has been determined by us and by G. and E. M. Burbidge (1959), resulting in positive total energy. Later, Limber and Mathews (1960) indicated that, under certain assumptions, when the mass of the components is supposed to be very high, the Quintet can have a negative total energy.

The sign of the total energy in a number of clusters of galaxies is elucidated in several recent investigations in detail. A number of difficulties are caused by uncertainty in the exact value of f for giant elliptical galaxies. This value is believed to lie in the range $30 < f < 70$. However, greater values ($f \sim 100$), although rare, are not excluded, particularly for the brightest supergiants ($M \sim -21.5$). No straightforward data exist that would enable us to estimate the value of f for these brightest supergiants. It is natural, therefore, to believe that the sign of the total energy is determined with greater reliability in those clusters and systems in which there are no supergiant elliptical galaxies. The case will be even better for systems with no giant elliptical galaxies either, and that is why the positive total energy of the nearby system of galaxies in Sculptor, as established by de Vaucouleurs (1959) is of paramount importance.

Of no lesser value is the result obtained by van den Bergh (1960) in respect to the cluster of galaxies in Canis Venatici, although studies toward determining the borders and identifying the members of this cluster should continue.

The Hercules cluster, investigated by G. and E. M. Burbidge (1959), contains but a small percentage of bright elliptical galaxies. To admit negative total energy of this cluster we have to ascribe to f_E the value of order of 300, which seems to be improbable. The contrast becomes still sharper in the case of the cluster in Virgo. Assuming the stability of this system we should have to acknowledge that $f_E > 1000$, as shown by de Vaucouleurs (1960).

It might be conjectured that perhaps the Coma cluster could have negative total energy if the modern distance scale of Sandage could be further changed by further diminution of the redshift constant. On the other hand, many members of this cluster are elliptical galaxies of moderate luminosity. The value f cannot be very high for them, so that a particularly high value of f must be ascribed to the remaining supergiant galaxies if the cluster has negative total energy.

Naturally, the sharp discrepancy between the summed luminosities of the clusters of galaxies and the masses found by applying the virial theorem has compelled some authors to favor the hypothesis of supplementary masses in the clusters which do not form part of member galaxies, i.e., intergalactic matter. Yet the data available on the upper limit of opacity in the clusters of galaxies, as well as the data on the 21-cm radiation, are not favorable.

There remains the assumption of a comparatively rich intergalactic *stellar* population in the clusters. Such a possibility has been contemplated in detail by de Vaucouleurs with reference to the Coma cluster. The result is negative if we refrain from an improbably large value of f for this intergalactic stellar population. This result apparently refers to other clusters as well.

Thus there is only one natural assumption left relating to the clusters cited above—they have positive total energies. It should be stressed that no *a priori* arguments can be advanced against this assumption.

A study of the structure of the irregular clusters of galaxies leads one to the conclusion that often they are made up of several superimposed groupings. An interesting example of such a grouping was pointed out by Markarian a few years ago: the chain of bright galaxies in the Virgo cluster containing NGC 4374, 4406, 4438 and others. This wonderful arc of eight bright galaxies is presumed to represent a physical grouping within the Virgo cluster. On the other hand, facts about the radial velocities of the members of this group undoubtedly establish its positive total energy.

Recently I looked through the abstract of van den Bergh's latest paper, in which the assumption that irregular clusters consist of separate subsystems and subclusters is made in the most general form. It is difficult to overestimate the importance of this phenomenon in comprehending the evolution of clusters of galaxies. In this case we apparently have consecutive formation of relatively independent subsystems (subclusters), the superposition of which brings about irregular clusters. It is probable that many of these subsystems have a positive total internal energy.

Of considerable interest are the results of determination of the average value of f from the differences of radial velocities in double galaxies as obtained by Page (1961), who obtained for spiral and irregular galaxies $f = 1/3$; for ellipticals and lenticulars $f = 94$. These values are derived on the assumption that in double galaxies the motion takes place in circular orbits. The value of f for spiral and irregular galaxies is even less than the value derived from rotations of single galaxies. This means that all or almost all of the observable narrow pairs of such galaxies constitute negative energy systems. Let us compare this with the unusually large values of f obtained on the basis of the virial theorem for clusters composed of spiral and irregular galaxies. The comparison leads to two inevitable conclusions:

(a) All explanations allowing for the negative total energy of clusters and groups made up of spirals and irregular galaxies become still more highly improbable, since the arguments adduced in similar cases are likewise applicable to the double galaxies.

(b) There are almost no systems with positive energy among the isolated double galaxies, for such systems can represent a *narrow* pair only for a very short duration of time (of the order of 10^8 years).

If this is so, then the double elliptical systems should also be regarded as systems possessing, for the most part, negative energies, and Page's value $f = 94$ can be taken as close to the real value. This brings us nearer to the conclusion about negative total energy of the Coma cluster.

It is interesting to note that when we pass from double galaxies to multiple configurations of the Trapezium type, the differences of the velocities of the components become much greater. For these configurations the assumption of negative total energy leads to too large values of f_E.

Speaking of instability in systems of galaxies, we should also touch upon the *radio galaxies* which, as a rule, occur in clusters of galaxies. Apparently the radio galaxies are always among the few brightest members of the corresponding clusters. The best example is furnished by the source Perseus A (NGC 1275), which is the brightest member of the cluster in Perseus.

In the radio galaxy NGC 4486, a jet is being ejected from the central nucleus with separate condensations, the luminosities of which resemble those of dwarf galaxies. Apparently these condensations contain an enormous number of relativistic electrons. However, it is difficult to refute the suggestion that these condensations also include a substantial amount of common matter in addition to the relativistic plasma. In particular they are likely to contain *sources* of relativistic electrons.

A strong argument in support of this view is provided by the galaxies NGC 3651 and IC 1182 the nuclei of which eject jets containing blue condensations. These galaxies with blue jets are also among the brightest members in the corresponding clusters. Finally, there are cases when blue components occur in the vicinity of other giant elliptical galaxies, which evidently represent a later stage in the evolution of the above blue condensations.

In all probability, the condensations in NGC 4486 represent an earlier stage of evolution of the same objects. In such cases the intensity of radio emission of the blue condensations and the blue companions is believed to have already weakened.

As revealed in Byurakan, a blue object of photographic magnitude $18^m.5$ is situated very close to the remote radio galaxy in Hydra. The color index of this object is about $-0^m.5$. Since its diameter is less than 2000 parsecs on our plates it looks like a star. If it can be shown that this object is in fact a physical companion of Hydra A, then a close connection between the two types of eruptive activities of the nuclei of supergiant galaxies will be established: ejection of plasma condensations and ejection of blue condensations. One way or another, all the data indicate that this activity is of great import in the origin of galaxies.

We conclude that there exist clusters which are in a particularly active phase of evolution when new galaxies originate within them. The existence of a radio galaxy is an indicator of this phase. It is possible that in such a phase the radio emission erupts now and then with varying intensity.

In the radio galaxy Perseus A it is well known that large relative velocities are observed, up to 3000 km/sec. Such velocities exceed the velocity of escape from the cluster and thus speak for themselves of instability.

It seems, therefore, that a study of the radio galaxies as systems from the nuclei of which large masses are ejected or which are in the process of division must throw new light upon the phenomena of instability in the clusters of galaxies.

REFERENCES

1. Ambartsumian, V. A., 1956, *Izvest. Akad. Nauk Armyan. SSR, Fis. Mat., Estestven. i, Tekn. Nauki*, SSR, 9, 23.
2. Burbidge, E. M. and Burbidge, G., 1959, *Astrophys. J.*, 130, 15.
3. —1959, *ibid.*, 130, 629.
4. de Vaucouleurs, G., 1959 *ibid.*, 130, 718.
5. —1960, *ibid.*, 131, 585.
6. Hodge, P., 1959 *Publs. Astron. Soc. Pacific*, 71, 28.
7. Holmberg, E., 1937 *Ann. Obs. Lund*, No. 6.
8. Limber, D. and Mathews, W., 1960, *Astrophys. J.*, 132, 286.
9. Page, T., 1961, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics, and Probability* (to be published).
10. Reaves, G., 1956, *Astrophys. J.*, 61, 69.

11. van den Bergh, S., 1960, *ibid.*, 131, 558.
12. Zwicky, F., 1957, *Morphological Astronomy*, (Springer-Verlag, Berlin), p. 220.

Byurakan
Astrophysical Observatory,
Academy of Sciences of Armenia.

ON THE NUCLEI OF GALAXIES AND THEIR ACTIVITY[1]

Introductory remarks

The present paper is my second report at the Solvay Conferences. The previous report of 1958 concerned the eruptive activity of the nuclei of galaxies, the ejection of large masses from those nuclei and other processes connected with the quick release of large quantities of energy in nuclei. The present communication will be devoted to the same subject: activity of nuclei.

The position of the speaker on this topic in 1958 was much more difficult than it is at present. At that time one had to argue, contrary to general opinion, that radio galaxies do not result from the collisions of pairs of galaxies but constitute stellar systems, in the nuclei of which giant explosions had taken place with the formation of large clouds of relativistic electrons. At that time only indirect evidence existed concerning ejections of large masses of *conventional matter*[2] from the nuclei of galaxies. But now the recent fine work of Sandage and Lynds relating to M 82 galaxy has left no doubt on that issue.

Now the available information on the galaxies of different morphological and physical types is much richer. It paves the way for a disclosure of the nature of these basic formations of the Universe.

As in 1958, I will try again to proceed not from preconceived notions, but to rely on observational data. Preconceived notions often hinder the reaching of right conclusions. In my first Solvay report, as well as in my 1961 Berkeley Invited Discourse, data were demonstrated evidencing the active and perhaps even basic role of the nuclei in the evolution of galaxies. Disregarding these facts, we still have attempts to explain the unusual

---

[1]Originally published by Interscience Publishers, a division of John Wiley & Sons, Ltd. © 1965 in *Proceedings of the Thirteenth Conference on Physics at the University of Brussels, September 1964: The Structure and Evolution of Galaxies*, 1965, pp. 1-14. Reproduced by permission of John Wiley & Sons Limited.

[2]By "conventional matter," Ambartsumian means "nonrelativistic plasma."

phenomena that we observe in the nuclei in terms of concepts of gradual concentration of the surrounding matter toward the nucleus. I think that the earlier we give up this idea, the sooner we arrive at the true explanations.

While new observations increasingly point to the outflow of matter from the center, to explosions, jet and ejections, some theoreticians speak in favor of condensation, implosions and collapses. In the meantime, no convincing and regular facts have been adduced to support condensation of large masses of the surrounding matter towards the nucleus of the galaxy.

Contrary to this, the problem of impact of the observed large-scale explosions and ejections from the nuclei on the life of the surrounding galaxy has not, so far, properly deserved attention of theoreticians.

It seems to us that an astronomer dedicated to the study and analysis of facts has to concentrate on two objectives:

(1) The study of the nature of nuclei and processes going on there;

(2) The influence of those processes on the evolution of the galaxy as a whole.

As to the theoretical explanation of the unusual phenomena occurring in the nuclei, we can think about two stages. The right interpretation of observations marks the first stage. When observational data are scanty, it is essential, first of all, to form a clear idea about the physical nature of the observed phenomenon. Next comes the second stage: after forming a general idea of what is going on, we try to find out the cause of the phenomenon.

Unfortunately, there is at times a tendency to skip the first stage. Particularly, such haste can be noticed in the problem of explosive processes in the nuclei of galaxies. I think, however, that at present we have to concentrate mainly on the first stage of the work leaving the explanations of the observed phenomena for the future.

Forms of activity of nuclei

Observations show that the nuclei of galaxies are not isolated systems. In addition to radiation they also emit conventional matter into the surrounding space. This process may occur in various ways, and there is reason to speak of the various forms or types of activity of nuclei. We list below some observed forms of activity including those that might be regarded as controversial.

(a) The quiet outflow of conventional gaseous matter from the area of the nucleus at the rate of tens or hundreds of kilometers per second. The best illustration of such an outflow is M 31 at λ3727. Similar outflow occurs in our own Galaxy and in the Small Magellanic Cloud.

(b) Continuous outflow of relativistic particles or other agents which produce high-energy electrons resulting in radio halos around the nucleus in the meter and decimeter wavelengths. Such a phenomenon is seen around the nucleus of our Galaxy. According to Mathewson and Rome, the radio frequency radiation in Sc galaxies in the decimeter range is concentrated in the area around the nucleus, and the diameter of the radio image is several times less than the diameter of the optical image (NGC 253, 4945, 5236 and also Sb galaxy 1068).

(c) Eruptive ejection of gaseous matter (examples: M 82 and possibly NGC 2685). The same phenomena are probably present in the radio galaxy NGC 1275 as well, where a gaseous cloud is observed moving at the speed of 3000 km/sec from the center of the galaxy.

(d) The eruptive ejection of dense relativistic plasma. Examples: NGC 4486, 5128 and many other radio galaxies.

(e) The ejection of rather dense blue concentrations having m_{pg} ranging from $-14^m.0$ to $-17^m.0$. These concentrations may be taken to be newly-born galaxies. Examples: NGC 3561 and IC 1182. The possible cases of the division of nuclei into two or more comparable components with a subsequent formation of double or multiple galaxies may also be classed with these phenomena.

(f) The outflow of matter from which the spiral arms are subsequently formed (a hypothetical form of activity).

(g) The ejection of the matter of bars in SB galaxies (hypothetical form of activity).

(h) The ejection of matter from which the stellar population of spherical subsystems is formed (hypothetical form).

It is quite possible that some of those processes represent different aspects of the same active process. Take, for instance, radio galaxy Hydra A in the immediate vicinity of which a very interesting blue object can be observed. It is rather likely that ejections of the radio-emitting cloud and of the blue object have taken place simultaneously. Though these phenomena are interrelated, it is also possible that they have occurred in some succession. All these forms of activity may also be supplemented by explosions that lead to the formation of quasi-stellar sources of the 3C 273 type. These phenomena exceed other forms of activity in scale. Such explosions may possibly mark the birth of a new galaxy or even of a whole cluster of galaxies.

On the nature of nuclei

From the point of view of the power of nuclei the observed galaxies may be grouped into five classes:

(1) Galaxies without any noticeable nuclei with no considerable condensation in the center. Many irregular galaxies belong to this class. The elliptic dwarf galaxies of the Sculptor type should also be included here.

(2) Galaxies having quiet nuclei of relatively low luminosity. This class may include cases where the nucleus is more than four magnitudes fainter than the integral luminosity of the galaxy. M 31, NGC 5194, M 33 and, possibly, our Galaxy fall into this class.

(3) Galaxies with quiet nuclei of high luminosity when the nucleus is fainter than the whole galaxy by 1.5–4 magnitudes. The spectra of nuclei in classes 2 and 3 are continuous. Emission lines 3727 and others may be present. Although these lines sometimes may attain a considerable degree of intensity, they show neither marked broadening nor division into components. Examples: NGC 4303, NGC 3162.

(4) The Seyfert galaxies with bright nuclei comprising a considerable portion of the luminosity of the entire galaxy. Numerous emission lines are present. The latter show either widening or splitting produced by the great speed of motion of the gaseous clouds inherent in the nucleus.

(5) Compact galaxies which may comprise the starlike radio galaxies as well as many other compact objects detected by optical means (Zwicky). In this case we can assume that the luminosity as a whole is concentrated in the nucleus of the galaxy.

Nuclei of class 2 are of small dimensions. Their diameters are several parsec or several tens of parsec. In classes 3, 4 and 5 we have nuclei of larger dimensions with diameters measured in hundreds of parsec. For instance, the nucleus of the galaxy of the SBb type NGC 3504 has a diameter of the order of 10 parsec, with some intensification of brightness toward the center. Other nuclei often show a more regular distribution of energy over the disk. However, this intricate problem of distribution of brightness over the disks of the nuclei requires higher resolving power of telescopes and remains completely unstudied.

The continuous spectrum of the nuclei of the galaxies of classes 2 and 3 indicates that the source of luminosity is the stellar population which differs but little from the stellar population of the central regions of such galaxies as M 31 and M 81. But the gaseous component is already present in these nuclei. Data relating to the lines $\lambda 3727$ in the area of the nucleus M 31 point to a comparatively quiet and continuous outflow of matter from such nuclei. Although the yield is poor, still a mass up to 10^8 $M_\odot$ may flow out over a long period of time. Hence the question of sources of the gaseous outflows arises.

We stress that the nuclei of Seyfert type (class 4) contain, apart from the stellar, a gas component too. The isolated discrete clouds of the latter escape from the nucleus at the speed of thousands of kilometers per second. Such great velocities leave no room for doubt that those discrete gas clouds were born within the nucleus. This invariably leads us to the conclusion that they were ejected from denser bodies, only a few tens of thousands of years ago. This means that such nuclei contain bodies that, at the present stage of evolution of nuclei, manifest a tremendous eruptive activity. Therefore, the Seyfert type nuclei of galaxies should be properly called excited nuclei. At the same time there is no reason to believe that the clouds have been ejected by the members of a common stellar population of the nucleus, particularly because the masses of certain clouds should be of the order of hundreds of $M_\odot$ and more. We inevitably come to the conclusion that such a nucleus contains one or more supermassive nonstellar bodies which eject the gaseous clouds.

As for the class 5 of compact objects, at least part of them contain supermassive bodies of nonstellar nature. Of course, we mean quasi-stellar radio galaxies. It is essential, however, that most of the radiation comes in this case directly from such a body. Judging by the spectral distribution of energy, the radiation that reaches us is nonthermal and is characterized by an ultraviolet excess.

The presence of an ultraviolet excess is also typical for the nuclei of the majority of the Seyfert galaxies (class 4). Moreover, Markarian has shown that many galaxies which should be classified in categories 2 and 3 also have an ultraviolet excess presumably of nonthermal origin. All this gives full reason to believe that nonstellar bodies exist also in the nuclei of these categories of galaxies although direct indications for this are by far less prominent than in categories 4 and 5. Particularly, the luminosity of the supermassive bodies in the visible part of the spectrum is faint as compared to the luminosity of the stellar component. The outflow of the gases is less powerful and is of a more quiet nature.

Therefore, an analysis of observational data brings us to the following conclusion: every nucleus contains a supermassive body which may be either in the state of eruption (quasi-stellar galaxies) or in an excited, active state (the Seyfert galaxies), or still in a state of weak activity (galaxies 2 and 3).

This signifies that the nucleus is made up of three components: *stellar population, gas and a supermassive body.* Dynamically, the nucleus evolves independently of the rest of the galaxy.

On the nature of the relationship between the nucleus and the galaxy

The assumption that every nucleus contains, as a rule, a supermassive nonstellar body is in full harmony with the view expressed in our report in Berkeley, according to which the nucleus plays an essential if not a dominant role in the evolution of each galaxy. Actually, there is no more arguing the idea that the origin and evolution of at least some of the subsystems forming the galaxy are due to the nucleus itself (for instance, the subsystem consisting of relativistic plasma so prominent in radio galaxies). The case of the M 82-type galaxies shows that in the evolution of the common (nonrelativistic) gas component, the nucleus can play a decisive part. However, assumptions made in the Berkeley report that both the spiral arms and the second type population originate from the matter ejected out of the nucleus remain unproved.

Two extreme points of view seem possible:

(1) Development of a nucleus is conditioned by the evolution of the galaxy itself. The subsequent evolution of the outer parts of the galaxy is practically independent of the nucleus.

(2) The formation of the various components of the galaxy is conditioned by the activity of the nucleus. The subsystems of stars once formed evolve henceforth loosely, depending on the nucleus as well as on other subsystems according to the laws of stellar dynamics.

Now the question arises: what would be the expected relation between the parameters characterizing the nucleus and the galaxy.

If the first hypothesis is true, the state of the galaxy should account for the state of the nucleus. In the case of the second hypothesis, the state of the nucleus should be, to a certain extent, independent of the state of the galaxy. To be more precise, in the latter case the state of a galaxy is to be explained in terms of the entire activity of the nucleus over the preceding period, that is the whole history of the nucleus. This means that the state of the galaxy should correlate with the present state of the nucleus, to the extent its history may be judged by the given state of the nucleus.

Our information concerning the nuclei is always very scanty. Nevertheless, in a number of cases where nearby galaxies contain relatively bright nuclei, we can roughly estimate some of their integral parameters especially luminosity and color index. Evaluation of the diameters of the nuclei is rarely possible. Therefore, we should look for a correlation of the state of the galaxies with values of those two nuclear integral parameters alone. But the values of these two parameters may not completely determine the whole history of the nucleus. We can expect no correlation between the state of the galaxies and the above-mentioned integral parameters of the nuclei.

During the past year, several hundred pictures of galaxies were taken in the Byurakan Observatory, with the aim of determining the characteristics of their nuclei. A scale has been applied to estimate the degree of prominence of a nucleus; it is now explained in Table 1.

In classes 3, 4, and 5 we regard the existence of a nucleus as definite, but photometric evaluations are possible only in 4 and 5. For the classes in the upper rows, it is possible to estimate only the upper limit of the luminosity of the nucleus that forms a part of the observed central condensation. The importance of the nucleus in galaxies of SB type is not very closely correlated with the morphological subtype of the galaxy. For instance, such subtypes as SBa, SBb often belong to classes 4 or 5, while for subtypes SBO and SBc this almost never occurs. As a rule, the SBc galaxies are apparently deprived of any bright nuclei.

Table 1
Prominence of Nuclei on the images of Galaxies

Class	Pattern	Interpretation
1	No appreciable condensation at the center	No nucleus present
2	Weak condensation at the center	Probably a nucleus exists
3	Strong concentration at the center, but no starlike image	A nucleus definitely exists but cannot be distinguished from the background
4	Starlike nuclear image at short exposures, but nebulous at long exposures	A nucleus is seen surounded by the dense part of the background
5	Starlike nuclear image even at moderate exposures	A bright nucleus clearly distinguished against the background

We have tried to detect a correlation between the absolute integral magnitude of the nucleus and the absolute magnitude of the galaxy for the entire group of galaxies SB and for galaxies Sc with nuclei belonging to classes 4 and 5 of Table 1. No significant correlation has been found in either case. This testifies to the relative independence of the state of the nucleus from the parameters that characterize the galaxy. The independence of the state of the nucleus from the luminosity of the galaxy is something which deserves particular mention. On the other hand, we have seen above that an explicit correlation of the nuclei with the morphological subtype exists in the SB class. Finally, in the case of giant elliptical galaxies, nuclei with low luminosity predominate, providing an example of a closer correlation. On the contrary, we can find nuclei of different luminosities, or even no nucleus at all in the elliptical galaxies of low luminosity (cf. M 32, NGC 205, 185, 147). The presence of close or loose correlation as conditioned by the class of the galaxies in question strongly supports the second hypothesis.

Let us assume that the giant galaxies begin their lives as elliptical systems in which the nuclei are also young and do not yet possess a significant stellar population. The greater the activity of the nuclei, the brighter their luminosity. At the same time, new subsystems are formed in the galaxy. Therefore, within the galaxies Sa, Sb, SBa and SBb nuclei of high luminosity are probable. Finally, galaxies Sc, SBc and those irregulars that contain population I (Magellanic Clouds and others) seem to be the oldest systems. Nuclei of high luminosity are rarely encountered in galaxies of the Sc type, while no nuclei are seen in SBc's and irregulars. At the last stage we have reduction of luminosity and disappearance of the nuclei.

Most astronomers engaged in the study of the evolution of galaxies proceed from the opposite end and consider the objects of the Magellanic Clouds type the youngest. They argue that the latter systems contain numerous young stars of high luminosity. It seems to me that those who hold this view ignore the fact that one should not confuse the youth of a galaxy with the youth of a certain part of its population. We know cities with histories dating back thousands of years, but the average age of their inhabitants is young. On the other hand think about a modern health resort with a population of patients of advanced age. In the course of time some industry may develop in this modern health resort and may attract a great number of young inhabitants.

Of course, this is only a rough comparison: I do not have much trust in the concept of formation of several consecutive generations of stars out of the same matter.

Thus, our starting point is the assumption that at its initial stage of development the stellar population is something like what we conventionally call population II, a young variety of population II. The formation of population I should be attributed to later stages when spiral arms form out of the matter ejected from the nuclei.

The initial stages of the evolution of galaxies

Now the question comes up as to whether there are galaxies consisting of population II with relatively direct evidence of their being young. At the 1958 Solvay Conference we mentioned that galaxy M 82 from the group M 81 displays a velocity probably greater than the escape velocity from the gravity center of this group. It naturally follows that the age of the galaxy must be of the order of 10^8 years (or 2×10^8 years). The well-known work of Sandage and Lynds drew attention to the galaxies of this type and I will dwell on this point in more detail.

De Vaucouleurs' list, comprising the new classification of 1500 bright galaxies, includes 12 objects of the M 82 type. Of these one galaxy has a southernmost position, and we do not have its photograph at our disposal. Instead, I have added galaxy NGC 520 which, no doubt, is of the same type. Three of these twelve galaxies, NGC 972, 3955 and 4753, are isolated objects. There is no galaxy of comparable luminosity or diameter in their surroundings. As to galaxies NGC 972 and NGC 4753, the radial velocities of which are known, it can be stated with confidence that no galaxy occurs over a surrounding area closer than 500,000 parsec in diameter that would be fainter than the one under consideration by four magnitudes or less. They seem to be, in fact, isolated objects of a fairly high absolute magnitude ($-20^m.0$).

Two of the nine nonisolated objects belong to double systems (NGC 5195 and NGC 3448). In both cases the other component is an Sc galaxy, the spiral arm of which stretches out to the object itself. Seven galaxies of the M 82 type enter into poor groups made up of four or five objects, apart from possible objects of very low luminosity. M 82, and NGC 3077 from M 81 group may serve as examples. These features are so outstanding that they may serve as a touchstone for hypotheses accounting for the origin of these galaxies.

We may suppose that at the earliest stage of evolution a galaxy has very low luminosity and an active nucleus. In the course of time the luminosity increases. If the groups of young galaxies have positive total energies, we will have younger objects among the groups of smaller linear size. Considering the M 82-type galaxies as such young objects, these galaxies will have lower luminosity in groups of smaller size, and higher luminosity in groups of larger linear size. After dispersion of the group we will have isolated M 82 galaxies of the highest possible luminosities. But this is what we actually observe. All M 82 objects of low luminosity enter into compact groups. Of the three galaxies with large luminosities, one (the NGC 520) is a member of a group very large in dimensions, and the two others are isolated objects.

On the nonthermal radiation of the nuclei

It is well known that objects of the 3C 273 type possess a spectrum that sharply deviates from the Planck curve. Apparently the distribution of energy in the spectrum of these objects can be better explained in terms of synchrotron radiation. Yet, one can believe that many other nuclei also exhibit a nonthermal component in their radiation. According to Minkowski, the nucleus of the radio galaxy NGC 6166 is particularly distinct in the ultraviolet. On the basis of an analysis of the colors of the central parts of the galaxies, in which the morphological features and the spectrum are incongruous, the conclusion is made (Markarian) that there is a blue excess of radiation of the nuclei of such galaxies. Using photographs of the blue jet from the galaxy NGC 3561, Zwicky has shown that the continuous spectrum extends far into the ultraviolet. In all these cases we can hardly expect to have blue stars in any considerable number in the nuclei of the galaxies. Therefore, the ultraviolet or blue excess should be ascribed to nonthermal radiation.

As established at Byurakan, a number of the SB-type galaxies with high luminosity nuclei are substantially redder than their nuclei. Sometimes the color index of the nuclei is +0.2. Galaxy NGC 3504 is an example. All this makes us believe that nonthermal radiation in the nuclei is a relatively frequent phenomenon. On the other hand, the occurrence of nonthermal radiation speaks for the activity of nonstellar bodies included in the nucleus.

Although it is difficult at present to judge the nature of the nonthermal radiation manifesting mainly in the form of an ultraviolet excess, I would like to make two remarks.

(1) We do not assume that this nonthermal radiation proceeds directly from a nonstellar massive body. On the contrary, it probably emanates directly from the diffuse matter within the nucleus. However, the source

of energy radiated by the diffuse matter can be the nonstellar body. This energy may be transferred to the diffuse matter, say, through the high energy particles, or, as in the case of the mechanism of relativistic electrons, it may radiate directly from those particles.

(2) The occurrence of a powerful nonthermal excess in the far ultraviolet may sometimes lead to the appearance of emission lines connected with fluorescence. Our astronomers have called attention to this fact in connection with the occurrence of H emission in a number of areas of M 82. One can imagine even more striking consequences of this phenomenon.

Conclusion

In conclusion I would like to say a few words about the theoretical explanation of the unusual phenomena that are connected with the nuclei of the galaxies. It is clear that very rapid transformations of energy play a substantial role in this case. Such rapid processes of transformation and release of energy originate in systems that are characterized by the instability of possible states. The relativistic theory of gravitation seems, from this point of view, to be most suitable. The first attempt at drawing up relativistic models that contain such local explosions was made by Novikov, who works with Prof. Zeldovitch.

Such models are, of course, very useful and they deserve more detailed study. But since work on interpreting the observed phenomena is not yet complete, it is hard to draw a line of comparison between the various models constructed and the reality.

Academy of Sciences,
Armenian S.S.R., Yerevan

PROBLEMS OF EXTRAGALACTIC RESEARCH[1]

Invited Discourse C given to participants in the General Assembly at 20^h00^m on Monday, August 21, 1962 in the Auditorium of Wheeler Hall on the Berkeley Campus of the University of California

Introduction

The present report deals with the basic facts of extragalactic astronomy. It should be noted that a true picture of the outer stellar systems (galaxies) was formed in astronomy only about forty years ago. Most basic problems concerning the outer galaxies remain unresolved. In the present report we formulate a number of problems concerning the outer galaxies. We shall try to keep as close to the facts as possible and handle, primarily, those problems the solution of which seems possible with the means available to us.

It is known that extragalactic astronomy comes in close contact with cosmology, that is with the theories attempting to describe the Universe as a whole. These theories are, undoubtedly, of definite usefulness insofar as solutions of the equations of Einstein's theory of gravitation are considered, and the question is raised as to the applicability of these solutions to the observed part of the Universe. At the same time some of these theories contain rough simplifications and unbounded extrapolations. It is beyond the scope of the present report to make an analysis of these theories and discuss the problems of their further development, although a critical review of the work carried out in this field should be extremely valuable. Nonetheless, the facts and the problems discussed below should be of importance to cosmological theories as well.

---

[1]Originally published by Academic Press, Inc. © 1962 in *Transactions of the International Astronomical Union*, Vol. XIB, 1962, pp. 145–160. Used here with the permission of Academic Press, Inc.

I. The Principal Facts on the Distribution of Matter in Extragalactic Space

The majority of the matter observed in space is concentrated in the stars. This is one of the properties of the surrounding world. Other bodies contain but a small part of the mass observed.

In the overwhelming majority of cases the observed stars are members of giant stellar groups called galaxies. This is the most significant fact of extragalactic astronomy.

The stellar population and the dimensions of the galaxies vary over an unusually wide range. The supergiant galaxies of the type of the two brightest galaxies at the center of the Coma cluster (NGC 4874 and NGC 4889) attain a photographic absolute magnitude of the order of -22. They contain hundreds of thousands of millions of stars, while the dwarf systems of the type of the nearby galaxy in Sculptor are of absolute magnitude $-11{\cdot}0$ and apparently contain only several million stars. However, stellar systems of much lower luminosity exist and might be called subdwarf galaxies. The galaxy in Capricorn of absolute magnitude $-6{\cdot}5$, discovered by Zwicky, is an illustration in point. This group seemingly comprises, at most, several tens of thousands of stars and is more than ten million times poorer than some supergiant galaxies. Moreover, it is inferior in population to many globular clusters.

The diameters of the galaxies, as a rule, are included in the range 50,000 pc for supergiants to 500 pc for subdwarfs.

The giant and the supergiant galaxies with diameters ranging from 5000 pc to 50,000 pc invariably possess high surface brightness (over 24.0 per square second of arc) and also show an extensive concentration of luminosity towards the center.

In dwarf galaxies objects with low surface brightness are encountered alongside those with high surface brightness. There exist dwarf galaxies with high gradient of surface brightness from the border to the center as well as dwarf galaxies with very low gradient. The latter systems look in pictures almost like uniform disks.

The dwarf stellar systems in Sculptor and Fornax discovered by Shapley are examples of galaxies with low density gradient in the local group. The surface luminosity of these systems is exceedingly low. Later, Baade showed that the galaxies NGC 147 and NGC 185, belonging to the Local Group, also contain a low density gradient. The surface luminosity of these two galaxies is considerably higher than that of the systems in Sculptor and Fornax. Of intermediate surface brightness are two other members of the Local Group: Sextans B ($9^h57^m.3$, + 5°34′; 1950·0) and Leo 2($10^h05^m.8$, + 12°33′; 1950·0). They also have a very low density gradient. Many objects of low surface brightness and low density gradient are found within the Virgo cluster. Their linear dimensions sometimes are of the same order as those of the galaxies of average luminosity. For instance, galaxy IC 3475 has very low surface brightness and a negligible density gradient while its diameter is about 5000 pc. Thus, this galaxy is by far superior in dimension to similar objects in the Local Group.

However, it is worthwhile to note that objects of relatively large dimensions with a small density gradient and low surface brightness are very rare. For instance, in the well-known cluster in Cancer, the largest of such galaxies has a linear diameter of about 2500 pc.

The fact that the absolute majority of stars belong to the galaxies becomes of primary importance if we take into account that the galaxies are, roughly speaking, isolated systems. Usually the distance between neighboring galaxies exceeds by far the diameters of their central, denser parts. At the same time the external, rarified parts of galaxies often overlap. Besides this spatial isolation, we must also indicate the mutual dynamical independence of the galaxies as stellar systems. The term "dynamical independence" implies the property under which the motion of the stars in each galaxy is determined, in the main, by its interaction with the other members of the same galaxy. It should be pointed out that this condition of dynamical independence holds usually only to a certain approximation. The cases of mutual perturbations of neighboring stellar systems, the ejections from the central parts of the galaxies, i.e., cases of violation of this independence, will be properly considered below. Galaxies form such systems as clusters, groups and multiple clusters.

Two decades ago it was assumed that apart from clusters and the groups of galaxies there exists a general field comprising the greater part of the galaxies — similar to the general star field in our stellar system which is overspread with stellar clusters and associations. At present the existence of the general field itself is in doubt. At any rate, it can be asserted that the clusters, the groups and the multiple systems contain the overwhelming majority of the galaxies of high luminosity.

The clusters we observe can be classed under two types: spherical clusters with a regular, symmetrical distribution of the galaxies about the center; and loose clusters with an irregular distribution of galaxies. The giant and the supergiant members of the spherical clusters comprise chiefly elliptical galaxies. The irregular clusters contain a high percentage of spirals. Groups of galaxies such as the Local Group or the groups around M 101 and M 81 are, in this respect, very much like the irregular clusters.

For instance, the groups of galaxies associated with M 101 and M 81 virtually contain no elliptical galaxies and are made up only of spirals and irregular galaxies. A group in Sculptor, studied by de Vaucouleurs, contains only galaxies of the Sc type and some irregulars as well. Our Local Group does not comprise elliptical galaxies of high luminosity either, but it possesses elliptical galaxies of low and moderate luminosity.

It is interesting to note that our Local Group in fact consists of two smaller groups approximately of the size of multiple galaxies. The first group contains our Galaxy, two Magellanic clouds and, apparently, some galaxies of the Sculptor type. The second group comprises the Andromeda nebula with its four satellites and M 33. However, such a division can be ascertained only for galaxies of high and moderate luminosity.

Referring to the dwarf galaxies, they might possibly fill in the space of the Local Group. The full mass of the whole Local Group is determined mainly by two galaxies which form actually the centers of these subgroups, that is M 31 and our own Galaxy. On the other hand, rich clusters of galaxies occur sometimes in twos and in threes, forming multiple clusters.

It has been mentioned above that the galaxies are isolated stellar systems. However, there are cases where this isolation is not observed strictly. Let us indicate the following three categories of such objects:

(a) *Interacting galaxies.* Two galaxies are close to each other and the presence of one deeply affects the structure of the other. Numerous instances of interacting galaxies are reproduced in the atlas of Vorontsov-Velyaminov. Two interpretations are possible: (i) tidal interactions; and

(ii) the division of one galaxy into two. In the latter case the "interaction" observed should be regarded as the aftermath of the process of division.

(b) *Couples of galaxies* interconnected by means of bridges and filaments. Many illustrations of this kind are cited in the articles of Zwicky, who has shown that the filaments are made up of stars. In addition, there are jets flowing out the central areas of some spherical galaxies. Such jets sometimes contain blue condensations which are known to be the dwarf galaxies. (The high luminosity galaxies NGC 3561 and IC 1182 possess jets containing condensations.)

It turns out that in this case, too, the jet joining the dwarf galaxy to the large one is a kind of filament. No doubt remains that the dwarf galaxy has been detached from the main body of the principal galaxy. It seems more plausible, therefore, to regard the bridges and the filaments as resulting from the process of formation of two galaxies from one.

(c) *Radio galaxies.* The radio galaxies were assumed to be cases of accidental collision of independent stellar systems. It was also assumed that the source of the energy of radio emission is the kinetic energy of gas masses contained in the colliding galaxies. The facts, however, contradict this hypothesis. All the data support the view that the radio galaxies form a stage, possibly of short duration, in the process of the internal evolution of the galaxies of very high luminosity.

Evidently, the radio-emitting activity of the galaxies is closely associated with the new formations within them, such as jets and ejections from the center (Virgo A), spiral arms and even complete new galaxies. In other words, in some cases a process of division of the main body of a galaxy and the emergence from it of a new one takes place. That is why radio galaxies often become very narrow pairs, consisting of the old galaxy and the new formation immersed in the old galaxy.

We note that all cases of violation of the isolation of galaxies taken together constitute only a small percentage of the total number of galaxies. There is every reason to believe that these violations occur only at a certain stage of the evolution of the galaxies, namely when new galaxies come into being.

Many important problems concerning space distribution of the galaxies are still unsolved. Here is one of them: do clusters of galaxies form, in their turn, systems of higher order of the type of superclusters or supergalaxies.

Our Local Group enters, no doubt, into a group of clusters with the great cluster of Virgo forming its main part. This great system was named the Supergalaxy by de Vaucouleurs; its dimensions are of the order of 20 million pc. However, nothing can be said so far as to the dynamical unity of this system or the existence of forces that might support such a unity.

As regards the problem of supergalaxies, one of two possibilities must eventually be realized: (i) the mutual distances between neighboring supergalaxies are great compared with the linear dimensions of the supergalaxies; or (ii) these mutual distances are of the same order of magnitude as the linear dimensions of the supergalaxies. In the former case a great many isolated formations must be clearly traced in projection on the sky. In the latter case we could trace out only a small number of isolated formations, and unless we make a deep study of the problem it would be hard to draw conclusions regarding the occurrence of distant supergalaxies.

The observations reveal irregularities in the distribution of clusters and groups of galaxies, which can be accounted for, to a certain extent, by the occurrence of supergalaxies. At the same time we observe only a few isolated cases of clouds consisting of a number of concentrations; for example, in the southern sky a vast cloud, extending from $l = 160°$ to $240°$ when $b = -40°$.

The irregularities in the distribution of the galaxies in the sky (independent of the absorption in our Galaxy) are clearly shown by the galaxies included in the catalogue of Shapley and Ames (limiting magnitude 13·0). This irregularity is stipulated mainly by the local Supergalaxy. The irregularities become more pronounced in the counts of Shain and Virtanen (limiting magnitude 18·4). The small-scale inhomogeneities are conditioned by the concentrations of the galaxies in clusters. But there are also inhomogeneities of greater dimensions brought about by the tendencies of the clusters to form groups similar to the supergalaxies discussed above.

According to Zwicky and other authors, the irregularities in the distribution of the galaxies extend up to the boundaries observable by the Palomar Schmidt telescope (almost to the 20th magnitude).

The huge clouds of galaxies in the region of the cluster in Corona Borealis are an illustration in point. However, the distribution of the cluster centers is of great interest when studying the problem of clustering of the clusters. Abel's study on the plates of the Palomar Atlas confirms the inhomogeneity in the distribution of the clusters.

Zwicky believes that the main cause of the inhomogeneity observed in the distribution of the clusters lies in the patchy distribution of the intergalactic dust. His arguments in favor of intergalactic absorptions in some directions are apparently convincing. But not all the deviations from homogeneity are accountable in this way. The real irregularity in the distribution of the most distant galaxies is something that should be properly taken into account.

These two facts show that the distances between the supergalaxies are of the same order as their diameters. Nevertheless, the existence of independent supergalaxies should be accepted. The following questions remain open:

(a) *What percentage of clusters of galaxies enters into the systems of higher order?* Is there an equally strong tendency towards clustering of the clusters in the two familiar types of clusters (spherical and diffuse)? Only more detailed photometric and statistical studies could provide answers to these questions.

(b) *To what extent do galaxies of low luminosity repeat the space distribution of those of high luminosity?* The concentration of galaxies in clusters is rather well established with respect to objects of high luminosity. However, beginning at a distance of several million parsecs, objects of low luminosity will be lost altogether among galaxies of remote background. Yet some conclusions as to the distribution of objects of low luminosity, especially galaxies of low surface brightness, can be made from the results of the work of Rijves. He established that the distribution of objects with low surface brightness in the Virgo cluster repeats roughly the distribution of galaxies with high luminosity. But we cannot assert whether the galaxies of exceedingly low surface brightness (systems of the Sculptor type or of the Zwicky object in Capricorn) make up one common metagalactic field or concentrate in clusters and groups.

(c) *The supergalaxies referred to above are objects with a diameter of the order of* 20 *million parsecs.* If they constitute the greatest inhomogeneities in the distribution of the galaxies, then space cells, 50 to 100 million parsecs in dimensions, are to be expected to contain approximately equal numbers of galaxies.

It is possible, however, that inhomogeneities of a larger scale occur. The question can be settled only by investigating the distribution of the faintest clusters of galaxies (up to $m = 21$), or by studying the distribution of extragalactic radio sources. The solution of this problem is extremely important for cosmological theories. For the time being there is no evidence to justify the usual postulate of homogeneity.

(d) *There are sound proofs in support of intergalactic dust.* In this connection, the need to study intergalactic matter should be stressed. The following kinds of intergalactic matter seem to be established as real:

(i) Bright intergalactic matter, sometimes filling up the central part of the volume occupied by a cluster of galaxies. All the data testify that the bright matter, as well as the frequently observable bridges and filaments, are made up of stars.

(ii) Intergalactic globular clusters, some of which are at a distance of over $100,000$ parsecs from us.

(iii) Huge clouds of relativistic electrons ejected from the interior of the galaxies. For instance, the radio source of Centaurus A consists of three such clouds, while the source Cygnus A consists of two. Each of these clouds surpasses in dimensions the normal galaxies. Doubtless, most of the clouds of this type have already dispersed in intergalactic space.

(iv) Absorbing dust. We have no data on the dimensions of separate dust clouds.

(v) Neutral hydrogen. The radiation emitted (in the line $\lambda = 21$ cm) has not been so far detected with certainty.

Each type of intergalactic matter deserves special study.

II. The Basic Facts Referring to the Kinematics and the Dynamics of the Systems of Galaxies

Our knowledge of the motions in the world of galaxies is restricted to the radial velocity of about one thousand galaxies. We have no data on transverse velocities. The data on radial velocities, obtained almost exclusively at the observatories of Mount Wilson, Palomar and Lick, raise the most difficult problems ever dealt with by astronomy.

The whole complex of galaxies observed represents one part of a grandiose system which we call the Metagalaxy. The fact of expansion of the Metagalaxy has been inferred from the knowledge of the radial velocities of the galaxies. Hubble's law

$$V_r = Hr$$

deduced from the empirical data, holds good for values of r up to almost two thousand million parsecs. This circumstance signifies the approximate homogeneity of the observed expansion.

All efforts to find an explanation for the redshift, other than the Doppler effect, have been futile.

Of course, Hubble's law is correct only for mean velocities. In addition to the velocity as defined by Hubble's formula, every cluster and every galaxy has its own peculiar velocity. Thus, in the Local Group, where the distance between the galaxies is small, the relative velocities are mainly determined by the peculiar motion of individual members. But even the nearest clusters of galaxies and groups are receding from us. This proves the smallness of their peculiar velocities in comparison to the regular velocities of recession given by Hubble's formula.

The numerical value of the constant H is of great importance, since it enables us to determine the absolute distance of the remotest clusters. Unfortunately, its precise value is unknown. It is very likely to be somewhere within the limits of

$$60\,\text{km/sec/megaparsec} < H < 140\ \text{km/sec/megaparsec},$$

and we can assume, with some risk, that it is included in the narrower interval

$$70\,\text{km/sec/megaparsec} < H < 100\ \text{km/sec/megaparsec}$$

in conformity with the results of Sandage (1958). In any case Hubble's law makes it possible to give a correct estimate of *relative distances.*

The second important fact regarding the motions of the galaxies is the dispersion of the velocities in each cluster, related to the internal motions of the galaxies within them.

If the clusters are stable or are going to become stable, then their total energy E must be negative

$$E = T + U < 0,$$

where T and U represent respectively the kinetic and potential energies of the system. On the other hand, if $E > 0$, then such a system cannot be stable and at least some of its members must escape from it.

Recent investigations have shown that for certain groups and multiple systems the kinetic energy of internal motions determined from the radial velocities far exceeds the probable value of the absolute magnitude of the potential energy, calculated on the assumptions that the whole mass of the cluster (or group) is concentrated in its galaxies and that the mass/luminosity ratio $f = M/L$ for the given type of galaxies is of the same order as the value found from the rotation of galaxies. It follows that some groups and clusters possess positive total energy and will be dispersed. Such a conclusion is valid, for example, for the clusters in Virgo and Hercules, and also for the nearby group of galaxies in Sculptor. The latter case, analyzed in detail by Vaucouleurs, is outstanding, for the kinetic energy exceeds the calculated absolute value of the potential energy by one and a half or two orders of magnitude.

Insofar as the positive energy must lead to the escape of members or to total disintegration of the cluster, it can be presumed that there is something common between the instability of a cluster on the one hand, and the expansion of the Metagalaxy on the other.

In this respect an intermediate part is to be played by the systems of the type of Local Supergalaxy. Its component parts are known to recede from one another. For instance, the cluster in Virgo or the M 81 group recedes from the Local Group of galaxies.

What has been said about the sign of the absolute internal energy of the clusters of galaxies holds good for multiple systems as well. Apparently some multiple systems also have a positive total energy. All these facts indicate that the corresponding galaxies are comparatively young — their ages being of the order of 10^9 years.

Another peculiarity of the population of multiple galaxies (triples, quadruples, etc.) attracts our attention, regardless of the sign of the total energy. It is known that the great majority of the *multiple stars* have configurations of the "usual" type, while the configurations of the "Orion Trapezium" type make up only $< 10\%$. On the other hand, nearly half of the *multiple galaxies* have Trapezium-like configurations. As the Trapezium-like systems are, as a rule, unstable we can infer that the time elapsed since the formation of these multiple groups is not greater than a few periods of revolution of such a multiple system, i.e., of the order of 10^9 to 5×10^9 years.

Finally, the assumption of negative energy for all double galaxies sometimes entails improbably great values for the mass of the components (Page). Hence there is reason to believe that some of the double galaxies also possess positive energy.

In very close pairs, such as the radio galaxies, considerable differences in the velocities of the components are noticed; thus in the radio galaxy Perseus A this difference reaches 3000 km/sec. In this way these pairs also possess positive energy. We apparently observe in this case the formation of a pair from one galaxy.

Further accumulation of data on the radial velocities of galaxies will make possible the solution of many unresolved questions of their kinematics and dynamics. Some of these questions are listed below:

(a) *More precise determination of the constant of the redshift law.* This will provide the scale of extragalactic distances.

(b) *Determination of the character of the redshift law for very large distances.* We must observe, no doubt, a violation of linear dependence. The sign of the deviation and its magnitude are important factors in the basic problems of cosmology.

(c) *Determination of the peculiar velocities of the centers of gravity of the separate clusters of galaxies*, that is the deviations of their observable velocities from Hubble's formula. This is of great importance for the problem of genetic relations between neighboring clusters and requires methods of more precise determination of the distances of the remote clusters, independently of Hubble's law.

(d) *To solve many questions of the dynamics of multiple galaxies and of clusters it is essential to determine their masses.* Unfortunately, we determine the masses of distant galaxies statistically, assuming negative energy and applicability of the virial theorem.

It is necessary to determine the masses of the galaxies, at least in the nearest clusters, irrespective of this assumption, as well as to evaluate, at least, the upper boundaries of the probable intergalactic mass in each system (clusters or groups).

(e) *The inconsistency between masses of systems, defined by the virial theorem, and masses, determined from the luminosities of the individual members of the system, is most prominent for some irregular clusters and groups of galaxies* (clusters in Virgo, Hercules, the group of galaxies in Sculptor, in Leo, etc.). On the other hand, according to Zwicky, rich spherical clusters show no signs of expansion.

To reach a final decision on these problems a greater number of radial velocities in some of the nearest rich spherical clusters should be obtained.

III. Some Facts on the Nature of Galaxies and their Clusters

The galaxies display a great diversity of form and internal features. It is important to have a complete, but simple, system of classification of galaxies. The deeper the physical meaning of the criteria forming the basis of the classification, the more useful it will prove.

The classification of Hubble is based upon the observed outer forms of the galaxies. It proved extremely useful, for, until quite recently, all information concerning the majority of the galaxies was confined to data on the form, the integral brightness and the apparent diameter. The last two parameters do not by themselves characterize a system in so far as the distance remains unknown. In recent years, however, we have learned to judge approximately the absolute brightnesses and the linear diameters of many galaxies in rich clusters, as it became known that the brightest members of these clusters are always supergiants with an absolute magnitude of order of $-21{\cdot}0$. If we compare this value with the apparent magnitude of the brightest members, we can make a very rough estimate of the distance and, thereby, of the luminosity and the absolute dimensions of all the remaining members. The range of the luminosities of the galaxies within the clusters is great. Gradually, it became clear that attribution of a given galaxy to the category of supergiants, giants, objects of moderate luminosity, dwarfs and objects of extremely low luminosity (of the type of the Zwicky object in Capricorn) is in most cases much more essential than the recognition of its form. Recall that a supergiant galaxy contains millions of times more stars than any galaxy of extremely low luminosity.

To understand better the properties of the galaxies, it is significant to study their central parts, particularly the dense central nuclei of small dimensions. New attempts at classification should give proper consideration to luminosity criteria. There are probably also other parameters determining the state of the galaxy.

The classification suggested recently by Morgan takes into account the degree of concentration of luminosity and is in line with one of these desiderata; but the assignment of a Morgan class leaves the luminosity indefinite. Recently, Van den Bergh has made an attempt to construct a parameter determined by the observable forms of the galaxy which, in essence, defines its luminosity. This is a valuable principle. However, Van den Bergh's classification is not universal and involves only late-type spirals. Perhaps new classifications will be proposed in the future which will take into account the most important parameters of each galaxy.

The concepts of subsystems within galaxies and of various types of stellar populations are remarkable achievements of the second quarter of our century (Lindblad, Kukarkin, Baade). In a number of galaxies, such as the systems of EO type, we notice a rather high degree of homogeneity of the population; in such cases it can be maintained that the whole galaxy is composed of only one subsystem. This is true particularly for such members of the Local Group as the one in Sculptor, and galaxies M 32 and NGC 147. Contrary to the view expressed by Baade, we do not observe systems consisting entirely of populations of the first type (the population of the spiral arms). In most cases, the galaxies represent superpositions of two or more subsystems, containing various types of populations.

Thus the lenticular galaxies (SO) consist of two subsystems, made up of stellar populations of both the spherical component and the disk. The giant spirals of the M 31 type are composed of the spherical component, the disk and the spiral arms. Possibly a more detailed division is needed; but at this point we should like to emphasize *the phenomenon of superposition* of the various subsystems.

The data available indicate that the various subsystems follow different paths of evolution. There is reason to believe that the average age of the stars of different subsystems varies too. If the dynamical interaction is not considered then each of the subsystems leads its own life, i.e., the galaxies are composite systems, resulting from the *mere superposition of the subsystems.*

There is no correlation between the degrees of development of the different subsystems of the galaxies. This verifies the relative independence of the different subsystems entering into the same galaxy.

Thus the spherical subsystem of M 31 does not sensibly differ in its richness and dimensions from the normal galaxy of EO type of absolute magnitude about $-19{\cdot}0$. Nevertheless, the latter has no population whatever of a flattened subsystem and spiral arms, whereas M 31 possesses powerful spiral arms and a rich disk population.

Systems of an intermediate position are also of interest from this standpoint; in such systems one of the subsystems is quite well developed while the other is comparatively poor. A striking illustration is NGC 5128 (radio source Centaurus A) which on overexposed plates looks like an elliptical galaxy but, as a matter of fact, contains in its central part a poorly developed flattened subsystem, including a lot of absorbing matter. The work of the Burbidges, based on measurements of radial velocities in this flattened subsystem, has revealed that the equatorial plane of the latter is nearly perpendicular to that of the elliptical subsystem; this bears testimony to the independence of the subsystems. NGC 3718 is another interesting illustration. The spiral arms of this galaxy are of small power but, unlike NGC 5128, extend far beyond the boundaries of the volume occupied by the spherical subsystem; the plane of concentration of the dark matter is inclined at about 25° to the equatorial plane of the elliptical subsystem, which is further evidence in favor of the independence of the subsystems.

Opposite examples might also be produced in which the spherical system is underdeveloped while the flattened system is very prominent. The Large Magellanic Cloud apparently is an example: there is a spherical subsystem in this cloud as suggested by the existence of at least three dozen globular clusters similar to those in our Galaxy and in M 31. Unfortunately, other objects of the spherical subsystem are very hard to uncover against a background consisting of the population of the flattened subsystem. It is therefore difficult to say which type of elliptical galaxy resembles the spherical component of the Large Magellanic Cloud; judging by the number and distribution of globular clusters this must be an elliptical galaxy of moderate luminosity ($M \sim -16$), having a low density gradient. It is well known that when passing from the supergiant elliptical galaxies to those of moderate and low luminosity, objects with low density gradient become more frequent.

Mention has been made above of the comparative independence of the subsystems in a galaxy. But there is a bond between the subsystems which is always observed with great strictness: the common center. The center of the spherical subsystem coincides with that of the disk and also with the region out of which the spiral arms emerge. As is evident from observations of the nearest galaxies of high luminosity, this center is usually the location of a nucleus only several parsecs in dimension (less than the diameter of the common globular cluster). It seems natural, therefore, that the formation of separate, almost independent subsystems is somehow linked with the nuclei.

No traces of nuclei have been reported in a number of galaxies; such is the case, for example, with NGC 185 and with Sculptor. But let us consider the luminosities of the nuclei. The absolute photographic magnitude of the nucleus of M 32 is $-11{\cdot}6$; in M 31 it is $-11{\cdot}1$ and in M 33 it is $-10{\cdot}3$. In NGC 147 it is $-5^{m}{\cdot}0$. One gathers the impression that the luminosity of the nucleus decreases with the diminution of the density gradient. Therefore, in NGC 185 and in the systems of the Sculptor type, as probably in the Magellanic Clouds, the nucleus may have a still lower luminosity than in the NGC 147. If it is of order -2, the nucleus will be lost among the stars. In the Magellanic Clouds the nucleus may remain unnoticed even if it is of magnitude -5; it is therefore premature to insist on the absence of nuclei in these systems. But if the nuclei do exist they must be of small power.

We have noted above that usually the concentricity of the subsystems in each galaxy is observed very strictly. However, there are cases of deviation from this rule; NGC 4438, in the Virgo cluster, is a pertinent example where the two subsystems are clearly shifted in respect to each other.

There is a certain similarity between the galaxies and the clusters due to the fact that the member galaxies in clusters can be attributed to two different types of population, just like the stars in galaxies. To the first type belong the spiral galaxies and the irregulars, while to the second belong the elliptical and the lenticular galaxies (SO).

Large spherical clusters of galaxies, such as the one in Coma, are rich in populations of the second type. The loose clouds of galaxies like the Ursa Major Cluster, one of the nearest to us, have almost no elliptical galaxies of high luminosity. The nearest to us, a group of galaxies in Sculptor ($m - M = 27{\cdot}0$) studied by de Vaucouleurs, is not only deprived of elliptical galaxies but contains no galaxies of SO, Sa and Sb types either. Only spirals of late subdivisions are included in this group. The large irregular cluster in Virgo contains giant elliptical galaxies as well as giant spirals.

Then comes the question: *is it possible to speak of the superposition of different subsystems of galaxies in one cluster*? We admit that combination of two quasi-independent subclusters into one is not observed; nevertheless, certain observations clearly favor this possibility. Thus in the Coma cluster one of the central galaxies (NGC 4874), which is a supergiant of the SO type, is apparently surrounded by a symmetrical cloud of elliptical galaxies of lower luminosity. Externally this group is very much like galaxy NGC 4486, surrounded by globular clusters. In this case, however, the latter are replaced by elliptical galaxies of moderate luminosity; and this dense group of elliptical galaxies, with NGC 4874 in the center, overlies a large cluster of galaxies, possessing a small density gradient.

For irregular clusters of galaxies we can presumably discover many more phenomena testifying to the superposition of separate groups. The chain of bright galaxies M 84, M 86, NGC 4435, 4438 and others in the Virgo cluster is an example in point; this chain is not an accidental formation, as stated by Markarian a few years ago, but overlies the Virgo cluster as an independent group.

It is highly probable that the irregular clusters of galaxies generally represent combination or superposition of a number of more or less independent groups.

We recall in this connection that there exist clusters (or groups) made up of one central galaxy surrounded by a number of objects of low luminosity. The group around M 101, for example, belongs to that category. We emphasize this fact because, in such cases, the common origin of the central galaxy and its faint satellites is beyond doubt. However, there are groups consisting almost exclusively of supergiants; Stephan's Quintet is a pertinent example. Unlike the M 101 group, no galaxies of low luminosity appear around these supergiants. It is possible that here we have a case of discontinuity of the luminosity function, and that this system contains some galaxies with an absolute magnitude fainter than the detection limit. All this in conjunction with the exceptional position of M 31 and our Galaxy in the Local Group emphasizes the great cosmogonic significance of the supergiant galaxies in clusters and in groups.

No less important are the comparatively poor groups. If isolated groups composed exclusively of galaxies of low luminosity are not abundant, this would signify that the cosmogonic processes going on in a galaxy of high luminosity are essential for the formation of surrounding dwarfs.

We have taken only the first steps in the investigation of the character of the stellar population of galaxies and various subsystems. We are in urgent need of further data on the composition of the population based on spectroscopic evidence (on the lines contemplated by Morgan and Mayall) and also on the quantitative analysis of the continuous spectrum (Markarian et al.).

Another question of importance is the analysis of the nature of the arms of galaxies. Among galaxies with arms of the same degree of openness and length we find objects both very rich and very poor in O-associations. To discover a correlation between the nature of the arms and the other parameters of the galaxies would get close to realizing the cause of these differences.

Of particular interest are the barred spirals (SB). To our regret, we do not fully realize the difference between the population of the bars and the arms. It is only known that usually the color of the bars is considerably redder than the color of the arms and that the latter, therefore, contain a larger number of stars of recent formation. It is especially important to find out to what extent the bars are rich in open clusters and supergiant stars.

IV. An Extended Understanding of the Phenomena of Superposition

We have already spoken of the particular cases in which the centers of the subsystems comprising the giant galaxy are shifted relative to one another. However, we know of other galaxies which are double but are in fact linked together by means of a material medium and can, therefore, be regarded as single systems. M 51 and NGC 7752–7753 are good examples. It is natural to think that in these instances we are concerned with diverging centers of subsystems of one single galaxy. In the case of IC 1613, on one side of the main mass of the galaxy, there is a superassociation which can be regarded either as part of the main galaxy or as a separate satellite galaxy. It is highly probable that this superassociation, consisting of hot giants, was formed much later than the remaining galaxy. (We have almost the same situation in galaxy IC 2574. To the north of the main part of this galaxy we have a bright superassociation connected with it by means of a faint arm.) It can be presumed in this connection that the evolution of the galaxy is stipulated by the consecutive formation of the various subsystems. One or another of the subsystems, and sometimes a group of subsystems with a new center, may become the satellite of the principal galaxy. This view leads to the idea that satellite formation and the origin of new subsystems within the limits of a given galaxy are related phenomena. Moreover, these phenomena can supposedly accompany one another. Thus, where the spiral arms join the center of the galaxy to the satellite it is natural to assume that the formation of the spiral arm and the satellite coincide in time.

After all, a satellite of the type of the dwarf system in Sculptor, revolving around some giant galaxy, differs but little in scale and nature of population from a globular stellar cluster evolving within the galaxy. Therefore, it is natural to assume that satellites of the Sculptor type have a similar origin.

V. Phenomena of Instability in the Galaxies

Up to now we have spoken of the galaxies as stable formations. But unstable phenomena of great interest also take place in galaxies, particularly in the supergiants.

We do not refer here to the processes of star formation in associations, although they are sometimes pronounced; we mean more rapid, readily-observable changes. It is interesting to note that the greater part of these unstable phenomena is linked with the nuclei of the galaxies and can even be regarded as the manifestation of activity of these nuclei.

(a) *Neutral hydrogen flows out from the central part of our Galaxy.* This phenomenon was uncovered by the Dutch astronomers through observations made at 21 cm. A similar phenomenon of gas outflow from the nucleus of M 31 was detected by Munch, following an investigation of the line $\lambda 3727$. In both cases the outflowing mass is of the order of one solar mass a year; oddly enough, this result is difficult to reconcile with the data on the mass of the galactic nuclei in question (of the order of $10^7 M_{\odot}$).

(b) *In a number of galaxies, with nuclei of high luminosity as stated by Seyfert, the emission line* $\lambda 3727$ *is remarkably widened.* This corresponds to velocities of motion of the order of several thousand kilometers per second; these velocities exceed those of escape. We have a powerful outflow of gas, ejected with high velocity from the center and then dispersing into infinity. Here, the amount of the outflowing gas supposedly far exceeds the corresponding amount for our Galaxy and M 31. The blue galaxies of Haro, in which the emission lines are intensive around the region of the nucleus, may be of a similar nature.

(c) *In the very center of the radio galaxy NGC 4486 we also observe* $\lambda 3727$ *and deduce a rather heavy outflow of gas with velocity of about* 500 *km/sec.* If we compare this with the radial jet, springing up from the center of the galaxy and containing condensations radiating in intensive radio emission, we come to the conclusion that these condensations are ejected with high velocities from the central nucleus of the galaxy. The polarization of light in these condensations points to the presence of high-energy electrons. These condensations are formations of much larger scale than the Crab nebula; the power of their radio emission, measured in absolute units, is tens of millions of times greater. The duration of the radio emission must be, in this case at least, a thousand times more. We conclude that the supplies of energy in these condensations exceed the total supply of the energy of the Crab nebula by thousands of millions of times. In other words, by energy and masses, these condensations must be objects representing small-scale galaxies, in agreement with their photographic absolute magnitudes.

Have these condensations been ejected from the nucleus of the galaxy as clouds of relativistic electrons or, which is more probable, are they some objects ejected from the nucleus forming continually new currents of such electrons? The very fact that such large-scale condensations may burst out from the nucleus of the giant galaxy is of great interest. This can hardly conform with our information about the masses of the nuclei of the galaxies.

(d) *The situation in other radio galaxies is much more difficult to interpret.* Even the galaxy NGC 1275 (Perseus A) belongs to the class of Seyfert galaxies in which the line $\lambda 3727$, observed in the central regions, is widened. In this case too there is an intensive outflow of matter from the nucleus. The existence of two nuclei in the radio galaxy Cygnus A points seemingly to a process of nucleus division that took place recently. According to the views expounded above, this will lead to the formation of subsystems with various centers, with a consequent formation of a double galaxy.

NGC 5128 (Centaurus A) also demonstrates that the nuclei of galaxies are able to eject either huge clouds of relativistic electrons or some condensations of matter capable of giving rise to such clouds in the future. The radio galaxies are systems in which the central nuclei exhibit tremendous activity — including sometimes the formation of new condensations, new subsystems and probably new galaxies. We can, therefore, assuredly speak of the *cosmogonic activity of the nuclei* although we are unaware at the expense of which masses such activity is displayed.

(e) *We know of giant galaxies with jets bursting out from the central regions.* In some cases these jets contain *blue galaxies* of absolute magnitude of order -15, that is of luminosity higher than the condensation in NGC 4486. NGC 3561 and IC 1182 serve as examples of such galaxies. The ejection of these condensations is one more form of cosmogonic activity of the nuclei of the galaxies.

(f) *The spiral arms arise from the nuclei of the galaxies, proving that their formation is also directly connected with the nucleus.*

(g) *Radio observations of the center of our Galaxy (Parijsky and others) show that the state of a nucleus composed mostly of late-type stars* *differs markedly from the state of other groupings of similar stars (such as the globular clusters).* The nucleus of our Galaxy is the source of thermal radio emission while the surrounding region, with a diameter of the order of 500 parsecs, is the source of a heavy nonthermal emission. These facts indicate that the physical state of the nuclei is quite different from that of common stellar groupings.

One of the important problems confronting us in the study of the outflow of matter and ejections from the nuclei of the galaxies is the quantitative evaluation of the ejected masses. This refers equally to galaxies the central parts of which show emission lines, to the radio galaxies and to other types with discrete ejections.

The few facts at our disposal show that these data may conflict with the law of conservation of energy (and matter) in its present form, and may perhaps require a generalization of this law.

VI. Conclusion

We note that the activity of the nuclei determines the most important processes in the life of large galaxies. This activity assumes several different forms as discussed above. However, two of them deserve special mention: formation of the spiral arms and formation of the stars and the stellar clusters of spherical components. These phenomena seem to occur at different stages of development and are accompanied by corresponding changes in the nuclei. The process of formation of a given type of subsystem varies under different circumstances. Thus M 32 does not apparently contain globular clusters while another satellite of the Andromeda nebula (NGC 205) comprises at least nine globular clusters. It is most surprising that globular clusters occur in the galaxies with a low density gradient. If we take for granted the hypothesis of the formation of galaxies from initially diffuse clouds, it would seem natural that dense formations, such as the globular clusters, should originate in systems with very high density regions, that is where high density gradients exist. Nonetheless, such qualitative reasoning cannot be considered quite adequate. *The number of globular clusters per unit of luminosity of the spherical population varies in different systems.* We have, therefore, an additional parameter to characterize the spherical systems and subsystems. How this parameter is related to others of the same system (integral luminosity, density gradient) will become evident from observations.

Statistical data concerning multiple galaxies and clusters of galaxies show that these systems could not form by capture of formerly independent galaxies. A common origin is to be ascribed to the components of these systems. This problem was dwelt upon in detail in my Solvay report of 1958.

In light of the data on the ejection of condensations from the nuclei and also on the division of the nuclei, one can presume the formation of multiple systems and whole groups consequent upon the splitting of one initial nucleus into several. This division is likely to recur.

Where in a group there is a central galaxy of high luminosity, the formation of faint galaxies must be related mainly to the activity of the nucleus of the high luminosity galaxy.

High activity of the nuclei of supergiant galaxies is proved by the fact that radio galaxies usually form one of the brightest members of the cluster to which they belong. If there is one dominant galaxy in the corresponding cluster this must be the radio galaxy itself.

Observations reveal that, although all the large clusters contain supergiant galaxies, only a small part of the latter are radio galaxies. Thus the radio-emitting activity must be a relatively short-lived phase in the evolution of the galaxies. The ejection of the radio-emitting agents is apparently a phenomenon accompanying the outflow from the nuclei of more powerful masses, and possibly taking place at a certain stage of the particular cosmogonic process.

Although extragalactic astronomy has already explored certain forms of activity of the nuclei, our information on the other types of this activity is scanty. Still less is our knowledge of the parameters characterizing the integral properties of these nuclei: luminosity, mass, color, dimensions, rotation. After all, we know nothing about the internal structure of these nuclei. In this respect, the widest range of exploration in the field of extragalactic astronomy is open. Here are some of the questions and problems still outstanding:

(a) *Do all the galaxies possess nuclei? If not, which are the features* *that characterize the non-nuclear galaxies?*

(b) *Determination of the integral characteristics of the nuclei for a possibly large number of galaxies.* We stress the technical difficulty of this problem for galaxies of high density gradient. At the same time, the nuclei of most galaxies of the Sc type are of such prominence that they can be independently explored.

(c) *Determination of the correlation between the parameters of the nuclei and the galaxies.*

(d) *Exploration of the spectra of the nuclei in galaxies, showing emission lines, phenomena of rotation and outflow.*

(e) *Investigation of the interrelationships between the nucleus and the bar in barred galaxies and between the bar and the phenomenon of outflow from the nucleus.*

(f) *Investigation of galaxies with multiple nuclei and of the radial velocities of the separate components of such nuclei.*

(g) *Dependence of the total number of globular clusters upon the nature of nucleus of the galaxy.*

In our formulation of conceptions concerning the formation of the galaxies, we have attempted to adhere to the facts and to refrain from unfounded speculations. The phenomena concerning the formation of the galaxies are so unusual that they cannot be predicted on the basis of any preconceived theory. We again are in a situation familiar in the history of astronomy: in the realm of new phenomena, new qualitative facts are found which transcend the limits of existing conceptions. Only an increase in observational data and more precise information about real objects and the structure of the various parts of the galaxies can help us in solving the difficult problems, some of which have been discussed here.

GALAXIES AND THEIR NUCLEI[1]

Our Extraordinary General Assembly is devoted to the memory of one of the greatest men of science, the great Polish astronomer Nicolaus Copernicus. The main service of Copernicus which made his name immortal was in finding the correct interpretation of the planetary motions we observe. Instead of the geocentric concept, which proved unable to explain the accumulated bulk of empirical data on the apparent motions of planets he put forward and advocated the concept of a *solar system*. The scientific revolution started by him was continued by Galileo and Kepler and was crowned with the great theoretical generalizations of Newton. A foundation has been created for the exact theories of motions in the solar system which were developed during the next centuries. These theories in their modern form make it possible to solve all the problems concerning the orbital motions of spaceships as well.

At this stage we can hardly boast that in the study of nuclei of galaxies and their activity we have reached a level comparable to the level of planetary astronomy even before Newton. Only 15 years have elapsed since the first attempt to formulate the idea of activity of nuclei of galaxies (Ambartsumian, 1958). During these years discoveries of greatest importance have been made. New unexpected discoveries occur almost each year. These discoveries influence decisively our ideas on the diversity of objects and phenomena in distant space, but they are still insufficient for the construction of adequate theories. Penetration into the nature of nuclear phenomena requires new observations, new measurements and new data. And if some optimists imagine that the time is ripe to build a general theory of these phenomena, more cautious astronomers would consider a tremendous success merely the satisfactory systematization of observational data concerning external physical processes accompanying the activity of the nuclei.

[1]Originally published in *Highlights of Astronomy*, G. Contopoulos (ed.), 1974, pp. 51–56. © IAU, 1974. Used here with permission of the International Astronomical Union.

These external processes often reach such a large scale that they influence the appearance and integral parameters of the galaxies. Therefore, the study of the nuclei and the study of the structure of galaxies with their evolutionary changes often cannot be separated.

In this report we consider some properties of galaxies immediately connected with the activity of nuclei and ultimately with the nature of the nuclei themselves.

1. Soon after the introduction of the concept of activity of nuclei, observations revealed some new forms of that activity. Now we can speak about the considerable diversity of external forms of activity of galactic nuclei. The following we consider more important:

(a) Ejection from the nucleus, and from the volume of the galaxy itself, of giant masses which transform into large clouds of relativistic plasma. Owing to this the galaxy transforms into a *radio galaxy.*

(b) Enhancement of the optical luminosity of the nucleus. Due to this form of activity the galaxy passes into class 5 of the Byurakan classification (Ambartsumian, 1966) or becomes Seyfert. In the case of a stronger increase of the luminosity we have an *N galaxy.* The extreme form of the same kind of activity are *quasars*, where the nuclei reach an absolute magnitude of −25 and even higher.

(c) Ejection and motion of gaseous clouds in and from the nuclei of Seyfert galaxies and from quasars.

(d) Great explosions which lead to the ejection of large gaseous masses of the order of 10^6 $M_\odot$, like the ejection that occurred in M 82.

(e) Relatively small but recurrent explosions in nuclei which manifest themselves as increases in radioluminosity at the centimeter range of wavelengths.

2. There are many indications that along with the above forms of activity which we observe immediately, there exist also the hypothetical forms:

(f) Ejection of condensations from the nuclei of supergiant galaxies which are capable of transforming into new galaxies (mostly into a satellite galaxy or a member of the group which surrounds the supergiant galaxy).

(g) Outflow of matter from the nucleus which can produce the spiral arms.

In the course of systematization of the data on galaxies and phenomena occurring in them it seems appropriate to assume that the forms f) and g) exist in reality. The possible (albeit improbable) fallacy of this assumption cannot discredit the results of such systematization.

3. Manifestations of nuclear activity are very unusual physical processes, and we are still far from understanding their real nature. Therefore, it is too early to build models explaining them. The majority of known galaxies having active nuclei are very distant. The nuclei and much larger central volumes in which the most important processes take place usually remain unresolved by our instruments. In the majority of cases we are not convinced that the radiation we detect is coming immediately from the central body which is the main source of activity. As a rule it is difficult to clarify *even the geometrical picture of the external processes*, let alone the mechanism of the active source or the structure of the true nucleus.

It seems that prior to construction of models a considerable amount of work must be done in order to find the empirical patterns. Such work requires a correct classification of objects and phenomena founded on direct observations of the external manifestations of nuclear activity. Only after the study of external manifestations may we find the way to the essence and causes of phenomena in the nuclei of galaxies.

4. At this stage the *systematization* of data on external manifestations of nuclear activity must become an important aim. During the last twenty years I have defended the opinion that each galaxy and its subsystems (the spherical component, disk, spiral arms) are the result of nuclear activity. In this sense manifestations of nuclear activity and the data on galaxies cannot be separated. However, the first step can be a classification based on direct forms of nuclear activity.

During our century several systems of classification of galaxies have been worked out and practically applied. In particular, classifications of Morgan and of the Byurakan Observatory put considerable emphasis on the properties of the structure of central circumnuclear regions of galaxies. Hubble's system takes into account only the presence and strength of spiral arms. In this report we are going to consider only a partial question related to the classification of a rather special category of galaxies. Within this category we will consider the parameters and properties connected with nuclear activity.

5. Let us concentrate our attention on galaxies of which the spherical component only (population II) has the total absolute magnitude $M_V < -21.0$ independently of the presence of other components (disk, spiral arms). If the stars of population I are present, the total luminosity of such galaxy will exceed the given limit $M_V = -21.0$. Thus we consider the galaxies of highest luminosity and mass.

These supergiant stellar systems are special in many respects, particularly in properties which depend on the activity of their nuclei. Our aim is to specify some essential parameters useful for classification of these systems.

We remark that though among all galaxies the supergiants are rare, their total mass represents a considerable part of the total mass of all galaxies (Ambartsumian, 1962).

6. We will consider the following parameters:

(a) *Radioluminosity* L_R. Radio galaxies are defined as systems emitting strong radio emission with radioluminosity $L_R > 10^{41}$ erg s^{-1}. At the same time they have high optical luminosities. Therefore, all radio galaxies belong to the category of galaxies under consideration. Curiously enough, choosing the galaxies according to the criterion $L_R > 10^{41}$ erg s^{-1} yields a sample for which the dispersion of optical luminosities is much smaller than the dispersion of radioluminosities.

For spiral galaxies which are sources of radio emission the radioluminosity is always lower than the indicated lower limit. Therefore, the radio galaxies are always E-systems. However, they sometimes contain significant quantities of dust and population I stars.

(b) *Optical luminosity of the nucleus.* The nuclei of galaxies in our category range from optically weak (NGC 4486 where $M_n > -15$) to the mini-quasars (the nuclei of N galaxies) and can even contain quasars with absolute magnitude reaching $M_n = -27.5$. Thus the range of optical luminosities of nuclei of systems under consideration is of the order of 10^5.

(c) *Presence, strength and degree of development of spiral arms* and generally the relative strength of stellar population I. It is desirable to choose again one numerical parameter, perhaps the ratio of the mass of neutral hydrogen in the given galaxy to its total mass. The use of another important parameter — the ratio L/M which according to the work of the Meudon group (Balkowski et al., 1973) changes abruptly when we pass from spiral and lenticular galaxies to ellipticals — seems not very practical. The value of $B - V$ of the galaxy can probably be of some use.

(d) *Presence, intensity and degree of development of the bar.* According to the surface photometry of giant SB galaxies (Kalloghlian, 1971) the surface brightness of the bar has some preferred value (the mean photographic surface brightness along the axis of the bar is about 21 magnitudes per square second of arc). Therefore, very roughly we can consider this parameter as attaining only two possible values (0 or 1 depending on the absence or the presence of a bar).

(e) Diameters of the spherical components of populations (which in the case of E-systems coincide with the galaxies) can vary greatly. For high luminosity systems, this implies different surface brightnesses. This corresponds to the division of supergiants according to the *degree of compactness.* At this stage it will be convenient to consider three types of galaxies: extended systems (with a diameter larger than 40,000 pc), normal systems (with a diameter between 15,000 and 40,000 pc), and compact galaxies (diameter less than 15,000 pc). To make the division more exact one must add to this some definition of the diameter of a galaxy.

Zwicky has suggested the mean surface brightness as an additional numerical parameter describing the degree of compactness. However, it is evident that the average

$$< i >_0 = \frac{\int i \, ds}{\int ds},$$

where i is the surface brightness, is not suitable since the domain of integration remains uncertain. One can propose instead a weighted average of the surface brightness, for example

$$< i >_1 = \frac{\int i^2 \, ds}{\int i \, ds}.$$

However, in this case the relative role of the central region becomes too large, and, therefore, it is necessary to know sufficiently well the exact behavior of i near the center of the galaxy, which is difficult since the angular resolution by photometric measurements is low. Probably the best practical alternative is to adopt an average of the type

$$< i >_2 = \frac{\int i\, f(i)\, ds}{\int f(i)\, ds},$$

where $f(i) = i_0$ when $i > i_0$ and $f(i) = i$ when $i < i_0$. Here i_0 is some conventional, fairly high surface brightness corresponding, for example, to 22nd mag. from a square second of arc.

Since the degree of compactness must be closely related to the ratio of the absolute value of gravitational energy of the system to the square of its total mass (this ratio is proportional to the radius of the system) it is clear that this degree of compactness must depend on the mechanism and conditions of formation of the spherical component of the galaxy, particularly on the mean kinetic energy of the member stars.

7. Compact galaxies attracted much attention after studies by Zwicky, who even discovered several clusters consisting of compact galaxies (Zwicky, 1971).

Recently Robinson and Wampler of the Lick Observatory have shown that the cluster Shakhbazian 1 is a *compact group of compact galaxies.* Shakhbazian (1973) has presented a list of similar groups consisting of compact galaxies.

Zwicky has expressed the view that quasars are the extreme cases of highly compact systems. There is now fairly good evidence that quasars generally have underlying galaxies. These underlying galaxies often are extended objects. Therefore, the fifth (compactness) and the second (luminosity of the nucleus) parameters should be considered separately. The question of statistical correlation or anticorrelation between them must be solved using observations. *One of these characteristics describes the distribution of stellar populations, the other — the state of the nucleus.*

8. The class of systems in question includes spirals with sufficiently luminous spherical subsystems. It is known that the ellipticals, which consist only of spherical subsystems, have an L/M ratio four times smaller than

the spirals. This means that in order to have a spherical subsystem of absolute magnitude $M_V = -21.0$ a spiral must have a total visual magnitude not lower than -22.5.

There are many ellipticals of such high luminosity, but the spiral systems with $M_V < -22.5$ are very few indeed. Among the possible candidates for this category is Markarian 10, for which $M_V \sim -23.0$. However, in order to place it in the category of systems we consider, it is necessary to show that its spherical subsystem has in fact a partial luminosity of $M_V < -21.0$. It may happen that the real number of such spirals is very small or they do not exist at all. In the latter case our conclusions are liable to some changes.

9. We see that the supergiant systems depend on at least five different parameters.

The question arises whether all possible combinations of different values of these parameters are represented among the real galaxies. We simplify this complex problem by means of very rough discretization:

(a) If the radioluminosity of a galaxy $L_R > 10^{41}$ erg s^{-1} we set $\alpha = 1$. If $L_R < 10^{41}$ erg s^{-1} we set $\alpha = 0$. For all radio galaxies $\alpha = 1$. For all other galaxies $\alpha = 0$.

(b) If the nucleus of a galaxy has an absolute photographic magnitude $M_{pg} > -21$ we put $\beta = 1$. If $M_{pg} \leq -21$ we set $\beta = 0$. For all other galaxies $\beta = 0$.

(c) If on the plates of sufficient resolution and density in photographic light a galaxy has noticeable spiral arms we put $\gamma = 1$. If they are unnoticeable, then $\gamma = 0$.

(d) If on the plates of sufficient resolution in photographic light a galaxy has a discernible bar we put $\delta = 1$. In the opposite case we set $\delta = 0$.

(e) If a galaxy is compact, i.e., if on the maps of the Palomar Sky Survey or equivalent, its radius is less than 15,000 pc, we put $\varepsilon = 1$. In the opposite case $\varepsilon = 0$. Incidentally, this definition of compactness based on diameter is correct only for high luminosity galaxies. For less luminous galaxies the limiting value of diameter must be smaller.

10. The five-digit binary number $S = \alpha\beta\gamma\delta\varepsilon$ roughly describes the state of a given galaxy. Correspondingly, in decimal numeration the state of a galaxy can be given by one of the numbers from $S = 0$ to $S = 31$. For example, $S = 0$ means a radioquiet noncompact galaxy with no quasar in its center, no spiral arms and no bar. *It is simply a normal elliptical galaxy.*

We can discuss the problem of independence of our parameters in two different ways.

(A) Do all 32 values of S have their counterparts among the galaxies? In other words, are all combinations of discrete quantities α, β, γ, δ, ε realized in the Universe? We have seen above that the combination 00000 corresponds to a normal elliptical and therefore is very frequent. On the other hand, some combinations, for example 10100 ($S = 20$), never occur. Unfortunately, we do not know whether the combination 11111 ($S = 31$) is realized anywhere, i.e., if there are quasi-stellar radio sources for which the underlying galaxy is of SB type and the spherical subsystem is compact.

(B) Can we represent the distribution function P_S of the values of S for galaxies in a unit volume as a simple product

$$P_S = \varphi_1(\alpha)\,\varphi_2(\beta)\,\varphi_3(\gamma)\,\varphi_4(\delta)\,\varphi_5(\varepsilon),$$

where $\varphi_1(\alpha)$ is the probability of the given value of α, each φ_k has a similar meaning.

Evidently the answer is negative. This follows from the fact that for some values of S we have $P_S = 0$. But this means that at some value of its argument at least one of the functions φ_i must vanish. But this cannot be the case, since this would mean that one of the two values of that argument is not realized in the Universe at all, while we have introduced our parameters on the ground that both values have been observed.

11. As regards the first problem (A), we can confine ourselves to the simpler question of compatibility of values of pairs of parameters (for example of α and β). From Table I containing the data on four extragalactic objects we can see that all four combinations of values of α and β exist in the Universe. This statement means that both the presence and absence of strong radio emission are compatible with the presence or absence of a quasar (or of a miniquasar) in the center of a galaxy.

Apparently we have a similar situation for the parameters β and ε, i.e., the presence of a quasi-stellar source and the compactness. This may be seen from Table II.

Table I

Object	Parameter	
	α	β
NGC 4889	0	0
Ton 256	0	1
NGC 4486	1	0
3C 371	1	1

Table II

Object	Parameter	
	β	ε
NGC 4889	0	0
IZw 94	0	1
Ton 256	1	0
Zw 0039 + 4003	1	1

However, statistically the compactness of a galaxy is rather anticorrelated with the presence of quasi-stellar objects in its central region.

The situation is more complicated for α and ε. Among the radio galaxies there are both extended stellar systems and systems of normal size. But we do not know any compact radio galaxy (with diameter less than 15,000 pc) nearer than 500 Mpc. Thus if $\alpha = 1$ we have $\varepsilon = 0$. But if $\alpha = 0$ the quantity ε can be equal either to 1 or to 0. Thus between the values of our two parameters no one-to-one correspondence exists.

We also cannot exclude the possibility that among the very distant (more than 500 Mpc) radio galaxies there are compact systems which have an external appearance similar to quasars. Therefore, before making any final conclusions we must wait for more refined data about the sizes of the optical images of quasi-stellar radio sources.

The survey of external forms of radio galaxies shows that none of them has developed any regular spiral arms. Strictly speaking this means that the values $\alpha = 1$, $\gamma = 1$ are incompatible. On the other hand, the radio galaxies NGC 5128 and 2175 show the presence of dust, gas and stellar population I. This is not equivalent to the presence of developed and regular spiral arms. Therefore, in this case too we cannot write $\gamma = 1$ or $\delta = 1$. These cases probably indicate that the radio galaxy phase precedes the phase of evolution of supergiant galaxies at which the developed spiral arms are formed.

12. We cannot conclude from the fact that the combination $\alpha = 1$, $\gamma = 1$ never occurs that the value of γ is determined by the value of α. In fact, when $\alpha = 0$ both $\gamma = 0$ and $\gamma = 1$ can happen. Again we have no one-to-one correspondence.

The question arises of the evolutionary interpretation of the chosen parameters and their combinations.

13. The diversity of forms and states of the galaxies is probably due to (a) differences of age and (b) differences of initial conditions. Among the initial conditions such quantities as the mass of a system, its total internal energy and the rotational momentum play an important part. Some significance can also be attached to the differences in the initial chemical composition. However, there is no doubt that during the life of a galaxy the chemical elements in it undergo essential evolutionary changes. If the initial dominant state of matter was of atomic type (and not of the type of nuclei or particles having masses of stellar or larger order), the simplest assumption would be the similarity of initial chemical composition. If, however, in the beginning the nuclear phase was predominant (of the type of the baryon star structures), it is probable that after the transition to atomic structure of matter approximately the same chemical composition emerged. Therefore, it seems possible to disregard the possibility of differences in the initial chemical composition.

For supergiant galaxies and systems which have masses of about the same order of magnitude the remaining three parameters are (a) age, (b) total energy, and (c) rotational momentum.

Thus we have a situation where the number of empirically determined parameters, which specify the different states of the systems, is larger than the number of parameters related to the diversity of initial conditions.

However, the activity of a nucleus sometimes takes such intense and cataclysmic forms that in a short time it can cause essential changes in the properties of the galaxy and even originate new temporary properties which should be described by new values of parameters.

This happens, for example, in the case of the sudden appearance of strong radio emission (formation of large clouds of relativistic charged particles), or X-ray emission or strong nonstellar optical emission (quasars).

The intervals of time during which different new properties are maintained can overlap. If for two given properties A and B there is partial overlap of time intervals τ (for example, the interval τ_B begins somewhere in the middle of τ_A), then the following cases are possible: (1) both properties are present in a galaxy, (2) a galaxy has only one of them, (3) a galaxy has none of the two. Exactly this happens for the pairs α, β and β, ε.

If the intervals do not overlap, both properties never meet in the same system. However, from the absence of one of them we cannot conclude about the presence of the second. Exactly such a situation occurs in the case of α, γ, i.e., the presence of spiral arms is incompatible with strong radio emission, but the absence of radio emission does not necessarily mean the presence of arms.

We can hope that further work on the classification of galaxies and measurements of the essential parameters will allow us to determine the lengths of time intervals in which the properties of interest are maintained.

14. Regarding the galaxies for which the spherical subsystem is considerably fainter than $M = -21.0$, we make the following remarks.

(a) Such galaxies are never strong sources of radio emission. But they can emit weak (normal spirals) or moderate radio emission (some Seyfert galaxies, for example NGC 1068, or the irregular M 82). For them always $L_R < 10^{41}$ erg s^{-1}.

(b) Such galaxies frequently have spiral arms.

(c) In this category, moderate radio emission and the presence of spiral arms are compatible.

(d) Many examples of Seyfert and Markarian galaxies show that these objects of low luminosity can have nuclei of relatively high luminosity. Such objects are sometimes the galaxies that host radioquiet quasars. However, many radioquiet quasi-stellar objects have high luminosity host systems, i.e., supergiant galaxies.

(e) There is a definite lower limit for the integral luminosity of the spherical component of galaxies capable of forming spiral arms of more or less regular shape. The exact value of this limit is not known but it is probably near $M = -14$. The galaxies with still fainter spherical subsystems can have some stellar population I and interstellar material of appreciable density. However, in such cases they have irregular shapes.

15. Taking as a starting point the assumption that the formation of spiral arms is the result of nuclear activity, we summarize these facts in the following way:

(a) If the spherical component has an integral absolute magnitude $M_V < -21.0$, then the nucleus is able to form large radio-emitting clouds, but rather seldom regular and bright spiral arms.

(b) If the luminosity of the spherical subsystem is within limits

$$-21.0 < M < -14.0$$

then the nucleus of such galaxy never forms strong radio-emitting clouds, but frequently forms regular spiral arms.

(c) For $M > -14$, the nucleus cannot form regular spiral arms but still is able to produce relatively abundant population I.

16. According to observations, the type of nuclear activity depends on the absolute magnitude and, therefore, on the mass of the spherical subsystem. On the other hand, it is clear that the spherical component of the galaxy can hardly have any direct influence on the properties of the nucleus. Therefore, there remain two possibilities:

(a) The spherical component itself is the result of the nuclear activity. Therefore, it is strongly correlated with the other external manifestations of the same activity.

(b) The nucleus and the spherical subsystem have been formed together. The properties of the nucleus and the mass of the spherical component are determined by the integral mass of the galaxy.

At this stage it is difficult to decide which of these alternatives corresponds to reality. However, the general considerations concerning the universal role of nuclear activity make the first possibility more likely.

17. No compact galaxies displaying strong radio emission are known. But compact galaxies sometimes possess a considerable population of type I and even form ejections and plumes which are similar to the spiral arms. As an example we have the galaxy NGC 1614. A very preliminary survey of *compact groups of compact galaxies* carried out in Byurakan has shown that some of them contain blue compact galaxies. Usually these blue members have almost elliptical appearances somewhat disturbed by the presence of absorbing matter. A study of color distribution in such galaxies may bring interesting results.

18. The study of clusters and groups of galaxies inevitably leads to the conclusion that *the supergiant galaxies play a particularly important role in the Universe.* Quoting from a recent paper of Sandage (1972): "The luminosity of the brightest cluster member does not depend strongly, if at all, on the luminosities of the fainter cluster members." The dispersion of absolute magnitudes of the brightest cluster members is of the order of 0.25 mag.

Except for a few compact groups, each cluster contains at least one member having a mass of the order of 5×10^{12} $M_{\odot}$. How can the theory of formation of clusters of galaxies from a large cloud of diffuse matter explain the existence of a definite upper limit for the masses of the parts into which the cloud splits and at the same time the existence of at least one part with mass of the order of that limit?

Under the alternative hypothesis of fragmentation of an initial dense and massive body, it is quite natural to suppose that during each step of such fragmentation a body divides into several pieces having masses of equal order. In this way at some stage dense bodies with masses of the order of 5×10^{12} $M_{\odot}$ will be formed. Then for some reason the division into almost equal masses stops and each part behaves as an active nucleus. Perhaps at this stage it is better to say "protonucleus." This means that each such part forms around itself a galaxy, consisting of stellar populations of different kinds. Moreover, ejection of secondary nuclei of smaller masses

(10^{11} $M_\odot$) is possible. Thus the nucleus of a supergiant galaxy contributes to the formation of the less massive galaxies of the cluster.

19. Zwicky has established the existence of several large *clusters consisting of compact galaxies.* Since it is difficult to judge the compactness of faint members, it is more correct to say that a number of bright galaxies in each such cluster are compact.

Among these clusters is Zw Cl 0152 + 33 which has the angular diameter of about one degree. Since the distance must be of the order of 5×10^8 pc (this corresponds to the radial velocity V_r = 26,300 km s^{-1} determined by Sargent (1972)) the linear diameter is of the order of 10^7 pc. Since the dispersion of radial velocities is of the order of 1000 km s^{-1} we arrive at the conclusion that to cross the cluster a galaxy needs time of the order of 10^{10} yr. Therefore, one can suppose that the differences in the ages of galaxies in a cluster are of this same order of magnitude. This means that compactness is not a transient property of a galaxy and lasts at least hundreds of millions, perhaps billions of years. This implies that these galaxies are in a stable state. It follows that these systems *will remain compact* also in the future, during the lifetime of stars from these systems.

Perhaps somewhat extrapolating we can suppose that compact galaxies as a rule are born and remain compact during their lifetime. In any case they are systems *sui generis* rather then some stages of evolution of normal galaxies.

The division of clusters and groups of galaxies into systems consisting of normal galaxies on one hand and of compact galaxies on the other is fundamental. This division must be intimately connected with the mechanism of formation of galaxies in clusters. It is very difficult to explain such a division on the basis of the hypothesis of formation of galaxies from diffuse matter.

20. *Compact groups of compact galaxies* are of great interest. Such systems usually have from half a dozen to two dozen members, though there are richer groups too. Typical representatives of such groups are No. 1 and No. 4 of Shakhbazian's list which will be published shortly. The first consists of 17, the second of 7 members. The linear sizes of these groups are of the order of 2×10^5 pc.

The first group was found at Byurakan in 1957 (Shakhbazian, 1957) during the study of the Palomar Sky Survey maps. Due to compactness of its members and of the group itself it looks very different from other groups of galaxies. This was the reason that after some hesitation we first supposed that it was a stellar cluster situated at some distance from our Galaxy. Later Kinman and Rosino (1962) found on large scale plates that some members of the group are galaxies. Since the other members looked like stars, they concluded that the group was a chance agglomeration of galaxies and stars on the sky. Only recently Robinson and Wampler (1973) have found from spectral observations that it is a physical group of compact galaxies. Meanwhile, new groups of similar type have been found at Byurakan.

The group Shakhbazian 1 has the redshift $z = 0.1$, i.e., it is at a distance of six hundred million parsecs from us. The brightest centrally located member of the group has an absolute magnitude of the order of $M_V = -23$. It is interesting that the brightest member of Zwicky cluster $0152 + 33$ has roughly the same luminosity.

Determination of dispersion of radial velocities of the members of Shakhbazian 1 from redshifts and application of the virial theorem have shown that the M/L ratio expressed in solar units is of the order of unity.

Thus in this case the virial theorem gives too small masses. In this sense the situation is opposite to what we have in usual clusters of galaxies.

However, there exist similar compact groups of compact galaxies, where the dispersion of radial velocities is not that small. According to unpublished observations by Khachikian[2], the dispersion of radial velocities in the remarkable compact group Shakhbazian 4 is of the same order of magnitude as in usual clusters. Probably in this case we again have an expanding group.

The compact clusters of compact galaxies differ from usual clusters (as catalogued by Zwicky and Abel) in that the integral magnitudes of member galaxies are contained in a narrow interval of stellar magnitudes and the difference of magnitudes between the first and the second brightest galaxies

---

[2]Ed. note: E. Ye. Khachikian, C. R. Linds, A. Amirkhanian, *IAU Colloquium*, No. 124, 1990.

is relatively small.

The search for new groups of Shakhbazian 1 type is now in progress at Byurakan. The number of groups already found has reached several dozen. Very preliminary statistics show that the number of such clusters with 18.5 red mag. for the brightest member must exceed one thousand.

Thus the compact groups of compact galaxies (CGCG) represent *one of the important constituents of the Metagalaxy.*

21. We conclude that the study of compact galaxies and their clusters may shed some light on the problem of the origin and evolution of galaxies and the nature of the activity of their nuclei.

The compact galaxies are not strictly isolated from the world of normal galaxies. Quite the opposite: there are cases when it is difficult to classify a given galaxy as compact or normal. It seems that the attention given to such intermediate cases will be rewarding.

The groups included by Shakhbazian in her first list contain almost exclusively compact galaxies. This is the result of intentional selection. If mixed systems exist, their study will help us understand opposing phenomena in the extragalactic world.

22. It is natural to suppose that there exist systems for which the ratio M^2/H, where H is the total internal energy, is smaller than what has been measured until now. Such systems would appear as less extended, compact galaxies. However it is remarkable that:

(a) There exist rich clusters of galaxies containing dozens of compact, but no extended, systems of high luminosity.

(b) In the observable part of the Metagalaxy there are thousands of compact groups of compact galaxies containing from five to ten compact systems but lacking normal or extended galaxies.

(c) Despite the differences in the nature of compact and normal galaxies and probable differences in the values of M/L (perhaps more than ten times), the upper limit of luminosities for the normal galaxies ($M_V = -23.7$) is apparently a sufficiently exact upper limit for the compact galaxies as well.

(d) The colors of compact galaxies apparently are not very different from the colors of some normal galaxies. However, no detailed study of the colors of compacts has been done.

23. Compacts apparently comprise only a small percentage of all galaxies. The nearest compact galaxies of high luminosity are at distances not less than 50 million parsecs from us. I am not quite sure that the Shapley-Ames catalogue contains even one high-luminosity compact. However, beginning with $m = 13.0$ they appear. Due to their high surface brightness, the compact galaxies of 13th or 14th apparent magnitude must have diameters smaller than $20''$. Therefore, their detailed morphological study will be difficult.

At the same time I should like to warn against unfounded pessimism. During the time since Copernicus astronomy has overcome the distance barrier from 10^{-4} pc to some billions of parsecs. In any case at distances of 800 million parsecs a contemporary extragalactic astronomer feels almost at home.

Now it is necessary to overcome the barrier of low angular resolution in optical observations.

This can be achieved by combination of optical interferometry with the use of observations from outer space. The fall of these barriers will open up new prospects in extragalactic research.

24. Finishing this discourse I would like to pay tribute to some astronomers who contributed to the solution of the above problems.

By tremendous observational work *Sandage* has reached the final conclusion that quasars are nuclei of supergiant elliptical galaxies. Having understood the significance of high-luminosity E-systems in the Metagalaxy he established the regularities concerning the brightest objects in a cluster of galaxies.

Essentially I agree with his important conclusions and they have been used above.

By his studies of compact galaxies *Professor Zwicky* has opened a new page in extragalactic astronomy. Every compact galaxy which has emission lines in its spectrum is of great individual interest as in the case of quasars and Markarian's galaxies. However, the totality of compact galaxies (the majority of which have no emission lines) presents much deeper interest and significance.

By attracting the attention of astronomers to the compact galaxies, Zwicky has shown once again how far from reality are those who think that we already know the composition and structure of the Universe, and that it is up to the theoreticians to put the last touches to their models. New types of objects and new kinds of processes in the Universe literally compel us not to follow such oversimplified views.

The idea of the inexhaustibility of the Universe has led modern astronomy to its great discoveries. This idea will perhaps remain a source of inspiration despite the increasing difficulty and depth of the arising problems. Then astronomers of 2473, celebrating the thousandth anniversary of Copernicus, will admit that the generation which lived halfway was not always sitting idle, but was sometimes unrestrained and fearless in the search for the unknown properties of the Universe.

REFERENCES

1. Ambartsumian, V.: 1958, *XI Solvay Conference Report*, Bruxelles, p. 241.
2. Ambartsumian, V.: 1962, *Trans. IAU*, XIB, 145
3. Ambartsumian, V.: 1966, *Trans. IAU*, XIIB, 578.
4. Balkowski, V., Botinelli, L., Gougenheim, L., and Heidmann, J.: 1973, "21 cm Neutral Hydrogen Line Study of Early Type Galaxies," in press.
5. Kalloghlian, A.: 1971, *Astrofizika*, 7, 189.
6. Kinman, T. D. and Rosino, L.: 1962, *Astron. J.*, 67, 644.
7. Robinson, L. B., and Wampler, E. J.: 1973, *Astron. J. Letters*, 179, 135.
8. Sandage, A.: 1972, *Astrophys. J.*, 178, 1.
9. Sargent, W. L.: 1972, *Astrophys. J.*, 176, 581.
10. Shakhbazian, R.: 1957, *Astron. Circ.*, No. 177, 11.
11. Shakhbazian, R.: 1973, *Astrofizika*, in press.
12. Zwicky, F.: 1971 *Catalogue of Selected Compact Galaxies and of Post Eruptive Galaxies*, Offsetdruck L. Speich, Zürich, Switzerland.

Academy of Sciences,
Armenia, U.S.S.R.

INTRODUCTION TO *NUCLEI OF GALAXIES*[1]

1. The great majority of galaxies have a density maximum somewhere near their center of inertia. We prefer to call this region of maximum density *the central part* of a galaxy.

In the astronomical literature this *central part* is sometimes called the *nucleus* of the galaxy, but we are going to avoid such usage. It is not this maximum density region with very indefinite borders which we are going to discuss.

In some nearer galaxies, for example M 31, M 32, M 33, we see that there is a starlike, or almost starlike, image superposed on this region of maximal density. In many distant galaxies the limits of angular resolution do not allow us to see similar starlike formations. They become lost in the bright background of the central part. However, in some distant galaxies these superimposed starlike formations have sufficient luminosity to be observed even in cases when the angular resolution of photographs is moderate ($1'' - 3''$). This is for example the case in Seyfert galaxies. Similar, but less prominent, starlike formations we observe in the photographic images of many other (mostly spiral) galaxies. I will apply the term "nucleus" just to these formations. They display a number of exceptionally interesting phenomena, the discovery of which affected the whole outlook of modern extragalactic astronomy.

2. The spectroscopic study of the more prominent nuclei shows that there are processes within the nuclei differing from phenomena in other parts of galaxies. We are going to speak about these processes a little later. Let us mention here only some of them: violent motions of gaseous clouds, considerable excess of radiation in the ultraviolet, relatively rapid changes of brightness, expulsion of jets and condensations.

---

[1]Originally published by Pontifica Academia Scientarium © 1971 in *Nuclei of Galaxies, Pont. Acad. Scient. Scripta Varia*, 35, 1971, pp. 1–12. Used here with permission of Pontifica Academia Scientarium.

The presence of at least one of these processes is described by the term *activity* of the nuclei. There are cases when no starlike discrete image is seen at the center of a galaxy, but there are clear signs of nuclear activity. It is natural to assume in such cases that a nucleus exists, but its total luminosity in the visible light is so low that its image is not resolved due to the bright background of the ordinary stellar population. In such cases higher resolution could probably reveal the presence of a small nucleus.

Nevertheless, no sign of a nucleus has been found in some of the *nearest* galaxies. Examples: SMC and the Sculptor System. We can only speculate on the possible presence of nuclei in the past history of these systems or on remnants of what at some time was the nucleus of a galaxy.

At the same time the existing data make it almost certain that all spiral galaxies as well as all ellipticals of high and intermediate luminosity have nuclei of different prominence. Putting it in another way, we can say that almost all galaxies of high and intermediate luminosity have nuclei, but it is possible that a large proportion of dwarf galaxies are deprived of nuclei.

Of course we do not know exactly where the boundary lies between the galaxies with and without a nucleus. There is no clearcut border and the difference is only in the luminosity and significance of nuclei of different galaxies. In any case, here is a problem for study which is very difficult.

3. There is a degree of similarity between quasars or QSRSs and active galactic nuclei. Parallel to QSRSs, which are comparatively rare objects, we also observe optical QSOs. According to Sandage and his collaborators, the number of QSOs of *a given apparent magnitude* is more than a hundred times larger than the number of quasars of the same apparent magnitude.

The ratio is much higher when we take the corresponding spatial densities of the same objects. We know that quasars have optical (photographic) luminosities between -24 and -26 absolute. The QSOs apparently have a somewhat larger dispersion of luminosities and their average luminosity must be of the order of -23. This makes it very probable that the spatial density of QSOs is more than one thousand times higher than that of quasars. This means that the optical QSOs are in some respects much more important than quasars. We can formulate the situation in the following way:

The QSOs have considerable dispersion of luminosities. The most luminous of them are also emitting intense radio-frequency radiation and are known as Quasars.

Such a large population of QSOs in the Universe is evidence against their short lifetime. It seems to me that the assumption that on the average QSOs live less than 10^9 years will imply many difficulties. However, if we consider the state of quasars as a special active phase in the evolution of QSOs then it is possible to suppose that the total duration of such phase is much shorter (of the order of 10^7–10^8 years, but hardly less).

For a QSO or a quasar of high luminosity ($M = -25$) such a long lifetime means that the total amount of energy emitted in the form of electromagnetic radiation including the strong infrared radiation must be of the order of 10^{63} ergs, which is equivalent to about $5.10^8 M_\odot$. If we suppose that the lifetime of very *high-luminosity* QSOs is shorter than 10^9 years, then even in the case of objects with $M = -22.5$ the problem of the energy sources will remain.

4. One of the most important things to be done in studies of active nuclei and of QSOs is to understand the interconnection between the different forms of activity.

The study of radio galaxies has opened the way to discovery of the phenomena we are now discussing. The radio sources form only a small part both of galaxies with active nuclei and of QSOs. At least this is the case when we speak about strong radio sources. It is quite possible that all active nuclei emit in radio frequencies, but apparently we are not able to detect such weak sources.

The activity of many galactic nuclei manifests an excess of ultraviolet radiation of nonthermal and *nonstellar origin* and the presence of emission lines. In almost all cases the strong emission lines originate by means of fluorescence processes similar to those in gaseous nebulae of our own galaxy. Therefore, let us concentrate on the continuous emission in the ultraviolet. In the case of QSOs, due to very large redshifts the presence of ultraviolet excess is quite clear. However, for the majority of galaxies the total radiation of the nucleus is weak (both absolutely and apparently) and only a small part of galaxies shows an ultraviolet excess coming from the nucleus. A careful search for galaxies with bright ultraviolet continua covering several thousand square degrees has been made at Byurakan Observatory. It has been found that about 2% of galaxies in the interval of apparent magnitudes 13.5 to 17.5 have comparatively bright ultraviolet continua. About 400 of such "ultraviolet" galaxies have already been found by Markarian and a list comprising 300 galaxies with ultraviolet excess has already been published. About one hundred galaxies of that number had already been observed by other observers (Khachikian, Weedman, Sargent, Arakelian, Dibaj, Esipov), and it is now clear that not less than 80% of Markarian's galaxies have strong emission lines. Thus the observations support the idea that the strong emission lines are strongly correlated with ultraviolet excess.

There is every reason to believe that the near-ultraviolet excess observed in these galaxies expands to the far-ultraviolet, as in the case of QSOs and that there exists a maximum of spectral intensities in wavelength scale $[I(\lambda)]$. We connect this fact with the observations of galaxies made by orbital astronomical observatories launched by the Americans. They have indicated that some normal galaxies (for example M 31) show an increase of intensity to the far ultraviolet, which suggests a maximum of intensity beyond 2000 Å. We can guess that the nuclear region of every galaxy is a source of nonthermal and nonstellar radiation, which has its maximum in the far ultraviolet. What we observe from the earth's surface is only a relatively faint wing of this radiation. In the cases when the excess is *large* (as in the case of Seyfert or of some N-galaxies), we can detect this wing. However, in the majority of cases the ultraviolet excess is faint and its near-ultraviolet wing is still fainter and we cannot detect it.

If this extrapolation is valid, we may suppose that all nuclei emit this kind of ultraviolet radiation, but in galaxies with active nuclei such emission is much more intense. It seems, therefore, that the observation of radiation of nuclei in the far ultraviolet is rather important.

All these questions are connected with the problem of low-level activity of the nuclei of normal galaxies. But even in normal galaxies violent events occur from time to time. The Dutch astronomers have shown recently using 21 cm observations that there are outward motions of some isolated clouds directed from the nucleus of our Galaxy at a considerable angle to the galactic plane.

As regards the source of ultraviolet radiation, there is no doubt that this continuum usually comes from a source of small diameter (less than 10^{17} cm), and the characteristic irregular variations speak in favor of this. How can we explain these variations? If the mechanism of radiation is of synchrotron nature, then probably the variations of radiation intensity are caused by variation in the flow of particles which are ejected from a central body which has a still smaller volume.

The infrared emission represents another important form of activity of some nuclei. There is evidence that dust is present in the nuclei of some of Markarian's galaxies. However, the dust is not the real cause of infrared emission.

5. Another form of nuclear activity is the ejection of gaseous clouds. In the case of less active nuclei we have an apparently stable outflow of matter from the nucleus. The loss of mass by active nuclei can be estimated.

In the case of NGC 4151, Anderson and Kraft (*Ap. J.* 158, 859, 1969) have calculated that the loss of mass is somewhere between 10 and 1000 $M_\odot$ per year. If we suppose the duration of the Seyfert phase to be $5 \cdot 10^7$ years and take the lower value for the loss per year, we obtain the total loss of the order of $5 \cdot 10^8 M_\odot$. Thus the activity must be connected with great changes in the state of the nucleus.

Another example is NGC 1275. Apparently the giant filamentary gaseous structure which we observe in this galaxy has a mass of the order of several times $10^8 M_\odot$.

Thus the outflow of gas from nuclei, in the form of either clouds or shells, indicates essential evolutionary changes in the masses of nuclei. At the same time we must suppose that at the initial stage of evolution the mass of the active nucleus forms a considerable part of the mass of the whole galaxy.

We can only guess about the further history of the gas. If the motions are extremely violent (more than 1000 km/sec) the galaxy loses the gas. In the case of low velocity outflow the gas can form some system of clouds around the nucleus.

Perhaps the constant escape of mass from such galaxies as NGC 4151 and Markarian 9 is the cause of the extreme faintness of the envelope surrounding the nuclei of these galaxies.

By definition, for Seyfert galaxies the allowed lines are much wider than the forbidden lines (N_1 and N_2). However, if the emission line spectra of Seyfert galaxies are explained as radiation of many gaseous clouds ejected from the nucleus, then the spectral property just mentioned means that a considerable part of the emission line radiation comes from clouds of small masses. An expanding cloud of small mass can give an appreciable amount of radiation only when it is dense, since the luminosity is proportional to M^2/V.

But at high density and small volume the forbidden lines cannot appear. When the density diminishes due to expansion the total radiation is too faint to be observed. Thus in such clouds we do not see any forbidden lines. For the massive clouds the opposite is true. When we have only massive clouds, then we observe both the allowed and forbidden lines. Everything depends on the velocity of expansion of these large clouds. If they have low velocity of escape, they produce narrow forbidden lines. If their expansion velocity is high we must observe wide forbidden lines. Now it is important that *there is a group of galaxies which show both the allowed and forbidden lines equally widened.* Examples are Markarian 3, 6 and 39. But this is exactly opposite to the case of Seyfert-type spectra. At the same time the physical causes are the same. Only the values of the masses of the clouds are different. Thus many non-Seyfert galaxies have active nuclei of the same kind as the Seyfert galaxies.

Many galaxies with Seyfert spectra are similar in their structure to N-galaxies introduced by Professor Morgan. Therefore, it seems more appropriate to discuss their morphology in connection with morphological properties exhibited by N-galaxies. Professor Morgan has some important new ideas on this matter, and I hope that he will tell us more on this later. In this connection I would like to discuss only one point which has been emphasized by Markarian and Arakelian recently.

In his survey Markarian has divided all ultraviolet objects in two classes. First, the s-galaxies which are strongly concentrated objects of spheroidal form, which have a spectral distribution like QSOs. Second, the d-objects, which have diffuse borders; as sources of emission lines, they occupy large volumes in the corresponding galaxies.

The redshifts of 42 CS objects (concentrated, spheroidal) are known at this stage; only 25 of them have been measured photoelectrically. For the latter we can determine the absolute magnitudes.

For the mean absolute magnitude and color of CSOs Markarian and Arakelian give $M_B = -19.2$, $B - V = +0.57$, $U - B = -0.28$, compared with published mean values for N-galaxies $M_B = -21$, $B - V = +0.9$, $U - B = -0.27$.

The s-galaxies of Markarian, though generally much nearer to us than N-galaxies, as a rule were not observed as radio sources.

Therefore, we can say that the CSOs of Markarian together with N-galaxies form one major class of objects. Optically the most luminous of these objects often are radio galaxies. The number of CSOs in a given volume is several hundred times larger that the number of N-galaxies.

We have the same situation in the case of QSOs, QSRs, D – E galaxies and corresponding radio sources.

6. *Radio-frequency emission.* Here we have one of the most important problems. We understand that a strong radio-frequency emission is always connected with the activity of nuclei. But the nature of this connection must be different in different cases.

In the cases of Quasars and N-galaxies the connection is apparently a direct one since the optical radiation in these cases comes from the nucleus. In the case of D or E radio galaxies, the radio emission often comes from the clouds of relativistic gas outside the galaxy and the optical luminosity is caused by the light of the stellar population. In these cases the connection is indirect and the explanation should be in the interconnection between the nucleus and the stellar population of the whole galaxy. Any theory aimed at explanation of the activity of the nuclei and the origin of radio galaxies must explain these simple facts.

Another important question is formation of clouds of relativistic electrons. The models supposing that the clouds were ejected directly from the nucleus meet some difficulties. Alternatively, we can suppose that the clouds have been formed by coherent bodies ejected from the nucleus. In this case we must assume that each of these ejected bodies behaves as an active center emitting relativistic electrons. Here is a challenge for theoreticians.

7. *Dense bodies ejected from nuclei.* On several occasions I have had the opportunity to speak on jets originating in the nuclei of some giant galaxies. The galaxy NGC 4486 is only one example. The jets in NGC 3561 and IC 1182 are similar in form but consist mainly of classical gas. The condensations in these jets have strong ultraviolet excess and emission lines. In this respect they resemble compact galaxies with active nuclei. There are some other condensations of this type, which differ from these examples only by absence of the jet which connects the condensation with the nucleus of the primary galaxy. Some of them show spectra similar to that of condensation in the jets of the above-mentioned two galaxies.

It is possible to argue that we have no direct evidence that these objects (condensations) are of the same nature as active nuclei of galaxies. But it seems that such an argument is not too strong. If we observe a star which has the same spectral properties as the Sun, we easily assume that such a star is a body of the same kind as the Sun.

Therefore, we must consider it very probable that these small blue objects have at least some (if not all) properties of active nuclei of galaxies.

We know that a nucleus can develop a gaseous envelope around itself. In the case of our Galaxy we are almost certain that the interstellar gas enriches itself by the flow coming from the nucleus of the Galaxy. Therefore, it is quite natural that we observe emission lines in these condensations.

Thus we come to the idea of fragmentation of nuclei and formation of new galaxies.

8. In order to explain the inconsistently large masses of galaxies, which we obtain by applying the virial theorem to clusters and groups of galaxies, the suggestion was made that the clusters of galaxies originate by means of successive fragmentation of some initial body and that the corresponding clusters and groups are systems with positive energy. During the 15 years that elapsed after this suggestion several attempts were made to postulate the presence of some hypothetical matter (for example, neutral hydrogen) in the clusters. However, these attempts were not successful. Therefore, my original suggestion remains valid. I have nothing to add to the original arguments except for the fact that both the fragmentation concept and the concept of the activity of nuclei were closely connected and now, when the second concept has been confirmed by direct observation, it is time to discuss the fragmentation concept very carefully.

But if we ascribe some kind of universality to the activity of nuclei, I think we must admit that each nucleus builds an environment by means of activity.

In this case the formation of globular clusters and generally of the type II population is one of the kinds of nuclear activity. We can suppose the same regarding the origin of spiral arms.

I would like to emphasize here that the risk connected with such a hypothesis is much less now than when we had had no idea about the energetics of nuclei. Usually the kinetic energy of the stars in a galactic system is of the order of 10^{59} ergs. Energies higher by one or two orders of magnitude are sometimes released by the nuclei.

9. What we observe is only a number of external manifestations of the activity of some massive bodies which lie hidden in the very centers of the nuclei. The long duration of active processes in nuclei makes it quite clear that no processes of collapse or accretion can explain such continuous activity.

At the present stage we know almost nothing about these central bodies. The only thing which is certain is that they are capable of producing very large amounts of energy in the form of both discrete portions and of continuous flow.

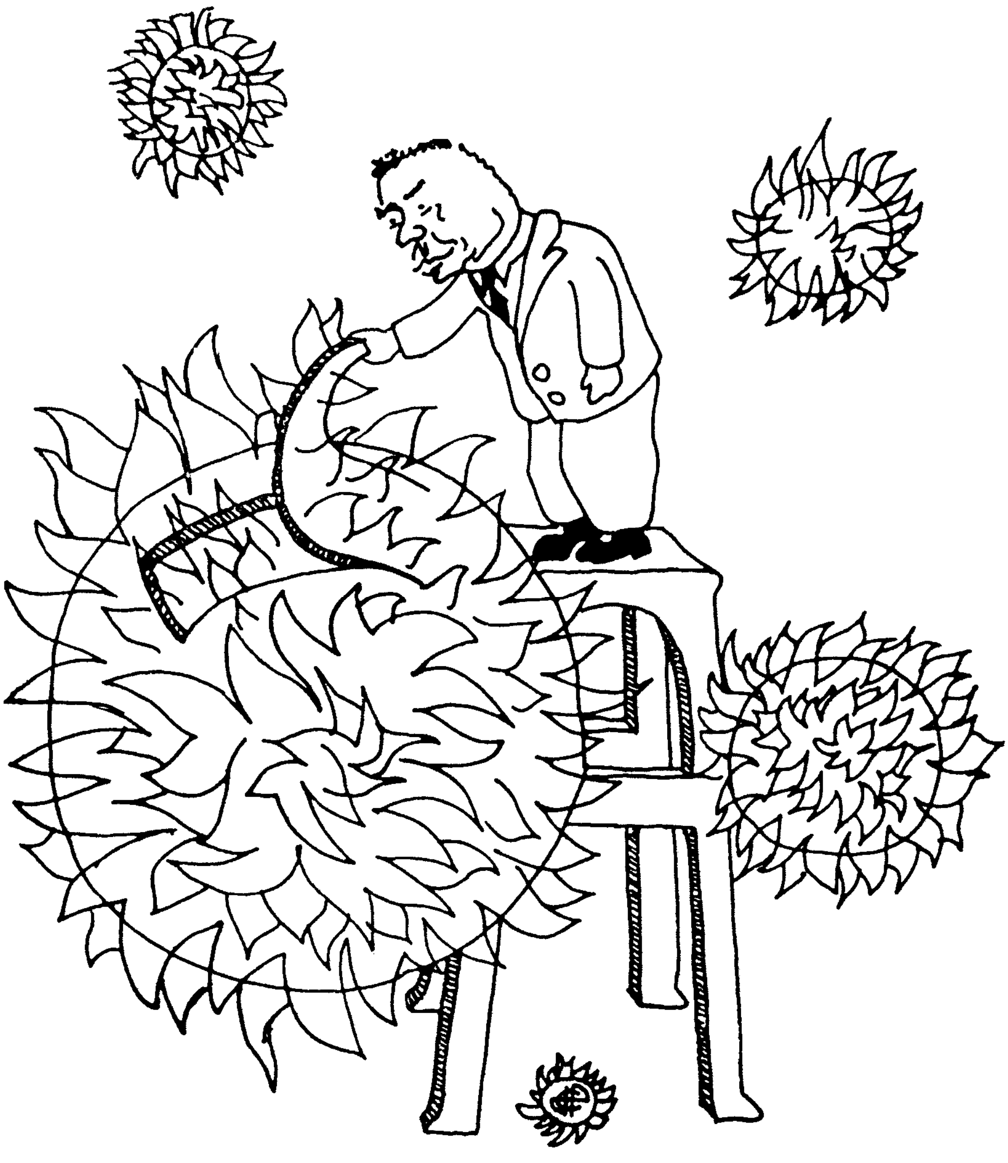
These bodies are apparently unstable, they change their physical state easily, but at the same time persist for a very long time. They sometimes eject great masses of the order of $10^8 M_\odot$, but after such ejection they continue their activity, perhaps in a less pronounced way. These central bodies of nuclei of galaxies and of QSOs represent a challenge to theoreticians.

10. As usually happens in astronomy when great new discoveries are made, theoreticians try to explain the new facts almost immediately. This time, however, we deal with very complex phenomena. It is even difficult to understand what is going on in the external parts of a nucleus which are transparent and open to observations. Therefore, some patience is necessary.

At this stage the problem is to better understand the external manifestation of the activity of nuclei and to obtain a correct general picture of the processes. Then will come the second stage when the theoreticians will give the explanation of the deep processes and of the physics of energy generation.

To make the first stage shorter we must emphasize observations and the systematization of the results of observations. The systematization and classification of objects is as important as the classification of relations we discover between different facts and forms of activity.

Nature is much more complicated and diverse than it seemed to us, who until recently had no idea about these wonderful processes. Let us study them with patience and base our conclusions mainly on the observational data.



Le principe d'invariance d'Ambartsumian.

Drawing by Jean-Claude Pecker.*

*From "La théorie des atmosphères stellaires: tendances actuelles" by Jean-Claude Pecker in *Problems of Physics and Evolution of the Universe*, Armenian Academy of Sciences Publishing House, Yerevan, 1978. Used with permission.

EPILOGUE[1]

During my studies at Leningrad University (1925–1928) I paid primary attention to astronomical and mathematical courses. Although I always realized the necessity of better knowledge of physics, at that time this discipline did not seem too attractive to me. The only exceptions I remember were quantum mechanics as well as some chapters in statistical physics. Now I feel that my knowledge of physics remained on a level insufficient for a theoretical astrophysicist.

Perhaps this circumstance, as well as lack of physical intuition, were the reasons that during the fifty years of my scientific work I concentrated mainly on directions where logical consistency is of greater importance than physical insight. At the same time, I have spent much time in the study of data obtained by observers.

Modern astrophysics deals with a great diversity and richness of observational data, with a huge variety of cosmic bodies and systems. At the same time, there is a great diversity of paths of scientific investigation and ways of thinking.

Nevertheless, my personal scientific efforts have been almost completely devoted to three main directions of theoretical work: (1) the invariance principles as applied to the theory of radiative transfer, (2) the inverse problems of astrophysics, and (3) the empirical approach to problems of the origin and evolution of stars and galaxies. In the following pages I present a short review of my results in these directions.

---

[1]Adapted from "On some trends in the development of astrophysics," published by Annual Reviews, Inc. in *Annual Review of Astronomy and Astrophysics*, volume 18, 1980, pp. 1-13. Adapted with permission from the *Annual Review of Astronomy and Astrophysics*, volume 18, © 1980 by Annual Reviews, Inc.

Invariance principles

The problem of scattering and absorption of light in a medium that consists of plane-parallel layers was considered in the classical works of Schwarzschild, Shuster, Eddington, Milne and Chandrasekhar. In essence, their method was based on a study of the balance of radiative energy in elementary volumes inside the medium. As a result the problem was reduced to a certain integral equation with an integral logarithmic kernel.

Even in the case of isotropic and monochromatic scattering the solution was not simple enough. But the general problem of anisotropic scattering with redistribution of frequencies (important for the theory of absorption lines) was connected with many more complications and difficulties.

As a university student, I tried to contribute to this field. My thesis was devoted to the integral equation of radiative equilibrium. However, the first essential results were achieved only in 1932–1933 when I developed a method of successive analysis of the Lyman-continuum and $\mathbf{L}_\alpha$ radiation fields in the problem of radiative equilibrium of planetary nebulae. I also found a simple approach to the monochromatic scattering problem in deep layers of a medium (for example, in deep layers of the ocean) with arbitrary indicatrix of scattering. But all this was within the framework of classical methods. Only in 1941 did I find other, much more effective, tools.

I am referring to the method of addition of layers, which is sufficiently described in the articles of the present book.

Because of the basic observation that the pattern of diffuse reflection from a semi-infinite layer remains unchanged when a supplementary layer is added to the boundary, the method was called the **invariance principle**.

In the simplest case of monochromatic and isotropic elementary acts of scattering the method enables one to replace the search for a family of solutions of a complicated linear integral equation by a numerical solution of a single and very simple nonlinear functional equation.

In my later work the invariance principle was applied to more complicated cases of finite optical thickness and anisotropic scattering.

In a wonderful way the principle of invariance has reduced the question of fluctuations of surface brightness of the Milky Way again to a very simple functional equation. The problem was treated further in a series of papers by S. Chandrasekhar and G. Münch in a much more complete form. No wonder the success of the method attracted the attention of researchers in adjacent fields. A far-reaching modification of the invariance principle was applied by Richard Bellman, under the name **invariant embedding,** to the solution of the most sophisticated problems of neutron transfer and others.

During the years after World War II, my younger colleagues attempted to apply the invariance principle to some *nonlinear problems* of radiative transfer. Some moderate success was achieved here too.

More recently, I learned that the invariance principle or invariant embedding was applied in a purely mathematical field of integral geometry, where it gave birth to a novel, combinatorial branch (see the book by my son R. V. Ambartsumian, *Combinatorial Integral Geometry* (John Wiley & Sons, Ltd., 1982)). The resources of the invariance principle seems to be immense indeed!

Inverse Problems

After my graduation from the University, my attention was attracted to the following question: to what degree does the totality of empirical data of atomic physics (the frequencies of spectral lines, the transition probabilities, etc.) define the system of laws and rules of quantum mechanics or, more specifically, the form of the Schrödinger equation? Very soon I came to the conclusion that rigorous solution of this problem was beyond my capabilities, and I decided to concentrate on a more modest problem of the same kind. For instance: *to what degree do the eigenvalues of an ordinary differential operator determine the functions and parameters entering into that operator?* I found that this problem was still too difficult. Finally, in 1929 I published in *Zeitschrift für Physik* a paper (the first paper in this collection) which contained the theorem that among all strings, the homogeneous string is uniquely determined by the set of its oscillation frequencies. Apparently during the fifteen subsequent years nobody took notice of that paper (when an *astronomer* publishes a *mathematical* paper in a *physical* journal he cannot expect to attract too many readers). However, beginning in 1944, that topic was developed by a number of outstanding mathematicians who have succeeded in obtaining many interesting results related to the "inverse Sturm-Liouville problem."

I myself tried persistently for many years to find other cases where one could directly derive natural laws from observational data or, as I now prefer to put it, to solve further inverse problems. There are many interesting examples of solutions of outstanding inverse problems in classical astronomy. The establishment of Kepler's laws of planetary motions from observations is an example of this. However, there had been rather few such cases in astrophysics.

In one of his popular papers Eddington asked the following question: is it possible to find the distribution function φ of the components of stellar velocities in the solar neighborhood *from radial velocities alone* without making any special assumption regarding the form of φ. This problem was solved in a paper that I wrote in 1935, which was presented by A. S. Eddington for publication in *Monthly Notices of The Royal Astronomical Society* (now in the present book).

It was shown in that paper that, mathematically, the problem reduces to the problem of finding the values of a function of three coordinates in the velocity space when the values of the integrals of this function over any plane in that space are given as a function of three parameters defining a plane. The problem was solved in a finite form, and the very first trials have shown the applicability of this method to the existing data on radial velocities. I think that now, when we have much richer catalogues of radial velocities, it is worthwhile to apply the solution again.

Can the problem of statistical evaluation of the number of flare stars in open clusters and associations be considered an inverse problem? My answer is "yes" and, in fact, any statistical problem I would attribute to this class. The philosophy of mathematical statistics is very close to and can be a clear illustration of the philosophy of inverse problems in general: given the observations, find the governing law.

The term "inverse problem" has now become increasingly fashionable in mathematical physics. One could expect that the twin expression "direct problem" should be used at least with the same frequency. However, this has not happened, and an obvious explanation for this asymmetry is that the authors do not feel the need to stress the direct nature of problems and avoid using the adjective "direct." Presumably, in a direct problem, given the governing law one tries to predict the result of observations.

The Empirical Approach to the Evolutionary Processes in the Universe

From the very beginning of my work in astrophysics I have been interested in the problems of the origin and evolution of stars and galaxies. It was clear to me that the old approach by means of global cosmogonic hypotheses could hardly bring serious results. It was clear that one must proceed from empirical data.

The evolutionary processes in the Universe are of an exceedingly complicated and diverse nature. Therefore, there is no chance of understanding them using a small number of speculative models or hypotheses. Instead of making more or less arbitrary assumptions, we must patiently analyze the empirical data and try to deduce from them conclusions on existing links between the evolutionary chains.

My idea was to find cases where it is relatively easy to deduce from the present state of an astronomical body or system the direction of its changes. In other words, I tried to find cases where we can conclude from simple considerations the evolutionary trend at a given phase, without the knowledge of all other phases. Of course, I do not claim this approach to be my invention. But I decided to follow this approach as strictly as possible.

I chose the *planetary nebulae*. For them Zanstra had concluded earlier that the only explanation of the unusual appearance of emission lines was expansion. My further studies soon made clear that the planetaries resulted from ejections from the outer layers of their central stars.

I analyzed the effect of interaction of members of a *stellar cluster* due to close passages during their motion. The inevitable conclusion was that the clusters are subject to the process of evaporation. In the case of open clusters, this process must be relatively rapid, having a time scale of the order of $10^8 - 10^9$ years. This is a short time compared to the time scale of the Galaxy.

Thus, it was shown that the open clusters that now exist in the Galaxy are relatively young and rapidly changing systems and that the general stellar field of the Galaxy is steadily growing in the number of stars at the cost of disintegration of clusters. At the same time, formation of clusters from individual field stars is practically impossible.

After World War II, I found that more extended groups of stars and of diffuse nebulae that have received the name *stellar associations* are much younger than the ordinary open clusters. They often contain hot giants (O and B stars) and always a large percentage of variable dwarfs (T Tauri variables and flare stars). The age of many associations is between 10^6 and 10^7 years. Their very existence proves two fundamental facts concerning the birth of stars in the Galaxy: (1) the formation of stars is a process continuing through the present epoch of the evolution of our Galaxy, and (2) the formation of stars proceeds in relatively large groups (associations and clusters).

The subsequent discovery of the fact that stellar associations contain multiple stars of a special type — the so-called Trapezium-type systems — has shown that in the associations subgroups exist that are younger than the associations as a whole (age between 10^5 and 10^6 years).

In the 1930s I tried to study the statistics of the elements of orbits of *double stars* in the Galaxy to obtain some indication about the direction of their dynamical evolution. My final conclusion was that wider pairs are rapidly disintegrating. Therefore, the existence of some very wide pairs puts an upper limit on the age of the Galaxy at least in its present state. This limit is quite independent of any cosmological consideration and is of the order of 10^{10} years.

My task was not merely to avoid the use of arbitrary models but also to dispel some *superstitions* remaining from classical cosmogonies, which at first glance appear to be quite natural assumptions. Take the idea that in the first phase of any process of formation of astronomical bodies or systems we always have nebular matter. Even now this opinion prevails among many theoreticians. However, it is difficult to find direct evidence for such an assumption in observational data.

All kinds of nebulae (and not only planetary or cometary) in our Galaxy, as well as in the external galaxies, are in a state of rapid change. Their lifetimes must be orders of magnitude shorter than the lifetimes of the majority of stars or planets. The radio nebulae, the best example of which is the Crab nebula, are results of supernovae explosions; they all dissipate rapidly. There is much evidence of expansion of some massive diffuse nebulae. The same is true for the so-called *compact H II regions*. The fact is that almost everywhere we observe, directly or indirectly, the formation of nebulae by way of ejections from stars and their groups. But the evidence in favor of the opposite processes (collapses of nebulae, accretion of nebular material) is infrequent and at times very dubious. At least the present-day picture of the Universe is dominated by processes of explosions, ejections from massive bodies, and subsequent formation of such short-lived objects as nebulae.

One of the most intriguing questions about stellar associations is that some of them are *expanding* or contain expanding groups of stars. In our first papers on stellar associations (1947–1951) it was predicted that expansion is a general phenomenon among associations. Studying proper motions in the association Perseus II, Professor A. Blaauw confirmed its expansion. Later he found the expansion phenomenon in a part of the Scorpio-Centaurus associations. At the same time, in many other associations no appreciable expansion has been found. However, these negative conclusions are definitive only for a number of nearby associations. Therefore, there are only one or two cases where we certainly have no simple expansion phenomenon. At the same time, the existence of at least some expanding groups is evidence of some kind of explosion processes connected with the birth or with the early stage of evolution of young stellar groups. Here again, the empirical data do not favor the theories of condensation of diffuse matter into the stars.

In the years 1955–1965 my attention was attracted by phenomena in and around the nuclei of galaxies. In the past, astronomers and, particularly, theoreticians showed little interest in the properties of the nuclei of the galaxies. In a report delivered to the Solvay Conference of 1958, I showed that these nuclei are often centers of large scale *activity* which proceeds in different forms. I suggested that the radio galaxies are not the products of collisions of galaxies, as was accepted at that time, but are systems in which ejections from the nuclei of tremendous scale take place. As a consequence of such ejections, clouds of high-energy particles are formed.

The subsequent discovery of *quasars* added one more form of nuclear activity by which a considerable part of liberated energy is emitted as the nonstellar optical radiation of the nucleus. In such cases, the luminosity of the nucleus often exceeds 10^{11} or 10^{12} times (sometimes even more) the luminosity of our Sun.

In another important development, the astronomers B. Markarian, E. Khachikian, and others who worked with me at Byurakan Observatory initiated a more systematic observational study of the optical manifestations of activity in galaxies such as ultraviolet excess and strong emission lines. One result of this work was a tenfold increase in the number of known Seyfert galaxies.

At the symposia organized in 1966 at Byurakan Observatory and in 1970 at the Vatican Academy of Sciences, different forms of activity of nuclei, including the phenomena in QSOs and in Seyfert galaxies, were thoroughly discussed. Since then a huge volume of observational work has been carried out. However, there has been little progress in theoretical interpretation as yet.

While the observed forms of the activity of nuclei speak directly in favor of the fundamental nature of explosion and expansion processes taking place in the central parts of galaxies, many theoreticians are still constructing models of nuclear phenomena in which the ejection processes are preceded by some form of collapse of great amounts of diffuse matter. According to such models, the ejections are only the secondary consequences of more fundamental processes of collapse. It is hardly necessary to say that I am very skeptical about such a mode of thinking. There is no evidence even for the possibility of such a course of events. It seems that such an approach is a remnant of the old notion that the evolutionary processes in the Universe are always proceeding in the direction of contraction and condensation.

Almost all the new interesting discoveries, which were extremely numerous during the last three decades, proved to be great surprises for existing formal theoretical models. Let me mention two cases of complete failure of the speculative approach.

(1) The existing theories have completely failed to predict such an important phenomenon as flare stars. There is no doubt now that the majority of stars after the period of their formation (T Tauri stage) go through this phase of evolution. Therefore, one of the first tasks of every evolutionary theory must be to explain of features of the flare processes.

(2) The situation is even worse with *fuors* (I use this term for stars of

FU Orionis type). The fuor stage is important in the life of at least some categories of stars, and this fact was rather fatal to many theories. But the situation is even more serious than it may seem at first glance. It now appears that there is a whole sequence of types which in their photometric behavior are more or less similar to fuors. The P Cygni star, which brightened almost four centuries ago, is an example. It is well known that in every spiral or irregular galaxy we have many supergiants of P Cygni type. Doubtless, brightenings of pre-P Cygni-type stars are significant in the evolution of supergiants.

It is natural to try to uncover the secrets of nature by *observing* the key points where they are hidden. We can hardly achieve this aim only by theorizing. Observations produce almost innumerable evidence in favor of ejections and explosions and are rather scanty regarding the processes of condensation and collapse. The facts are pronouncing an indictment against the ideas connected with the condensation process: in the observable Universe the opposite phenomena, i.e., expansion and diffusion, are responsible for the majority of changes now taking place.

Yerevan, 1994