ACTA PHYSICA

ACADEMIAE SCIENTIARUM HUNGARICAE

EDITORIAL BOARD
R. GÁSPÁR, K. NAGY, L. PÁL, A. SZALAY, I. TARJÁN

EDITOR-IN-CHIEF
I. KOVÁCS

VOLUME 53 NUMBERS 1–2

AKADÉMIAI KIADÓ, BUDAPEST
1982

ACTA PHYS. HUNG. APAHAQ 53 (1–2) 1–300 (1982)

ACTA PHYSICA

A JOURNAL OF THE HUNGARIAN ACADEMY OF SCIENCES

EDITED BY

I. KOVÁCS

Acta Physica publishes original papers on subjects in physics. Papers are accepted in English, French, German and Russian.

Acta Physica is published in two yearly volumes (4 issues each) by

AKADÉMIAI KIADÓ

Publishing House of the Hungarian Academy of Sciences

H-1054 Budapest, Alkotmány u. 21

Subscription information

Orders should be addressed to

KULTURA Foreign Trading Company

1389 Budapest P.O. Box 149

or to its representatives abroad.

ACTA PHYSICA

ACADEMIAE SCIENTIARUM HUNGARICAE

EDITORIAL BOARD
R. GÁSPÁR, K. NAGY, L. PÁL, A. SZALAY, I. TARJÁN

EDITOR-IN-CHIEF
I. KOVÁCS

VOLUME 53

AKADÉMIAI KIADÓ, BUDAPEST
1982

ACTA PHYS. HUNG.

CONTENTS

Volume 53

70th birthday of Professor Imre Tarján 3

NUCLEAR PHYSICS

L. Lovas, M. Rogge, U. Schwinn, P. Turek, D. Ingham and Erika Lovas: A Search for Giant Resonances by Inelastic Proton Scattering on ^{24}Mg 217

ATOMIC AND MOLECULAR PHYSICS

D. Aslanian, Gy. Rontó and K. Tóth: Raman Study of Isolated and "in situ" T7 Phage DNA: Conformation and Possible Interaction with the Proteins 25

L. Szalay, E. Tombácz, Z. Várkonyi and Á. Faludi-Dániel: Detergent Effects on an Albumin-Chlorophyll Complex, Model of Photosynthetic Protein-Pigment Complexes 225

S. Damjanovich and J. Szöllősi: Molecular Spectroscopy as a Tool in Studying Biophysical Problems 237

R. Gáspár and Á. Nagy: Comparison of the Orbitals of Neon, Argon and Krypton Calculated by the Hartree—Fock and the Xα Methods with Several Values of α 247

CONDENSED MATTER

R. Voszka, G. Gévay, I. Földvári and S. Keszthelyi-Lándori: Growth and Characterization of $Bi_4Ge_3O_{12}$ Single Crystals 7

A. Mecseki, I. Földvári and R. Voszka: Mathematical Statistical Method for Fitting the Urbach Parameters of TeO_2 15

Z. Morlin: Some Group Theoretical Remarks on the Symmetry of Polymorphic Crystals 33

A. Watterich and K. Raksányi: Model for the Mg_i Centre in NaCl 39

J. Janszky and Yu. Ya. Yushin: On the Quantum Theory of Parametric Frequency Converter .. 51

L. Malicskó: On the Role of Dislocation Groups in the Formation of Dominating Growth Hills 59

L. Malicskó, L. Jeszenszky, E. Hartmann, L. Kovács, Á. Péter and B. Vajna: The Effect of Dopants on the Growth and Properties of TGS Crystals 67

A. Tóth, P. Kálmán, T. Keszthelyi and J. Sárközi: Low Temperature Yield Stress Increment Due to Impurity-Vacancy Dipoles in NaCl Crystals 75

J. Sárközi, A. Tóth, K. Orbán, T. Keszthelyi and P. Kálmán: The Effect of Surface Processes on the Flow Stress of Doped NaCl Crystals Heat Treated in Air 81

T. Siklós: The Anharmonic Linear Chain 89

B. K. Vainshtein, L. A. Feigin and D. I. Svergun: The Investigation of the Structure of Biomolecules by Means of Small-Angle Scattering 105

L. M. Beljaev, V. V. Nabatov and Yu. Ya. Yushin: On the Kinetics of Growth of Conductivity Electron Concentration in Dielectric Crystals Irradiated by Intensive Light 121

M. Hársy, F. Koltai, I. Gyuró, T. Görög and E. Lendvay: The Growth and Texture Study of GaSb Ingots 133

I. Kirschner: Preface to Cryosem 81' 303

K. Fischer, S. Rohr, J. Mönch, W. Holzhäuser and Ch. Frenzel: V_3Ga Superconductors with a High Critical Current Density 305

I. Kirschner and K. Sajó: Determination of Microscopic Parameters by Macroscopic Measurements in Superconductors 313
A. Pevala: Application of the "Boson Method" to the Study of Superconducting Order Parameter 321
J. Hoellrich and G. Kath: Investigation of the Influence of the Heat Treatment and Forming Regimes on the Critical Current and Pinning Force of NbTi-Superconductors 331
S. Takács: Statistical Theory of Pinning in Superconductors Including Correlations in the Defect Distribution 337
I. Kirschner and K. Martinás: The Pinning of Vortex Lines in Very Dilute Superconductors 347
J. Bánkuti and É. Farkas: Experimental Determination of Pinning Forces 353
P. Debreceni, I. Kirschner, M. F. Kotkata, G. Kovács and T. Porjesz: On the Exact Calculation of the Specific Electrical Conductivity of Semiconducting Slabs 361
A. Handstein, J. Schneider and J. Kadlecová: Temperature and Magnetic Field Dependence of Magnetization and Magnetoresistance in Amorphous Fe – Ni Based Alloys 371
A. Gladun, C. Gladun and H. Vinzelberg: Thermal Conductivity of Niobium and Molybdenum Single Crystals at Low Temperatures 383
P. Debreceni, I. Kirschner and T. Porjesz: Crystal Symmetry Effects of the Conduction Tensor . 395
G. Pompe, M. Gaafar, M. Falz and P. Büttner: Thermal Conductivity of Amorphous Iron Alloys at Low Temperatures 401
T. Porjesz: New Perspective in Semiconductor Physics Based on Advanced Technology 411
K. Fröhlich, D. Machajdik, S. Benacka and V. Cernusko: On the Relation between T_c Measurements and Structure Inhomogeneity of Nb_3Ge Superconductor 419
J. Kruzliak, P. Hutka, P. Kovác and P. Setina: Some Physical Properties of Multifilamentary Nb_3Sn Superconductors Prepared by Bronze Technology 425
F. Gömöry: AC Loss Measurement of Superconducting Magnets. Influence of the Compensating Scheme 433
J. Bánkuti and L. László ffy: Superconducting Laboratory Magnets 445
A. Juhász, I. Fözy, P. Tasnádi, I. Kovács and I. Vitányi: Grain Boudary Motion in Vanadium .. 455

INTERDISCIPLINARY

L. Keszthelyi, P. Ormos and G. Váró: Fast Components of the Electric Response Signal of Bacteriorhodopsin Protein 143
J. Bacsó, M. Kis-Varga, P. Kovács, J Pálvölgyi, D. Berényi, J. Hideg, R. A. Tigranjan and T. A. Viting: Investigation of Change of Mineral Metabolism of Cosmonauts by X-ray Fluorescence Method 159
Cs. M. Buczkó, J. Csikai and J. Nemeskéry: Application of Nuclear Methods in Paleosciences 165
L. Bozóky: Investigations into the Mechanism of Action of Ionizing Radiation 179
L. B. Sztanyik and I. Nikl: Contribution of Cosmic Rays to Radiation Exposure of the Population 189
G. Varjas and O. Haiman: The Dose Gradient Vector Concept and its Usefulness in Planning Dose Distributions for Radiation Teletherapy 201
G. Müller, D. Noack, R. Schorcht, S. Gáspár and L. Herényi: Mathematical Modelling of Segregation Processes in Microbial Populations Containing a Single Plasmid Species ... 255
F. Deák and G. Marx: Computer Models for Some Aspects of Evolution 263
L. Pál: Mysteries of Noise 279

GEOPHYSICS

A. Szalay: Gas Carrier and Degassed Magmatic Rocks and the Primordial Atmosphere 291

CONTENTS

70th birthday of Professor Imre Tarján 3

NUCLEAR PHYSICS

I. Lovas, M. Rogge, U. Schwinn, P. Turek, D. Ingham and *Erika Lovas:* A Search for Giant Resonances by Inelastic Proton Scattering on ^{24}Mg 217

ATOMIC AND MOLECULAR PHYSICS

D. Aslanian, Gy. Rontó and K. Tóth: Raman Study of Isolated and "in situ" T7 Phage DNA: Conformation and Possible Interaction with the Proteins 25

L. Szalay, E. Tombácz, Z. Várkonyi and Á. Faludi-Dániel: Detergent Effects on an Albumin-Chlorophyll Complex Model of Photosynthetic Protein-Pigment Complexes 225

S. Damjanovich and J. Szöllősi: Molecular Spectroscopy as a Tool in Studying Biophysical Problems 237

R. Gáspár and Á. Nagy: Comparison of the Orbitals of Neon, Argon and Krypton Calculated by the Hartree—Fock and the Xα Methods with Several Values of α 247

CONDENSED MATTER

R. Voszka, G. Gévay, I. Földvári and S. Keszthelyi-Lándori: Growth and Characterization of $Bi_4Ge_3O_{12}$ Single Crystals 7

A. Mecseki, I. Földvári and R. Voszka: Mathematical Statistical Method for Fitting the Urbach Parameters of TeO_2 15

Z. Morlin: Some Group Theoretical Remarks on the Symmetry of Polymorphic Crystals 33

A. Watterich and K. Raksányi: Model for the Mg_1 Centre in NaCl 39

J. Janszky and Yu. Ya. Yushin: On the Quantum Theory of Parametric Frequency Converter .. 51

L. Malicskó: On the Role of Dislocation Groups in the Formation of Dominating Growth Hills 59

L. Malicskó, L. Jeszenszky, E. Hartmann, L. Kovács, Á. Péter and B. Vajna: The Effect of Dopants on the Growth and Properties of TGS Crystals 67

A. Tóth, P. Kálmán, T. Keszthelyi and J. Sárközi: Low Temperature Yield Stress Increment Due to Impurity-Vacancy Dipoles in NaCl Crystals 75

J. Sárközi, A. Tóth, K. Orbán, T. Keszthelyi and P. Kálmán: The Effect of Surface Processes on the Flow Stress of Doped NaCl Crystals Heat Treated in Air 81

T. Siklós: The Anharmonic Linear Chain 89

B. K. Voinshtein, L. A. Feigin and D. I. Svergun: The Investigation of the Structure of Biomolecules by Means of Small-Angle Scattering 105

L. M. Beljaev, V. V. Nabatov and Yu. Ya. Yushin: On the Kinetics of Growth of Conductivity Electron Concentration in Dielectric Crystals Irradiated by Intensive Light 121

M. Hársy, F. Koltai, I. Gyuró, T. Görög and E. Lendvay: The Growth and Texture Study of GaSb Ingots 133

INTERDISCIPLINARY

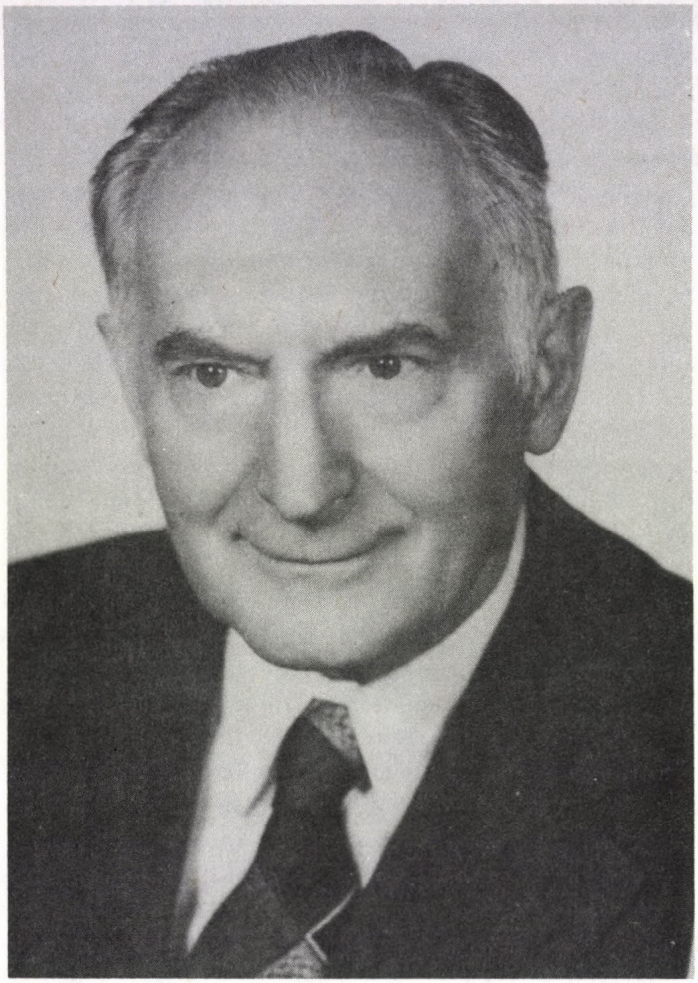
L. Keszthelyi, P. Ormos and G. Váró: Fast Components of the Electric Response Signal of Bacteriorhodopsin Protein 143

J. Bacsó, M. Kis-Varga, P. Kovács, J. Pálvölgyi, D. Berényi, J. Hideg, R. A. Tigranjan and *T. A. Viting:* Investigation of Change of Mineral Metabolism of Cosmonauts by X-ray Fluorescence Method 159
Cs. M. Buczkó, J. Csikai and J. Nemeskéry: Application of Nuclear Methods in Paleosciences . 165
L. Bozóky: Investigations into the Mechanism of Action of Ionizing Radiation 179
L. B. Sztanyik and I. Nikl: Contribution of Cosmic Rays to Radiation Exposure of the Population 189
G. Varjas and O. Haiman: The Dose Gradient Vector Concept and its Usefulness in Planning Dose Distributions for Radiation Teletherapy 201
G. Müller, D. Noack, R. Schorcht, S. Gáspár and *L. Herényi:* Mathematical Modelling of Segregation Processes in Microbial Populations Containing a Single Plasmid Species 255
F. Deák and G. Marx: Computer Models for Some Aspects of Evolution 263
L. Pál: Mysteries of Noise 279

GEOPHYSICS

A. Szalay: Gas Carrier and Degassed Magmatic Rocks and the Primordial Atmosphere 291

The final manuscript was received by the Editorial Office on 25 February 1982.



70TH BIRTHDAY OF PROFESSOR IMRE TARJÁN

On 26th July 1982 Professor Imre Tarján, Member of the Hungarian Academy of Sciences and President of its Section for Mathematical and Physical Sciences, celebrated his 70th birthday.

Professor Tarján is a scientist who has achieved outstanding results in two apparently largely different fields; moreover, in both fields he founded scientific schools. One of his fields is solid state physics, or more precisely, crystal physics; the other is molecular biophysics. Professor Tarján was among the first to recognize and to carry out pioneering work in order to develop the idea that in biological

macromolecular systems as well as in solids a basic understanding of the physical properties and biological functions can be achieved only by a thorough knowledge of the atomic and molecular order and of the structural defects acting against this order.

While still a student of Zoltán Gyulai, in the thirties he began to investigate structural defects generated by ionizing radiation, especially by X-rays, in alkali halide crystals. Developing these studies he made conclusions concerning the mechanism of generation and the structure of crystal defects, their interaction with other lattice defects (dislocations, impurities), and he pointed out the role of the defect structure in bringing about some macroscopic properties. He was successful together with his co-workers in producing extremely pure alkali halide crystals; this success became the starting point for obtaining further interesting results both in Hungary and abroad.

One of his outstanding results was realized in the early fifties, when, together with Zoltán Gyulai, he was among the first in the world to grow artificial quartz crystals. At about the same time he and his co-workers produced NaI(Tl), anthracene, naphthalene and other single crystals for the detection of nuclear radiation.

In the mid-sixties he extended methods and approaches usually applied in solid state physics to the investigation of biological macromolecular systems, e.g. he was one of the first to use stochastic models to characterize processes of photodamage in the nucleoproteins of bacteriophages and also to describe the interaction between ions or antibiotics and the membranes of bacteria. One of the conclusions of this work was that the protection against photodimerization of a nucleic acid having a double-helix structure is linked with particularities of this structure.

Professor Tarján always considered the application to practical purposes of results obtained in basic research as a most important aspect of his work. He is one of the first representatives, on an international scale, to advocate application-oriented crystal growth. He contributed towards the foundation of nuclear medicine in Hungary and has been responsible for many of its developments. Two prime examples of the latter were the transfer to industry of the technology developed together with his co-workers for producing NaI(Tl) single crystals for gamma-ray detection, and the elaboration of models for a number of instruments in nuclear medicine. A result of particular importance obtained in recent years by him and his team in applied biotechnology is represented by a process developed for fast quantitative characterization of the mutagenic activity of chemicals and automatic equipment for its realization; this is applicable in environment protection in the pharmaceutics industry, in the food industry and in agriculture.

Scientific workers and physicians have good reason to be grateful to Professor Tarján for his outstanding role in their education. He founded one of the bases of crystal physics in Hungary, the Research Laboratory for Crystal Physics of the Hungarian Academy of Sciences; the leading scientific staff of this laboratory know him as one of their former teachers. In the field of biophysics he had a pioneering role in the organization of interdisciplinary teams and in finding and shaping the ways and the attitudes in the collaboration of variously trained research workers. He was

responsible for the founding of the Research Laboratory for Biophysics of the Hungarian Academy of Sciences where the leading scientists — experts in biology, physics and chemistry — were also his students. In the field of medical education he not only participated in the training of generations of physicians by inspiring an exact scientific way of thinking but also developed an internationally accepted system and textbook for the biophysical education of medical students.

Professor Tarján's 70th birthday should be seen purely as a milestone, since his rich oeuvre is continually being augmented by his activities in biophysical education, in the research of biological macromolecules (as partly ordered systems) and in scientific public life.

His colleagues and students, Hungarian physicists and biophysicists, together with the Hungarian scientific and medical communities wish him good health and creative activity in the coming years.

György Rontó *Rudolf Voszka*

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 7—13 (1982)

GROWTH AND CHARACTERIZATION OF $Bi_4Ge_3O_{12}$ SINGLE CRYSTALS*

R. VOSZKA, G. GÉVAY, I. FÖLDVÁRI

RESEARCH LABORATORY FOR CRYSTAL PHYSICS OF THE HUNGARIAN ACADEMY OF SCIENCES
1502 BUDAPEST, HUNGARY

and

S. KESZTHELYI-LÁNDORI

GAMMA WORKS, BUDAPEST, HUNGARY

Experiences obtained in the growth of $Bi_4Ge_3O_{12}$ single crystals together with characteristic data on the crystals are described. The tail of the self-absorption near the absorption edge is shown to depend sensitively on crystal quality. This property can be used for crystal characterization.

1. Introduction

In the fifties one of the predecessors of our Laboratory was engaged in developing NaI(Tl) scintillators, and introduced the manufacturing of these crystals at the Gamma Works, where they are now being produced on large scale in excellent quality. However, the techniques of nuclear measurement are steadily developing and require scintillators of various properties. Miniaturization raised the need for scintillators having large atomic numbers. Furthermore, the development of computed tomography (CT) demands scintillators of short afterglow in order to increase the signal processing rate. In Table I we have selected the most important scintillators and presented some of their characteristic data. The first column presents the density, the second the effective atomic number, the third the index of refraction (measured at the wavelength of the emission maximum), the fourth the afterglow (which is the emitted energy after the first 3 milliseconds relative to the total emitted energy of scintillation), the fifth the wavelength of maximum emission, the sixth the decay time (defined as the time in which the decreasing light intensity attains the e-th fraction of the initial intesity), finally in the last column the light output relative to that of NaI(Tl) is presented. These data clearly show the advantage of bismuth germanate manifested by its great density, large effective atomic number and short afterglow. The density and atomic number determine the high absorption coefficient. This enables one to decrease 9—16 times the volume of the crystal which is of considerable advantage for example in

* Dedicated to Prof. I. Tarján on his 70th birthday.

Table I
Characteristic data of scintillators

	Density [g/m^3]	Z_{eff}	n at λ_{Emax}	Afterglow % >3 ms	$\lambda_{E_{max}}$	Decay time [μs]	Rel. light output [%]
NaI(Tl)	3.67	49.8	1.85	0.5–5	410	0.23	100
CaF(Eu)	3.18	16.6	1.44	0.3	435	0.94	50
CsJ(Tl)	4.51	54.0	1.80	0.5–5	565	1.00	45
$Bi_4Ge_3O_{12}$	7.13	71.7	2.15	0.005	480	0.30	8
$CdWO_4$	7.90	61.2	2.3	.0.005	540	5	38

geophysical measurements. Taking into account that bismuth germanate, because of the $^3P_1 \rightarrow {}^1S_0$ transition of the Bi^{3+} ion, is an intrinsic scintillator, no problems arise due to activator distribution. Moreover, the crystal is chemically inert, nonhygroscopic, its afterglow is short, consequently it is reasonable to state that bismuth germanate is one of the best scintillators for CT application.

2. Products of crystals

The first paper on bismuth germanate crystal growth was published by Nitsche [1] in 1965. His crystals were still yellow. After this growth work started at several places and in 1971 Philipsborn [2] reported already on colourless crystals. In 1975 Nestor and his coworkers [3] had measured the scintillator properties of the crystal that led to CT applications. Literature is poor in describing growth techniques of practically important materials, so we have made some effort to arrive at a result in a short time by realizing our own conception.

The tasks that had to be solved are summarized as follows:

a) First it was necessary to test the basic materials of various purity and origin. We have found that in order to get colourless crystals substances of the purity of at least 5N are required. However, in order to preserve the Pt crucible the use of Bi_2O_3 and GeO_2 of 6N purity is more advantageous. With 4N purity the crystals have a yellow tinge.

b) It became necessary to develop the optimal rate of the solid state reaction in order to maintain stoichiometry. This was obtained by a special temperature program.

c) In connection with the Czochralski method a series of furnace constructions had to be tested and for every furnace construction the optimal rate of pulling and rotation had to be established. This was rather difficult because in the literature one finds extreme values ranging from 0 to 100 rpm. It was found that at medium rotation rates the crystal tends to form inclusions due to the flow system appearing in the melt. Further on it has also been established that inclusions diminish at low rotation rates.

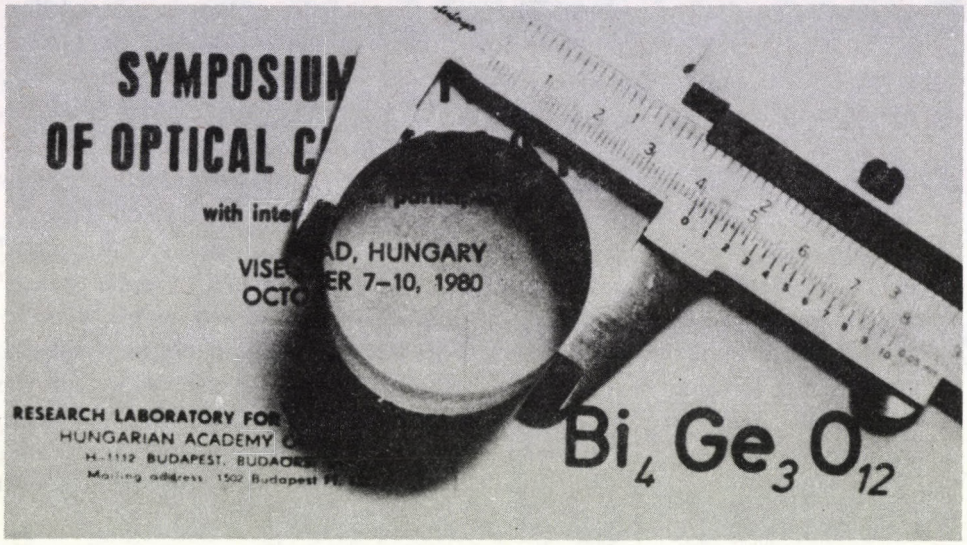


Fig. 1. $Bi_4Ge_3O_{12}$ single crystal

(Transition to Kyropoulos growth). They also diminish on passing over to higher rotation rates where the convexity of the growing front decreases. The crystals were generally grown on seeds of ⟨100⟩ orientation.

d) Most of our experiments have been carried out by means of some temperature control. It was necessary to work out an optimal program for a given furnace construction in order to grow crystals close to the required uniform diameter. Since we did not always succeed we tried our invention developed for the growth of TeO_2 and to be patented presently, which is a simple balance control system. In this way we succeeded to pull $Bi_4Ge_3O_{12}$ single crystals with a uniform and predetermined diameter of 40 mm and a length also of 40 mm. The processed crystal is depicted in Fig. 1.

e) For scintillation purposes the crystals must be processed. It was found that the crystals can be processed by simple turning, whereas the optical window for the scintillation light output can be prepared by the usual polishing technique applied to glasses of smaller hardness. Finally the crystals are capsulated and provided with a light reflector in a way used for NaI(Tl).

3. Characterization of crystals

a) Visual observation

Observing visually the as-grown crystal boule after cooling some statements can be made. There are ridges on the crystal surface starting from the seed, their perfectness proves the single-crystal character of the crystal and their symmetry shows the crystallographic directions. The ridges tend to extend to facets, so that the crystal takes

a typical habit, especially if the thermal gradient is small due to furnace construction. By visual observation the colour and inclusions of the crystal become directly apparent. Both are connected to impurities. Pulling with high rate gives rise to a very large number of inclusions so that the crystal is non-transparent, with a dark grey colour. Slow pulling results in a better purification effect, the crystal is generally colourless and transparent, however, because of the large index of refraction there is hardly any way to observe the inside of the crystal through its curved surfaces. Thus it is useful to polish two flat faces of optical quality on both ends of the crystal perpendicular to the rotation axis. In this way the inside of the crystal can be investigated. In some cases an ordered inclusion structure forms starting from the cylinder jacket around the rotation axis. This may be probably connected to the growth front having some unfavourable shape and to the flow structure of the melt. The melt frequently contains well observable rotating flow cells. We succeeded to produce inclusion-free crystals of small diameter only from materials purified by pre-pulling, using low pulling and high rotation rates.

b) Absorption measurement

Absorption has been measured using polished samples cut from crystals in dimensions of $10 \times 20 \times 5$ mm, near the selfabsorption edge applying a Perkin-Elmer 554 type spectrophotometer. In Fig. 2 the wavelength is recorded on the horizontal and the absorption coefficient on the vertical axis. The curves show rather well that the

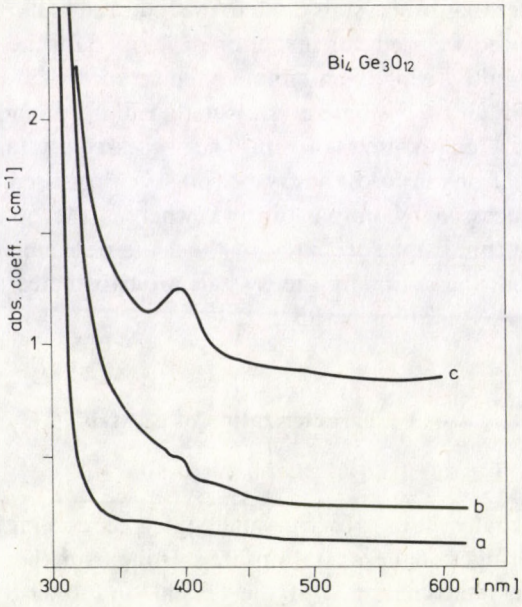


Fig. 2. Absorption curves

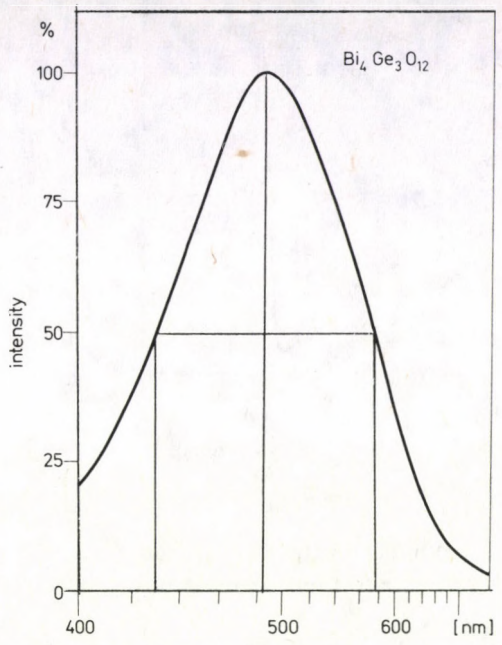


Fig. 3. Luminescence emission spectrum

absorption changes sensitively with impurities and inclusions. Curve "a" shows the absorption of an inclusion-free colourless crystal, curve "c" belongs to a crystal of yellow colour and containing inclusions, whereas curve "b" belongs to a crystal having intermediate properties. Attempts at the identification of the impurity responsible for the maximum at 400 nm have been up to now unsuccessful. Research work in this direction is still in progress.

c) Luminescence measurement

The luminescence emission spectrum has also been measured on samples prepared for absorption measurement using a measuring system developed in our Laboratory [4]. The crystals have been excited by X-rays. Fig. 3 shows a typical emission spectrum where the emitted intensity is given in relative units. The maximum is near to 480 nm. This is close to the value published in the literature (Table I). In the case of yellow crystals the maximum is shifted towards longer wavelengths, which can be partly explained by the self-absorption of such crystals.

d) Measurements of scintillator characteristics

The resolution of our best crystal of 25 mm diameter and 25 mm length has been measured in the Gamma Works using the 0.661 MeV line of ^{137}Cs. The spectrum is shown in Fig. 4. The channel number is recorded on the horizontal axis whereas the

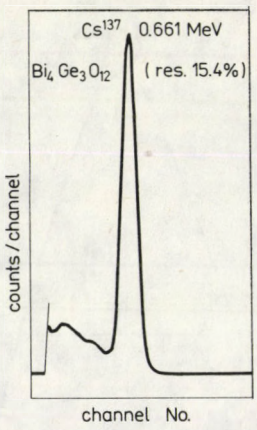


Fig. 4. The γ-spectrum of ^{137}Cs

vertical axis displays the number of impulses per channel. The resolution, which is defined as the halfwidth relative to the maximum position, is in this case 15.4%. The value given in the literature is 15% (Table I). The resolution is rather sensitive to changes of crystal quality. We have measured values as large as 43% on yellow crystals full of inclusions. The light output as related to NaI(Tl) has also been measured, in the case of our best crystal it proved to be 7.2% which is in good agreement with the value 8% given in the literature (Table I).

4. Conclusion

As shown by the above discussion bismuth germanate possesses properties making this material rather important from the practical point of view. Though activator distribution problems do not arise, experiments carried out so far indicate a strong inclination to form inclusions as well as a susceptibility to impurities, requiring a proper growth technology to be worked out reproducing the dimensions and quality of the crystals. Absorption measurements appear to be useful for the characterization of crystals because the tail of the self-absorption near the absorption edge shows sensitive changes due to impurities and inclusions. A separate investigation of the influence of various impurities on the scintillation properties would be of interest. A better knowledge of our material resulting from such investigations would help to produce more perfect crystals by either improving the purification methods or by applying dopants positively influencing the crystal properties.

References

1. R. Nitsche, J. Appl. Phys., **36** (8), 2358, 1965.
2. H. von Philipsborn, J. Cryst. Growth, **11,** 348, 1971.
3. O. H. Nestor and C. Y. Huang, IEEE Trans. Nucl. Sci., **NS-22,** 68, 1975.
4. A. Watterich and R. Voszka, Acta Phys. Hung., **33,** 323, 1973.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 15—23 (1982)

MATHEMATICAL STATISTICAL METHOD FOR FITTING THE URBACH PARAMETERS OF TeO_2*

A. MECSEKI, I. FÖLDVÁRI and R. VOSZKA

RESEARCH LABORATORY FOR CRYSTAL PHYSICS OF THE HUNGARIAN ACADEMY OF SCIENCES
1502 BUDAPEST, HUNGARY

The intrinsic absorption edge of TeO_2 was measured between RT and 20 K for the ordinary ray. A special statistical model was developed for fitting the experimental data. In this way a correct decision function was used avoiding the ad hoc separation of the Urbach equation. The obtained Urbach parameters are $E_0 = 4.254 - 4.257$ eV; $\hbar\omega_0 = 2.273 - 2.275 \times 10^{-2}$ eV, $\sigma_{S0} = 0.632 - 0.635$. The effectiveness of the method was demonstrated and the reliability of the parameters obtained was discussed.

Introduction

Nowadays the interest and requirements for laseroptic single crystals have steadily been increasing. Active optical elements like deflectors, modulators and tunable filters can be produced using the acousto-optic properties of the crystals. Presently in the visible region the paratellurite (TeO_2) is the preferred superior acousto-optic crystal. Its M_2 acousto-optic figure of merit (1.2×10^{-15} s^3g^{-1} for transversal acoustic waves) is the largest known value for optically transparent materials, and its other properties as chemical stability, mechanical durability, wide transmission range also suggest its application in acousto-optic devices.

To diminish the acoustic and optic losses the single crystals have to be prepared free from imperfections and impurities. The investigations have shown that both the acoustic attenuation and optical absorption of the TeO_2 crystals grown in our laboratory are lower than those published earlier [1]. The purity of the crystals allows to study their intrinsic absorption properties. There are two papers in the literature on the intrinsic absorption of TeO_2, however, one is confined only to room temperatures [2] and the other's data indicate a moderate quality of the crystal [3].

The shape of the absorption edge of most insulator crystals is exponential and shows a characteristic dependence on temperature which was first observed by Urbach on AgBr [4]. Acceptable explanation of the Urbach rule was given by Toyozawa [5] on the basis of exciton-phonon interaction theory. The long wavelength side of the exciton band is described by

$$\mathbf{A} = A_0 \exp\left[-\sigma_s(T) \cdot (E_0 - E)/k_B T\right], \tag{1}$$

* Dedicated to Prof. I. Tarján on his 70th birthday.

where A [cm] is the absorption coefficient versus photon energy E [eV]. E_0 is practically the same as the maximum of the lowest energy exciton transition at $T=0$ K. As the shape of the exciton band near its top does not obey the Urbach rule, A_0 is only an extrapolated value. The exciton-phonon interaction is taken into account by $\sigma_s(T)$ according to Eq. (2):

$$\sigma_s(T)=\sigma_{s0}\cdot\frac{2k_BT}{\hbar\omega_0}\cdot\tanh\frac{\hbar\omega_0}{2k_BT}, \tag{2}$$

where σ_{s0} is a material constant, ω_0 is the average frequency of the optical phonons.

Determining the Urbach parameters from the experimental data by fitting seems to be a complicated mathematical problem because

a) the dependence of the function value on some of the parameters is rather weak;

b) the equation cannot be transformed into a linear one for all parameters;

c) the function values are spread over some orders of magnitude.

For solving this problem a special fitting procedure was applied, the effectiveness of which had previously been checked for similar tasks [6].

Experimental

The TeO_2 single crystals were grown by the Czochralski technique. The purification of the raw material and the crystal growth are discussed in detail elsewhere [7, 8].

The oriented crystals were cut and the destroyed layer was removed by chemical polishing, then the surfaces were high-polished mechanically. For spectroscopic investigations samples with their polished planes perpendicular to the optical axis ⟨001⟩, were used. To cover the possible absorption range the sample thickness was varied from 0.3 to 15 mm.

The spectroscopic investigations were performed on a Perkin-Elmer 554 spectrophotometer equipped with an Air Products Displex cryostat. Under the circumstances used a vapor film originating from the diffusion pump oil was produced only during 20 K measurements. Following the kinetics of vaporization the oil-free absorption could be extrapolated at this temperature, too. The multiple reflection from the surfaces was taken into account by the equation

$$I=I_0=\frac{(1-R)^2e^{-Ax}}{1-R^2e^{-2Ax}}, \tag{3}$$

where I_0 and I are the incident and transmitted intensity, respectively, and x [cm] is the sample thickness. The reflectivity data (R) were taken from the results of Takizawa [3]. The absorption coefficient values were determined by an iteration method from Eq. (3) using a minicomputer program.

Model for fitting

The sample values measured are used to obtain unknown parameters by fitting. Instead of using ad hoc procedures [9] it is much more useful to apply the well elaborated general statistical method [10, 11]. So we have to specialize the following general notions: the elementary event space $\mathbf{\Omega}$, the σ-algebra of events $\mathbf{S}$, the parameter space $\mathbf{\Theta}$, the set of probability measure $\mathbf{P} := \{P_\vartheta | \vartheta \in \mathbf{\Theta}\}$, the sample space $\mathbf{F}$, the sample $\mathbf{\Phi}$, the decision space $\mathbf{D}$ and the decision function d. The symbol " : = " means definition so the letter introduced like that can be considered as an abbreviation. We are not going to use unit transformations. So the SI unit system extended for eV will be fixed and, for the sake of simplicity, each quantity will be characterized by means of a real number omitting the unit.

Let the function measured be continuous

$$F_{\mathbf{p}} : \mathbf{X} \to \mathbf{Y}; \qquad \mathbf{u} \mapsto F_{\mathbf{p}}(\mathbf{u}), \tag{4}$$

where $\mathbf{u} := (u_1, u_2, \ldots, u_r) \in \mathbf{R}^r$ is the independent variable vector of r-dimension $\mathbf{p} := (p_1, p_2, \ldots, p_k) \in \mathbf{R}^k$ is a k-dimensional parameter vector. The result of a measurement is always a finite real series so the accidental effects are given by the following definition

$$\mathbf{\Omega} := (\mathbf{X} \times \mathbf{Y})^{\mathbf{N}}, \tag{5}$$

where $\mathbf{N}$ is the set of natural numbers: $\{1, 2, 3, \ldots\}$.

The event space $\mathbf{S}$ should be a σ-algebra over $\mathbf{\Omega}$. The exact definition traces back to the measurable subsets of $\mathbf{R}^{(r+1)n}$, where $n \in \mathbf{N}$ is a given number which shows at how many values of the argument $\mathbf{u}$ the function was measured. Let f be the following map:

$$f : \mathbf{B} \to \{\text{the power set of } \mathbf{\Omega}\}; \qquad B \in \mathbf{B}$$

$$f(B) := \{\omega | \omega \in \mathbf{\Omega} \text{ and } (\omega_1^{(x)1}, \omega_1^{(x)2}, \ldots, \omega_1^{(x)r}, \omega_1^{(y)},$$

$$\omega_2^{(x)1}, \omega_2^{(x)2}, \ldots, \omega_2^{(x)r}, \omega_2^{(y)}, \ldots, \omega_n^{(x)1}, \omega_n^{(x)2}, \ldots, \omega_n^{(x)r}, \omega_n^{(y)}) \in B\}, \tag{6}$$

where $\mathbf{B}$ is the set of measurable subsets of $\mathbf{R}^{(r+1)n}$, $i \in \mathbf{N}$; $i \mapsto \omega_i$; $\boldsymbol{\omega}_i^{(x)} := (\omega_i^{(x)1}, \omega_i^{(x)2}, \ldots, \omega_i^{(x)r}) \in \mathbf{X}$; $\omega_i^{(y)} \in \mathbf{Y}$. Now $\mathbf{S}$ can be defined like this

$$\mathbf{S} := f(\mathbf{B}). \tag{7}$$

f can be proved to be bijective, and $\mathbf{S}$ is really a σ-algebra.

We want to have values for k different parameters ($k \in \mathbf{N}$) and the relative variance $V > 0$. V is assumed to be constant. So the parameter space of the statistical field is chosen in the following manner:

$$\mathbf{\Theta} \subset \mathbf{R}^k \times (\mathbf{R}^+ - \{0\}) \text{ and it is a finite closed interval,} \tag{8}$$

where $\mathbf{R}$ is the set of real numbers, $(\mathbf{R}^+ - \{0\})$ is the set of positive real numbers.

Since the measurement is influenced by numerous random effects it is appropriate to assume normal distributions. More exactly

$$P_\vartheta : \mathbf{S} \to [0, 1]$$

$$P_\vartheta(S) := \int_{f^{-1}(S)} \frac{1}{(2\pi V)^n \left[\prod_{i=1}^{n} \left(\prod_{\rho=1}^{r} \sigma_\rho^{(i)}\right) \cdot F_{\mathbf{p}}(\mathbf{x}^{(i)})\right]} \cdot$$

$$\cdot e^{-\frac{1}{2}\sum_{i=1}^{n}\left[\sum_{\rho=1}^{r} \frac{(x_\rho^{(i)} - u_\rho^{(i)})^2}{\sigma_\rho^{(i)2}} + \frac{(y^{(i)} - F_{\mathbf{p}}(\mathbf{x}^{(i)}))^2}{F_{\mathbf{p}}^2(\mathbf{x}^{(i)}) \cdot V^2}\right]} \prod_{i=1}^{n} \left(\prod_{\rho=1}^{r} \mathrm{d}x_\rho^{(i)}\, \mathrm{d}y^{(i)}\right), \tag{9}$$

where $S \in \mathbf{S}$; $\vartheta := (p_1, p_2, \ldots, p_k, V)$; $\sigma_\rho^{(i)} \in \mathbf{R}$ $(\rho = 1, 2, \ldots, r;\ i = 1, 2, \ldots, n)$ are given mean deviations. $\mathbf{u}^{(i)} := (u_1^{(i)}, u_2^{(i)}, \ldots, u_r^{(i)}) \in \mathbf{R}^r$ are those values at which the measurements are thought to be performed. It can be proved that $P_\vartheta(f(\mathbf{R}^{(r+1)n})) = 1$ is satisfied. By means of definition (9)

$$\mathbf{P} := \{P_\vartheta \mid \vartheta \in \mathbf{\Theta}\}. \tag{10}$$

As we measure at n different r dimensional points the sample space will be as usual

$$\mathbf{F} := \mathbf{R}^n. \tag{11}$$

The sample is a map of $\mathbf{\Omega}$ onto $\mathbf{F}$ as follows

$$\mathbf{\Phi} : \mathbf{\Omega} \to \mathbf{F}; \qquad \mathbf{\Phi}(\omega) := (\omega_1^{(y)}, \omega_2^{(y)}, \ldots, \omega_n^{(y)}). \tag{12}$$

From the sample we would like to obtain the parameter values realized in our measurements, so the $k+1$ dimensional Euclidean space would seem to be appropriate as a decision space. However, a point estimate like that could be given only in a sophisticated way. That is why we choose an appropriate set of intervals of $k+1$ dimension:

$$\mathbf{D} := \{J^{(k+1)} \mid J^{(k+1)} \subset \mathbf{\Theta} \text{ and } J^{(k+1)} \text{ is an interval}\}. \tag{13}$$

Especially the intervals may consist of only one element.

The application of Urbach rule means to determine four parameters (and in addition the variance) by fitting. To solve that task usually ad hoc decision functions are used [9] but without a precise knowledge of the properties of the applied method. It would not be useful to study an ad hoc method as it could not be applied for other cases because of its speciality. The original Gaussian decision function (the least squares method) or its logarithmic variant cannot be used because both the variance and its

logarithm vary while measuring the function $F_{\mathbf{p}}(\mathbf{u})$. The relative variance, however, was constant. Let us introduce the following notations:

$$\vec{F} := (F_1, F_2, \ldots, F_n) \in \mathbf{F}\,, \tag{14}$$

$$V_{\mathbf{p}}^2 := \frac{1}{n-k} \sum_{i=1}^{n} \left[\frac{F_i - F_{\mathbf{p}}(\mathbf{u}^{(i)})}{F_{\mathbf{p}}(\mathbf{u}^{(i)})} \right]^2, \tag{15}$$

where $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \mathbf{u}^{(3)}, \ldots, \mathbf{u}^{(n)} (\in \mathbf{R}^r)$ were defined earlier (Eq. (9)); but can be characterized also as the mean value of the following random variable:

$$\xi^{(i)} : \boldsymbol{\Omega} \to \mathbf{R}^r; \qquad \xi^{(i)}(\omega) := \omega_i^{(x)} \qquad (i = 1, 2, \ldots, n)\,. \tag{16}$$

Obviously, $V_{\mathbf{p}}^2 > 0$ for all $\vec{F} \in \mathbf{F}$ and $\mathbf{p} \in \mathbf{R}^k \cap \boldsymbol{\Theta}$ so the function $\mathbf{p} \mapsto V_{\mathbf{p}}^2$ at any fixed $\vec{F}$ has certainly a minimum over $\mathbf{R}^k \cap \boldsymbol{\Theta}$ because of (4) and (8). However, the minimum value $V_{\min}^2$ can be taken at more than one value of the arguments. Let us denote by $J_{\mathbf{p}\min}^{(k)}$ the smallest k dimensional interval which contains the set $\{\mathbf{p} | \mathbf{p} \in \mathbf{R}^k \cap \boldsymbol{\Theta}$ and $V_{\mathbf{p}}^2 = V_{\min}^2\}$. Then our decision function is given by:

$$\mathrm{d} : \mathbf{F} \to \mathbf{D}; \qquad \mathrm{d}(\mathbf{F}) := J_{\mathbf{p}\min}^{k} \times \{V_{\min}\}\,. \tag{17}$$

It is worth mentioning that the explicit values of the mean deviations $\sigma_\rho^{(i)}$ $(\rho = 1, 2, \ldots, r;\ i = 1, 2, \ldots, n)$ are not required for using that decision function, however, they are important for examining its properties (confidence interval, bias etc.)

Method for calculation

If during measurement an elementary event ω is realized, the sample value $\vec{F} = \boldsymbol{\Phi}(\omega)$ will be obtained and the decision $\mathrm{d}(\vec{F}) = \mathrm{d}(\boldsymbol{\Phi}(\omega))$ will be made for the unknown parameters. The main difficulty for that procedure is to find the interval $J_{\mathbf{p}\min}^{(k)}$ and $V_{\min}$. In our Laboratory a computer program has been elaborated for that purpose. Its rough flowing chart can be seen in Fig. 1, where $R_{\mathbf{p}}$ denotes the register series for $\mathbf{p}$ in the computer. If we have no idea at all for an initial value $\mathbf{p}_{\text{guess}}$ the program can find one by itself. The familiar gradient method does not need any explanation, only a remark: the derivatives need not be programmed, they are calculated from the $\mathbf{p} \mapsto V_{\mathbf{p}}^2$ function. After the gradient part a checking subroutine begins to work. We called its algorithm lattice method because it takes a k-dimensional cubeshape lattice network around $\mathbf{p}_{\min}$ found by the gradient subroutine, and checks if there is a smaller $V_{\mathbf{p}}^2$ value at any lattice point. Lattice constant means the distance of two neighbouring lattice points. The limit precision is an input datum. If the lattice subroutine has found two lattice points at which $V_{\mathbf{p}}^2$ is the same, it stops even if it has not reached the precision required and gives a message: it is not worth continuing. Occasionally after a few restarts we can make sure what $J_{\mathbf{p}\min}^{(k+1)}$ has to be taken. The program can find a "very flat" minimum as well, and

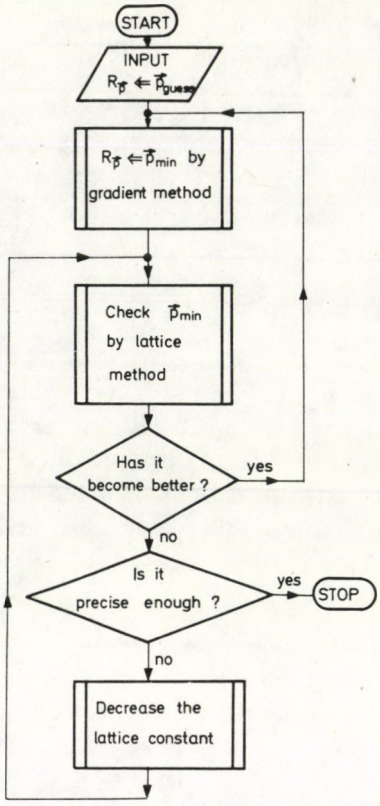


Fig. 1. Flow chart of the computer program for fitting the Urbach parameters

the dependence of the $\mathbf{p} \mapsto V_{\mathbf{p}}^2$ function on any parameter component can be "very weak" in comparison with the others. That is the module of one component of the gradient vector can be smaller even by many orders of magnitude than the other ones.

Results and discussion

The absorption data corrected for reflected light are collected in Table I and are also plotted in Fig. 2. The TeO_2 crystals grown in our laboratory show lower absorption in their transparent region than those investigated earlier [2, 3]. The absorption data plotted on a logarithmic scale deviate from the straight line only for values below 1 cm^{-1} of the absorption coefficient indicating the perfectness of the crystals. The low level of both light losses and acoustic attenuation was attributed to the relatively small amount of crystal defects (1). The absorption values below 0.1 cm^{-1} seem to be uncertain because of the limited exactness of the reflection data. Takizawa's statement that the reflectivity was independent of the temperature could be accepted as

Table I

Absorption data versus photon energy for TeO_2 after correcting for reflected light

Data used for fitting are labelled by asterisks

E [eV]	Abs. coeff [cm^{-1}]			
	20 K	100 K	200 K	300 K
3.10				0.024
3.54				0.054
3.65				0.109
3.67				0.310*
3.69				0.561*
3.71				1.071*
3.73				1.883*
3.76		0.038	0.203	3.752*
3.78				7.519*
3.80			0.291	14.118*
3.81				20.096*
3.82				27.23 *
3.84				39.26 *
3.85				53.90 *
3.86				66.53 *
3.87	0.223	0.224	1.10 *	84.63 *
3.88				105 *
3.89				128 *
3.90		0.262	2.902*	
3.92		0.344	7.833*	
3.94			13.304*	
3.95		0.464	21.68 *	
3.96			37.44 *	
3.97		0.895*	58.40 *	
3.99			88.58 *	
4.00	0.573	2.692*	137.46 *	
4.03	1.268*	8.502*		
4.04		14.584*		
4.05	4.496*	33.04 *		
4.06	8.28 *	57.95 *		
4.08	14.577*	91.50 *		
4.09	25.0 *	151 *		
4.10	45.0 *			
4.11	112 *			

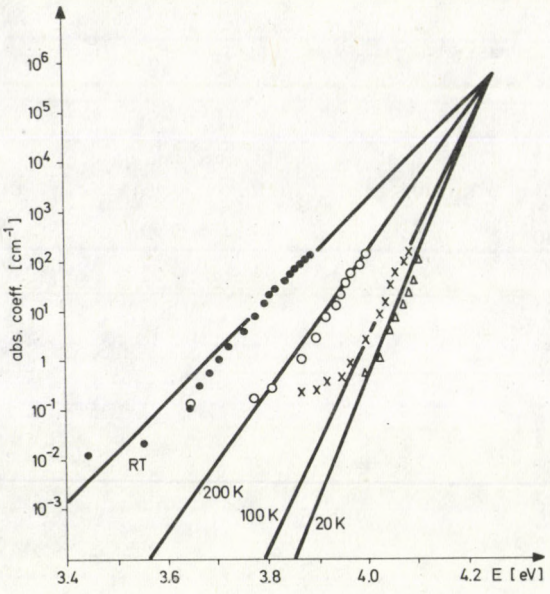


Fig. 2. Experimental absorption data and computed Urbach lines for TeO_2 single crystals

a first approximation [3]. Moreover, the effect of optical activity on the reflectance may cause further errors [12]. Nevertheless the incorrectness of the reflection correction might be neglected at the higher absorption values which were used in the computer fitting of the Urbach equation.

In our case, using the definitions in Eqs (1) and (2)

$$r=2; \qquad \mathbf{u}=(E, T),$$

$$k=4; \qquad \mathbf{p}=(A_0, \sigma_{s0}, E_0, \hbar\omega_0),$$

$$F_{\mathbf{p}}(\mathbf{u})=A,$$

$$n=39.$$

The independent variables $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \ldots, \mathbf{u}^{(39)}$ and the sample $(F_1, F_2, \ldots, F_{39})$ are given in Table I. The parameter space $\boldsymbol{\Theta}$ can be naturally chosen according to the smallest and largest positive number of the computer.

The computer fitting yielded the following results

$$^*J_{A_{0\min}} \quad \text{is inside of} \quad [5.2\times10^5, 5.3\times10^5]\ [\text{cm}^{-1}]$$

$$J_{\sigma_{S0\min}} \quad \text{is inside of} \quad [0.6324, 0.6354]$$

$$J_{E_{0\min}} \quad \text{is inside of} \quad [4.254, 4.257]\ [\text{eV}]$$

$$J_{\hbar\omega_{0\min}} \quad \text{is inside of} \quad [2.273\times10^{-2}, 2.275\times10^{-2}]\ [\text{eV}]$$

$$J^{(4)}_{\mathbf{p}_{\min}} = J_{A_{0\min}} \times J_{\sigma S 0\min} \times J_{E_{0\min}} \times J_{\hbar\omega_{0\min}}$$

$$V_{\min} = 32.446\,894\%$$

The intervals could have been decreased by running our program for a longer time, but it did not seem to be necessary, as the variances would have been probably much larger. Our statistical model allows us to calculate a confidence interval of 95% for the parameter vector **p** but because of computation time limitations we would like to test the method by a function having less than four parameters. The "weakest" parameter is A_0 and the "strongest" is $\hbar\omega_0$. The gradient subprogram changed them from the initial values 2×10^7, 1.74×10^{-2} to 7.26×10^5, 2.35×10^{-2} respectively, and $V_{\mathbf{p}}$ changed from 4592% (!) to 32.48% so we can see that the lattice subprogram could not be neglected.

The physical reality of the computed data could hardly be checked by comparing them to literature data. The Urbach shape analysis was performed by Takizawa [3] only for extraordinary rays. For ordinary rays he published reflectivity spectra which contain a shoulder near 4.30 eV at RT. That shoulder roughly corresponds to our E_0 parameter. The ω_0 phonon frequencies have been measured by IR and Raman spectroscopy [13, 14] but the average energy of the phonons combining the exciton transitions with the given polarization direction have not been calculated.

References

1. I. M. Silvestrova, Yu. V. Pisarevskii, I. Földvári, Á. Péter, R. Voszka and J. Janszky, Phys. Stat. Sol. (a) **66,** K55, 1981.
2. M. Uchida, Phys. Rev., B **4,** 3736, 1971.
3. T. Takizawa, J. Phys. Soc. Japan, **48,** 505, 1980.
4. F. Urbach, Phys. Rev., **92,** 1324, 1953.
5. T. Toyozawa, Progr. Theoret. Phys., **20,** 53, 1958; **22,** 445, 1959.
6. A. Mecseki, I. Várnai, I. Földvári, Prognosztika 19, 1981. (in Hungarian).
7. K. Raksányi et al. to be published in Acta Chim. Acad. Sci. Hung.
8. I. Földvári, K. Raksányi, R. Voszka, E. Hartmann and Á. Péter, J. Crystal Growth, **52,** 561, 1981.
9. E.g. T. Tomiki, T. Miyata, H. Tsukamoto, Z. Naturforsch., **29a,** 145, 1974.
10. I. Vincze, Mathematical Statistics, Tankönyvkiadó, Budapest, 1972 (in Hungarian).
11. A. Wald, Statistical Decision Functions, J. Wiley, New York, 1961.
12. T. Takizawa, J. Phys. Soc. Japan, **50,** 3054, 1981.
13. A. S. Pine and G. Dresselhaus, Phys. Rev., **B5,** 4087, 1972.
14. D. M. Korn, A. S. Pine, G. Dresselhaus and T. B. Reed, Phys. Rev., **B8,** 768, 1973.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 25—32 (1982)

RAMAN STUDY OF ISOLATED AND "IN SITU" T7 PHAGE DNA: CONFORMATION AND POSSIBLE INTERACTION WITH THE PROTEINS*

D. ASLANIAN

LABORATOIRE DE PHYSIQUE DES SOLIDES ET SPECTROSCOPIE DES SOLIDES
ASSOCIÉ AU CNRS LA154 UNIVERSITÉ P. ET M. CURIE
75230 PARIS CEDEX 05, FRANCE

GY. RONTÓ and K. TÓTH

INSTITUTE OF BIOPHYSICS, SEMMELWEIS MEDICAL UNIVERSITY
BUDAPEST, HUNGARY

The Raman spectrum of T7 phage is shown to be dominated by the vibrations of the constituent DNA (backbone chain and bases). These vibrations are compared with their counterparts in the spectrum of isolated T7 DNA. Inspite of the great similarity between the two spectra there are some differences which suggest that the bases play an important role in the DNA-protein interaction. The sites of these interactions appear to involve the CH_3 and C=O groups of thymine, C=O groups of cytosine and N_7—C_8 group of guanine.

Some preliminary observations are given on the UV induced changes in isolated DNA.

Introduction

T7 bacteriophage is one of the small, well-examined nucleoprotein systems which contains double-stranded DNA and several proteins in a functional unity.

The phage head is icosahedron-like with a diameter of ~ 60 nm and it has a short tail [1]. The molecular weight of the total phage is about $51 \cdot 10^6$ daltons [2], the half of it is that of the DNA [3]. In the head, the DNA is wound around a protein core and is enclosed by a protein capsid. The winding represents a regular higher structure, kept by interactions between the DNA and proteins and by internal interactions of DNA [1].

The knowledge of the DNA fine structure in the phage head, related to other nucleoproteins (viruses, chromosomes), is important if one wants to develop model systems to study the responses to external agents such as radiations or chemicals. The effect of the UV irradiation on T7 bacteriophage studied by biological and several optical spectroscopic methods are reported [4—7]. Nevertheless, results on the fine changes in the molecular vibrations are missing.

It has been shown that Raman spectroscopy is a sensitive tool to study the conformation of nucleic acids and proteins ([8, 9] and ref. therein) as well as of virus

* Dedicated to Prof. I. Tarján on his 70th birthday.

[10, 11, 12], and their conformational changes. Raman studies also show that the interactions of histidine or protamine with DNA do not lead to structural changes of DNA, which remains in a stable B form [13, 14].

In this work we report on the Raman spectrum of T7 phages and of their normal and UV-irradiated DNA. Evidence is presented that the spectra contain useful information on the DNA structure and on its interactions with the phage proteins.

UV-induced conformational changes in the DNA molecule have been studied and some preliminary results are reported.

Experimental

Samples

T7 bacteriophage was grown on chemostated E. coli B host cells, using our earlier experiments obtained by this method [15]. The purification and concentration of the phages was carried out by a modified method of Strauss and Sinsheimer [16]. The final phage-concentration was about 70 mg/ml in 0,1 mol/l phosphate buffer [M9].

T7 DNA was isolated according to Mandell [17], and it was stored in ethanol. Before measurements DNA was separated from the ethanol and dried slightly. It was resuspended in some drops of the same buffer as the phages. To have a sufficiently high concentration, the buffer quantity was limited and thereby the percentage of the remaining ethanol could be significant. The final DNA concentration was about 15 mg/ml, in a gel-like state.

UV irradiation was performed by a 15 W Tungsram germicidal lamp. Samples in quartz capillary tubes (diameter 1 mm) were at 5 cm distance from the lamp for 1 hour. The incident dose at the surface was about 3 kJ/m^2 (at 254 nm), but in average it is much smaller because of the high absorptivity of the DNA.

Raman spectra

Raman spectra were recorded by a Coderg PH 1 Spectrometer. An argon laser at $\lambda = 488$ nm and $P \sim 600$ mW was used. The spectral slits were 6 cm^{-1}. Quartz capillary tubes held in vertical position in the middle of a specially constructed brass cell were employed for the samples (10 µl). The temperature ($\sim 10\,^{\circ}C$) was controlled by a thermoelectric cooling system.

Results and discussion

1. *Raman spectra of DNA*

The Raman spectrum of T7 phages is shown in Fig. 1a. It is dominated by the Raman scattering of the constituent DNA. This corresponds well to the 50% attribution of DNA in the total molecular weight of the T7 phage [2] considering that

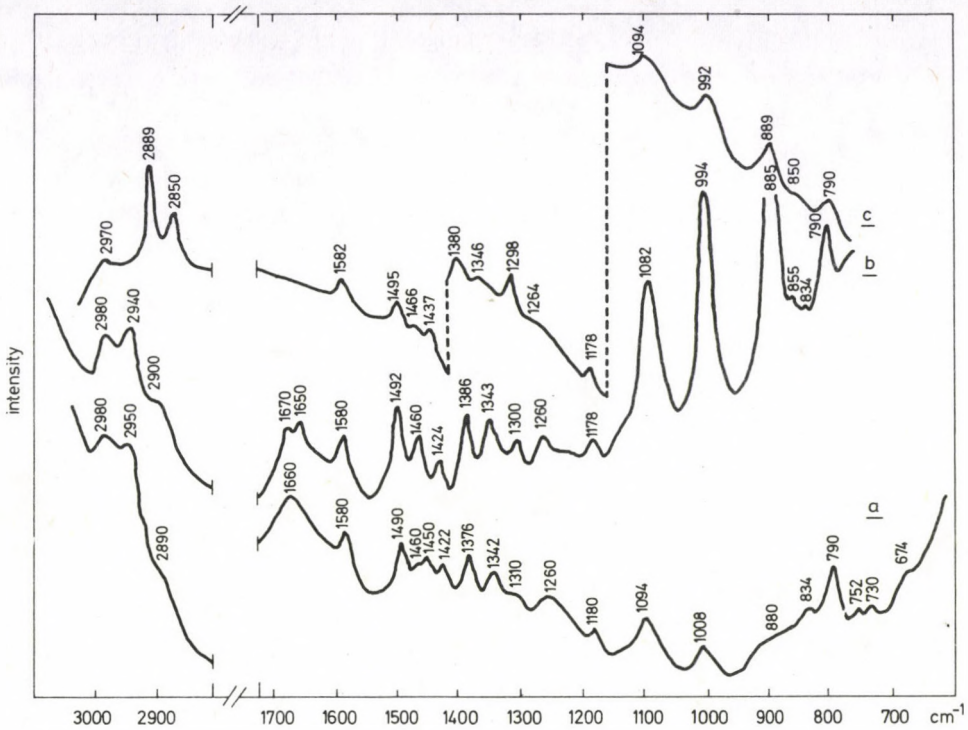


Fig. 1. Raman spectra of (a) T7 bacteriophage in 0.1 mol/l phosphate buffer at pH = 7 ($c \sim 70$ mg/ml); (b) isolated T7 bacteriophage DNA in the same buffer ($c \sim 15$ mg/ml); (c) UV-irradiated, isolated T7 DNA ($\lambda = 254$ nm; incident dose ~ 3 kJ/m^2)

the Raman effect of the proteins is much weaker than that of the DNA. Thus our discussion concerns only the DNA spectrum.

The Raman spectra of isolated T7 DNA, native and UV-irradiated are presented in Figs 1b and c. The attributed frequencies and their assignment are included in Table I.

The Raman spectrum of DNA consists of vibrations which come from the two main parts of the molecule: the backbone chain and the pyrimidine and purine bases.

1.1. *Backbone chain vibrations.* The sugar-phosphate backbone of nucleic acids possesses only two Raman-active vibrations which are clearly identifiable in terms of normal modes [8]: the diester stretching, ν(—O—P—O—), and dioxy symmetric stretching, $\nu(PO_2)$. Their frequencies and intensities are directly related to the A, B or C form of the molecule [8, 18, 19].

In the Raman spectrum of T7 phages (Fig. 1a) the ν(—O—P—O—) mode appears at 834 cm^{-1} as a weak shoulder of the strong vibration at 790 cm^{-1}. The latter one has been assigned primarily to a cytosine band [20]. Later, it was found [21] that the same band changes substantially in intensity when DNA passes from B form to A form. It has been suggested that, in fact, this band corresponds to the ν(—O—P—O—)

vibrations of the B-type structure when accompanied by the weak band at 834 cm^{-1}. Thus the presence of the two vibrations at 834 cm^{-1} and 790 cm^{-1} in the spectrum of T7 phages (Fig. 1a) indicates that its DNA "in situ" is in the B form. This is confirmed by the dioxy symmetric stretching vibration of the phosphate group, $\nu(PO_2^-)$, situated at 1094 cm^{-1}. This frequency has been indicated to be also characteristic for the B conformation of DNA [18]; when the backbone is in form A it lies at 1101 cm^{-1}.

In her CD-study on the T7 phage DNA "in situ" Tóth [6] came to the conclusion that the conformation of DNA in the phage is intermediate between B and C forms. Small-angle X-ray scattering results [22] led to the assumption that DNA is in a B-like conformation in the phage head. The Raman spectrum of the C form of DNA studied by Goodwin and Brahms [19] has indicated a band at 865—870 cm^{-1} as being characteristic of this form. Such a vibration is not visible in the spectrum in Fig. 1a. Indeed, there is a large, not well defined shoulder at about 880 cm^{-1}; thus it is may be possible that the CD-spectrum is due to a mixture of B and C forms. However, the Raman spectrum favours strongly the B conformation of DNA "in situ" as the largely preponderant one.

Concerning the conformation of the isolated DNA, the spectrum in Fig 1b shows the bands at 790 cm^{-1} and 834 cm^{-1} as well as a third one at about 855 cm^{-1}. The last two vibrations appear as shoulders on the slope of the strong ethanol band situated at 885 cm^{-1}. There are two other strong bands at 994 and 1082 cm^{-1} which are not characteristic either for DNA or for ethanol and remain at present non assigned.

Herbeck et al [23] have recorded Raman spectra of DNA in 60, 70 and 80% ethanol. They concluded that the DNA is in the B conformation at 60 and 70% ethanol but it is in the A form at 80% ethanol. Using CD measurements Girod et al [24] have suggested that in 60% ethanol DNA is in the C form. Herbeck et al [23] cannot confirm the differences between the B and C forms. It seems that in our Raman spectrum (Fig. 1b) the shoulder near 855 cm^{-1} indicates the presence of the C structure. Consequently, we may conclude that the sample contains between 60 and 70% ethanol as it was presumable on the basis of preparation procedure.

1.2. *Vibrations of the bases.* The more intense lines of the nucleic acid bases are believed to be due to the in-plane vibrations of the rings, involving particularly the stretching of C—C, C=C and C=N bonds [20]. They are found mainly in the frequency range 1800—1200 cm^{-1} together with the external ν(C=O) and ν(C—N) vibrations. Their assignment, based on previous studies [8, 19, 25, 26] is given in Table I. Thus the vibration at 1670 cm^{-1} corresponds to ν(C=O) of thymine and that at 1650 cm^{-1} to ν(C=O) of cytosine. The strong band at 1580 cm^{-1} is attributed to the purine rings of adenine and guanine, mainly due to the C_4=C_5 double bond vibrations along with the NH_2 scissoring effect. The very strong peak at 1490 cm^{-1} is assigned predominantly to the guanine in-plane vibration, involving the $\nu(N_7—C_8)$, with a small contribution from adenine. To the cytosine ring is attributed the strong vibration at about 1380 cm^{-1} due to $\nu(N_3=C_4)$ and $\nu(C_4—C_5)$ whereas that one at 1343 cm^{-1} is characteristic of the adenine ring $\nu(C_8=N_7)$, $\nu(C_2=N_3)$ and $\nu(C_8—H)$ vibrations. The

next vibration at 1300 cm^{-1} belongs to the adenine and cytosine rings and the peak at 1260 cm^{-1} correponds to the adenine $\nu(N_7{=}C_8)$ and $\nu(C_8{-}N_9)$ vibrations.

1.3. *C—H stretching modes (2800—3000 cm^{-1}).* In the spectra of isolated and in situ" DNA (Figs 1a and b) only CH_2 deoxyribose groups and CH_3 of thymine can be expected to show stretching vibrations in this range. Thus the shouler at 2900 cm^{-1} is most probably an $\nu_s(CH_2)$ symmetric stretching vibration [20]. The two other vibrations at 2940 and 2980 cm^{-1} are associated with the CH_3 group and are assigned as symmetric, $\nu_s(CH_3)$ and symmetric, $\nu_a(CH_3)$, methyl vibrations.

2. Interaction of DNA with proteins

Some information on the interaction of DNA with the phage proteins can be given comparing the spectrum of T7 phages or of DNA "in situ" (Fig. 1a) to the spectrum of isolated DNA (Fig. 1b). The following observations have been made:

— the peaks at 1670 cm^{-1} (ν(C=O) of thymine) and at 1650 cm^{-1} (ν(C=O) of cytosine) practically disappear in the phage spectrum;

— the vibration at 1490 cm^{-1} ($\nu(N_7{=}C_8)$ of guanine) is clearly far weaker in the phage DNA than in the isolated DNA spectrum;

— the peak at 1460 cm^{-1} (δCH_3 of thymine) almost disappears in the phage spectrum; a new vibration at 1450 cm^{-1} appears, which is not visible in the isolated DNA spectrum and is due probably to the proteins CH_2 and CH_3 vibrations;

— the $\nu_s CH_3$ (2940 cm^{-1}) and $\nu_a CH_3$ (2980 cm^{-1}) vibrations appear with inversed intensities in the phage spectrum.

All these phenomena indicate clearly that the thymine, guanine and cytosine bases play an important role in the interactions of DNA with the phage proteins, involving the C=O and CH_3 groups of thymine, the C=O group of cytosine and the $N_7{=}C_8$ group of guanine.

3. UV-induced conformational changes in isolated T7 DNA

Because of the great experimental difficulties due to a very high fluorescence background of the UV-irradiated samples the quality of the spectrum in Fig. 1c is not the same as that of the other two spectra in the same Figure (Figs. 1a and b). Nevertheless some alterations are clearly seen and we may report them. Thus the strong peak at 2940 cm^{-1} ($\nu_s CH_2$) disappears and another very strong one appears at 2889 cm^{-1}. The shoulder at 2900 cm^{-1} ($\nu_a CH_2$) increases in intensity. The two vibrations at 1670 cm^{-1} (ν(C=O) of thymine) and 1650 cm^{-1} (ν(C=O) of cytosine) disappear also. The peak at 1460 cm^{-1} (δCH_3 of thymine) and 1343 cm^{-1} (adenine) diminish their intensity. On the contrary the intensity of the 1300 cm^{-1} (adenine, cytosine) vibrations increases.

All these changes concern primarily the thymine. This is not surprising because of the known effect dimerization of the thymine bases under UV irradiation. There are

Table I

Characteristic Raman frequencies of T7 phages and T7 DNA alone: normal and UV-irradiated (incident dose 3 kJ/m^2)

Frequencies [cm^{-1}]				
T7 phage (DNA "in situ")	DNA alone		Assignment	Probable origin
	normal	UV-irradiated		
674 sh			G	ring stretching
730 vw			A	ring stretching
752 vw				
790 s	790 s	790	C, phosphate	diester symmetric
834 sh	834 sh		phosphate	stretching
	855 sh	850	phosphate	
	885 vvs	889	ethanol	
	994 vvs	992		
1008 m			protein	Phe
	1082 vs			
1094		1094	phosphate	dioxy symmetric stretching
1180 w	1178 vw	1178		
1260 m(L)	1260 w	1264	A	ring stretching: $\nu(N_7{=}C_8)$ and $\nu(C_8{-}N_9)$
1310 sh	1300 sh	1298	A, C	ring stretching

1342 m	1343 s	1346	A	ring stretching: $\nu(C_8{=}N_7)$ $\nu(C_2{=}N_3)$ and $\nu(C_8{-}H)$
1376 m	1386 s	1380	C	ring stretching: $\nu(N_3{=}C_4)$ $\nu(C_4{-}C_5)$
1422 w	1424 w	1437	A, G	ring stretching
1450 m			protein	$\delta CH_3, \delta CH_2$ bending
1460 vw	1460	1466	T, deoxyribose	$\delta CH_3, \delta CH_2$
1490 s	1492 s	1495	G, A	ring stretching in-plane: $\nu(N_7{=}C_8)$ of G
1580 s	1580 s	1582	G, A	ring stretching $\nu(C_4{=}C_5)+NH_2$ sciss.
1660 L (H_2O)	1650 s	–	C	$\nu(C{=}O)$
	1670 s	–	T	$\nu(C{=}O)$
		2850 s		
2890 sh	2900 sh	2889 vs	deoxyribose	$\nu_s(CH_2)$
2950 sh (vs)	2940 vs		T	$\nu_s(CH_3)$
		2970 vw		
2980 vs	2980 sh		T	$\nu_a(CH_3)$

A – adenine, C – cytosine, G – guanine, T – thymine, s – strong, w – weak, m – medium, v – very, L – large, sh – shoulder ν – stretching vibr., ν_s – symmetric stretching, ν_a – asymmetric stretching, δ – bending

changes also in guanine, adenine and cytosine vibrations as well as in the deoxyribose CH_2 groups.

It seems that the backbone may be in a C form, but this needs confirmation. Further studies on UV-irradiated isolated DNA and T7 phages are under way.

References

1. Gy. Rontó, M. M. Agamalyan, G. M. Drabkin, L. A. Feigin and Yu. A. Lvov (to be published in Biophys. J.)
2. P. Camerini-Otrero et al., Biochemistry, **13,** 960, 1974.
3. J. A. Harpst, Biophys. Chem., **11,** 295, 1980.
4. Gy. Rontó, Theses for D.S.C. Budapest, 1979.
5. A. Fekete and Gy. Rontó, Studia biophys., **80,** 165, 1980.
6. K. Tóth, Theses, Université P. et. M. Curie, Paris, 1981.
7. S. Dobos, J. J. Fidy and Zs. Laczkó, J. Mol. Structure, **61,** 251 1980.
8. W. L. Peticolas, in Infrared and Raman Spectroscopy of Biological Molecules (Theo M. Theophanides ed.) NATO Advanced Institutes Series, D. Reidel Publishing Company, Doidrecht, Holland, 153—165, 1979.
9. M. Tsuboi, Proceedings of 5th Intern. Confer. on Raman Spectroscopy, Freiburg, 1976 (E. D. Schmidt et al. ed.) Hans Ferdinand Schulz Verlag, D-7800 Freiburg im Breisgau,135—143, 1976.
10. K. A. Hartman, N. Clayton and G. J. Thomas Jr., Biochem. Biophys. Research Communic., **50,** 942, 1973.
11. G. J. Thomas Jr. and P. Murphy, Science, **188,** 1205, 1975.
12. G. J. Thomas Jr., B. Prescott, P. E. McDonald-Ordzie and K. A. Hartman, J. Mol. Biol., **102,** 103, 1976.
13. T. T. Herskovits and J. Brahms, Biopolymers, **15,** 687, 1976.
14. J. Liquier, J. Taboury, E. Taillandier and J. Brahms, Biochemistry, **16,** 3262, 1977.
15. S. Gáspár, Gy. Rontó and G. Müller, Teitsch. Allg. Microbiol., **19,** 163, 1979.
16. J. M. Strauss and R. L. Sinsheimer, J. Mol. Biol., **7,** 43, 1963.
17. I. Mandell and A. D. Hershey, Anal. Biochem., **1,** 66, 1960.
18. S. C. Erfurth, E. J. Kiser and W. L. Peticolas, Proc. Natl. Acad-Sci. USA, **69,** 938, 1972.
19. D. C. Goodwin and J. Brahms, Nucleic. Acids. Research, **5,** 835, 1978.
20. R. C. Lord and G. J. Thomas, Spectrochim. Acta, **23A,** 2551, 1967.
21. S. C. Erfurth, P. J. Bond and W. L. Peticolas, Biopolymers, **14,** 1245, 1975.
22. Yu. A. Rolbin, D. I. Svergun, L. A. Feigin, S. Gáspár and Gy. Rontó, Dokl. Akad. Nauk, **255,** 1497, 1980.
23. R. Herbeck, T. J. Yu and W. L. Peticolas, Biochemistry, **15,** 2656, 1976.
24. J. C. Girod, W. C. Johnson Jr., S. K. Huntington and M. F. Maestre, Biochemistry, **12,** 5092, 1973.
25. L. Chinsky, P. Y. Turpin, M. Duquesne and J. Brahms, Biopolymers, **17,** 1347, 1978.
26. M. Tsuboi, J. Raman Spectroscopy, **2,** 609, 1974.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 33—37 (1982)

SOME GROUP THEORETICAL REMARKS ON THE SYMMETRY OF POLYMORPHIC CRYSTALS*

Z. MORLIN

RESEARCH LABORATORY FOR CRYSTAL PHYSICS OF THE HUNGARIAN ACADEMY OF SCIENCES
1502 BUDAPEST, HUNGARY

Some group-theoretical aspects of polymorphic phase transitions are discussed. The equilibrium symmetries contain as a subgroup in a preformed structure that phase which is in equilibrium at a different temperature. If the polymorphic structure consists of two phases the transition takes place through a subgroup common to both phases. This subgroup is an intersection of the two groups in the group-theoretical sense. The phase transition actually consists of the dissymmetrization and symmetrization of the subgroup, whereby the subgroup first becomes dissymetrized followed by symmetrization from which finally the other equilibrium phase emerges. The dissymmetrization-symmetrization processes are reflected in a symmetrical change of physical properties.

The equilibrium phase of polymorphic crystals always contains in some preformed way volume elements of the non equilibrium structure [1, 2, 3]. From the group-theoretical viewpoint this means the coexistence of two (eventually more) groups of various degree of symmetry. As an example consider a P4 tetragonal structure embraced by a hexagonal body with P6 symmetry (Fig. 1). The two symmetries are described by the cyclical groups

$$\mathrm{P4} = \{1, A_{\pi/2}, A^2_{\pi/2}, A^3_{\pi/2}\}\,, \tag{1}$$

$$\mathrm{P6} = \{1, A_{2\pi/6}, A^2_{2\pi/6}, A^3_{2\pi/6}, A^4_{2\pi/6}, A^5_{2\pi/6}\}\,. \tag{2}$$

A common structure of G_1 and G_2 is

$$H = \{1, A_\pi\} \tag{3}$$

which is the only symmetry of the composite structure. H is an intersection of G_1 and G_2, and has a lower symmetry, which is generally the case if two higher symmetries are combined to result in a common structure (2, 4):

$$H = G_1 \cap G_2\,. \tag{4}$$

Any transition from the subgroup H to one of the groups G_1 or G_2 is called symmetrization, whereas Eq. (4) represents a dissymmetrization of G_1 and G_2, respectively.

The phase transformation consists essentially of the symmetrization subsequent to a transitional dissymmetry. This becomes manifest only in the course of the

* Dedicated to Prof. I. Tarján on his 70th birthday.

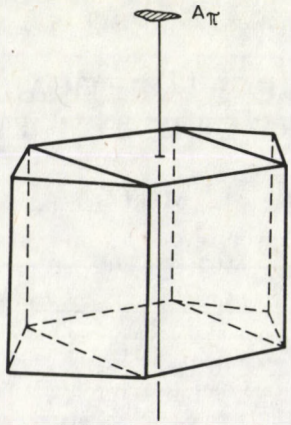


Fig. 1. Composite structure consisting of P4 and P6 symmetries. The common rotation axis denoted by A_π reduces both symmetries to P2

transformation process, since in equilibrium the preformed non-equilibrium state does not violate the symmetry of the superstructure. The superstructure represents a mathematical group with the preformed structure as a subgroup. Any crystallographic group may be built up by the union of its subgroups or by coset formation [5, 6].

Fig. 2a depicts a simple example, the equilibrium structure of a CsCl-α type $\left(I\frac{4}{m}\bar{3}\frac{2}{m}\right)$ crystal denoted by G_α. The point group $\frac{4}{m}\bar{3}\frac{2}{m}$ is built up from te skeleton group $\{432\}$ which contains only rotational elements:

$$\begin{aligned} G=\{432\} = \{1, A_{\pi/2}, A^2_{\pi/2}, A^3_{\pi/2}; B_{\pi/2}, B^2_{\pi/2}, B^3_{\pi/2}; \\ C_{\pi/2}, C^2_{\pi/2}, C^3_{\pi/2}; \\ {}^0D_{2\pi/3}, {}^0D^2_{2\pi/3}; \quad {}^1D_{2\pi/3}, {}^1D^2_{2\pi/3}, {}^3D_{2\pi/3}, {}^3D^2_{2\pi/3}; \\ {}^0F_\pi, {}^1F_\pi, {}^2F_\pi, {}^3F_\pi, {}^4F_\pi, {}^5F_\pi\}, \end{aligned} \tag{5}$$

where A, B and C are 4-fold, D are 3-fold, and F represent 2-fold axes (Fig. 3).

If this group is combined with the inversion group $G=\{1, \bar{1}\}=\{1, i\}$ one obtains the point group $\frac{4}{m}\bar{3}\frac{2}{m}$ which contains besides the rotational elements an inversion centre and mirrors [5, 6]. All the elements of this point group are distributed among 25 subgroups. The CsCl-β (G_β) is the equilibrium structure above 460 °C. It consists of the same symmetry elements as CsCl-α (G_α), but belongs to a different space group $F\frac{4}{m}\bar{3}\frac{2}{m}$. From the viewpoint of phase transformation the diagonal three dimensional 3-fold axes (see Fig. 2) are of interest, because these axes connect the alternate anions and cations. This is the preformed high temperature phase embraced by the G_α

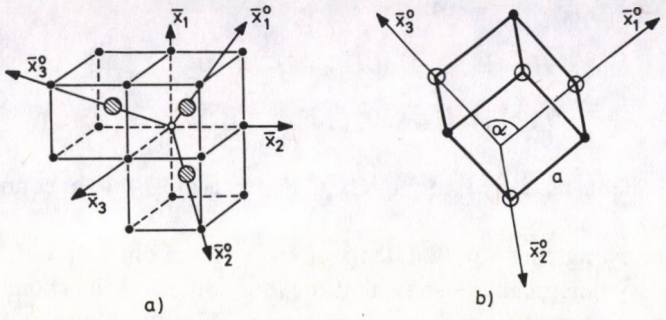


Fig. 2a. The three non-orthogonal axes of a rhombohedron within the CsCl-α structure; the $\bar{x}_1^0$, $\bar{x}_2^0$, $\bar{x}_3^0$ rhombohedral axes are simultaneously 3-fold rotation axes of the CsCl-α symmetries
Fig. 2b. The antisymmetric rhombohedral symmetry within the CsCl-α phase

group. Three 3-fold axes of the G_α group define a non-orthogonal co-ordinate system within the orthogonal G_α system (1). Since the angles of the three axes are the same $\alpha = \beta = \gamma = 70°31'$ (for every CsCl-type crystal), and the ions are at equal distances from each other in the three directions defined by the non-orthogonal coordinates, i.e. $\bar{a} = \bar{b} = \bar{c}$, this system within the group constitutes a rhombohedral symmetry, a substructure of G_α. The rhombohedral group denoted by H may belong either to the space group $R\bar{3}$ or

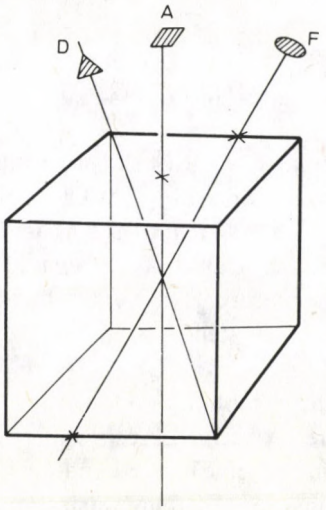


Fig. 3. Rotation axes of the {432} point group
A: 4-fold axes (altogether three)
D: 3-fold axes (altogether four)
F: 2-fold axes (altogether six)
(For simplicity only one of each type is drawn.)

R32 with the elements

$$H_{\bar{3}} = \{1, D_{2\pi/3}, D^2_{2\pi/3}, i, D_{2\pi/3}i, D^2_{2\pi/3}i\}\,, \tag{6}$$

$$H_{32} = \{1, D_{2\pi/3}, D^2_{2\pi/3}, A_\pi, {}^1C_\pi, {}^2C_\pi\}\,, \tag{7}$$

where the D elements are 3-fold, the A, ${}^1C$, ${}^2C$ elements are 2-fold axes and i stands for inversion [5, 6].

It is interesting to note that the primitive rhombohedral CsCl transitional structure is antisymmetric in the sense that the four corners of the rhombohedron are alternately occupied by Cs^+ and Cl^-, respectively (Fig. 2b). No rhombohedral face centred cells exist. The antisymmetry becomes manifest either by inversion in the $R\bar{3}$ or by the 2-fold rotations in the R32 symmetry. In the first case one type of ions is inverted into the second type, in the R32 case one type is rotated around the 2-fold axes into the position of the other, (cation $\rightleftarrows$ anion). This corresponds to Shubnikov's black and white symmetries [2]. Including this antisymmetric property the rhombohedral symmetries within the G group will be further on denoted by a stroke: $R\bar{3}'$ and R32', respectively.

The actual phase transformation of CsCl-α into CsCl-β may be described by the group transitions

$$R\bar{3}' \rightarrow G_\beta$$

or (8)

$$R32' \rightarrow G_\beta.$$

It is quite clear that the H subgroup is an intersection of the groups G_α and G_β

$$H_{\bar{3}}, H_{32} = G_\alpha \cap G_\beta\,. \tag{9}$$

Geometrically the transformation consists of an orthogonalization and dilatation (contraction) of the rhombohedral cell [1]. At first sight this would mean a sudden change (jump) of the steady state system at the phase transformation. Experimental evidence, however, shows that the transition is prepared by a continuous change of the vibrational spectrum of the structure. This eventually results in a gradual disruption of the rhombohedral nucleus which step by step becomes orthogonalized by diffusion controlled kinetics [7].

Consequently, with the exception of a few cases in extremely small regions, no rapid change of the structure is observed. According to X-ray and electron diffraction evidence [8, 9], the reflections become blurred and weakened in the temperature range of transformation due to transitional dissymmetrization, even the co-existence of both phases is observed. Finally, the reflections of the equilibrium symmetry emerges, again with strong and sharp reflection spots or lines. Also electrical conductivity measurements showed a prolonged change at transformation [10, 11, 12]. In this way the

geometrical symmetries are reflected by physical symmetry properties. In the diffraction experiments one has the process:

Strong reflections→transitional blurring→strong reflections.

Similarly also the electrical conductivity properties show a symmetrical change as described in earlier papers cited in the references.

References

1. Z. Morlin, Acta Phys. Hung., **24,** 75, 1968.
2. A. V. Shubnikov and A. V. Koptsik, Symmetry in Science and Art (translated from Russian by D. Archard), I—XXIII, 1—420, Plenum, New York, 1977.
3. Z. Morlin, E. Hartmann, L. Malicskó, Kristall und Technik, **13,** K50, 1978.
4. E. Hartmann, Magyar Fizikai Folyóirat, **21,** 357, 1973 (in Hungarian).
5. M. Buerger, Elementary Crystallography, I—XXIII. 1—528, the MIT Press, London, 1978.
6. Z. Morlin, Magyar Fizikai Folyóirat, **30,** 465; 591, 1982 (in Hungarian).
7. I. Gaál, Z. Morlin and I. Tarján, Acta Phys. Hung., **27,** 25, 1969.
8. Z. Morlin and J. Tremmel, Acta Phys. Hung., **24,** 75, 1966.
9. J. W. Menary, A. R. Ubbelohde and I. Woodward, Proc. Roy. Soc., **A 208,** 158, 1951.
10. Z. Morlin, Acta Phys. Hung., **21,** 137, 1966.
11. J. Arends and H. Nijboer, Solid State Comm. **5,** 163, 1967.
12. P. J. Harvey and J. Hoodless, Phil. Mag., **6,** 543, 1967.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 39—50 (1982)

MODEL FOR THE Mg_1 CENTRE IN NaCl*

A. WATTERICH and K. RAKSÁNYI

RESEARCH LABORATORY FOR CRYSTAL PHYSICS OF THE HUNGARIAN ACADEMY OF SCIENCES
1502 BUDAPEST, HUNGARY

Optical and thermal properties of the Mg_1 centre generated by illumination in the F band of X-irradiated NaCl:Mg crystals have been investigated. A new model of this centre being essentially a dimer which has captured two electrons and two anion vacancies is proposed.

1. Introduction

In Mg^{2+} doped NaCl four different electron excess centres were produced by X-irradiation and subsequent F- and thermal bleaching [1]. The structure of the Mg_1 centre is the least known among them. The Mg_1 centres transform into Mg_2 centres at RT, whereas at higher temperatures the $Mg_2 \rightarrow Mg_3$ centre conversion takes place [1]. Thus the Mg_1 centre can be considered as a primary centre and it would be important to know more about its structure. That is why we tried to follow in more detail the formation kinetics of the Mg_1 centre together with its thermal conversion kinetics. In order to do that the first task was to determine more exactly the optical parameters of the overlapping absorption bands and the optimal conditions and characteristics of the Mg_1 centre formation.

2. Experimental procedures

$MgCl_2$ doped NaCl crystals were grown OH-free by a zone-melting method [2] and cooled down slowly ("as-grown" samples). Some crystals were annealed at 400 °C with subsequent quenching to RT (quenched samples). X-irradiation was carried out at RT using a THX-250 deep-therapy apparatus working at 200 kV and 20 mA yielding a dose rate of ~ 2 C/kg. min. For the bleaching of the F-band at RT a tungsten-iodine lamp equipped with BG-19 and GG-3 filters was used. The optical absorption measurements were recorded at RT by a UNICAM SP-700 spectrophotometer. For the calculations a CDC-3300 computer was used.

* Dedicated to Prof. I. Tarján on his 70th birthday.

3. Experimental results

3.1. The optical parameters

During coloration of "as-grown" crystals in the first stage (i.e. for total irradiation times $t_{irr} < 30$ min) the increase of the F and a V-type band at 209 nm could be observed with a proportionality between the optical densities of both bands (Fig. 1, curve 1).

In the second coloration stage the peak position of the V-band shifted to longer wavelengths and the proportionality changed (Fig. 1, curve 2) indicating the appearance of a new V-type band at longer wavelengths. Also the M band appeared in the optical absorption spectrum.

If the sample was previously annealed at 400 °C for about 10 min and quenched to RT a short X-irradiation and F-bleaching resulted in an optical absorption spectrum showing the presence of a manifold of V bands, the F band (these bands decreased during F-bleaching), and a group of Mg bands between them which increased during F-bleaching (Fig. 2, curve 1). An additional anneal at 100 °C for 10 min yielded curve 2 of Fig. 2. Fig. 2, curve 3 shows the difference spectrum. The group of V band obviously consists of at least two bands: one is unstable at 100 °C (~229 nm), the other (209 nm) being more stable. In the group of Mg bands the Mg_1 band at 332 nm and the one at 390 nm are unstable at 100 °C, and so is the F′ band at 510 nm. Stable bands (Fig. 2, curve 2) in addition to the V band (209 nm) are the Mg_2 (~300 nm) and F (467 nm) bands.

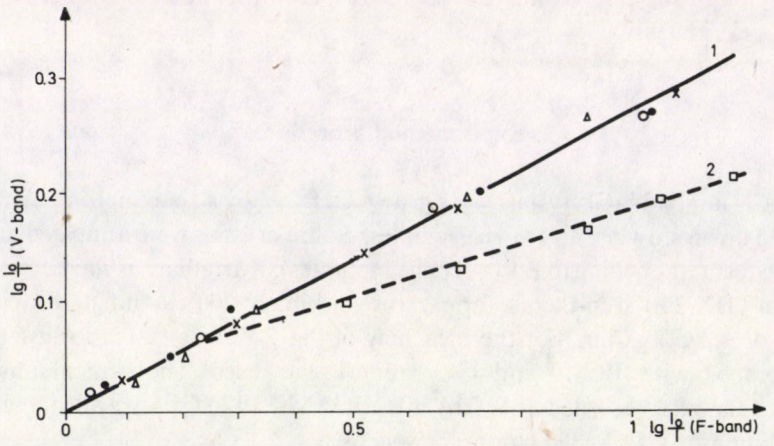


Fig. 1. The optical density of the V band at 209 nm versus that of the F band for the "as-grown" NaCl:Mg crystal

curve 1: First stage of coloration (total irradiation time < 30 min; ○, ●, △ and × standing for different samples)

curve 2: Second stage of coloration (the first value was measured after 1 hour of X-irradiation; □)

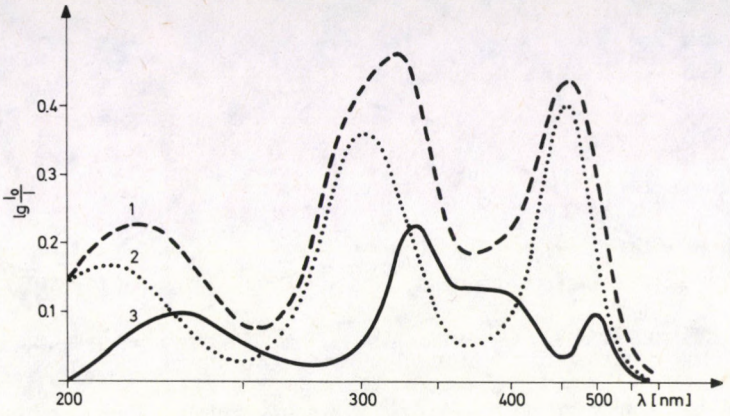


Fig. 2. Curve 1: Optical absorption spectrum of quenched NaCl:Mg after 10 min of X-irradiation and subsequent 10 min of F-bleaching (– – –)
curve 2: The same after an additional 10 min anneal at 100 °C (. . .)
curve 3: The difference of curve 1 and 2 (———)

The original spectrum (Fig. 2, curve 1) was decomposed by an iterative FORTRAN program based on the least square method. The best decomposition was found by assuming seven Gaussian curves (Fig. 3).

In order to determine the parameters of the Mg_3 and Mg_4 bands we used the optical absorption spectrum published earlier [1] (see Fig. 4). In this case the F, V, Mg_2, Mg_3 and Mg_4 bands should be taken into consideration by the decomposition.

All parameters concerning the Mg and V bands obtained from the decompositions are collected in Table I.

The spectrum shown in Fig. 5, curve 1 is mainly the superposition of Mg_3 and Mg_4 bands. Curve 2 belongs to the same sample but annealed at some higher

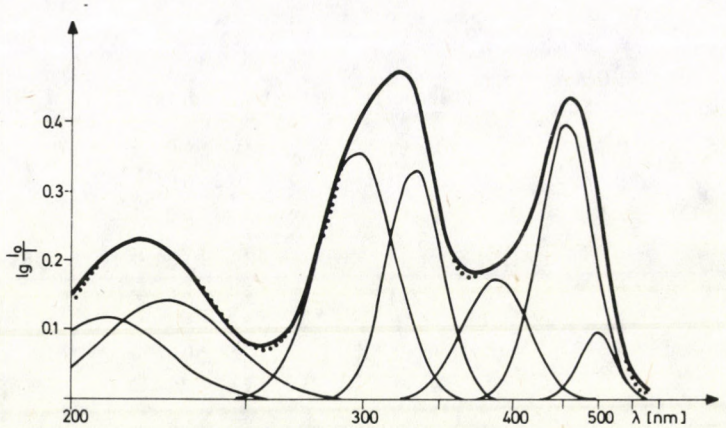


Fig. 3. The Gaussian decomposition of Fig. 2, curve 1 assuming seven Gaussian curves. (Experimental curve: ———, the calculated sum of Gaussian curve: . . . . .)

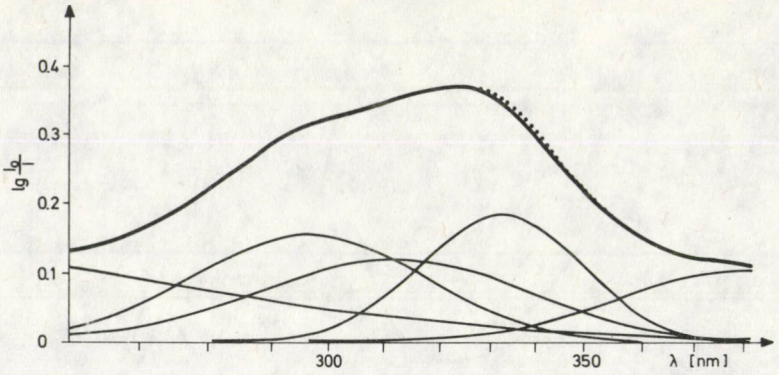


Fig. 4. Optical absorption spectrum of NaCl:Mg after 30 min X-irradiation, F-bleaching for 13 min and annealing at 120 °C for 10 min. (Experimental curve: ———, the calculated sum of Gaussian curves:. . . . .)

temperature (180 °C) where the Mg_3 band becomes unstable. The parameters of the difference (curve 3) are in good agreement with those of the Mg_3 band obtained from the Gaussian decomposition. The calculated parameters of the Mg_4 band are also in good agreement with those given in [3] where the Mg_4 band was prepared independently without the presence of any other interfering Mg band by electrolytical coloration and subsequent "decoloration".

Finally it can be stated that the presence of all the above mentioned absorption bands is experimentally proved and not only assumed for the sake of a good decomposition.

Table I

Parameters of the optical absorption bands obtained from Gaussian decomposition

Centre	Optical absorption peak	
	Position [nm]	Halfwidth [eV]
Mg_1	332±1	0.48±0.01
Mg_2	299±1	0.62±0.01
Mg_3	332±1	0.45±0.01
Mg_4	312±1	0.75±0.01
?	390±1	0.54±0.01
V	209±0.5	1.00±0.01
	226±0.5	1.07±0.01

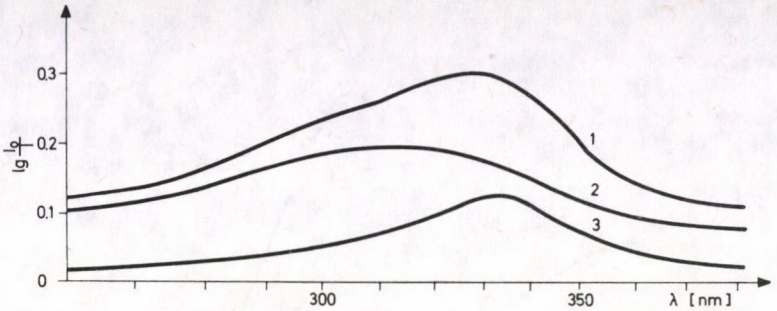


Fig. 5. Curve 1: Optical absorption spectrum of the same crystal as in Fig. 4, but annealed at 120 °C for 30 min
curve 2: The same but heated up to 180 °C
curve 3: The difference of curve 1 and 2

3.2. Optimal conditions and characteristics of the formation of the Mg_1 centre

In order to find out the optimal conditions for a study of the formation kinetics of Mg_1 centres we reinvestigated the "as-grown" crystals.

In "as-grown", X-irradiated and F-bleached crystals (Fig. 6, curve 1) a V band at 209 nm (corresponding to the stabler V band shown in Fig. 3) and a band at ~ 320 nm, i.e. essentially the Mg_1 band were found. The fact that the latter appears to be asymmetric, shifted and broader than the Mg_1 band found previously by decomposition (Table I) indicates the presence of the Mg_2 band. The extinction of this Mg_2 band, however, is much lower than in the quenched crystal treated similarly (Fig. 6, curve 2). In the "as-grown" crystal neither the F′ nor the 390 nm band appeared.

The spectrum of an "as-grown" crystal X-irradiated as far as the second coloration stage and F-bleached afterwards (Fig. 7) shows only the V, Z_4 like [4], Mg_3 and Mg_4 bands but neither the Mg_1 nor the Mg_2 are present since the Mg_1 band should

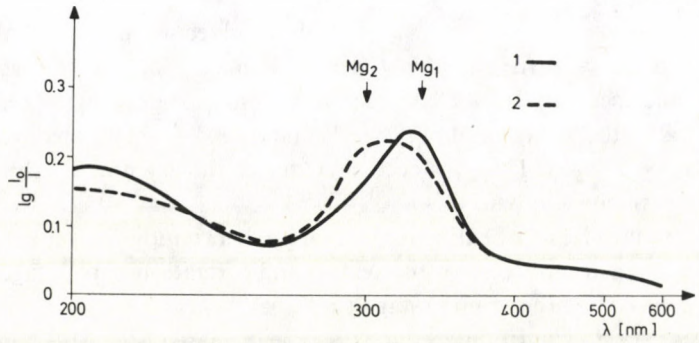


Fig. 6. Optical absorption spectrum of
curve 1: "as-grown" crystal after 10 min of X-irradiation and 10 min of subsequent F-bleaching
curve 2: the same sample but annealed at 400 °C before being treated in the same way as above

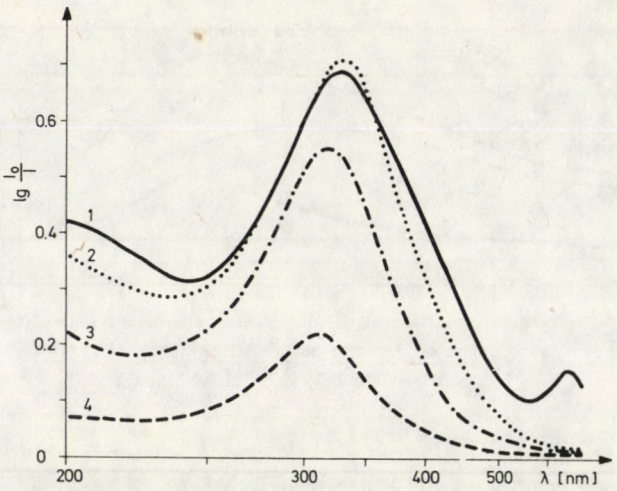


Fig. 7. The optical absorption spectrum of an "as-grown" sample
curve 1: after X-irradiation for 2 hours and total F-bleaching
curve 2: the same with an additional anneal at 100 °C
curve 3: the same with an additional anneal at 150 °C
curve 4: the same with an additional anneal at 180 °C

transform into the Mg_2 band and the Mg_2 band into the Mg_3 band upon annealing at 100 °C and 150 °C, respectively [1], which is not the case in second stage coloured crystals (Fig. 7, curves 1, 2 and 2, 3, respectively). The dominating presence of the Mg_3 centre as opposed to the Mg_2 centre is supported by ESR (their spectra were identified in [1]).

Consequently, the formation kinetics of the Mg_1 band can be best studied using "as-grown" crystals X-irradiated in the first coloration stage ($t_{irr} < 30$ min). In this case only the changes of the V (209 nm), Mg_1, Mg_2 and F bands during F-bleaching have to be taken into account.

Thus we chose such crystals with $t_{irr} = 20$ min and recorded the optical extinction at the peak maxima shown in Table I between short F-bleaching steps. For the evaluation of the spectra a FORTRAN decomposition program was used assuming Gaussian forms for the V, Mg_1 and Mg_2 bands and the real form (as observed before F-bleaching) for the F band. This was necessary due to the asymmetric shape of the F band at the centre concentration used.

The interdependence of extinctions in Mg and F bands during F-bleaching is shown in Fig. 8. In the process of the Mg_1 band formation three stages could be distinguished: in the first (beginning) stage the F band decreases rapidly and the Mg_1 band increases very slowly. In the second stage the increase of the Mg_1 band becomes more rapid, in the third (final) stage slower again. The smaller the Mg concentration the slower the formation rate in the last stage (Figs 8 and 9a).

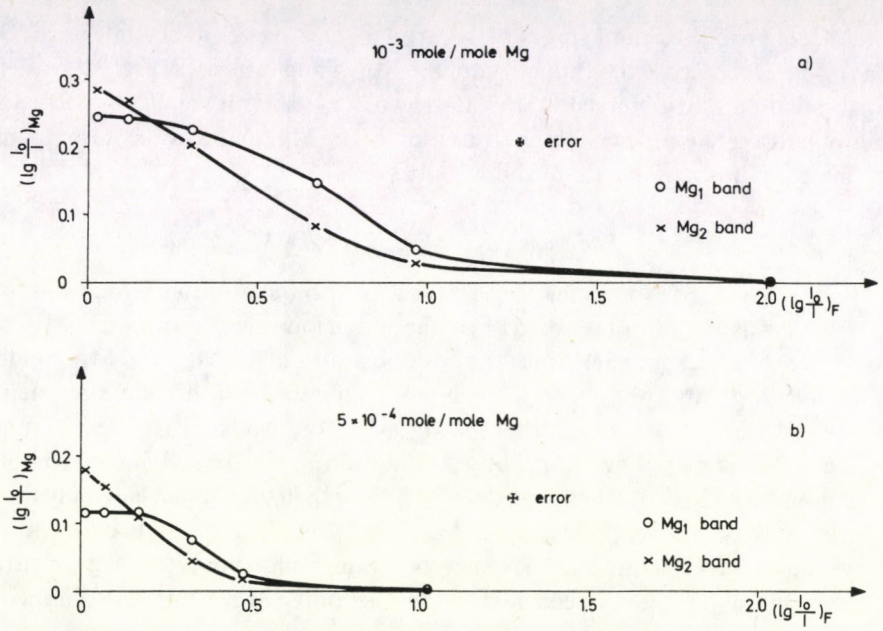


Fig. 8. Optical density of the Mg_1 and Mg_2 bands versus that of the F band during F-bleaching of "as-grown" and X-irradiated NaCl:Mg. (Decomposition-corrected values in the band maxima listed in Table I)

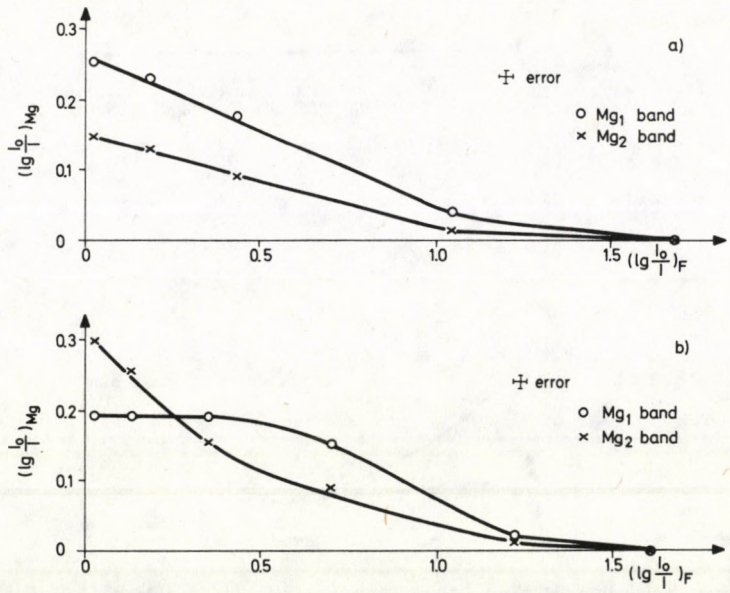


Fig. 9. Optical density of the Mg_1 band versus that of the F band during F-bleaching. The same sample was used in "as-grown" a) and quenched b) cases. The Mg concentration was 3×10^{-3} mole/mole

Comparing the third stages of "as-grown" and quenched crystals it can be seen that in the second case the extinction of the Mg_1 band saturates (Fig. 9). This effect is very similar to that observed in an "as-grown" crystal with small Mg concentration (Fig. 8b). In a quenched crystal the saturation of the Mg_1 extinction is accompanied by the increased formation rate of the Mg_2 band.

3.3. The thermal conversion of the Mg_1 band

The $Mg_1 \rightarrow Mg_2$ thermal conversion was also quantitatively followed.

As it was already established [1] at the end of this conversion a small Mg_3 band is also present in the spectrum. Since the peak positions of the Mg_1 and Mg_3, on the one hand, and Mg_2 and Mg_4 bands, on the other hand, are rather close to each other (Table I), it is hard to decide whether the Mg_3 and Mg_4 bands exist immediately after F-bleaching. We know, however, that X-irradiation in the second stage promotes the formation of the Mg_3 and Mg_4 bands (Fig. 7) though in the present case only first stage coloration was applied. Moreover, Mg_1 and Mg_3 bands have different optical conversion characteristics: as a result of optical bleaching the Mg_1 centres are converted into F centres, in contrast to the case of the Mg_3 centres which recombine with V centres [1].

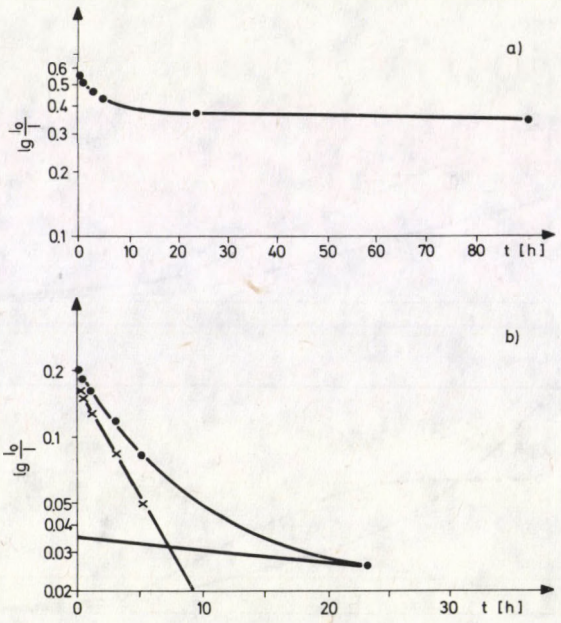


Fig. 10. a) The optical density of the Mg_1 band versus storing time. (The sample was kept in the dark at RT). b) The last measured value in Fig. 10a was subtracted from all the other values (. . .). This curve could be easily fitted by the sum of two linear functions

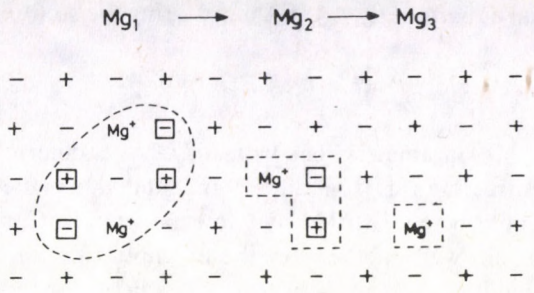


Fig. 11. The models of the Mg centres

Taking into account these properties an upper limit for the extinction of the Mg_3 band immediately after F-bleaching could be given: no more than 15% of the whole extinction at 332 nm belongs to the Mg_3 band.

During thermal bleaching the whole extinction was measured at 332 nm (E_{332}) and drawn versus storing time (Fig. 10a). This consists of the whole extinction of the Mg_1 and Mg_3 bands and a certain contribution (a) of that of the Mg_2 band:

$$E_{332} = E_1 + aE_2 + E_3 , \tag{1}$$

where E_i's are the extinctions of the corresponding Mg bands at the band maxima.

The conversions are assumed to be of first order. This seems to be quite plausible for $Mg_2 \rightarrow Mg_3$ (see the models in Fig. 11), but questionable for $Mg_1 \rightarrow Mg_2$. Some justification for this assumption is provided by the successful fitting procedure to be presented.

The facts that no back-conversions could be detected and both the Mg_1 and Mg_2 decompositions appeared to be full indicate that back-reactions can be neglected at the present stage.

The changes of the extinctions were described by the following differential equations:

$$\frac{dE_1}{dt} = -k_1 E_1 , \tag{2}$$

$$\frac{dE_2}{dt} = \alpha k_1 E_1 - k_2 E_2 , \tag{3}$$

$$\frac{dE_3}{dt} = \beta k_2 E_2 , \tag{4}$$

where the k_i's are the rate constants of the corresponding conversion processes, the parameters α and β relating the extinctions of various bands during conversion.

Solving the differential equations (2—4) and substituting the solution into (1) the final solution is

$$E_{332}(t)=E'e^{-k_1t}+E''e^{-k_2t}+E''', \tag{5}$$

where E', E'' and E''' are parameters. The value of E''' was determined from the last point of Fig. 10a. Subtracting this from all experimental values the curve in Fig. 10b was obtained. This could be easily fitted by two exponentials. Consequently, Eq. (5) described the process quite well and therefore the assumption concerning the first order kinetics of the Mg_1 band conversion is not in contradiction with the experiment.

4. Discussion

Table II contains all characteristics which should be taken into consideration in the creation of a new model for the Mg_1 centre, more suitable than the one proposed in [1].

The decrease of the formation rate in the last stage cannot be explained only by the instability of the Mg_1 band at RT. This effect together with its dependence on Mg concentration indicates the presence of a precentre obviously connected with the dopant. These precentres might exist in the crystal prior to F-bleaching and are used up during illumination simultaneously with the increase of the formation rate of the Mg_2 band.

Applying a large dose of X-irradiation the Mg_1 band formation saturates very early during F-bleaching either because of the small concentration of the precentres

Table II

Main characteristics of the Mg_1 centre

The Mg_1 centre	
is not paramagnetic	
can be converted by	optical bleaching →F centre X-reirradiation →F centre thermal bleaching →Mg_2 centre (first order kinetics)
formation is promoted by the circumstances	large Mg concentration "as-grown" crystal first stage X-irradiation
formation has stages	I slow II rapid III slow

compared to the F centre concentration, or because long irradiation destroys the precentres themselves.

On the other hand the slow first stage in the formation of the Mg_1 centre indicates that a second precentre should also exist which should be formed during F-bleaching.

The Mg^{2+} ions, like other divalent impurities are built in the crystal together with charge-compensating cation vacancies forming dipoles. The dipoles aggregate into polymers, in the simplest case into dimers the existence of which is proved [5—8].

The experimental facts that:

1. the formation of the Mg_1 centres saturates in quenched crystals (Fig. 9),

2. the long X-irradiation does not promote the formation of Mg_1 centres suggest that the above mentioned first precentre is some kind of aggregated dipoles. In fact, annealing decreases the concentration of polymers [9], and X-irradiation destroys in several cases the divalent metal aggregates [10]. The results concerning this disaggregation effect for the NaCl:Mg system are as follows: X-irradiation

1. disaggregates the $MgCl_2$ precipitate,

2. decreases the concentration of Mg—Mn dimers,

3. probably dissociates the dipoles. (This conclusion may be drawn from the fact that in NaCl:Mg crystals V_F-type centres could be measured by ESR similar to that found in NaCl:Ca [11]. The formation of V_F-type centres is probably connected to the appearance of free cation vacancies.)

The simplest assumption for the first precentre is a dimer. Dimers exist in "as-grown" crystals but annealing and long X-irradiation destroy them. In these cases Mg_1 centres are formed in smaller concentration.

The most plausible assumption for the second precentre (which should be formed during F-bleaching) is an anion vacancy. This view is supported by the observation that the Mg_2 centre formation also requires this precentre (see Figs. 8 and 9) and indeed the anion vacancy is one of the components of the Mg_2 centre according to our model based on other experiments.

Taking into account the above mentioned precentres and the lack of ESR signal the new and more convenient model for the Mg_1 centre is: one dimer + two anion vacancies + two electrons (Fig. 11). Electron excess dimer centres (including or not cation vacancies) are well known also in other systems like Ag_2^+ in KCl [12], $(Cd^+)_2$ in KCl [13] and Tl_2^+ and Ga_2^+ in KCl [14].

The above model is supported by the experience concerning the first order kinetics of the thermal decomposition of the Mg_1 centre and in conversion into Mg_2 centre: one Mg_1 centre dissociates into two Mg_2 centres.

The larger stability of the Mg_2 centre compared to that of the Mg_1 centre can be explained also very well on the basis of the above model. The jump of a cation vacancy needs smaller energy than the moving of a whole vacancy-pair [15]. A vacancy jump can dissociate the Mg_1 centre whereas the same vacancy jump does not lead to

structural change in the Mg_2 centre, only the orientation of the vacancy pair changes. The decomposition of the Mg_2 centre occurs when the whole vacancy pair moves away.

The new model of the Mg_1 centre explains all the experiments without any contradiction.

References

1. A . Watterich and R. Voszka, Acta Phys. Hung., **33,** 323, 1973.
2. R. Voszka, I. Tarján, L. Berkes and J. Krajsovszky, Kristall und Technik, **1,** 423, 1966.
3. R. Voszka and A. Watterich, phys. stat. sol. (b), **55,** 787, 1973.
4. A. Watterich, Phys. Stat. Sol. (b), **88,** K51, 1978.
5. R. Capelletti, Proceedings of the Symp. on Thermal and Photostimulated Currents in Insulators (ed. by D. M. Smyth) p. 1, 1976.
6. M. Ikeya and J. H. Crawford, Phys. Stat. Sol. (b), **58,** 643, 1973.
7. S. Unger and M. M. Perlman, Phys. Rev., B **10,** 3692, 1974.
8. G. Berg, F. Fröhlich and S. Siebenhüner, Phys. Stat. Sol. (a), **31,** 385, 1975.
9. M. Hartmanová, phys. stat. sol. (a), **7,** 303, 1971.
10. A. Watterich and R. Voszka, Phys. Stat. Sol (b), **93,** K161 (1979).
11. A. Watterich, M. Gécs and R. Voszka, Phys. Stat. Sol., **31,** 571, 1969.
12. R. A. Zhitnikov, P. G. Baranov and N. I. Melnikov, Phys. Stat. Sol. (b), **59,** K111, 1973.
13. L. Bern-Dor, A. Glasner and S. Zolotov, J. Sol. St. Chem., **4,** 4, 1972.
14. V. S. Osminin and S. G. Zazubovich, Opt. i Spektr., **29,** 924, 1975.
15. M. Beniere, M. Chemla and F. Beniere, J. Phys. Chem. Solids, **37,** 525, 1976.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 51—57 (1982)

ON THE QUANTUM THEORY OF PARAMETRIC FREQUENCY CONVERTER*

J. JANSZKY

RESEARCH LABORATORY FOR CRYSTAL PHYSICS OF THE HUNGARIAN ACADEMY OF SCIENCES
1502 BUDAPEST, HUNGARY

and

YU. YA. YUSHIN

INSTITUTE OF CRYSTALLOGRAPHY, ACADEMY OF SCIENCES OF THE USSR, MOSCOW, USSR

A quantum mechanical treatment is given of the acoustooptic parametric conversion processes in dielectric crystals when the signal acoustical and idle light waves are transformed into each other in the presence of intensive optical pumping. The approximate Heisenberg equations of motion are found and solved for the creation and annihilation operators of signal and idle modes with due regard to the interaction of these modes with other light and vibratory modes of the crystal ("the thermostat"). It is shown that the thermostat influence results in noise and attenuation effects. These persistent noises are also converted from one mode into another and vice versa. Threshold conditions and asymptotic levels of noise are discussed.

One of the most interesting fields of quantum optics is the investigation of parametric processes of generation, amplification and frequency conversion. In the present paper we shall consider the interaction of three boson modes in dielectric crystals supposing that their frequencies and wave vectors fulfil the phase matching conditions $\omega_1 = \omega_2 + \omega_3$ and $\mathbf{k}_1 = \mathbf{k}_2 + \mathbf{k}_3$. Periodic boundary conditions on the crystal surface are assumed, thus the normal coordinates correspond to running waves. Let mode 3 be intensively excited by an external source. In this case mode 1 will be parametrically converted to mode 2 and vice versa. Such processes play an important role in physical phenomena as interaction between light and ultrasound, coupling of laser and infrared radiation. In the last case the parametric converter can be used for the detection of infrared radiation (up-conversion).

In [1] a quantum mechanical model of parametric conversion was given. In this model the interaction of the signal and idle modes with the remaining electromagnetic and vibratory modes of the crystal was not taken into account (the set of these modes we shall call "thermostat" in the present paper), consequently this model did not deal with noise and attenuation. However, if the parametric frequency converter is used for information processing the attenuation of the signal and the signal-noise ratio are important characteristics of the device. The parametric converter is an example of coupled vibrations. Investigation of noise and damping in such systems is of both theoretical and practical interest.

* Dedicated to Prof. I. Tarján on his 70th birthday.

In this paper we shall consider the quantum theory of the parametric frequency converter taking into account the interaction of signal and idle modes with the thermostat in a microscopic way. Without pumping this interaction causes attenuation and noise, excluding the coherent signal, the average number of quanta in a mode is determined by the temperature of the crystal according to Planck's formula. As soon as the pumping starts the modes get coupled. As a result, both signals and noises are transferred from one mode into another. A quantitative description of noise transformation in the process of parametric frequency conversion is the main result of this paper. Another consequence of mode coupling is the change in the character of attenuation.

The interaction of the signal and idle modes with the modes of the thermostat depends on the type of the modes. To be more specific we shall consider acousto-optical processes in dielectric crystals when the pumping and idle modes are light waves and the signal is a quasi-transverse acoustical wave.

We shall take into account the interaction between these modes proceeding from the Hamiltonian $H=H_0+V_1+V_2$, written in the representation of secondary quantization where H_0 describes non-interacting light and vibratory modes, V_1 is the cubic anharmonicity term and V_2 corresponds to the photon-phonon interaction. We shall find solutions of the approximate equations of motion for time dependent creation and annihilation operators which allow us to study the role of attenuation and noise in the considered process and investigate the statistical properties of the transformed signals.

1. Equations of motion and their approximate solution

Let us consider the interaction between a quasitransverse acoustical and two electromagnetic modes (signal, idle and pumping waves, respectively) described by a coupling constant p in V_2. The crystal temperature is assumed to be so low that the Landau—Rumer criterion is fulfilled for the ultrasonic wave.

For the intensive pumping light we shall use the parametric approximation [1—3], i.e. we shall neglect its damping and substitute C-numbers α_3^* and α_3 in place of its creation and annihilation operators a_3^+ and a_3. Terms in H, which describe the interaction of the optical idle (1) and acoustical signal (2) modes with the thermostat can be written as

$$\lambda_i \hbar (a_i Q_i^+ + a_i^+ Q_i),$$

where $i=1,2$ and the operators Q_1, Q_2 are given:

$$\lambda_1 Q_1 = \hbar^{1/2} \sum_{\mathbf{l},i,j} \varkappa_{\mathbf{k},il,j(\mathbf{k}_1-\mathbf{l})} a_{i\mathbf{l}} [b_{j(\mathbf{k}_2-\mathbf{l})} + b_{j(\mathbf{l}-\mathbf{k}_1)}^+], \tag{1}$$

$$\lambda_2 Q_2 = \hbar^{1/2} \sum_{\mathbf{k},i,j} U_{-\mathbf{k}_2,i\mathbf{k},j(\mathbf{k}_2-\mathbf{k})} [b_{i\mathbf{k}} + b_{j(-\mathbf{k})}^+] [b_{j(\mathbf{k}_2-\mathbf{k})} b_{j(\mathbf{k}-\mathbf{k}_2)}^+]. \tag{2}$$

Here $a_{\mathbf{l}}$ and $b_{j\mathbf{k}}$ are the annihilation operators of the thermostat, photons and phonons, respectively, $\varkappa$ and U are photoelectric and anharmonic coefficients, $\mathbf{l}$ and $\mathbf{k}$ are wave vectors. The small parameters λ_1 and λ_2 are introduced to take into account explicitly the weakness of anharmonicity and of the interaction between light and sound.

Let us assume that the pumping starts at $t=0$. For $t \geqq 0$ from the Hamiltonian H we obtain the Heisenberg equations of motion for the creation operator of a photon of the idle mode $a_1^+(t)$ and that of a phonon of the signal mode $a_2^+(t)$

$$\frac{\mathrm{d}a_1^+(t)}{\mathrm{d}t} = i\omega_1 a_1^+(t) + ig^* e^{i\omega_3 t} a_2^+(t) + i\lambda_1 Q_1^+(t),$$

$$\frac{\mathrm{d}a_2(t)}{\mathrm{d}t} = i\omega_2 a_2^+(t) + ige^{-i\omega_3 t} a_1^+(t) + i\lambda_2 Q_2^+(t). \tag{3}$$

Here $g = -\hbar^{-1}\alpha^* p$ and $p = \hbar^{3/2}\varkappa_{\mathbf{k}_3,\mathbf{k}_1,\mathbf{k}_2}$. Eq. (3) can be supplemented with equations of motion for $Q_i(t)$.

Knowing the solution of Eq. (3) the normally ordered characteristic functions $\chi_i(t,\eta) = \mathrm{Tr}\{\rho e^{\eta a_i^+(t)} e^{-\eta^* a_i(t)}\}$ can be obtained [3] where ρ is the density operator of the whole system in the Heisenberg representation. These characteristic functions make it possible to evaluate the mean value of any operator of $[a_i^+(t)]^m[a_i(t)]^n$ type and they contain information about the statistical properties of the converted signals.

Solving Eq. (3) exactly with respect to g we shall use perturbation theory regarding the small parameters λ_1 and λ_2.

Let us set some initial conditions at $t=0$. For large t we retain only those terms of the perturbative solution of Eq. (3), which correspond to secular terms in the expansion of $\chi_{1,2}(t,\eta)$. To be more detailed: we sum up those terms in the characteristic function which are of $(\lambda_1^2 t)^m (\lambda_2^2 t)^n$ type $(m, n = 1, 2, \ldots)$ and neglect those of $\lambda_i^2 (\lambda_1^2 t)^{n-1} (\lambda_2^2 t)^m$ type. Taking into account the fact that the macroscopic system of the thermostat is only weakly perturbed by the modes under consideration and neglecting the intrinsic anharmonicity of the thermostat we obtain from (3):

$$\frac{\mathrm{d}a_1^+(t)}{\mathrm{d}t} = i(\omega_1 + \Delta\omega_1 + i\Gamma_1)a_1^+(t) + ig^* e^{i\omega_3 t} a_2^+(t) + i\lambda_1 Q_1^{+(0)}(t),$$

$$\frac{\mathrm{d}a_2^+(t)}{\mathrm{d}t} = i(\omega_2 + \Delta\omega_2 + i\Gamma_2)a_2^+(t) + ige^{-i\omega_3 t} a_1^+(t) + i\lambda_2 Q_2^{+(0)}(t), \tag{4}$$

$$\Gamma_2 = \frac{\hbar\pi}{2} \sum_{i,j,\mathbf{k}} |U_{-\mathbf{k}_2, j\mathbf{k}, i(\mathbf{k}_2-\mathbf{k})}|^2 \{[1 + \hat{n}_{j\mathbf{k}} + \hat{n}_{i(\mathbf{k}_2-\mathbf{k})}]$$

$$\delta(\omega_2 - \omega_{j\mathbf{k}} - \omega_{i(\mathbf{k}_2-\mathbf{k})}) +$$

$$+ 2[\hat{n}_{j\mathbf{k}} - \hat{n}_{i(\mathbf{k}_2-\mathbf{k})}]\delta(\omega_{i(\mathbf{k}_2-\mathbf{k})} - \omega_{j\mathbf{k}} - \omega_2)\}, \tag{5}$$

$$\hat{n} = b^+ b.$$

The expression for $\Delta\omega_2$ can be obtained from (5) by substituting $P\left(\frac{1}{x}\right)$ for $\delta(x)$ where $P(x)$ symbolizes the Cauchy principal value. Analogous expressions can be given for Γ_1 and $\Delta\omega_1$. In Eq. (4) and (5) Γ, $\Delta\omega$ and $Q^{(0)}(t)$ are operators acting in the Fock space of states. Each state in this space is characterized by fixed numbers of photons and phonons of the thermostat modes. The above operators shall be evaluated neglecting the interaction of the thermostat with the signal and idle modes.

The solution of the system of operator equations (4) is found to be of the form

$$a_2^+(t) = a_2^+(0)C(t) + a_1^+(0)S(t) + f(t), \tag{6}$$

where

$$C(t) = \exp\left[\left(i\omega_2 + i\Delta\omega_2 - \frac{\Gamma_1+\Gamma_2}{2}\right)t\right]\left(\cos rt + \frac{\Gamma_1-\Gamma_2}{2r}\sin rt\right), \tag{7}$$

$$S(t) = \frac{ig}{r}\exp\left[\left(i\omega_2 + i\Delta\omega_2 - \frac{\Gamma_1+\Gamma_2}{2}\right)t\right]\sin rt, \tag{8}$$

$$f(t) = i\int_0^t \mathrm{d}t'\,[C(t-t')\lambda_2 Q_2^{+(0)}(t') + S(t-t')\lambda_1 Q_1^{+(0)}(t')e^{-i\omega_3 t'}], \tag{9}$$

$$r = \left[|g|^2 - \frac{(\Gamma_1-\Gamma_2)^2}{4}\right]^{1/2}.$$

The expression for the operator $a_1^+(t)$ can be found with subscripts 1 and 2 interchanged in (6—9).

By direct calculation one can check that in the sense of the above mentioned approximation the obtained solutions obey the commutation relations:

$$[a_i(t), a_j^+(t)] = \delta_{ij}, \qquad i,j = 1,2. \tag{10}$$

If the attenuation is introduced into the equation of motion of $a_i^+(t)$ in a phenomenological way, the solutions of such a system of equations do not obey the commutation relation (10).

2. Statistical properties of the signals

Let us assume that at $t=0$ the density operator of the whole system is $\rho = \rho_1\rho_2\rho_T$, where ρ_T is the density operator of the thermostat. It is also supposed that at $t=0$ the density operators of the idle and signal modes ρ_1 and ρ_2 are described in the P-representation of the coherent states [3] by weight functions $P_1^{(0)}(\beta)$ and $P_2^{(0)}(\alpha)$, respectively. The characteristic function $\chi_2(\eta, t)$ of the signal mode is given:

$$\chi_2(\eta, t) = \iint \mathrm{d}^2\alpha\,\mathrm{d}^2\beta\, P_2^{(0)}(\alpha)P_1^{(0)}(\beta)G(\eta, t; \alpha, \beta), \tag{11}$$

where the kernfunction

$$G(\eta, t; \alpha, \beta) = tr_T\{\rho_{T2} <\alpha|_1 <\beta| e^{\eta a_2^{\dagger}(t)} e^{-\eta^* a_2(t)} |\beta>_1 |\alpha>_2\} \tag{12}$$

is the characteristic function of the signal mode if both the signal and idle modes were in pure coherent states $|\alpha>_2$ and $|\beta>_1$ at $t=0$. Using (6—8) and retaining only the secular terms one can evaluate $G(\eta, t; \alpha, \beta)$

$$G(\eta, t; \alpha, \beta) = \exp\{-|\eta|^2 \mu(t) + \\ + \eta[\alpha^* c(t) + \beta^* s(t)] - \eta^*[\alpha c^*(t) + \beta s^*(t)]\}, \tag{13}$$

where $s(t)$, $c(t)$ can be obtained from $S(t)$ and $C(t)$ by the replacement of operators by C-numbers:

$$\Gamma_{1,2} \to \gamma_{1,2} = tr_T(\rho_T \Gamma_{1,2}); \quad \Delta\omega_{1,2} \to tr_T(\rho_T \Delta\omega_{1,2});$$

$$\mu(t) = 2 \int_0^t dt' \{\gamma_2 n_2(T) |c(t')|^2 + \gamma_1 n_1(T) |s(t')|^2\}. \tag{14}$$

Here $n_{1,2}(T)$ are the occupation numbers of the idle and signal modes in thermodynamical equilibrium. Let us suppose that at $t=0$ each of the considered modes is in a state that corresponds to the superposition of coherent signal with equilibrium thermal noise:

$$P_{1,2}^{(0)}(\alpha) = \frac{1}{\pi n_{1,2}(T)} \exp\left\{-\frac{|\alpha - z_{1,2}|^2}{n_{1,2}(T)}\right\}. \tag{15}$$

Substituting (13) and (15) into (11) one finds the characteristic function of the signal mode

$$\chi_2(\eta, t) = \exp\{-|\eta|^2 N(t) + \eta R^*(t) - \eta^* R(t)\}, \tag{16}$$

and the corresponding weight function

$$P_2(\alpha, t) = \frac{1}{\pi N(t)} \exp\left\{-\frac{|\alpha - R(t)|^2}{N(t)}\right\}, \tag{17}$$

where

$$R(t) = z_2 c^*(t) + z_1 s^*(t) \tag{18}$$

and

$$N(t) = n_2(T)|c(t)|^2 + n_1(T)|s(t)|^2 + \mu(t). \tag{19}$$

One can see that Eqs (16, 17) describe a superposition of a coherent state having a parametrically transformed amplitude $R(t)$ with a Gaussian noise with variance $N(t)$.

3. Discussion

As it can be seen from Eqs (6—9) there are two different dynamic conditions of the parametric converter depending on the relation between $|g|$ and $|\gamma_1-\gamma_2|/2$. Let us assume that there is a pure coherent signal z_1 in mode 1 at $t=0$. If $|g|>|\gamma_1-\gamma_2|/2$ the signal will periodically transform from mode 1 to mode 2 and vice versa with a damping amplitude. In this case the damping constant is equal to $(\gamma_1+\gamma_2)/2$. If, on the contrary, $|g|<|\gamma_1-\gamma_2|/2$, there is also some conversion but the converted signal will attenuate without oscillation. In this case at $g\to 0$ Eqs (16—19) transform into a form describing non-interacting modes.

To understand the cause of the secondly mentioned dynamic duty of the parametric converter let us consider two oscillators with weak and strong dampings, respectively. Let us assume that at $t=0$ only the oscillator with weak damping is excited and a weak parametric coupling begins at that moment. The energy transfer from the first oscillator to the second one will be slower than the potential velocity of energy loss by the second oscillator. Therefore there will be only a one way energy conversion between these oscillators.

For $|g|>|\gamma_1-\gamma_2|/2$, using the characteristic function, let us evaluate the average number of quanta in mode 2 if at $t=0$ there were signals in both modes. From Eq. (17) we have:

$$\langle a_2^+(t)\, a_2(t)\rangle = |R(t)|^2 + N(t)\,, \tag{20}$$

The first term $|R(t)|^2$ in Eq. (20) describes the kinetics of the signal intensity, the second term $N(t)$ shows the level of noise in mode 2.

The expressions $N(t)$ in Eq. (19) can be divided into two parts:

$$N_i(t) = N(t) - \mu(t) \quad \text{and} \quad \mu(t)\,. \tag{21}$$

The quantity

$$N_i(t) = n_2(T)\exp\left[-(\gamma_1+\gamma_2)t\right]\left(\cos rt + \frac{\gamma_1-\gamma_2}{2r}\sin rt\right)^2 +$$

$$+ n_1(T)\frac{|g|^2}{r^2}\exp\left[-(\gamma_1+\gamma_2)t\right]\sin^2 rt \tag{22}$$

originates from the initial levels of noise of the signal and idle modes. It should be noticed that in this expression the temperature T can be different from the crystal temperature (e.g. in case of ballistic phonons [4]). In this case if $z_{1,2}=0$ Eq. (19) describes parametric conversion of input signals with Gaussian statistics.

The term

$$\mu(t) = N_a + N_t(t)\,, \tag{23}$$

where

$$N_a = (|g|^2+\gamma_1\gamma_2)^{-1}\left[n_2\gamma_1\gamma_2 + \frac{\gamma_1 n_1+\gamma_2 n_2}{\gamma_1+\gamma_2}|g|^2\right] \tag{24}$$

and

$$N_t(t) = -e^{-(\gamma_1+\gamma_2)t}\left\{\frac{|g|^2(n_2\gamma_2+n_1\gamma_1)}{r^2(\gamma_1+\gamma_2)} + \right.$$

$$+\frac{\cos 2rt}{2(|g|^2+\gamma_1\gamma_2)}\left[\frac{|g|^2(\gamma_1+\gamma_2)(n_2\gamma_2-n_1\gamma_1)}{2r^2} + n_2\gamma_2(\gamma_1-\gamma_2)\left(1-\frac{\gamma_1^2-\gamma_2^2}{4r^2}\right)\right]+$$

$$\left.+\frac{\sin 2rt}{2(|g|^2+\gamma_1\gamma_2)}\left[\frac{|g|^2(n_1\gamma_1-n_2\gamma_2)}{r} + n_2\gamma_2(\gamma_1-\gamma_2)\left(1-\frac{\gamma_1-\gamma_2}{2r}\right)\right]\right\} \tag{25}$$

describe Gaussian noise in mode 2 connected with the persistent influence of the thermostat on modes 1 and 2 in the interval of time from 0 to t. The quantity $N_t(t)\to 0$ in the limit $t\to\infty$ and corresponds to rather complicated transient conditions of noise level. The term N_a describes the asymptotic level of noise in mode 2 in the limit $t\to\infty$.

From Eq. (24) in the limit $g\to 0$, as it can be expected, we have $N_a\to n_2(T)$. On the contrary, if $|g|\gg\gamma_1\gamma_2$, then

$$N_a \approx \frac{\gamma_1 n_1(T)+\gamma_2 n_2(T)}{\gamma_1+\gamma_2}. \tag{26}$$

In the last case (i.e. $|g|\gg\gamma_1\gamma_2$) the number of quanta of noise in mode 1 has the same limit. This limit will be between the equilibrium numbers of quanta in modes 1 and 2; $n_1(T)$ and $n_2(T)$. In case of acoustooptical conversion the asymptotic level of noise in the optical idle mode will be higher than that which corresponds to Planck's formula and the asymptotic level of noise in the acoustical signal mode will be correspondingly lower than in thermal equilibrium.

References

1. W. H. Louisell, A. Yariv and A. E. Siegman, Phys. Rev., **124,** 1646, 1961.
2. W. H. Louisell, Coupled Mode and Parametric Electronics, John Wiley and Sons, New York, 1960.
3. B. R. Mollow and R. J. Glauber, Phys. Rev., **160,** 1076, 1967.
4. A. A. Kaplyanskii, Colloques intern. CNRS, No 255, 137, 1977.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 59—66 (1982)

ON THE ROLE OF DISLOCATION GROUPS IN THE FORMATION OF DOMINATING GROWTH HILLS*

L. MALICSKÓ

RESEARCH LABORATORY FOR CRYSTAL PHYSICS OF THE HUNGARIAN ACADEMY OF SCIENCES
1502 BUDAPEST, HUNGARY

The layer structure of the hills on growth faces and of slightly evaporated cleavage faces of KCl crystals grown from solutions were studied by the surface decoration replica technique in a transmission electron microscope. The layer structure of the growth hills and the evaporation pits found is described and the role of the collective activity of dislocations as layer sources in the formation of the dominating growth hills is discussed.

Introduction

Optical microscopic investigations often showed the presence of flat layered hills on the growth faces of polyhedral crystals grown from solutions or vapour phases (see e.g. [1—3]). The central regions of these hills, the so-called growth centres, seemed to be the most active sources of growth layers on the crystal faces. From the point of view of growth these centres play a positive role. The formation of these hills showing often spiral terrace forms was explained by the dislocation mechanism, i.e. by the so called BCF theory of crystal growth [4]. For the interpretation of the great step heights found in some cases at the spiral terraces the existence of dislocations having Burgers vectors much larger than the lattice parameters had to be assumed in the middle part of the spirals [5]. But this assumption involves different energetic problems.

In the dislocation theory of the crystal growth it was also pointed out that groups of screw dislocations belonging to some grain boundary of twist character can operate during growth as collective sources of atomic layers with an "activity" which differs from that of the single screw dislocations [4, 6]. "Activity" means here the number of the atomic layers emitted by a source in unit time.

It was shown that beside the positive role of the growth centres these centres may also be sources of compositional inhomogeneities and mechanical stresses during growth [7, 8], which means that from the point of view of crystal quality they can play a negative role. So, it is an important practical question: where and how are the growth hills formed?

* Dedicated to Prof. I. Tarján on his 70th birthday.

The aim of the present investigations carried out on KCl crystals was to give further contributions to the knowledge about the role of dislocations in the crystal growth, especially in the formation of dominating growth hills.

Experiments

The alkali halides are classical model materials for the investigation of crystal defects. For the present experiments KCl single and bicrystals were grown from solutions by the temperature decreasing or evaporation method described earlier [7, 9]. To facilitate the growth the solutions were doped with 0.1 mole% $PbCl_2$. Seeds of dimensions of about $15 \times 7 \times 3\ mm^3$ were cleaved from nominally pure melt-grown crystals and selected in two groups. The first series of seeds contained only single dislocations randomly distributed along the starting cleavage faces according to previous optical microscopic investigations. The second series of seeds showed in addition to single dislocations also straight and curved trace lines of randomly misoriented grain boundaries.

The growth of crystals occurred with rates in order of 10^{-4} μm/s. As finish of the growth process the crystals were quickly lifted from the solutions, dried on very fine filtering paper and then stored in desiccators. The layer structures of such as-grown and dried faces were previously studied in an optical microscope. Surface parts having only one growth hill and also inner parts of the crystals were selected and cleaved out as samples of dimensions of about $10 \times 8 \times 2\ mm^3$. After appropriate drying and evaporation of these samples in high vacuum surface gold decoration carbon replicas were made in the usual way for transmission electron microscopic studies [10].

Results and discussion

The KCl crystals grown from solutions were mainly bordered by the {100} habit faces as it is shown in Fig. 1. On almost all {100} faces of the crystals grown from both series of seeds one could observe growth hills. On the optical micrograph in Fig. 2 one can see such a growth hill being formed on a (100) face of a crystal without the operation of any grain boundary. The number of the hills was generally small: from about 1 up to 10 per cm^2. On the basis of other investigations it was known that the surface density of single dislocations on the growth faces could amount to about $10^6\ cm^{-2}$. Considering the discrepancy between the surface density of growth hills observed and that of single dislocations it follows that only a small part of the dislocations emerging onto the growing face is responsible for the formation of the dominating hills. This fact suggests that the growth centres are layer sources consisting possibly of more dislocations which form collectively the hills.

According to earlier theoretical considerations [4, 6] it is to be expected that dislocations arranged in lines, e.g. in grain boundaries, can form together step sources

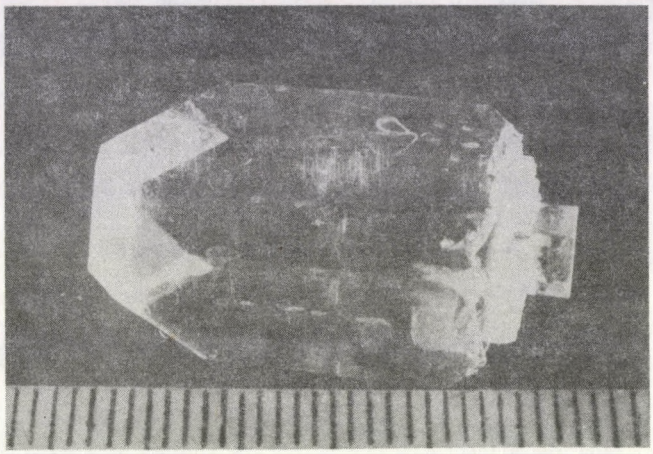


Fig. 1. The form of a solution-grown KCl crystal bordered mainly by {100} habit faces. The dimensions are marked by the mm scale

of great activity. So, if the seed contains a grain boundary, its full trace line is theoretically favourable for the growth of the crystals. Consequently, the trace lines remain more or less visible also on the growing face. This is demonstrated by the optical micrograph in Fig. 3. In this way the formation of dominating growth hills could be observed to occur mainly at some sections along the trace lines of the grain boundaries. Not all parts of a given grain boundary emerging in the crystal face are uniformly active from the point of view of forming growth hills. The observable break points or sections in the directions of the trace lines seemed to be mostly active, as it can be seen also in the case of Fig. 3.

The evaporation processes are analogous to the growth ones. Therefore the layer structure of the evaporated cleavage faces of the KCl crystals was also investigated by TEM using surface gold decoration carbon replicas. In order to be able to compare the activities of the evaporation layer sources on the basis of the step structure belonging to them, some known characteristic step structures which can be found on the evaporated cleavage faces are shown in Fig. 4.

One can see that the activity of the single screw dislocation and that of the single egde one as surface step sources (centres) is about the same (see Figs 4a and 4b). The step structure around the strongly poisoned dislocations in Fig. 4c verifies their greater source activity compared to single dislocations (Figs 4a and 4b).

The fact that only certain parts of a given grain boundary are active to form dominating step sources was verified also on the atomic scale. This is demonstrated by the atomic evaporation step arrangement which is formed along the closed trace line of a grain boundary in Fig. 5. Only the grain boundary sections and break point marked by arrows formed evaporation pits (evaporation centres) analogous to those in Fig. 4.

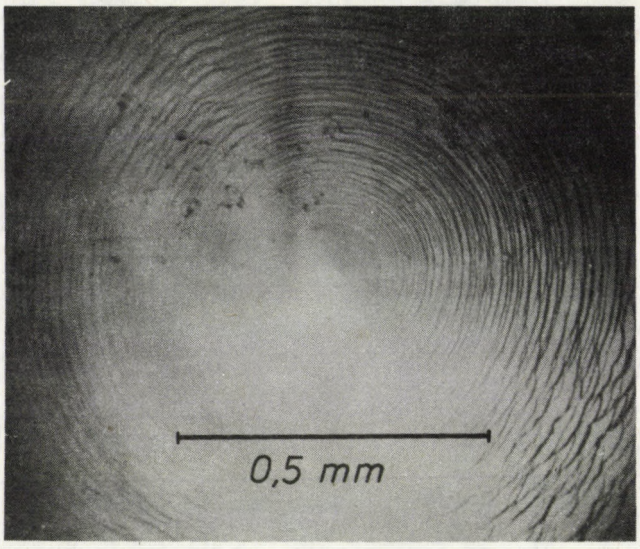


Fig. 2. A layered flat hill on a (100) growth face of a KCl single crystal grown from solution. Optical micrograph

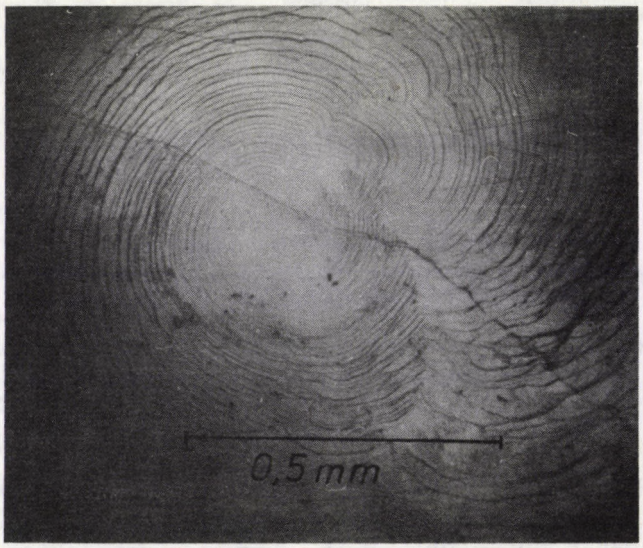


Fig. 3. A dominating growth hill formed at a break section of the trace line of a grain boundary on a (100) face of a KCl crystal. Optical micrograph

Depending on the local dislocation arrangement of the grain boundaries their "active" sections can have very different step source activity. In Fig. 6b a break point of a grain boundary trace line can be seen. Around this break point an evaporation pit was formed with a step structure like that of a single dislocation (cf. Figs. 4a, 4b and 6a). That means its activity was about the same as that of the single dislocations. In contrast to that Fig. 6c shows a straight line section of a grain boundary along which an evaporation pit was formed with approximately quadratic steps of one and more atomic heights (compare with Fig 4c) verifying a very great collective activity of this boundary section. In rough outlines this step system shows a spiral form like the etch spirals having high step heights described and observed earlier (e.g. [11, 12]).

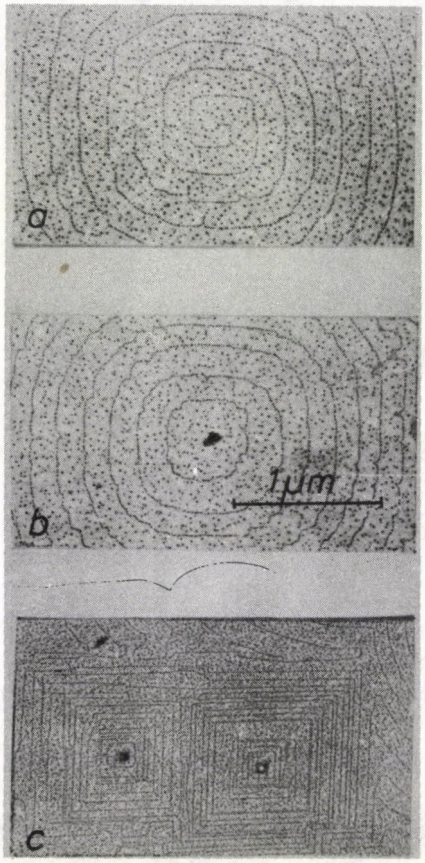


Fig. 4. Transmission electron micrograph of some typical evaporation layer structures decorated with gold crystallites on a (100) face of a KCl crystal:
a — atomic spiral step caused by a single screw dislocation,
b — concentric circular atomic step system caused by a single edge dislocation and
c — systems having nearly quadratic steps of one or more atomic heights around strongly poisoned dislocations

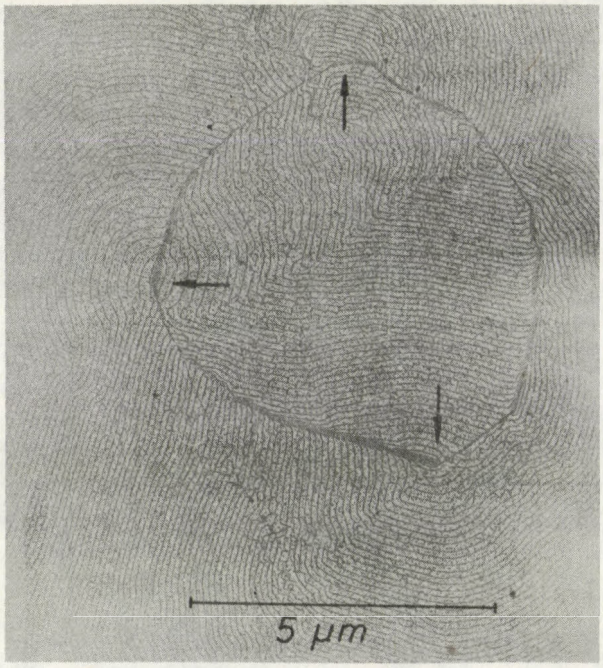


Fig. 5. Transmission electron micrograph of the decorated atomic step structure along the closed trace line of a small grain embedded in its surroundings with a small angle misorientation. Only the grain boundary sections marked by the arrows, acted as step sources like those in Fig. 4

On the replicas made from the central parts of the growth hills formed without the presence of any grain boundary after a slight evaporation of the original growth faces groups of small hills could be observed. Such square shaped small hills are shown in Fig. 7a. All such small hills forming together the big growth hill consist generally of a few single dislocations as it is shown in Fig. 7b. That means, at a given growth condition the single dislocations (screw and edge*) distributed favourably in a small area of the growing crystal face can act through groupings as a collective layer source of very high activity and so they form a dominating growth hill during growth.

From the results one can conclude that small surface regions containing groups of single screw and edge dislocations favourably arranged either in a linear distribution predicted by the theories or in a surface distribution as found now can have a collective step source activity which is higher than that of the single dislocations. In consequence of their very high collective activity they build together dominating growth centres forming the big flat growth hills observable in the light microscope or in some cases even with the naked eye.

* *Note added in proof:*

The fact that the edge dislocations can act as sources of monomolecular *growth* steps ("transverse step sources"), was directly proved by E. Bauser and H. Strunk in J. Cryst. Growth, **51**, 362, 1981.

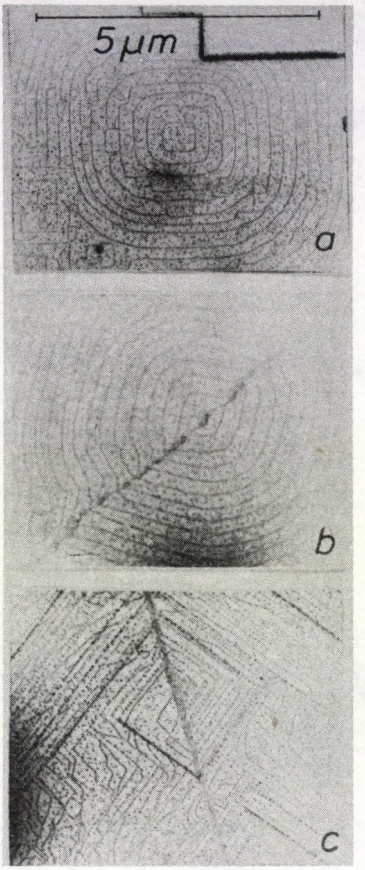


Fig. 6. Transmission electron micrographs of decorated evaporation step arrangements in surroundings of step sources of different activity:
a — concentric circular atomic steps formed around a single edge dislocation (for comparison),
b — a break section on a grain boundary trace line acting as a step source having an activity like a single dislocation, and
c — a step arrangement around a straight section of a grain boundary forming a collective step source of very high activity. The quasiquadratic steps of one or more atomic heights form together an interesting spiral evaporation pit

Acknowledgement

The author would like to thank Prof. Dr. R. Voszka for supporting the investigations, and also Mrs. M. Kiss, Mrs. G. Böszörményi-Nagy and Mrs. K. Farkas for their assistance in the experimental work.

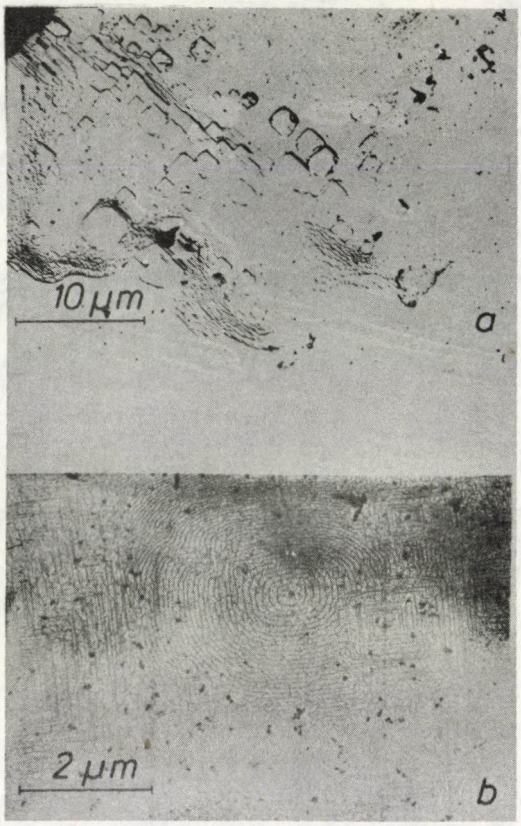


Fig. 7. Transmission electron micrographs about the surface structure in the central part of an optical microscopically visible growth hill:
a — the group of square shaped small hills in the central part,
b — the decorated step structure caused by the group of single dislocations forming one of the small groth hills shown in Fig. 7a

References

1. C. W. Bunn, in H. E. Buckley, Crystal Growth, J. Wiley and Sons Inc., New York, 1951. (in Russian, Izd. Inostr. Lit., Moscow, 1954. p. 157).
2. G. G. Lemmlejn and N. V. Gliki, Dokl. AN USSR, **94,** 473, 1954.
3. A. Williams, Phil. Mag., **2,** 635, 1957.
4. W. Burton, N. Cabrera and F. Frank, Phil. Trans., *A***243,** 299, 1951.
5. H. G. Van Bueren, Imperfections in Crystals, North-Holl. Publ. Co. Amsterdam, 1960. (in Russian, Izd. Inostr. Lit., Moscow, 1962. p. 410).
6. A. R. Verma, Crystal Growth and Dislocations, Butterworths Sci. Publ., London, 1953.
7. L. Malicskó and L. Jeszenszky, J. Cryst. Growth, **15,** 243, 1972.
8. L. Malicskó, Krist. u. Technik,**14,** 691, 1979.
9. L. Malicskó and L. Jeszenszky, J. Cryst. Growth, **7,** 13, 1970.
10. H. Bethge, Phys. Stat. Sol., **2,** 755, 1962.
11. S. Amelinckx, W. Bontinck and W. Dekeyser, Phil. Mag., **2,** 1264, 1957.
12. B. Jeszenszky, Acta Phys. Hung., **8,** 147, 1957.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 67—73 (1982)

THE EFFECT OF DOPANTS ON THE GROWTH AND PROPERTIES OF TGS CRYSTALS*

L. MALICSKÓ, L. JESZENSZKY, E. HARTMANN, L. KOVÁCS, Á. PÉTER and B. VAJNA

RESEARCH LABORATORY FOR CRYSTAL PHYSICS OF THE HUNGARIAN ACADEMY OF SCIENCES, 1502 BUDAPEST, HUNGARY

Triglycine sulphate (TGS) crystals were grown from pure and doped solutions. The results referring to the influence of the dopants on the growth, the domain structure, the lattice parameters and the different dielectric parameters are reported and discussed.

Introduction

Triglycine sulphate (TGS) is well known as a ferroelectric material [1]. Its properties are generally known [2]. Because of its low Curie temperature of about 49 °C TGS is a convenient ferroelectric model material from which single crystals of good quality can be grown. The crystals are also used in the practice as pyroelectric displays and infrared detectors. The technical applications demand single crystals of high quality. Therefore, the study of the relation between the growth conditions and the properties of these crystals are of practical importance, too. The aim of the present paper was to give some contributions to the effect of dopants on the growth in macroscopic scale and to the ferroelectric behaviour of the TGS crystals.

Experiments

Chemical preparations

The triglycine sulphate was synthetized from hot aqueous solutions containing in stoichiometric composition glycine and sulphuric acid of high purity [3]. In order to decrease the unwanted metal impurities, first of all that of the Fe ions, the glycine raw material was prepurified by recrystallizing two times. The steady composition of the starting material was controlled by potentiometric titration [4]. As dopants pure $CuSO_4 \cdot 5\,H_2O$ and $KCr(SO_4)_2 \cdot 12\,H_2O$, respectively, were used whose concentrations within the solutions were varied between 2 and 25 mole% in order to grow crystals of rather large dopant contents.

* Dedicated to Prof. I. Tarján on his 70th birthday.

The concentrations of Cu dopants incorporated into the TGS crystals were determined by complexometric titration [5]. As reagent 10^{-3} M solution of SELECTON C (Complexon IV) produced by Reanal were used in specimen solutions of constant 0.5 M TGS concentrations and at constant 5.8 pH values. As indicator PAN was applied. The concentrations of Cr dopants within the crystals were measured spectrophotometrically [6, 7] using 0.1 M solutions of SELECTON B2 (EDTA) produced by Reanal in specimen solutions of constant 0.5 M TGS concentrations, at constant 6.00 pH values. The Cr concentrations were determined from the absorption peak heights measured at 548 nm.

For revealing the domain structure on the {010} cleavage faces of the crystals instead of the usual water (e.g. [3, 8]) cc. propion acid doped with a small amount of $BaCO_3$ was used as etchant.

Crystal growth

The crystals for the investigations were grown from pure and doped aqueous solutions by the well-known techniques of the solvent evaporation at constant temperature or of a program regulated decreasing of the temperature. The growth processes of two steps occurred below the Curie temperature of the crystals, that means, the growth history took place in the ferroelectric state. In the first step going out from rod-like seeds nearly uniform small crystals of dimensions of about 2 cm were grown from pure solutions. They were usually bordered by {001} and {110} habit faces and {100}, {111}, {121}, {011} and {010} faces of small sizes. In the second step these small crystals were grown on.

Measurements of the dielectric parameters

From the crystals small specimens were cleaved out the {010} cleavage faces of which were covered with silver paste as electrodes. Then at an electric field strength of 100 Vm^{-1} and a frequency of 1 kHz the relative dielectric permittivity (ε_r) and the dielectric loss (tg δ) of the specimens were measured as function of the temperature between about 20 and 60 °C with an accuracy better than ± 0.1 °C.

The ferroelectric hysteresis loops were measured in a modified Sawyer-Tower circuit at an electric field strength of $2 \cdot 10^5$ Vm^{-1} and a frequency of 50 Hz. In this way the spontaneous polarization P_s and the coercitive force E_k belonging to the temperature of 20 °C were determined.

Measurements of the lattice parameters

On powder specimens prepared from pure and copper doped crystals the lattice parameters were measured with Debye-Scherrer method. On cleavage pieces of the pure and doped crystals precision lattice parameter measurements were carried out with the Bond method [9, 10].

Results and discussion

In growing the crystals using the seeds mentioned before in pure solutions the bordering, the so-called *growth form* of the crystals remains the same as that of the seeds (Fig. 1a). After the growth from solutions doped with copper or chromium and potassium, respectively, a change in the size ratios of the faces were observed (Fig. 1b). In the bordering the {001} and {110} faces remain the habit faces, but the other ones become relatively smaller or some of them even disappear. This growth form modification caused by the selected effect of the dopants on the growth of the crystallographically different faces is bigger in the case of the Cr and K dopants than in that of Cu dopants.

Depending on the growth conditions the dopant particles present in the solutions are incorporated into the crystals during the growth. By measuring the dopant concentrations in the crystals the effective segregation coefficients of Cu and Cr dopants were determined. Increasing the growth rate from 0.2 up to 10 mm/day measured in the ⟨010⟩ direction the effective segregation coefficients have shown a rising tendency between $5.0 \cdot 10^{-3}$ and $3.6 \cdot 10^{-2}$ in the case of Cu dopants and between $2 \cdot 10^{-2}$ and $2.3 \cdot 10^{-1}$ in the case of Cr ones. The segregation coefficients for both kinds

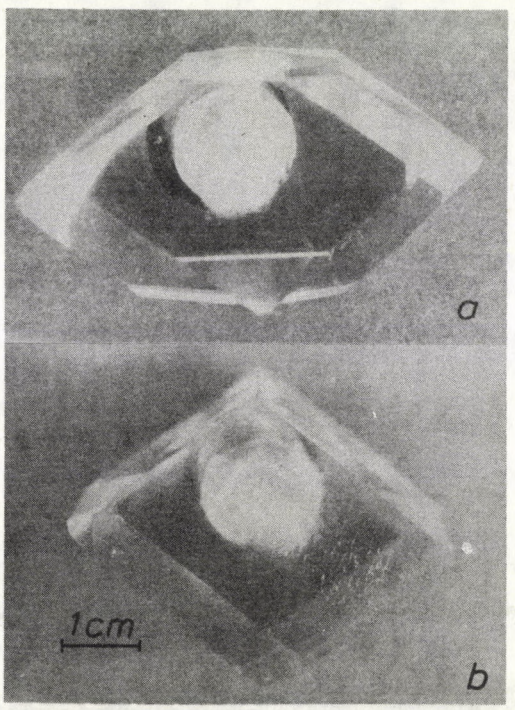


Fig. 1. The growth forms of TGS crystals grown from a pure solution (a) and from that doped with 10 mole% Cu (b)

of dopants are smaller than unity, and their values lie in that order of magnitude which was found also by other authors [11].

According to the etching experiments the crystals in the as-grown state consist only of a few of the ferroelectric domains. After a thermal treatment above the Curie temperature the number of the newly formed domains, especially in the case of the doped crystals, increases significantly. The domain form appearing by etching on the {010} cleavage faces of the crystals shows certain differences according to the dopant species. As Fig. 2 demonstrates on the faces of the pure crystals the domains are generally lens-like (Fig. 2a) and on those of the doped crystals they have generally elongated forms (Fig. 2b). In the case of Cr dopants whisker forms were also observed (Fig. 2c).

The lattice parameters measured on TGS crystals doped with Cu in concentrations between 0.05 and 0.6 mole% were: $a=(0.942\pm0.010)$ nm, $b=$ $=(1.262\pm0.010)$ nm, and $c=(0{,}573\pm0{,}007)$ nm. They do not differ from those measured on the pure crystals and given in the International Tables. That means that the possible changes of the parameters are within the given limit of error. The precision

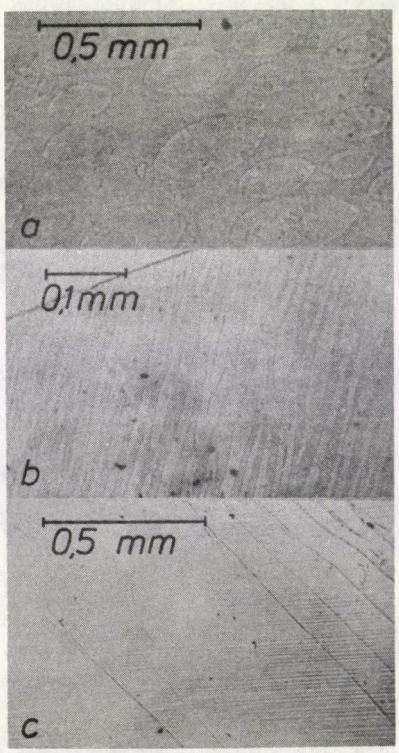


Fig. 2. Some typical forms of ferroelectric domains after a thermal treatment, revealed by etching on the {010} cleavage faces of different TGS crystals: a — in a pure crystal, b — in a Cu doped crystal, and c — in a Cr and K doped crystal

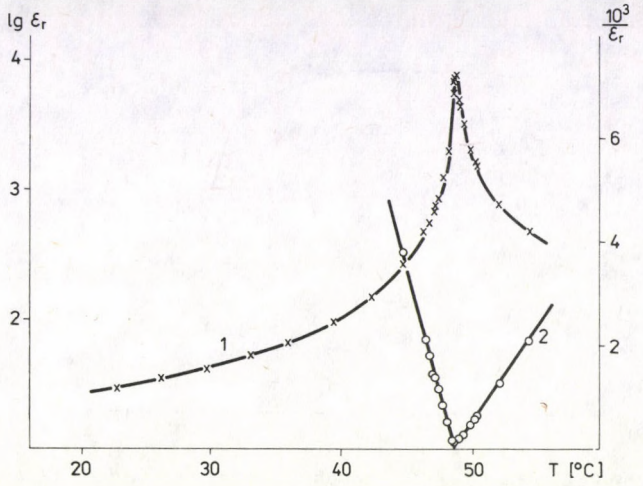


Fig. 3. The temperature T dependence of the relative dielectric permittivity ε_r of a Cu doped TGS crystal: 1 — in a $\lg \varepsilon_r$ versus T representation; 2 — in a $1/\varepsilon_r$ versus T representation

measurements of the parameter b show, however, certain differences: in the undoped crystal $b=(1{,}264\,824\,9\pm5\cdot10^{-7})$ nm, in the crystal doped with 0.1 mole% Cu $b=$ $=(1.264\,828\,7\pm5\cdot10^{-7})$ nm and in the crystal doped with 0.2 mole% Cr (together with K in unknown concentration) $b=(1.264\,732\,3\pm5\cdot10^{-7})$ nm. From this especially the Cr and K dopants seem to be effective in decreasing the parameter b of the TGS crystals.

The relative dielectric permittivities of the crystals have shown the well-known temperature dependent course. This is demonstrated by the curve 1 in Fig. 3 for the case of a TGS crystal doped with Cu in a heating course. At the phase transition temperature T_m a very high maximum ($\varepsilon_{r,\max}$) appears as compared with the starting

Table I

The dielectric parameters of TGS crystals, measured at heating (h) and cooling (co), respectively. C^-, C^+ are the Curie constants in ferroelectric and paraelectric phase, respectively. $\varepsilon_{r,\max}$ is the maximum relative dielectric permittivity at the phase transition temperature T_m. T_c is the Curie temperature

Dopant	C^- [10^3 °C]		C^+ [10^3 °C]	$\varepsilon_{r,\max}$ [10^3]	T_m [°C]	T_c [°C]
—	0.90±0.01	(h, co)	3.03±0.04 (h, co)	8.6±0.2	48.9±0.1	48.6±0.1
Cu	1,06±0.02	(h)	2.87±0.02	7.7±0.1	48.8±0.1	48.4±0.1
	(0.9...1.7)	(co)	(h, co)			
Cr (and K)	0,90±0,01	(h)	2.63±0.03	8.0±0.1	49.1±0.1	48.7±0.1
	(0.9...1.75)	(co)	(h, co)			

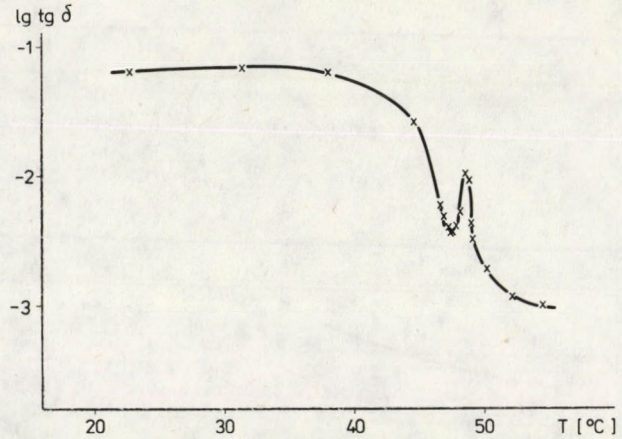


Fig. 4. The temperature T dependence of the logarithmic dielectric loss lgtg δ of a Cu doped TGS crystal

value of the dielectric permittivity at room temperature. In the as-grown crystals having a prehistory only below the Curie temperature these starting values were found to be about 30. Both the ferroelectric and the paraelectric parts of the temperature T dependence of the dielectric permittivity in the vicinity of the Curie point T_c may be described by the Curie—Weiss law: $1/\varepsilon_r = (T - T_c)/C$, where the Curie constant C for the ferroelectric phase C^- differs from that for the paraelectric phase C^+.

The dielectric parameters found are summarized in Table I. The data of Table I present the following details. The Curie constants measured in the paraelectric phase are reproducible independent of the heating or cooling course of the thermal treatment. The Curie constants measured in the ferroelectric phase show generally a larger or smaller practically irreproducible hysteresis. Such hystereses were already found on TGS by other authors [12].

Both in the undoped and doped TGS crystals the dielectric loss showed a strong decrease with an increasing temperature, and a sharp maximum appeared around the ferroelectric phase transition temperature (Fig. 4). The values of the spontaneous polarizations and the coercitive forces determined from the measurements of the hysteresis loops of the crystals are listed in Table II.

Considering that the maximum of the dielectric permittivity and the Curie temperature can be influenced by the defects within the crystals [13] the values of $\varepsilon_{r,\max}$, T_m and T_c listed in Table I are in good agreement with those mentioned in the literature [14, 15]. The hysteresis and the differences in Tables I and II or found in the literature are believed to lead back to the complex interactions between the lattice defects, e.g. impurities, and the domain structure. As it was already mentioned the as-grown domain structure can differ considerably from that after thermal treatments. Consequently, the reproducibility of the dielectric parameters is the question of the reproducibility of the domain structure of the ferroelectric crystals.

Table II

The values of the spontaneous polarizations P_s and the coercitive forces E_k at 20 °C, determined from the hysteresis loops investigated on (010) oriented cleavage pieces of pure and doped TGS crystals

Dopant	P_s [10^{-2} Cm^{-2}]	E_k [10^4 Vm^{-1}]
—	2.8 ± 0.2	9.1 ± 0.5
Cu	1.7 ± 0.1	7.5 ± 0.3
Cr (and K)	1.6 ± 0.1	8.3 ± 0.4

Summary

By measuring the dielectric parameters the crystals produced in our laboratory were controlled. Furthermore, the present investigations have given some new contributions to the close connection among the growth conditions, the real structure and the physical properties of the TGS crystals.

Acknowledgement

The authors would like to thank Dr. J. Stepien-Damm for her help and advice in carrying out the precision lattice parameter measurements at the Institute of Low Temperature and Structure Research of the Polish Academy of Sciences, Wrocław, furthermore Mrs. K. Farkas and Mrs. S. Tóth for their technical assistance.

References

1. B. T. Matthias and J. P. Remeika, Phys. Rev., **103,** 262, 1956.
2. Landolt-Börnstein, New Series III/11. Berlin. 1979.
3. V. P. Konstantinova, J. M. Silvestrova and K. S. Alexandrov, Kristallografiya, **4,** 69, 1959.
4. J. Sörensen, Biochem. Z., **7,** 43, 1907.
5. G. Schwarzenbach and H. Flaschka, Die Komplexometrische Titration, Enke, Stuttgart, 1965.
6. R. Pribil and J. Klubakova, Collection Czechosl. Chem. Comm., **15,** 42, 1950.
7. V. Komeswara Rao, D. S. Sundar and M. N. Sastri, Z. Anal. Chem., **218,** 93, 1966.
8. L. Taurel and F. Gilletta, Proc. Int. Meeting on Ferroelectricity, Vol. II., p. 43, Prague, 1966.
9. W. L. Bond, Acta Cryst., **13,** 814, 1960.
10. K. Lukaszevicz, D. Kucharczyk, M. Malinowski and A. Pietraszko, Kristall u. Techn., **13,** 561, 1978.
11. M. S. Tsedrik and E. M. Kravchenya, Kristall u. Techn., **11,** 49. 1976.
12. J. Mazur and A. Jaskievicz, Acta Phys. Pol., **34,** 859, 1968.
13. Z. Malek, M. Polcarova, J. Strajblova and J. Janta, Phys. Stat. Sol. (a), **11,** 195, 1972.
14. A. Levstik, M. Burgar and R. Blinck, Phys. Stat. Sol. (a), **18,** K113, 1973.
15. Z. Malek, J. Moravec and J. Strajblova, Czechosl. J. Phys. *B***19,** 1185, 1969.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 75—80 (1982)

LOW TEMPERATURE YIELD STRESS INCREMENT DUE TO IMPURITY-VACANCY DIPOLES IN NaCl CRYSTALS*

A. TÓTH, P. KÁLMÁN, T. KESZTHELYI and J. SÁRKÖZI

DEPARTMENT OF EXPERIMENTAL PHYSICS, INSTITUTE OF PHYSICS, TECHNICAL UNIVERSITY
1521 BUDAPEST, HUNGARY

The yield stress increment of NaCl crystals caused by impurity-vacancy dipoles containing Mg^{2+}, Sr^{2+} and Ba^{2+} impurity ions is calculated on the assumption that it is essentially controlled by the interaction between the dipoles and screw or edge dislocations. For calculation of the elastic strength of the dipoles the ionic displacements around them are calculated using the semiclassical Born—Mayer theory of ionic solids. There is no fitting parameter applied throughout the calculations.

Introduction

It is generally accepted that divalent cation impurities significantly influence the mechanical properties of NaCl type crystals [1, 2]. A theoretical calculation of the low temperature yield stress increment in NaCl crystals due to rigid Ca^{2+} ion—cation vacancy dipoles was given in [3] recently. Following the method described there further calculations have been made to obtain the yield stress increment for the dopants Mg, Sr and Ba in NaCl crystals. Since the method of calculation has been detailed in [3], only a brief account of the main points is given here. The lattice distortion around the dipole, i.e. the displacements of the chlorine ions nearest to the dipole, are determined on the basis of the semiclassical Born—Mayer theory of ionic solids. From the displacements obtained in this way the mean strain tensors ascribed to dipoles of different impurity ions are calculated. The interaction energy and thus the force acting between the dipole and a dislocation are determined with the use of the continuum theory. Once a general formula is found for the force acting on the moving dislocation from the dipoles of different orientations with respect to and at different distances from the slip plane, the yield stress increment is straightforward to determine by Friedel's formula [4]. The calculation is carried out for screw and edge dislocations as well.

Method of computation

In the lattice distortion calculation the crystal is divided into two regions. Region I includes the divalent impurity ion, the cation vacancy and their nearest neighbour chlorine ions (Fig. 1). Here all the variables (ionic displacements and ionic

* Dedicated to Prof. I. Tarján on his 70th birthday.

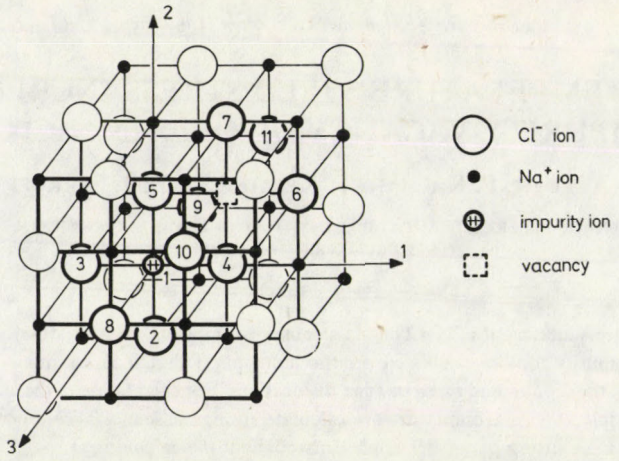


Fig. 1. The frame of reference and region I used in lattice distortion calculation. The divalent impurity and all the ions numbered are in region I

polarizabilities) are calculated simultaneously. Region II contains the rest part of the crystal, where only the ionic polarizations are taken into account [5] using rigid boundary condition.

The equilibrium displacements and dipole moments (66 variables) of the ions in region I are determined by the solution of coupled nonlinear force balance and polarization equations. The number of independent variables and equations can be reduced to 22 through the symmetry relations of the n.n. dipole. The symmetry relations, the system of the remaining nonlinear equations and the two-body potentials together with the parameters used are given in paragraph 2 of [3].

Taking the dipole to be point-like, i.e. neglecting the variation of the stress field of the dislocation over the defect volume, the dipole—dislocation interaction energy

$$E_1^n = V_d \bar{\varepsilon}_{ij}^n \sigma_{ij}^D(\mathbf{x}_d), \tag{1}$$

where V_d is the defect volume (in our case $8a_0^3$; a_0 is the cation—anion distance), $\sigma_{ij}^D(\mathbf{x}_d)$ is the stress of the dislocation at the centre $\mathbf{x}_d$ of the dipole and the mean strain tensor characteristic of the dipole $\bar{\varepsilon}_{ij}^n$ is the following

$$\bar{\varepsilon}_{ij}^n = \frac{1}{V_d} \int_{V_d} \varepsilon_{ij}^n \, \mathrm{d}V . \tag{2}$$

The suffix n refers to the orientation of the dipole. When calculating the mean strain tensor from formula (2) the volume integral can be transformed into a surface integral of the displacement field over the defect surface, i.e. the surface covering the defect volume presented in Fig. 2. Furthermore, since the displacement field is given by the

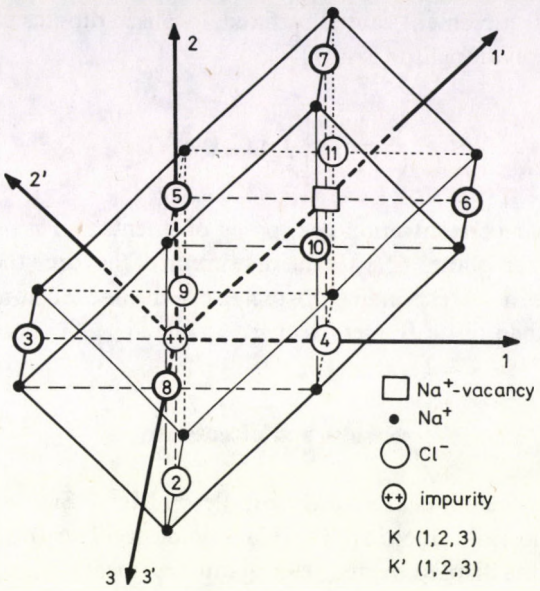


Fig. 2. The frames of reference, the defect volume and surface used in the calculation of the mean strain tensor of the impurity—vacancy dipole

discrete values of the ionic displacements the integral is approximated by a simple sum of the products of the displacements and the corresponding surface element vectors. As the mean strain tensor of the dipole obtained in this way is given in its principal axes system K' (Fig. 2), it has to be transformed into the system of the dislocation, where the stress field σ_{ij}^{D} is known [6]. The transformed mean strain tensor is denoted by $\bar{\varepsilon}_{ij}^{n}$. The suffix n referring to the orientation of the dipole can take on integer values 1—12 according to the twelve possible orientations the dipole can have at a given location of the impurity ion. The explicit formulae for the appropriate dipole—dislocation interaction energies for edge and screw dislocations can also be found in [3].

The force f_x^n acting on a moving dislocation can be deduced from the interaction energy $E_I^n(x, y)$:

$$f_x^n(x, y) = -\frac{\partial}{\partial x} E_I^n(x, y) \tag{3}$$

with the dislocation moving along axes x in its slip plane at distance y from the centre of the dipole of orientation n. The maxima $f_m^n(y)$ of the force can be computed from

$$\frac{\partial}{\partial x} f_x^n(x, y) = 0 . \tag{4}$$

The yield stress increment caused by fixed, localized dipoles can be given by the generalized Friedel relationship [3, 4, 7]

$$\Delta\tau = \frac{1}{(2T)^{1/2} b}\left[\sum_{n,i}(f_m^n(y_i))^3 c_{pi}^n\right]^{1/2}. \tag{5}$$

Here c_{pi}^n is the planar concentration of dipoles of orientation n in the i-th plane at distance y_i from the slip plane, $f_m^n(y_i)$ is the maximum of the force the dislocation must overcome as it passes a corresponding obstacle, T is the line tension of the dislocation and b is the magnitude of its Burgers vector.

Results and discussion

The ionic displacements around the dipoles of Mg^{2+}, Sr^{2+} and Ba^{2+} impurity ions are similar to those obtained in [3] for Ca^{2+} impurity. The only difference appears in the magnitude of the displacements yielding different mean strain tensors for dipoles containing different impurity ions. The principal strains are given in Table I.

In the calculation of the yield stress caused by dipoles of given impurity the effects of those dipoles have been taken into account, which give a maximum force f_m^n greater than one tenth of the greatest one. The problem of intersected dipoles is dealt with as before [3].

Rearranging Eq. (5) the yield stress increment can be expressed through the atomic concentration c of the dipoles in the general form:

$$\Delta\tau = \beta \cdot c^{1/2}. \tag{6}$$

The different values obtained for β in the case of screw and edge dislocation for different impurity ions are given in Table II. It can be seen that similarly to the case of the Ca^{2+} impurity [3] the β values and thus the calculated yield stress values are much higher for edge dislocation than for screw dislocation.

In the following we discuss only the screw dislocations, firstly because the yield stress calculated for Ca^{2+} ion—cation vacancy dipoles interacting with screw

Table I

The principal strains of the dipoles with different impurity ions

Principal strain	Impurity ion			
	Mg^{2+}	Sr^{2+}	Ba^{2+}	Ca^{2+}
$\bar{\varepsilon}_{11}$	0.034	0.031	0.028	0.032
$\bar{\varepsilon}_{22}$	0.041	0.229	0.192	0.066
$\bar{\varepsilon}_{33}$	−0.092	−0.118	−0.101	−0.088

Table II

The β values in formula (6) for edge and screw dislocations and for different impurity ions

		Impurity ion			
		Mg^{2+}	Sr^{2+}	Ba^{2+}	Ca^{2+}
β (MPa)	Screw	1900	6000	4700	2100
	Edge	5100	65500	50100	8000

dislocations [3] are in agreement with the low temperature yield stress values measured in NaCl crystals doped with Ca [8], and secondly because the mechanism giving the lower yield stress is expected to control the process.

The calculated results are valid at 0 K, and can only be compared with those of low temperature measurements. Such experiments have been made in NaCl:Ca system [8], but no data of this kind are available for the other three impurities. However, some rough consideration can be made in order to compare our results with experimental data obtained at room temperature. We cannot compare of course the yield stress values themselves, because in these crystals the yield stress at room temperature is thought to be caused mainly by Snoek effect [9, 10] and not by the Fleischer-type [1] rigid obstacles used in our calculations. But one can use the tetragonality values ($\Delta\lambda$) characterizing the dipoles as obtained from room temperature yield stress measurements [11] on the basis of Frank's formula for the Snoek effect [12]. We can also evaluate the tetragonality values from our calculated yield stress data as if they were experimental ones using the expression for Fleischer-type interaction [1]. Since the tetragonality is in proportion to our β value, we can use it for a rough comparison provided we compare only the relative quantities

$$\beta_r = \frac{\beta}{\beta_{Ca}}, \tag{7}$$

$$\lambda_r = \frac{\Delta\lambda}{\Delta\lambda_{Ca}} \tag{8}$$

for different impurities. Here β_{Ca} and $\Delta\lambda_{Ca}$ means the appropriate quantities for Ca impurity. The β_r and λ_r values are given in Table III. As can be seen, the dependence of tetragonalities of both kinds on the type of the impurity ion has the same character.

Table III

Relative tetragonalities obtained from experimental (λ_r) and theoretical (β_r) yield stress data

	Impurity ion			
	Mg^{2+}	Sr^{2+}	Ba^{2+}	Ca^{2+}
λ_r	0.8	2.1	1.5	1.0
β_r	0.9	2.9	2.2	1.0

References

1. R. L. Fleischer, J. Appl. Phys., **33**, 3504, 1962.
2. T. E. Mitchell and A. H. Heuer, Mat. Sci. Engng., **28**, 81, 1977.
3. P. Kálmán, A. Tóth, T. Keszthelyi and J. Sárközi, Mat. Sci. Engng. to be published.
4. J. Friedel, Les Dislocations, Gauther-Villars, Paris, p. 205, 1956.
5. A. S. Lidiard, Handbuch der Physik, **20**, Springer Verlag, Berlin p. 246, 1957.
6. J. P. Hirth and J. Lothe, Theory of Dislocations, McGraw-Hill, New York, 1968.
7. J. Koppenaal and D. Kuhlmann-Wilsdorf, Appl. Phys. Lett., **4**, 59, 1964.
8. J. Hesse, Reinststoffprobleme, **3**, Akademie-Verlag, Berlin p. 413, 1967.
9. J. Snoek, Physica, **8**, 711, 1941.
10. F. Appel, Phys. Stat. Sol. a, **31**, 607, 1974.
11. J. Sárközi, A. Tóth, K. Orbán, T. Keszthelyi and P. Kálmán, Acta Phys. Hung., **53**, 81, 1982.
12. W. Frank, Phys. Stat. Sol., **29**, 391, 1968.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 81—87 (1982)

THE EFFECT OF SURFACE PROCESSES ON THE FLOW STRESS OF DOPED NaCl CRYSTALS HEAT TREATED IN AIR*

J. SÁRKÖZI, A. TÓTH, K. ORBÁN, T. KESZTHELYI and P. KÁLMÁN

DEPARTMENT OF EXPERIMENTAL PHYSICS, INSTITUTE OF PHYSICS, TECHNICAL UNIVERSITY
1521 BUDAPEST, HUNGARY

The effect of surface processes due to heat treatment on the flow stress of NaCl crystals doped with different divalent impurities is investigated. The results show that in the surface layer of the crystals annealed in air some chemical reactions take place leading to substantial changes in surface conditions and thus in the flow stress of the crystals. The effect may dwarf the flow stress increment caused by dipole—dislocation interaction in the bulk, and may be the reason for some contradictory results in the literature.

Introduction

The flow stress of alkali halide crystals is mainly determined by divalent cation impurities [1]. When having individual impurity vacancy dipoles in crystals, the flow stress depends on the type of impurity ion forming the dipoles [2]. But in crystals even with low impurity concentration the dipoles tend to aggregate, and to investigate the role of individual dipoles one has to quench them in by annealing at higher temperature and subsequent cooling to room temperature. In spite of the fact that the importance of some surface treatments in plastic behaviour is well known (e.g. Joffe effect, Rebinder effect [3, 4]), there are several investigations [5, 6, 7], where the annealing of samples is carried out in air, and no attention is paid to the surface effects arising in this way. The aim of the present paper is to investigate the role the surface processes play in the flow stress in doped NaCl crystals heat treated in air.

In order to avoid any misleading effects we used specially grown, OH^--free NaCl: M^{2+} crystals at our disposal, which practically did not contain other impurities [8]. It is of primary importance because unwanted impurities, especially OH^-, can chemically react with the dopant during heat treatment and thus make its effect on flow stress inexplicable [9].

Experimental results

In our investigations the flow stress and microhardness of heat treated crystals were measured at room temperature. The microhardness measurements were to trace the hardening effect of the heat treatment in vacuum and air on crystal surfaces. In

* Dedicated to Prof. I. Tarján on his 70th birthday.

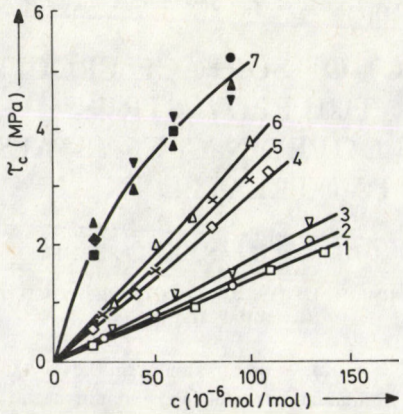


Fig. 1. The dependence of resolved flow stress increment τ_c on the impurity concentration c in NaCl crystals. The dopants: 1 — Mg, 2 — Mn, 3 — Ca, 4 — Ba, 5 — Pb, 6 — Sr, 7 — different points correspond to different kinds of impurities mentioned above. Samples 1—6 were annealed and quenched in vacuum, samples 7 were annealed and quenched in air.

addition to this, the change in crystal surface state was also followed by means of electron microscope. Gold decorating method was applied, which is based on the fact that decorating particles deposited on the surface of a sample map the places enriched in impurity. If a thin carbon layer is evaporated onto the gold decorated surface it will image the distribution of gold particles. On removing the carbon layer from the surface it can be put in electron microscope to make a picture of the impurity distribution. In the studied NaCl:Mg, NaCl:Mn, NaCl:Ca, NaCl:Sr, NaCl:Pb, NaCl:Ba crystals the impurity concentration ranged from 10 to 150 ppm whereas in the "pure" reference crystal it was less than 0,2 ppm. The heat treatment was carried out in an oven of low heat capacity. The size of the samples and quenching rate were chosen so that thermal stress and plastic deformation could be avoided. Each sample was annealed at 600 °C for 2 h and quenched at a rate of 10 °C/min. The flow stress of samples quenched in vacuum and air vs impurity concentration is presented in Fig. 1. The curves show that the dependence of flow stress on the type and concentration of contamination is fundamentally determined by the kind of heat treatment applied. The impurity dependent flow stress increment ($\tau_c = \tau - \tau_0$ where τ and τ_0 denote the flow stress of doped and pure crystals, respectively) of samples annealed in vacuum varies linearly with impurity concentration depending on its type (curves 1—6). It is of particular interest that impurity ions replacing Na^+ ions, with ionic radius smaller than or equal to that of the Na^+ ion (curves 1, 2, 3 for Mg^{2+}, Mn^{2+}, Ca^{2+}) have a weaker flow stress increasing effect than those having larger ionic radius (curves 4, 5, 6 for Sr^{2+}, Pb^{2+}, Ba^{2+}). As a result of annealing in air, however, the flow stress no longer depends on the type of impurity (curve 7) and at the same time τ_c becomes proportional to $\sqrt{c}$ instead of c.

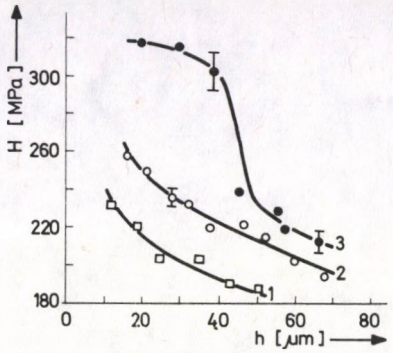


Fig. 2. The dependence of microhardness H on the depth of indentation h in NaCl crystals: 1 — pure NaCl, 2 — NaCl doped with Pb and annealed in vacuum, 3 — NaCl doped with Pb and annealed in air

The great difference between the effects of annealing in air and vacuum calls attention to the question of the surface conditions of crystals. Plastic behaviour of the surface was studied by measuring microhardness in dependence on indentation depth. This dependence for NaCl : Pb as an example is demonstrated in Fig. 2. Curve 1 belongs to pure crystals and serves as a reference curve, which is in agreement with expectation [10]. Curve 2 that shows the microhardness of doped samples after annealing in vacuum against indentation depth is the same as curve 1 if displaced in parallel. Consequently, the hardness increment caused by impurity is independent of the indentation depth. The indentation depth dependence of microhardness for NaCl : Pb annealed in air (curve 3) is essentially different from what is described by curves 1 and 2. According to curve 3 hardness drops at $h \approx 40$ μm and beyond ≈ 50 μm it runs parallel to curve 2. All this means that there must be an approximately 40 μm thick, rather hard layer on the surface of the samples quenched in air. This layer presumably develops by chemical interaction between the atmosphere and impurity diffused to the surface during heat treatment. In order to prove this assumption we have taken decorated pictures about the surface of pure NaCl and NaCl : M^{2+} crystals annealed in vacuum and air. Here too, similarly to the presentation of hardness vs indentation depth, we show pictures of NaCl : Pb^{2+} crystal as an example, because all the other impurities look the same. Information on the surface conditions of extremely pure crystals quenched from high temperature in vacuum and air is given in Figs 3a and 3b, respectively. The two ways of heat treatment do not have essentially different effects, the only difference appears in the arrangement of thermally etched steps. Figs 3c and 3d show the surface of Pb doped crystals annealed in vacuum and air, respectively. Contamination in doped samples diffuses to the surface during tempering in vacuum [11] and the thermal etching maps the enriched areas. As it can be seen in Fig. 3c the surface roughens through formation of small square pits. The effect of annealing in air is shown in picture 3d which was taken after a mere half hour of

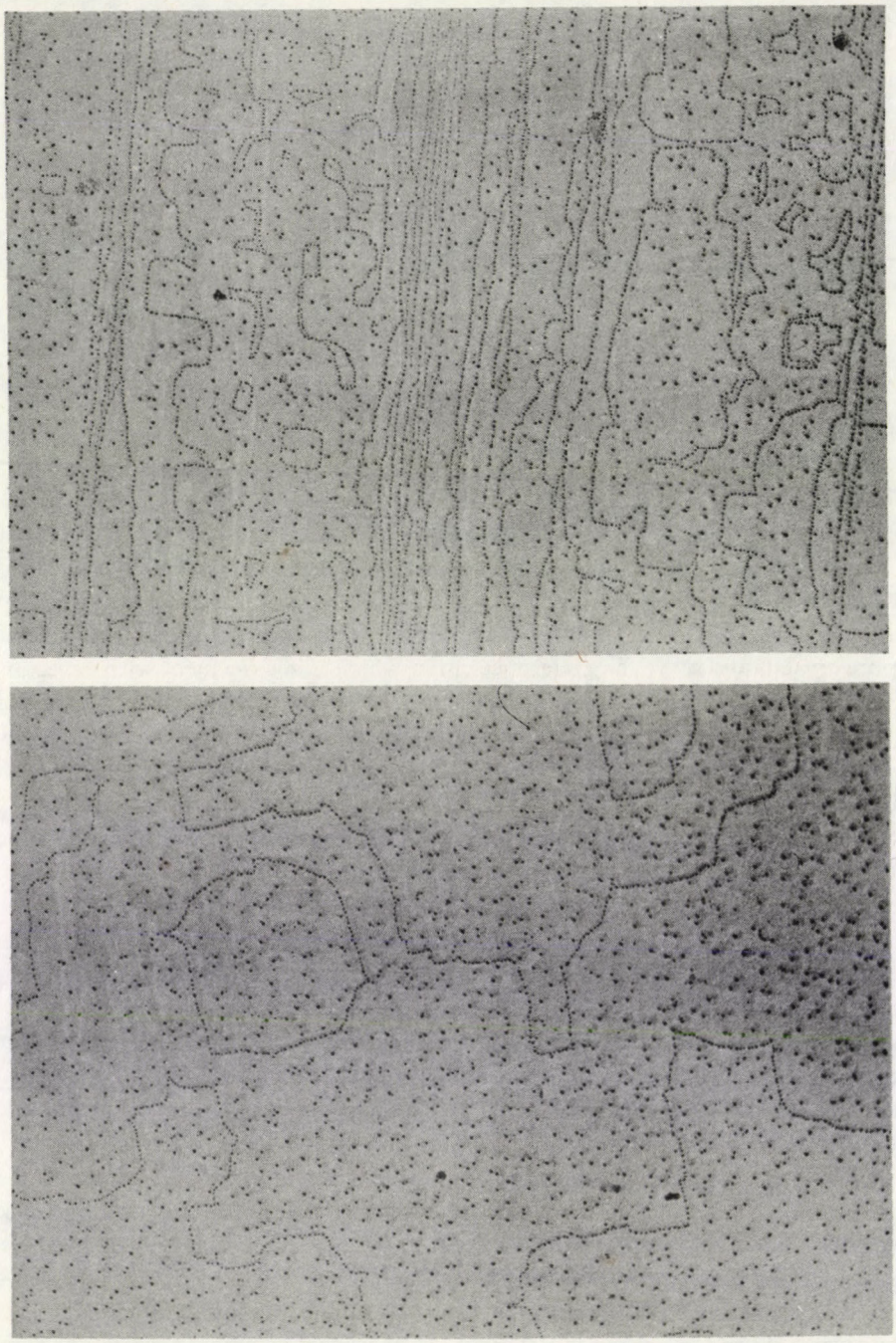
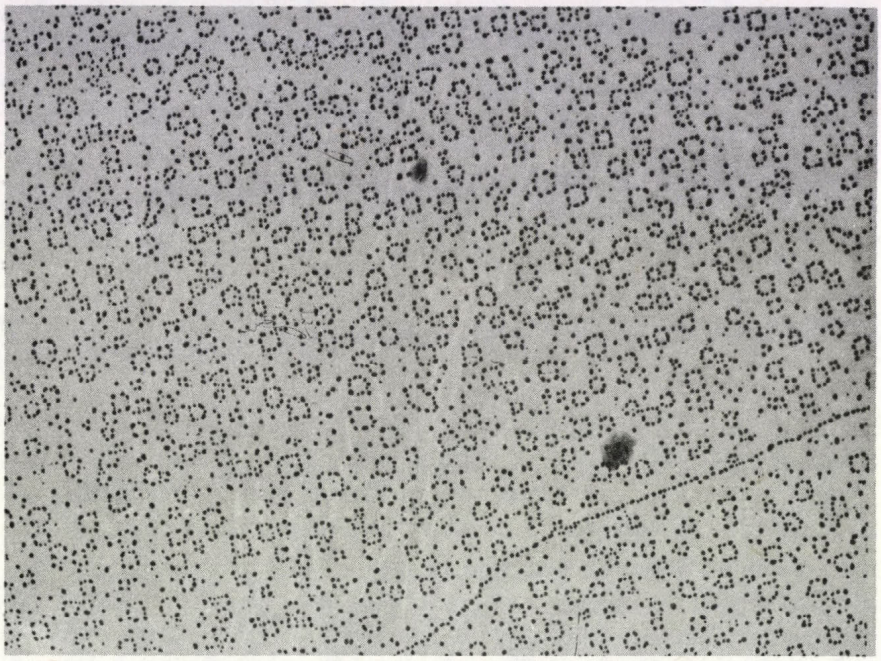
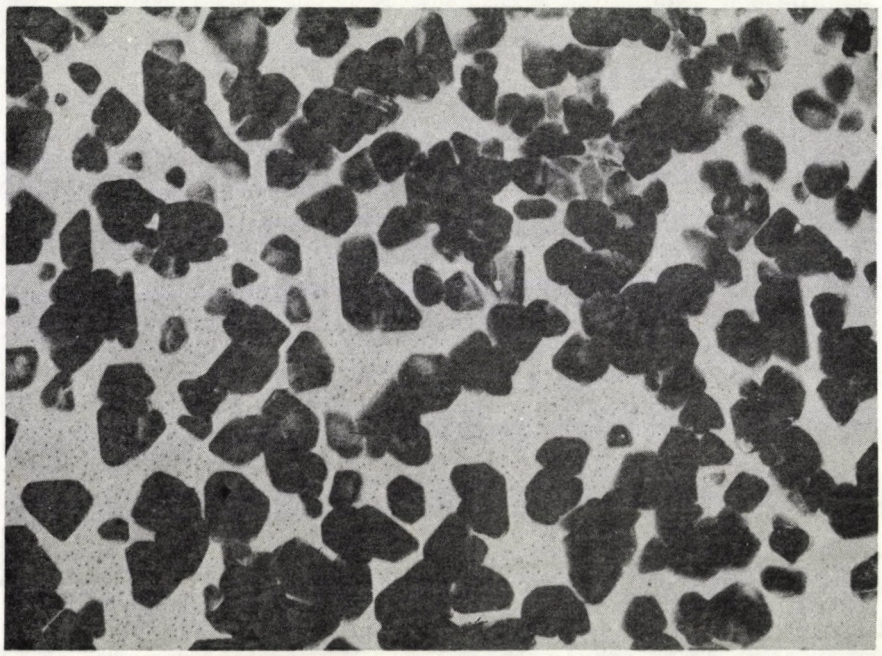


Fig. 3. Electron micrographs showing the surface structure of heat treated NaCl crystals. *a* — pure crystal annealed in vacuum, *b* — pure crystal annealed in air, *c* — Pb doped crystal annealed in vacuum (next page), *d* — Pb doped crystal annealed in air (next page). (Magnification: 1 000 X)



c)



d)

annealing instead of several hours as usual. This picture shows small separate crystalline grains which in the course of a longer heat treatment multiply to such an extent that they grow into a continuous layer. The survey of its chemical composition (probably some sort of lead oxide) would require further investigation — here we confine ourselves to proving its presence.

Discussion

As mentioned in the introduction heat treatment for the purpose of experimental investigation of flow stress increment due to impurity in alkali halides is often carried out in air. According to our experiments this way of heat treatment is rather unsuited for clearing up the role the impurity-vacancy dipoles play in increasing the flow stress. The layer developing on the surface of samples quenched in air adds to the flow stress so much that it dwarfs the increment caused by dipole—dislocation interactions. In association with it the direct proportionality of flow stress increment to impurity concentration in crystals quenched in vacuum is replaced by its variation with $\sqrt{c}$. This square root-like dependence is usually explained by Fleischer's theory for rigid, tetragonal point defects (dipoles) [12]. In our opinion, for the description of the interaction between dislocations and tetragonal point defects the linear dependence ($\tau_c \sim c$) obtained after quenching in vacuum is acceptable. Such a dependence can be explained by assuming that the dipoles are not rigid but can rotate and it is through the Snoek effect that they hinder the moving dislocations [13]. It is worth special mentioning that only good quality crystals annealed in vacuum can show the explicit difference between flow stress increments caused by different impurities. This dependence is blotted out by annealing in air.

Acknowledgement

The authors are indebted to Dr. R. Voszka, Head of the Laboratory of Crystal Physics of the Hungarian Academy of Sciences, for the crystals used in the work. Thanks are also due to Mrs. Burján and Mrs. Kiss for their assistance in taking the electron micrographs.

References

1. T. E. Mitchell and A. H. Heuer, Mat. Sci. Engng., **28,** 81, 1977.
2. J. Sárközi, A. Tóth, T. Keszthelyi and P. Kálmán, to be published.
3. P. Haasen and G. Leibfried, Fortschr. Phys., **2,** 73, 1954.
4. Proc. Conf. Environment Sensitive Mechanical Behavior Mater., Gordon and Breach Sc. Publ., New York, 1966.
5. B. H. Kear and P. L. Pratt, Phil. Mag., **4,** 56, 1959.

6. J. S. Dryden, S. Morimoto and J. S. Cook, Phil. Mag., **12,** 379, 1965.
7. R. P. Harrison and C. W. Newey, Phil. Mag., **17,** 525, 1968.
8. R. Voszka, I. Tarján, L. Berkes and J. Krajsovszky, Kristall und Technik, **1,** 423, 1966.
9. P. Grau, M. Czapelski, G. Berg, G. Klahre and M. Suszynska, Phys. Stat. Sol. (a), **61,** 189, 1980.
10. J. Sárközi, A. Tóth and Z. Morlin, Acta Phys. Hung., **42,** 283, 1977.
11. E. Pedrero, T. Ocaña, A. Gómez and M. José Yacamán, Thin Solid Films, **27,** 149, 1975.
12. R. L. Fleischer, J. Appl. Phys., **33,** 3504, 1962.
13. W. Frank, Phys. Stat. Sol., **29,** 391, 1968.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 89—104 (1982)

THE ANHARMONIC LINEAR CHAIN*

T. SIKLÓS

CENTRAL RESEARCH INSTITUTE FOR PHYSICS
1525 BUDAPEST, HUNGARY**

A simple model of the crystals, the anharmonic linear chain with nearest neighbour interaction will be investigated briefly in the first order of the self-consistent phonon field theory (SCPT). In this case we can obtain a simple explicit solution which helps to clarify some aspects of the SCPT and the phenomenon of the lattice instability.

1. Introduction

The theory of lattice dynamics founded on the classical works by Debye, Born and Kármán in its most simple approximation (in the harmonic approximation) is considered to be a well established theory capable to describe many of the physical properties of the crystals in terms of independent normal modes—phonons. For a more precise description the anharmonicity of lattice vibrations or the interaction between the phonons should be taken into account and usually the ordinary perturbation theory, considering the cubic and quartic interaction is quite appropriate for this purpose (see e.g. [1], [2], [3], [4]).

The investigations of the past years showed, however, that this approach cannot be applied in certain cases: near the phase transition points, e.g. melting point; for the quantum crystals with large zero-point energy; for the light impurities with small binding energy etc., when the anharmonic effects are not small (see, e.g. [5]).

Thus some modification of the well established Born—Kármán theory of lattice dynamics is needed in considering the highly anharmonic crystals: the quantum crystals and the crystals at high temperatures, $T > (0{,}3 - 0{,}5)\, T_m$, where T_m is the melting temperature. In order to obtain a theory which is convenient for these highly anharmonic crystals it is necessary to take into account all orders of the anharmonic interactions applying a self-consistent method. This very natural idea of introducing the self-consistent collective modes was already proposed by Born [6].

In recent years Born's idea was rediscovered and the self-consistent phonon theory (SCPT) of anharmonic crystals was elaborated simultaneously and independently by several authors by a variety of techniques. In one of these approaches the SCPT is based on the variational principle, the most elegant treatment of which is given by Werthamer [7]. A selective resummation of diagrammatic perturbation theory was

* Dedicated to Prof. I. Tarján on his 70th birthday.

** Address: H-1525 Budapest, POB 49, Hungary.

used in another group of papers, the most detailed description of this approach being presented in Choquard's book [8]. The SCPT based on the thermodynamic double-time Green's function method was proposed independently in [9], [10] and was summarized in an up-to-date formulation in [11]. It was shown in [7], [12] that all of these three variants of the SCPT are equivalent.

Today the SCPT is considered to be a well established theory and it has been applied for the investigation of dynamical, thermodynamical and elastic properties of various crystals (see. e.g. [5]).

In the present paper the SCPT is applied for the investigation of the properties of an anharmonic linear chain in the first order of SCPT. In the next Section a self-consistent system of equations is obtained which determine the properties of the anharmonic linear chain. In Section 3 the phenomenon of the instability is discussed. In Section 4 the properties of the anharmonic linear chain in the case of small tension is investigated. Some conclusions are presented in the last Section.

2. The self-consistent system of equations

Let us consider an anharmonic linear chain of length L which consists of $N+1$ identical atoms with mass M. Taking into account only nearest neighbour interaction, the Hamiltonian in the adiabatic approximation [1] reads:

$$\mathcal{H} = H + H_1 = \sum_{n=0}^{N} \frac{p_n^2}{2M} + \frac{1}{2} \sum_{n=1}^{N} \varphi(R_n - R_{n-1}) + H_1 , \tag{2.1}$$

where p_n and R_n are the momentum and position operators for the n-th atom. The interaction potential between the neighbouring atoms is denoted by $\varphi(R_n - R_{n-1})$. In the case of a one-dimensional chain the effect of the external forces can be described by the external tension P which acts on the ends of the chain:

$$H_1 = P(R_N - R_0) = P \sum_{n=1}^{N} (R_n - R_{n-1}). \tag{2.2}$$

It is convenient to introduce the equilibrium separation l between the neighbouring atoms and the relative displacement operatos by the following definition:

$$R_n - R_{n-1} = \langle R_n - R_{n-1} \rangle + u_n - u_{n-1} \equiv l + u_n - u_{n-1} , \tag{2.3}$$

where the statistical average $\langle \ldots \rangle$ is calculated for the equilibrium state of the system described by the Hamiltonian (2.1):

$$\langle \ldots \rangle = \mathrm{Sp}\{e^{-\mathcal{H}/\theta} \ldots\} / \mathrm{Sp}\{e^{-\mathcal{H}/\theta}\} \qquad (\theta = kT). \tag{2.4}$$

The equilibrium separation l in the one-dimensional case can be obtained from the equation

$$P = -\frac{1}{2} < \frac{\partial}{\partial R_n} \varphi(R_n - R_{n-1}) > = -\frac{1}{2} \langle \varphi'(l) \rangle , \tag{2.5}$$

which shows that the average force acting on an arbitrary atom in the equilibrium position is equal to zero.

It is convenient to introduce explicitly the displacement operators in the Hamiltonian (2.1) by the Fourier transformation:

$$\varphi(R) = \sum_q \varphi(q) e^{iqR}; \qquad \varphi(q) = \frac{1}{L} \int_{-L/2}^{+L/2} \mathrm{d}R \, \varphi(R) e^{-iqR} . \tag{2.6}$$

In this representation the Hamiltonian of the linear chain (2.1) takes the form:

$$\mathscr{H} = \sum_n \frac{p_n^2}{2M} + \frac{1}{2} \sum_n \sum_q \varphi(q) e^{iql} e^{iq(u_n - u_{n-1})} + H_1 . \tag{2.7}$$

For the calculation of the correlation function of nearest neighbours and the frequency of the lattice vibration we apply the method of thermodynamic double-time GF [13]. We use the following one-phonon GF:

$$G_{nn'}(t-t') = \langle\langle u_n(t); \quad u_{n'}(t') \rangle\rangle = -i\theta(t-t') \langle [u_n(t); u_{n'}(t')] \rangle \tag{2.8}$$

in usual notations [14].

To obtain the equation of motion for the GF (2.8) we differentiate it twice with respect to time t and employ the equation of motion for the Heisenberg operators $u_n(t)$ and $p_n(t)$. In this manner we get:

$$Mi^2 \frac{d^2}{dt^2} G_{nn'}(t-t') = \delta(t-t') \delta_{nn'} +$$

$$+ \frac{1}{2} \sum_q \varphi(q) e^{iql} iq \langle\langle \{ e^{iq(u_n - u_{n-1})} - e^{iq(u_{n+1} - u_n)} \}; u_{n'}(t') \rangle\rangle . \tag{2.9}$$

The multiphonon GF on the r.h.s. of Eq. (2.9) describes an uncorrelated propagation of phonons in an averaged phonon field. We use the first order or renormalized harmonic approximation of the SCPT here, in which the processes connected with the damping of phonons are not considered, but the renormalization of

the energy of phonons in the phonon self-consistent field is taken into account. In this approximation the multiphonon GF can be written in the form:

$$\langle\langle e^{iq(u_n-u_{n-1})};u_{n'}\rangle\rangle=\sum_{s=1}^{\infty}\frac{1}{s!}\langle\langle\{iq(u_n-u_{n-1})\}^s;u_{n'}\rangle\rangle\approx$$

$$\approx\sum_{s=1}^{\infty}\frac{1}{s!}\langle\{iq(u_n-u_{n-1})\}^{s-1}\rangle s\langle\langle iq(u_n-u_{n-1});u_{n'}\rangle\rangle=$$

$$=\langle e^{iq(u_n-u_{n-1})}\rangle iq\langle\langle(u_n-u_{n-1});u_{n'}\rangle\rangle. \tag{2.10}$$

For the calculation of the correlation function on the r.h.s. of Eq. (2.10) we use the same approximation. We introduce the following function:

$$F(\lambda)=\langle e^{\lambda q(u_n-u_{n-1})}\rangle;\qquad F(0)=1. \tag{2.11}$$

Differentiating it on λ and using similar approximations as in Eq. (2.10) we get:

$$\frac{\partial F}{\partial\lambda}=\langle q(u_n-u_{n-1})e^{\lambda q(u_n-u_{n-1})}\rangle=\langle q(u_n-u_{n-1})\sum_{s=1}^{\infty}\frac{(q\lambda)^s}{s!}(u_n-u_{n-1})^s\rangle\approx$$

$$\approx\lambda q^2\langle(u_n-u_{n-1})^2>F(\lambda). \tag{2.12}$$

The integration of this equation over λ from $\lambda=0$ to $\lambda=i$ gives us:

$$\langle e^{iq(u_n-u_{n-1})}\rangle=e^{-\frac{1}{2}\langle(u_n-u_{n-1})^2\rangle}=e^{-\frac{1}{2}q^2\overline{u^2}}, \tag{2.13}$$

where we take into account that the correlation function of nearest neighbour atoms does not depend on n:

$$\overline{u^2}=\langle(u_{n+1}-u_n)^2\rangle=\langle(u_n-u_{n-1})^2\rangle. \tag{2.14}$$

Now we introduce the Fourier transform for the GF (2.8):

$$G_{nn'}(t-t')=\frac{1}{2\pi}\int_{-\infty}^{\infty}d\omega\, e^{-i\omega(t-t')}G_{nn'}(\omega) \tag{2.15}$$

and take into account that it depends only on the difference of lattice sites $(n-n')$:

$$G_{nn'}(\omega)=\frac{1}{MN}\sum_k e^{ikl(n-n')}G_k(\omega). \tag{2.16}$$

Then Eq. (2.9) takes the form:

$$\omega^2G_k(\omega)=1+\frac{1}{2M}\sum_q\varphi(q)e^{iql}(iq)^2e^{-\frac{1}{2}q^2\overline{u^2}}2(1-\cos kl)G_k(\omega), \tag{2.17}$$

where Eqs. (2.10), (2.13) have been used. The solution of Eq. (2.17) reads

$$G_k(\omega)=\frac{1}{\omega^2-\omega_k^2} \tag{2.18}$$

as in the harmonic approximation except for the renormalization of the force constant:

$$\omega_k^2=\frac{4f\theta,l)}{M}\sin^2\frac{kl}{2}=\frac{f(\theta,l)}{f}\omega_{0k}^2\equiv\alpha^2\omega_{0k}^2, \tag{2.19}$$

where ω_{0k} is the harmonic frequency of vibration and f stands for the harmonic force constant. The renormalized force constant $f(\theta, l)$ according to Eq. (2.17) can be written as:

$$f(\theta,l)=\frac{1}{2}\sum_q\varphi(q)e^{iql}(iq)^2e^{-\frac{1}{2}q^2\overline{u^2}}=\frac{1}{2}\tilde{\varphi}''(l), \tag{2.20}$$

where we introduced the self-consistent potential

$$\langle\varphi(R_n-R_{n-1})\rangle\approx\tilde{\varphi}(l)=\sum_q\varphi(q)e^{iql}e^{-\frac{1}{2}q^2\overline{u^2}}=$$

$$=\sum_{s=0}^{\infty}\frac{1}{s!}\left(\frac{\overline{u^2}}{2}\right)^s\varphi^{(2s)}(l). \tag{2.21}$$

In obtaining Eq. (2.21) we decomposed the function $\exp\left(-\frac{1}{2}q^2\overline{u^2}\right)$ into the series of $\overline{u^2}$ and integrated it over q. The self-consistent potential can also be written in the form

$$\tilde{\varphi}(l)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{d}x\,e^{-x^2/2}\varphi(l+x\sqrt{\overline{u^2}}), \tag{2.22}$$

where $x=R/\sqrt{\overline{u^2}}$. It is easy to see that in Eq. (2.22) the renormalization of the potential due to the vibrations of the atoms is taken into account by averaging it over the small region $R\sim\sqrt{\overline{u^2}}\ll l$ with the Gaussian function $\exp(-x^2/2)$ which describes the effect of the phonon self-consistent field. Owing to this function only the shape of the potential $\varphi(R)$ at the bottom of the potential well is of importance.

The correlation function of nearest neighbours in Eqs. (2.21), (2.22) can be obtained from the spectral theorem [13] [14] for the GF:

$$\langle u_n u_n\rangle=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{d\omega}{e^{\omega/\theta}-1}\{-2\,\mathrm{Im}\,G_{nn}(\omega+i\delta)\}. \tag{2.23}$$

From Eq. (2.18) we get:

$$\overline{u^2} = \langle (u_n - u_{n-1})^2 \rangle = \frac{1}{Nf} \sum_k \frac{\omega_{0k}^2}{2\omega_k} \coth \frac{\omega_k}{2\theta} =$$

$$= \frac{\omega_{0k}}{\pi f} \frac{1}{\alpha} \int_0^{\pi/2} d\varphi \sin \varphi \coth \frac{\alpha \sin \varphi}{2\tau}, \tag{2.24}$$

where in the second line of Eq. (2.24) we have replaced the sum over k by an integral over $\varphi = kl/2$. The maximum value of the vibrational frequency of the chain in the harmonic approximation is denoted by $\omega_{0L} = (4f/M)^{\frac{1}{2}}$ and $\tau = \theta/\omega_{0L}$ stands for the reduced temperature. In the high temperature ($\tau \gg 1$) and low temperature ($\tau \ll 1$) limit the integral in Eq. (2.24) becomes

$$\frac{\pi f}{\omega_{0L}} \alpha^2 \overline{u^2} = \pi\tau \left\{ 1 + \frac{1}{24} \left(\frac{\alpha}{\tau} \right)^2 \right\} + O(\tau^{-3}) \qquad (\tau \gg 1), \tag{2.25}$$

$$\frac{\pi f}{\omega_{0L}} \alpha \overline{u^2} = 1 + \frac{\pi^2}{3} \left(\frac{\tau}{\alpha} \right)^2 + O(\tau^4) \qquad (\tau \ll 1). \tag{2.26}$$

In addition to the temperature τ the properties of the linear chain are determined also by the length of the chain $L = Nl$ or by the external tension P. According to Eqs (2.5), (2.21) these parameters satisfy the following equation:

$$P = -\frac{1}{2} \langle \varphi'(R_n - R_{n-1}) \rangle = -\frac{1}{2} \tilde{\varphi}'(l). \tag{2.27}$$

The internal energy is given in our approximation by the equation

$$\frac{1}{N} E_1 = \frac{1}{N} \langle H \rangle = \left\langle \frac{p_n^2}{2M} \right\rangle + \frac{1}{2} \langle \varphi(R_n - R_{n-1}) \rangle =$$

$$= \frac{1}{2N} \sum_k \frac{\omega_k}{2} \coth (\omega_k/2\theta) + \frac{1}{2} \tilde{\varphi}(l) = \frac{1}{2} \{ \tilde{\varphi}(l) + f(\theta, l) \overline{u^2} \}. \tag{2.28}$$

Then for the free energy of the anharmonic linear chain in our approximation one gets:

$$\frac{1}{N} F_1 = \frac{\theta}{N} \sum_k ln \{ 2 \sinh (\omega_k/2\theta) \} + \frac{1}{2} \{ \tilde{\varphi}(l) - f(\theta, l) \overline{u^2} \}. \tag{2.29}$$

In this way we have a closed system of self-consistent equations (2.19), (2.20), (2.21) or (2.22), (2.24), (2.27), (2.28), (2.29) which determine the dynamical, thermodynamical and elastic properties of the anharmonic linear chain in the renormalized harmonic approximation. This self-consistent system of equations is determined by the self-consistent potential (2.21) or (2.22), which can be obtained if the form of the interaction potential in the Hamiltonian (2.1) is known.

3. The instability

Let us take the Morse potential as a model potential

$$\varphi(R)=D\{[e^{-a(R-r_0)}-1]^2-1\}\,, \tag{3.1}$$

where r_0 is the average distance between the neighbouring atoms in the harmonic approximation: $\varphi'(r_0)=0$ and D is the depth of the potential: $\varphi(r_0)=-D$. The force constant in the harmonic approximation is given by $f=1/2\varphi''(r_0)=Da^2$.

Applying the expansion of Eq. (2.21) or taking the integral of Eq. (2.22) we get the following expression for the self-consistent potential:

$$\tilde{\varphi}(x)=D\{e^{-2ar_0x}e^{2y}-2e^{-ar_0x}e^{y/2}\}\,, \tag{3.2}$$

where $y=a^2\overline{u^2}=(ar_0)^2(\overline{u^2}/r_0^2)$ and $x=(l/r_0)-1$. In Fig. 1 the self-consistent potential $\tilde{\varphi}(x)/D$ is presented for some values of y. Since y depends on the temperature and, as we will see, in the quantum limit on the energy of zero-point vibrations, the larger values of y mean higher temperature or larger energies of the zero-point vibrations. The diagrams of the self-consistent potential in Fig. 1 show the thermal expansion of the crystals and the decrease of the binding energy of lattice atoms when the temperature or the energy of the zero-point vibrations increases. So for larger y values the self-

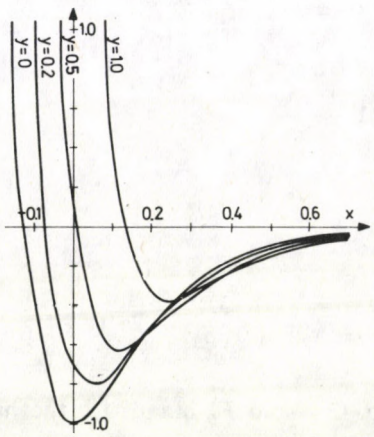


Fig. 1. The self-consistent potential $\tilde{\varphi}(x)/D$

consistent potential sufficiently differs from the interaction potential that means that the behaviour of the crystals differs sufficiently from that calculated in the harmonic approximation.

Eq. (2.5) for the Morse potential reads:

$$P^* = 4\{e^{-2ar_0x}e^{2y} - e^{-ar_0x}e^{y/2}\}, \tag{3.3}$$

where we introduced the reduced tension $P^* = (4/ar_0)P(r_0/D)$. We note that the introduced reduced tension differs from that used in works [15], [16], [17]. The renormalization factor of the frequency, according to Eqs. (2.19), (2.20), for the Morse potential takes the form:

$$\alpha^2(y) = \frac{1}{2}\{P^* + e^{-y}[1 + \sqrt{1 + P^* e^y}]\}. \tag{3.4}$$

Taking into account (3.2) we can rewrite Eqs. (2.24), (2.25), (2.26) as equations for y:

$$\lambda\alpha(y)y = \int_0^{\pi/2} \mathrm{d}\varphi \sin\varphi \coth\left(\frac{\alpha \sin\varphi}{2\tau}\right), \tag{3.5}$$

$$\alpha^2(y)y = T^*\left\{1 + \frac{1}{24}\left(\frac{\pi\alpha}{\lambda T^*}\right)^2\right\} \qquad \tau \gg 1, \tag{3.6}$$

$$\lambda\alpha(y)y = 1 + \frac{\pi^2}{3}\left(\frac{\tau}{\alpha}\right)^2 \qquad \tau \ll 1, \tag{3.7}$$

where $\lambda = (\pi D/\omega_{0L})$ is the dimensionless coupling constant and $T^* = \theta/D = \tau\pi/\lambda$ is the reduced temperature.

The expressions for the equilibrium separation of neighbouring atoms l, the internal energy (2.28) and the free energy (2.29) in the case of Morse potential can be written as follows

$$\frac{l}{r_0} = 1 + \frac{3}{2ar_0}y - \frac{1}{ar_0}\ln\frac{1}{2}\{1 + \sqrt{1 + P^* e^y}\}, \tag{3.8}$$

$$\frac{1}{N}E_1 = \frac{D}{4}\left\{\frac{3P^*}{2} + 2\alpha^2(y-1)\right\}, \tag{3.9}$$

$$\frac{1}{N}F_1 = \frac{1}{N}F_0 + \frac{D}{4}\left\{\frac{3P^*}{2} - 2\alpha^2(y+1)\right\}, \tag{3.10}$$

where α^2 is given by Eq. (3.4) and F_0 stands for the harmonic free energy: $F_0 = \theta\sum_k \ln(2\sinh(\omega_k/2\theta))$.

The system of self-consistent equations (3.3), (3.4), (3.5) determines the properties of the anharmonic linear chain, when λ, τ and P^* or l are given, see Eqs. (3.8)—(3.10).

Let us now consider the equation (3.5) in the high temperature limit, Eq. (3.6). Taking into account only the first term in the r.h.s. of Eq. (3.6) and using (3.4) the self-consistent equation can be written as follows

$$F(y)=1-\frac{y}{2T^*}\{P^*+e^{-y}[1+\sqrt{1+P^*e^y}]\}=0\,. \tag{3.11}$$

The dependence of the solution of Eq. (3.11) on the reduced temperature T^* and reduced tension P^* is given in Fig. 2. At sufficiently low temperature: $T^*<T_c^*=0.578$, and pressure: $P^*<P_c^*=0.222$ there are several real solutions for $y(T)$, the thermodynamically stable ones are shown by the full lines. The lowest line, $y_1(T)\leqq 1$ corresponds to a crystalline state with small vibrational motions. In the limit of small anharmonicity (or lower temperature), this solution gives the harmonic correlation function. But as $T^*\to T_s^*(P)$ (denoted by the full dots in Fig. 2) $y_1(T)$ becomes unstable: $T_s^*(P)$ is the instability temperature. As can be seen from Fig. 2 and from the PV diagram corresponding to Eq. (3.3) or (3.8) (for $ar_0=6$), shown in Fig. 3 at $T^*=T_s^*(P)$ there is a first order phase transition that drives the chain from the state with $y=y_1(T)$ to some other state with $y=y_2(T)$, shown by the upper full line in Fig. 2. The latter is also a "crystalline" state (due to the restrictions imposed by the theory) but the vibrations of the atoms are rather large, $y_2(T)>1$ and are defined by the external pressure: in the limit $p\to 0$ the solution $y_2(T)\to\infty$. At sufficiently high pressure $P^*>P_c^*$ or

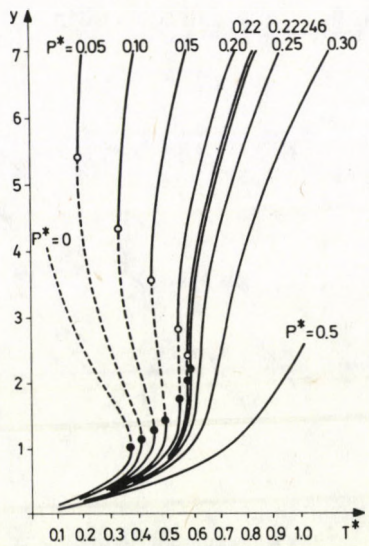


Fig. 2. The real solution of the self-consistent equation for $\tau\gg 1$

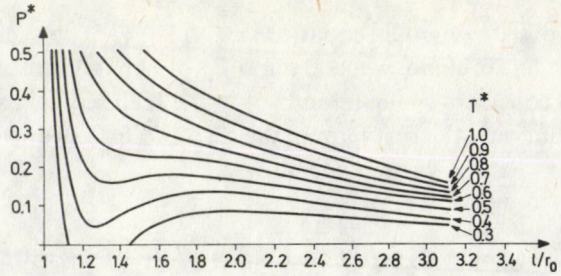


Fig. 3. The PV-diagram of the anharmonic linear chain for $\tau \gg 1$

corresponding high temperature $T^* > T_c^*$ this type of vibrational instability disappears. The two solutions at P_c^* coincide and there is only one stable solution for the correlation function $y(T)$ for $P^* > P_c^*$. Physically it means that the external forces become more efficient than the interatomic ones and the former determine the lattice dynamics: they are strong enough to localize the atomic vibrations.

At low temperatures ($\tau \ll 1$) the solution of the self-consistent equation (3.7) for the correlation function $y(\lambda, T)$ leads to the same results [17] as in the high temperature limit. The only difference is that at low temperatures the amplitude of atomic vibrations is given mostly by the zero-point energy, proportional to λ^{-1}. As a result, a highly anharmonic chain with small coupling constant $\lambda < \lambda_c = 1.207$ can be unstable even at $T = 0$ K, the critical value of the tension in this case is $P_c^* = 0.037$. In Fig. 4 the instability temperature τ_s is presented as a function of reduced tension P^* for some values of λ. The critical curve τ_c is denoted by the dotted line.

The self-consistent equation (3.5) which is valid in the whole temperature region can be solved only numerically. In Fig. 5 the dependence of the instability temperature

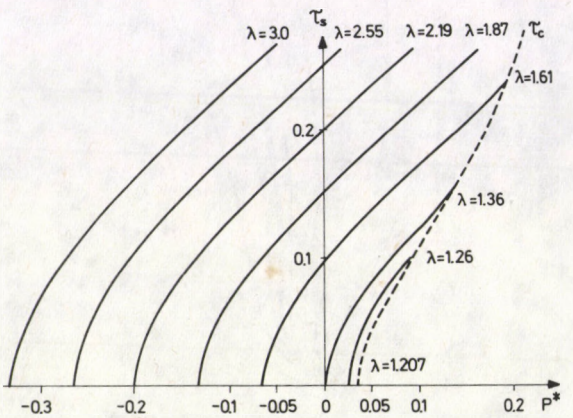


Fig. 4. The dependence of the instability temperature τ_s on the reduced tension P^* for some values of the dimensionless coupling constant λ in the low temperature limit $\tau \ll 1$

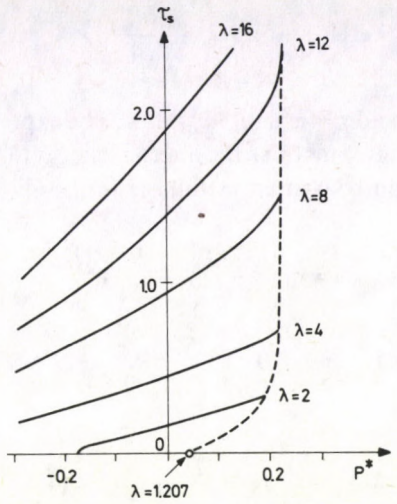


Fig. 5. The dependence of the instability temperature τ_s on the reduced tension P^* for some values of the dimensionless coupling constant λ

τ_s on the reduced tension P^* is shown for some values of λ. The critical curve τ_c is denoted by the dotted line. Using the solutions of Eq. (3.5) and Eqs (3.8)—(3.10), we can calculate the "physical" properties of the anharmonic linear chain. But it is more interesting to elucidate better the physical meaning of the instability temperature T_s. It was shown that at sufficiently high pressure $P^* > P_c^*$ the vibrational instability disappears. But the critical value of the tension is rather small even at high temperatures therefore it seems sufficient to investigate the small tension limit $P^* \ll 1$, when the results of the calculations can be presented in analytical form.

4. Properties of the anharmonic linear chain in the case of small tension

If $P^* \ll 1$ the renormalization factor of the frequency, according to Eqs. (2.19), (2.20), (3.4) takes the form

$$\alpha^2(y) = e^{-y}\left\{1 + \frac{3P^*}{4}e^y\right\}, \tag{4.1}$$

and Eqs (3.6), (3.7) using (4.1) can be rewritten as

$$\frac{D}{\theta}\left(y - \frac{D}{\theta}\eta\right)\left\{1 + \frac{3P^*}{4}e^y\right\} = e^y, \qquad (\tau \gg 1) \tag{4.2}$$

$$\lambda^2 y^2 = e^y \left\{1 + \lambda^2 y^2 \left(2\gamma - \frac{3P^*}{4}\right)\right\} \qquad (\tau \ll 1), \tag{4.3}$$

where $\eta = (\omega_{0L}^2/24D^2) \ll 1$ and $\gamma = (\pi^2\theta^2)/(3\omega_{0L}^2) \ll 1$. The expressions for the equilibrium separation of neighbouring atoms (3.8), the internal energy (3.9) and free energy (3.10) in the small tension limit can be written as follows ($ar_0 = 6$):

$$l = r_0 \left\{1 + \frac{1}{4} y - \frac{P^*}{24} e^y\right\}, \tag{4.4}$$

$$\frac{1}{N} E_1 = -\frac{D}{2} \left\{e^{-y}(1-y) - y\frac{3P^*}{4}\right\}, \tag{4.5}$$

$$\frac{1}{N} F_1 = \frac{1}{N} F_0 - \frac{D}{2} \left\{e^{-y}(y+1) + \frac{3P^*}{4} y\right\}. \tag{4.6}$$

Let us investigate the high and low temperatures separately.

4.1. High temperature ($\tau \gg 1$) limit

Equation (4.2) has a real solution only if $\theta \leqq \theta_s$, where θ_s is the instability temperature. The solution $y_s = y(\theta_s)$ is obtained as the simultaneous solution of Eq. (4.2) and its derivative. The calculation gives:

$$\theta_s \approx \frac{D}{e} \left\{1 + e\left[\frac{3P^*}{4} - \frac{1}{24}\frac{\omega_{0L}^2}{D^2}\right]\right\}, \tag{4.7}$$

$$y(\theta \lesssim \theta_s) \approx 1 + e\left(\frac{3P^*}{4} - \eta\right) - \sqrt{2(1-\theta/\theta_s)}\,. \tag{4.8}$$

The vibrational frequency ω_k at $\theta \lesssim \theta_s$ is given by

$$\omega_k^2(\theta \lesssim \theta_s) \approx \frac{\omega_{0k}^2}{2} \{1 - e\eta + \sqrt{2(1-\theta/\theta_s)} + (1-\theta/\theta_s)\}\,. \tag{4.9}$$

It becomes complex if $\theta > \theta_s$, which shows the instability of the system. It means that the state of the linear chain, which was stable at θ/θ_s, becomes unstable in the temperature region $\theta > \theta_s$.

The length of the chain, the internal energy and also the free energy, which can be obtained from Eqs. (4.4), (4.5), (4.6), respectively, remain finite at $\theta \to \theta_s$. But the coefficient of the linear thermal expansion

$$\alpha_T = \frac{k}{L}\frac{\partial L}{\partial \theta} \approx \frac{k}{l}\frac{3}{2aD}\frac{e}{\sqrt{2(1-\theta/\theta_s)}}, \tag{4.10}$$

and the specific heat at constant pressure

$$c_p = \frac{k}{N}\left[\frac{\partial}{\partial\theta}\left(E + \frac{DaP^*}{4}L\right)\right]_{P^*=\text{const}} = k\left\{1 + \frac{0.3}{\sqrt{1-\theta/\theta_s}}\right\} \tag{4.11}$$

tend to infinity if $\theta \to \theta_s$. It should be emphasized that the relative displacement of the atoms at the instability point is rather small:

$$\frac{\sqrt{\overline{u_s^2}}}{l_s} = \frac{\sqrt{y(\theta_s)}}{al(\theta_s)} = \frac{1}{ar_0 + \frac{3}{2}} \approx 0.13 \qquad (ar_0 = 6)\,. \tag{4.12}$$

In the range of temperatures where $\omega_{0L} \ll \theta \ll D$ the solution of Eq. (4.2) has the following form:

$$y = \frac{\theta}{D}\left\{1 + \frac{\theta}{D} + \frac{1}{24}\frac{\omega_{0L}^2}{\theta^2}\left(1 + \frac{\theta}{D}\right) - \frac{3P^*}{4}\right\}. \tag{4.13}$$

In this case the renormalized frequency ω_k, the equilibrium separation of the neighbouring atoms l, the linear thermal expansion α_T, the internal energy $(1/N)E$, and the specific heat at constant pressure c_p are given by the following formulae

$$\omega_k^2 = \omega_{0k}^2\left\{1 + \frac{3P^*}{4} - \frac{\theta}{D}\left[1 + \frac{1}{2}\frac{\theta}{D} + \frac{1}{24}\frac{\omega_{0L}^2}{\theta^2} - \frac{3P^*}{4}\right]\right\}, \tag{4.14}$$

$$l = r_0\left\{1 - \frac{P^*}{24} + \frac{1}{4}\frac{\theta}{D}\left[1 + \frac{1}{24}\frac{\omega_{0L}^2}{\theta^2} + \frac{\theta}{D} - \frac{11}{12}P^*\right]\right\}, \tag{4.15}$$

$$\alpha_T = \frac{k}{l}\frac{r_0}{4D}\left\{1 + 2\frac{\theta}{D} - \frac{1}{24}\frac{\omega_{0L}^2}{\theta^2} - \frac{11}{12}P^*\right\} \tag{4.16}$$

$$\frac{1}{N}E = -\frac{D}{2} + \theta\left\{1 + \frac{1}{4}\frac{\theta}{D} + \frac{1}{24}\frac{\omega_{0L}^2}{\theta^2} - \frac{3P^*}{8}\left(1 + 2\frac{\theta}{D}\right)\right\}, \tag{4.17}$$

$$c_p = k\left\{1 - \frac{1}{24}\frac{\omega_{0L}^2}{\theta^2} + \frac{1}{2}\frac{\theta}{D}\left(1 + \frac{3}{2}P^*\right)\right\}. \tag{4.18}$$

The coefficient of the linear thermal expansion, Eq. (4.16) and also the average of the quadratic displacement, Eq. (4.13) coincide with the result obtained in [18] using the regular perturbation theory, but the higher order terms in the expression of the specific

heat at constant pressure do not coincide with the result obtained in [18] because in the renormalized harmonic approximation all the odd term, thus e.g. the cubic term in the anharmonic interaction, are missing.

4.2. Low temperature ($\tau \ll 1$) limit

Equation (4.3) has a real solution for $y>0$ in the range where $\lambda>\lambda_0$ and $\theta<\theta_s$. These parameters defined previously in the case of high temperatures are now:

$$\theta_s = \frac{\omega_{0L}}{\pi e}\sqrt{\frac{6}{e}(\lambda-\lambda_0)}\,; \qquad \lambda_0 = \frac{e}{2}\left\{1-\frac{3}{2}\left(\frac{e}{2}\right)^2 P^*\right\}. \tag{4.19}$$

The solution of Eq. (4.3) at $\theta \lesssim \theta_s$ is given by

$$y(\theta \lesssim \theta_s) = 2\left\{1+\frac{3e^2}{4}P^* - \frac{4}{e}(\lambda-\lambda_0) - \right.$$

$$\left. - \sqrt{\frac{4}{e}(\lambda-\lambda_0)(1-\theta^2/\theta_s^2)}\left[1-\frac{8}{e}(\lambda-\lambda_0)\right] + \frac{4}{e}(\lambda-\lambda_0)(1-\theta^2/\theta_s^2)\right\}. \tag{4.20}$$

The renormalized frequency ω_k near the instability point can be expressed as

$$\omega_k^2 = \left(\frac{\omega_{0k}}{e}\right)^2\left\{1-\frac{3e^2}{4}P^* + \frac{8}{e}(\lambda-\lambda_0) + \frac{4}{\sqrt{e}}\sqrt{(\lambda-\lambda_0)(1-\theta^2/\theta_s^2)}\right\}. \tag{4.21}$$

The frequency becomes complex if $\lambda<\lambda_0$ or $\theta>\theta_s$, which shows the instability of the system. It is worth-while to emphasize that the chain becomes unstable even at zero temperature $T=0$ K if the zero-point energy is sufficiently high $(\omega_{0L}/2\pi)>(D/e)$. Such a situation can occur for a chain of light atoms with small binding energy ("quantum chain").

In the vicinity of the instability temperature $\theta \lesssim \theta_s$ the length of the chain, the internal energy and also the free energy, which can be obtained from Eqs (4.4), (4.5), (4.6), respectively, using solution (4.20), remain finite. But the coefficient of the linear thermal expansion

$$\alpha_T = \frac{k}{al}\frac{\theta}{\omega_{0L}^2}\frac{12\pi^2}{\sqrt{(\lambda-\lambda_0)(1-\theta^2/\theta_s^2)}} \tag{4.22}$$

and the specific heat at constant pressure

$$c_p = k\frac{2\pi}{3}\frac{3\theta}{\omega_{0L}}\left\{1+2\sqrt{\frac{\lambda-\lambda_0}{1-\theta^2/\theta_s^2}}\right\} \tag{4.23}$$

tend to infinity if $\theta \to \theta_s$. The relative displacement of the atoms at the instability point in the low temperature limit is also rather small

$$\frac{\sqrt{u_s^2}}{l_s} = \frac{\sqrt{y(\theta_s)}}{al(\theta_s)} = \frac{\sqrt{2}}{ar_0+3} \approx 0.13 \qquad (ar_0 = 6). \tag{4.24}$$

For $\lambda \gg 1$ the solution of Eq. (4.3) has the following form:

$$y = \frac{1}{\lambda}\left\{1 + \frac{1}{2\lambda} + \gamma\left(1 + \frac{2}{\lambda}\right) - \frac{3P^*}{8}\left(1 + \frac{2}{\lambda}\right)\right\}, \tag{4.25}$$

from which we see that $y \ll 1$.

Using this solution the renormalized frequency ω_k, the equilibrium separation of the neighbouring atoms l, the linear thermal expansion α_T, the internal energy $(1/N)E$, and the specific heat at constant pressure c_p can be written as

$$\omega_k^2 = \omega_{0k}^2\left\{1 + \frac{P^*}{3} - \frac{1}{\lambda}\left(1 + \gamma - \frac{3}{2}P^*\right)\right\}, \tag{4.26}$$

$$l = r_0\left\{1 - \frac{P^*}{54} + \frac{1}{4}\frac{\omega_{0L}}{\pi D}\left[1 + \frac{\omega_{0L}}{2\pi D} + \frac{\pi^2}{3}\frac{\theta^2}{\omega_{0L}^2}\left(1 + \frac{2\omega_{0L}}{\pi D}\right) - \frac{13}{24}P^*\right\}, \tag{4.27}$$

$$\alpha_T = \frac{kr_0}{6l}\frac{\theta}{\omega_{0L}}\frac{\pi}{D}\left(1 + \frac{2\omega_{0L}}{\pi D}\right), \tag{4.28}$$

$$\frac{1}{N}E = -\frac{D}{2} + \frac{\omega_{0L}}{\pi}\left\{1 - \frac{1}{4}\frac{\omega_{0L}}{\pi D} + \frac{\pi^2\theta^2}{3\omega_{0L}^2}\left(1 + 2\frac{\omega_{0L}}{\pi D}\right)\right\}, \tag{4.29}$$

$$c_p = k\frac{2\pi}{3}\frac{\theta}{\omega_{0L}}\left\{1 + 2\frac{\omega_{0L}}{\pi D}\right\}. \tag{4.30}$$

The expressions (4.26)—(4.30) characterize the behaviour of a weakly anharmonic linear chain in the low temperature limit. Using the ordinary perturbation theory these expressions can be obtained.

5. Conclusions

The SCPT can be used for investigating the properties of strongly anharmonic crystals in a wide range of temperature and external pressure. It is obvious also that in the limit of weak anharmonicity the SCPT gives the same results as the usual perturbation theory. It is interesting to point out that the method of double-time thermodynamic Green's function in the SCPT appears to be more simple and effective in the derivation of the main results in comparison with the other methods, cited in

Section 1. We think that the investigation of a simple model, the anharmonic linear chain with nearest neighbour central force interaction in the first order of SCPT, given in the Sections 2, 3 and 4 has confirmed our opinion. This simple model was investigated also in the second order of SCPT [19]. The results obtained in the second order coincidence with those of the renormalized harmonic approximation except for numerical coefficients, which are somewhat different. In second order the instability temperature becomes smaller and the dimensionless coupling constant becomes higher than in the first order of SCPT.

This variant of the SCPT was applied for the investigation of a simple model of three-dimensional anharmonic crystals, too (see [11] and references quoted therein). It is worth-while to point out also several applications of the SCPT to more complicated systems, where the interaction of lattice vibrations with other subsystems plays an essential role in the lattice instability, e.g. in the electron—phonon system of metals and semiconductors, magnetic systems with spin—phonon interaction, molecular crystals with rotational degrees of freedom, ferroelectric crystals, etc.

References

1. M. Born and K. Huang, Dynamical Theory of Crystal Lattices, Oxford Univ. Press (Clarendon), New York and London, 1954.
2. A. A. Maradudin, E. W. Montroll, G. H. Weiss and I. P. Ipatova, Solid State Physics, Supplement 3, second ed., Theory of Lattice Dynamics in the Harmonic Approximation, Academic Press, New York and London, 1971.
3. G. Leibfried, in Handbuch der Physik, edited by S. Flügge, 2-nd ed. Vol. 7., part 1., p. 104, Springer-Verlag, Berlin—Göttingen—Heidelberg, 1955.
4. G. Leibfried, W. Ludwig, in Solid State Physics, edited by F. Seitz and D. Turnbull, Vol. 12., p. 275, Academic Press, New York and London, 1961.
5. G. K. Horton, Rare Gas Solids — A Century of Excitement and Progress, in Rare Gas Solids, edited by M. L. Klein and J. A. Venables, Vol. 1. Chap. 1, p. 1, Academic Press, London—New York—San Francisco, 1976.
6. M. Born, in Festschrift zur Feier des zweihundertjährigen Bestehens der Akademie der Wissenschaften in Göttingen: I. Mathematisch-physikalische Klasse, Springer-Verlag, Berlin—Göttingen—Heidelberg, 1951.
7. N. R. Werthamer, Am. J. Phys., **37,** 763, 1969; Phys. Rev., **B1,** 572, 1970.
8. Ph. F. Choquard, The Anharmonic Crystals, W. A. Benjamin Inc., New York, 1967.
9. N. M. Plakida and T. Siklós, Acta Phys. Hung., **25,** 17, 1968.
10. N. M. Plakida and T. Siklós, Phys. Stat. Sol., **33,** 103, 1969.
11. N. M. Plakida and T. Siklós, Acta Phys. Hung., **45,** 37, 1978.
12. N. M. Plakida, Teor. i Mat. Fiz., **12,** 135, 1972.
13. N. N. Bogoliubov and S. V. Tyablikov, Dokl. Akad. Nauk SSSR, **126,** 53, 1959.
14. D. N. Zubarev, Sov. Phys. — Uspekhi, **3,** 320, 1960; Non-equilibrium Statistical Mechanics, § 16., Plenum Press, New York, 1974.
15. N. M. Plakida and T. Siklós, Phys. Lett., **26A,** 342, 1968; Acta Phys. Hung., **26,** 387, 1969.
16. T. Siklós, Acta Phys. Hung., **30,** 181, 1971.
17. T. Siklós, Acta Phys. Hung., **30,** 301, 1971.
18. J. S. Dugdale and D. K. MacDonald, Phys. Rev., **96,** 57, 1954.
19. N. M. Plakida and T. Siklós, Phys. Stat. Sol., **33,** 113, 1969.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 105—119 (1982)

THE INVESTIGATION OF THE STRUCTURE OF BIOMOLECULES BY MEANS OF SMALL-ANGLE SCATTERING*

B. K. VAINSHTEIN, L. A. FEIGIN and D. I. SVERGUN

INSTITUTE OF CRYSTALLOGRAPHY, ACADEMY OF SCIENCES OF THE USSR
MOSCOW, USSR

The general features of the diffraction techniques used for the investigations of biological macromolecules are considered. The possibilities of small-angle scattering as a method of structure determination are analysed. The new method of direct structure analysis of small-angle scattering data is described and an example of its practical application is presented.

The study of the structure of biological macromolecules, i.e. proteins, nucleic acids, nucleoproteins, polysaccharides and their associations into more complex systems such as viruses, ribosomes, chromosomes, membranes and so on represents a rather important problem in molecular biology. The physical methods to treat these problems are based on the scattering of X-rays, electrons and neutrons, respectively. After discussing the general principles of diffraction techniques, their possibilities and limitations, the paper discusses several methods of structure determination by means of small-angle scattering (SAS). Further on some data on the structure of a few biomolecules will be presented.

Our ultimate task aims at obtaining a pattern of the spatial structure $\rho(\mathbf{r})$ of the object under investigation. By means of optics as, for instance, in a transmission or electron microscope, one obtains directly an image of the object. The image formation can be presented according to Abbe's scheme as demonstrated in Fig. 1. However, depending on the nature of radiation used for analysing the structure under investigation the realization of the full optical pattern is sometimes impossible, though its first stage, i.e. the observation of scattered, diffracted radiation may be always accomplished. In diffraction experiments one first measures the intensities $I(\mathbf{s})$ or moduli of amplitudes $F(\mathbf{s})$ of the scattered radiation, subsequently the diffracted beam is collected with the aid of lenses to form an image of the object, or, to be more precise, to present an image of its projection $pr_{\mathbf{k}}\rho(\mathbf{r})$ along the vector $\mathbf{k}$ of the initial wave (whenever possible). In the diffraction structure analysis the transition from the intensities to the structure of an object $\rho(\mathbf{r})$ is not realized physically, but it may

* Dedicated to Prof. I. Tarján on his 70th birthday.

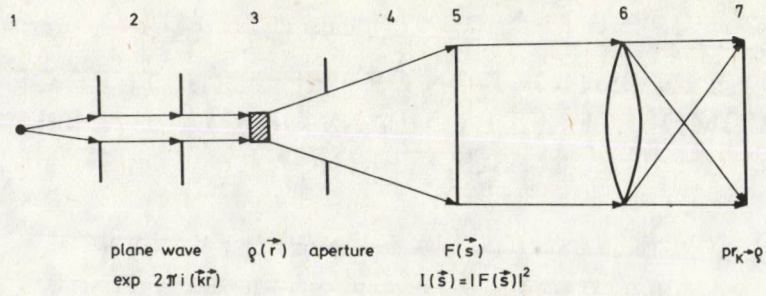


Fig. 1. Formation of the image in diffraction experiment
1 — radiation source, 2 — collimation and formation of the initial beam; **k** — the vector of the initial wave, 3 — object, 4 — diffraction, 5 — diffraction plane; vector $\mathbf{s}=\mathbf{k}-\mathbf{k}''$, $\mathbf{k}''$ is the vector of diffracted wave, 6 — optical system, 7 — image

performed, with some other degree of reliability and accuracy; to do this mathematically is the ultimate goal of this kind of research.

Let us consider this scheme in more detail going from left to right, i.e. from a radiation source to the image.

Various sources of radiation, X-rays, electrons and neutrons are used. All of them have wavelength values which are suitable for structural investigations: their wavelengths ranging from several tenths of nanometers down to several hundredths, which result in a resolution of atomic scales. However, due to a different nature of interaction with various substances, the "scattering matter" appears to be different for each of them: for X-rays the electron density, for electrons the electrostatic potential, and for neutrons the nuclear and spin density of the object. It should be noted that when we are not interested in the structure of an object at the level of atomic resolution, as in the case of SAS, this difference is not essential. However, it may become important combining diffraction data of one and the same object obtained from various radiations.

Passing to the scattering object itself, the character of its order determines, to a large extent, the possibility of deriving the structure from a diffraction experiment. With higher order the separation of the object image $F(\mathbf{s})$ in the reciprocal (Fourier) space

$$F(\mathbf{s})=\mathscr{F}[\rho(\mathbf{r})]$$

becomes more distinct in principle, the derivation of the structure of an "immobile" non-averaged object from diffraction data is more readily obtained. With decreasing the ordering, textures of different type are obtained whose ordering character may be described by some averaging operator B (Vainshtein [1]). In this case the intensities in the reciprocal space are mixed up being averaged according to the same operator and it is impossible to extract a pure module of the scattering amplitude $|F(\mathbf{s})|=\sqrt{I(\mathbf{s})}$ from

the given point in the reciprocal space. The intensities are the Fourier transforms of self-convolution of an object specifically averaged:

$$Q(\mathbf{r}) = \rho(\mathbf{r}) * \rho(-\mathbf{r}) .$$

The general scheme is given in Fig. 2. In the case of SAS by biomolecules in solution a random orientation exists (B is a spherical averaging) and most difficulties in the structure determination are due to this reason.

With the question of the properties of the object the problem of the information inherent in a scattering pattern is closely connected. As it is clear from Fig. 2 one may pass over from the intensities to the structure $\rho(\mathbf{r})$ using either the amplitudes $F(\mathbf{r})$ or the distance function $Q(\mathbf{r})$. Table I illustrates the possibilities of deriving the structure of an object from experimental data. It demonstrates the principal role of X-ray structure analysis of crystals in the investigation of matter and, particularly, of biomolecules.

At present, this method yielding a resolution of 0.3—0.15 nm has been used for investigating more than 100 globular proteins having molecular weights of 10 000 to 300 000. This made it possible to establish the main regularities in the packing of polypeptide chains as well as to reveal the mechanism of some enzymatic reactions. The structure of transport RNA and the protein coat of two small spherical viruses have been also found (Blundell and Johnson [2]).

The structure of larger proteins as well as of high polymeric DNA and RNA at high resolution is still unknown. This may be associated with the very large dimensions of these molecules and with difficulties (or sometimes the impossibility) in obtaining such objects in the form of single crystals. The multicomponent systems such as ribosomes, membranes, viruses, chromatines and so on are to be regarded as the most complex biological objects in structural investigations.

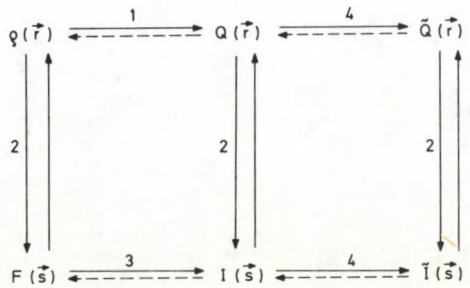


Fig. 2. Connections between real and reciprocal space patterns $\rho(\mathbf{r})$ — electron density, $F(\mathbf{s})$ — amplitude of scattering, $Q(\mathbf{r})$ — Patterson function, $I(\mathbf{s})$ — intensity of scattering, $\tilde{Q}(\mathbf{r})$ — Patterson function after averaging procedure, $\tilde{I}(\mathbf{s})$ — corresponding intensity of scattering, 1—4 — some mathematical treatments: 1 — self-convolution, 2 — Fourier-transform, 3 — multiplication by complex-conjugated value (square), 4 — averaging—transformation. Full arrows show the possibility of unambiguous transition. Dashed arrows show that the transition is not direct or ambiguous

Acta Physica Academiae Scientiarum Hungaricae 53, 1982

Table I

The possibilities of the derivation of the structure from diffraction data

Methods and objects	Construction of $Q(\mathbf{r})$	Finding $F(\mathbf{s})$	Mathematical determination of phases	Physical measurement of phases	Physical image	Model building
X-ray diffraction						
Neutron diffraction						
Electron diffraction						
for crystals	+	+	$+^{(1)}$	$+^{(2)}$	−	$+^{(4)}$
for textures	+	(+)	$(+)^{(1)}$	−	−	$+^{(4)}$
for random systems	+	(+)	$(+)^{(5)}$	−	−	$+^{(4)}$
Electron-microscopy	−	−	−	+	$+^{(3)}$	−

+ possible, − impossible, (+) means the possibility in special cases or the possibility of ambiguous transition

1 — direct methods based on the relationships between modules of amplitudes; 2 — insertion of heavy atoms, dynamical scattering; 3 — holographic methods, calculation from an image; 4 — application of scattering invariants, various radiations, changes in contrast, minimizing functionals; 5 — direct method using spherical harmonics

Very valuable information on the structure of biological macromolecules has been obtained by means of electron microscopy (Vainshtein [3]). However, these studies are restricted by their own limitations. The space resolution presents the first difficulty. Despite the fact that the instrumental resolution of the modern electron microscopes is 0.15 nm biological samples must be prepared by staining with heavy-atom compounds; this reduces the resolution to 3—2 nm. The second limitation is the transition from an image which is a projection $p(\mathbf{r}_k) \rightarrow \rho(\mathbf{r})$ to the three-dimensional structure.

Thus the methods of X-ray diffraction and electron microscopy cannot be applied in all instances, especially in dealing with complex biological objects; with this type of samples one has to resort to SAS in solutions. Here the molecules are in their natural, native state, though due to spherical averaging a large amount of diffraction information is lost. Nevertheless many important characteristic features of an object may be found, and as it will be shown below, in some cases one can even directly determine the inner structure of a biomolecule.

The X-ray SAS owes its origin to the classical works of Guinier [4]. The name itself — SAS — is associated with the observation of diffuse scattering near the primary beam.

Although SAS has been applied for more than 40 years to many inhomogeneous systems such as coals, glasses, alloys, catalyzers and so on, the revival of the method is doubtlessly associated with biological objects.

The possibility of the study of dilute solutions of biomolecules is an impetus for the development of theoretical methods to extract structural information from scattering curves and for the development of new experimental methods. The success of X-ray small-angle studies of biological objects is associated with the development of X-ray experimental and calculation methods involving computers as well as the possibility of obtaining homogeneous preparations in large amounts.

In the last ten years the experimental SAS methods have, in principle, acquired new forms. New possibilities arose as a result of synchrotron radiation, powerful neutron beams, position-sensitive detectors. Let it be noted that the detailed investigation of the structure of biological objects, i.e. the search for the shape and inner structure of the molecules require precise intensity measurements at angles where the intensity of scattering by the solution decreases by 3—5 orders of magnitude as compared with the scattering at zero angle, and is only slightly different from the scattering by a pure solvent. Consequently, one of the main experimental requirements is the measurement of a wide range of small scattering angles. The position-sensitive detectors adopted from nuclear physics are rather suitable for this task.

Let us now consider the problem of deriving the structure from an experiment. The intensity of SAS (to be denoted by $I(s)$, where $s = 4\pi \sin \psi/\lambda$, λ = wavelength and 2ψ is the scattering angle) for any predetermined system may be calculated either analytically or by numerical methods. The problem of calculating the function $I(s)$ from the function $\rho(\mathbf{r})$ presents no difficulties. In principle, the classical Debye formula

(Guinier and Fournet [5])

$$I(s) = \int\int \rho(r)\,\rho(r')\,\frac{\sin s|\mathbf{r}-\mathbf{r}'|}{s|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}v_1\,\mathrm{d}v_2$$

is the basic equation and may be applied in most of the SAS cases.

However, when analysing an unknown object the investigator is naturally interested in the opposite problem, which consists in finding the distribution $\rho(\mathbf{r})$ from a given scattering intensity $I(s)$. This problem has no unambiguous solution due to considerable loss of information as a result of the spherical averaging. Up to now there have been no direct methods of structure determination based on SAS data, and therefore the principal method is still the comparison of experimental intensities (or their density distribution functions) with the scattering by model structures.

In the calculations of the scattering intensity many configurations from simple geometrical forms and their aggregates to compact and unfolded macromolecules, or particles with nonuniform scattering density distribution are considered. In some cases strict analytical solution can be obtained, however, in many cases only approximations with the aid of a computer are possible. The solution should fit the integral characteristics of an object, i.e. its invariants which are determined directly from the intensity: these are the radius of gyration R_g, the volume V, the surface S, the maximal size $l_{\max}$ and some others. These invariant quantities serve as the basis of a model both in the framework of the simplest geometrical bodies and in the homogeneous density approximation.

An important step forward was the development of the modelling method (Kratky [6]; Feigin [7]). We can now build up a model of a homogeneous body of a complicated form, filling its volume with certain elementary bodies, e.g. with cubes, balls, etc. By varying the form of the model one obtains different curves $I(s)$, which may be compared with the tentative form. It turned out that the theoretical SAS curves are rather sensitive to any changes in the model.

Thus the model method is a peculiar version of the trial-and-error method in classical structural crystallography. The choice of an initial model is, to some extent, an intuitive process, since any, even a very simple model of a molecule, is a function of many parameters. Here the superfluous details at the initial stage may only impede any subsequent research. The choice of an initial model (satisfying the invariants) is completed by outside information extracted from electron microscopy, biochemistry and physico-chemical data.

It is difficult to state unambiguously which of the numerous models is the most preferable one. The integral discrepancy factor may serve as an objective criterion. As a rule, one uses the following R factor

$$R_I = \frac{\int_{s_1}^{s_2} [I_{\text{model}}(s) - I(s)]^2 s^4\,\mathrm{d}s}{\int_{s_1}^{s_2} I^2(s) s^4\,\mathrm{d}s}, \qquad (1)$$

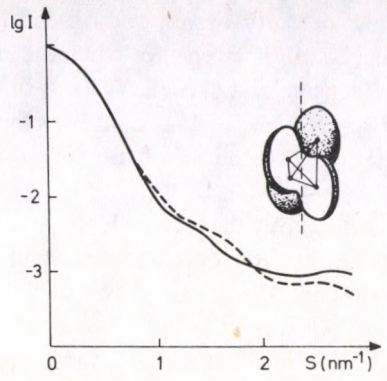


Fig. 3. The model of catalase
Full line — experimental curve, Dashed line — the curve of SAS by the model

where (s_1, s_2) is the interval whithin which the comparison is made. The derivation of the model from the SAS data may be compared with the search for a tentative model in the structure analysis of crystals.

As an example of applying the model method the SAS investigation of catalase is presented.

Catalase is an enzyme (mol. weight 250 000) which catalyzes the decomposition of hydrogen peroxide into water and oxygen. According to X-ray analysis and electron microscopy data catalase has the tetrahedral symmetry 222, which has been used as a starting structure to construct the model. Fig. 3 shows the best model found for catalase as well as its theoretical intensity curve together with the experimental curve. The model consists of four identical ellipsoids of revolution ($a = 1.7$ nm, $c = 4.7$ nm) spaced in the tetrahedral vertices with an edge length of 5.9 nm. The gap between the ellipsoids is also filled by scattering density (Vainshtein et al [8]).

More recent examples of the model method by the precise modelling of the form of a biomolecule may be represented by the studies of histidinedecarboxylase (Gonchar et al [9]) and the bacteriophage T7 (Rolbin et al [10]). The parametrisation of the model and the use of the discrepancy integrals will probably render it possible to find, in the future, an algorithm for looking for the "best" model by the method of non-local search or the least-square technique.

The objects with inhomogeneous density such as various nucleoproteins (viruses, ribosomes, etc.) are becoming important for small-angle investigations. With these samples the decisive role is played by the possibility of varying the solvent density which changes markedly the s dependence of I as well as the possibility of combining various radiations. The method known as "contrast variation" (Stuhrmann and Kirste [11]) allows after performing measurements in a series of solvents, the separation of three scattering functions

$$I(s) = I_F(s) + I_{FS}(s) + I_S(s)\,,$$

where $I_F(s)$ is the scattering intensity corresponding to the homogeneous electron density, $I_S(s)$ is the scattering by inhomogeneities when the medium electron density is equal to the mean electron density of a particle, $I_{FS}(s)$ is the interference term, which may be both positive and negative. For biological objects the contrast variation is best achieved for neutron scattering with the aid of solvents of different concentration of heavy water.

Some useful information about the specific features of biomolecules can also be obtained with the aid of the recently developed methods of heavy-atom markers (Vainshtein et al [12]), anomalous SAS (Stuhrmann [13]), and the methods using real-space information (Glatter [14]).

It should be stressed, however, that all these methods, except a very few special cases, do not provide directly the structural information about a particle. Only some of its general parameters can be specified without modelling. The relationship between the theoretical and experimental possibilities in the small angle studies of biological objects is rather complicated: partly because up to now the experiment is devoid of sufficient accuracy, so that some parameters which may be found theoretically cannot be calculated from experimental data; and partly because even the high precision experimental scattering curves do not yield sufficient information; consequently, we still fail to provide direct information about the structure. So one of the most important problems in the field of SAS to solve biomolecular objects is the creation of a direct method to find their structure, i.e. to obtain the density distribution $\rho(\mathbf{r})$.

It is clear, however, that the task is to obtain a three-dimensional function $\rho(\mathbf{r})$ from the one-dimensional function $I(s)$, which has an infinite variety of solutions. This means that any direct method should be regarded as an approach which enables to narrow down this variety as much as possible by means of imposing some physically justified restrictions. We have developed the direct method of structure analysis in SAS using the information about the symmetry of a particle, its dimensions and the range of scattering density (Svergun et al [15]). The main features of the method as well as an example of its application to the concrete structure investigation are given below. To choose a convenient class of possible solutions the apparatus of spherical harmonics was used. Such mathematical approach was introduced in SAS by Stuhrmann [16]. If one represents the density $\rho(\mathbf{r})$ as a series

$$\rho(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \rho_{lm}(r) Y_{lm}(\theta, \varphi) = \sum_{l=0}^{\infty} \rho_l(\mathbf{r}), \tag{2}$$

where r, θ, φ are spherical coordinates, $Y_{lm}(\theta, \varphi)$ are spherical harmonics, $\rho_{lm}(r)$ are radial functions, and the equation

$$\rho_l(\mathbf{r}) = \sum_{m=-l}^{l} \rho_{lm}(r) Y_{lm}(\theta, \varphi) \tag{3}$$

defines the partial multipole densities, the intensity of the SAS is given by:

$$I(s)=2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} |A_{lm}(s)|^2 . \tag{4}$$

The functions $\rho_{lm}(r)$ and $A_{lm}(s)$ are connected by the Hankel transform of order l:

$$A_{lm}(s)=i^l \sqrt{2/\pi} \int_0^{\infty} \rho_{lm}(r) j_l(sr) r^2 \, dr , \tag{5}$$

$$\rho_{lm}(r)=(-i)^l \sqrt{2/\pi} \int_0^{\infty} A_{lm}(s) j_l(sr) r^2 \, dr . \tag{6}$$

Here $j_l(sr)$ are spherical Bessel functions. Stuhrmann has shown that the function $I(s)$ is invariant to independent real space rotations of any partial density $\rho_l(\mathbf{r})$. So the manifold distributions $\rho(\mathbf{r})$ are determined by the same functions $\rho_{lm}(r)$ corresponding to the decomposition (4), and the set $\rho_{lm}(r)$ can be regarded as the variety of possible solutions introduced above.

It is clear that one can decompose $I(s)$ into a sum of squares of the amplitudes $A_{lm}(s)$, generally speaking, quite arbitrarily. Thus to use this approach it is nessesary to obtain the decomposition (4) corresponding to the real structure of a particle. Few attempts were made to solve the problem (Marguerie and Stuhrmann [17]; Stuhrmann and Fuess [18]), but the general algorithm has not been found yet.

Obviously, the summands cannot be separated from the sum unless some assumptions are made. First of all let us suppose that the density $\rho(\mathbf{r})$ is satisfactorily represented by a finite sufficiently small number of harmonics. Furthermore, as it results from formulae (5), functions $A_{lm}(s)$ are independent of the value of m upon the fixed functions $\rho_{lm}(r)$. This fact means that the contributions of harmonics with different m and the same l are in principle inseparable in the scattering intensity $I(s)$. Thus the formulae (4) can be rewritten:

$$I(s)=2\pi^2 \sum_{l=0}^{\infty} |A_l(s)|^2 , \tag{7}$$

where

$$|A_l(s)|^2 = \sum_{m=-l}^{l} |A_{lm}(s)|^2 . \tag{8}$$

In this way the problem consists of obtaining the decomposition (7), i.e. harmonics with different l values. Each term of the sum (7) determines according to transform (6) a radial function $\rho_l(r)$. In the case of an axially symmetrical particle ($\rho_{lm}(r) \equiv 0$ if $m \neq 0$) the sum over m vanishes and one has simply $\rho_l(r)=\rho_{l0}(r)$. In the general case the decomposition (7) holds, but the functions $\rho_l(r)$ have no explicit physical meaning, they are superpositions of the functions $\rho_{lm}(r)$.

Whatever the meaning of the functions $\rho_{lm}(r)$, there are strictly specified angular non-uniformities depending on the number of the harmonic l, inherent to any partial density (3), and therefore there are definite regions in the reciprocal space where the scattering intensity $|A_l(s)|^2$ contributes maximally to $I(s)$ (in accordance with the order of the Hankel transform (5)). Consequently, the changes of the relative contribution of different harmonics to the total intensity $I(s)$ are connected with their numbers.

Besides, the possible mode of $A_l(s)$ is restricted by the fact that any biomolecule is finite in space, i.e. there exists an R value so that

$$\rho(\mathbf{r}) \equiv \rho_{lm}(\mathbf{r}) \equiv 0, \text{ if } r > R. \tag{9}$$

The peculiarities of the behaviour of the functions $A_l(s)$ examined above can be regarded as the physical foundation for developing an algorithm to decompose the scattering intensity into functions (7).

Actually let us specify the $L+1$ functions $\rho_l^{(k)}(r)$, obeying the space restrictions and determining the functions $A_l^{(k)}(s)$ which in turn determine, according to (7), the intensity $I^{(k)}(s)$. This $I^{(k)}(s)$, generally does not coincide with the true intensity $I(s)$.

In order to fit the intensity $I(s)$, by keeping the relative contributions of different harmonics the same, one may renormalize the amplitudes as follows:

$$\tilde{A}_l^{(k)}(s) = A_l^{(k)}(s) \sqrt{\frac{I(s)}{I^{(k)}(s)}}. \tag{10}$$

The amplitudes $\tilde{A}_l^{(k)}(s)$ will obey Eq. (7), but the set $\tilde{\rho}_l^{(k)}(r)$ specified by $\tilde{A}_l^{(k)}(s)$ will generally not obey (9). Let us therefore set

$$\tilde{\rho}_l^{(k+1)}(r) = \tilde{\rho}_l^{(k)}(r) \Pi(r-R) = \begin{cases} \rho_l^{(k)}, & r \leqq R, \\ 0, & r > R. \end{cases} \tag{11}$$

The amplitudes of the next approximation $A_l^{(k+1)}(s)$ are determined by the function (11) according to (5).

Thus, with each step of the iterative process determined by Eqs (10) and (11) the partial amplitudes are redistributed in accordance with the number of the harmonic l, the value of R and the scattering intensity $I(s)$. Since no a priori information about the radial functions, except condition (9), is available, the step functions

$$\rho_l^{(0)}(r) = \Pi(r-R) = \begin{cases} 1, & r \leqq R \\ 0, & r > R \end{cases} \tag{12}$$

seem to be convenient for the first approximation.

It is possible to show that the procedures (10)–(11) converge to the true intensity $I(s)$. To estimate the deviations in reciprocal space between $I(s)$ and $I^{(k)}(s)$ the same R

factor has been used as for the model method. The convergence of the process over the R factor is uniform.

The efficiency of this procedure has been verified by means of model examples. The transforms (5)—(6) have been computed by the Simpson rule, the termination effects in (6) (the intensities have been calculated up to some finite value s_{max}) were reduced by the generalized Steklov filter (Rolbin et al [10]).

In the simplest case $L=0$ (spherically-symmetric particle) the problem is reduced to the determination of the true set of signs of the function $\sqrt{I(s)}$. When the change of signs stops, the process is interrupted. 5—6 iterations are usually quite enough to obtain the solution, function $\rho(r)$ is restored practically completely. This is an example of a quite unfavourable case of the multistep function $\rho(r)$ shown in Fig. 4. The restoring appears to be good, the R factor is comparable with one of the transforms by true signs, i.e. with the termination effects.

The most important practical case, however, corresponds to $L>0$. In this case the convergence of the process to the true intensity $I(s)$ does not guarantee the convergence to the true radial functions $\rho_l(r)$. The incorrectness of the problem (one should restore the summands from the sum of their squares) may lead to considerable distortions of these functions. To reduce them natural physical conditions were used,

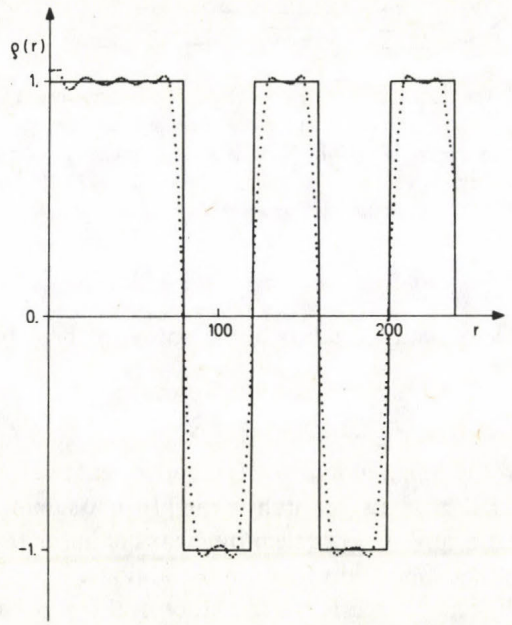


Fig. 4. The restoring of the structure of a spherically-symmetrical particle
Full line — true density distribution, Dotted line — restored distribution, Values $R=24$ nm, $s_{max}=5\ \text{nm}^{-1}$; the result of the 7th iteration is shown, $R_I=2.5\times10^{-4}$

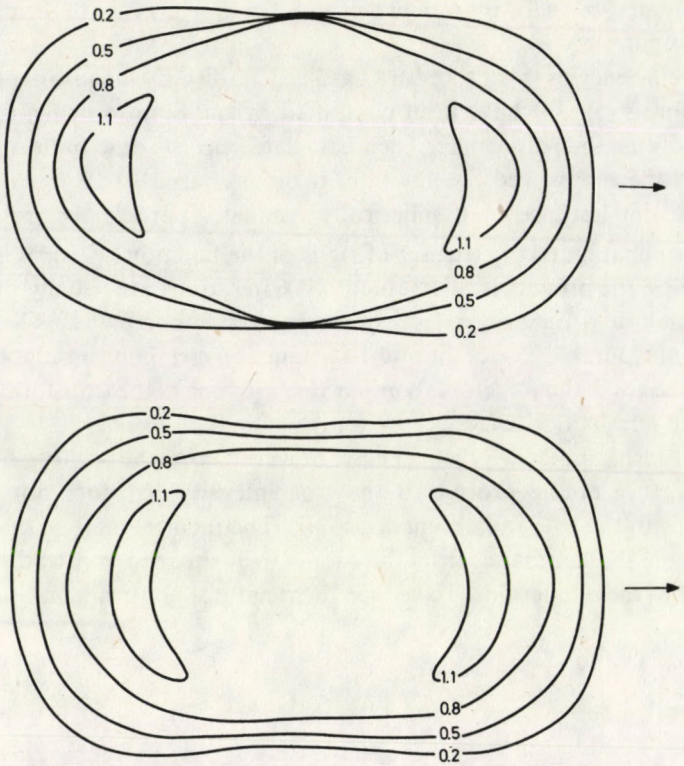


Fig. 5. The restoring of the model of an axially-symmetrical structure
a — true structure, b — restored structure
The distribution in the cross-section containing the axis of axial symmetry (indicated by an arrow) is shown; values $R = 10$ nm, $s_{max} = 5\ \text{nm}^{-1}$. The numbers refer to density levels in $10^{-3}\ \text{nm}^3$. The result of the 9th iteration is presented, $R_I = 7.8 \times 10^{-3}$

namely the existence of limiting values of the scattering density

$$\rho_{min} \leqq \rho(\mathbf{r}) \leqq \rho_{max} . \tag{13}$$

In fact, one can introduce the corrections $\rho_l(r)$ to the radial functions ensuring the correctness of (13) with a minimal R factor. It is convenient to express the corrections as a series of Laguerre polynomials (see Stuhrmann [16]). Assuming the corrections to be small enough, one can solve the problem by means of linear programming (Dantzig [20]). We have used the modified simplex method (ibid). It should be noted that the decreasing of the R factor suggests the relevance of this procedure.

An example of using procedures (10)—(11) with the restrictions (13) is presented in Fig. 5. The axially-symmetrical model structure (Fig. 5a) is constructed from three first non-zero radial functions of an ellipsoid of revolution with the density $\rho(\mathbf{r}) = 1$,

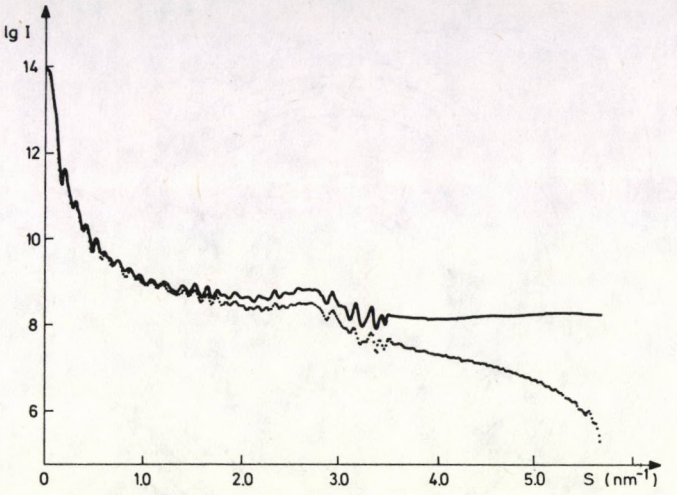


Fig. 6. The scattering curves of bacteriophage T7
Full line — experimental curve, Dotted line — the SAS curve from the restored structure

half-axes $a = 5$ nm, $c = 10$ nm. The structure shown in Fig. 5b is constructed from the intensity calculated for the previous one. The agreement seems to be quite satisfactory.

The examples shown above illustrate the efficiency of the procedure using parameters R, $\rho_{\min}$, $\rho_{\max}$ and a set of exactly defined harmonics. In practice, these parameters are as a rule known with some errors. We have verified the stability of the procedure with respect to the deviation of these input parameters in the model examples. It has been proved that the method yields suitable solutions with deviations up to 10% in R and 20% in $\rho_{\min}$ and $\rho_{\max}$. The errors in the set of harmonics may be connected mainly with the fact that any particle is generally represented by an infinite series of harmonics, whereas one has only a finite number of terms. It may be noted, however, that the spherical harmonics constitute a complete system of functions, and a few first terms in the series would describe the whole structure, as a rule, quite sufficiently.

It should be stressed that the procedure holds also for particles with arbitrary density distribution. In the general case in order to select a class of possible solutions one should perform a new separation, the separation of functions $\rho_{lm}(r)$ from $\rho_l(r)$, with the aid of some additional (non SAS) informations. Selecting one (or several) relevant solutions one should analyze independent rotations of different partial structures $\rho_l(\mathbf{r})$ (the problem is much easier in the axially-symmetrical case since any rotation of the partial density other than π violates the given axial symmetry).

As an example of a practical application of the method to certain structural investigation some results are presented which were obtained from the SAS curve of the

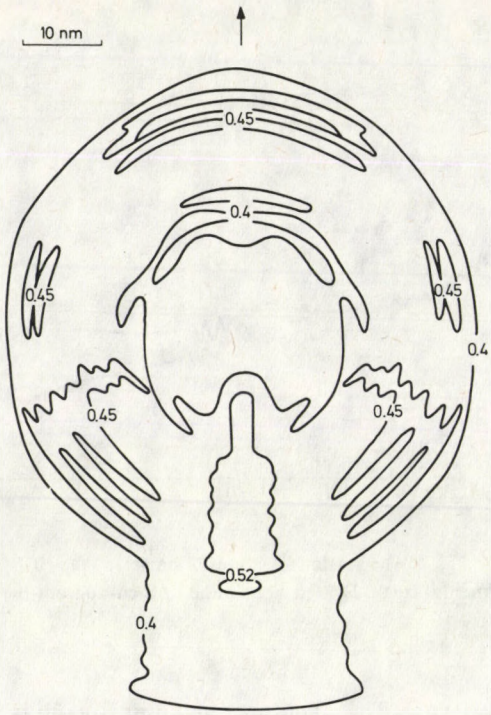


Fig. 7. The map of electron density of bacteriophage T7. The cross-section containing the axis of axial symmetry (indicated by an arrow) is shown. Level 0.4 corresponds to the protein, 0.45 — to strongly hydrated DNA, 0.52 — slightly hydrated DNA. The numbers refer to density levels in 10^{-3} nm^3. The result of the 8th iteration is given, $R_I = 2.2 \times 10^{-2}$

bacteriophage T7. This large bacterial virus has been studied also by various physical and chemical methods. It possesses an approximately axially-symmetrical constitution (isometric polyhedral head and cylindrical tail). Highly refined solutions of the phage T7 for our SAS studies were extracted at the Semmelweis Medical University, Budapest, under the guidance of Prof. I. Tarján and Prof. G. Rontó. We recently measured the X-ray SAS curve of the phage quite precisely in a wide range of angles, a number of general parameters of the phage have been determined and its model has been suggested (Rolbin et al [19]). We have applied the direct method to interpret the part of the curve shown in Fig. 6. The parameters R, $\rho_{\min}$ and $\rho_{\max}$ have been selected by using general phage parameters determined from this curve. An iterative procedure has been applied and the result is presented in Fig. 7.

This figure is a map of the electron density of the phage T7 in cross-section containing the axes of axial symmetry, the radial resolution is 1.2 nm. One can see the projection of the phage head to be embraced by a six-fred symmetry with the corresponding edges equal to 35 nm. The tail appears as a circular cylinder with a radius 11 nm and a height of 18 nm. This is a protein core in the central part of the head

with a diameter of about 24 nm, in good agreement with the data of neutron scattering (Agamalian et al [21]). There is a cylindrical region of higher DNA concentration near the tail, whereas the phage DNA as a whole is strongly hydrated leading to an almost uniform density within the phage. Further on, there are hints in favour of a regular arrangement of DNA, which is revealed by the arcs in Fig. 7.

The scattering curve of the restored density distribution (Fig. 6) coincides with the experimental one well enough up to the scattering angles responsible for the dimensions in a particle compared with the value of radial resolution. All T7 features specified are in good accord with the data obtained (as a rule, by indirect methods) from other physical and chemical techniques.

The results suggest that this method will be quite useful for the structure analysis of biological macromolecules in solution by SAS studies.

It is a honour for the authors to publish this paper in the issue devoted to the 70th birthday of Prof. I. Tarján, who made valuable contributions both to solid state physics as well as biophysics. The results presented in this paper on bacterial viruses would have been impossible without our collaboration with Prof. I. Tarján and his coworkers.

References

1. B. K. Vainshtein, Diffraction of X-rays by Chain Molecules, Elsevier Publ. Comp., Amsterdam, 1966.
2. T. L. Blundell and L. N. Johnson, Protein Crystallography, Acad. Press, New York, 1976.
3. B. K. Vainshtein, Advances in Optical and Electron Microscopy, **7**, 281, 1978.
4. A. Guinier, Compt. Rend., **206**, 1374, 1938; **206**, 1641, 1938.
5. A. Guinier and G. Fournet, Small-angle Scattering of X-rays, J. Wiley, New York, 1955.
6. O. Kratky, Progress in Biophysics, **13**, 105, 1963.
7. L. A. Feigin, Kristallografia, **16**, 711, 1971.
8. B. K. Vainshtein, S. Ya. Karpuhina, N. I. Sosfenov and L. A. Feigin, Doklady AN SSSR, **207**, 1336, 1972.
9. N. A. Gonchar, Yu. M. Lvov, T. G. Samsonidze, L. A. Semina and L. A. Feigin, Biofizika, **23**, 768, 1978.
10. Yu. A. Rolbin, D. I. Svergun, L. A. Feigin and B. M. Schedrin, Kristallografia, **25**, 1125, 1980.
11. H. B. Stuhrmann and R. G. Kirste, Z. Phys. Chem., **46**, 247, 1965.
12. B. K. Vainshtein, L. A. Feigin, Yu. M. Lvov, R. I. Gvozdev, S. A. Marakushev and G. I. Likhtenshtein, FEBS Letters, **116**, 107, 1980.
13. H. B. Stuhrmann, Kristallografia, **26**, 956, 1981.
14. O. Glatter, J. Appl. Cryst., **12**, 166, 1979; **14**, 101, 1981.
15. D. I. Svergun, L. A. Feigin and B. M. Schedrin, Doklady AN SSSR, **261**, 887, 1981; Kristallografia, **26**, 1171, 1981.
16. H. B. Stuhrmann, Acta Crystallogr., **A26**, 297, 1970.
17. G. Marguerie and H. B. Stuhrmann, J. Mol. Biol., **102**, 143, 1976.
18. H. B. Stuhrmann and H. Fuess, Acta Crystallogr., **A32**, 67, 1976.
19. Yu. A. Rolbin, D. I. Svergun, L. A. Feigin, Sh. Gashpar and G. Ronto, Doklady AN SSSR, **255**, 1497, 1980.
20. G. B. Dantzig, Linear Programming and Extensions, Princeton Univ. Press, New Jersey, 1963.
21. M. M. Agamalian, G. M. Drabkin, A. A. Dovgikov, T. I. Krivshich, Yu. M. Lvov, L. A. Feigin, G. Ronto and Sh. Gashpar, Kristallografia, **27**, 92, 1982.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 121—131 (1982)

ON THE KINETICS OF GROWTH OF CONDUCTIVITY ELECTRON CONCENTRATION IN DIELECTRIC CRYSTAL IRRADIATED BY INTENSIVE LIGHT*

L. M. BELJAEV, V. V. NABATOV and YU. YA. YUSHIN

INSTITUTE OF CRYSTALLOGRAPHY, ACADEMY OF SCIENCES OF THE USSR
MOSCOW, USSR

The kinetics is studied of electron avalanche in insulating crystals under the influence of intense light beams. To obtain solutions of kinetic equations, the method of successive approximations is developed. The multistage character of optical damage in insulators is discussed.

1. Introduction

It is well known that the electronic structure of dielectric crystals is such that there are no electrons in the conduction band, which is detached from a completely filled valence band by an energy gap ΔE. Let us suppose that such a crystal is illuminated by a beam of light with frequency $\omega < \Delta E/\hbar$, where $\hbar$ is Planck's constant. If the problem of interaction between the beam of light and the crystal is treated in linear approximation in respect of the intensity of light, then the pure dielectric crystal must not absorb light at all. At the same time there is experimental evidence that laser radiation can give rise to a damage of a transparent insulator if the intensity of light is high enough [1].

One of the possible mechanisms of the optical damage of dielectric crystals is a development of conductivity electron avalanche [1—6]. This process can be imagined as follows. As a consequence of a multiphoton absorption or thermal fluctuations a few electrons appear in the conduction band. Absorbing photons from the incident radiation these primary electrons can gain an amount of energy greater than the minimum energy I which is necessary for impact ionization. The concentration of conductivity electrons increases as a consequence of their multiplication and after all that results in crystal damage.

In theoretical papers dedicated to the avalanche development the concentration of the conductivity electrons $n(t)$ is usually represented as $n_0 e^{\gamma t}$ and the avalanche development rate γ is actually calculated with small values of $n(t)$. As a consequence, nonlinear in $n(t)$ effects remain in the background.

In the present paper the kinetics of the electron avalanche is studied in the case of intensity of incident light close to the damage threshold with due regard to the above

* Dedicated to Prof. I. Tarján on his 70th birthday.

mentioned nonlinear effects. The frequency of light is supposed to belong to the visual range. The following physical effects which are nonlinear in concentration of conductivity electrons are taken into account. Firstly, the excitation of vibratory modes of the crystal by conductivity electrons and holes should be taken into consideration if only this process can cause melting of some portion of the crystal. In its turn, this melting is an important intermediate stage of the damage process as a whole.

Secondly, at definite values of $n(t)$ ($\sim 10^{15}$ cm^{-3}) electron—electron collisions become essential [7]. Such collisions result in a redistribution of energy between electrons and, in consequence, an electron distribution function approximates the Boltzmann one.

Finally, the recombination processes should be taken into account. They reduce the concentrations of electrons and holes. In absence of recombination processes the optical damage of transparent pure crystals has actually no intensity threshold provided that the pulse duration of the incident light is unlimited.

The direct radiative electron—hole recombination cannot apparently be an effective mechanism of reduction of the charge carrier concentration because the cross-section of this transition is too small ($\sim 10^{-19}$ cm^2). The latter can be found from the experimental data on light absorption in crystals making use of the relation of van Roosbroek and Shockley [8]. But recombination transitions can also occur through intermediate states (excitons in the present case). The appropriate elementary process is binding of an electron and a hole in an exciton [9—11]. It has a rather large cross-section $\sigma_{ex} \sim 10^{-13}$ cm^2. The exciton loses energy when scattering on phonons and at length transforms into a photon [11, 12] that leaves the crystal. This process proves to be an effective mechanism of decrease of charge carrier concentration in spite of the competition of a reverse process. This is exciton dissociation to an electron and a hole due to its interaction with a phonon [9].

In Section 2 we present the approximate method of solution of kinetic equations for electron and phonon subsystems of the crystal. That is the method of successive approximations. In Section 3 we shall briefly discuss the multistage character of optical damage of dielectric crystals.

2. Approximate solution of kinetic equations

Let us consider a dielectric crystal in which conduction and valence bands are the standard parabolic bands. We assume also that $m_e \ll m_h$, where m_e and m_h are effective masses of a conductivity electron and a hole, respectively. Suppose that the crystal is irradiated by the light beam, starting from the time moment $t = 0$. We shall describe the plane light wave propagating in the crystal by the vector potential

$$\mathbf{A} = A_0 \mathbf{a} \cos(\mathbf{qr} - \omega t),$$

where A_0 is the amplitude of the potential, $\mathbf{a}$ is the unit polarization vector,

$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t}$. Coupling of this wave with conduction electrons can be characterized by the dimensionless parameter $y = (\hbar\omega m_e)^{-1} e^2 A_0^2$, e is the electron charge. We shall consider such intensities of light when $y \ll 1$.

As far as absorption and emission of visible-light photon by a conductivity electron have quantum nature we shall use quantum theory, proceeding from the Hamiltonian $H = H_0 + H_{\text{int}}$. Here H_0 is the Hamiltonian of noninteracting quasi-particles (electrons, holes, phonons and photons), and H_{int} describes the Coulomb interaction between charge carriers, coupling of electrons with photons and phonons, and anharmonic phonon—phonon interaction.

Kinetic equations for occupation numbers of electrons, holes, and phonons, averaged with crystal density matrix $\hat{\rho}$ such as $n_{\mathbf{k}}(t) = \text{Tr}\,(\hat{\rho}(t)\hat{n}_{\mathbf{k}})$, (master equations) can be obtained within the framework of nonequilibrium quantum statistical mechanics with the aid of Kadanoff—Baym equations and summing up the secular terms in perturbation theory series [13, 14]. Here Tr denotes the trace operation. The problem under consideration is rather specific: there are some small parameters and, besides, absorption and emission of a photon by a conductivity electron must be accompanied by absorption or emission of a phonon to obey the conservation laws of energy and quasi-momentum.

In the course of avalanche development the Fermi gas of conductivity electrons remains nondegenerate. In this case the master equation for electron occupation probabilities has the form

$$\frac{\partial f_e(\mathbf{p}, t)}{\partial t} = (\hat{S}_{ph} + \hat{S}_r + \hat{S}_i + \hat{S}_{\text{ex}} + \hat{S}_c) f_e(\mathbf{p}, t)\,. \tag{1}$$

where $f_e(\mathbf{p}, t)$ is the conduction electron distribution function and $\hat{S}$ are transition probabilities. $\mathbf{p}$ is electron quasi-momentum. Distribution function of holes in the valence band $f_h(\mathbf{p}, t)$ satisfies the analogous equation. The terms $\hat{S}f$ have the following common structure. Each of them is the sum of products of the normalizing factor, modulus of the matrix element squared, deltafunction standing for energy conservation law in the elementary process, electron and hole distribution functions in the initial states, and factor $N_{\mathbf{k}}(t)$ or $N_{\mathbf{k}}(t) + 1$ in the case of absorption or emission of phonon. It is too cumbersome to write down the explicit expressions here and we shall refer the readers to the relevant literature. It is essential that in the case of nonpolar crystals the interaction of electrons with acoustic and optical phonons (scattering on the acoustic and optical deformation potentials) is taken into consideration in H_{int}. If the crystal has polar optical vibratory modes, we suppose that their interaction with conductivity electrons is weak. It is also taken into account in H_{int}.

The term $\hat{S}_{ph} f_e$ describes phonon absorption and emission by conductivity electrons [7, 15]. Absorption of photons from the light beam by electron and photon

emission in the same mode with simultaneous scattering on phonons is presented by $\hat{S}_r f_e$ [15]. Due to $y \ll 1$ we restrict ourselves to one photon processes. The term $\hat{S}_i f_e$ describes the impact ionization of electron and hole both with [3] and without photon absorption [16, 17]. The term $\hat{S}_{ex} f_e$ corresponds to the binding of electrons and holes in excitons [9] and $\hat{S}_c f_e$ takes into consideration electron—electron and electron—hole collisions in Born approximation [14].

Strictly speaking, Eq. (1) should be supplemented by the kinetic equation for excitons in which the following processes are taken into account: exciton dissociation, exciton—phonon scattering and luminescence of excitons (polaritons) [11, 12]. We shall consider these processes in a phenomenological way multiplying σ_{ex} by a coefficient $A < 1$. Accordingly, Eq. (1) contains an effective probability of the disappearance of electron and hole as a final result of the binding process.

Kinetic equations for the average numbers of phonons $N_{\mathbf{k}j}(t)$ of the mode j with quasi-momentum $\mathbf{k}$ can be written as

$$\frac{\partial N_{\mathbf{k}j}(t)}{\partial t} = (\hat{L}_a + \hat{L}_e) N_{\mathbf{k}j}(t) . \tag{2}$$

Here $\hat{L}_a N_{\mathbf{k}j}(t)$ represents the variation of $N_{\mathbf{k}j}(t)$ with time as a result of phonon—phonon interaction, and $\hat{L}_e N_{\mathbf{k}j}(t)$ describes the phonon absorption and emission in the course of the above electronic transitions [7, 15].

A kinetic equation such as (1) was considered in [3] with some difference: the last two terms on the r.h.s. of Eq. (1) were omitted and, accordingly, the above mentioned nonlinear effects were not considered. For the approximate solution of the kinetic equation the method of slowly varying amplitudes was useful in [3]. It had been worked out earlier in the paper [18] where the problem of gas breakdown by laser radiation was studied. In [3] the function $f_e(\mathbf{p}, t)$ was presented as $\exp(\gamma t) f(\mathbf{p})$. In doing so the impact ionization was considered as a slow process in contrast to the quick establishment of energy distribution in the electron system.

We intend to solve the system of equations (1), (2) at light intensity values close to the threshold of optical damage. These equations are nonlinear: operators $\hat{S}$ depend on $N_{\mathbf{k}j}(t)$, $\hat{L}_e$ contains $f_{e,h}(\mathbf{p}, t)$ and, besides, the last two terms on the r.h.s. of (1) are the bilinear expressions in respect of $f_{e,h}$. Keeping in mind the arguments given in the Introduction we shall generalize the method of slowly varying amplitudes and look for an approximate solution of (1) in the form $n(t)\tilde{f}_e(\mathbf{p}, n(t))$. Here $n(t)$ is the concentration of conductivity electrons at the time t and the function $\tilde{f}_e(\mathbf{p}, n(t))$ satisfies

$$(\hat{S}_{ph} + \hat{S}_r + \hat{S}_c) \tilde{f}_e(\mathbf{p}, n(t)) = 0 . \tag{3}$$

Then Eq. (1) reduces to

$$\frac{\mathrm{d}n(t)}{\mathrm{d}t} = a(t)\, n(t) - b(t)\, n^2(t) , \tag{4}$$

where $a(t)$ and $b(t)\,n(t)$ are the probabilities of impact ionization and electron—hole binding in an exciton averaged with $\tilde{f}_e(\mathbf{p}, n(t))$ and the respective function for holes $\tilde{f}_h(\mathbf{p}, n(t))$.

The system of equations (2)—(4) is still too complicated to be solved exactly. As it was shown in [14], the conservation laws should be taken into consideration when looking for approximate solutions of kinetic equations, i.e. the approximate solutions must satisfy them. When solving (2)—(4) we propose to use the method of successive approximations. We obtain the first approximation $f^{(1)}_{e,h}(\mathbf{p}, t)$, if we put in (3)—(4) $N_{\mathbf{k}j}(t)$ to be equal to their values at $t=0$. The latter are determined by Planck's formula for the initial temperature T_0 of the crystal. Then we put the solutions of the resulting equations into Eq. (2) and find the time dependent average occupation numbers of phonons $N^{(1)}_{\mathbf{k}j}(t)$. Now we can write down equations for $f^{(2)}_{e,h}(\mathbf{p}, t)$ by substituting $N^{(1)}_{\mathbf{k}j}(t)$ in (3), (4) for $N_{\mathbf{k}j}(t)$ and so on. It is obvious how to continue the iteration procedure.

If this procedure is quickly convergent, we shall have a rather precise solution after a few steps already. In any case, even $N^{(1)}_{\mathbf{k}j}(t)$ reveals crystal lattice heat-up produced by conduction electrons and holes. And in $f^{(2)}_{e,h}(\mathbf{p}, t)$ the reverse influence of this heat-up on the distribution functions of charge carriers is duly taken into consideration.

The main obstacle to the solution of Eq. (3) consists in the quantum nature of absorption and emission of light by electrons: as a result of these transitions the electron energy $\varepsilon(\mathbf{p})$ changes by a quantity $\sim\hbar\omega$ which is commensurable with the impact ionization threshold I. Making use of the transition probabilities averaged on polarization directions of the incident light we shall look for the solution of Eq. (3) holding that $\tilde{f}_e(\mathbf{p}, n(t))$ depends on $|\mathbf{p}|$ only. Then this solution can be expanded in terms of the complete orthonormal set of functions $\varphi_n(x)=L_n(x)\exp\left(-\frac{x^2}{2}\right)$ as follows

$$\tilde{f}_e(\mathbf{p}, n(t)) = \sum_{n=0}^{\infty} c_n(t)\varphi_n(x), \quad x = \sqrt{\frac{|\mathbf{p}|^2}{m_e k T_e}}. \tag{5}$$

Here $L_n(x)$ are Laguerre polynomials, and k is the Boltzmann constant. The advantage of such an expansion consists in the fact that $\varphi_0(x)=\exp[-|\mathbf{p}|^2/2m_e kT_e]$. At $n(t)>n_c\sim 10^{15}\,\mathrm{cm}^{-3}$ collisions between electrons will result in the redistribution of energy and all $c_n(t)$, besides $c_0(t)$, will become zero. Also the expansion (5) contains a free parameter, that is the electronic temperature T_e. Its value can be determined in such a way that the energy conservation law is fulfilled for the approximate solutions.

Let us find out an approximate distribution function of conductivity electrons by putting in (5) all $c_n(t)$, besides $c_0(t)$, to be equal to zero for all time values. Note that we can look for corrections to such an approximation also with the aid of the expansion (5).

In the course of the iteration process we shall determine a value of $T_e(t)$ from the energy balance condition: the energy $n(t)\left\langle -\frac{\mathrm{d}\varepsilon}{\mathrm{d}t}\right\rangle$ lost in unit of volume by conductivity

electrons per unit time must be equal to the energy SK obtained by them in light absorption and emission processes. Here S in the intensity of incident light and K is the light absorption coefficient. In the case of nonpolar crystals this condition is

$$-\left\langle \frac{d\varepsilon}{dt} \right\rangle_a - \left\langle \frac{d\varepsilon}{dt} \right\rangle_{0d} = \frac{1}{n(t)} S(K_a + K_{0d}). \tag{6}$$

Here the quantities averaged with the function φ_0 appear and

$$-\left\langle \frac{d\varepsilon}{dt} \right\rangle_a = \frac{8\sqrt{2}}{(\pi)^{3/2}} \frac{E_1^2 m_e^{5/2}}{\hbar^4 \rho} (kT_e)^{3/2} \left(1 - \frac{T}{T_e}\right).$$

Here $T(t)$ and ρ are the temperature and the density of the crystal, respectively, E_1 is the deformation potential constant [7]. Also

$$-\left\langle \frac{d\varepsilon}{dt} \right\rangle_{0d} = \frac{m_e^{3/2} D^2 \sqrt{\hbar\omega_0}}{\pi^{3/2} \hbar^2 \rho} \frac{\sqrt{\lambda z}\, K_1(\lambda z)}{sh(z)} sh[(1-\lambda)z],$$

where ω_0 is the maximum frequency of optical phonon, $\lambda = T/T_e$, $z = \dfrac{\hbar\omega_0}{2kT}$, and D is the optical deformation potential constant [15]. From now on $K_n(x)$ is the modified Bessel function, $n = 0, 1, 2$. The contribution to the right hand side of Eq. (6) from the above mentioned radiative transitions accompanied by absorption or emission of an acoustical phonon is given by

$$\frac{1}{n(t)} SK_a = A_0^2 \omega C (2m_e \lambda k T)^{1/2} sh\left(\frac{\hbar\omega}{2kT_e}\right) K_2\left(\frac{\hbar\omega}{2kT_e}\right),$$

where $C = e^2 E_1^2 (3\pi^{3/2} \hbar \rho u_e^2)^{-1}$ and u_e is the sound velocity in the crystal [15]. This is absorbed radiation energy attributed to one electron. The same quantity in the case when an optical phonon is emitted or absorbed is [15]

$$\frac{SK_{0d}}{n(t)} = A_0^2 \omega^{-1} B (2kT_e)^{3/2} \frac{sh(z_+ - z) z_+^2 K_2(z_+) + sh(z_- + z) z_-^2 K_2(|z_-|)}{sh(z)},$$

where

$$B = \frac{e^2 D^2 m_e^{1/2}}{6\pi^{3/2} \hbar^4 \rho \omega_0}, \qquad z_\pm = \frac{\hbar\omega \pm \hbar\omega_0}{2kT_e}.$$

Eq. (6) defines T_e as a function of A_0^2 and T.

Let us turn our attention to polar insulating crystals. If the dimensionless polar constant $d_p \ll 1$, then T_e can be found from the energy balance condition

$$-\left\langle \frac{d\varepsilon}{dt} \right\rangle_a - \left\langle \frac{d\varepsilon}{dt} \right\rangle_{0d} - \left\langle \frac{d\varepsilon}{dt} \right\rangle_{0p} = \frac{1}{n(t)} S(K_a + K_{0d} + K_{0p}) . \tag{7}$$

Here [15]

$$-\left\langle \frac{d\varepsilon}{dt} \right\rangle_{0p} = \frac{2^{3/2} dp(\hbar\omega_0)^2 \sqrt{\lambda z}}{\pi^{1/2}\hbar} \frac{sh[(1-\lambda)z] K_0(\lambda z)}{sh(z)},$$

$$\frac{1}{n(t)} SK_{0p} = A_0^2 \frac{dpe^2(\hbar\omega_0)^{3/2}}{3\pi^{1/2}\hbar^2\omega m_e} (kT_e)^{1/2} [sh(z)]^{-1} \times$$

$$\times [sh(z_+ - z) z_+ K_1(z_+) + sh(z_- + z) |z_-| K_1(|z_-|)] .$$

The analogous equations determine a temperature of holes (we remind that $m_e \ll m_h$).

The first approximation of the iteration method can be obtained if we put $T = T_0$ in Eqs. (6), (7). It is obvious that quantities $T_e^{(1)}$ and $T_h^{(1)}$ determined in such a way are time-independent. Now we can evaluate the coefficients a and b in Eq. (4). In doing so we shall average the probabilities of impact ionization $w_0(\varepsilon)$ and $w_1(\varepsilon)$ (without the participation of photons [16, 17] and with the absorption of one photon [3], respectively) and the quantity $|\mathbf{v}_e - \mathbf{v}_h| \sigma_{ex}$ with functions $\tilde{f}_{e,h}^{(1)}(\varepsilon)$ ($\mathbf{v}_e$ and $\mathbf{v}_h$ are velocities of electron and hole, respectively)

$$a = \langle w_0(\varepsilon) + w_1(\varepsilon) \rangle , \qquad b = \langle |\mathbf{v}_e - \mathbf{v}_h| \sigma_{ex} \rangle A .$$

The solution of Eq. (4) is ($n_0 = n(0)$)

$$n_1(t) = \frac{n_0 a \exp(at)}{a - n_0 b + n_0 b \exp(at)} . \tag{8}$$

The asymptotic behaviour of (8) is $n_1(t) \to a/b$, when $t \to \infty$. Turning to Eq. (2) we shall bear in mind that as a result of the anharmonic interaction the energy received by a phonon subsystem from conductivity electrons and holes tends to be quickly redistributed between the vibratory modes. Accordingly, $N_{\mathbf{k}j}(t)$ will be given by Planck's formula with time-dependent temperature $T(t)$, which in turn satisfies

$$\frac{dT(t)}{dt} = \frac{n(t)}{c_v} \left[-\left\langle \frac{d\varepsilon_e}{dt} \right\rangle - \left\langle \frac{d\varepsilon_h}{dt} \right\rangle \right], \tag{9}$$

where c_v is the specific heat of solid, $T(0) = T_0$ and the magnitude $-\left\langle \frac{d\varepsilon_e}{dt} \right\rangle$ is given by

the left hand side of Eq. (6) or Eq. (7). Let us substitute $n_1(t)$ given by Eq. (8) for $n(t)$ in Eq. (9) and integrate. The result is

$$T(t) = T_0 + (c_v b)^{-1} \ln\left[\frac{b}{a} n_0 (e^{at} - 1) + 1\right]\left\{-\left\langle \frac{d\varepsilon_e}{dt}\right\rangle - \left\langle \frac{d\varepsilon_h}{dt}\right\rangle\right\}. \tag{10}$$

The second approximation for $f_e(\mathbf{p}, t)$ can be obtained by substitution of Eq. (10) for T in (6), (7). By solving the equations we find the time-dependent quantity $T_e(t)$. Also the coefficients a and b in (4) will be time-dependent. Due to inequality $\frac{d}{dt}\left[\ln\left(\frac{n_0 b(t)}{a(t)}\right)\right] \ll a(t)$ the approximate formula can be obtained which describes the time-dependence of conductivity electron concentration with due regard to crystal lattice heat-up

$$n_2(t) = n_0 \frac{a(t) \exp\left(\int_0^t a(\tau)\, d\tau\right)}{a(t) - n_0 b(t) + n_0 b(t) \exp\left(\int_0^t a(\tau)\, d\tau\right)}. \tag{11}$$

Note that $n(t)$ can also be given by (11) in such a case when light intensity depends on time.

Now we shall give some numerical estimates making use of the parameter values which are characteristic of dielectrics with wide forbidden band: $m_e \sim 10^{-27}$ g, $u_e \sim 7 \cdot 10^5 \frac{\text{cm}}{\text{s}}$, $\hbar\omega_0 \sim 0{,}1$ eV, $I \approx 6$ eV, $E_1 \approx 11$ eV (these values refer to sapphire), $D = 5 \cdot 10^8 \frac{\text{eV}}{\text{cm}}$. At $T = 300$ K we find from Eq. (6) that $kT_e \sim 0.5$ eV corresponds to vacuum electrical field strength in the incident beam of order $10^5 \frac{\text{V}}{\text{cm}}$. We shall assume that at $\varepsilon > I$ $w_0(\varepsilon)$ linearly depends on their difference [16] as $r(\varepsilon - I)$. In this case $\langle w_0(\varepsilon)\rangle \approx \frac{2r}{\sqrt{\pi}} (kT_e I)^{1/2} \exp(-I/kT_e)$. If $r \sim 0.5 \cdot 10^{14}\ \text{eV}^{-1}\,\text{s}^{-1}$ then $a \sim 10^{10}\ \text{s}^{-1}$. And at $A\sigma_{ex} \sim 10^{-14}\ \text{cm}^2$ we obtain $n(\infty) \sim 10^{16}\ \text{cm}^{-3}$.

At the initial stage of avalanche development in alkali halide crystals our treatment cannot be used directly because the polar constant $\alpha_p > 1$. It can be considered only as a qualitative one. But at high values of $n(t)$ the electron—phonon interaction is screened, i.e. $\alpha_p = \alpha_p(n)$ [19]. Owing to this, we can use our treatment starting from such values of $n(t)$ when $\alpha_p(n(t)) < 1$. Also we must assign initial conditions $n(t_0) = n_0$ at some moment $t_0 > 0$. However, some modifications are necessary. The first approximation for $n(t)$ is given by a formula like (11) where now $a(t)$ and $b(t)$ must be determined in the self-consistent way. And $T(t)$ takes approximately the form of Eq. (10) with the substitution of e^{at} by $\exp\left(\int_{t_0}^{t} a(\tau)\, d\tau\right)$. Note that the same formula with $t_0 = 0$ describes $T(t)$ in the case of time-varying light intensity.

3. Multistage character of optical damage process

In experiments on optical damage only some part of the crystal is usually under the influence of intensive light. If use is made of the focussed light beams, then it is the caustic region. In such a case the results of Section 2 should be referred to such a small region only.

Here we want to lay emphasis on the fact that laser damage in insulating transparent crystals has multistage character, although it is a fleeting phenomenon. The stages differ from each other both by the phase state of matter in the irradiated region and the light absorption mechanism and they are dominant physical processes. The electron avalanche development discussed in Section 2 can be considered as the first stage of the entire optical damage process. In due time the irreversible variations take place in the crystal at high values of $n(t)$. Let t_1 be the solution of equation $T(t) = T_m$, where T_m is the melting temperature of the crystal and $T(t)$ is given by (10). Then at $t > t_1$ the crystal melts in the above small region. There is, at the same time, a considerable concentration of free charge carriers (electrons and holes).

The arrangement of ions in the liquid state has been described by Bernal as "random close packing". Therefore the localization of some amount of valency electrons on the outer shells of ions, as it was in the crystalline state, becomes energetically unfavourable. It seems reasonable to suppose that in a small part of the solid a new, e.g. a liquid metal phase appears, with the free charge carriers concentration of the order of $10^{20} - 10^{21}$ cm^{-3}. This results in an ever increasing absorption of incident light (photon absorption by the electron is accompanied by electron—ion scattering). The temperature in the considered region also increases quickly.

The next stage of optical damage consists in the formation of electron—ion plasma whose temperature is fairly high and the density of free electrons is $\gtrsim 10^{23}$ cm^{-3}. At this stage the continuous radiation is emitted from the crystal [20, 21]. Its origin is probably connected with electron bremsstrahlung and recombination radiation. The experimental study of spectrum of radiation emitted from alkali halide crystals in the course of their laser damage gives the value of plasma temperature up to 27 000 K [20, 21]. Again the lines of doubly ionized aluminium were identified when the damage region in ruby was investigated spectroscopically [21].

The volume filled with plasma increases as a consequence of the energy transfer from plasma particles to ions of crystal phase on the interface. The low heat conductivity of the insulating crystal results in the onset of considerable temperature gradients in the frontier of the crystal, inducing thermal stress [22].

When the thermal stress reaches its critical value, the stress crack arises in the cleavage plane of the crystal. The crack is developing, while the electron—ion plasma is penetrating into its cavity. The total absorption of light falls. A movement of the crack front is accompanied by the same processes as in the case of the brittle failure under

mechanical influence: the electron emission from the newly formed solid surfaces [23], the triboluminescence [24], and the radio-frequency radiation [25].

The damage process ends with the thermochemical processes in the damage region: plasma cools down and its matter precipitates on the new solid surfaces. The precipitation is accompanied by considerable crystal disordering in close proximity to the above surfaces (formation of vacancies, interstitial atoms and so on). This supposition can be indirectly corroborated by the coloration that takes place in the alkali halide crystals near the zone of fracture [26].

4. Discussion

When the kinetic equation is solved in the diffusive approximation upon the energy variable [2, 4, 5], then function $f(\varepsilon+\hbar\omega)$ is substituted by $f(\varepsilon)+\hbar\omega f'(\varepsilon)+\frac{(\hbar\omega)^2}{2}f''(\varepsilon)$ for all values of ε. This is equivalent to the neglect of all the other terms in the Taylor expansion. In order to clarify the question for what types of functions $f(\varepsilon)$ this procedure is legitimate, we shall write down the remainder of Taylor series in the Lagrange form as $R_2=\frac{(\hbar\omega)^3}{3!}f'''(\varepsilon+\theta\hbar\omega)$, where $0<\theta<1$. Thus the terms neglected in the diffusive approximation are negligible indeed if the inequality

$$\max|f'''(\varepsilon)| \ll \frac{3!}{(\hbar\omega)^2}\left|f'(\varepsilon)+\frac{\hbar\omega}{2}f''(\varepsilon)\right|$$

is valid for all energy values. On the left side here there is a maximum value of $|f'''(\varepsilon)|$ within the energy interval under consideration. If the photon energy $\hbar\omega>1.5$ eV, it is comparable to the impact ionization threshold I. Then in interval $0<\varepsilon<I$ the inequality is satisfied only by functions $f(\varepsilon)$ obeying $f'''(\varepsilon)\approx 0$, i.e. $f(\varepsilon)\approx c_1+c_2\varepsilon+c_3\varepsilon^2$. Again the energy balance condition must be kept in mind. For its fulfilment the average number of electron transitions per unit time, with an incident photon absorbed, must exceed the average number of transitions with induced photon emission. This rules out functions $f(\varepsilon)$ monotonically increasing with the growth of energy.

As soon as $n(t)$ reaches the values when electron—electron collisions become essential, function $f(\varepsilon)$ approximates to the Boltzmann distribution function. Then the above inequality takes the form $\frac{2}{3}\left(\frac{\hbar\omega}{2kT_e}\right)^2 \ll \left|\frac{\hbar\omega}{2kT_e}-1\right|$. It is not fulfilled if the intensity of light is close to its damage threshold value and $\hbar\omega>1.5$ eV.

In this paper heat diffusion from the part of the crystal which is exposed to the incident radiation was left out of account. It is of importance at $t>\tau\sim 10^{-3}$ s. If the asymptotic value of $n(t)$ does not cause irreversible variations in the crystal at $t<\tau$, then at $t>\tau$ some quasi-stationary magnitude of T is set in. In fact, it does not differ much

from T_0. Photoconductivity and exciton luminescence can be recorded in these subthreshold conditions.

The plasma frequency $\omega_p = \left(\frac{4\pi n(t)e^2}{m_e}\right)^{1/2}$ can also be evaluated by means of the expressions obtained here. A discovery of light scattering on plasma oscillations at subthreshold intensities could bear out the electron avalanche mechanism of optical damage in insulating crystals.

The authors consider it an honour to be able to publish this paper in the special issue of the Journal dedicated to the 70th birthday of Professor I. Tarján, who has made important contributions to the physics of dielectric and, particularly, ionic crystals. We are happy to mention that our research cooperation with Professor I. Tarján has been in progress for more than two decades.

References

1. N. Bloembergen, J. Quantum Electron., **10,** 375, 1974.
2. A. G. Molchanov, Fiz. Tv. Tela, **12,** 954, 1970.
3. V. N.Strekalov, Fiz. Tv. Tela, **15,** 1373, 1973.
4. A. S. Epifanov, JETPh, **67,** 1805, 1974.
5. S. I. Zaharov, JETPh, **68,** 2167, 1975.
6. S. Brawer, Phys. Rev., **B20,** 3422, 1979.
7. E. M. Conwell, High Field Transport in Semiconductors, Academic Press, New York, 1967.
8. J. I. Pankove, Optical Processes in Semiconductors, Prentice-Hall, Inc., New Jersey, 1971.
9. A. A. Lipnik, Fiz. Tv. Tela, **3,** 2322, 1961.
10. B. Goodman and O. S. Oen, J. Phys. Chem. Solids, **8,** 291, 1959.
11. R. S. Knox, Theory of Excitons, Solid State Phys., **5,** 1, 1963.
12. Y. Toyozawa, Progr. Theor. Phys., **20,** 53, 1958.
13. S. Fujita, Introduction to Non-equilibrium Quantum Statistical Mechanics, W. B. Saunders Company, Philadelphia, 1966.
14. L. P. Kadanoff and G. Baym, Quantum Statistical Mechanics, W. A. Benjamin, Inc., New York, 1962.
15. K. Seeger, Semiconductor Physics, Springer-Verlag, Wien, 1973.
16. L. V. Keldysh, JETPh, **37,** 713, 1959.
17. W. Franz, Dielektrischer Durchschlag, Handbuch der Physik, **17,** 155, 1956.
18. Ya. B. Zeldovich and Yu. P. Rizer, JETPh, **47,** 1150, 1964.
19. H. Ehrenreih, J. Phys. Solids, **8,** 130, 1959.
20. L. M. Beljaev, A. N. Golovistikov and V. V. Nabatov, Fiz. Tv. Tela, **10,** 3733, 1968.
21. J. F. Ready, Effects of High-Power Laser Radiation, Academic Press, New York, 1971.
22. V. L. Indenbom and E. M. Shefter, JETPh, Pisma, **4,** 258, 1966.
23. L. M. Beljaev, Yu. N. Martyshev and Yu. Ya. Yushin, Acta Phys. Hung., **33,** 307, 1973.
24. L. M. Beljaev, Yu. N. Martyshev, Phys. Stat. Sol., **34,** 57, 1969.
25. L. M. Beljaev, V. V. Nabatov and Yu. N. Martyshev, Kristallografiya, **7,** 576, 1962.
26. V. V. Nabatov, L. M. Beljaev, N. N. Dymenko and R. Voszka, Acta Phys. Hung., **47,** 107, 1979.
27. P. M. Platzman and P. A. Wolff, Waves and Interactions in Solid State Plasmas, Sol. State Phys. Suppl., **13,** 1, 1973.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 133—141 (1982)

THE GROWTH AND TEXTURE STUDY OF GaSb INGOTS*

M. HÁRSY, F. KOLTAI, I. GYURÓ, T. GÖRÖG and E. LENDVAY

RESEARCH INSTITUTE FOR TECHNICAL PHYSICS, HUNGARIAN ACADEMY OF SCIENCES
1325 BUDAPEST, HUNGARY

GaSb was synthetized from Ga of 6N and Sb of 5N5 purity. After the synthesis the crystallization took place in Bridgman-type growth systems. Terrestrial and space experiments aboard Salyut-6 have been performed. In terrestrial experiments depending on the growth parameters (pulling rate, ampoule diameter, etc.) macro-, or polycrystalline ingots showing a characteristic texture have been formed. It has been observed that in wider ampoules boules of better quality have grown, the diameter of individual crystals ranging from 1—3 cm. In the space experiment a high-quality bicrystal has been formed. Examining the texture of terrestrial samples, the orientation of crystallites, the number of low and wide angle grain boundaries have been determined using electron channelling (EC) patterns.

1. Introduction

In the last few years GaSb has become one of the most attractive semiconducting materials. Apart from its optoelectronic application as laser diodes or detectors in the near infrared range [1—3], the single crystals of GaSb are excellent substrate materials for developing different ternary and quaternary layer structures. The GaSb, for instance, has a close lattice match with AlGaSb, AlGaAsSb and InGaAsSb. These hetero systems grown epitaxially onto GaSb substrates are used for photodetectors, LED and laser devices working in the visible and the near infrared regions of the spectrum [4—6].

The most essential requirement for these applications is the production of perfect single crystalline GaSb. To prepare usable GaSb crystals is, however, not a simple problem. Numerous publications describing different growth processes and aimed at improving the quality of GaSb crystals have been published, but the materials obtained so far often contain microscopic and macroscopic defects as well [7—11]. The crystals usually show twinning and also microfaceted growth is frequently observed. Boules pulled by the widely used Czochralski technique exhibit macroscopic heterogeneity of impurities due to the so-called "facet effect" [12].

In this paper we present some results of our program to study the growth properties of GaSb and if possible to improve the quality of crystals. Luckily we had a chance to perform a GaSb crystallization experiment during the first Soviet—Hungarian joint space flight in 1980. Some of the findings will also be discussed.

* Dedicated to Prof. I. Tarján on his 70th birthday.

2. Experimental part

The synthesis and the subsequent crystallization of GaSb samples have been carried out in a Bridgman-type vertical arrangement. The gallium and arsenic in a stoichiometric ratio were placed into a conically tipped quartz tube which was later sealed under vacuum. Heating the ampoule to a temperature slightly over 730 °C the reaction of the components was completed within an hour. After the synthesis the ampoule was lowered slowly along the decreasing temperature gradient of the furnace. The lowering rate varied between 0.3 and 18 mm/hour, and ampoules of different diameters (8—30 mm) and shapes were tried. Fig. 1 illustrates some of the ampoule shapes. The lengths of the ingots were 40—50 mm.

In the space experiment a polycrystalline ingot of a length of 39 mm and a diameter of 8 mm was melted, then pulled out with a rate of 0.188 mm/min to the cooler part of the tube furnace where the material gradually crystallized. The experimental set-up is demonstrated in Fig. 2.

For the investigation, the upper top of the ingot was removed, then the remaining part was cut to halves along its length. The flat surfaces were polished and etched in the 1:1:2 mixture of $HCl:H_2O_2:H_2O$. The latter made the wide angle grain boundaries visible.

The fine structure measurements took place in JEOL JSM-35 type three lense Scanning Electron Microscope (SEM) using the Electron Channelling (EC) effect (Fig. 3).

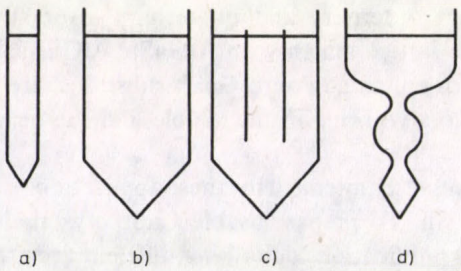


Fig. 1. Different ampoule shapes used for GaSb growths

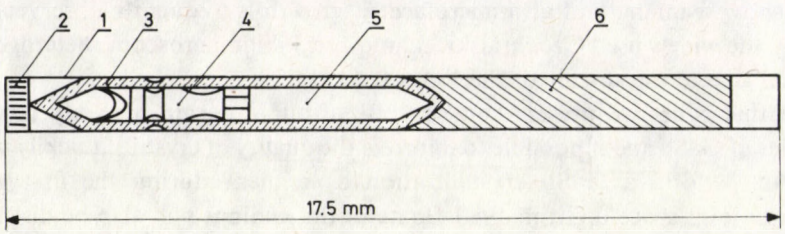


Fig. 2. GaSb space capsule. 1. Stainless steel tube, 2. screwed in cap, 3. quartz ampoule, 4. graphite spacer, 5. polycrystalline GaSb, 6. asbestos wool

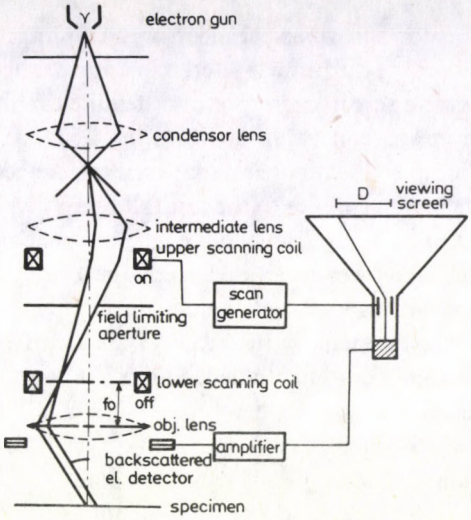


Fig. 3. EC measuring set-up

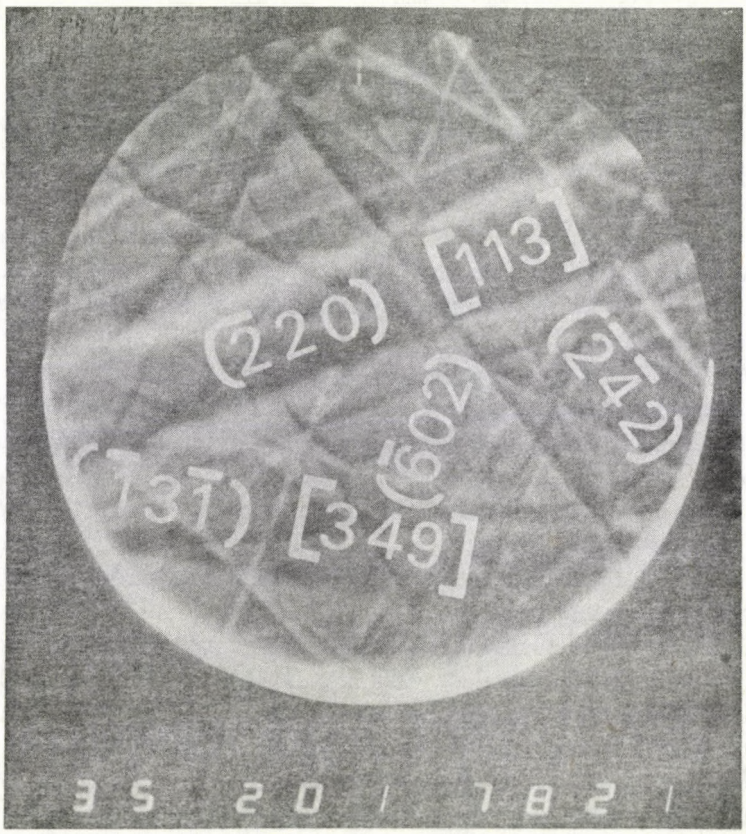


Fig. 4. A typical EC pattern

An EC pattern is formed when the electron beam impinges a small area of the sample, while the direction of the beam is altered in appr. 8° surroundings of the optical axis. Points of the viewing screen correspond to definite directions of the beam and a Kikuchi pattern like image consisting of white lines can be observed (Fig. 4). The brightness is proportional to the number of backscattered electrons, and their number is a function of the angle of incidence to the crystallographic planes. The cross-section of the Rutherford scattering is higher when the angle of incidence is somewhat smaller than the Bragg angle, and lower when it exceeds the Bragg angle.

The accelerating voltage and the beam current used were 35 kV and appr. 10^{-7} A, respectively. The diameter of the beam at the sample surface was 10—30 μm, therefore rather small areas could be studied. The crystallographic planes corresponding to the pairs of lines and bands were determined by a trial-and-error procedure, then the laboratory coordinates of the [001], [010] and [100] directions were computed. Using the orientation matrix composed from these data the laboratory indices can be converted to normal Miller indices or vice versa. Applying this computation, both the texture and the grain boundary properties could be examined.

3. Results and discussion

Figs 5—6 show some representative samples.

The ingots are solid and free of inclusions but the surface of terrestrial ones frequently contain semi-spherical depressions due to vacuum residues between the quartz wall and the molten material. The space sample shows a peculiar ridged surface which suggests that the crystallization was quite free of wall effects and slow fluid motions, not detectable under ordinary gravitational conditions, determine the shape of the growing interface [13].

The crystallites of the terrestrial ingots were elongated in some cases to lengths a hundred times their widths, and positioned more or less parallel to each other, along the ampoule length. Other samples consisted of only a few crystallites of dimensions near those of the ingots.

Moving the electron beam perpendicularly to the ingot length the changes in EC patterns indicated the individual grains, and in this manner the number of grain boundaries could be counted. Table I summarizes some growth conditions and boundary numbers found in selected samples. The Table shows that in wider ampoules the nucleation is lower. The sizes of individual grains grow with decreasing pulling rate and this effect is more significant for wider ampoules. It can be supposed that the wall effect and some kind of stress, arising when the cooling of the boule is not uniform, play an important role in propagating the nucleation. Sample VII shows a good example of stress effect (Fig. 7). In a D-type ampoule the diameter of the interface was alternatively changed during the growth process, consequently the heat transport conditions varied, too. Applying a slow lowering rate, similar to that of Sample III, many crystallites have

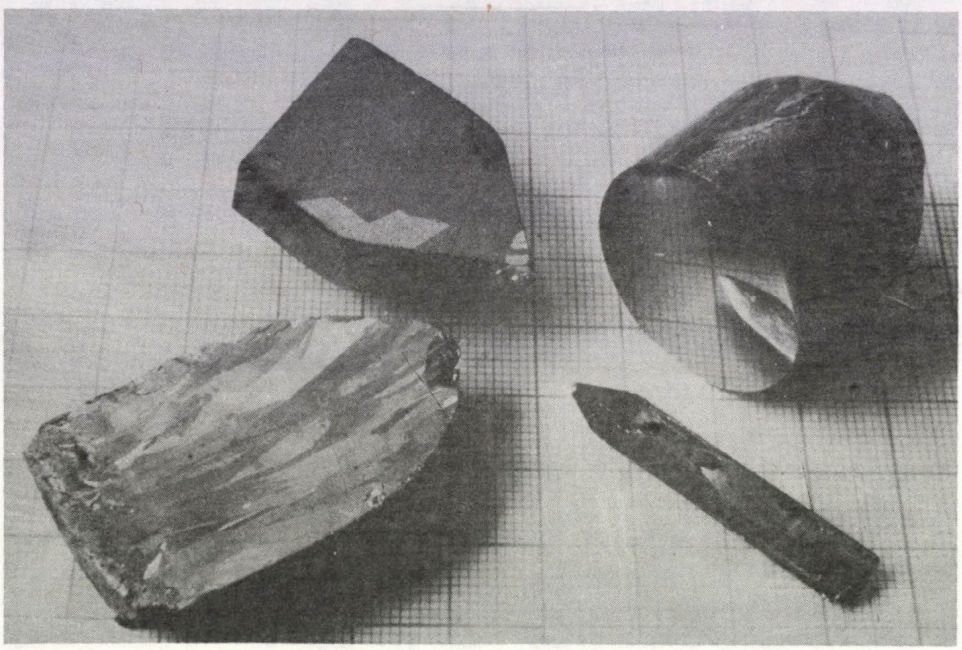


Fig. 5. Different GaSb ingots

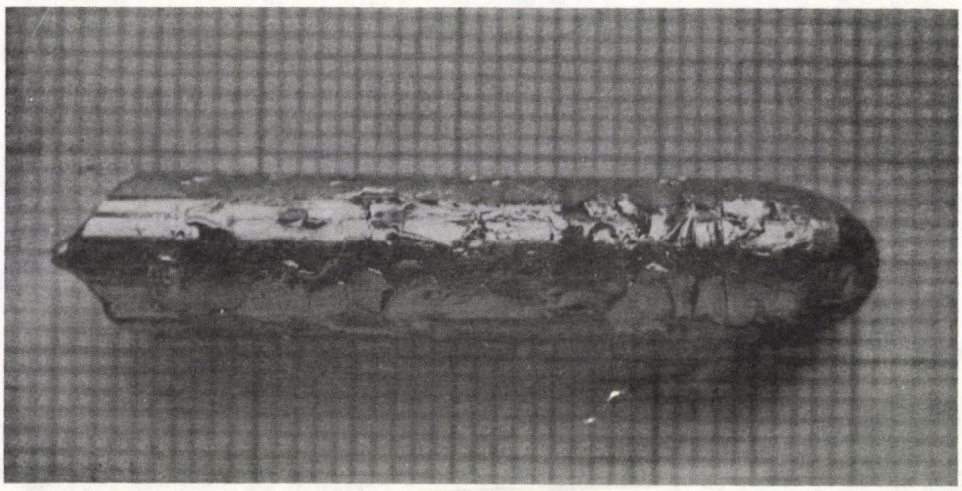


Fig. 6. Space-grown GaSb ingot

Table I

Growth parameters and number of boundaries/cm found perpendicularly to ingot length [14]

Sample	Diameter [mm]	Pulling rate [mm/h]	Boundaries [cm]		
			law angle	wide angle	total
I-3A	8	0.3	6	4	10
I-3B	8	18	20	10	30
II	28	18	16	8	24
III	28	3.6	7	0–1	7–8

formed, especially in the bulk of the balls, indicating the nucleation prone sections of the ingot.

The space sample gives an example to the contrary. In microgravity conditions the crystallization took place practically free of wall effect, resulting in a bicrystal with only three small twin lamellas penetrating into it (Fig. 8).

As has been already mentioned, the grains are elongated more or less parallel with the pulling direction. The chemically developed twin lamellas, in ingots consisting

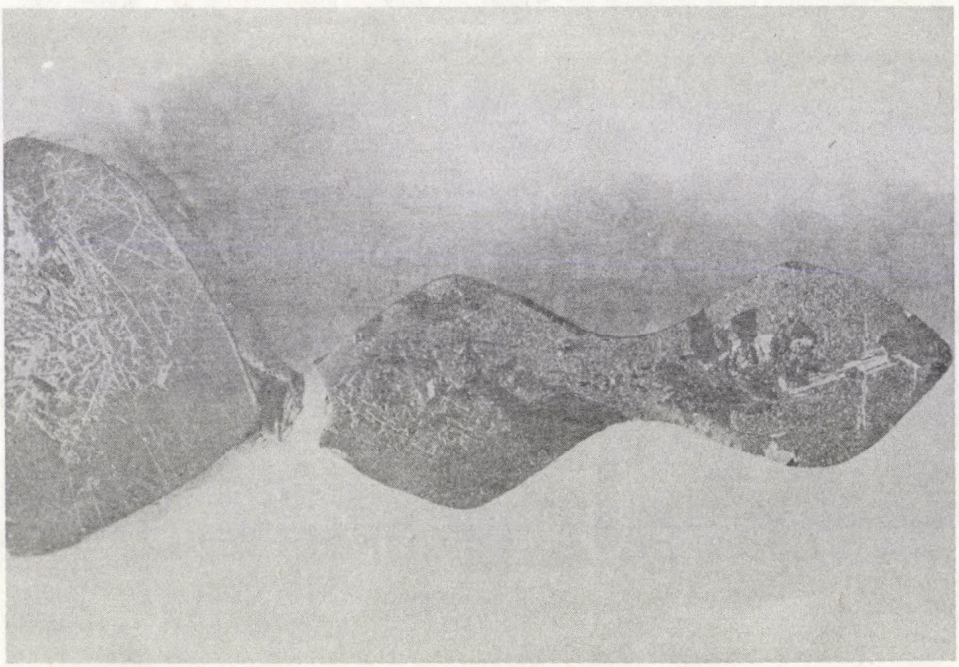


Fig. 7. Cross-section of Sample VII showing high density of grains in the bulk

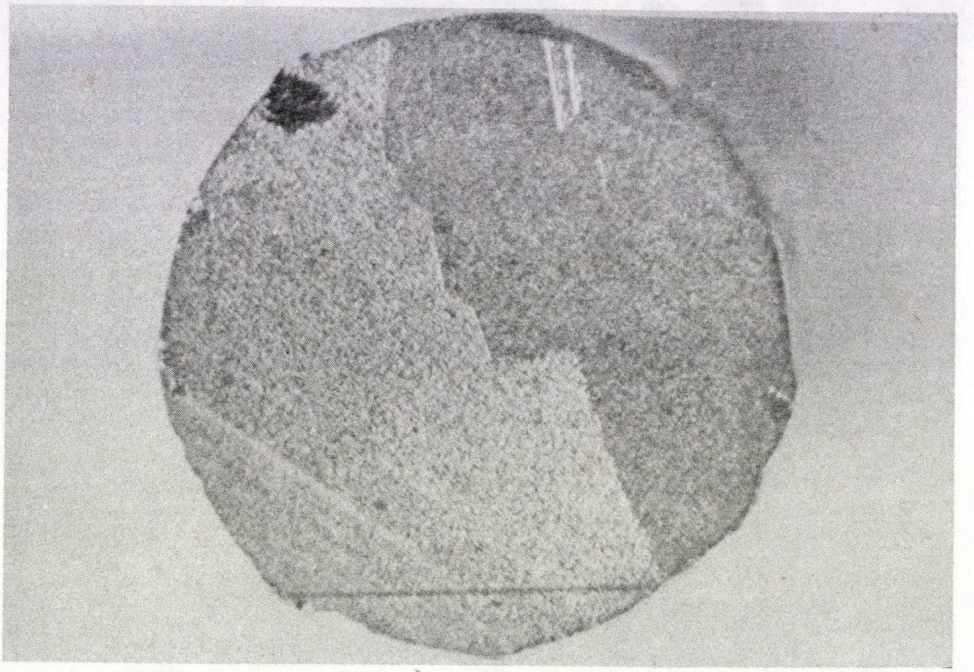


Fig. 8. The cross-section of space sample

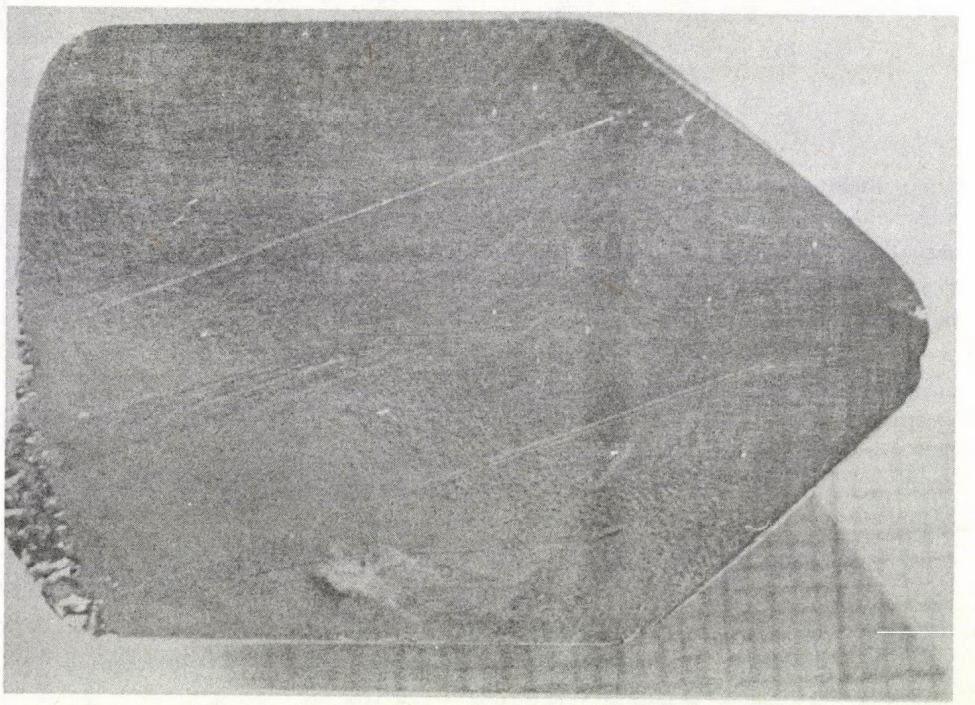


Fig. 9. Parallel twin lamellas in GaSb III

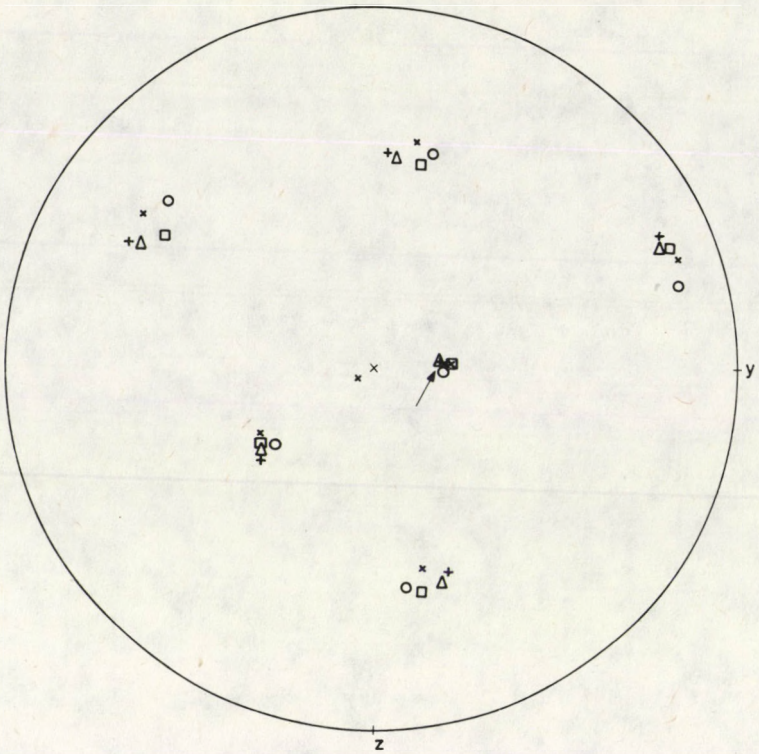


Fig. 10. Sample 1-3B. ⟨110⟩, directions of grains in the laboratory coordinate system

of only a few macroscopic crystallites, form clearly visible parallel lines (Fig. 9). According to our calculations, these lines indicate the ⟨110⟩ directions of the grains. If these ⟨110⟩ directions are placed into the stereographic projection of the laboratory coordinate system in which the centre point, i.e. the line rising perpendicularly upward from it represents the main axis of the ingot, they are grouped within a small circle near the centre. This means that the ingot has a definite texture in this direction.

In Sample VI, which contains numerous wide angle boundaries, the ⟨110⟩ directions of the grains are positioned along a circle, indicating that the individual crystallites are rotated randomly around their ⟨110⟩ axes. Those crystal sections in which only low angle boundaries can be found, show the same texture in ⟨110⟩ direction, but, in this case, their other crystallographic directions are also relatively close to each other, meaning that the crystallites are turned around the ⟨110⟩ axis by small angles (Fig. 10) only.

The deviation of textures from the ingot main axis is varied in different samples. In narrow ingots it was found about 5—7 degrees while the wider boules showed some 20 degrees deviation. More detailed texture analyses will be published later [15].

References

1. F. Capasso, M. B. Panish and S. Sumski, IEEE J. Quant. Electronics, QE-**17,** 273, 1981.
2. N. Kitamura, M. Kakehi, J. Shen and T. Wada, Jap. J. Appl. Phys., **20,** 995, 1981.
3. F. Capasso, M. B. Panish, S. Sumski and P. W. Foy, Appl. Phys. Letters, **36,** 165, 1980.
4. T. Sukegawa, T. Hiraguchi, A. Janaka and M. Hagino, Appl. Phys. Letters, **32,** 316, 1978.
5. L. M. Dolginov, L. V. Druzhinina, P. G. Elisev, A. N. Lapshin, M. G. Milvidskij and B. N. Sverdlov, Soviet J. Quantum Electr., **8,** 416, 1978.
6. H. Kano and K. Sugiyama, Electr. Letters, **16,** 146, 1980.
7. M. A. Dasevskij, G. V. Kukuladze and M. S. Mirgalovskaja, Soviet Inorg, Mater., **5,** 1141, 1969.
8. M. Kumagawa, J. Cryst. Growth, **44,** 35, 1979.
9. K. W. Benz and G. Miller, J. Cryst. Growth, **46,** 35, 1979.
10. T. Migauchi and H. Sonomura, Jap. J. Appl. Phys., **4,** 317, 1969.
11. G. Wolf, P. H. Keck and J. D. Broder, Phys. Rev., **94,** 753, 1954.
12. K. F. Hulme and J. B. Mullin, Phil. Mag., **4,** 1286, 1959.
13. E. Lendvay, M. Hársy, T. Görög, I. Pozsgai and J. Gyulai (to be published).
14. M. Hársy, T. Görög, E. Lendvay and F. Koltai, J. Cryst. Growth, **53,** 234, 1981.
15. F. Koltai, M. Hársy and E. Lendvay, (to be published).

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 143—157 (1982)

FAST COMPONENTS OF THE ELECTRIC RESPONSE SIGNAL OF BACTERIORHODOPSIN PROTEIN*

L. KESZTHELYI, P. ORMOS and G. VÁRÓ

INSTITUTE OF BIOPHYSICS, BIOLOGICAL RESEARCH CENTER, SZEGED, HUNGARY

Fast electric signals corresponding to bR→*K*, *K*→*L* and *L*→*M* transitions in the bacteriorhodopsin photocycle were measured in cases of oriented purple membranes in solution and dried samples. In the latter case the effect of external electric field was to increase (positive field) or decrease (negative field) the transition lifetimes. Based on the existing data in visible, UV, Raman spectroscopy and electric signal a model of the bacteriorhodopsin proton pump was constructed.

1. Introduction

Bacteriorhodopsin (bR) is an integral membrane protein in the plasma mebrane of *Halobacterium halobium.* Its function is to absorb light and to convey this light energy into membrane potential and a pH gradient by pumping protons from inside the cell to its exterior [1]. These membrane gradients are then consumed by various bacterial energy requiring systems such as the ATP-ase enzymes located in the plasma membrane which synthesise ATP molecules. The so-called chemiosmotic process [2] involves essentially the same phenomena in mitochondria and chloroplasts where food energy and light energy are transformed by proton pumps into membrane potentials and pH gradients. Because proton pumps are so critical in bioenergetic processes, an understanding of their molecular mechanism is very important. The simplest systems to study are the light driven bR molecules.

The bR protein has a retinal chromophore bound to a lysine side chain (either at lysine 216 or 41) which absorbs at 570 nm and undergoes a photocycle: new absorption bands appear with different lifetimes (Fig. 1). During the cycle, protons are released at the external surface and other protons are picked up on the internal side of the plasma membrane in which bR is embedded. The bR molecules are localized in approximately 500 nm diameter discs in the so called purple membranes (pm). These pm-s can be separated from the bacteria.

The kinetic characteristics of proton movement have been studied by many authors on not well defined systems [3, 4]. A new method has been developed in our laboratory for the study of proton movement [5, 6, 7]. We have demonstrated [5] that pm-s in suspension can be oriented by low electric fields (10—15 V/cm), and that after a laser flash, protons move parallel to the field and induce measurable electrical signals in an external circuit (protein electric response signal — PERS). From the analysis of the

* Dedicated to Prof. I. Tarján on his 70th birthday.

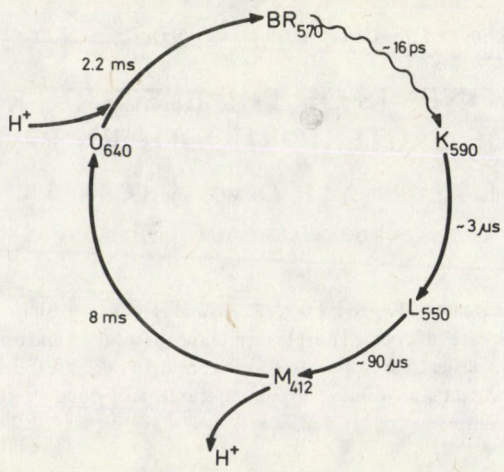


Fig. 1. The photocycle of bR. The first very fast transition to the *K*-state is light driven, the remaining ones are the consequences of thermal agitation

time course of the PERS, it has been unambiguously stated that these proton currents have the same components as the photocycle [6, 7].

Subsequently dried oriented pm samples have been produced [8] which enable us to perform other meaningful experiments. In this paper, we report the results of studies on the very fast components of the electric signal in both oriented suspension and dried samples. In the latter case an external electric field was also applied to mimic the in vivo situation: the proteins in the cell work in the very high electric field of the membrane potential ($\sim 10^5$ V/cm). The effect of membrane potential on the lifetimes of the bR photocycle in bacteria has already been characterized by Dancsházy et al [9].

2. Materials and methods

The pm-s used in the measurements were separated by the usual procedure from Halobacterium halobium strain NRL R_1M_1 [10]. The apparatus utilized to measure the signals in solution was the same as in [5] with one important exception. The Keithley 604 differential amplifier was changed to a home made much faster differential amplifier. Fig. 2 shows response of this unit to a 1 μs square pulse switched in series with the pm solution. It can be seen that the time resolution of the amplifier is 0.2—0.3 μs if the signal is taken from a 10 kΩ resistance.

The preparation of dried, oriented samples was described in [8]. Minor modifications of the technique were necessary to enable us to conduct fast timings: the transparent, conducting SnO_2 layer had a resistance of $\sim 50\,\Omega$, and an A1 layer was evaporated onto the dried pm layers. An external voltage of ~ 140 V was applied to the SnO_2 and A1 electrodes when needed. Because ~ 2000 pm layers were in the sample,

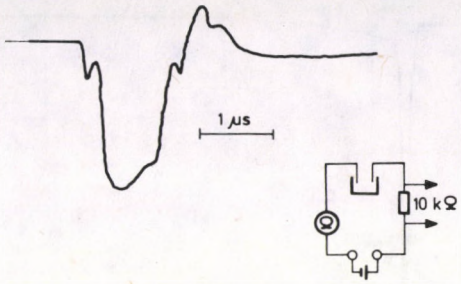


Fig. 2. Amplifier response to 1 μs square voltage switched in series with the pm solution (insert). Time resolution 0.1 μs

~70 mV potential was on a pm. The electric signals were registered at a resistance of ~1 kΩ switched parallel to the layer. Opton dye laser was used for excitation at a wavelength of 590 nm, a duration of 1 μs, and and energy ~10 mJ.

3. Results

a. pm-s oriented in suspension

The improved resolution of the electric circuit has made it possible to resolve the first negative component of the protein electric response signal (PERS) reported in [6] into the two components seen if Figs 3a, b. The temperature of the solution was maintained at 5 °C to obtain a clear discrimination between the first fast negative component and a second negative component having the lifetime of the $K \rightarrow L$ transition. The latter has been checked by observing the K-decay in light absorption measurement. Fig. 4 shows the fast component in an extended time scale. The signal increases during the 1 μs laser pulse and falls with the time constant of the electronics (~0.2—0.3 μs) after the light pulse is terminated.

The positive component in Figs. 3a, b corresponds to the $L \rightarrow M$ transition in accord with the previous assignment [6].

b. pm-s in dried oriented samples

The fast components of the PERS for dried oriented sample are very similar to the components in suspension. Fig. 5 shows the signal when no external field is applied to the sample. A large negative signal and a small positive signal appear, the latter containing two components. The time constants of these components differ from the corresponding time constants of bR-s in suspension. The same time constants appear in light absorption measurements, too.

In Fig. 5b a negative signal with very small amplitude and long lifetime is seen. Its area is nearly equal to the area of the positive signal in Fig. 5a. The signal does not have

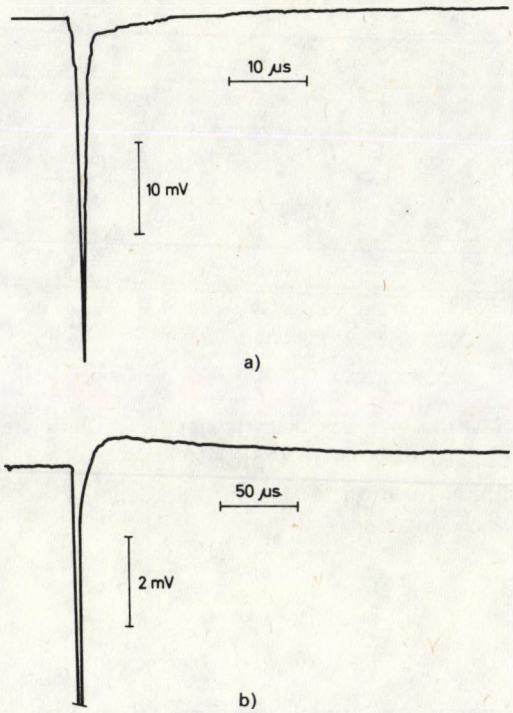


Fig. 3. The fast components of the PERS with different amplifications a, b. Orienting field $E = 15$ V/cm, absorbance of the pm suspension $A = 0.8$, temperature $T = 5$ °C. Recording from one flash

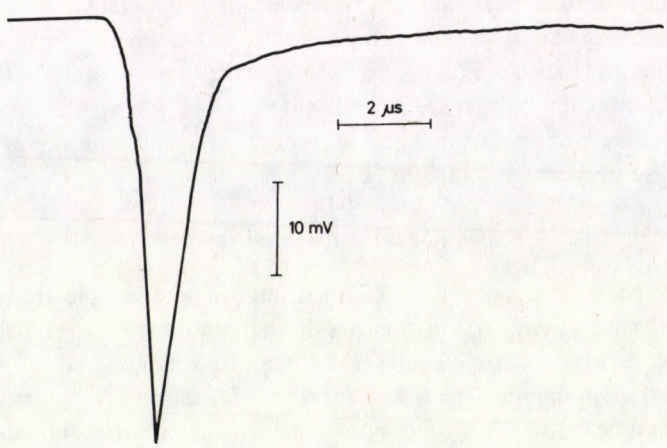


Fig. 4. The fast components of the PERS with extended time scale. The same record as in Fig. 3

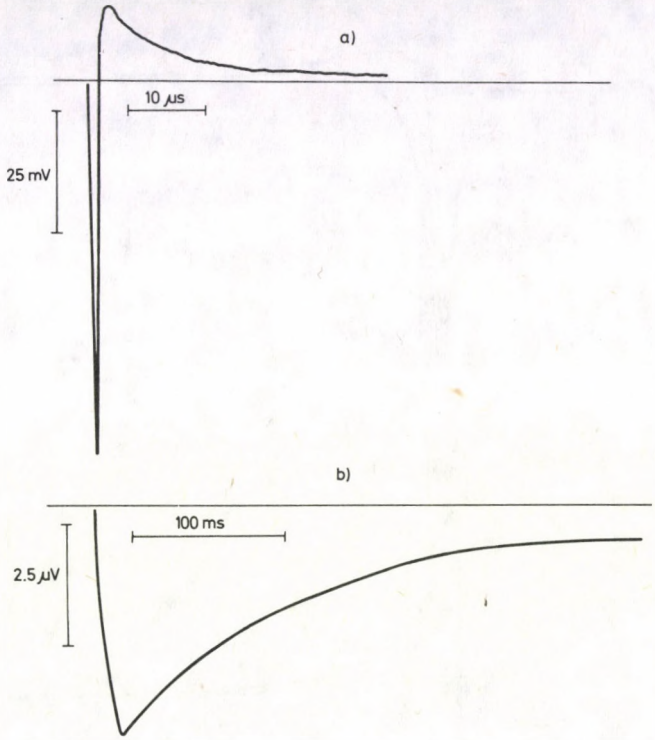


Fig. 5. The fast a) and slow b) components of the PERS in dried oriented sample. Signal taken from 1 kΩ resistance switched parallel to the sample

an analogue in the photocycle. This PERS can be understood as a backflow of protons from *M* state.

More details of the fast component can be observed in Figs 6a, b. In this series of measurements, data were produced in an externally applied electric field in the positive direction (+) (i.e. in the appropriate in vivo direction of proton flow) and in opposite direction (−). For comparison, we also measured the zero field situation. The electric field influences the PERS components: the first negative component has a larger amplitude in positive (+) and a smaller one in negative applied field (−) than the zero field signal (0). The peak positions and the decay times were also different. The decay time differences appear also in the positive components: they are longer in a (+) field and shorter in a (−) field than in the field free case. The data are summarized in Table I.

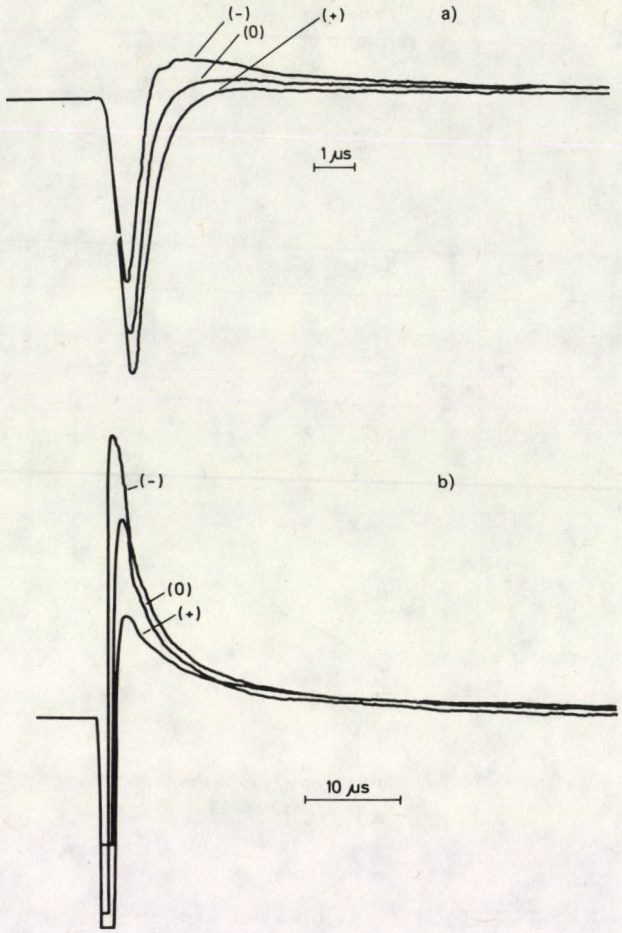


Fig. 6. The fast components of the PERS in external electric field. a, b, different amplifications and time scales. The voltage on ~2000 dried oriented layers was ±140 V. Signals taken from 1 kΩ resistance. (+), (−) and (0) signs refer to the field directions (see text) and to zero field

Table I

Amplitudes and time constants of the fast PERS components depending on electric field ((+), (−), (0) means field directions against and parallel to proton movement and zero field). A_1 amplitude, τ_1 decay time of the negative component, A'_3, A''_3, τ'_3, τ''_3 the same for the positive components. A-s in relative units, τ-s in μs. Area of the positive PERS components is $A'_3\tau'_3 + A''_3\tau''_3$ in relative units.

Field	A_1	τ_1	A'_3	τ'_3	A''_3	τ''_3	Area
+	90	0.62	2.2	5	1.35	39	64
0	78	0.54	4.5	4	1.3	31	60
−	62	0.45	6	3.1	1.9	26	68

4. Discussion

a. The protein electric response signal (PERS)

The PERS components were previously established [6] as the current which is induced in the external circuit by protons moving during the bR photocycle. More specifically, it is assumed that the proton at the protonated Schiff base of the lysine and retinal bond jumps to the external medium in two steps ($L \rightarrow M$, $M \rightarrow 0$ transitions) and that the missing proton from the Schiff base is then replaced from the internal side ($0 \rightarrow$ bR transition). The negative PERS component has been assigned the to $K \rightarrow L$ transition as a reversed jump of the proton. This component has now been resolved into two components offering the possibility of a finer understanding of the first steps in the proton pumping mechanism of bR. For this reason a short recapitulation and extension of the theoretical basis for PERS originally given in [6] is needed. (A rather detailed theory of PERS is in preparation [11]).

The proton movement in bR causes induced current in the external circuit. If a charge Q moves with a velocity v between electrodes over a distance D then

$$i = \frac{Q}{D} v \tag{1}$$

is the value of the displacement current.

We assume that the velocity v is very large (since v has really no meaning for protons jumping from one position to another) and calculate the induced charge by integrating Eq. (1) for time:

$$Q_{\text{ind}} = \frac{Q}{D} \int_0^\infty v \, dt = \frac{Qd}{D}, \tag{2}$$

where d is the displacement of the charge. The measuring circuit is an RC circuit, and if a single Q_{ind} arrives, then a time course of voltage

$$V(t)_{(1)} = \frac{Qd}{D} \cdot \frac{1}{C} e^{-\frac{t}{RC}} \tag{3}$$

could be measured. In a laser flash many bR-s are excited, therefore many protons move (N). The number of moving protons per time unit ($\rho(N)$) in the case of an exponential decay is

$$\rho(N) = Nke^{-kt}, \tag{4}$$

where k is the rate constant. To get the wave form for N particles, Eqs. (3) and (4) must be folded:

$$V(t)_N = \frac{NQdk}{DC} \int_0^t e^{-kt'} e^{-\frac{t-t'}{RC}} \, dt' = \frac{NQdk}{D} \cdot \frac{R}{1-kRC} [e^{-kt} - e^{-\frac{t}{RC}}]. \tag{5}$$

Two limiting cases are important:

$$V(t)_N = \frac{NQdkR}{D} \cdot e^{-kt}, \quad \text{if} \quad k \ll \frac{1}{RC}, \tag{6a}$$

$$V(t)_N = \frac{NQd}{D} \frac{1}{C} e^{-\frac{t}{RC}}, \quad \text{if} \quad k \gg \frac{1}{RC}. \tag{6b}$$

In the above derivation it was tacitly assumed that the pm-s are in a medium of high resistance. This is the case in dried, oriented samples where the external R and C determine the circuit time constant. An $RC \sim 10^{-7}$ s can be estimated. In the case of the suspension R is the resistance and C is the capacitance of the electrolyte between the two sides of a pm. It can be estimated that $RC \sim 10^{-9}$ s. A realistic treatment must consider the data for pm-s, the R and C values of the solution and the measuring resistance [11]. For the present case, it is enough that the realistic treatment has the same time dependence as given in Eqs. (6a, b) and only the R and C values are somewhat modified. It is a good approximation that the RC of the external circuit (RC of the electronics, being $\simeq 0.2$—0.3 μs) is determining the measurements in suspension.

The lifetime of the bR→K transition is 16 ps (Fig. 1). The first negative PERS component should correspond to this transition because the second negative component in Fig. 3 has a lifetime of about 7 μs which corresponds to the $K \rightarrow L$ transition at 5 °C. Naturally we do not know without appropriate measurements that this charge movement also happens within 16 ps. Nevertheless, we may accept that this movement is much faster than the time constant RC, therefore, for the very fast component Eq. (6b), and for the following ones, Eq. (6a) may be applied.

The length of the laser flash is ~ 1 μs in our case and this is reflected in the ~ 1 μs continuous growth of the fast negative signal in Fig. 4. The folding of the moving charge per time unit for bR→K transition $\rho(n_K)$ and Eq. (3) is needed.

We assume, according to [12], that an S state appears after the photon-absorption which decays by the rate constant k_S to the K state.

Therefore

$$\begin{aligned} \dot{n}_{bR} &= -\sigma\phi n_{bR}, \\ \dot{n}_S &= +\sigma\phi n_{bR} - k_S n_S, \\ \dot{n}_K &= k_S n_S. \end{aligned} \tag{7}$$

Here n_{bR}, n_S, n_K are the normalized numbers of states, σ is the cross section to produce the state S, ϕ is the photon flux. Solving Eq. (7)

$$\dot{n}_K = \rho(n_K) = \frac{k_S \sigma\phi}{\sigma\phi - k_S} (e^{-k_S t} - e^{-\sigma\phi t}). \tag{8}$$

From the folding

$$V_N(t) = \frac{NQd}{DC}\frac{k_S}{\sigma\phi - k_S}\left\{\frac{RC}{1-k_S RC}(e^{-k_S t} - e^{-\frac{t}{RC}}) - \frac{RC}{1-\sigma\phi RC}(e^{-\sigma\phi t} - e^{-\frac{t}{RC}})\right\}. \quad (9)$$

If $k_S \gg \frac{1}{RC}$ and $\sigma\phi \sim \frac{1}{RC}$, then

$$V_N(t) = \frac{NQd}{D}\frac{\sigma\phi R}{1-\sigma\phi RC}(e^{-\sigma\phi t} - e^{-\frac{t}{RC}}). \quad (10)$$

Eq. (10) is valid for 1 μs, after which the signal falls by the time constant RC. $V_N(t)$ has a flat maximum at $\sim 2\tau_{\min}$, where $\tau_{\min}$ is the smallest of time constants $\frac{1}{\sigma\phi}$ and RC. For $RC \simeq 0.2-0.3$ μs, and $\sigma\phi \simeq 1$, Eq. (10) describes Fig. 4 satisfactorily.

Eq. (10) and Fig. 4 could also be used to estimate d, the displacement of the charge in the bR$\rightarrow K$ transition. It is assumed that a single charge moves and excites the negative signal. The PERS of the $L\rightarrow M$ transition was also measured (Fig. 3) and the displacement for this transition is known to be $d_{L\rightarrow M} = 0.5$ nm from [6]. Using this value as a normalization for $\frac{NQ \cdot R}{D}$ (in the calculations Eq. (10) was applied with $RC = 0.3$ μs, $\sigma\phi = 1$ μs) $d_{\mathrm{bR}\rightarrow K} \simeq 0.16$ nm. It is emphasized that this value is an estimated value only.

The assignment of the positive components of PERS in the dried, oriented samples is simple: they correspond to the $L\rightarrow M$ transition. The existence of two lifetimes (Table I) seems to reflect a branching of K decay to two different L states. Both τ_3', and τ_3'' are smaller in dried samples than the τ of the $L\rightarrow M$ transition in suspension. The external field accelerates (−) and decelerates (+) the transition.

Eq. (6a) can be used for the evaluation of the signal because $k_3' = \frac{1}{\tau_3'}$ and $k_3'' = \frac{1}{\tau_3''}$ are smaller than $\frac{1}{RC}$. The amplitude is proportional to the rate constant but the area of the signal must be constant if the number of excited bR-s N and the displacements d do not change. Table I contains the area of the positive signals for the different external fields as calculated from amplitudes and time constants.

The results show that the external fields change only the transition times and not the area proportional to Nd. We assume that d is given by the structure of the molecule which is not influenced by the electric field therefore the number of excited states N or with other words the cross section for bR$\rightarrow K$ transition σ is not influenced by the external fields (the constancy is within ±8%). This would mean a constant amplitude of the negative signal according to Eq. (10) because even RC of the circuitry is

Table II
Distances determined from PERS
(data in nm)

Transition	d [nm]
1. br→K	−0.16
2. K→L	−0.05
3. L→M	+0.5
4. M→0	3.0
5. 0→bR	1.5

independent of the external field. The values of A_1 in Table I, however, contradict the above requirements.

A closer examination of the curves in Fig. 6a and Fig. 4, however, resolves the problem. It is clear that the transition from negative to positive peak needs time. This is the time constant of the $K \rightarrow L$ transition. In Fig. 3 and $4\tau_{K\rightarrow L} = 7$ μs is much longer than the 1 μs laser flash, therefore it does not disturb the negative peak. In dry samples the time constants are much shorter, $\tau_{K\rightarrow L}$ is $\simeq 0.5$ μs, which can be seen in Table I. The recharging already occurs during the flash therefore Eq. (10) should be modified by adding an exponential with opposite sign. The electric field influences $\tau_{K\rightarrow L}$ also in qualitatively the same way as τ_3' and τ_3'' (Table I). The amplitude of an exponential is proportional to $k = \frac{1}{\tau}$ (Eq. 6a) therefore a larger amplitude should be subtracted for field (−) than for field (+). The data A_1 and τ_1 in Table I agree qualitatively with this explanation. τ_1 in Table I is the resultant of $\tau_{K\rightarrow L}$, RC and $\sigma\phi$. To determine $\tau_{K\rightarrow L}$ requires a better time resolution by the electronics and a shorter laser pulse. For the present, we may state that the electric field changes the time constants of the transitions $K \rightarrow L$, $L \rightarrow M$ in a similar way and does not change the cross section σ. The time constant of bR→K may also be changed, but this is beyond the limit of our time resolution.

b. A model of bR proton pump

The data extracted from the measurement of PERS [6, 7 and present work] indicate charge movement which occurs in 5 steps in the protein. The distances were determined by assuming that the PERS-s are the consequences of proton movement and not the movement of some charged amino acid side chains. In Table II we recapitulate the data for these distances including the presently determined distance for the bR→K transition.

One can see that the distances represent two distinct groups: (1) small ones (<0.5 nm, transitions 1., 2., and 3.,) and (2) large ones (4., 5.,). It is generally accepted that the protons originate from the protonated Schiff base which is at a distance of

1.4—1.5 nm from the internal surface of the membrane. Therefore the movements in group (1) must occur deep in the protein. We call these movements internal movements. In transitions 4. and 5., the protons are released to and enter from the environments, they may be called external movements. In dried samples, the internal time constants become shorter and the external constants longer. These facts support our grouping of the transitions.

We can speculate about the nature of the proton donating and accepting sites. To do this we must consider other data in addition to the optical data of the photocycle in visible and PERS-data:

— Laser Raman spectroscopy indicates that the Schiff base is deprotonated during $L \rightarrow M$ decay, and it is reprotonated when bR ground state is reestablished [1].

— UV absorption changes show that both a tyrosine and a tryptophan side chain participate in the proton translocation function of bR [13, 14, 15]. More specifically, a tryptophan is disturbed by a proton approaching it [15]. The rise time of this event coincides with the $K \rightarrow L$ transition [14], the decay time with the $M \rightarrow 0$ transition or the $M(\rightarrow 0) \rightarrow$ bR transition, but we cannot decide which from the existing data [14, 15]. We assume it happens after the M decay. A tyrosine becomes deprotonated with rise time of the $L \rightarrow M$ transition [14] and reprotonated with the same timing as the tryptophan.

The events in the five transitions may be described by the following model (Fig. 7):

(1) In the bR ground state the retinal is bound by the protonated Schiff base lysine 216 (or 41). The N^+ is compensated for by the negative charge of a close residue (A_1^-) which could belong to an aspartic or glutamic acid. The distance between H^+ and A_1^- is ~ 0.2 nm. Near the Schiff base is a TRP and also a TYR residue. For the following discussion we assume the existence of another residue, A_2^-, which is protonated in bR ground state A_2H. The situation is depicted in Fig. 7 (1).

The ground state retinal is in an all-trans configuration, and this is transformed into the 13-cis configuration by photon absorption. The trans-cis isomerization involves an inward movement (Fig. 7 (1) (2)) of the H^+N group with the salt bridge between H^+N and A_1^- being broken and the new position is stabilized [16, 17]. This means that the energy barrier for cis-trans transition is high. From the geometry it can be calculated that the inward displacement of the H^+ $d_{1\text{geom}||}$ is 0.16 nm, and the lateral one is $d_{1\text{geom}\perp}$ 0.2 nm. The fast negative portion of the PERS corresponds to the inward motion of the proton. The estimated value of charge displacement $d_1 \sim 0.16$ nm is in good agreement with $d_{1\text{geom}||}$. A large lateral movement is expected in accord with Fig. 7 (1). We have H^+N and A_1^- in a distance of $\simeq 0.5$ nm. The energy transformed into electrical energy and stored is $\simeq 0.2$ aJ ($\sim 60\%$ of the energy of absorbed photon).

(2) In the $K \rightarrow L$ transition the proton from A_2H moves a small distance d_2 inward and disturbs the TRP residue (Fig. 7 (2)). After this transition, in the L state there are four uncompensated charges in the molecule, A_1^-, A_2^-, $TRPH^+$ and H^+N.

(3) We have to assume that the $L \rightarrow M$ transition has two proton movements: both the Schiff base and a TYR are deprotonated simultaneously. We postulate that

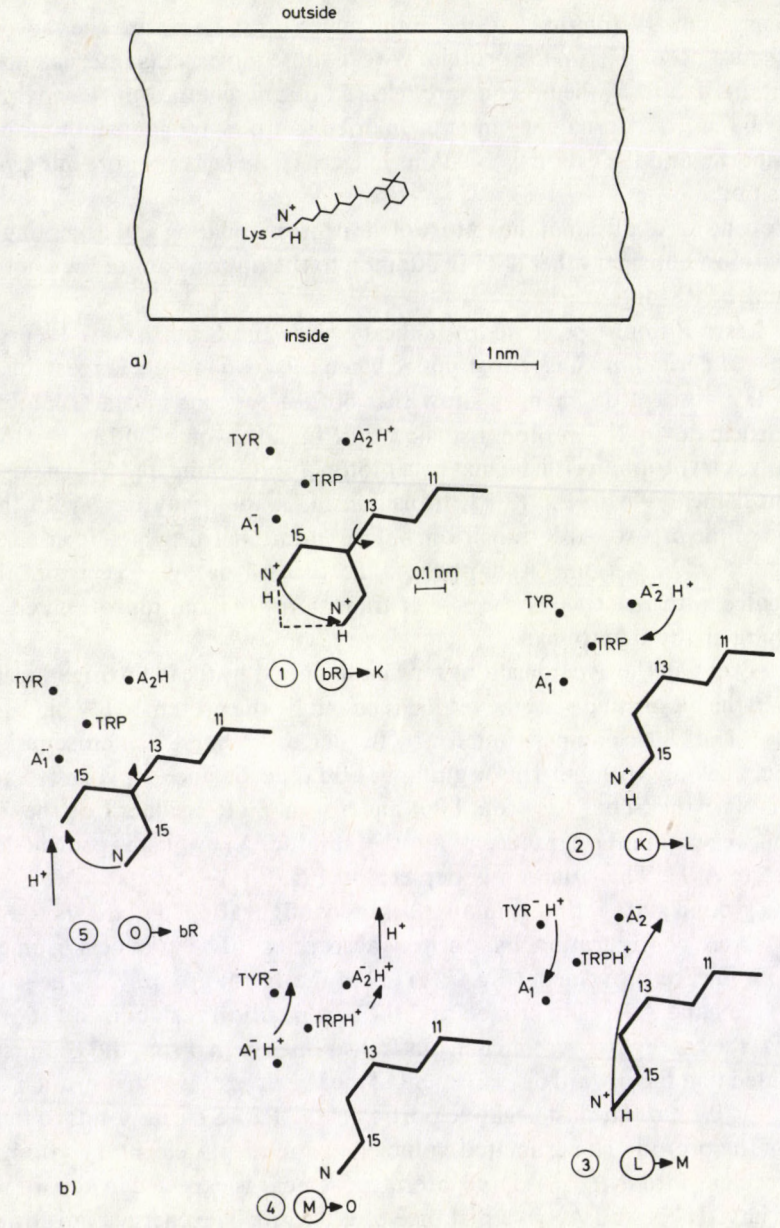


Fig. 7. Geometrical model of the bacteriorhodopsin proton pump. a. The location of the retinal in the bR molecule. b. The magnified (5 times) neighbourhood of the Schiff base (1), (2), (3), (4) and (5) represent the bR ground state, *K*, *L*, *M* and *O*-states. Arrows show the assumed transitions from the given states

the Schiff base proton moves to the A_2^- while the TYR proton moves to the A_1^- forming A_2H, A_1H and TYR^-. The algebraic sum of the two distances $d_3 = d_{3\text{Schiff}} + d_{3\text{TYR}} = 0.5$ nm. We think that the $d_{3(\text{Schiff})}$ is the dominant shift and it is directed outward (i.e. it is positive) and that $d_{3(\text{TYR})}$ is smaller and negative. The M state is characterized then by a $TRPH^+$ and a TYR^-. Note that the original dipole NH^+ and A_1^- in K state (1) is shifted outwards.

(4) The M decay is the most important step during proton pumping and also the most complicated in the present model. Three proton movements must happen during this rearrangement.

The dipole field of $TRPH^+$ and TYR^- expels the proton from the A_2H barrier to the external surface. This displacement is $d_4 \sim 3$ nm. The assumed small proton movements of $TRPH^+ \rightarrow TRP + A_2H$ and $A_1H \rightarrow TYR + A_1^-$ are negligible in comparison with d_4.

The small internal proton movements can probably be accounted for as jumps between neighbouring positions. The 3 nm external movement, however, requires further clarification. The long lifetime of the M state $\tau_M \simeq 10$ ms indicates the low probability with which the proton will leave the A_2^- barrier. Once the proton has overcome the $A_2H - A_2^-$ barrier its pathway is a straightforward move to the surface using the so-called proton wire which has been substantiated as a network of hydrogen bonded side chains [18, 19, 20, 21]. The proton current is conducted similar to the current which occurs in ice. An upper time limit estimate for this conduction is ~ 1 μs, which is negligible when compared with τ_M.

We now suggest another mechanism for proton conductance which does not require the specific structure of hydrogen bounded chains. It is well known that fluctuations in protein structure can make any part of a protein accessible to water molecules [22, 23]. This is a rather fast process (in some cases it is in the ms range) as indicated from measurement of H – D exchange of peptide hydrogens in proteins [24]. The H can be exchanged only when a D_2O molecule is in very close proximity.

We assume that the A_2H can only be deprotonated when an H_2O appears and binds the proton forming a hydroxonium ion (H_3^+O), which is then expelled by the electrostatic forces of the existing dipole. τ_M then reflects both the time necessary for an H_2O to appear and the height of the barrier for protons to leave A_2^-. After proton exchange the time for the H_3^+O to move to the surface is short [25]. This type of proton conducting mechanism can explain the tremendous increase (100—1000 times) of the M lifetime in the dried samples and the missing proton extrusion (Fig. 5). This explanation is in line with the current concept of protein fluctuations [26].

In the O state, one proton is outside at the external surface of the bR molecule and one charged group A_1^- remains inside.

(5) The $O \rightarrow$ bR transition is characterized by a cis-trans isomerization of the retinal. The N group moves back to its original position and is protonated from inside the membrane. This protonation is also by external interaction and occurs by the same mechanism outlined in (4). The height of the barrier to this step is determined by the

barrier of the cis-trans isomerization and the availability of a proton conducted as H_3^+O.

The model outlined above explains all the known facts of the bR photocycle available from visual, UV, Laser Raman spectroscopy and the measured PERS-s. The effects of the electric fields on the lifetimes measured by Dancsházy et al [9] and in this work (Figs 5, 6) are easily understood: the barriers for proton jumps are increased (+) or decreased (−).

The back photeraction, i.e. the termination of proton pumping and the reappearance of bR ground state from M state under the effect of blue light which is connected with spectral [27] and PERS changes [7] is accommodated by the model: In M state, the retinal absorbs a photon and undergoes a rapid cis-trans isomerization. Note that the Schiff base is neutral in M state. This induces the release of a proton from $A_1^-H^+$ to TYR^- causing a small fast positive electric signal. This has actually been observed in [7]. The next step is the reprotonation of the Schiff base from A_2H leaving A_2^- to be reprotonated from $TRPH^+$. If these steps are separated two additional components can exist in visual light spectroscopy as observed in [14].

The retinal in *M* state may have a spontaneous cis-trans transition with a very long lifetime of ~ 100 ms. If the *M* state is unable to decay in the dry sample because of the absence of H_2O then the back reaction which is initiated by spontaneous cis-trans transition shortcuts the proton pump. This explains the slow negative signal observed in Fig. 5b.

An inherent difficulty in this model is the assumption of multiple events in some of the transitions (2 in (3), 3 in (4) and 2 in (5)). This could, however, actually be advantageous in explaining the existence of two time constants measured for the *L*-decay here, and also reported by many authors, e.g. [14]. These brachings are reported to exist under unusual circumstances such as dryness, low temperature, non physiological pH, modified bR, ...).

There are a growing number of papers which indicate that two protons are probably pumped per photocycle (for a review see [29]). Our model has enough flexibility to accommodate this requirement by making small modifications: one proton may transfer via the Schiff base and A_2^-, while the other proton is transported via the TYR^- side chain and A_1^-.

This model is largely dependent on structural informations. It requires the existence of a TYR, TRP and two carboxyl groups (A_1^-, A_2^- from aspartic our glutamic acid) at the appropriate distances from the Schiff base.

One should find the characteristic absorption changes for $A_1^- + H^+$, $A_2^- + H^+$ with the forecasted time constants in the short wavelength UV region. The lateral motions of these charges could confirm the structural assignments. The easiest distance to measure seems to be $d_{1\,\mathrm{geom}\perp}$ which must exist in a given relationship to $d_{1\,\mathrm{geom}\|}$ which has already been measured.

The time of appearance of the pumped proton in the solution is crucial. It should happen in the *M* decay.

Variations in photocycle due to specific modifications of bR side chains could also help in critical checks of the model.

We may consider the velocity of proton movements in the external transitions ($M \rightarrow O$, $O \rightarrow$ bR) as a possibility to decide between the different proton wire models (i.e. hydrogen bonded chains or H_2O channel produced by protein fluctuation). In Eq. (2) the velocity was assumed to be very large for all transitions. Final velocity, however, can be included in Eq. (3) and finally in Eq. (5) [11] offering a means for experimental control.

Discussions related to the proposed model with Prof. W. Stoeckenius are highly acknowledged.

References

1. W. Stoeckenius, R. H. Lozier and R. A. Bogomolni, Biochim. Biophys. Acta, **505,** 215, 1979.
2. P. Mitchell, Nature, **191,** 144, 1961.
3. L. A. Drachev, A. D. Kaulen and V. P. Skulachev, FEBS Lett., **87,** 161, 1978.
4. F. T. Hong and M. Montal, Biophys. J., **25,** 465, 1979.
5. L. Keszthelyi, Biochim. Biophys. Acta, **598,** 429, 1980.
6. L. Keszthelyi and P. Ormos, FEBS Lett., **109,** 189, 1980.
7. P. Ormos, Zs. Dancsházy and L. Keszthelyi, Biophys. J., **31,** 207, 1980.
8. Gy. Váró, Acta Biologica Hung., **32,** 301, 1981.
9. Zs. Dancsházy, S. Helgerson and W. Stoeckenius, to be published.
10. D. Oesterhelt and W. Stoeckenius, Methods in Enzymology, **31,** 667, 1974.
11. P. Ormos and L. Keszthelyi, to be published.
12. M. L. Applebury, K. Peters and P. Rentzepis, Biophys. J., **23,** 375, 1978.
13. R. A. Bogomolni, in Bioelectrochemistry ed. by H. Keyzer and F. Guttman, Plenum Publishing Corp. p. 83., 1980.
14. B. Hess and P. Kuschmitz, FEBS Lett., **100,** 334, 1979.
15. J. Czégé, A. Dér, L. Zimányi and L. Keszthelyi, Proc. Natl. Acad. Sci. **79,** 7273, 1982.
16. B. Honig, T. Ebrey, R. H. Callender, U. Dinur and M. Ottolenghi, PNAS, **76,** 2503, 1979.
17. A. Warshel, Photochem. Photobiol. **30,** 285, 1979.
18. J. F. Nagle, M. Mille and H. J. Morowitz, J. Chem. Phys., **72,** 3959, 1979.
19. J. F. Nagle and M. Mille, J. Chem. Phys., **74,** 1367, 1981.
20. W. Stoeckenius in Membrane Transduction Mechanisms, ed. by R. A. Cone and J. E. Dowling, Raven, New York, p. 39, 1979.
21. H. Merz and G. Zundel, Biochem. Biophys. Res. Comm., **101,** 540, 1981.
22. F. M. Richards, Ann. Rev. Biophys. Bioenerg., **6,** 151, 1977.
23. F. M. Richards, Abstracts of the VII. International Biophys. Congress, Mexico City, p. 14, 1981.
24. J. Rose, J. Mol. Biol., **113,** 143, 1979.
25. P. Läuger, Biochim, Biophys. Acta, **552,** 143, 1979.
26. H. Frauenfelder, G. A. Petsko and D. Tsernoglou, Nature, **280,** 558, 1979.
27. O. Kalinsky, U. Lachish and M. Ottolenghi, Photochem. Photobiol., **28,** 261, 1978.
28. F. F. Litvin and S. P. Balashov, Biofizika, **22,** 1111, 1977.
29. R. A. Bogomolni, R. A. Baker, R. H. Lozier and W. Stoeckenius, Biochemistry, **19,** 2152, 1980.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 159—164 (1982)

INVESTIGATION OF CHANGE OF MINERAL METABOLISM OF COSMONAUTS BY X-RAY FLUORESCENCE METHOD*

J. BACSÓ, M. KIS-VARGA, P. KOVÁCS, J. PÁLVÖLGYI and D. BERÉNYI

INSTITUTE OF NUCLEAR RESEARCH OF THE HUNGARIAN ACADEMY OF SCIENCES
H-4001 DEBRECEN, HUNGARY

J. HIDEG

MEDICAL SERVICE OF THE HUNGARIAN PEOPLE'S ARMY, BUDAPEST, HUNGARY

R. A. TIGRANJAN and T. A. VITING

INSTITUTE FOR MEDICAL BIOLOGICAL PROBLEMS OF THE MINISTRY OF HEALTH, MOSCOW, USSR

Micro-element determination was carried out in the blood serum and in the hair samples of the first Soviet—Hungarian space team, by using the techniques of X-ray fluorescence analysis. The concentration of Cl, K, Ca, Br, Cu and Zn was determined before and after the space flight.

Introduction

Blood circulation transports nutritive materials to the tissues and organs of the human body. This transport for nutritive materials and for the end products of metabolism has evolved under terrestrial conditions (in the presence of the force of gravity). Thus, its undisturbed operation is guaranteed under earthly conditions.

Under the conditions of weightlessness the circulation resistance of the circulatory system of cosmonauts changes because of the absence of gravity, and this results in a change in the blood circulation. There is a stagnation in the lower extremities, at the same time an enchanced circulation in head and trunk. This affects the household of water and mineral materials of the organism, i.e. the metabolism of the human organism [1—3]. Investigations into the modification of metabolism as well as its knowledge in order to prevent and to avoid its harmful effects are very important from the point of view of space research.

Really effective studies on the metabolic processes would need the continuous registration of the changes but so far there is no suitable instrument; however, the human organism offers some possibility.

Human hair is a special end product of metabolism. The constituents of hair are determined during its development by metabolism. However, after the development (a

* Dedicated to Prof. I. Tarján on his 70th birthday.

few days) [4], the internal composition does not change. The speed of the growth of human hair is 0.3—0.5 mm/day. In this way the human hair stores the changes of metabolism taking place in the immediate environment of the hair follicle in chronological order. The measurement of macro- and micro-element concentration in hair, supposing that it was not affected by some external contamination, is suitable for investigating certain changes in human metabolism. The advantage of this method is that sample collection is painless, the samples do not decompose, thus storage does not require special precautions. The results obtained in this way can be compared with a similar analysis of blood serum.

The micro-element determination carried out on the blood serum and on hair samples of the two cosmonauts of the first Soviet—Hungarian space team are given in the present paper.

Sample preparation, measurement

Four blood serum samples were taken from both cosmonauts. Samples were taken on 21 April, 14 May, 4 June and 9 June 1980. The space mission lasted from 26 May to 3 June. The samples (approx. 0.5 ml each) were received in plastic test tubes sealed with paraffin. They were stored in a refrigerator (at $-20\,^{\circ}C$) before they were delivered to us. Other special treatment is not known. The samples were dried (0.1 ml) into spots of about 1 cm diameter. Two parallel preparations were made of every sample thereby enabling parallel measurements.

Measurements of the hair and beard samples were carried out in the natural, original form of the samples, i.e. without washing them before measurement.

The samples were analysed by X-ray fluorescence method. The schematic arrangement of the measuring set-up can be seen in Fig. 1. The energy resolution (FWHM) of the instrument is 190 eV. The Cl, K and Ca concentration of the samples

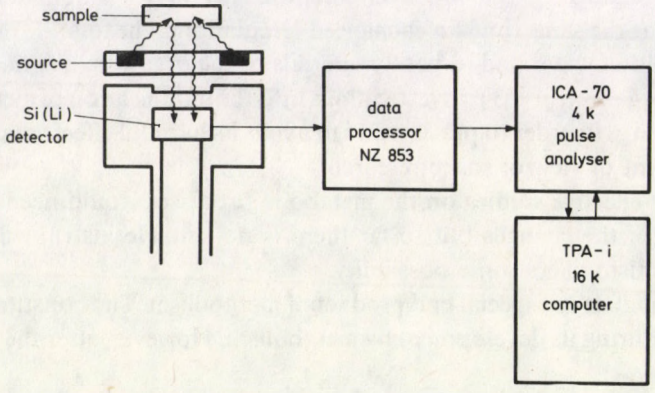


Fig. 1. Set-up of the instrumentation used for the measurements

was measured with Fe-55 (750 MBq) source the Cu, Zn and Br concentration was measured with I-125 (370 MBq) exciting source. The evaluation of X-ray spectra was made by a TPAi computer.

Results

The I-125 source induced X-ray spectra of the serum samples of the two cosmonauts before and after the flight can be seen in Fig. 2. The dates (duration) of the flight are also given in the Figure. The Cu, Zn and Br values were determined from the measurement with I-125 (the Fe line in the spectrum comes from the instrument).

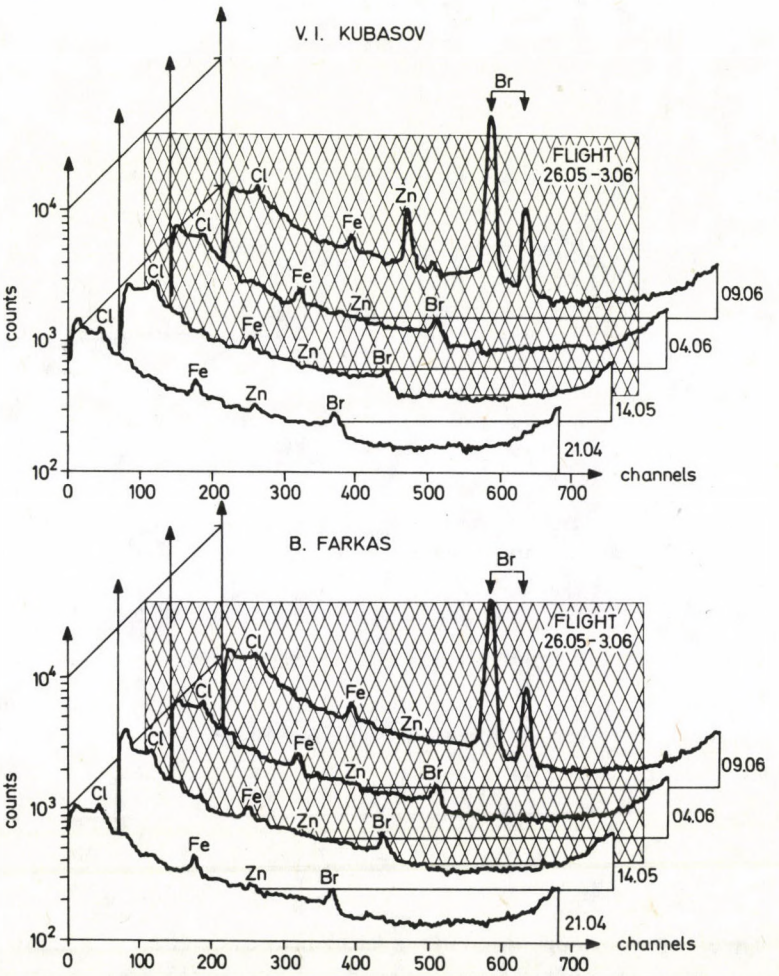


Fig. 2. Spectra of blood serum samples

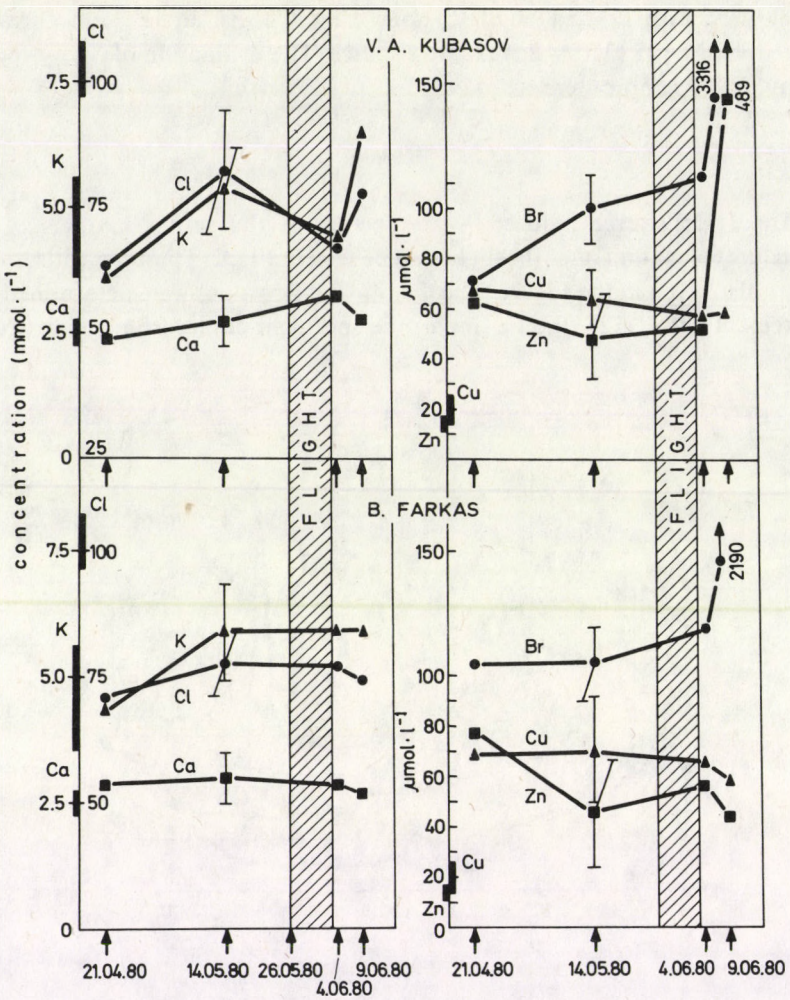


Fig. 3. Summary of results for blood serum samples

The concentrations of mineral elements measured in the serum samples of the two cosmonauts are given as a function of time (see Fig. 3). The dates (duration) are also indicated in this Figure; the normal values of certain elements are shown on the vertical axis [5].

Discussion

The K and Ca values found in the serum before and after space flight correspond to the normal values and remain unchanged even if taking into account the error limits (Fig. 3). However, the measured Cl content does not reach the normal value, it is about

65—80% of the normal value and its level did not change before and after the space flight.

The measured Zn and Cu levels are many times higher than the normal value [5] in the serum of both cosmonauts. This is presumably due to the fact that the detection limit with 0.1 ml serum is just above the upper limit of the normal value and small Zn and Cu contaminations are present in parts of the measuring equipment. (In Fig. 2 the Zn line can be seen only in "Kubasov's last spectrum".) The Zn and Cu values under the detection limit cannot be evaluated. The Zn value for Cosmonaut Kubasov is 28.3 ± 1.3 times the normal upper limit in the sample taken on the fifth day after landing; this value is unchanged for Cosmonaut Farkas.

Br levels, with the exception of the sample taken on the fifth day after landing, can be compared with the serum level of other healthy individuals. In the last sample (see Figs 2 and 3) the Br level increases to a value which is $\simeq 30$ times higher than the former values for both cosmonauts.

In Fig. 4 the X-ray spectrum of Farkas's beard sample taken before the space flight can be seen. It is obvious from the Figure that the Zn concentration can be measured with considerably higher sensitivity while Br can be measured with the same sensitivity if we use hair samples for analyses instead of blood serum samples.

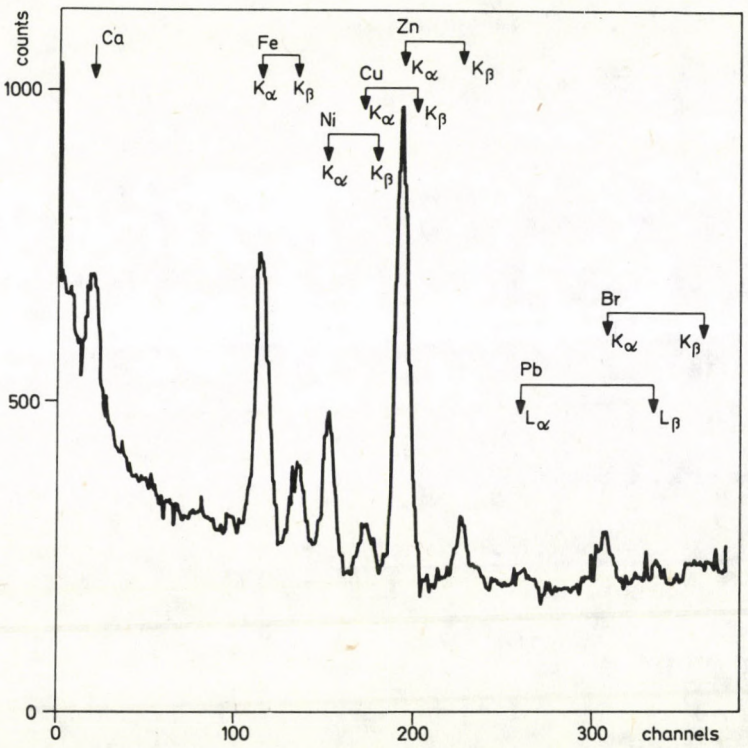


Fig. 4. Spectrum of beard samples of B. Farkas before his space flight (22—26 May 1980)

Had the beard samples been collected separately day by day before, during and after the space flight, it would have been possible to analyse the changes in Zn and Br content for every day and probably to clear up the origin of or reason for the changes. On the basis of the above, we suggest that the investigation of changes of mineral element metabolism in cosmonauts by analysing beard samples is worthy of greater consideration.

References

1. J. P. Henry, O. H. Gauer and J. I. Reves, Circ. Res., **4,** 85, 1956.
2. O. H. Gauer, J. P. Henry and H. O. Sieker, Prog. Cardiovasc. Dis., **4,** 1, 1961.
3. O. H. Gauer, P. Eckert, D. Kaiser and H. J. Lincenbach, 2nd Int. Symp. on Basic Environmental Problems in Space. Paris, 1965, Springer, New York, 1967. p. 212.
4. H. C. Hopps, Sci. Tot. Environ., **7,** 71, 19771.
5. L. Goreczky and J. Sós, Klinikai kémiai laboratóriumi zsebkönyv (Laboratory pocket manual of clinical chemistry, in Hungarian), Medicina, Budapest, 1974.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 165—178 (1982)

APPLICATION OF NUCLEAR METHODS IN PALEOSCIENCES*

Cs. M. BUCZKÓ

ISOTOPE LABORATORY
KOSSUTH LAJOS UNIVERSITY, 4010 DEBRECEN, HUNGARY

J. CSIKAI

INSTITUTE OF EXPERIMENTAL PHYSICS
KOSSUTH LAJOS UNIVERSITY, 4010 DEBRECEN, HUNGARY

and

J. NEMESKÉRY

INSTITUTE OF ZOOLOGY AND ANTHROPOLOGY
KOSSUTH LAJOS UNIVERSITY, 4010 DEBRECEN, HUNGARY

Concentrations of N, F, Na, Mg, Al, Si, P, K, Ca, Mn, Fe, Zn, Sr and Ba were determined in fossil bones using neutron activation analysis and X-ray fluorescence methods. Analysis of the data led to the conclusion that a correlation exists between the concentration of nitrogen and the temperature at the burial time of bones. This correlation enabled the temperature for different eras to be estimated. Correlation was also found between the Al, Si and Fe infiltered into the organic part of bones.

Introduction

Certain authors (Middleton [1], Kiszely [2], Vonach [3], Szöőr [4], Eisenbarth and Hille [5]), have stated that a definite correlation exists between the chemical composition and the age of fossil bones. Mainly the nitrogen content or organic fractions, and fluorine and uranium concentrations were used for dating. According to these investigations the nitrogen content decreases with time because of the degradation of the organic materials. This observation was supported by laboratory experiments (Abelson [6]) measuring the decay rates of amino acids at temperatures higher than those at which fossilization occurs in nature. It is supposed that the increase of the uranium and fluorine contents of fossil bones is linked with the transformation of hydroxyapatite into other types of apatites.

Other authors (Protsch [7], Cook and Heizer [8]) have observed a large spread in the concentration of chemical elements even for bones of the same age. This fact made the dating based on the chemical composition of bones doubtful. The spread in the concentration was attributed to the effect of the soil in which the bones were embedded (Oakley [9]), therefore, only samples discovered in the same geological horizons have been used for relative dating.

The aim of this work is to study the chemical composition of fossil bones by systematic measurements on a large number of samples. In our experiments the

* Dedicated to Prof. I. Tarján on his 70th birthday.

Table I
Nuclear methods used for the determination of elements

Element	Method	Reaction
N	FNAA	$^{14}N(n, 2n)^{13}N$
F	FNAA	$^{19}F(n, 2n)^{18}F$
Na	TNAA	$^{23}Na(n, \gamma)^{24}Na$
Mg	TNAA	$^{26}Mg(n, \gamma)^{27}Mg$
Al	TNAA	$^{27}Al(n, \gamma)^{28}Al$
Si	FNAA	$^{29}Si(n, p)^{29}Al$
P	XRFA	–
K	TNAA	$^{41}K(n, \gamma)^{42}K$
Ca	XRFA	–
Mn	TNAA	$^{55}Mn(n, \gamma)^{56}Mn$
Fe	FNAA	$^{56}Fe(n, p)^{56}Mn$
Fe	XRFA	–
Zn	XRFA	–
Sr	XRFA	–
Ba	XRFA	–

nondestructive neutron activation analysis and X-ray fluorescence methods were utilized to determine the elements listed in Table I, which also summarizes the corresponding methods.

According to an earlier measurement of Buczkó [10] the concentration of elements in the different parts of fossil skeleton deviates significantly; therefore, only human vertebrae were selected for these investigations. Bones from the past 9000 years discovered in Hungary, Yugoslavia and Bulgaria were studied. The age of these bones was determined either by ^{14}C or by morphological methods.

Nitrogen content as a paleothermometer

Using the available methods significant differences were obtained for the decay rate of organic components of bones in time. On the basis of experimental data measured by Vonach [3] for the bone samples from the past 100 million years, Buczkó and Vas [11] deduced the following analytical relation for the decreasing of nitrogen content (N) as a function of age (t)

$$\lg N(t) = -0.135 \lg t + \lg N_0 , \tag{1}$$

where t is given in years and N_0 denotes the nitrogen content in weight per cent at $t = 1$ year. It is considered that N_0 is the initial value of the decreasing nitrogen content. In

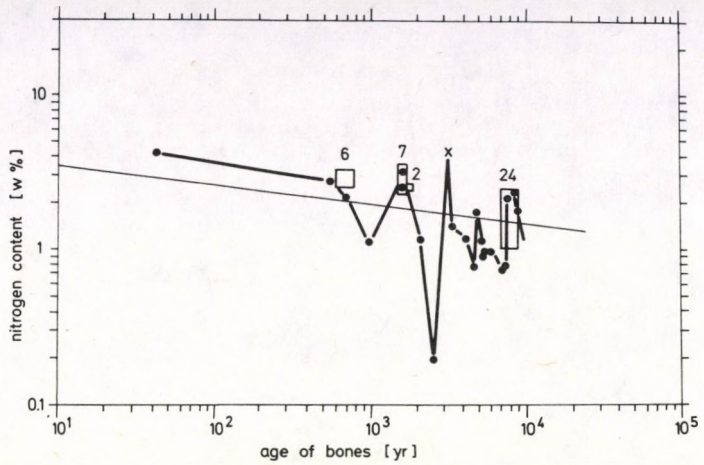


Fig. 1. Nitrogen content of bone as a function of age (●: single measurements, □: a set of samples)

addition to Eq. (1) there are a few expressions for the function of $N(t)$ based on samples from short time intervals. Measurements on a large number of dated samples from the past 9000 years made it possible to check the validity of Eq. (1). Data obtained in our experiment show large fluctuations around the gross trend determined by Eq. (1). As can be seen in Fig. 1 the nitrogen contents deviate by an order of magnitude at 2600 and 3100 years resulting in a large error (10^6 years) in the determination of age. Buczkó and Vas [11] found that the nitrogen content changes with time with a periodicity of about 2000 years. This fluctuation cannot be attributed to edaphic effects because the bones from an 8000 years old graveyard used for 2000 years in Vlasac (Yugoslavia) show the same behaviour (Buczkó et al [12]).

Starting from the fact that the decomposition of protein depends strongly on temperature, it was supposed that the nitrogen content of fossil bones is governed by two independent parameters: age and climate. This supposition is supported by the curves of climatic change for the past 10 000 years (see Fig. 2). Although the shapes of curves given by different authors (Sugden and John [13], Flohn [14], Ambrózy et al [15]) somewhat deviate from each other, the trends of fluctuations are similar. As can be seen in Fig. 2, a definite correlation exists between the nitrogen content of bones and the climate. The minima in nitrogen content coincide fairly well with the climatic maximum. There is a discrepancy in the curves of climate at 1300 ys B. P. The nitrogen content observed for this age supports the existence of a cold period.

For the separation of the two effects (age and climate) on the nitrogen content, the N_0 values were determined by Eq. (1) in the knowledge of N and t. To describe the correlation between the initial nitrogen content and the climate by an analytical expression the temperature was replaced by the sea-level difference. On the basis of the data for sea-level (Fairbridge [15]; Járai-Komlódi [17]) and temperature (Ambrózy

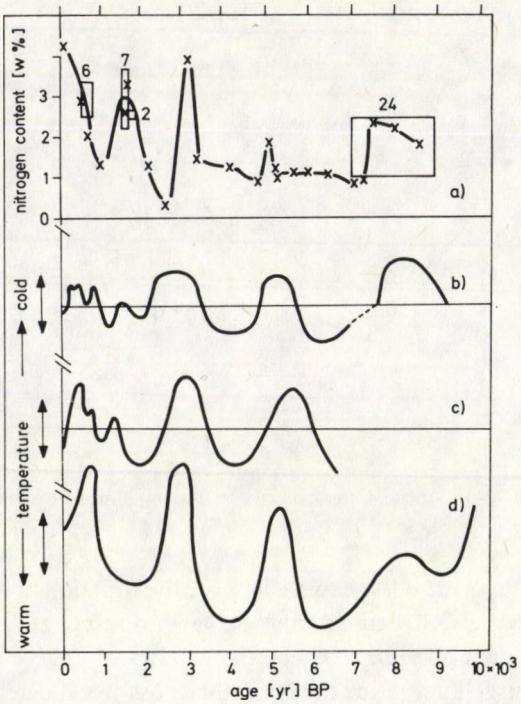


Fig. 2. Correlation between nitrogen content and temperature (× : single measurements, □: a set of samples). Curves b, c and d were given by Flohn [14], Sugden and John [13] and Ambrózy et al [15], respectively

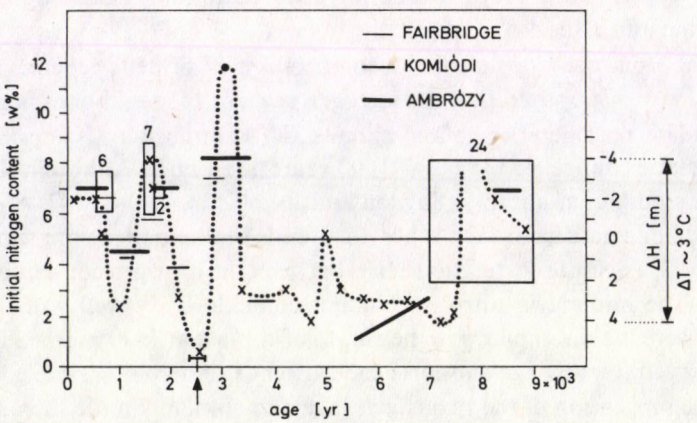


Fig. 3. Initial nitrogen content of bones (× : single measurements, □: a set of samples) and sea-level differences (—) in different eras

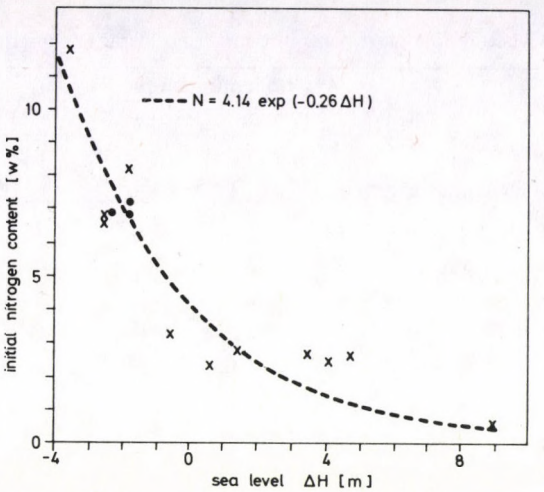


Fig. 4. Nitrogen content of bones at burial time as a function of sea-level differences. × : single measurements, ●: average value for a short time interval belonging to the squares represented in Figs 1, 2, and 3

[15]), a value of $\Delta H/\Delta T \simeq 2.6$ m/°C was estimated, if a linear correlation is accepted (Brooks [18]) between ΔH and ΔT. The negative value of ΔH indicates a lower temperature than the present one. The sea-level differences ΔH normalizing to the present one and the N_0 data are indicated in Fig. 3 as a function of age. An extremely high value for sea-level (9m) was given by Brooks [18] at 2600 yr B. P. As can be seen in Fig. 3, the correlation between N_0 and ΔH is more expressive than that indicated in Fig. 2. This observation means that the decay of nitrogen content in time cannot be neglected and Eq. (1) deduced from the data given by Vonach [3] seems to be a good approximation.

If one plots the initial nitrogen content as a function of ΔH, the data show a monotonic trend which can be well approximated by the following relation (dashed curve in Fig. 4):

$$N_0 = 4.14 \exp(-0.26\,\Delta H), \tag{2}$$

where N_0 is in w% and ΔH is in meters. This correlation means that the nitrogen content of bones is mainly influenced by the climate obtaining a few years after the burial time. The nitrogen content formed in these years decays continuously in time independently of the climatic change during the thousands of following years. The correlation expressed by Eq. (2) proves the negligible edaphic effect on the actual nitrogen content of bones, too.

A further item of evidence for the governing effect of the temperature on the nitrogen content of bones was gained by measuring a large number of samples from Vlasac. The nitrogen content of these samples together with their F, Ca and P concentrations are summarized in Table II.

Table II
Concentration of elements in bones from Vlasac (Yugoslavia)

Site location	Concentration [w%]							
	N	F	Ca	P	Fe	Al	Si	Mn
2	1.03	0.12	27.2	13.5	0.95	1.90	5.47	0.016
6	1.39	0.08	43.4	8.6	0.49	0.86	2.06	
7	1.09	0.11	23.4	7.5	1.32	2.75	7.09	0.024
16	0.74	0.16	24.8	6.9	1.74	3.26	9.71	0.017
18/A	1.55	0.12	27.8	10.9	1.15	1.86	5.20	
23	2.03	0.08	43.0	18.2	0.13	0.20	0.66	
27	2.36	0.02	34.0	15.1	0.38	0.90	3.02	
28	1.43	0.13	20.6	9.4	1.41	2.85	6.87	
29	1.84	0.10	42.2	20.3	0.32	0.55	1.60	
31	1.80	0.08	32.5	14.6	0.64	1.16	3.81	
32	2.17	0.12	30.8	12.4	0.54	0.72	–	
38	1.29	0.10	36.5	16.7	0.44	1.07	3.16	
46	1.69	0.12	27.8	11.6	0.86	1.67	4.06	
50	1.51	0.05	34.8	12.9	0.53	1.21	3.57	
50/B	1.106	0.11	45.2	17.0	0.24	0.40	1.03	
54	1.14	0.09	34.2	15.5	0.64	1.27	3.35	
55	1.72	0.06	36.6	13.0	0.54	1.11	2.88	
60	1.55	0.06	36.8	13.6	0.58	1.15	4.07	
63	2.08	0.09	41.8	18.4	0.26	0.30	0.85	
66	1.09	0.13	14.0	6.8	2.17	4.02	10.50	0.024
67	1.92	0.10	35.8	15.2	0.44	0.92	2.04	
69	1.69	0.15	22.3	11.3	1.07	2.06	5.21	
70	2.13	0.12	21.5	8.6	1.09	2.13	6.26	
71	1.18	0.22	33.4	18.2	0.76	1.79	4.93	
72	1.58	0.10	38.0	19.1	0.30	0.70	1.98	
78/A	1.63	0.09	31.5	15.3	0.56	0.68	1.91	
82	1.63	0.09	25.8	12.8	0.82	1.88	–	
84	1.56	0.16	–	–	0.27	0.60	1.27	

The data in Table II show large spreads for each element, e.g. in the case of nitrogen a factor of three was obtained, proving a definite change in the temperature for this period. From Eq. (1) a change of 1.2% caused by age can be expected only for N.

Figure 5 shows the nitrogen and organic contents of snake vertebrae of different ages discovered in a cave. The shapes of curves are similar proving the correlation between the nitrogen and organic contents. For the determination of organic content

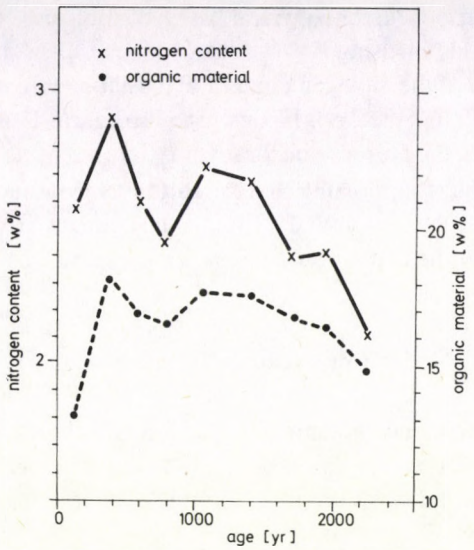


Fig. 5. Nitrogen and organic material contents of bones from different ages

the thermoanalytical method was used, the age was estimated stratigraphically. The gross trends in the curves are similar to those obtained for human vertebrae. The ratio of the organic part to the nitrogen content was found to be 6.5 in good agreement with the value for protein.

Judging by the results mentioned above the nitrogen content of fossil bones can be used as a paleothermometer. On the basis of the measured nitrogen content for

Table III

Sea-level differences determined by different methods

Site location	Age of bone [yr]	Number of samples	Measured nitrogen content [w%]	Initial nitrogen content [w%]	Estimated sea-level difference [m]	Sea-level given by	
						Fairbridge [16]	Járai-Komlódy [17]
Kavarna Chirakman	550±100	2	2.84	6.57	−1.77	−2	−2.5
Kabile II (Jambol)	680±150	7	2.87	6.92	−1.96	−2	−2.5
Pernik	700± 50	2	2.15	5.15	−0.83	−2	−2.5
Kabile I (Jambol)	1580±100	6	2.69	7.24	−2.12	−1	−2.5
Stara Gora	1650±200	3	2.58	6.98	−2.00	−1	−2.5
Arbitus (Razgrad)	1780±100	2	2.50	6.84	−1.98	−1	−2.5

dated bones from various archaeological sites of Bulgaria the temperature was estimated (Buczkó et al [19]) using Eqs (1) and (2). The results summarized in Table III, prove the applicability of the nitrogen content of fossil bones in paleoclimatology. The data determined by different methods (Fairbridge [16], Járai-Komlódy [17]) indicate a lower temperature than the present one for several time intervals of the past 2000 years.

There are a few facts supporting that the Piltdown remains is a hoax. One of these is the relatively high N content. It should be noted, however, that the high concentration of N can be explained if a low temperature was present at the time of burial.

Correlation of elements

The applied radioanalytical methods give a possibility to determine simultaneously a number of elements in fossil bones. The results indicated in Table IV show that there is no correlation between the age and the concentration for F, P, Ca, Fe, Al, Si, Na, K, Mn, Mg, Sr, Zn and Ba.

Plotting the concentrations obtained for Al, Si, Fe and Ca as a function of age, we found fluctuations for these elements with the same period as for nitrogen. The behaviour of the concentrations shown in Fig. 6 cannot be attributed to the soil environment because large spreads are present for samples from the same sites (see Table II). An opposite trend is shown for Si, Al and Fe to that for N and Ca. As can be seen in Figs 7 and 8 a definite correlation exists between the Al, Si and Fe contents in fossil bones. The atomic ratios of these elements are: $Si/Al = 2.44$; $Al/Fe = 3.56$, independently of the age and the site of the remains. This suggests that these elements are present in a well defined chemical compound. There are a number of clay minerals

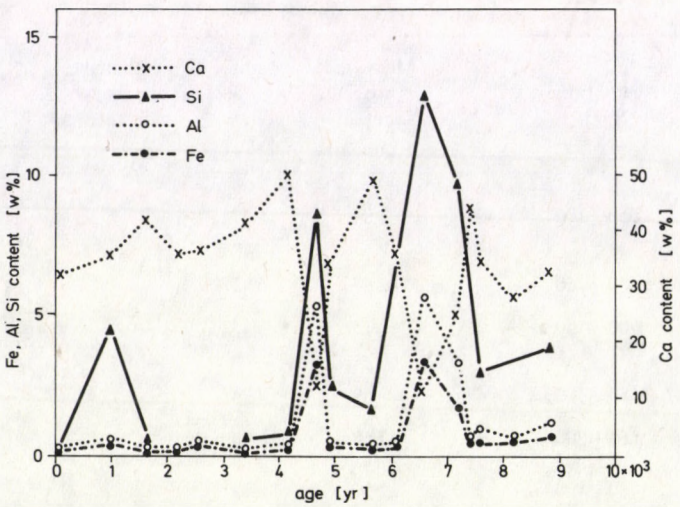


Fig. 6. Concentration of Ca, Si, Al and Fe in bones as a function of age

Table IV

Concentration of elements in fossil bones

Site location	Archaeological period	Age [yr]	Concentration [w%]										ppm		
			N	F	Ca	P	Fe	Al	Si	Na	K	Mn	Mg	Sr	Zn	Ba
Budapest	Present Age	40	4.22	0.07	32.5	12.6	0.12	0.04	—	0.28	<0.05	—	—	100	232	196
Rakamaz	10th century	100	1.13	0.10	35.6	15.6	0.29	0.60	4.62	0.18	0.08	—	—	280	42	140
Unknown	Roman period	1630	3.23	0.22	42.0	19.0	0.10	0.20	0.40	0.21	0.13	—	—	190	47	70
Pilis	Unknown	2180	1.18	0.09	36.0	10.6	0.15	0.22	—	—	—	—	—	80	260	50
Mátraszele	Iron Age	2580	0.20	0.01	36.5	18.2	0.49	0.42	—	—	—	—	—	110	187	120
Egyek	Late Bronze Age	3430	1.40	0.04	41.0	20.6	0.10	0.10	0.54	0.09	<0.05	—	—	310	86	140
Bodrogkeresztúr	Copper Age	4230	1.19	0.11	50.0	18.3	0.12	0.18	0.70	0.15	0.44	—	—	140	750	70
Polgár-Basa-tanya II.	Late Neolithic-Aeneolithic	4730	0.78	0.13	12.2	5.1	3.28	5.27	8.70	0.09	2.41	0.03	~4–5	200	92	460
Polgár-Basa-tanya III.	Late Neolithic-Copper-Age	4980	1.81	0.21	34.1	15.1	0.31	0.40	2.38	0.22	<0.05	—	—	1600	77	260
Lengyel	Neolithic	5310	1.17	0.05	—	—	0.12	0.18	—	0.13	0.52	—	—	110	106	160
Szob	Neolithic	5320	0.92	0.05	—	—	0.07	0.06	—	—	—	—	—	—	—	—
Unknown	Neolithic	5730	1.04	0.16	49.0	16.1	0.22	0.36	1.70	0.16	0.28	—	—	170	40	530
Bánhida	Early Neolithic	6130	0.98	0.07	35.8	15.5	2.26	0.40	—	—	—	—	—	70	73	—
Zsáka	Early Neolithic	6580	1.02	0.59	11.2	2.5	3.25	5.58	12.93	0.23	2.06	0.02	~4–5	0	71	490
Vlasac 16	Mesolithic	7230	0.74	0.16	24.8	6.9	1.74	3.26	9.71	—	—	—	—	110	86	330
Aszód	Late Neolithic	7480	0.84	0.10	44.0	16.5	0.37	0.60	—	0.38	1.03	—	—	380	50	50
Vlasac 27	Mesolithic	7630	2.36	0.02	34.0	15.1	0.38	0.90	3.02	0.13	<0.05	—	—	170	98	190
Vlasac 32	Mesolithic	8230	2.17	0.12	30.8	12.4	0.54	0.72	—	0.17	<0.05	—	—	160	223	140
Vlasac 31	Mesolithic	8830	1.80	0.08	32.5	14.6	0.64	1.16	3.81	0.12	0.52	—	—	220	113	240

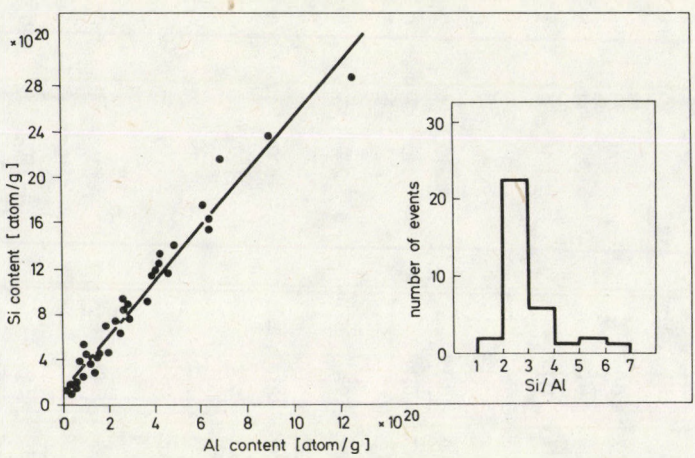


Fig. 7. Correlation between the Si and Al contents infiltered in the organic part of fossil bones

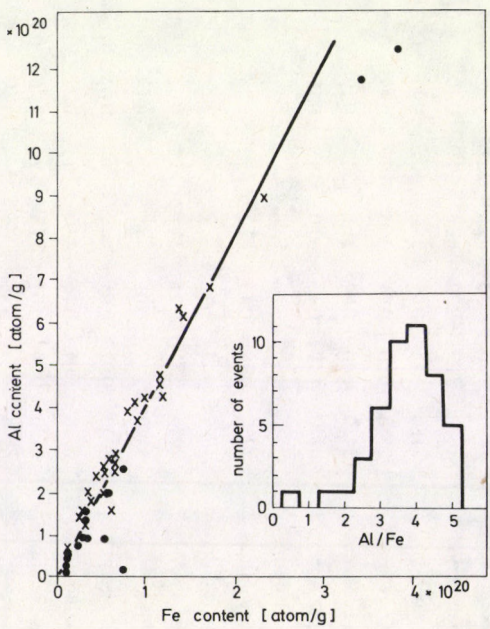


Fig. 8. Correlation between the Al and Fe contents infiltered in the organic part of fossil bones

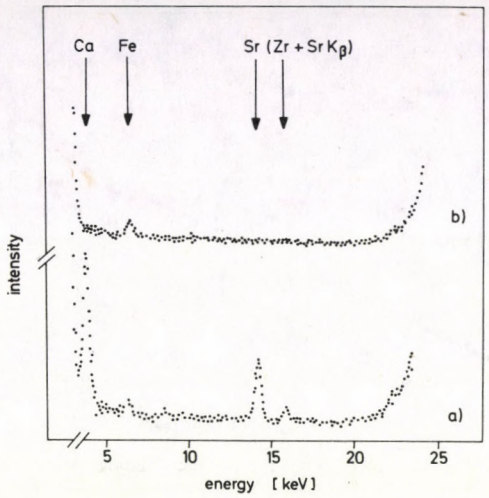


Fig. 9. X-ray spectra of fossil bone (a) and its organic parts (b)

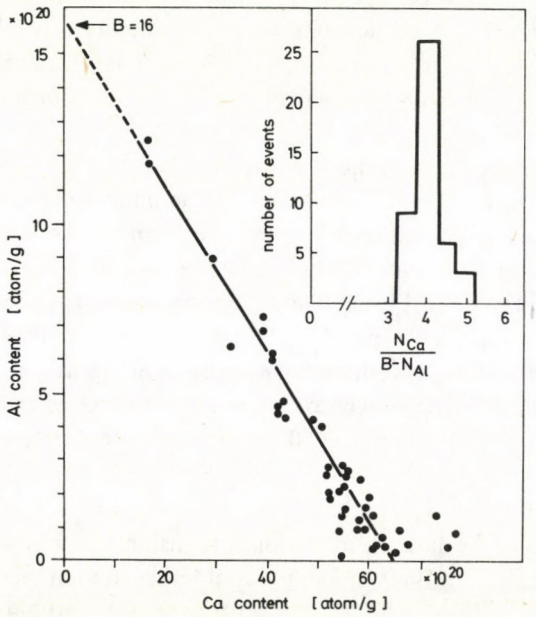


Fig. 10. Correlation between the Al and Ca contents of fossil bones

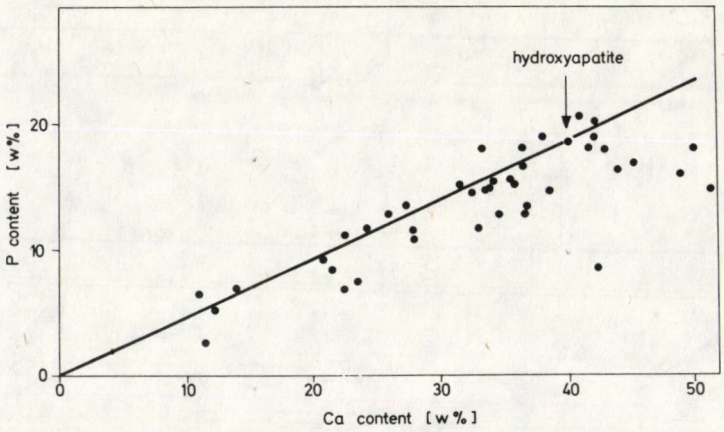


Fig. 11. Correlation between the Ca and P contents of fossil bones

containing Si, Al and Fe. The atomic ratios of these elements for Mg-parahalloysite (Nemecz [20]) are in good agreement with those obtained for fossil bones. However, the same values are found for a mixture of vermiculite and montmorillonite with a ratio of 1/3 to 2/3, respectively. Judging from the conclusion of Bowen [21], these minerals are present in soils bound strongly to organic matter. The X-ray spectra of bone and its organic fraction proves the presence of Fe in the organic part, too (see Fig. 9).

It was found that the amount of clay minerals in bones is inversely proportional to the Ca content, which can be explained by the mineralization process. Figure 10 demonstrates the inverse correlation between the Al and Ca contents. The atomic ratio is Ca/Al = 4, which means that each infiltrated Al is followed by the loss of four Ca atoms. As can be seen in Fig. 11, the P and Ca contents of fossil bones vary over a wide range. The solid line represents the ratio of these in hydroxyapatite. Our measurements show a slight increase in the Ca content which can be attributed to the infiltration of Ca into the bones from soil. The concentration of clay minerals is extremely high in those remains which were buried around 6500 yr B. P. — the estimated time of the flood.

The concentrations of 14 elements in fossil bones compared with recent ones are summarized in Table V.

On the basis of the data the following conclusions can be drawn: a) the content of Mg, Al, Si, Mn, Fe, Sr and Ba is higher in fossil bones than in recent ones; b) a definite decreasing trend was observed for the concentration of N and Na, whereas the content of other elements, such as F, K, Ca, P and Zn, fluctuates randomly.

Further investigations are needed to clarify the details of the fossilization mechanism. Special attention should be paid to those biochemical processes which govern the temperature sensitivity of bones.

Table V

Concentration of elements in recent and in fossil bones

Element	Recent bone [ppm]	Fossil bone [ppm]
N	45 000	4 000—42 000
F	1500	100—5 900
Na	14 000	900—3 800
Mg	1 700	40 000—50 000
Al	<8	400—56 000
Si	105	4 000—130 000
P	118 000	25 000—206 000
K	1 500	500—24 000
Ca	257 000	1 120 000—500 000
Mn	0.2	160—260
Fe	200	800—34 000
Zn	120	40—260
Sr	64	700—1 600
Ba	6.9	50—530

References

1. J. Middleton, Proc. Geol. Soc. of London, **4,** 431, 1844.
2. I. Kiszely, Derivatographic research of subfossile bones, A Móra Ferenc Múzeum Évkönyve, 217, 1969.
3. H. Vonach, Fast Neutron Physics at the Institut für Radiumforschung und Kernphysik, ATOMKI Közlemények, **18,** 247, 1976.
4. Gy. Szöőr, Acta Mineralogica-Petrographica, **22,** 61, 1975.
5. P. Eisenbarth and P. Hille, J. of Radional, Chem., **40,** 203, 1977.
6. P. H. Abelson, Geochemistry of Amino Acids, Organic Geochemistry (ed. I. A. Berger) International Series of Monographs on Earth Science, **16,** 431, 1963.
7. R. Protsch, Comparison of Absolute Bone Dates by Radiocarbon and Amino-Acid Dating on Upper Pleistocene Hominids, UISPP IX. Congress, Nice 1976, p. 180—211.
8. S. F. Cook and R. F. Heizer, Am. Antiquity, **4,** 354, 1953.
9. K. P. Oakley, Bulletin of the British Museum (Natural History), Geology Series, **34,** 1, 1980.
10. Cs. M. Buczkó, Nukleáris módszerek alkalmazása fosszilis csontok elemösszetevőinek meghatározására, (Use of nuclear methods for the determination of chemical composition of fossil bones). Thesis, Debrecen, 1978 (in Hungarian).
11. Cs. M. Buczkó and L. Vas, Nature, **269,** 792, 1977.
12. Cs. M. Buczkó, J. Nemeskéry and L. Vas, Age Dating of Bones from Vlasac Based on their Nitrogen and Fluorine Contents, VLASAC a Mesolithic Settlement in the Iron Gates, Vol. II. p. 75—76, Geology-Biology-Anthropology, Beograd, 1978.
13. D. E. Sugden and B. S. John, Glaciers and Landscape, London, 1976.
14. H. Flohn, Stehen wir vor einer Klima-Katastrophe?, Umschau **77,** Heft 17, 1977.

15. P. Ambrózy, R. Czelnai and G. Götz, Fizikai Szemle, **27,** 54, 1977 (in Hungarian).
16. Rh. W. Fairbridge, Phys. and Chem. of the Earth, **4,** 99, 1961.
17. M. Járai-Komlódy, Botanikai Közlemények, **58,** 131, 1971 (in Hungarian).
18. C. E. P. Brooks, Climate Through the Ages, London, 1950.
19. Cs. M. Buczkó, N. I. Ilkov, A. Borbély, Radiochem. Radioanal. Letters, **35,** 175, 1978.
20. E. Nemecz, Agyagásványok, (Clay Minerals), Akadémiai Kiadó, Budapest, 1973 (in Hungarian).
21. H. I. M. Bowen, Trace Elements in Biochemistry, London, 1966.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 179—188 (1982)

INVESTIGATIONS INTO THE MECHANISM OF ACTION OF IONIZING RADIATION*

L. BOZÓKY

NATIONAL ONCOLOGICAL INSTITUTE
BUDAPEST, HUNGARY

Model tests were performed to gain further information on the mechanism for inducing malignancies by ionizing radiation. Precise variation of the physical parameters of grid irradiations made a separate study of particular test conditions feasible. Barley seeds were sown in dry and in wet soils and they were exposed to grid and to open-field radiations of varying dose; the different effects were measured at different times. Macroscopic migration is supposed. Such migration might be appropriate to explain the differences found between the biological effects of open-field and grid irradiations in man.

Knowledge of the mechanism for inducing malignancies by ionizing radiation is, for the time being, rather scanty. The number of phenomena that cannot be explained even today, after over eighty years of routine applications of ionizing radiations in therapy, is surprisingly high. On the other hand, it was precisely investigations into such manifestations that appeared to be the most suitable to lead to the discovery of further general tendencies of the mechanism of radiations effects.

One such manifestation observed by several researchers is that of grid irradiation. We do not wish to assess the by no means unequivocal results of therapeutic grid irradiations, but only intend to lay emphasis upon the empirical finding generally admitted, according to which the sensitivity of the skin to radiation substantially decreases with grid irradiation. The extent of the decrease in sensitivity is a function of numerous factors, furthermore the conditions for obtaining the results were not the same.

Some authors observed the appearance of the standard erythema of the skin in the case of grid irradiation after the application of doses three, four, or five times higher than in the course of irradiation without a grid, that is to say, at 600 cGy skin tolerance dose generally adopted.

The essentials of the phenomenon can be summarized thus: a skin surface of say, 1 cm^2, will show the same extent of skin erithema only by a significantly larger dose when it is surrounded by other, non irradiated skin portions, than when the same skin surface of 1 cm^2 makes part of a larger, uniformly irradiated skin surface, that is to say, encircled all around by irradiated skin.

* Dedicated to Prof. I. Tarján on his 70th birthday.

This fact is usually explained by stating that in case of a grid irradiation the portion of the irradiated skin can regenerate from the neighbouring, non irradiated skin surfaces.

The process cannot be studied directly on living humans, and there is no information available in the literature on animal experiments. It seems that, for the moment the difficulties of test procedures prevent the carrying out of such experiments.

In order to attempt to make a step towards the discovery of the mechanism of radiation effects, we have tried to perform conclusive model experiments by applying exact biophysical methods:

A *model* was constructed in which the biological matter itself and its environment could be varied independently. Thus the effects of ionizing radiation could be studied under variable conditions, and the correctness or incorrectness of certain theoretical assumptions ascertained.

The biological matter in the model consisted of various seeds, mainly barley, the intercellular matter of sand, with a bed of pebble under it, through which the sand layer of 5 cm thickness could be saturated promptly and uniformly with water. Barley was obtained from the Tápiószentkereszt Research Institute as a material of 100 per cent viability. On every occasion, 900 seeds were sown in sand, previously dried into cubes of 1 cm^2 of a square grid of 30 by 30 cm, with the germinating end of the seeds pointing upwards. The test box was placed into another box. (Fig. 1). As water was let flow into the second box, the liquid level rising in the pebble bed wetted the sand simultaneously and uniformly, without displacing the seeds, in a perfectly reproducible way.

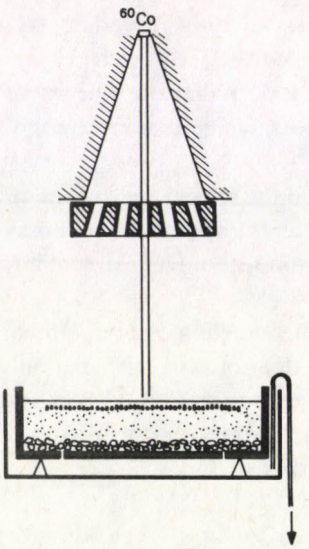


Fig. 1. The 50 per cent grid-filter made of lead, with openings of 25 mm diameter

A feature of great importance described above is that in every phase of the experiment the model accurately simulates the natural conditions.

The model was irradiated by a double-tank system Rotacert cobalt unit developed in our Institute. The charge of the unit was ^{60}Co of 130 TBq activity. After the irradiation treatment the models were transferred to our (constant temperature) laboratory where observations and measurements were performed and photos taken.

The task

Our aim was to check whether during irradiation a macroscopic migration takes place in the biological system and if so, what is its mechanism. We wished to study the problem by a model test with grid irradiation.

Test methods

The Rotacert cobalt unit [1], [2] was, first of all, completed by a projection grid [3] that could be rigidly fixed to the irradiation head at a distance of 50 cm from the source, and could project the gamma field of the grid assembly upon a larger surface. The grid filter and model are illustrated in Fig. 1. The openings, directed towards the 20 by 20 mm cobalt source, are 25 mm in diameter. Half of the total area is screened. The thickness of the lead plate is 45 mm, the screened parts letting 10 per cent of the ^{60}Co radiation through. Thus, grid irradiation gave strongly irradiated areas (100 per cent at the grid points) and slightly irradiated ones (10 per cent through the screened areas) (Fig. 2).

Open-field irradiations were performed in flower pots, grid irradiations in the wooden boxes described above. The irradiations were directed vertically downwards, perpendicularly to the sown areas. In order to limit the risk of errors. irradiations in dry and in wet soil were performed for comparison at the same time and by a single exposure.

The barley seeds previously germinated, sprouted on the fourth day. Observations and measurements were continued for a fortnight.

Aspects of evaluation

We considered the following:

— The number of seeds sprouted in a given area, e.g. in a grid opening area.

— The summed heights of the plants sprouting in a given surface area.

— The colour of the sprouting plants.

An issue of importance was the comparison of the radiation effects brought about in dry and in wet soil. In view of this the wet germinated seeds had to be sown

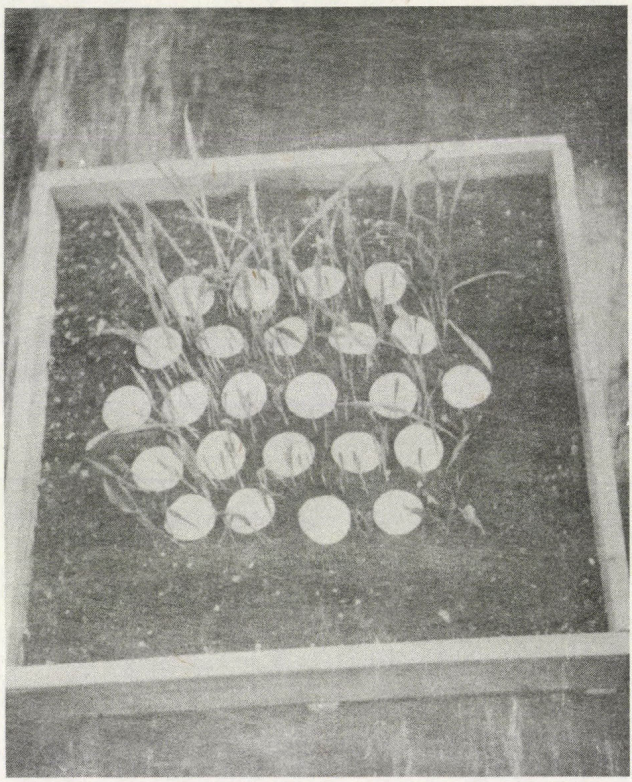


Fig. 2. Plants that sprouted on the tenth day after grid irradiation. The seeds failed to sprout at the grid points, where the white paper circles lie smoothly upon the soil

partly in dry, partly in wet soil. Thus, the conditions of development were variable even without irradiation, as the dry soil gradually absorbed the moisture of the seeds. The latter by and by became desiccated, thereby causing the growth rate to slacken, as compared with the rates of growth of the seeds in wet sand.

Assuming the time taken by the sowing of 900 barley seeds under the conditions outlined above to be 2 to 3 hours, and the irradiation taking a half to two hours, we had to reckon with a significant percentage drying. Therefore we additionally examined the influence of the drying process as a function of time, also looking for practicable ways to reduce the time.

A significant shortening of time was possible by sowing barley seeds soaked in water (germinated), instead of dry seeds. The fact that quickly cleaving cells in quick cleavage are as a rule more sensitive to radiation than cell populations multiplying by bipartition at normal rates is well known. The delivery of doses several times smaller took significantly shorter irradiation times. For our purposes the optimum soaking time for barley was found to be 20 hours so this was the time we used.

The barley seeds soaked in water for 20 hours were sown in groups of 20 in dry sand, and the groups were watered after 1/4, 1/2, 1, 2, 3 and 6 hours. The effects of drying were ascertained by measuring the heights of the plants on the sixth day. Our measurements showed that the slackening of the growth rates of the plants, watered within the first 2 hours still keeps below 5 per cent. This means that the added times of sowing and irradiation had to be brought down to below 2 hours. Even though this raised rather thorny problems, yet the discussion of such details will not be entered into here.

Test results

The results can be summed up as follows:

The 900 barley seeds were sown in dry sand corresponding to a square network of 1 cm mesh. The sown area of 30 by 30 cm was split into two equal parts by a vertical aluminium plate. In the first part of the box the sand was wetted. Immediately after that the box was placed under the Rotacert cobalt unit and the two parts were irradiated at the same time with 6,500 cGy gamma dose measured at the grid points.

Dosimetry was performed with an instrument of our own construction containing a small-sized ionization chamber [4]. After exposure the second part of the

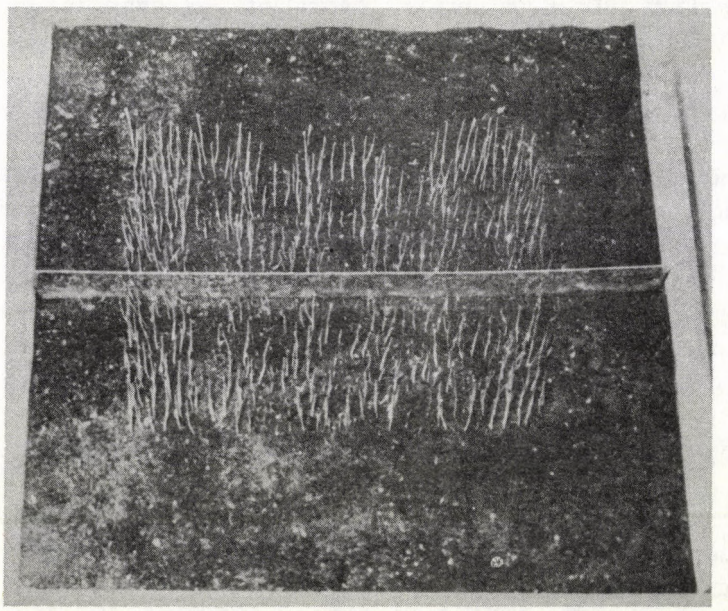


Fig. 3. Plants that sprouted after irradiation with a dose of 8000 cGy dose, on the eight day. At the time of irradiation the soil in the upper half of the box was wet, that in the lower part dry

Table I

Data on barley sown in dry and in wet soil. The data were taken on the seventh day after a grid irradiation at five grid points for each soil.

Experimental conditions	Number of plants sprouting at five grid points	Summed heights of plants	Dose at grid point	Ratio of heights
Irradiation in wet soil	42	605 mm	5500 cGy	$\frac{605}{172} = 3.5$
Irradiation in dry soil	27	172 mm		
Irradiation in wet soil	17	92 mm	8000 cGy	$\frac{92}{30} = 3.1$
Irradiation in dry soil	14	30 mm		

box was also wetted and the boxes transferred to the laboratory for further observations and for the performance of measurements.

The plants that came up on the fifth day at the grid points were counted every third day and their summed heights obtained. In the screened regions all barley seeds were sprouted (doses of 700 cGy), and the plants were grown similarly to the controls (Fig. 3). The test was then repeated with doses of 8000 and 20 000 cGy (Fig. 3).

The results obtained are given in Table I. According to the findings in case of irradiation in wet soil both the numbers of the plants sprouting at the grid point and their summed heights were found to come three times greater than those of the plants sprouting from seeds irradiated in dry soil.

Thus, between the effects of radiation in dry and in wet soils quite significant differences were found with grid irradiation. It is manifest that the cause of this must be the presence of water. The phenomenon can be explained by a sort of "poisonous matter", unknown for the time being, produced in the barley seeds by the irradiation. This matter cannot migrate in dry soil, whereas it can easily do so in water, by diffusion. Thus the poisonous matter produced in the barley grains exposed to large doses can diffuse from the grid points towards the screened areas receiving ten times lower doses only and so also the originally produced concentration of the poisonous matter may be surmised to be of ten times lower concentrations there. This causes the damaging of the seeds at the grid point to be less severe than that suffered by the seeds subjected to radiation in dry soil where migration is impossible. The tests have indicated a factor of three.

It is evident, on the other hand, that the damaging of the seeds under the screened areas, on account of the migration will be more severe than that of the seeds irradiated by the same doses but in dry soil.

It shoud be expected that, in the case of exposures to larger doses of grid irradiation performed in wet soil, the picture of the grid pattern simply vanishes from

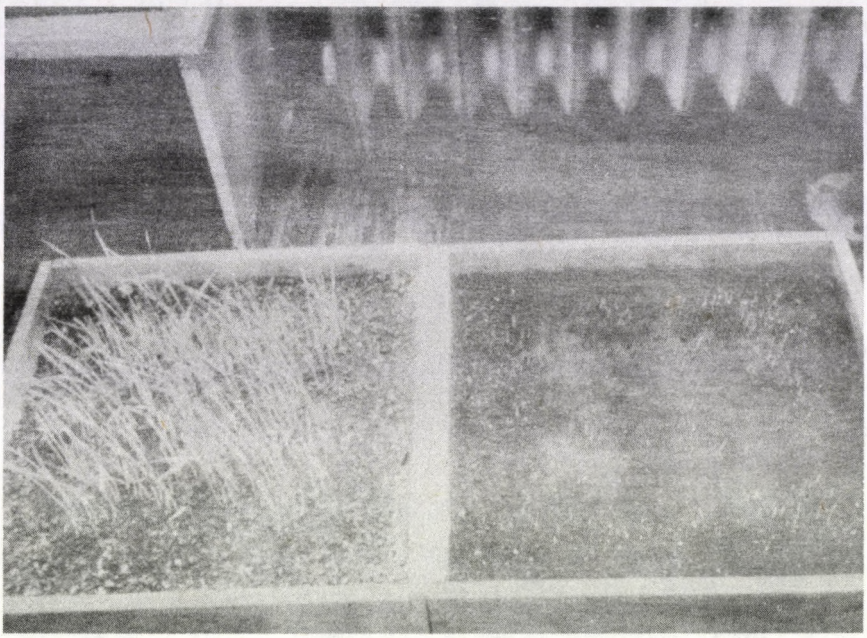


Fig. 4. Plants on the tenth day after irradiation with a dose of 80 000 cGy in a dry soil (left-hand side box) and in wet soil (right-hand side box)

the sown area and homogeneous damaging appears, because the grid points suffer lower, the screened areas suffer more severe damage on account of migration than those that would be motivated only by the dose of irradiation in these regions without migration, as in the case of dry soil. This is also substantiated by the photos of Fig. 4.

This inference was fully verified by further experiments. After exposure to irradiation of 40 000 cGy dose the grid pattern is still clearly visible in both dry and wet sands, like in Fig. 3, with a dose of 100 000 cGy, however, though in dry sand the holes corresponding to the grid formation still do appear; with wet sand the picture of the grid formation completely vanishes, severe uniform damage by radiation is found, and the few dwarfish plants sprouting here and there, will not grow higher than 2—3 mm, unhealthy looking greyish, distorted formations appear, and even the soil displays conspicuous discolouration.

For final conclusions we examined the effects of open-field irradiations without a grid, under various conditions.

In order to secure identical test conditions barley was sown in flower pots in which the soil was separated into three parts by aluminium plates, and the pots were simultaneously irradiated with doses of 10 000 cGy of gamma rays.

In the pots (Fig. 5) each upper third part contains the plants sprouting of 20 dry barley seeds sown in dry sand; the left-hand side sections of the pots those sprouting from 20 seeds sown in dry sand after a previous germinating for 4.5 h and for 20 h,

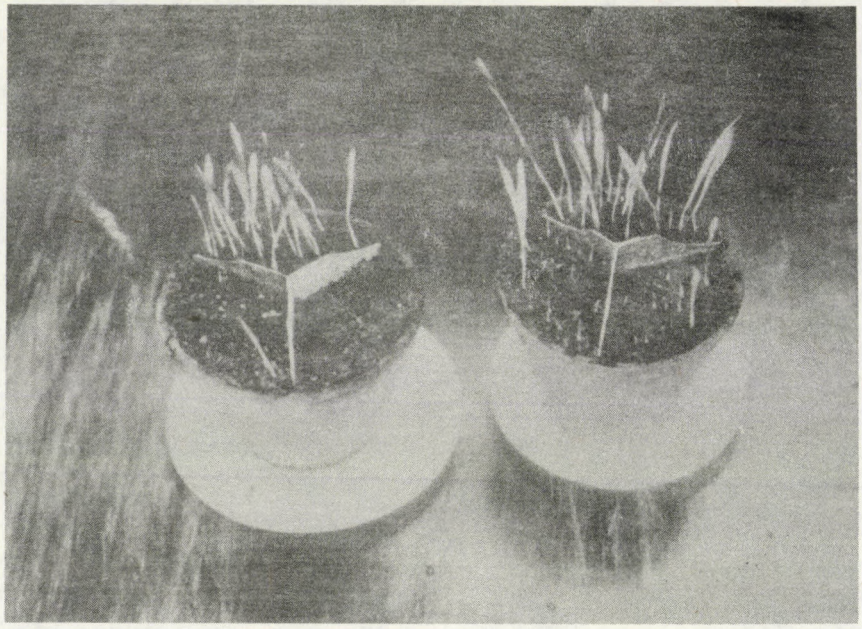


Fig. 5. Plants on the seventh day after irradiation with a dose of 10000 cGy. The seeds were preliminarily germinated for 4.5 h (left-hand side spot) and for 20 h (right-hand side spot), respectively. In the two top sectors dry seeds were irradiated in dry sand, and in the two left-hand side sectors preliminarily germinated seeds in dry sand, in the two right-hand side sectors germinated seeds in wet sand were irradiated

respectively. The right-hand side sections of the pots contain the plants sprouting from 20 seeds sown in wet sand, after germination for 4.5 h and 20 h, respectively, on the seventh day. It is clearly seen that:

a) between the dry seeds and those preliminarily germinated, differences of 20—50 times are found. The dry seeds sown in dry sand all sprouted, the plants are of a nice green colour. Also they later displayed the same rates of growth as the controls.

b) The plants from germinated seeds sown in dry and in wet sand respectively, and irradiated (the two bottom sectors in Fig. 5) are practically the same, in contrast with the differences up to 300 per cent found after grid irradiation. The seeds germinated for 4.5 hours (right-hand side pot) are perhaps less damaged than those subjected to 20-hour germination — and so obtaining somewhat greater sensitivity (left-hand side pot) where the seeds irradiated in wet sand died off completely. The differences are rather small.

This observation gives substantial support to our hypothesis, namely that, in the case of larger irradiated areas, the damaging of the application of germinated seeds with identical doses is the same, irrespectively of the application of irradiation in a dry or a wet soil, simply for the reason that the surmised "poisonous matter" produced in the barley seeds keeps stationary in both cases. In dry sand the dryness of the soil itself,

in a wet soil the "poisonous matter" present all around in uniform concentration, will prevent diffusion.

Similar tests were performed and the same results obtained with alfalfa and lettuce seeds.

Conclusions

1. The model tests of some phenomena of radiobiology seem to give a properly reproducible, exact basis for the development of a hypothesis suitable for bringing us nearer by a step to the knowledge of the mechanism of action of radiation. From our well-proved model tests showing that lesions produced at grid points are much less severe when radiation is applied in wet sand than those exposed in dry sand, the assumption can be made that with radiation in the presence of water a migration takes place, e.g. through diffusion, that cannot occur in dry sand. This migration directed from the grid opening toward the screened areas, takes away a part of the "poisonous matter" surmised to be produced by the re-unification of the molecule particles breaking down as a consequence of the ionizing radiation. Because of this, the remainder of the "poisonous matter" causes damage corresponding to smaller doses than that which would be produced in dry sand, where the displacement of the poisonous matter toward the surroundings and so the drop of concentration involved in this cannot follow.

2. As can be seen, the results of the model tests performed without a grid fully agree with this hypothesis. The reason why, for an open field irradiation with wet sand, no differences are found compared with those obtained in dry sand is that the poisonous matter does not move from the inner points towards the surroundings, not even in the case of wet sand, because in consequence of irradiation with the same dose, the poisonous matter is already present in uniform concentration all around.

3. We may state that the fact experienced with grid irradiations, namely that doses 3 to 4 times higher are required to produce the same skin reaction as in an open-field irradiation is essentially in agreement with the results of our model tests. Thus, it is not excluded from the very first that our hypotheses surmising a macroscopic migration of a poisonous matter, which was proved by our model tests, could be attempted to be applied also to the mechanism of radiation effects produced in humans and in other animals, to try to prove the existence of the said poisonous matter at a place and time that seems propitious for the performance of such work.

References

1. L. Bozóky and I. Martos, New System Cobalt Irradiation Unit. Hungarian Scientific and Technical Days, London, E.C.I., No. 7., 1967.
2. New Technique for Full Radiation Protection of Personnel in Telecobalt Treatments in Proc. First International Congress of Radiation Protection, Pergamon Press, Oxford 1968. Vol. II. pp. 1597—1601.
3. L. Bozóky and I. Rodé, Acta Phys. Hung., **28,** 83, 1970.
4. L. Bozóky and J. Toperczer, Progress in Radiophysics, Twentyfive Years in the Fight Against Cancer, Reports of the State Oncological Institute, Akadémiai Kiadó, Budapest, 1966. p. 79.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 189—200 (1982)

CONTRIBUTION OF COSMIC RAYS TO RADIATION EXPOSURE OF THE POPULATION*

L. B. SZTANYIK and I. NIKL

"FRÉDÉRIC JOLIOT-CURIE" NATIONAL RESEARCH INSTITUTE FOR RADIOBIOLOGY AND RADIOHYGIENE, BUDAPEST, HUNGARY

To evaluate the exposure of the Hungarian population to cosmic rays, the absorbed dose rate in air of cosmic radiation was directly measured by high pressure ionization chamber at ground level on the surface of different bodies of water and at various altitudes on board an aircraft. From the dose rates measured in this way, the outdoor dose equivalent rate from the ionizing components of cosmic radiation to people living at sea level would be 300—325 μSv per year. Taking into account the altitude distribution of the population, the average weighted dose equivalent is about 320, μSv per year. At Kékestető, the highest peak of the Mátra Mountains, (the highest altitude in Hungary), the annual dose equivalent is about 50 per cent higher than on the Great Hungarian Plain.

Introduction

Radiation in the environment from natural sources is the major contributor to radiation exposure of man. It is frequently used, therefore, as a standard of comparison for exposures to man-made sources of ionizing radiation, in particular, for those due to nuclear power generation. It is for this reason that interest in the knowledge and evaluation of exposure of the population to natural background radiation has increased during the last quarter of the century. Several reports have been published on nation-wide surveys in various countries [1—5] and on global averages of natural radiation exposure of mankind. The most comprehensive have been the 1972 and 1977 reports of the United Nations Scientific Committee on the Effects of Atomic Radiation UNSCEAR [6—7].

In the mid-70s, after the decision to develop nuclear power in Hungary had been taken, our Institute embarked on a research programme to examine and evaluate the studies performed by the various institutions and scientists in our country [8], and to initiate additional investigations regarding sources of ionizing radiation in the environment that may contribute to the overall exposure of the population. This paper, the first in a series of publications envisaged under this programme, is concerned with assessing the dose to the population from cosmic rays. Preliminary results have already been published at various national and international meetings [9—10].

* Dedicated to Prof. I. Tarján on his 70th birthday.

1. Cosmic radiation

The term cosmic radiation refers both to the high energy particles that enter the earth's atmosphere from outer space (primary cosmic rays), and to the particles and electromagnetic radiation generated by the interactions of primary particles with target atoms in the atmosphere (secondary cosmic rays). By virtue of these interactions, the atmosphere serves as a shield against cosmic radiation. The primary particles are almost completely absorbed in the upper atmosphere, so that the cosmic radiation detected at ground level consists mainly of the highly penetrating muon and associated decay and collision electron components of secondary radiation. There are also some smaller components such as neutrons and gamma photons. The composition and intensity of the cosmic radiation field within the atmosphere vary markedly with altitude, and there is a less important variation with geomagnetic latitude and solar cycle.

The ionization rate per unit volume in free air is a measure of the flux density of the total charged-particle component of the cosmic ray field and is usually expressed as the number of ion pairs formed per second in each cubic centimetre of air at normal temperature and pressure (NTP). The values of the sea level cosmic ray ionization rate reported after 1960 show a relatively good agreement and scatter between 1.9 and 2.6 $cm^{-3} s^{-1}$. Assuming that each ion pair requires 33.7 eV to be produced in air, the absorbed dose rate for an ionization rate of 1 $cm^{-3} s^{-1}$ is $1.50 \cdot 10^{-8}$ $Gy \cdot h^{-1}$. Using the values of the ionization rate at sea level, the absorbed dose rate in air is between 2.85 and $3.90 \cdot 10^{-8}$ $Gy \cdot h^{-1}$. For the average, a value of $3.2 \cdot 10^{-8}$ $Gy \cdot h^{-1}$ has been accepted by UNSCEAR [7].

About 75 per cent of the dose is from muon collision electrons, 15 per cent from muon decay electrons and 10 per cent from other electron, proton and neutron processes. Flux density data of cosmic ray neutrons at ground level are in serious disagreement [2]. A value of $8 \cdot 10^{-3}$ $cm^{-2} s^{-1}$ has been adopted for the purpose of estimating dose rates at sea level and latitudes above 40°. Using a conversion factor from neutron flux density to absorbed dose rate of $5 \cdot 10^{-8}$ $Gy \cdot h^{-1} cm^2 s$, the absorbed dose rate would be $4 \cdot 10^{-10}$ $Gy \cdot h^{-1}$ [7]. Apparently, the neutron absorbed dose rate in air is negligible in comparison with that from charged particles.

The extensive and traditional cosmic ray studies performed in Hungary have been mainly directed to characterizing particular components and little attention has been paid to determining their ionizing properties and none to their dose rates produced either in air or in human tissues [11—13]. It is intended that this gap be filled by our measurements.

2. Equipment and methods

Ionization chamber and thermoluminescence dosimeters have been used to measure the absorbed dose rate in air of the ionizing components of cosmic rays. Since the expectable neutron dose rate is about two orders of magnitude lower than that of

the charged particles and since we had no neutron detector of appropriate sensitivity, no neutron component measurements were made.

2.1. High pressure ionization chamber

Field measurements, especially those that take longer than a few hours, are best performed with systems which operate on selfcontaining battery packs, maintain their characteristics over long periods, and are relatively insensitive to climatic conditions. With these features in mind, the Area Monitor Model RSS-111 (Reuter-Stokes Instruments, Inc. Cleveland, Ohio, USA) was chosen as the basic instrument for our dose rate measurements.

The standard detector of the instrument is a spherical, 25.4 cm diameter, high pressure ionization chamber (HPIC). The gas filling is ultra high purity argon at a pressure of 25 atmospheres. The 2.4 $g \cdot cm^{-2}$ thick chamber shell is made of stainless steel. The spherical shape of the chamber provides an essentially omnidirectional response to the incident radiation, and the spherical collecting electrode ensures a more uniform electric field configuration than the usual rod electrodes.

A 300 V battery connected between the guard ring and the outer shell is routinely used for all measurements. Saturation studies have shown that the ionization chamber exhibits essentially complete collection of ionization currents in fields up to $1 \cdot 10^{-5}$ $Gy \cdot h^{-1}$ with the 300 V polarizing potential.

The ionization chamber is connected electrically and mechanically to an electrometer designed to give a temperature-compensated signal. This system with digital display, strip chart recorder and magnetic tape recorder is capable of resolving changes of less than $1 \cdot 10^{-9}$ $Gy \cdot h^{-1}$ in the field, using a computer technique [14].

2.2. Thermoluminescence dosimeters

Thermoluminescence dosimeters (TLD's) are widely used for environmental radiation monitoring, personnel dosimetry of professional radiation workers and special dosimetry studies. Processing techniques and procedures have been developed that make measurements of radiation exposures at environmental levels ($1 - 10 \cdot 10^{-5}$ Gy per week) accomplishable with an accuracy better than 5 per cent [15].

Various possible phosphors were checked regarding sensitivity, stability and fading, energy and directional response, manageability, etc. On the basis of the suitability for our environmental monitoring, LiF:Mg/Ti phosphor designated TLD-700 (Harshaw Chemical Co., USA) and $CaSO_4$:Dy/Tm phosphor (made in Hungary) [16] were selected and used in powder form throughout the programme. The phosphors were put into waterproof and light-tight capsules made of teflon and copper. Dosimeter readout was by a commercial reader (Harshaw 2000 AB).

For calibration of the dosimeters, different known exposures from ^{60}Co and ^{137}Cs sources were given, measured with a secondary standard dosimeter.

Field measurements with HPIC at a particular location provide quick information on exposure rate and its variations within minutes; measurements with TLD's give total exposure data integrated over an extended period. The advantages of environmental monitoring with TLD's are their relatively low price and applicability in large numbers over an extensive area.

3. Results of measurements

3.1. Measurements at ground level

At ground level, man is subjected simultaneously to all types of natural radiation. The contribution of the individual components of radiation to overall exposure can only be established if these components are measured separately. This is why a major part of our measurements of the cosmic ray absorbed dose rate have been carried out on the surface of bodies of water whose depth is sufficient to absorb radiation emitted by the radionuclides present in the earth's crust. A water layer of 2 m reduces the intensity of this radiation by about three orders of magnitude, while a layer of 4 m by more than six orders of magnitude [17].

Measurements were performed on the surface of Lake Balaton with HPIC at two points and with TLD's at one point (Fig. 1). Additional measurements were made on Lake Szelidi and on the River Danube with HPIC at one point on each. The most

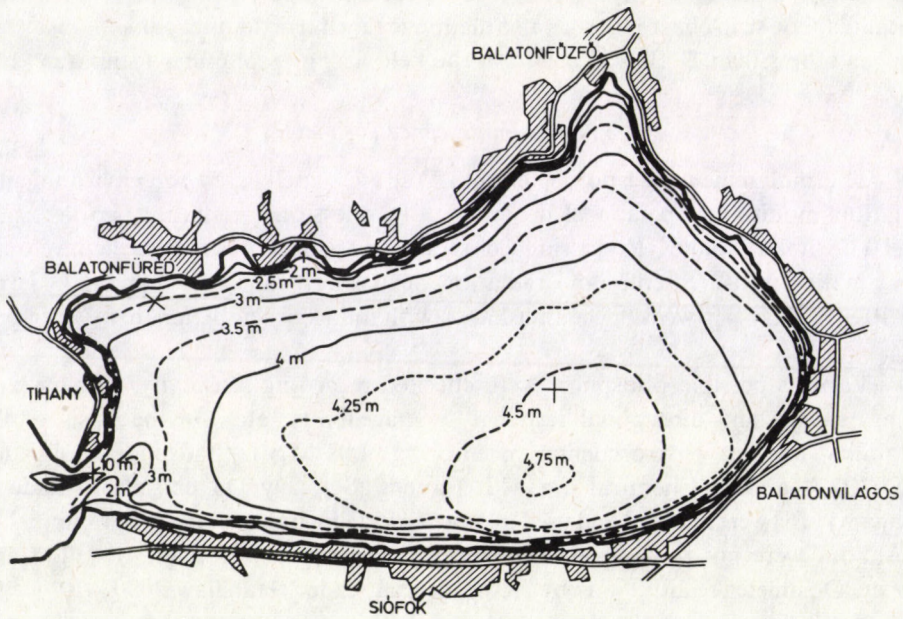


Fig. 1. Sketch of the northern basin of Lake Balaton showing water depths and the points of measurement with HPIC (+) and TLD's (×)

Table I
Absorbed dose rates in air measured at different locations

Location of measurements	Approximate distance from shore [km]	Average depth of water [m]	Overall duration of measurements [h]	Absorbed dose rate measured [nGy · h^{-1}]
Lake Balaton (Tihany)	0.4	9.5—10.5	35.2	36.7 ± 3.7
Lake Balaton (Siófok)	4	4.6	27.5	37.7 ± 3.6
Lake Balaton (Füred)	0.25	2.5	1008	36.0 ± 2.9
Lake Szelidi	0.2	5	24	40.4 ± 3.6
River Danube (Tétény)	0.2	6	1	35.6 ± 3.4

Note: All measurements were carried out with HPIC except the one at Balatonfüred that was performed with TLD's. The measured values are corrected to sea level.

important characteristics of these points, the total duration of measurements at each point and the absorbed dose rates, can be seen in Table I.

Measurements have been carried out both in summer and winter, and under various conditions (Fig. 2). It might be interesting, therefore, to see the results of individual series of measurements separately (Table II). It appears from these data that there might be a moderate seasonal variation in cosmic ray absorbed dose rate. The values obtained in winter time (28.8 ± 4.5 nGy · h^{-1}) are somewhat lower than those in summer (37.6 ± 3.6 nGy · h^{-1}). We should be cautious, however, concerning this conclusion, since the total time of measurements in winter (3.67 h) is about 16 times shorter than that in summer (59 h).

The overall average of cosmic ray absorbed dose rate in air measured with HPIC on water surface and extrapolated to sea level is found to be 37.1 ± 3.6 nGy · h^{-1}. This is about 15 per cent higher than the value assumed by UNSCEAR [7] as the world average, but it is still within the range measured and published by several authors [18—20]. Nonetheless, a potential interference of radiation from naturally occurring radionuclides should also be analysed and excluded.

Activity concentration of ^{40}K in Balaton water varied between 0.24 and 0.59 Bq · l^{-1} during the last few years [21]. This may result in an increment of absorbed dose rate in air of 0.010—0.025 nGy · h^{-1}, that is negligible.

The contribution to the absorbed dose rate in air of all the radionuclides produced by cosmic rays (cosmogenic radionuclides) is of the order of 0.20 nGy · h^{-1}; this is also insignificant compared with the direct contribution of cosmic rays [7].

More significant changes with time of the absorbed dose rate in air might be associated with variations in the atmosphere concentrations of ^{222}Rn decay products. As a result of variations in stability conditions of the lower atmosphere, the night-time concentrations of ^{222}Rn decay products in air near the ground are usually a few times higher than those existing during the day. It is believed that the absorbed dose rate



Fig. 2. Measurements of cosmic ray absorbed dose rate with HPIC on the surface of Lake Balaton

Table II
Dose rates measured with HPIC on the surface of Lake Balaton

Date of measurement	Placing of instrument	Location of measurements			
		Tihany		Siófok	
		Duration, [h]	Dose rate [nGy · h^{-1}]	Duration, [h]	Dose rate [nGy · h^{-1}]
23.01	ice	2	27.4 ± 5.4	–	–
15—17.08	pier	6	36.6 ± 3.4	–	–
15—17.08	boat	1	37.7 ± 3.9	11	37.2 ± 3.9
10—11.09	boat	25	37.7 ± 3.6	16	38.2 ± 3.4
20.12	boat	0.5	28.9 ± 2.0	0.5	39.4 ± 2.8
20.12	ship	0.67	32.3 ± 5.1	–	–
	average		36.7 ± 3.7		37.7 ± 3.6
overall average			37.1 ± 3.6		

Note: All measured values are corrected to sea level.

in air from atmospheric ^{222}Rn decay products might usually be approximately 1—2 $nGy \cdot h^{-1}$ [7]. Our long term measurements with HPIC at Tihany and Siófok (Lake Balaton) have indeed shown a diurnal variation of the dose rate in air ranging from about 36 to 43 $nGy \cdot h^{-1}$. Maximum values have been observed in the early morning and minimum values in the early afternoon. In order to clarify the role of this factor, parallel radon concentration and dose rate measurements have been initiated.

It should also be mentioned that the absorbed dose rate obtained with TLD's is corrected according to O'Brien. This author has established that TLD's, if calibrated with gamma-ray sources such as ^{60}Co, ^{137}Cs and radium in air, register 0.83 of the muon produced ionization. The ratio between the total cosmic ray ionization and the ionization inferred from the TLD calibration is 0.865 ± 0.017 [22]. Therefore the dose rate of 31.1 ± 2.5 $nGy \cdot h^{-1}$ measured with TLD's at Balatonfüred has been corrected with this factor (Table I).

3.2. Measurements at various altitudes

Alteration of the cosmic ray absorbed dose rate with altitude above sea level (i.e. with atmosphere depth) has been measured on board a JAK-40 type aircraft of the Hungarian Airlines MALÉV (Fig. 3). For these measurements, flights were made above Lake Balaton, in the air-space of the first Hungarian nuclear power station under

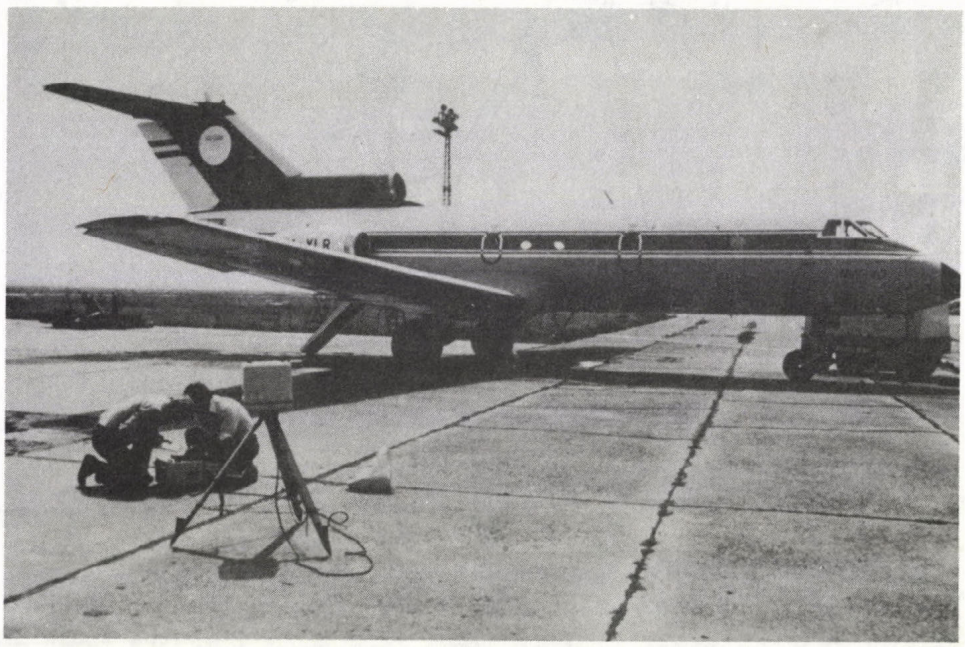


Fig. 3. HPIC in position of field measurement in front of the aircraft used for measuring cosmic dose rates in air at various altitudes

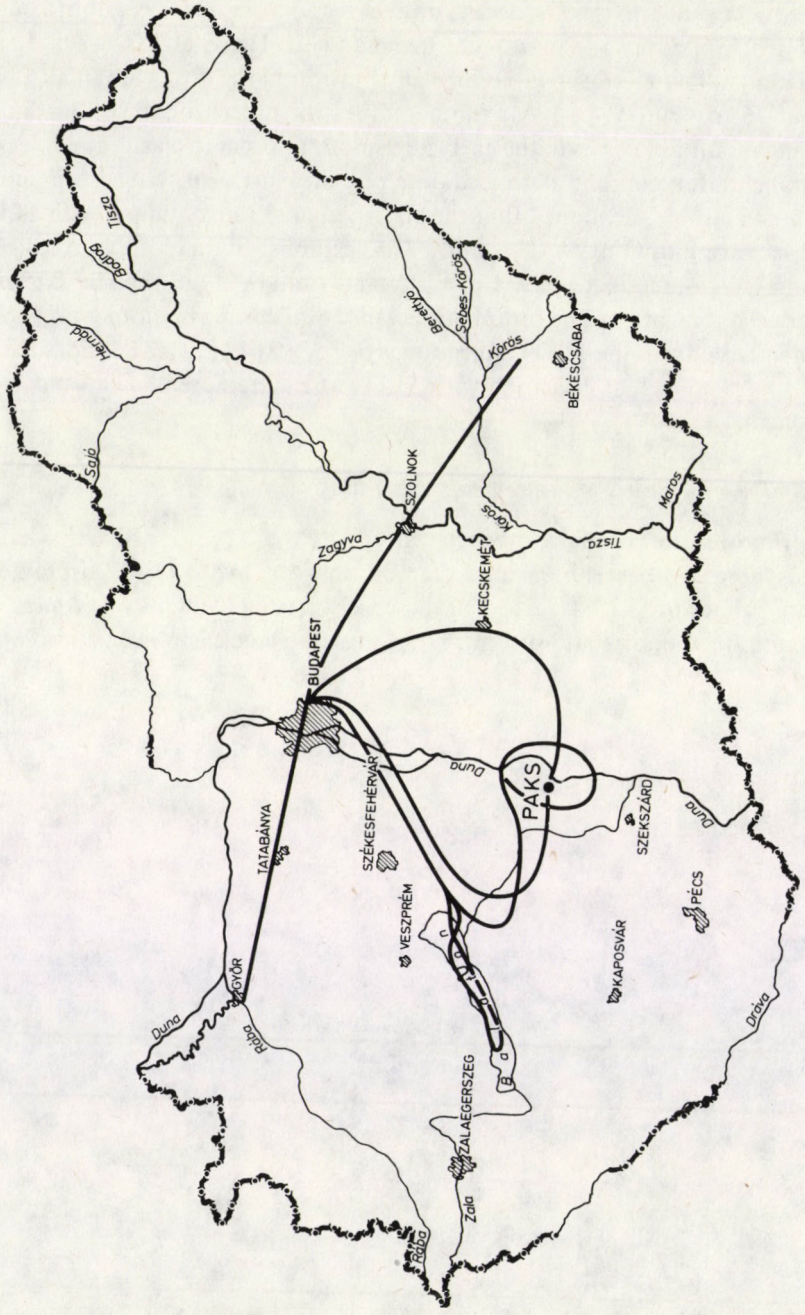


Fig. 4. Map of the country with the routes of the aircraft flown during the measurements of cosmic dose rates in air

Table III

Variation of cosmic dose rate with altitude in the lower atmosphere*

Height above sea level [m]	Number of data registered	Dose rate in air, measured [nGy · h^{-1}]	Dose rate obtained from regression curve [nGy · h^{-1}]
0		37.1 ± 3.6**	34.2
900	200	52.5 ± 4.3	50.3
1 110	750	56.5 ± 2.9	55.0
1 910	200	72.4 ± 4.5	77.3
2 940	900	114.1 ± 5.8	119.9
4 100	250	200.5 ± 3.2	196.6
4 600	750	251.4 ± 11.5	243.3
4 750	200	257.7 ± 11.2	259.4

Notes:

* These measurements were performed with HPIC on board an aircraft on 9th and 15th April, and 26th July 1979.

** Overall average of dose rates measured on the surface of Lake Balaton in 1978 and 1979. This value was omitted when fitting the data.

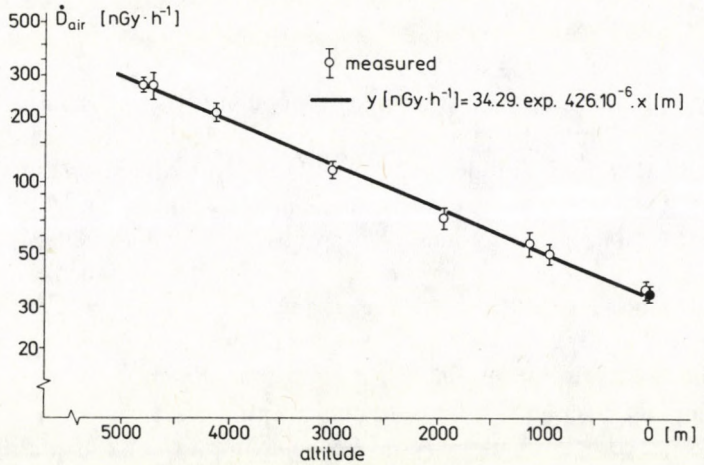


Fig. 5. Change of cosmic ray absorbed dose rate in air with altitude

construction at Paks, and along the main civil air travel service (ATS) routes of our country (Fig. 4). The measurements were performed with HPIC at altitudes between 900 and 4750 m (Table III). No additional radiation from the structural elements of the aircraft could be detected on board at the point the HPIC was fixed. The absorbed dose rate values measured have been fitted and a regression curve obtained (Fig. 5). The correlation coefficient of this curve is 0.99. It is worth noting that the absorbed dose value obtained from the regression curve for sea level is 34.3 $nGy \cdot h^{-1}$, that is, somewhat below the measured average.

4. Average dose equivalent rate to the population from cosmic rays

The stopping power of muons and associated fast electrons in the body is barely different from that in air. Therefore, the air absorbed dose rate can be taken without modification to infer the absorbed dose rate of cosmic radiation to a person in the outdoor environment. The mean quality factor of muons and electrons is also close to unity, so that the same figures may be used for calculating the cosmic ray charged particle dose equivalent rate. Protons constitute less than one per cent of the charged particles in the lower atmosphere [2].

Assuming that the outdoor absorbed dose rate at sea level varies between 34.3 $nGy \cdot h^{-1}$ (as a lower bound value that follows from the regression curve of dose rates measured at various altitudes) and 37.1 $nGy \cdot h^{-1}$ (as an upper bound value which is the overall average of HPIC measurements on the surface of Lake Balaton), the annual absorbed dose equivalent in human tissues from ionizing components of cosmic rays at sea level is 300—325 μSv.

The average annual dose equivalent to the population from cosmic radiation can be estimated by combining the data on the absorbed dose rate in air (or rather the dose equivalent rates) at various altitudes with the altitudinal distribution of the population. Unfortunately, the latter is not readily available in our country. Therefore, the rough fractional distribution of the national territory has been used as first approximation (Table IV). The annual dose equivalent from cosmic radiation averaged over the whole population is found to be 320 μSv. Citizens living in the highest part of the country (Kékestető, Mt. Mátra) receive about 50 per cent higher doses from cosmic rays annually than those who live on the Great Hungarian Plain.

In reality, the average annual dose equivalent is somewhat higher than given above, due to the contribution of cosmic ray neutrons. Taking the quality factor of these neutrons to be 5 or 6 as given in the literature [2, 7], the annual dose equivalent is to be increased by 18—21 μSv (that is 5—7 per cent). On the other hand, there is another factor which results in a decrease of the actual exposure. Individuals spending a large part of their time in buildings are exposed to lower cosmic radiation levels than the outdoor exposure level, because of the effect of structural shielding. A 10 per cent or higher structural attenuation factor should be applied depending on the thickness of

Table IV

Cosmic ray dose rates at various altitudes and approximate fractions of the national territory affected

Height above sea level [m]	Average cosmic ray dose rate in air or tissue		Approximate fraction of the national territory [%]
	[$nGy \cdot h^{-1}$]	[$\mu Gy \cdot y^{-1}$]	
0—200	35.8	313	79
200—400	39.0	341	18
400—600	42.2	371	2
600—800	46.2	405	0.5
800—1000	50.3	441	0.3

overhead shielding. This effect is due to be considered for a future publication dealing with exposure of the population to natural radiation sources inside and outside buildings.

Acknowledgements

The authors are indebted to all those who helped in accomplishing this programme. Mr. K. Zsdánszky and J. Hizó of the National Bureau of Standards assisted in receiving a HPIC on loan from the Dosimetry Laboratory of the Österreichische Studiengesellschaft für Atomenergie GmbH, headed by Mr. K. E. Duftschmid. Water-craft were provided by the Biology Research Institute of the Hungarian Academy of Sciences, Tihany, and the Balatonfüred Unit of the Hungarian Ship and Crane Works. Measurements on board an aircraft were made possible by the Civil Aviation Directorate of the Ministry of Transport and Posts.

Particular thanks are due to the crew of the ship and aircraft and to several co-workers of the authors' institute who were willing to participate in the field measurements, sometimes even under unfavourable and risky conditions.

References

1. H. Ohlsen, Bestimmung der mittleren Bevölkerungsbelastung durch natürliche äussere Strahlung auf dem Gebiet der DDR. Report SZS-14/69. Staatliche Zentrale für Strahlenschutz, 1969.
2. Natural Background Radiation in the United States. NCRP Report No. 45. National Council on Radiation Protection and Measurements, Washington, D.C., 1975.
3. Strahlenkarte Österreichs. Mittlere Bevölkerungsdosen im Freien durch terrestrische und kosmische Strahlung. Bundesministerium für Gesundheit und Umweltschutz, Wien, 1975.
4. Die Strahlenexposition von aussen in der Bundesrepublik Deutschland durch natürliche radioaktive Stoffe im Freien und in Wohnungen unter Berücksichtigung des Einflusses von Baustoffen. Der Bundesminister des Inneren, Bonn, 1978.
5. F. E. Taylor and G. A. M. Webb, Radiation Exposure of the UK Population, NRPB-R77, Harwell, 1978.
6. Ionizing Radiation: Levels and Effects. UNSCEAR 1972 Report, United Nations, New York, 1972.
7. Sources and Effects of Ionizing Radiation. UNSCEAR 1977 Report, United Nations, New York, 1977.

8. Present level of environmental radiation and radioactivity in the environs of Paks Nuclear Power Station. Status report. Ed. L. B. Sztanyik et al. NIMDOK-81450, Budapest, 1980. (in Hungarian).
9. I. Nikl and L. B. Sztanyik, 10th Conference on Biophysics, Tihany, 20—22 September, 1979. Abstracts of Papers. Akadémiai Kiadó, Budapest, 1979, p. 42.
10. I. Nikl and L. B. Sztanyik, International Workshop on Environmental Monitoring around Nuclear Installations. Dobogókő, Hungary, 9—11 September, 1980. Abstracts of Papers, p. 24.
11. L. Jánossy, Cosmic Radiation, Művelt Nép Könyvkiadó, Budapest, 1954 (in Hungarian).
12. G. Benkó et al., The 1979—80 Yearbook of the Central Research Institute for Physics, Budapest, 1980, pp. 9—11.
13. A. Somogyi et al., High Energy Particles in the Surrounding of the Sun and in the Interplanetary Space. Presented at the "Special Session of 10 Years of Hungarian Space Research" Budapest, 2—3 June, 1981 (in Hungarian).
14. EML Procedures Manual. HASL-300, C-01. DOE Environmental Measurements Laboratory, New York, 1981.
15. EML Procedures Manual. HASL-300, C-04. DOE Environmental Measurements Laboratory, New York, 1981.
16. I. Nikl and I. Kása, III. National Summer School on Luminescence. Abstracts of Papers. Spectroscopy Subcommittee of the Hung. Acad. Sci., Budapest, 1980, p. 196.
17. Technical and Sanitary Radiation Protection. Ed. J. Bonta. Műszaki Könyvkiadó, Budapest, 1962. (in Hungarian).
18. M. J. George, J. Geophys. Res., **71,** 3693, 1970.
19. K. Katsurayama et al., Health Phys., **36,** 558. 1979.
20. I. R. McAulay and P. A. Colgan, Health Phys., **39,** 821, 1980.
21. I. Kurcz-Csiky (personal communication).
22. K. O'Brien, J. Appl. Rad. a. Isotopes, **29,** 735, 1978.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 201—216 (1982)

THE DOSE GRADIENT VECTOR CONCEPT AND ITS USEFULNESS IN PLANNING DOSE DISTRIBUTIONS FOR RADIATION TELETHERAPY*

G. VARJAS

CLINLCAL INSTITUTE OF THE NATIONAL ONCOLOGICAL INSTITUTE
BUDAPEST, HUNGARY

and

O. HAIMAN

INSTITUTE FOR ATOMIC PHYSICS OF THE ROLAND EÖTVÖS UNIVERSITY,
BUDAPEST, HUNGARY

The method of characterizing scalar fields by their gradient vectors rather than by the scalar values themselves is used by the authors to define conditions that must be fulfilled by radiation fields used in the radio-teletherapy of malignant tumours. The usefulness of some general relations formulated in terms of dose gradient vectors is stressed and beams of currently used irradiation units described in these terms.

Using a few very simple and easily visualized rules, the general lay-out of an irradiation field for the treatment of lung tumours is sketched. Subsequent computer aided planning yielded a dose distribution which met the basic requirements of therapy.

Introduction

The treatment of patients suffering from malignant tumours is one of the most urgent concerns of public health, and work aiming at the solution of its problems is given high priority. At present complex tumour treatment is usually applied, in which radiotherapy plays a major role. The improvement of devices and methods for optimally applying ionizing radiations for this purpose is, therefore, of great importance and timeliness.

The treatment of tumours situated in various parts of the human body requires different irradiation fields which are planned bearing in mind the main aim of delivering a dose as high as needed to destroy the tumour and its immediate surroundings (the target volume), at the same time keeping adverse radiation effects in adjacent healthy tissues at a minimum level.

Radiation therapeutical dose distributions should, accordingly, be characterized by 1. a high and homogeneous dose intensity within the target area, and 2. a fast decrease of dose intensity outside the target area, the decrease being especially abrupt

* Dedicated to Prof. I. Tarján on his 70th birthday.

in directions pointing to organs with high radiation sensitivity (spinal cord, kidneys, eyes, lung, etc.) and in directions where large body volumes are present.

In trying to comply with the above demands, the radiotherapeutist-physicist has quite a choice of radiation sources (and their configurations) at his disposal, most of which can be treated using quite simple mathematics in the initial step of determining the general lay-out.

1. Dose and dose gradient vector distributions of point sources

The absorbed radiation dose in a homogeneous medium surrounding a point source (Fig. 1), as a function of distance **r**, is given by the formula (valid only for radiation equilibrium)

$$D = c\frac{e^{-\mu r}}{r^2}, \tag{1}$$

where c is a constant whose value depends upon both the kind of radiation and the medium and μ is a similar constant called the attenuation coefficient.

Dose distributions are scalar-vector functions, the gradient vector of which is easily found, e.g. for the field given by (1):

$$[\text{grad } D(\mathbf{r})]_k = \frac{\partial}{\partial x_k} c\frac{e^{-\mu r}}{r^2}, \tag{2}$$

where

$$k = 1, 2, 3; \qquad \mathbf{r} = \mathbf{r}(x_1, x_2, x_3) \tag{3}$$

and

$$r = \sqrt{x_1^2 + x_2^2 + x_3^2}.$$

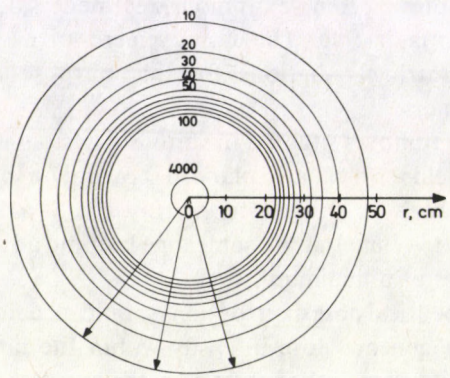


Fig. 1. Distribution of doses (in arbitrary units) around a point radiation source. Medium: water, $\mu = 0.03\,\text{cm}^{-1}$. Isodose curves are drawn in 10 unit intervals between 10 and 100. Radial straight lines represent directions of fastest dose decrease

Carrying out the differentiation, we get

$$[\text{grad } D(\mathbf{r})]_k = \frac{\partial}{\partial r}\left(\frac{ce^{-\mu r}}{r^2}\right)\frac{\partial r}{\partial x_k} = \left(-\frac{c\mu e^{-\mu r}}{r^2} - 2\frac{ce^{-\mu r}}{r^3}\right)\frac{2x_k}{2r} \tag{4}$$

which may be summarized by writing

$$\text{grad } D(\mathbf{r}) = -c\frac{e^{-\mu r}}{r^2}\left(\mu + \frac{2}{r}\right)\frac{\mathbf{r}}{r}. \tag{5}$$

This gradient vector is seen to point toward the source. It has, however, become common usage [14, 15, 16, 17] to define the dose gradient vector in radiation therapy as the opposite vector, pointing in the direction of fastest decrease.

Let us divide the module of the dose gradient vector by the dose value

$$\frac{|\text{grad } D(\mathbf{r})|}{D} = \mu + \frac{2}{r}. \tag{6}$$

This quotient becomes, for

$$\mu \gg \frac{2}{r} \quad \text{i.e. for distances} \quad r \gg \frac{2}{\mu} \tag{7}$$

$$\frac{|\text{grad } D(\mathbf{r})|}{D} \approx \mu, \tag{8}$$

which means that the module of the newly defined dose gradient vector and the dose itself show the same spatial distribution of distances

$$r > 70\text{ cm} \quad \text{for} \quad \mu = 0.03\text{ cm}^{-1}; \tag{9a}$$

$$r > 40\text{ cm} \quad \text{for} \quad \mu = 0.05\text{ cm}^{-1}. \tag{9b}$$

(7) is usually satisfied in teletherapy work. Teletherapy might even have been defined as satisfying the inequalities (7) and consequently (8), too.

Equation (6) can be shown to be an important relation helping to determine, among others, the degree of dose distribution deformation caused by oblique incidence [11] and by inhomogeneities in medium composition or density (e.g. lung, muscle, bone) [27].

Let us next investigate the simplest dose distribution brought about by simultaneously (or consecutively) using two point sources (Fig. 2).

The dose at any point in this field is obtained by simply adding two dose values (scalars) corresponding to the two point sources, each radiating for itself. The resulting

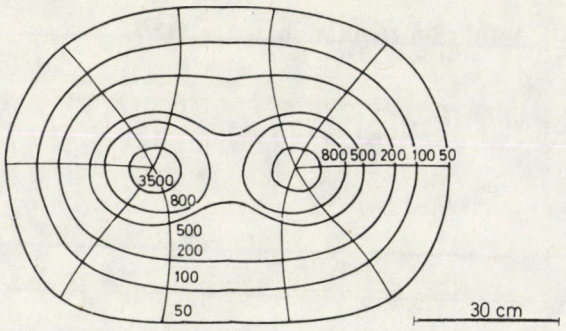


Fig. 2. Distribution of doses (in arbitrary units) around two point sources (source separation 30 cm). Medium: water, $\mu = 0.03\ \mathrm{cm^2 g^{-1}}$. Curves originating in the source points are obtained by proceeding in the direction of fastest dose decrease

dose gradient vector, however, must be determined by adding three pairs of components. The result of this vectorial addition is

$$\operatorname{grad} D(\mathbf{r}) = -\left[\left(\frac{\partial D_{\mathrm{I}}}{\partial x_1} + \frac{\partial D_{\mathrm{II}}}{\partial x_1}\right)\mathbf{i} + \left(\frac{D_{\mathrm{I}}}{\partial x_2} + \frac{\partial D_{\mathrm{II}}}{\partial x_2}\right)\mathbf{j} + \right.$$
$$\left. + \left(\frac{\partial D_{\mathrm{I}}}{\partial x_3} + \frac{\partial D_{\mathrm{II}}}{\partial x_3}\right)\mathbf{k}\right] = -\left[\frac{\partial D}{\partial x_1}\mathbf{i} + \frac{\partial D}{\partial x_2}\mathbf{j} + \frac{\partial D}{\partial x_3}\mathbf{k}\right], \tag{10}$$

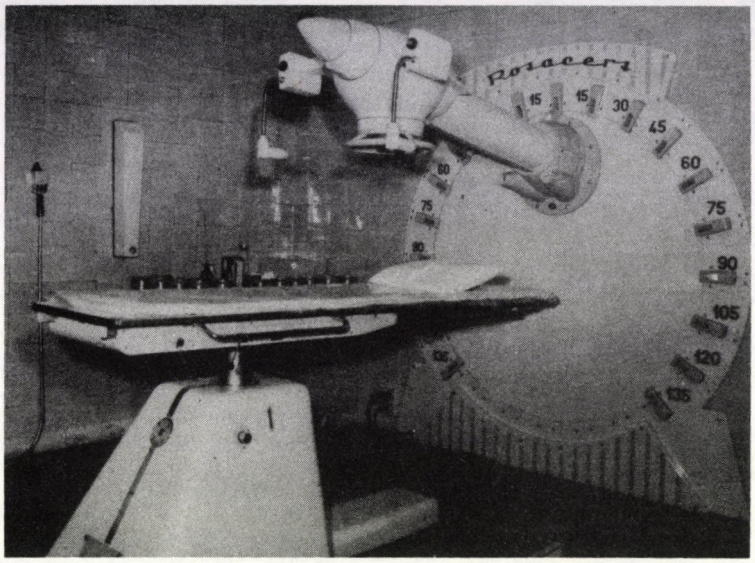
where $D = D_{\mathrm{I}} + D_{\mathrm{II}}$.

Fig. 2 shows that isodose curves and dose gradient vector chains or "lines" strongly resemble field maps with their equipotential curves around point charges and their "lines of force". The analogy, of course, is not a perfect one, since radiation fields are not source-free in between radiation point sources.

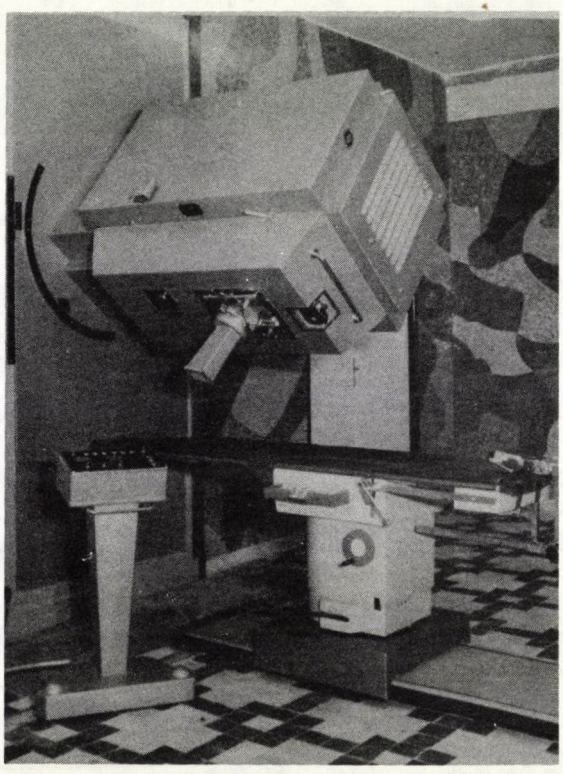
It is, however, not only possible, but often very helpful to conceive radiation fields used is radiotherapy in terms of their vector gradient distributions rather than their dose distributions. The obvious reason for this is that many of the requirements of radiotherapy are formulated — as we did in the Introduction — in terms of spatial dose variations rather than in terms of dose values.

2. Dose and dose gradient vector distributions in single fields of currently used radiation sources

Most requirements of radiotherapy cannot in general be fulfilled by using one or several point sources with fields as simple as to be characterized by Eq. (1). However, the evaluation [30] of clinical experience and usage shows that if the main features of typical dose distributions in beams of currently used radiation sources are clearly recognized (see later in this Section) and beams (or sources) appropriately combined



a)



b)

Fig. 3. Teletherapy apparatus 3a. Cobalt unit, ROTACERT type, mfg. by MEDICOR, Budapest, 3b. Betatron, B5M25 type, made in USSR

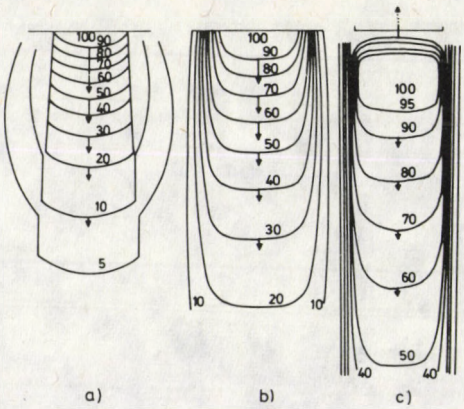


Fig. 4. Dose distributions in 6 × 8 cm beams of various teletherapy apparatus. Isodose curves have been plotted in a plane containing the central beam, parallel to the 6 cm side. Dose gradient vectors are represented by arrows whose length is proportional to their module. 4a. 200 kV X-ray (FSD = 50 cm, HVL = 1 mm Cu). 4b. Cobalt (SSD = 60 cm). 4c. 29 MV X-ray (FSD = 100 cm)

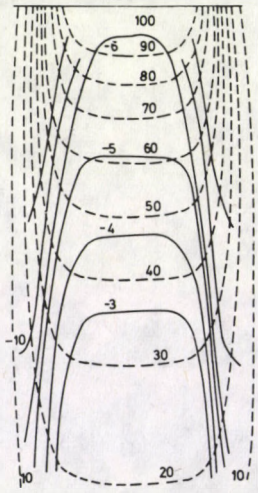


Fig. 5. Isodose curves (dashed lines) and isogradient curves (solid lines) in the radiation field of a cobalt beam

(e.g., see Section III), it is often not too difficult to figure out a nearly optimal lay-out for many specific purposes. We shall endeavour to show that this work is largely helped by using the dose gradient vector concept.

Let us recall some features of almost all beams used in teletherapy:

1. The limitation of beams by diaphragms in order to reduce needless irradiation of large body volumes. This collimation brings about typical dose distribution profiles in planes perpendicular to the central beams.

2. The presence of air between the source and the surface of the irradiated human body. With certain radiations, and near medium boundaries, scattering and dose build-

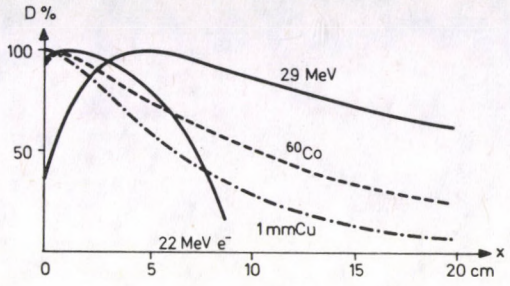


Fig. 6. Dose distributions (per cent depth-dose curves) along the central beams of the radiations listed in Fig. 4 and of 22 MV electron radiation (FSD = 90 cm, field size 12 × 12 cm)

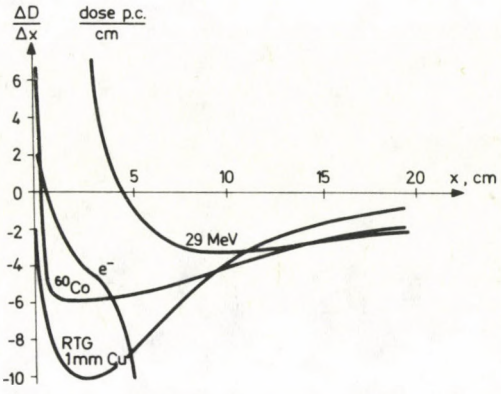


Fig. 7. Plots of $\Delta D/\Delta x$ values (see Table I) vs depth for the same beams as in Fig. 6

up effects often lead to transients in the dose vs depth plots taken along the central beam; "attenuation" curves may show maxima before the transition into ranges of monotonic decrease.

3. The application of wedges protruding into the beams. Wedging results in the tilting of isodose surfaces (or curves in two-dimensional plots) and is frequently used to compensate for the obliqueness of incident beams and/or the curvedness of body surfaces.

Let us next study typical dose distributions in beams of three radiation sources frequently applied in therapeutical work at the Oncological Institute: a 200 kV X-ray apparatus, a cobalt unit (Fig. 3a), and a betatron (Fig. 3b).

Dose distributions in body tissue irradiated with these sources are shown in Fig. 4. Fig. 5 shows, in addition to isodose curves obtained with our cobalt unit beam, a set of isogradient curves in the same field.

Plots of dose distributions along the central beams (depth-dose curves) are presented in Fig. 6, whereas Fig. 7 shows gradient values plotted against depth for the same kinds of teletherapy beams. Fig. 8 shows the dependence of gradient vs depth

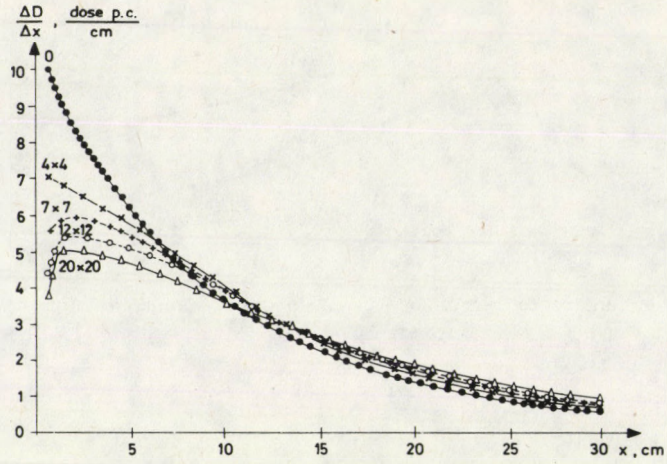


Fig. 8. Plots of $\Delta D/\Delta x$ values (see Table II) vs depth for various cobalt field sizes (SSD = 60 cm)

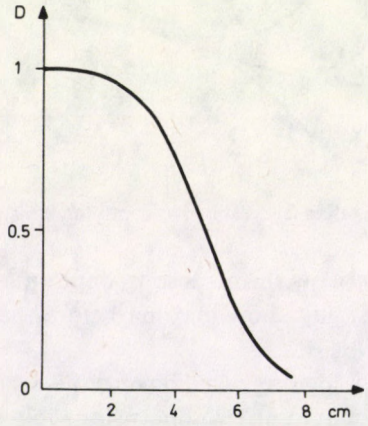


Fig. 9. Dose distributions obtained within a tissue equivalent phantom in a depth of 10 cm under irradiation with a 6 × 6 cm cobalt ray field. Relative dose values were measured along a line perpendicular to the central ray and parallel to one of the 6 cm sides

curves on field size for Co radiation. Fig. 9 gives an example of a dose profile in a depth of 10 cm plotted perpendicularly to the central beam, proceeding outwards.

Betatron data and the cobalt profile shown in Fig. 9 were obtained by Reischl [23], and all other data relating to cobalt radiation by Bozóky [7]. Isodose curves for 200 kV X-ray beams were taken from the Tsien atlas [26] and evaluated graphically to yield dose gradient values. Also used were data published in [8].

For 200 kV X-rays, cobalt γ-radiation and betatron 29 mV X-rays depth dose curves can be approximated [18, 19, 29] by the formula

$$D = D_0 \left(\frac{f + x_0}{f + x}\right)^2 [C_\mu \exp(-\mu x) - C_{\mu s} \exp(-\mu_s x)], \tag{11}$$

where D_0 is a normalization factor, f is the source—surface distance (SSD), x_0 the depth below the surface where maximum dose intensity is measured. μ and μ_s are empirical coefficients related to primary and scattered radiations, C_μ and $C_{\mu s}$ likewise.

Typical values of these constants for the three kinds of radiation listed below (field size 6 × 8 cm; beam incidence perpendicular to tissue equivalent phantom surface) [29] are:

200 kV X-ray beam (HVL = 1 mm Cu, FSD = 50 cm, $f = 50$ cm, $x_0 = 0$ cm):

$$D_0 = 3.68 \text{ cGy}, \ \mu = 0.14 \text{ cm}^{-1}, \ \mu_s = 0.53 \text{ cm}^{-1},$$

$$C_\mu = 36.4, \ C_{\mu s} = 9.19.$$

Cobalt γ-radiation beam (SSD = 60 cm, $f = 60$ cm, $x_0 = 0.5$ cm):

$$D_0 = 10 \text{ cGy}, \ \mu = 0.047 \text{ cm}^{-1}, \ \mu_s = 5.0 \text{ cm}^{-1},$$

$$C_\mu = 10.5, \ C_{\mu s} = 2.2.$$

29 MV X-ray beam (FSD = 100 cm, $f = 100$ cm, $x_0 = 4.7$ cm):

$$D_0 = 10 \text{ cGy}, \ \mu = 0.018 \text{ cm}^{-1}, \ \mu_s = 0.517 \text{ cm}^{-1},$$

$$C_\mu = 11.7, \ C_{\mu s} = 8.5.$$

Electron beam radiation gives rise to another type of depth dose curve [1, 20, 21] which, for $10 \exp(\mu x) < 110$, may be approximated by the formula

$$D = 110 - 10 \exp(\mu x). \tag{12}$$

Specifically, for 22 mV electron beams $\mu = 0.25 \text{ cm}^{-1}$.

For cobalt radiation, dose distribution profiles along a line perpendicular to the central beam and parallel to any side of the rectangle field cross section are approximately given by

$$D = \exp(-(A|y|)^B), \tag{13}$$

where y is the distance from the central beam, A and B are empirically determined constants whose numerical values for our ROTACERT cobalt unit, 6 × 6 cm field size, and 10 cm depth are $A = 0.0012 \text{ cm}^{-1/B}$ and $B = 4.06$.

An important phenomenon in teletherapy is the deformation of isodose curves resulting from oblique incidence of beams in air-tissue boundaries. Isodose curves are

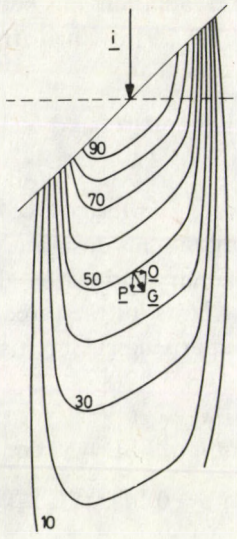


Fig. 10. Dose distribution obtained within a tissue equivalent phantom. Note the obliqueness of incident beam **i** with respect to the plane separating free air and tissue. Dose gradient vectors **G** are shown as superpositions of primary dose gradients **P** and of gradients **O** due to oblique incidence

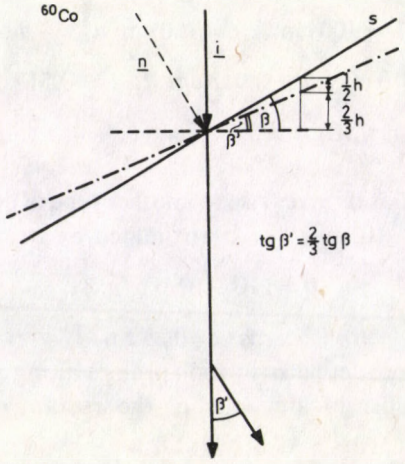


Fig. 11. Obliqueness β' of dose gradient vectors with respect to central beams, as derived from the obliqueness β of incidence with respect to the normal n of surface s: $\operatorname{tg}\beta' = F \operatorname{tg}\beta$, the specific F value for cobalt being 2/3

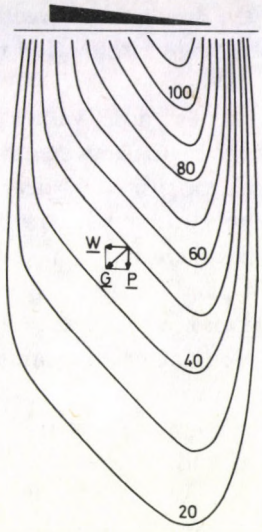


Fig. 12. Dose distribution in a wedged cobalt beam. Dose gradient vectors are shown as superpositions of primary dose distribution gradients **P** and of gradients due to wedging **W**

shifted in the direction of the beam (or in opposite direction), where the surface of the body is below (or above) the plane drawn perpendicularly to the central beam through the point of incidence (Fig. 10).

Dose gradient vectors in the deformed distribution are no longer pointing in the direction of the beam but tilted by an angle β' (Fig. 10 and Fig. 11). This angle may be calculated using the formula

$$\operatorname{tg} \beta' = F \operatorname{tg} \beta \,, \tag{14}$$

where β is the angle between the normal of the body surface (or its tangent plane) and the beam incident on it. F is called the shift factor. Its value for cobalt radiation is about 2/3 [11]. It may be shown that Eq. (14) holds only in ranges where $\frac{|\operatorname{grad} D(\mathbf{r})|}{D}$ does not change appreciably in the direction of the beam.

If it so happens that obliqueness of incidence, curvedness of body surface or other factors result in dose distributions with gradients pointing in the wrong (wrong in the therapeutical sense) direction it is possible to tilt them back by wedging (Fig. 12). If, specifically, the dose gradient vector **G** in Fig. 11 is to point in the beam direction again, a wedge of angle β' (see Eq. (14)) is needed.

3. The use of the dose gradient vector concept in the planning of irradiation fields

Fig. 2 exemplifies that highest dose values are obtained and best dose homogeneity is accomplished in the region between the two point sources where dose gradient vectors are, by and large, opposite and nearly cancel.

The general idea is accordingly, to apply several beams crossing in the target area in order to obtain high dose values there and to have the dose gradient vectors of the single beams cancel either by pairs or after overall summation. This may be accomplished by an appropriate choice of radiation sources, beam directions, beam weight factors, and by suitable wedging of the beams.

In the following, we give an example for irradiation field planning based on considerations relating to the dose gradient vectors in this field.

Fig. 13 illustrates the beams and the field used for irradiating a lung tumour. In view of the high radiation sensitivity of healthy half lung and the spinal cord it is not permissible to apply irradiation through these organs. Although the target volume is rather centrally located, its irradiation cannot be performed in the traditional way of using symmetrical beams crossing each other; one beam which would pass through the healthy half lung has to be omitted. This means that one of the four dose gradient vectors which usually cancel by pairs is no more present, and dose homogeneity no more warranted.

In order to re-establish homogeneity, one or two beams have to be modified in a way which brings about a dose gradient vector which, roughly speaking, cancels the uneven gradient vector. This is accomplished by using wedges. Let us calculate angles and orientations of wedges needed for cancelling.

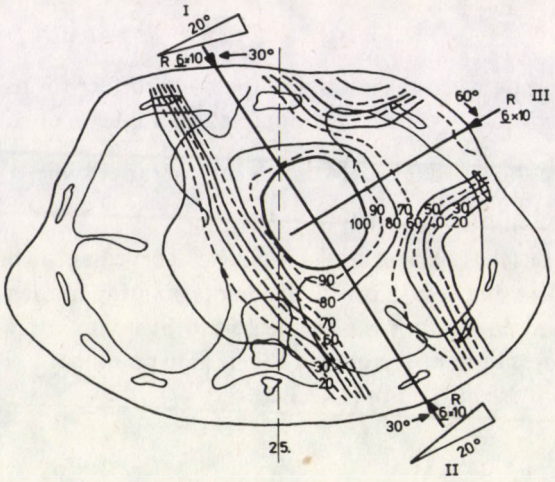


Fig. 13. Irradiation of a lung tumour with three 6 × 10 cm cobalt fields. *R* stands for our ROTACERT cobalt unit (see Fig. 3a); SSD = 60 cm. One 30° and one 20° wedge applied

At the entry point of beam I, the angle of incidence is about 30°. The resulting obliqueness deformation of the dose distribution might, if necessary, be compensated with an edge filter of angle β' where

$$\operatorname{tg}\beta' = \frac{2}{3}\operatorname{tg} 30° \quad \text{and} \quad \beta' \approx 20° .$$

The crossing point of central beams I, II and III lies about 12 cm below the entry point of central beam I. Evaluating Fig. 8*, we get, for the dose gradient vector at $x = 12$ cm

$$|\mathbf{G}_\mathrm{I}| = 3.5 \text{ cGy/cm} .$$

The angle of incidence for beam II is about 20°. By a calculation similar to the foregoing, we obtain a wedge angle of about 15° which would be needed to compensate the obliqueness effect.

The crossing point of the three central beams lies about 14.5 cm below the entry point of central beam II, but 6 cm of lung tissue (0.3 g/cm^3) is between. According to literature [9, 25, 27] the shift factor for this kind of radiation is about 0.4. Isodose curves must therefore be shifted by about $0.4 \cdot 6 \text{ cm} \approx 2.5$ cm after the 6 cm of lung tissue. This is equivalent to saying that the beam has to pass through 14.5 minus 2.5 cm = 12 cm of unit density body tissue, bringing about, at the crossing point, a dose gradient vector of absolute value

$$|\mathbf{G}_\mathrm{II}| = 3.5 \text{ cGy/cm}$$

which is opposite to the gradient vector of the wedged beam I. This means that the dose gradient vectors of beams I and II happen to cancel, there being no need for compensation by adjustment of weight factors.

Beam III is nearly perpendicular to the body surface at its entry point, thus it is not necessary to compensate for obliqueness by wedging.

The crossing point of the beams lies about 12 cm below the entry point of beam III. 5 cm of this being lung tissue, we have to subtract $0.4 \cdot 5 \text{ cm} = 2$ cm. Evaluating Fig. 8, we get for the dose gradient vector of beam III at the crossing point

$$|\mathbf{G}_\mathrm{III}| = 4 \text{ cGy/cm} .$$

As $\mathbf{G}_\mathrm{III}$ points in a direction perpendicular to beams I and II, its compensation requires filtering of one or both of these beams with wedges which bring about a resulting dose gradient vector opposite to it.

The evaluation of the parallelogram in Fig. 12 shows that if this compensation were to be accomplished with one wedge, its angle α would be given by

$$\operatorname{tg}\alpha = \frac{|\mathbf{W}|}{|\mathbf{G}_\mathrm{I}|} = \frac{4}{3.5} \quad \text{i.e. } \alpha \approx 50° .$$

* Since the values shown in Table II and plotted in Fig. 8 are dose percent values, the hypothetical value $|\mathbf{G}_\mathrm{I}| = 3.5$ cGy/cm refers to a beam of a source which delivers 100 cGy at the peak dose point.

Table I

Depth dose and dose gradient at some kind of teletherapy irradiation

X-ray HVL = 1 mm Cu			^{60}Co		29 MV X-ray		22 MeV electron beam	
Depth	Depth dose	Dose gradient	Depth dose	Dose gradient	Depth dose	Dose gradient	Depth dose	Dose gradient
cm	p.c.	dose p.c. / cm	p.c.	dose p.c. / cm	p.c.	dose p.c. / cm	p.c.	dose p.c. / cm
0	100.0		95		35		98	
		−5.6		+10		+30		+2
0.5	97.2		100.0		50		99	
		−8.2		−5.6		+34		0
1	93.1		97.2		67		99	
		−9.9		−5.9		+19		−2
2	83.2		91.3		86		97	
		−9.8		−5.9		+9		−4
3	73.4		85.4		95		93	
		−10.2		−5.8		+4		−5
4	63.2		79.6		99		88	
		−9.2		−5.5		+0.5		−7
5	54.0		74.1		99.5		81	
		−8.0		−5.5		−1.5		−12
6	46.0		68.6		98		69	
		−6.7		−5.0		−2		−14
7	39.3		63.6		96		55	
		−6.2		−5.0		−3		−19
8	33.1		58.6		93		36	
		−5.2		−4.6		−3		−24
9	27.9		54.0		90		12	
		−4.2		−4.4		−3		
10	23.7		49.6		87			
		−3.5		−3.8		−3		
11	20.2		45.8		84			
		−3.0		−3.6		−2.5		
12	17.2		42.2		81.5			
		−2.8		−3.4		−2.5		
13	14.4		38.8		78			
		−2.1		−3.0		−2.5		
14	12.3		35.8		75.5			
		−2.1		−2.8		−2.5		
15	10.2		33.0		73			
		−1.5		−2.6		−2.5		
16	8.7		30.4		70.5			
		−1.3		−2.4		−2.5		
17	7.4		28.0		68			
		−1.2		−2.1		−2.5		
18	6.2		25.9		65.5			
		−0.9		−2.0		−2		
19	5.3		23.9		63.5			
		−0.8		−1.9		−2		
20	4.5		22.0		61.5			

Now we have all data needed to determine what wedging must be used in beams I and II in order to ensure a total cancellation of dose gradient vectors at the crossing point of the three central beams.

Dose gradient vectors resulting from the obliqueness of beams I and II are opposite and nearly equal. The wedge angles needed for their compensation (each by itself) must therefore be subtracted and there remains an angle of $20° - 15° = 5°$.

The dose gradient vector due to beam III requires, for its compensation, a 50° wedge, as determined above.

In view of the fact that the resultant dose gradient vector of beams I and II is parallel to $\mathbf{G}_{III}$, we get a final value of $50° + 5° = 55°$ for the total wedging angle.

Table II
Absolute values of $\Delta D/\Delta x$

Field size [cm] Depth [cm]	0	4×4	5×5	6×6	7×7	8×8	10×10	12×12	15×15	20×20
0.5	10	7.0	6.4	6.0	6.5	5.2	4.6	4.4	4.0	3.8
1	11	6.8	6.5	6.2	5.9	5.7	5.6	5.4	5.8	5.1
2	8.1	6.5	6.3	6.0	5.9	5.7	5.5	5.4	5.2	5.0
3	7.2	6.2	6.0	5.9	5.8	5.7	5.4	5.2	5.1	4.9
4	6.5	6.0	5.8	5.7	5.5	5.5	5.3	5.2	5.0	4.8
5	5.9	5.6	5.6	5.5	5.5	5.3	5.2	5.0	4.9	4.6
6	5.3	5.3	5.2	5.2	5.0	5.1	5.1	5.0	4.7	4.5
7	4.8	5.0	5.1	4.9	5.0	4.9	4.8	4.8	4.5	4.3
8	4.3	4.7	4.5	4.6	4.6	4.6	4.5	4.3	4.3	4.0
9	4.0	4.3	4.4	4.3	4.4	4.3	4.2	4.1	4.1	3.9
10	3.6	3.9	3.9	3.9	3.8	3.9	3.9	3.9	3.9	3.6
11	3.2	3.5	3.6	3.6	3.6	3.6	3.6	3.6	3.6	3.4
12	2.9	3.2	3.2	3.3	3.4	3.3	3.3	3.3	3.2	3.1
13	2.7	3.0	3.1	3.1	3.0	3.1	3.1	3.0	3.0	3.0
14	2.4	2.7	2.6	2.8	2.8	2.8	2.8	2.8	2.8	2.8
15	2.2	2.	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6
16	2.0	2.2	2.2	2.3	2.4	2.4	2.4	2.4	2.4	2.4
17	1.8	2.0	2.0	2.1	2.1	2.2	2.2	2.2	2.2	2.2
18	1.7	1.9	2.1	2.0	2.0	2.0	2.0	2.0	2.1	2.1
19	1.5	1.8	1.8	1.9	1.9	1.9	2.0	2.0	2.0	2.0
20	1.4	1.6	1.6	1.6	1.7	1.7	1.7	1.7	1.7	1.9
22	1.1	1.3	1.3	1.4	1.4	1.4	1.4	1.5	1.5	1.6
24	0.9	1.1	1.1	1.1	1.2	1.2	1.3	1.3	1.3	1.4
26	0.8	0.9	1.0	1.0	1.0	1.0	1.1	1.1	1.2	1.2
28	0.6	0.8	0.8	0.8	0.8	0.8	0.9	1.0	1.0	1.0
30										

In Fig. 13 the compensation is shown to be accomplished by using two wedges of angles 30° and 20°, respectively. This choice is motivated by the following fact: The Van de Geijn program [5, 10] fed into the Honeywell/Bull 60/66 computer of the National Computerized Irradiation Planning System [6] provides for only three wedge angles of 20°, 30° and 45°, respectively, which are used at the Oncological Institute.

Starting from basic irradiation field data as worked out above and data relating to body contour, to contours and densities of inhomogeneities within the body as well as so-called fix data relating to our irradiation facilities, the computer determined the dose distribution shown in Fig. 13.

It may be seen that this dose distribution fulfils the requirements outlined in the Introduction: dose values in the target volume are high, homogeneous, and fall off quickly in the directions of highly radiation sensitive organs and of large body volumes.

References

1. H. Azhigaev, Primenenie betatrona v onkologii, Izdatel'stvo Kazakhstan, Alma-Ata, 1977.
2. L. Bozóky, Strahlentherapie, **112,** 629, 1960.
3. L. Bozóky, Fizikai Szemle, **19,** 352, 1969 (in Hungarian).
4. L. Bozóky, Fizikai Szemle, **21,** 59, 1971 (in Hungarian).
5. L. Bozóky, B. Kanyár, G. Varjas, J. Zábráczky, Magyar Radiológia, **27,** 368, 1975 (in Hungarian).
6. L. Bozóky, G. József, L. Gyarmathy, B. Kanyár, Zs. Makra, J. Petrányi, Gy. Reischl, G. Varjas, Számítógépes besugárzás tervezési országos hálózat (Országos Onkológiai Intézetben rendezett tanfolyam jegyzete), Budapest, 1978. (Lecture notes of computerized treatment planning network; in Hungarian).
7. L. Bozóky, private communication.
8. Brit. J. Radiol., Suppl. No. 10, 1961.
9. D. Greene and J. R. Stewart, Brit. J. Radiol., **38,** 378, 1965.
10. J. van de Geijn, Comp. Progr. Biomed., **1,** 47, 1970.
11. P. H. Van der Giessen, Brit. J. Radiol., **46,** 978, 1973.
12. L. Gyarmathy, Gy. Reischl, Magyar Biofizikai Társaság Értesítője, 1974 (Yearbook of Hungarian Biophysical Society; in Hungarian).
13. Gy. Gyenes, Thesis, 1968.
14. M. D. Hamilton, J. Laurie and J. S. Orr, Brit. J. Radiol., **41,** 687, 1968.
15. C. S. Hope and J. S. Orr, Phys. Med. Biol., **10,** 365, 1965.
16. C. S. Hope and O. Cain, Comp. Progr. Biomed., **2,** 221, 1972.
17. ICRU Report 24, Washington, 1976.
18. H. E. Johns, E. K. Darby, R. N. N. Halsam, L. Katz and E. L. Harrington, Am. J. Roentgenol., **62,** 257, 1949.
19. J. S. Laughlin, J. W. Beattie, J. E. Lindsay and R. S. Harvey, Am. J. Roentgenol., **65,** 787, 1951.
20. J. S. Laughlin, J. Ovadia, J. W. Beattie, W. J. Henderson, R. S. Harvey and L. L. Haas, Radiology, **60,** 165, 1953.
21. J. S. Laughlin, Physical aspects of betatron therapy, C. C. Thomas, Springfield, 1954.
22. J. Petrányi, Thesis, 1980.
23. Gy. Reischl, private communication.
24. I. Rodé, L. Gyarmathy, J. Petrányi, Magyar Onkológia, **23,** 26, 1979 (in Hungarian).
25. L. Sundbom, Acta Radiol., **3,** 342, 1968.
26. K. C. Tsien and M. Cohen, Isodose Charts and Depth Dose Tables for Medium Energy X-rays, Butterworths, London, 1962.
27. U. D. Tripathy, Strahlentherapie, **156,** 164, 1980.
28. G. Varjas, Kórház és Orvostechnika, **18,** 175, 1980 (in Hungarian).
29. G. Varjas, G. József and G. Reischl, Proc. Symp. of Radiological Physicists, Bratislava, 1981.
30. G. Varjas, Orvos és Technika, **13,** 103, 1975 (in Hungarian).
31. R. J. Wilks and J. F. Sutcliffe, Phys. Med. Biol., **22,** 737, 1977.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 217—223 (1982)

A SEARCH FOR GIANT RESONANCES BY INELASTIC PROTON SCATTERING ON ^{24}Mg*

I. LOVAS[1], M. ROGGE, U. SCHWINN[2], P. TUREK and D. INGHAM[3]

INSTITUTE FOR NUCLEAR PHYSICS
NUCLEAR RESEARCH ESTABLISHMENT JÜLICH, D-517 JÜLICH, W. GERMANY

and

ERIKA LOVAS

CENTRAL RESEARCH INSTITUTE FOR PHYSICS
1525 BUDAPEST, HUNGARY

A straightforward method is proposed for the observation of giant resonances based on the phase shift analysis of the inelastic transition between a 0^+ ground state and a weakly excited 0^+ state. The method is tested by analysing the differential cross section of inelastic proton scattering leading to the 0^+ ($E_{ex}=6.43$ MeV) excited state of ^{24}Mg.

1. Introduction

One of the most effective tools for the excitation of giant resonances is inelastic hadron scattering. The various possibilities for the observation of giant resonances by means of inelastic scattering can be summarized as follows. In the course of the inelastic collision the target nucleus is excited to some high lying states and then the giant resonances show up in the energy spectrum of the inelastically scattered particles as pronounced maxima above a continuous background [1, 2, 3, 4]. In this case the resonating states are excited as the final states of the target nucleus. Another possibility is the excitation of the giant resonances as intermediate states of the compound system. In this case the excitation probability of a given low lying state is measured as a function of the bombarding energy and the maxima of this excitation function are identified with the giant resonances of the compound (target plus projectile) system. In order to increase the sensitivity of this method it is advisable to select such an inelastic transition where the direct excitation is hindered because of some specific reason. In

* Dedicated to Prof. I. Tarján on his 70th birthday.

Present address:

[1] Central Research Institute for Physics, H-1525 Budapest, Hungary.

[2] Deceased

[3] Now at MIT, Cambridge, Mass. USA

such a case the contribution of a two-step mechanism involving the high lying states as intermediate states, is relatively enhanced.

The direct excitation of the low lying non-normal parity states is hindered, because it may take place only by a spin-flip mechanism. Consequently, the excitation function of the non-normal parity states provides a rather sensitive tool for the study of the giant resonances [5, 6, 7, 8].

In this paper our aim is to explore another possibility for the study of giant resonances. Namely we note that in addition to the non-normal parity, a number of other reasons may lead to the suppression of the direct transition and the relative enhancement of the two-step mechanism. For example if the wave function of the excited state contains with appreciable amplitude many-particle many-hole excitations compared to the ground state then the one-step direct transition is obviously suppressed. If the inelastic cross section of a given low lying state is considerably lower than that of the neighbouring states having the same spin and parity, then it is justified to assume that the transition takes place predominantly by a multi-step mechanism involving highly excited states of the compound system. The measurement of the excitation function of such a weakly excited state provides the possibility to study the resonating high lying states. Of the weakly excited low lying states the 0^+ states are the most appropriate for such a study since in the $0^+ \rightarrow 0^+$ transition the partial waves can be associated uniquely to the multipolarity of the giant resonances. In order to test the above suggestion we analyse the inelastic proton scattering on ^{24}Mg leading to the 0^+ excited state, situated at $E_{ex} = 6.43$ MeV.

2. Experimental results and phase shift analysis

The inelastic scattering of protons on ^{24}Mg has been measured at the isochronous cyclotron JÜLICH at proton energies:

$$E = 22.50,\ 23.25,\ 24.00,\ 24.75,\ 25.00,\ 25.25,\ 25.50,\ 25.75,\ 27.00 \text{ and } 28.50 \text{ MeV}.$$

The angular dependence of the scattering cross sections has been observed between 20° and 160° in 5° intervals for the 7 lowest lying states. The description of the experimental details and the tabulated cross section values can be found in [7] and in [8], respectively.

Searching for giant resonances of the compound system the excitation function of the weak inelastic transition to the 0^+ ($E_{ex} = 6.43$ MeV) excited state has been studied by means of phase shift analysis, using the standard formula:

$$\frac{d\sigma'_{pp}}{d\Omega} = \frac{1}{8k^2} \sum_{lJJ'LL'} Z^2(L\,JL'\,J', 1/2l)\,.$$

$$\mathrm{Re}\,(S_L^{J*}\,S_{L'}^{J'})\,P_l(\cos\Theta)\,,$$

where the wave number of the relative motion is denoted by k, the argument of the Legendre polynomial P_l contains the scattering angle Θ, measured in the centre of mass system of reference. The coefficient Z can be expressed in terms of Racah and Clebsch-Gordan coefficients. The S-matrix elements S_L^J are parametrized in the usual way:

$$S_L^J = \eta_L e^{2i\delta_L}.$$

It was checked that the partial waves higher than $L_{max}=6$ give negligible contribution in the energy interval studied here. The spin splitting of the partial waves is neglected. The measured values of the inelastic cross section together with the results obtained from the phase shift analysis are shown in Fig. 1.

The absolute squares of the S-matrix elements η_L^2 are exhibited in Fig. 2 as functions of the bombarding proton energy E.

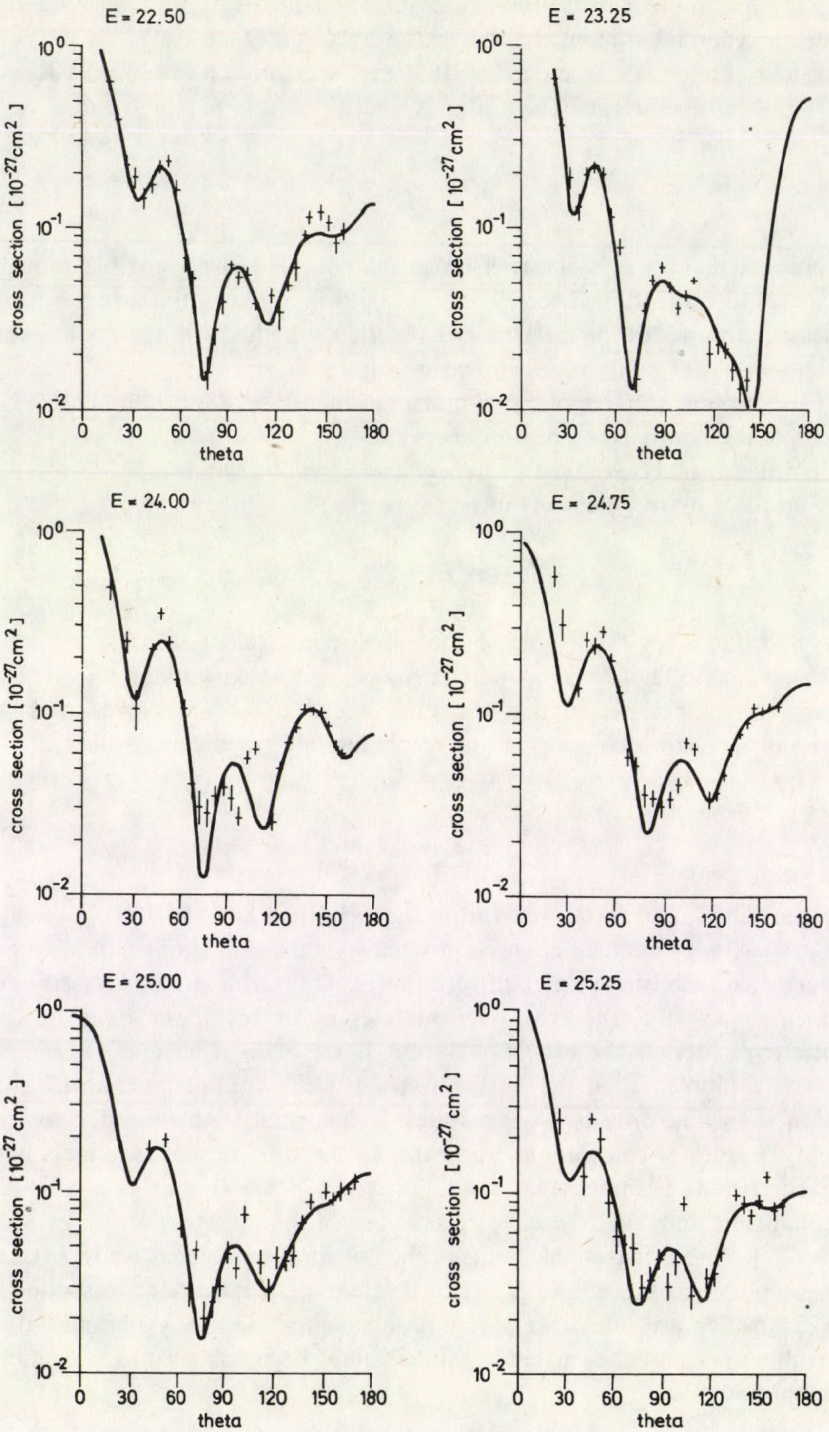
In order to analyse the energy dependence of the transition amplitude it is useful to split up the S-matrix elements into two parts:

$$S_L^J(E)=(S_L^J(E)-\langle S_L^J\rangle)+\langle S_L^J\rangle,$$

where $\langle S_L^J\rangle$ stands for the average S-matrix elements taken over the whole energy range studied here. The average S-matrix can be associated with the one-step (direct) processes and it is expected that the difference $S_L^J(E)-\langle S_L^J\rangle$ corresponds to the multistep processes which may lead to resonating intermediate states.

3. Discussion

According to our basic assumption the inelastic transition 0^+ ($E_{ex}=0.00$ MeV) $\rightarrow 0^+$($E_{ex}=6.43$ MeV) takes place predominantly via the excitation of high lying states, consequently it is sensitive for giant resonances. This assumption is based upon the experimental fact that the cross section is considerably lower than that of the neighbouring states. If the excited 0^+ state, however, is a member of a collective vibrational multiplet, then the possibility of a weak multi-step channel coupling excitation involving only low lying states, is not ruled out. In order to test this possibility a series of calculations using the method of Coupled Channels has been performed. It was assumed that the 2^+ ($E_{ex}=4.24$ MeV) state is a one-phonon vibrational state and the 0^+ excited state is one of the members of the two-phonon multiplet [9]. Within reasonable limits of the potential parameters we were unable to obtain acceptable cross sections. The typical values of the inelastic cross section for the $0^+\rightarrow 0^+$ transition are one order of magnitude higher than the experimental values. This result shows that the excited 0^+ state cannot be interpreted as a member of a vibrational multiplet.


E = 22.50
E = 23.25
E = 24.00
E = 24.75
E = 25.00
E = 25.25
cross section [10^-27 cm^2]
theta
10^0
10^-1
10^-2
0 30 60 90 120 150 180

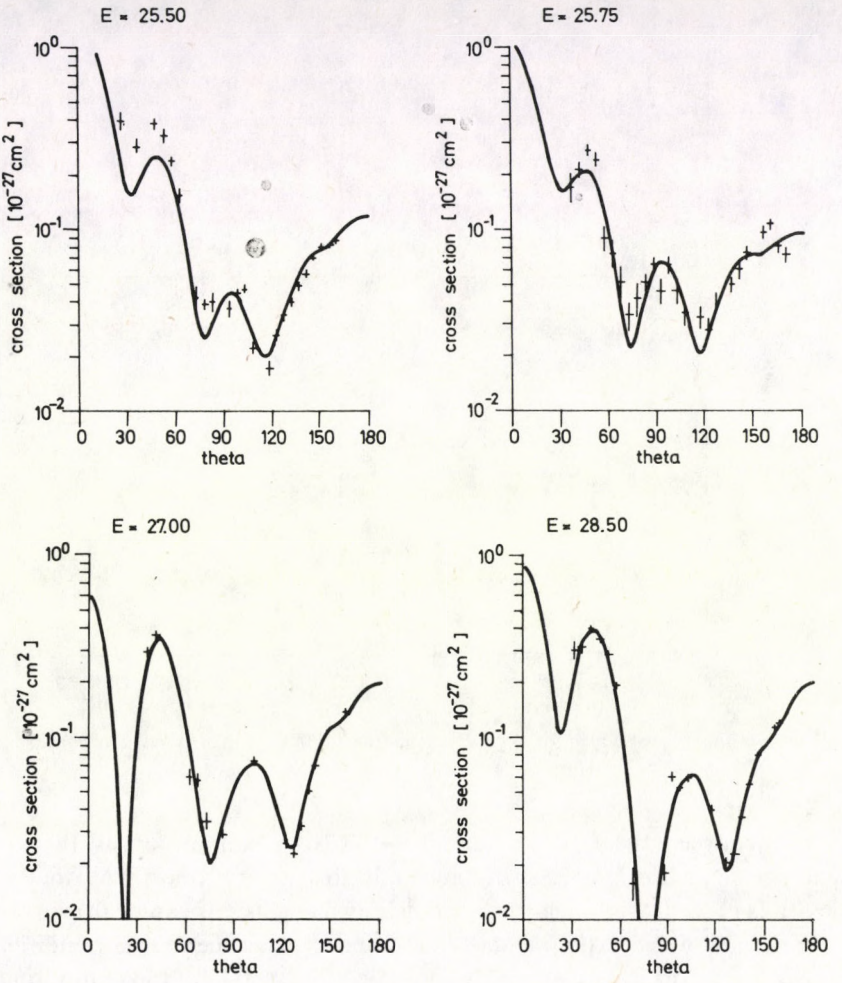


Fig. 1. Differential cross section of inelastic proton scattering on ^{24}Mg, leading to the excited state $0^+(E_{ex}=6.43$ MeV) measured at proton energies between 22.50 and 28.50 MeV. The error bars correspond to the statistical error of the counting rate. The solid lines are obtained by phase shift analysis

The splitting of the S-matrix elements into an average and a fluctuating part has been carried out, as it was indicated in the previous Section, and the fluctuating part was parametrized in the following way:

$$S_L^J(E) - \langle S_L^J \rangle = G_L(E) e^{i\varphi_L(E)} .$$

The quantities $G_L(E)$ and $\varphi_L(E)$ can be compared with the results of the analysis carried out for the $0^+ \to 3^+$ transition in [7]. From this comparison one can see a qualitative

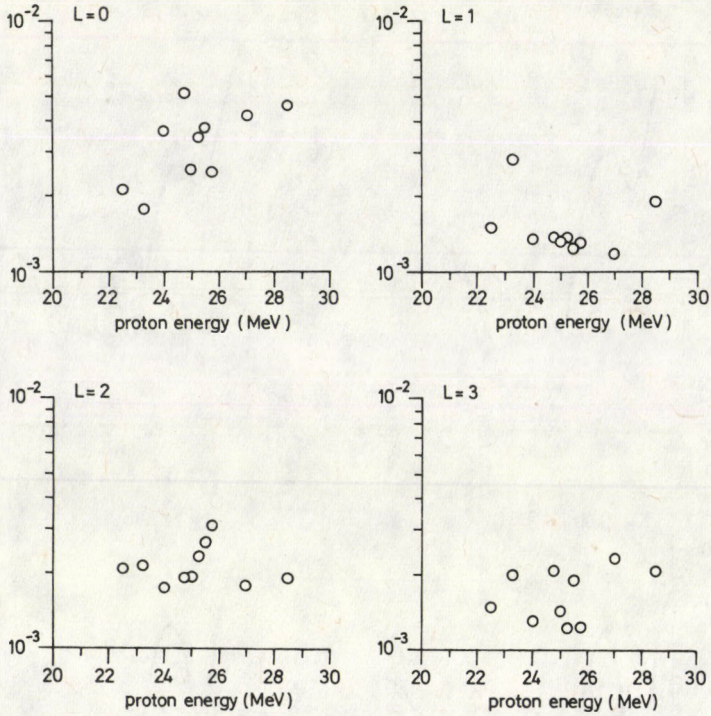


Fig. 2. The transition probability η_L^2 for angular momentum $L=0$, 1, 2 and 3 obtained by phase shift analysis

similarity between the $0^+ \rightarrow 3^+$ and $0^+ \rightarrow 0^+$ transition as far as the excitation probability of giant resonances with low multipolarity are concerned. However, from the analysis of the $0^+ \rightarrow 0^+$ transition no definite conclusions can be drawn due to the poor statistics of the experimental data. In our experiments the conditions were optimised for the measurement of the inealstic cross section of the non-normal parity state 3^+, consequently the counting rates for the $0^+ \rightarrow 0^+$ transition were too low.

In order to be able to draw quantitative conclusions on the strength distributions of the giant resonances experimental data of improved quality are needed. More specifically:

1) longer measurements are desirable because of the low cross section values;

2) in order to facilitate the phase shift analysis, measurements at angles sensitive to this particular angular distribution are required;

3) to check the maxima of the strength distributions, measurements at some additional proton energies are necessary.

The authors are indebted to Prof. C. Mayer-Böricke for valuable discussions.

References

1. M. B. Lewis and F. E. Bertrand, Nucl. Phys., **A196,** 337, 1972.
2. A. Kiss, C. Mayer-Böricke, M. Rogge, P. Turek and S. Wiktor, Phys. Rev. Lett., **37,** 1188, 1976.
3. C. Mayer-Böricke, W. Oelert, A. Kiss, M. Rogge, P. Turek and S. Wiktor, Nucl. Phys., **A293,** 189, 1977.
4. H. P. Morsch, C. Sükösd, M. Rogge, P. Turek and C. Mayer-Böricke, Phys. Rev., **C22,** 489, 1980.
5. H. V. Geramb, B. Sprickmann and G. L. Strobel, Nucl. Phys., **A199,** 545, 1973.
6. H. V. Geramb, K. Amos, R. Sprickmann, K. T. Knöpfle, M. Rogge, D. Ingham and C. Mayer-Böricke, Phys. Rev., **C12,** 1697, 1975.
7. I. Lovas, M. Rogge, U. Schwinn, P. Turek, D. Ingham and C. Mayer-Böricke, Nucl. Phys., **A286,** 12, 1977.
8. I. Lovas, M. Rogge, U. Schwinn, P. Turek, D. Ingham and C. Mayer-Böricke, Jül-1520, 1978.
9. H. P. Morsch and P. J. Ellis, Phys. Lett., **64 B,** 386, 1976.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 225—235 (1982)

DETERGENT EFFECTS ON AN ALBUMIN-CHLOROPHYLL COMPLEX MODEL OF PHOTOSYNTHETIC PROTEIN–PIGMENT COMPLEXES*

L. SZALAY, E. TOMBÁCZ, Z. VÁRKONYI

INSTITUTE OF BIOPHYSICS, JÓZSEF ATTILA UNIVERSITY, SZEGED, HUNGARY

and

Á. FALUDI-DÁNIEL*

INSTITUTE OF PLANT PHYSIOLOGY, BIOLOGICAL RESEARCH CENTER
HUNGARIAN ACADEMY OF SCIENCES, SZEGED, HUNGARY

The absorption, fluorescence, fluorescence polarization and circular dichroism spectra of an artificial complex human serum albumin–chlorophyll-a were measured at 20 °C in buffer solution (pH 7.2) and the effect of detergents (digitonin, sodium lauryl sulfate, above critical micelle concentration) was studied. A strong chlorophyll–chlorophyll interaction found in the complex is removed by detergent treatment. The ionic detergent sodium lauryl sulphate induces pheophytinization and conformational changes in the protein part of the complex leading to changes in the absolute configuration of the transition moments of the pigment. Profound changes in the fluorescence polarization spectra are caused by both detergents. The effects of detergents suggest that the possibility of similar phenomena in treating plant material should be rigorously considered.

Introduction

Molecules involved in the primary processes of photosynthesis, absorption of light, migration of excitation energy to the reaction centre, and charge separation at the centre can be grouped as follows: *lipids* (forming the membranes), *pigments* (chlorophyll-a and accessory pigments, constituents of the light harvesting system, reaction centre and the chain of the electron excitation energy) and *proteins* (metalloproteins and complexes with chlorophylls, constituents of the oxidation-reduction electron transfer chain).

According to recent knowledge pigment–pigment and pigment–protein interactions play the major roles in photosynthetic processes. The growing interest in pigment–protein complexes is well documented in two recent reviews [1] and in the Proceedings of the recent International Photosynthesis Congress with over one hundred relevant papers [2].

* Dedicated to Prof. I. Tarján on his 70th birthday.

An interesting point in all studies of these *complexes isolated from different algae and higher plants* is the universality in apparent molar weights and spectroscopical properties [3]. These complexes have different molar weights above 20 kD, absorption maxima around 436 and 673 nm (chlorophyll-a bands), and at about 280 nm (protein band). *Different complexes isolated from the same plant* (maize) *have very similar absorption spectra* though the apparent molar weights differ widely (from 20 to 250 kD). It is worth mentioning that no significant amounts of free chlorophylls were found in this plant [4], which shows that practically all the chlorophyll exists as chlorophyll–protein complexes demonstrating the importance of these complexes in photosynthetizing systems.

Owing to the universality of chlorophyll–protein complexes artificial model complexes with well-defined composition can be prepared. These complexes are very similar to those existing in vivo, and can therefore be used for studying environmental effects and for determining roles existing under in vivo conditions. The modeling of pigment–protein complexes is not new; milk proteins [5], casein [6], bovine [7] and human serum albumin have been used to study the interactions between pigment and protein molecules, bonding states, photochemical activities and the aggregation of pigments in the complexes. A recently prepared chlorophyll–bovine serum albumin complex has been investigated spectroscopically [8] and its luminescence has been studied [9]. An improved method [10] resulted in a human serum albumin complex with an apparent molar weight of about 230 kD, containing 3 albumin molecules each carrying about 3 chlorophyll-a molecules. This stoichiometry is remarkably similar to that of a recently isolated complex [1, 4] with a molecular weight of 250 kD and composed of two or more subunits, each probably containing three chlorophyll-a molecules. It is remarkable that all of these artificial complexes show spectroscopic properties very similar to one another as well as to those isolated from plant material, i.e. to the in vivo complexes. There are models in which the chlorophyll–chlorophyll interactions have been eliminated by providing complexes of one molecule protein (apomyoglobin) with one molecule chlorophyll [11]. In another model these interactions do exist, but can be regulated [12]. These models are very useful to find the consequences of the association of protein and chlorophyll molecules. If, however, we are interested in the interaction of the complex molecule as a whole with its environment, then models containing more subunits and several chlorophyll molecules which are reminiscent of in vivo complexes seem more reliable.

In order to study the protein–pigment complexes present in plant material, the complexes first need to be isolated. An important step in the isolation process is the solubilization of the complex using detergent (digitonin, triton, sodium dodecyl sulphate, etc.). This is followed by fractionation (gel filtration, electrophoresis). With this procedure complexes of different apparent molar weights may be separated. However, the presence of detergent molecules may bring about new interactions and in addition to the above-mentioned molecular interactions detergent–pigment, detergent–protein and detergent–lipid interactions may appear and produce artifacts

in studies performed on the isolated medium. The importance of this possibility has already been recognized by several authors and very recently Argyroudi-Akoyunoglou [13] found that the binding of chlorophyll to protein in the complexes is specific and not an artifact due to sodium lauryl sulphate treatment. However, depending on the detergent concentration, certain changes occur in the absorption and fluorescence spectra of the light–harvesting pigment–protein complex isolated from pea chloroplasts [14]. Among the complexes isolated in some laboratories, there is an unstable one, named CPa (a complex of chlorophyll-a), which is considered to be the reaction centre of the second photochemical system (PS-2) of photosynthesis. This statement has been confirmed recently by several authors [13]. However, this complex is labile to detergent treatment therefore it cannot be safely isolated [15]. All these reports on chlorophyll-a–protein complexes, and on the effect of detergent treatments point out the importance of obtaining more information about chlorophyll-a–protein complexes and about the effects of different environments, particularly those created by detergent molecules. We have already experimented on the effects of detergents on chlorophylls and on other dyestuffs [16] and on proteins [17]. We decided to investigate an artificial human serum albumin–chlorophyll-a complex recently prepared in our laboratory and already used successfully in spectroscopic studies [10]. The following main problems were studied: a) the condition of the pigment in the complex; b) the effect of detergent treatment on a pigment–protein complex; c) the relevance of these effects on in vivo complexes extracted by a similar treatment.

Materials and methods

A chlorophyll-a–human serum albumin (Chl-HSA) complex prepared in our laboratory was used. The details of the preparation are published elsewhere [10]. The main steps are the following. Lecithin dissolved in chloroform and Chl dissolved in diethyl ether are mixed and evaporated to dryness. The dry material is dissolved in Britton-Robinson buffer and is sonicated to produce liposomes containing Chl. HSA is added to the suspension and after shaking and waiting for the complexation of Chl and HSA a Sephadex column is used for separating the complex. The fractions may have different absorption coefficient ratios $k(280\,\text{nm})/k(680\,\text{nm})$ depending on the number of Chl molecules per molecule of HSA. If fractions with ratios from 0.95 to 1.10 are collected, the apparent molar weight of the complex is about 230 kD and the number of Chl molecules per molecule HSA is 3 to 1 depending on the concentration of the solutions. (If the ratio is lower or higher than 1:1, more or less Chl-s, respectively, are attached to one HSA molecule.)

For detergent treatment sodium lauryl sulphate (SLS, an ionic detergent, product of Fluka AG, Buchs SG, Switzerland) and digitonin (D, a non-ionic detergent, product of Reanal, Hungary) were used. These detergents are both very often applied in the separation of pigment–protein complexes. Above a critical detergent concentration

micelles are formed (in the case of SLS this concentration is about $2.5 \cdot 10^{-3}$ M, in non-ionic detergents it is lower). In order to follow the procedures used in solubilizing in vivo complexes their concentrations were well above the critical concentration; in all cases $6.6 \cdot 10^{-3}$ M detergent solutions were used.

The absorption spectra were measured with a Specord UV-VIS (Zeiss) or (especially in the Chl absorption region) with an SF-18 recording spectrophotometer (USSR). The fluorescence spectra and fluorescence polarization spectra were recorded with a Perkin-Elmer (Type MPF-44A) spectrofluorimeter. Correcting for instrument error the degree of polarization was calculated from the formula

$$p = \frac{I_{||} - \rho I_{\perp}}{I_{||} + \rho I_{\perp}},$$

where $I_{||}$ is the intensity of fluorescence measured with both the observing and the exciting polarizers vertically oriented, and $I_{\perp}$ is the intensity measured with the exciting polarizer vertically oriented and the observing polarizer horizontally oriented. The correction factor is the ratio of fluorescence intensities obtained with the exciting polarizer horizontally oriented and with the observing polarizer both vertically and horizontally oriented. The ratio $\rho = I_{\text{vertical}}/I_{\text{horizontal}}$ takes the systematic error of the instrument into account.

The circular dichroism (CD) spectra were recorded with a JASCO 40 C spectropolarimeter.

All measurements were carried out at room temperature. All manipulations with chlorophyll-containing materials were carried out in dim light; the preparations were kept under air in the dark and in a refrigerator at 4 °C. In these circumstances the optical properties remained reasonably constant for 3—4 days.

Results

Absorption spectra

The absorption spectra were recorded from 220 nm to the red end at about 700 nm. The spectra are shown in Fig. 1 from 350 nm only, because the absorption peak at 274 nm (belonging to HSA absorption) published in [10] did not change except for a slight shift towards longer waves and the change in the absorption intensity was less than 5 per cent. The detergent treatment in the absorption region of Chl (Fig. 1) causes considerable change when the ionic detergent SLS is used. The peak of the absorption of the complex at 440 nm disappears and a new peak at 418 nm appears together with the maxima at 512 and 550 nm, which should be attributed to an extensive pheophytinization of chlorophylls. In the red band a shift of the maximum from 674 to 677 nm is observed. The non-ionic detergent D does not induce such spectral changes, but causes a decrease of the extinction, $E(\lambda)$. (By using proper sample thicknesses the

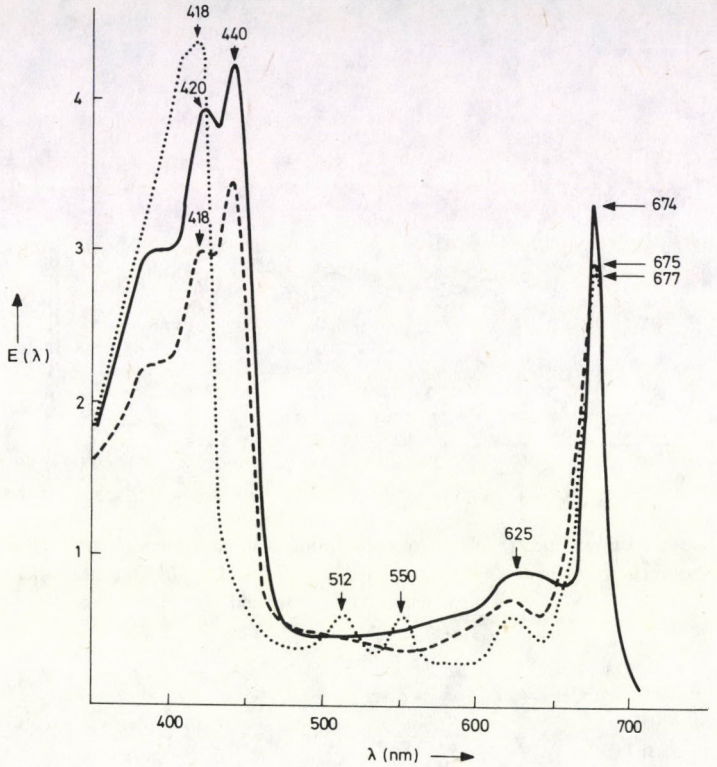


Fig. 1. Absorption spectra of the complex human serum albumin–chlorophyll-a in Britton–Robinson buffer (pH 7.2; solid line), complex $+6.6 \cdot 10^{-3}$ M sodium sulphate (dotted line), complex $+6.6 \cdot 10^{-3}$ M digitonin (broken line)

actual measured values of extinction never exceeded 0.8. In Fig. 1 all $E(\lambda)$-values refer to 1 cm layer-thickness.)

Since the incubation time with detergent has been found to be an important factor of the detergent effect, the absorption spectra after 1 and 24-hour incubation were recorded (Fig. 2). (Note the different scales for SLS- and D-treated solutions.) The effect of detergents on the absorption spectrum of the complex is apparent after 1-hour incubation.

Fluorescence spectra

The spectral distribution of the intensity of protein fluorescence excited at 280 nm and that of the chlorophyll-a fluorescence excited at 435 nm are shown in Fig. 3. Relative intensities $F_{rel} = F/F_{max}$ are plotted; $F_{max} = 1$ is taken arbitrarily for both the HSA and the Chl part of the spectrum. The intensity scale, F_{rel}, is different for the fluorescence of HSA-Chl complex solutions and for the solutions treated with D and

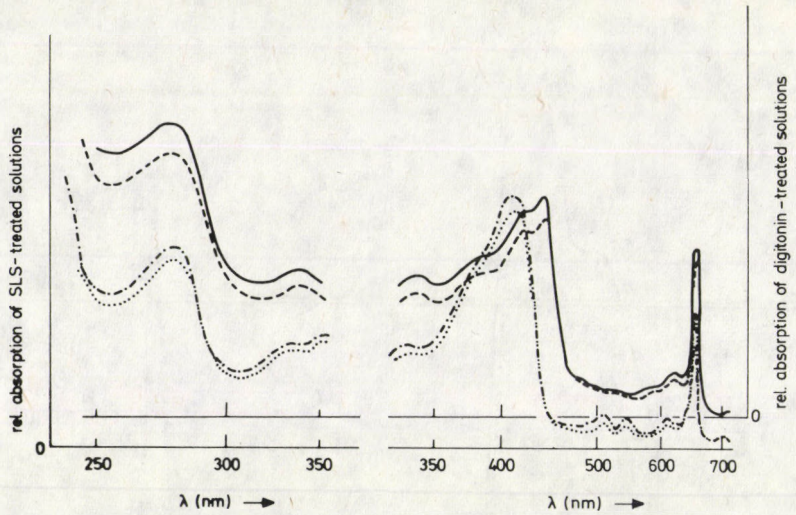


Fig. 2. Relative absorption spectra of the complex human serum albumin–chlorophyll-a treated with $6.6 \cdot 10^{-3}$ M sodium lauryl sulphate for 1 hour (dotted line) and for 1 day (dash-dotted line), and treated with $6.6 \cdot 10^{-3}$ M digitonin for 1 hour (solid line) and for 1 day (broken line)

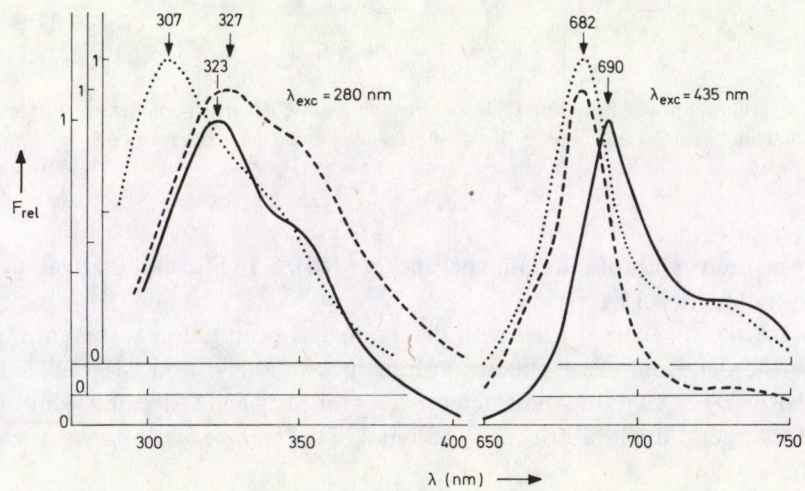


Fig. 3. Relative fluorescence spectra of the complex human serum albumin–chlorophyll-a and the spectra of detergent-treated samples (notation as in Fig. 1)

SLS detergents. The protein fluorescence peak of SLS-treated solution is shifted far to the ultraviolet part (to 307 nm). The peak of chlorophyll fluorescence is shifted from 690 to 682 nm after detergent treatment. The absolute intensities of the fluorescence are greatly increased in the Chl band after detergent treatment: relative to the untreated sample the maximum height of the peaks at 682 nm is about 50 times higher. With excitation at 280 nm (HSA absorption peak) the fluorescence of chlorophyll-a appears and the spectral distribution is similar to that shown in Fig. 3.

Circular dichroism spectra

These spectra exhibit very pronounced changes after detergent treatment (Fig. 4). The CD-spectrum of the HSA-Chl complex in the red band of absorption $E(\lambda)$ (shown in Fig. 1) has a typical shape attributed to the presence of Chl dimers. The positive and negative peaks of almost equal heights and the intersection at 694 nm on the wavelength axis are characteristic of the CD spectrum of the complex.

After D treatment a single negative band appears with maximum at 677 nm. After SLS treatment a single positive band is obtained with peak at 683 nm.

Fluorescence polarization spectra

Relative fluorescence polarization spectra are shown in Fig. 5, where p_{max} denotes the maximum value of p in the appropriate spectral range. For comparison the absorption spectrum $E(\lambda)$ is shown (right scale). This presentation gives information on

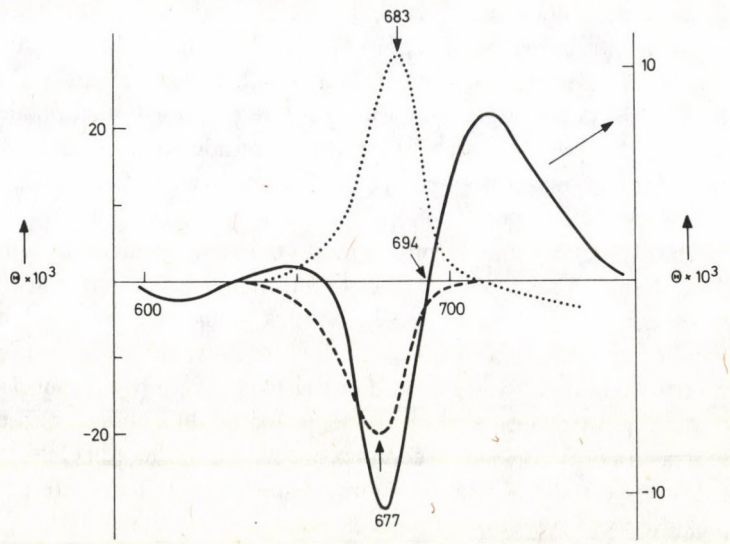


Fig. 4. Circular dichroism spectra of the complex human serum albumin–chlorophyll-a and the spectra of detergent-treated samples (notation as in Fig. 1)

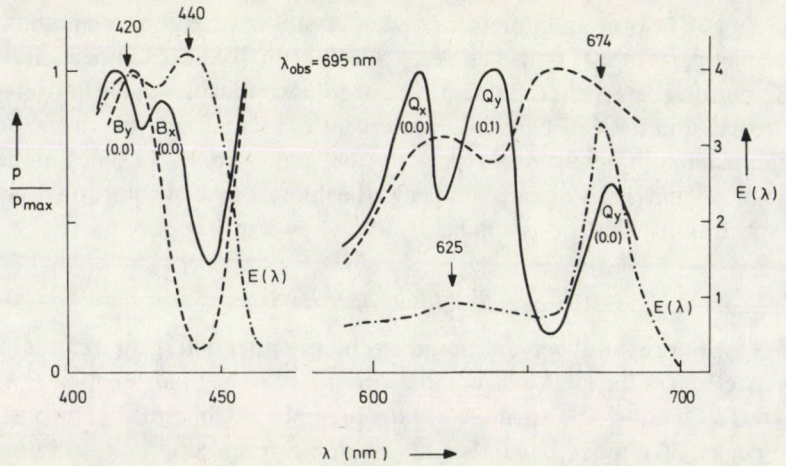


Fig.5. Relative fluorescence polarization spectra of the complex human serum albumin–chlorophyll-a and the spectra of detergent-treated samples (notation as in Fig. 1)

the spectral distribution of p. The distribution was reproducible: from 3 samples prepared independently the peaks were at 415, 436, 615, 640 and 677 nm within ± 3 nm accuracy. However, the absolute values of p varied for some peaks by over 100 per cent.

Discussion

From the experimental results conclusions can be drawn: a) on the composition (and partly on the structure) of the complex prepared for modeling plant pigment–protein complexes, especially for that of CPa, which cannot be easily isolated from plants; b) on the effect of detergents on the structure of our artificial pigment–protein complex; c) on the points which should be considered in the evaluations of spectroscopic data of plant pigment–protein complexes obtained with detergent treatment.

In [9] the evaluation of absorption and fluorescence spectroscopic data leads to the conclusion that HSA does not induce essential changes in the spectral and energetic parameters of the pigment part of the complex. This statement is valid only if spectroscopy means strictly fluorescence spectroscopy, and the red shift of the absorption (10—15 nm) and the increase of the band width (by about 5 nm and the appearance of long-wave fluorescence) are neglected. In this sense our absorption and fluorescence spectra of the HSA–Chl complex are similar to those of Chl-solutions, too. Low-temperature spectra, however, are more sensitive to the changes of pigment entity after complexation and they suggest the presence of Chl associates in the complexes. In [12] casein–Chl complexes with a peak of associated Chl molecules appeared at 672 nm if the number of Chl molecules per complex molecule ≈ 5—6. Though the

absorption and fluorescence spectra, especially at ambient temperatures, do not show definite pigment–pigment interactions in all cases, the CD spectra reveal the presence of associates without any doubt. As Fig. 4 indicates, the CD spectrum of HSA–Chl has an exciton signal with extrema at 677 (−) and 715 nm (+) in the main red absorption band of Chl, which means that dimeric Chl molecules are present in our HSA–Chl complex. The prerequisite of pigment association is given, since from earlier evidence [10] we know that about 9 Chl molecules are bound to each HSA unit (of 3 HSA molecules). Since, however, the pigment is non-covalently linked to protein, only an average number of chlorophyll molecules per protein molecule can be given [21].

In D-treated complex solutions the absorption spectra do not change much in accordance with the above statement about the low sensitivity of these spectra (Fig. 1). In the CD spectrum a negative band appears with a maximum at 658 nm indicating the presence of monomeric Chl. (A similar negative band is seen in diethyl ether solution, where only monomeric forms can be present.) D is a non-ionic detergent having a slight effect, but still strong enough to break the Chl associates, most probably by incorporating Chl molecules into D micelles.

In SLS-treated complex solutions the absorption spectrum (Fig. 1) exhibits very pronounced changes: the shift of the blue band (from 440 to 418 nm) and the appearance of new peaks between 500 and 600 nm indicate a pheophytinization of Chl. A profound change in the CD spectrum (Fig. 4) is the reversal of the single band with maximum at 662 nm from a negative (obtained with D-treatment) to a positive band. One should not forget, however, that this CD-band belongs to pheophytinized Chl.

This behaviour has been found very recently in bacteriochlorophyllides [11]. The association of protein and chlorophyll in defined (1 : 1) structures was studied using apomyoglobin (as protein) and different chlorophyllides (as chlorophyll analogs). In this system the heme group of the protein is replaced by a single pigment molecule. If bacteriochlorophyllide was used for replacement, several transitions in the CD spectrum became entirely reversed. The reversal can be explained by the change of the absolute configuration of the transition moments. In our case, the SLS treatment should induce a conformational change of the HSA matrix. This change transfers a monomer linked to a given site into another orientation. In micellar solutions the concentration of monomeric detergent molecules is a saturation concentration therefore the number of monomeric detergent molecules is always sufficiently high to interact with proteins.

According to [18] monomeric forms can only interact with proteins and they often induce conformational changes or even disruption of the protein macrostructure. The conformational change of HSA is corroborated by the great shift of the fluorescence maximum from 327 to 307 nm, which can be explained by conformational changes leading to tyrosine fluorescence (hindering the excitation transfer from tyrosines to tryptophan) or to changes of the environment of tryptophan.

Detergents are able to solubilize Chl and can easily form micellar Chl solutions. The effect seems to take place comparatively quickly. After one hour of incubation the

effect on the absorption spectrum appears (Fig. 2); not much further change can be seen after one day of incubation. This might be different when chloroplast fragments of larger size are treated [14].

The general features of the relative fluorescence polarization spectrum of the complex (Fig. 5) are similar to those obtained earlier for Chl-s or Chl analogues. The polarization spectrum of Chl-a in castor oil showed a strong overlap of two bands around 400 nm and a further overlap around 650 nm [23]. The potassium chlorophyllide fluorescence polarization spectrum in glycol had maxima at 408 and 415 nm and an increase beyond 430 nm [24]. The Chl-a fluorescence polarization spectrum in ethanol at 77 K shows three smaller peaks from about 400 to 480 nm, a sharp peak at about 630 nm and an increase toward longer waves [20].

In connection with our spectra two statements can be made: a) the reproducibility of the absolute values of the degree of fluorescence polarization was very bad; b) the D treatment induced profound changes in the spectral distribution of the degree of polarization. The first statement can be easily understood. Since Chl molecules are non-covalently linked to different amino acid residues, in spite of their similar spectroscopic properties and energetically similar build-up, the fluorescence polarization may be different in different samples. The degree of polarization reflects the mobility of the pigment molecule in its microenvironment, which is highly sensitive to minor changes.

The second statement cannot be easily explained. The peaks at around 430 and 640 nm disappear in D-treated samples. Two possible explanations can be suggested: a) The electronic transitions with perpendicular transition moments Q_x and Q_y [24] interact differently with detergent molecules. The transitions of the moments Q_y and Q_x lie around 674 ($Q_y(0, 0)$), 625 nm ($Q_y(0, 1)$) and around 615 nm ($Q_x(0, 0)$); the transitions of the moments B_x and B_y are at about 440 ($B_x(0, 0)$) and 240 nm ($B_y(0, 0)$). The detergent molecules affect the transitions $B_x(0, 0)$ and $Q_y(0, 1)$ only. This picture is not clear however. It is generally known that Q_y is perturbed more strongly by environmental effects [24], but one cannot accept that $Q_y(0, 0)$ and $B_y(0, 0)$ are not concerned. b) Two equilibrating ionic forms of the complex [19] interact with the HSA–Chl complex differently. At pH 7.2 there exist two forms, a cationic form with a more pronounced absorption peak at about 420 nm and a neutral form with higher maximum at around 440 nm. Since the peak of the fluorescence polarization spectrum at longer wavelength in this region (at 432 nm) disappears after D treatment, the neutral form of the complex seems to be influenced by D molecules. In the red band of Chl absorption, however, this explanation cannot be satisfactorily applied either, because in this region the bands of the neutral and cationic forms strongly overlap, just as the fluorescence originating from the excitation of these forms [19].

The conclusions can be briefly summarized as follows. a) The artificial HSA–Chl complex incorporates Chl dimers and can therefore be a candidate for modeling CPa. b) The detergent treatment of this complex removes the dimers. The ionic detergent SLS induces pheophytinization and conformational changes of the complex; further, it

causes changes in the absolute configuration of the transition moments. Non-ionic detergents, such as D, may change the spectral distribution of fluorescence polarization both by removing some characteristic bands and by shifting the maxima. c) Under in vivo conditions, using the same types of detergents in similar concentration, one should consider the possibility of detergent–complex interactions which were found with this artificial complex. SLS treatment needs careful consideration, and the data of polarization should be very carefully analysed. The effect of detergent on the protein part of the complex deserves special attention, because it is believed that detergent micelle formation diminishes the concentration of monomers and prevents protein denaturation.

References

1. J. P. Thornber, T. P. Markwell and S. Reinman, Photochem. Photobiol., **29,** 1205, 1979.; J. Barber, ibid., **30,** 203, 1980.
2. Proceedings of the Vth International Congress on Photosynthesis, Kallithea, Greece, Sept. 1980. Ed. G. Akoyunoglou, Balaban International Science Service, Philadelphia, 1981.
3. R. G. Hiller, in Proc. Vth International Congress on Photosynthesis, see Ref. [2].
4. J. P. Markwell, J. P. Thornber and R. T. Boggs, Proc. Natl. Acad. Sci. USA, **76,** 1233, 1979.
5. Yu. E. Giller, G. B. Krasichoa and D. I. Saposhnikov, Biofizika (Moscow) **15,** 38, 1970; Yu. E. Giller, L. N. Yuhanova and S. K. Abdullaeva, Dokl. Akad. Nauk SSSR, **207,** 1475, 1972; M. Dinant and J. Aghion, Photochem. Photobiol., **17,** 25, 1973; M. Widart, M. Dinant, F. Téchy and J. Aghion, Photobiochem. Photobiophys., **1,** 103, 1980.
6. A. Meister and E. Brecht, Biochem. Physiol. Pflanzen, **174,** 305, 1979.
7. G. M. Sherman, Nature, **224,** 1108, 1963.
8. E. Tombácz and E. Vozáry, Acta Biol. Univ. Szeged, **26,** 33, 1980.
9. E. I. Zenkevich, G. A. Kochubeev and A. M. Kulba, Biofizika (Moscow) **26,** 389, 1981.
10. E. Tombácz, Z. Várkonyi and L. Szalay, Zh. Prikl. Spectr. (Minsk) **36,** 64, 1982.
11. S. G. Boxer, 9th Ann. Meeting. Amer. Soc. Photobiol. Program and Abstract. June 14—18, p. 61. 1981.
12. Yu. E. Giller, Mol. Biol. (Moscow), **14,** 1057, 1980.
13. A. Wild, Z. Pflanzenphysiol., **100,** 1, 1980; J. M. Anderson, FEBS Lett., **117,** 327, 1980; J. H. Argyroudi-Akoyunoglou, A. Castorinis, Arch. Biochem. Biophys., **200,** 326, 1980.
14. M. D. Illina, E. A. Kotova and A. Yu. Borisov, Biochim. Biophys. Acta, **636,** 193, 1981.
15. A. C. Stewart, FEBS Lett., **114,** 67, 1980.
16. D. B. Knaff, R. Malkin, J. C. Myron and M. Stoller, Biochim. Biophys. Acta, **459,** 402, 1977; T. Szitó, J. Hevesi and E. Bálint, Zh. Prikl. Spectr. (Minsk) **33,** 751, 1980; E. Lehoczki and K. Csatorday, Biochim. Biophys. Acta, **396,** 86, 1975; E. Lehoczki, ibid. **408,** 223, 1975.
17. E. Vozáry and L. Szalay, Acta Phys. Chem. Szeged, **23,** 17, 1976.
18. A. Helenius and K. Simons, Biochim, Biophys. Acta, **415,** 29, 1975.
19. L. Szalay, Z. Várkonyi and E. Tombácz, Biofizika (Moscow), **26,** 628, 1981.
20. P. Koka and S. Pill-Soon, Biochim. Biophys. Acta, **495,** 220, 1977.
21. I. N. Stadnichuk and M. V. Guseva, Photosynthetica, **15,** 221, 1981.
22. G. P. Gurinovich, A. N. Sevchenko and K. N. Solovjev, Spectroscopy of Chlorophylls and Related Compounds. Translated from Russian AEC-tr-7199, Chemistry (TID-4500) US Atomic Energy Commission, Div. Techn. Inform. p. 241, 1971.
23. D. Frackowiak and Z. Kojro, Proc. Int. Conf. Luminescence, Budapest, Akadémiai Kiadó, 301, 1966.
24. R. K. Clayton, Photosynthesis: Physical Mechanisms and Chemical Patterns. IUPAB Biophysics Series; 4. Cambridge University Press, p. 147, 1980.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 237—246 (1982)

MOLECULAR SPECTROSCOPY AS A TOOL IN STUDYING BIOPHYSICAL PROBLEMS*

S. DAMJANOVICH and J. SZÖLLŐSI

DEPARTMENT OF BIOPHYSICS, UNIVERSITY MEDICAL SCHOOL OF DEBRECEN
DEBRECEN, HUNGARY

A joint application of flow cytometry, a fast optical analysis of animal cells, and the Forster-type resonance energy transfer technique provided a suitable tool for the study of proximity relationships on the cell surface.

Fluorescence cell sorting and analysis made it possible to study the distribution pattern of cell membrane components such as the $H-2K^k$ antigen. The relative distribution of the concanavalin A receptor and the $H-2K^k$ antigen has also been studied.

The steady development of modern physical theories and experimental methods for the last 30 years has influenced biology to such an extent that nowadays we may speak of a revolutionary development of this science. Molecular spectroscopy has provided a particularly powerful means to study biophysical problems.

Three of the recently developed luminescence spectroscopic methods seem to be extremely versatile. These methods are in the chronological order of their introduction to biophysics: fluorescence activated cell sorting and analysis [1, 2, 3], fluorescence redistribution after photobleaching (FRAP) [4, 5] and time-dependent delayed fluorescence and phosphorescence spectroscopy [6, 7] at room temperature.

All these experimental techiques have contributed to the rapid development of cellular and molecular biophysics more than any other experimental technique so far, except for X-ray crystallography. These luminescence techniques also have the great advantage of enabling us to investigate living material, including living cells. If compared with some of the resonance spectroscopic methods, like NMR, ESR or Mössbauer spectroscopy, the latter have much lower sensitivity and demand more expensive hardware.

After the rapid and really spectacular success of biochemistry in studying biological material, its molecular compositions and chemical reactions, more and more information has been obtained indicating the importance of molecular dynamics in the living material. The rotational and translational mobility of macromolecules reflects the combined effects of size, shape, internal flexibility, environmental fluidity and stereochemical constraints, static and dynamic interaction, proximity relationships, etc. Luminescence techniques have opened up a new world, a time window

* Dedicated to Prof. I. Tarján on his 70th birthday.

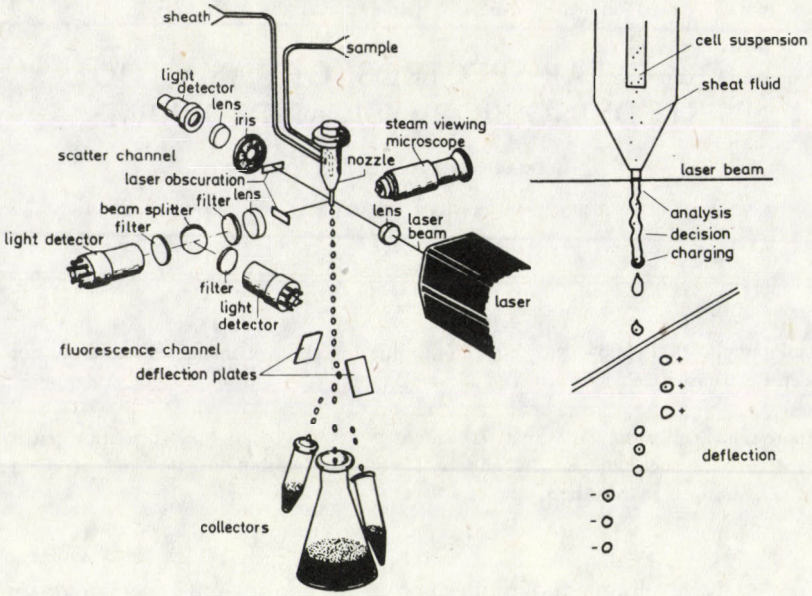


Fig. 1. Details of the optical interrogation and sorting of the flow cytometer, FACS III (Becton-Dickinson)

ranging from $10^{-9} - 10$ s, making the study of such motions and interactions possible. However, the statistical nature of the biological interaction, i.e. the slight individual differences if we study only a limited number of biological objects like cells, may misinform us regarding minor, although sometimes essential details.

In this communication we intend to deal briefly with fluorescence-activated cell sorting and analysis, with some references to the work carried out in our Laboratory. Fluorescence-activated cell sorting and analysis was the first method to give spectroscopic information about individual cells at a speed of 1—10^4 cell/s. Fig. 1 shows the basic principle of the instrument. A cell suspension is driven through a capillary system. Suitable adjustment of the air pressure in the sample and sheath containers allow a sample of fluorescently tagged cells to be injected into the centre of the flowing stream of compatible sheath liquids, establishing a laminar coaxial flow within the nozzle-transducer assembly, until the thin fluid stream has a diameter comparable to the diameter of a single cell. A laser beam orthogonal to the cell stream is centred so that the scattered light and the fluorescence signals can be collected and centred on photosensitive detector systems, like photomultipliers or photodiodes. As Fig. 2 shows, the signals can be individually collected either by a microprocessor, or in a computer. This system allows data collection and analysis at high speed with great precision. The scattered light is proportional to the particle or cell size and shape and also viability [8, 9, 10]. The fluorescence signal (or signals) can give information on any of the cell constituents labeled specifically by fluorescence dyes. In the case of multiple

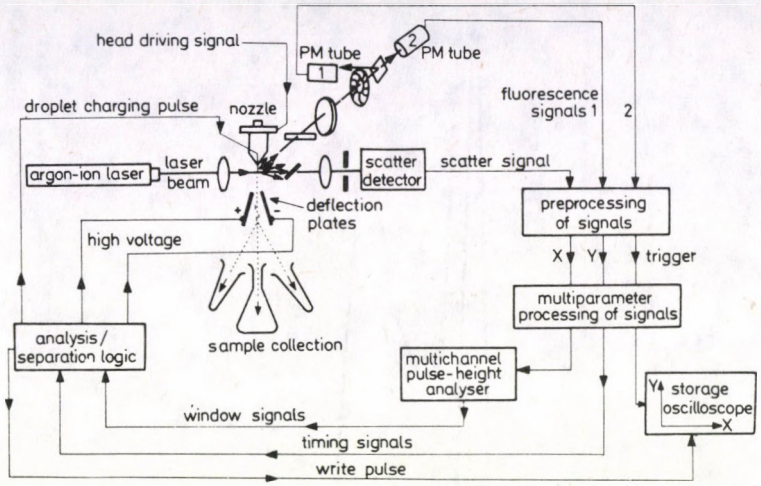


Fig. 2. Schematic block diagram of the signal processing of the flow cytometer, FACS III (Becton-Dickinson)

detection and double excitation, anisotropy data or resonance energy transfer processes can be traced equally well [11].

The immediate application of the high speed analysis of the cells, giving information on the cell size distribution, DNA and/or protein content, number and distribution of specifically labeled ligand binding sites, etc., is plausible in cell biology, hematology, immunology, gynecology, oncology, etc.

Naturally, analysis in itself, although very important, is still not the ultimate possibility of this system. A properly applied piezoelectric transducer is capable of presenting uniform droplet formation. If the analysis of the optical signals occurs at the unbroken jet of the fluid stream where the droplets are not yet formed, and the electronic analysis of the signals is fast enough compared to the flow, then applying a pulse of proper shape to the saline in the flowchamber at the time the wanted (analyzed) cell is about to be pinched off into a droplet, the cells giving certain signals can be separated. Generally three droplets are charged, making the separation of the cells much safer.

Figs 3 and 4 demonstrate some of our data on the DNA content of normal and Gross virus leukemic lymphocytes. The proportion of cells having large DNA content is higher in the case of leukemic lymphocytes than in the case of normal ones, i.e., the leukemic lymphocytes divide faster than normal cells. Other significant differences between normal and leukemic cells were found when samples stained with fluorescein diacetate (FDA) were analyzed in the flow cytometer (Figs. 5 and 6). FDA is hydrolysed inside the cells by esterase enzymes and the liberated fluorescein accumulates in the cells. The analysis of scattered light and fluorescence intensity distribution histograms of cells revealed that the fluorescein content of large leukemic cells is three times higher

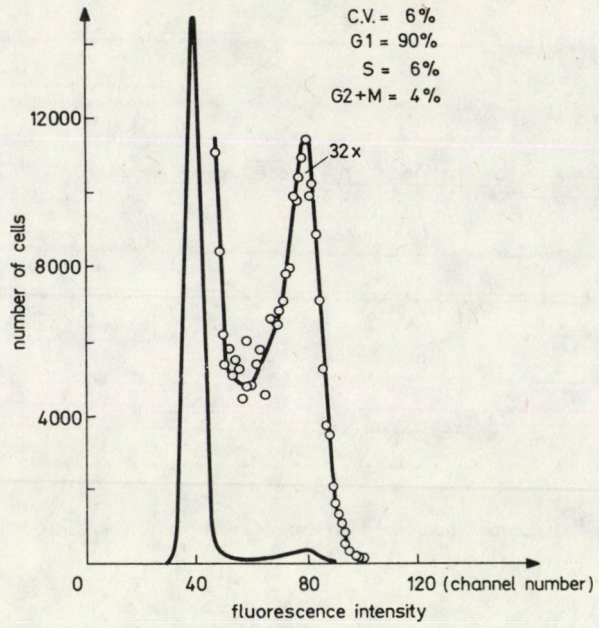


Fig. 3. Fluorescence intensity distribution histogram of normal mouse lymphocytes. The cells were fixed with alcohol and stained with Hoechst 33 342 dye which is specific for DNA. The cell cycle distribution of cells is shown in the figure

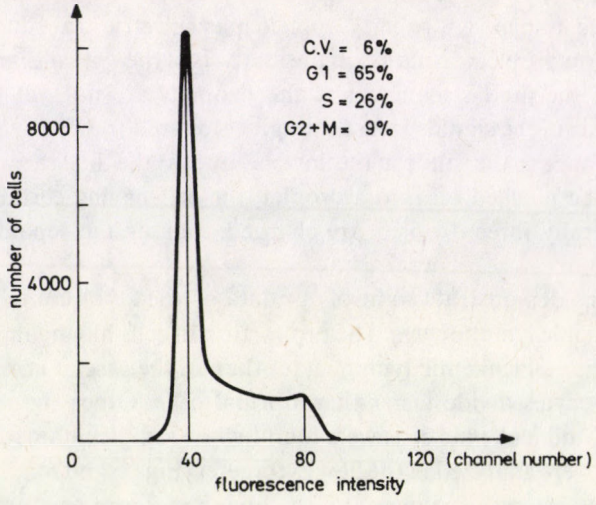


Fig. 4. Fluorescence intensity distribution histogram of leukemic mouse lymphocytes. The cells were fixed with alcohol and stained with Hoechst 33 342 dye which is specific for DNA. The cell cycle distribution of cells is shown in the figure

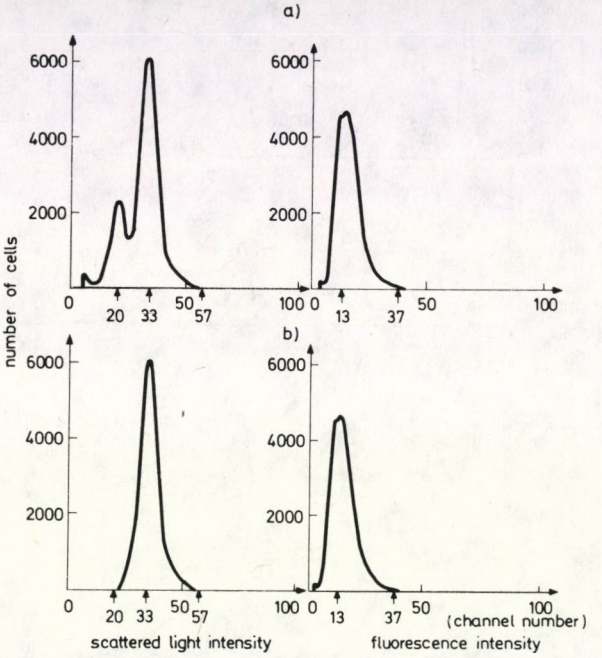


Fig. 5. Frequency distribution histograms of scattered light and fluorescence intensities of normal C3H/He-mg × AKR/F_1 hybrid mouse lymphocytes. For the analysis the lymphocytes were prepared from mesenteric lymph node. The distribution histograms were recorded in the 6th min of the FDA hydrolysis. Data collection was triggered with the scattered light (*A*) or with the fluorescence intensity (*B*). The first and second peaks of light scattering distribution histogram (*A*) belong to the red blood cells and normal lymphocytes, respectively

than that of the small, normal ones. The observed differences both in size and fluorescein content of normal and leukemic lymphocytes enable one to distinguish unequivocally between these cells [12, 13, 14].

Fig. 7 gives a more advanced version of a fluorescence flow cytometer compared to the system shown in Figs. 1 and 2. In this system the dual laser excitation provides a new possibility for the study of cell surface receptors and their proximity relationships. The binding sites for the ligands, Con *A* and anti-$H-2K^k$, on the cell surface of T-41 lymphoma cells were simultaneously labeled with fluorescein izothiocyanate (FITC) and tetramethyl rhodamine izothiocyanate (TRITC) conjugated ligands. The two dyes fulfil the Förster requirement for fluorescence energy transfer as the emission spectra of the donor dye (FITC) and the absorption spectra of the acceptor dye (TRITC) substantially overlap (Fig. 8).

The transfer efficiency of a Förster type fluorescence resonance energy transfer depends on the negative sixth power of the donor–acceptor distance (R) and parameter

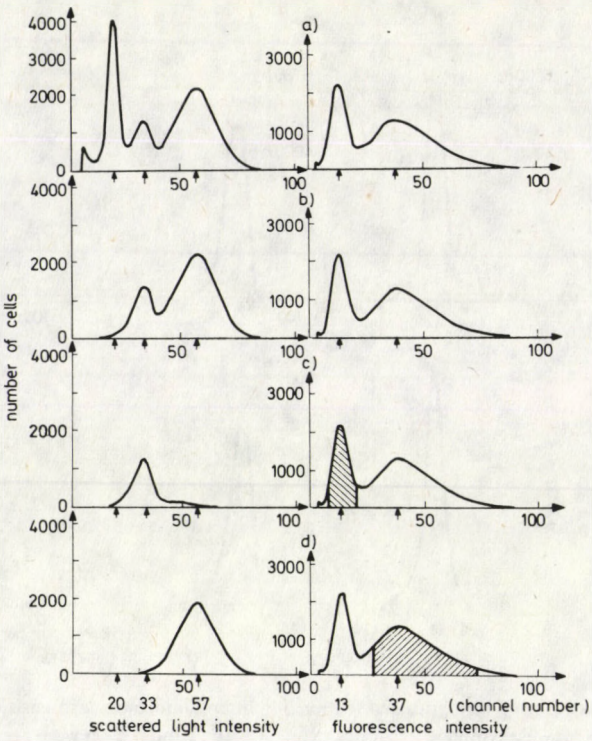


Fig. 6. Frequency distribution histograms of scattered light and fluorescence intensities of leukemic C3H/He-mg × AKR/F_1 hybrid mouse lymphocytes. For the analysis the lymphocytes were prepared from mesenteric lymph node. The distribution histograms were recorded in the 6th min of the FDA hydrolysis. Data collection was triggered with scattered light (*A*), fluorescence intensity (*B*), or the hatched region of fluorescence intensity (*C*, *D*) shown in the Figure. The first, second and third peaks of the light scattering distribution histogram (*A*) belong to the red blood cells, normal lymphocytes and the leukemic lymphocytes, respectively

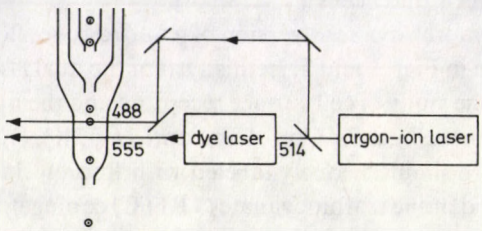


Fig. 7. Schematic diagram of a flow cytometer with dual laser excitation

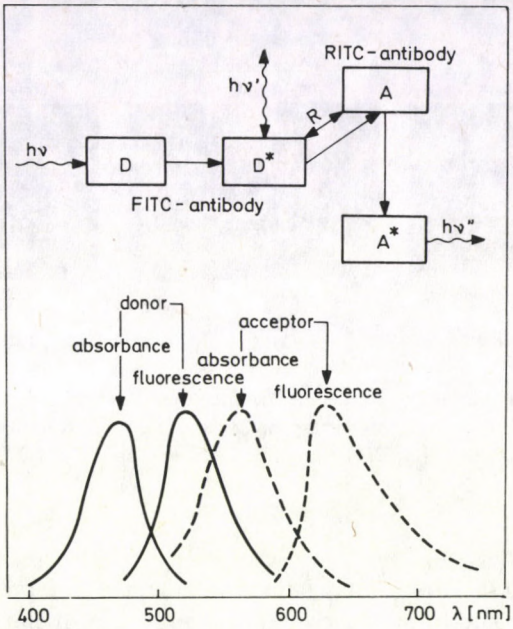


Fig. 8. Fluorescence excitation and emission spectra of FITC and TRITC conjugated ligands

R_0. R_0 is the distance between the donor and the acceptor where the probability of the emission of the excitation energy by the donor and acceptor is

$$ET = \frac{R_0^6}{R_0^6 + R^6}, \tag{1}$$

$$R_0^6 = 8.8 \cdot 10^{-25} \cdot Q_D \cdot \kappa^2 \cdot n^{-4} \cdot \int_0^\infty \varepsilon_A(\lambda) F_D(\lambda) \lambda^{-4} \, d\lambda \, \text{cm}^6,$$

where $\varepsilon_A(\lambda)$ is the molar extinction coefficient of the acceptor, $F_D(\lambda)$ is the normalized fluorescence spectrum of the donor, λ is the wavelength in cm, Q_D is the quantum yield of the donor, κ^2 is an orientation factor determined by the spatial relationships of the donor and acceptor, i.e. by the angle between their respective absorption and emission oscillators, and n is the refractive index of the medium. ET can efficiently be determined between donor and acceptor separated within 2—8 nm.

The transfer efficiency (ET) is related to the emission of the acceptor excited through the donor by the following expression:

$$ET = EA_D^A - \alpha ED_D^D - \beta EA_A^A. \tag{2}$$

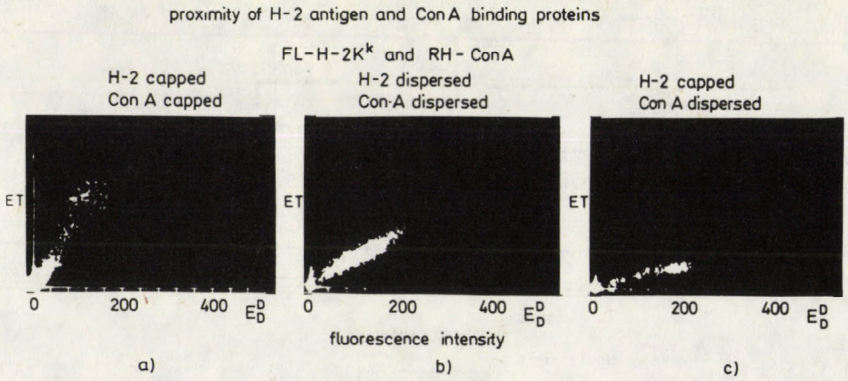


Fig. 9. Fluorescence resonance energy transfer measurements on single cell basis. The dots on photographs represent cells and the coordinates of dots correspond to parameters of cells shown on the $X(E_D^D)$ and $Y(ET)$ axis. The cells were stained with FITC conjugated Con *A* and TRITC conjugated anti-$H-2K^k$ immunoglobulin simultaneously

EA and *ED* represent the fluorescence intensity of the acceptor and donor, respectively. The superscripts and subscripts stand for donor and acceptor emission and excitation, respectively. *A* represents the emission or excitation of the acceptor, while *D* denotes the donor excitation and emission peaks. The emission of the donor is excited at the excitation site of the donors, detecting the emission at the donor emission site (the donor excitation is provided by the 488 nm argon-ion laser line, the emission is detected through a 515 nm narrow bandpass interference filter in front of the photomultiplier). ED_D^D will be proportional also to the "unwanted" emission of the donor at the acceptor emission site (>590 nm). EA_A^A, the emission of the acceptor excited at 555 and detected above 590 nm is proportional to the direct excitability of the acceptor at the excitation wavelength of the donor. The λ and β correcting factors, multiplied by the ED_D^D and EA_A^A and subtracted for the EA_D^A value result in a value directly proportional to the transfer efficiency. Fig. 9 shows a plot of our experiments where the proximity relationships of cell surface receptors like Con *A* and $H-2K^k$ are represented by the slope of the fitted curve. Con *A* is a protein of a molecular weight of 120 000, which binds specifically to a certain kind of oligosaccharide on the cell surface. $H-2K^k$ is an antigen determining the self (i.e. characterizing) of a mouse cell. This method recently developed by Drs T. M. and Donna Jovin has served as a tool to detect an uneven distribution between cell surface ligand binding sites at nm level [15, 16].

The number of Con *A* receptors renders it possible to have the fluorescein labeled proteins (FITC-Con *A*) at a distance from the rhodamine labeled other cell surface binding ligand, the anti-$H-2K^k$ antibody, so that it also fulfils the distance requirement of resonance energy transfer.

As Fig. 10 shows, certain cell surface elements may have a distribution having the specifically labeled ligand binding sites collected at one pole of the cell like a "cap". This

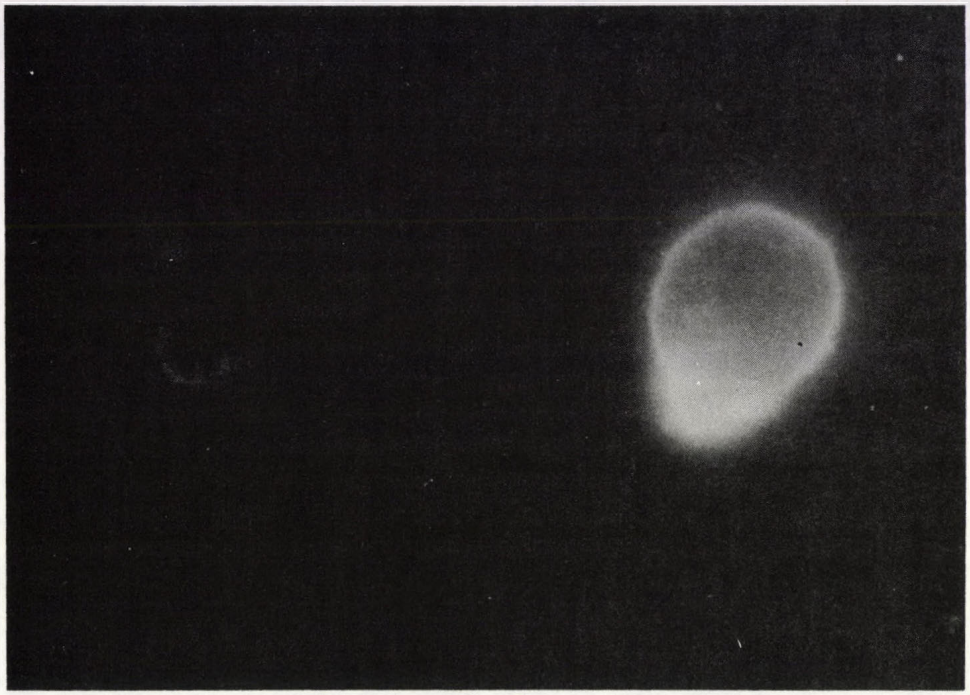


Fig. 10. Cell surface pattern of fluorescence of the same cell at green excitation (left — cap formation) and blue excitation (right — uniformly distributed fluorescence). The cells were labeled with FITC conjugated Con *A* and TRITC conjugated anti-$H-2K^k$ immunoglobulins

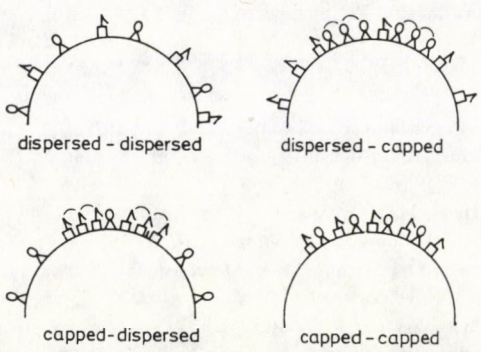


Fig. 11. Schematic drawings of the distribution of binding sites in dispersed and cap forming state on the cell surface

phenomenon, called capping, with quite a number of interesting cell-biological implications, is perfectly suitable to alter the relative distribution of the cell surface elements in case they are "capped" separately. Fig. 11 demonstrates a simplified scheme of the capping and also the disperse distribution of the ligands, and, consequently, the ligand binding sites.

Though the phenomenon outlined above has already been known for ten years, the exact biophysical investigation of the possible distribution patterns of cell surface receptors and other ligand binding sites is still missing [17]. The use of fluorescence energy transfer as a slide rule at molecular level, and the application of the cell sorter and analyzer giving individual information at single cell level, without losing the advantage of the cell populations, bring us really closer to a better understanding of cellular biophysics.

Acknowledgements

Part of the work has been elaborated during one of the authors' (S.D.) guest professorship at the Max-Planck Institut für Biophysikalische Chemie, Göttingen, FRG. Thanks are due to the Deutsche Forschungsgemeinschaft for sponsoring the work.

A more detailed account on some of the experiments outlined here briefly will be published elsewhere.

References

1. L. A. Herzenberg and L. A. Herzenberg, Handbook of Experimental Immunology, Vol. 2, Chapter 22, D. M. Weir (ed.), Blackwell Scientific Publications, Oxford, England, 1978.
2. J. Schlammadinger and R. Gáspár, Jr., Acta Biol. Acad. Sci. Hung., **30,** 283, 1979.
3. R. Gáspár, Jr., S. Damjanovich and B. Somogyi, MTA Biol. Oszt. Közl., **23,** 205, 1980 (in Hungarian).
4. R. Peters, Cell. Biol., **5,** 733, 1981.
5. D. E. Koppel, Biophys. J., **28,** 281, 1979.
6. D. D. Thomas, Biophys. J., **24,** 439, 1978.
7. C. Moore, D. Boxer and P. Garland, FEBS Letters, **108,** 161, 1979.
8. G. C. Salzman, P. F. Mullaney, B. J. Price, Flow Cytometry and Sorting, M. R. Melamed, P. F. Mulaney, M. L. Mendelsohn (eds.) Wiley Medical Publication, New York, 1979.
9. M. R. Loken and L. A. Herzenberg, Annals New York Academy of Sciences, **254,** 163, 1975.
10. G. Szabó, Jr., A. Kiss and S. Damjanovich, Cytometry, **2,** 20, 1981.
11. T. M. Jovin, Flow Cytometry and Sorting, M. R. Melamed, P. F. Mullaney, M. L. Mendelsohn (eds.) Wiley Medical Publication, New York, 1979.
12. J. Szöllősi, P. Kertai, B. Somogyi and S. Damjanovich, J. Histochem. Cytochem., **29,** 503, 1981.
13. I. Ember, I. Arany, P. Rády, J. Szöllősi, Zs. Tatár-Kiss, V. Thomázy and P. Kertai, Kísérletes Orvostudomány, **33,** 101, 1981 (in Hungarian).
14. J. Szöllősi, I. Ember, L. Trón and P. Kertai, Cytometry, **3,** 110, 1982.
15. S. S. Chan. D. J. Arndt-Jovin and T. M. Jovin, J. Histochem. Cytochem., **27,** 56, 1979.
16. T. M. Jovin and D. J. Arndt-Jovin, Proc. 8th Internatl. Congr. on Photobiology, July 20—25, 1981, Strasbourg, Plenum, New York, 1982. (in press).
17. S. Damjanovich, B. Somogyi and L. Trón, Neural Communication and Control, Gy. Székely, E. Lábos, S. Damjanovich (eds), Adv. Physiol. Sci., **30,** 9, 1980.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 247—254 (1982)

COMPARISON OF THE ORBITALS OF NEON, ARGON AND KRYPTON CALCULATED BY THE HARTREE–FOCK AND THE Xα METHODS WITH SEVERAL VALUES OF α *

R. GÁSPÁR and Á. NAGY

INSTITUTE OF THEORETICAL PHYSICS, KOSSUTH LAJOS UNIVERSITY,
4010 DEBRECEN, HUNGARY

The orbitals of Neon, Argon and Krypton were calculated by the Xα method with parameters α theoretically determined by one of the authors [1—3].

It is pointed out by calculating expectation values of some powers of the radius that the orbitals of this method agree well with those of the Hartree–Fock method. For comparison values for the orbitals calculated with $\alpha = 1$ and 2/3 are also presented and some experimental values are included.

Introduction

In 1951 Slater [4] was the first to suggest an approximate exchange potential proportional to the one-third power of the local electronic density. In 1954 Gáspár [5] and later in 1965 Kohn and Sham [6] proposed another exchange potential which was the two-thirds of Slater's. Since that time calculations with several numerical coefficients in the exchange potential have been carried out and this new theory using local exchange potential has been denoted by the term "Xα method".

For the nearly three decades of the Xα method tremendous progress has been made in this theory but the problem of the parameter α still seems to be an unsettled one. In his previous papers [1—3] Gáspár introduced a new, theoretical way of calculation of the exchange parameters α. In this paper a comparison of orbitals of Ne, Ar and Kr calculated with several values of α is presented. It is shown that the orbitals of Gáspár's new method are the nearest in approximating the Hartree–Fock orbitals.

* Dedicated to Prof. I. Tarján on his 70th birthday.

The theory

The total energy of an electron system is given by

$$\langle EX\alpha\rangle = \sum_i n_i \int u_i^*(1) f_1 u_i(1)\,\mathrm{d}v_1 + \frac{1}{2}\int \rho(1)\,\rho(2) g_{12}\,\mathrm{d}v_1\,\mathrm{d}v_2$$
$$+ \frac{1}{2}\int [\rho_\uparrow(1)\,U_{X\uparrow}(1) + \rho_\downarrow(1)\,U_{X\downarrow}(1)]\,\mathrm{d}v\,, \tag{1}$$

where f_1 is the one-electron operator acting on the first electron, i.e. the sum of the kinetic energy and its potential energy in the field of the nuclei. g_{12} is the two-electron operator, the Coulomb interaction energy of the first and second electrons. The charge densities of the electrons with spin up and down have the form

$$\rho_\uparrow = \sum_{i\uparrow} n_i u_i^* u_i \tag{2}$$

and

$$\rho_\downarrow = \sum_{j\downarrow} n_j u_j^* u_j\,, \tag{3}$$

where u_i are the spin orbitals. The total charge density is

$$\rho = \rho_\uparrow + \rho_\downarrow\,. \tag{4}$$

The last term in the total energy expression provides the exchange energy.

Using the variational principle the Xα one-electron equations are

$$[f_1 + \int \rho(2) g_{12}\,\mathrm{d}v_2 + V_{X\alpha\uparrow}(1)]\,u_{i\uparrow}(1) = \varepsilon_{iX\alpha\uparrow}\,u_{i\uparrow}(1) \tag{5}$$

and similar equations for spin orbitals with spin down. The exchange potential is given by

$$V_{X\alpha\uparrow}(1) = \frac{1}{2}\left[U_{X\uparrow}(1) + \rho_\uparrow(1)\frac{\delta U_{X\uparrow}(1)}{\delta\rho_\uparrow(1)}\right]. \tag{6}$$

In the Xα method the exchange potential in the total energy expression is proportional to the 1/3 power of the charge density

$$U_{X\alpha\uparrow}(1) = -9\alpha\left[\frac{3}{4\pi}\rho_\uparrow(1)\right]^{1/3}. \tag{7}$$

(The energy is in Rydbergs.) The exchange potential in the one-electron equations is

$$V_{X\alpha\uparrow}(1) = -6\alpha\left[\frac{3}{4\pi}\rho_\uparrow(1)\right]^{1/3}. \tag{8}$$

$\alpha=1$ gives the exchange potential proposed by Slater [4] and $\alpha=2/3$ provides the exchange potential suggested by Gáspár [5] and Kohn–Sham [6].

The Xα method of Slater contains a free parameter α. This permits to adjust the Xα method to the Hartree—Fock theory. The parameter α can be chosen so that the total Xα energy should be equal to the Hartree—Fock total energy. These parameters α_{HF} were calculated by Schwarz [9] and a method nearly as good as the Hartree—Fock method itself was obtained. There are, however, some problems with the parameter in the Xα method. Because the value of α is determined by adjusting, the Xα method is often considered as a semiempirical one. The parameter α_{HF} depends on the atomic number Z, it is not quite straightforward how to choose the value of α when the method is used in molecules.

To overcome these shortcomings of the Xα method some refinements have been made. Here the Xαβ method by Herman, Van Dyke and Ortenburger [7, 8] is discussed only. It is argued by these authors that the dependence of the parameter α on the atomic number Z is due to the inhomogeneity in the density distribution and a gradient term is introduced. The main advantage of the Xαβ method is that the parameters α and β are not Z dependent allowing a natural application to molecules and solids. But it has the disadvantage of having two adjustable parameters and a much more difficult way of computation.

One of the authors [1—3] has suggested a new way of theoretical introduction of the parameter in the Xα method. The exchange potential of an electron gas is given by

$$V_{X\uparrow}(1)=-8F(\eta)\left[\frac{3}{4\pi}\rho_{\uparrow}(1)\right], \tag{9}$$

where

$$F(\eta)=\frac{1}{2}+\frac{1-\eta^2}{4\eta}\ln\left|\frac{1+\eta}{1-\eta}\right| \tag{10}$$

and

$$\eta=\frac{k}{k_F} \tag{11}$$

is the reduced momentum of the electron considered and

$$E_{F\uparrow}=k_{F\uparrow}^2=(6\pi^2\rho)^{2/3} \tag{12}$$

is the Fermi energy. By an averaging process in the k space the averaged exchange potential is obtained

$$\bar{V}_{X\uparrow}=-8\left[\frac{3}{4\pi}\rho_{\uparrow}(1)\right]^{1/3}\frac{\int_{\eta_1}^{\eta_2}F(\eta)\eta^2\,d\eta}{\int_{\eta_1}^{\eta_2}\eta^2\,d\eta}. \tag{13}$$

This is an Xα exchange potential with the parameter

$$\alpha = \left\{\frac{1}{2}(\eta^3+\eta) - \frac{1}{4}(\eta^2-1)^2 \ln\left|\frac{1+\eta}{1-\eta}\right|\right\}_{\eta_1}^{\eta_2} [\{\eta^3\}_{\eta_1}^{\eta_2}]^{-1} . \tag{14}$$

If we average over all occupied states, i.e. $\eta_1 = 0$ and $\eta_2 = \frac{k_F}{k_F} = 1$, the parameter introduced by Slater ($\alpha = 1$) is obtained. On averaging over a thin shell near k_F, i.e. $\eta_1 = \frac{k_F - \varepsilon}{k_F}$ and $\eta_2 = \frac{k_F}{k_F} = 1$ with $\varepsilon \to 0$, $\alpha = 2/3$ is obtained. Between these two exchange potentials another exchange potential may be constructed by the introduction of an averaging process near the Fermi surface for a layer containing $\nu_\uparrow$ electrons in the unit volume, i.e. $\eta_1 = (1 - \nu_\uparrow/\rho_\uparrow)^{1/3}$ and $\eta_2 = 1$. $\nu_\uparrow = u_i^* u_i$ is the density of the reference electron. The exchange potential is given by

$$V_{X\alpha\uparrow} = V_{X\alpha shell\uparrow} = -6\alpha_{shell}\left(\frac{3}{4\pi}\rho_\uparrow\right)^{1/3} , \tag{15}$$

where

$$\alpha_{shell} = \frac{\rho_\uparrow}{\nu_\uparrow}\left\{1 - \frac{1}{2}\eta^3 - \frac{1}{2}\eta + \frac{1}{4}[\eta^2 - 1]\ln\left|\frac{1+\eta}{1-\eta}\right|\right\} \tag{16}$$

and

$$\eta = \left(1 - \frac{\nu_\uparrow}{\rho_\uparrow}\right)^{1/3} . \tag{17}$$

These exchange potentials are different for different shells. It is possible to carry out another averaging process for the electron system with a weight factor n_j

$$\alpha'_\uparrow = \frac{\sum_{j\uparrow} n_j \alpha_{shell}}{\sum_j n_j} , \tag{18}$$

where n_j is the number of electrons in the shell j. Now α'_1 is the same for each electron with the same spin but it depends on the position (1). It is possible to get a constant parameter α_G so that the mean squared deviation of $V_{X\alpha_G\uparrow}$ and $V_{X\alpha'\uparrow}$ be a minimum. The parameter α_G can be determined in a self-consistent way, i.e. starting from a trial atomic (or molecular) potential the spin orbitals u_i are obtained by solving the Xα equations. From the spin orbitals the new exchange parameter α_G is obtained and then the new potential is calculated. This process has to be carried out until self-consistency is achieved.

Results and discussion

The expectation values of powers of the radius $\langle \frac{1}{r} \rangle$, $\langle r \rangle$ and $\langle r^2 \rangle$, in atomic units, are calculated by the Hartree—Fock and the Xα method with parameters $\alpha = 1$, $\alpha = 2/3$, α_{HF} and α_G and they are presented in Tables I—III, which show that there is a good agreement between these orbitals. It is the results gained with the parameter α_G that are closest to the Hartree—Fock values [10]. The expectation values with α_{HF} are nearly as close to the Hartree—Fock values as those calculated with α_G. Results with $\alpha_S = 1$ and $\alpha_{GKS} = 2/3$ are farther from the Hartree—Fock ones, though the expectation value obtained with α_{GKS} are sometimes rather close to the Hartree—Fock ones.

These expectation values show the behaviour of the orbitals at different values of the radius. The values $\langle \frac{1}{r} \rangle$ are sensitive to the behaviour of the orbitals near the nucleus. The expectation values $\langle r \rangle$ and $\langle r^2 \rangle$ show how good the orbitals are in the middle and outer regions of the atom.

The expectation value $\langle r^2 \rangle$ of the atom is proportional to the experimentally measurable diamagnetic susceptibility. These values calculated by different methods and gained from experiment can be found in Table IV. A comparison shows a good agreement between the results obtained by Hartree—Fock and Xα calculations with α_G and α_{HF} and the experiment. Unfortunately, the uncertainties in experimental values do not make it possible to decide which orbitals are the best.

Table I

Expectation values of powers of the radius for some orbitals in Ne (in atomic units)

		$\langle 1/r \rangle$	$\langle r \rangle$	$\langle r^2 \rangle$
1*s*				
HF		9.618 09	0.157 63	0.033 47
Xα	$\alpha = 1$	9.674 32	0.156 84	0.033 15
	$\alpha = 2/3$	9.558 68	0.158 98	0.034 09
	α_{HF}	9.581 28	0.158 65	0.033 90
	α_G	9.588 70	0.158 42	0.033 84
2*s*				
HF		1.632 56	0.892 11	0.967 08
Xα	$\alpha = 1$	1.674 14	0.874 38	0.928 68
	$\alpha = 2/3$	1.624 51	0.906 00	1.004 67
	α_{HF}	1.633 58	0.899 97	0.989 88
	α_G	1.636 77	0.897 91	0.984 86
2*p*				
HF		1.435 35	0.965 27	1.228 46
Xα	$\alpha = 1$	1.514 19	0.912 78	1.097 49
	$\alpha = 2/3$	1.431 03	0.989 98	1.326 81
	α_{HF}	1.447 22	0.973 66	1.275 64
	α_G	1.452 79	0.968 24	1.259 00

Table II

Expectation values of powers of radius for some orbitals in Ar (in atomic units)

		$\langle 1/r \rangle$	$\langle r \rangle$	$\langle r^2 \rangle$
1*s*				
HF		17.553 31	0.086 10	0.009 96
Xα	$\alpha = 1$	17.611 85	0.085 82	0.009 89
	$\alpha = 2/3$	17.500 58	0.086 43	0.010 03
	α_{HF}	17.519 25	0.086 33	0.010 01
	α_G	17.517 09	0.086 34	0.010 01
2*s*				
HF		3.555 33	0.412 28	0.201 22
Xα	$\alpha = 1$	3.574 70	0.410 78	0.200 01
	$\alpha = 2/3$	3.511 07	0.418 14	0.207 47
	α_{HF}	3.521 58	0.416 90	0.206 20
	α_G	3.520 38	0.417 04	0.206 34
2*p*				
HF		3.449 99	0.375 33	0.174 34
Xα	$\alpha = 1$	3.507 87	0.369 77	0.169 57
	$\alpha = 2/3$	3.437 86	0.378 41	0.178 02
	α_{HF}	3.449 44	0.376 95	0.176 57
	α_G	3.448 15	0.377 12	0.176 74
3*s*				
HF		0.961 98	1.422 17	2.350 41
Xα	$\alpha = 1$	1.013 57	1.370 21	2.179 14
	$\alpha = 2/3$	0.984 33	1.417 94	2.347 82
	α_{HF}	0.988 95	1.410 09	2.319 55
	α_G	0.988 45	1.410 96	2.322 68
3*p*				
HF		0.814 07	1.662 96	3.310 85
Xα	$\alpha = 1$	0.875 21	1.563 80	2.918 47
	$\alpha = 2/3$	0.824 67	1.682 10	3.435 21
	α_{HF}	0.833 15	1.660 68	3.337 15
	α_G	0.832 22	1.663 04	3.347 93

In conclusion, we can state that the new way in the calculation of the exchange parameter α in the Xα method provides good orbitals. Although the orbitals obtained with the parameter α_{HF} are nearly as good as those calculated with α_G, this new method has two very important advantages. First, this method contains no adjustable parameters and so it can be regarded as an ab initio method. Second, this method can be applied to any system of molecules or solids without further theoretical difficulties.

Table III

Expectation values of powers of radius for some orbitals in Kr (in atomic units)

		$\langle 1/r \rangle$	$\langle r \rangle$	$\langle r^2 \rangle$
1*s*				
HF		35.498 28	0.042 44	0.002 41
Xα	$\alpha = 1$	35.558 67	0.042 36	0.002 40
	$\alpha = 2/3$	35.448 83	0.042 51	0.002 42
	α_{HF}	35.461 54	0.042 49	0.002 42
	α_G	35.457 28	0.042 50	0.002 42
2*s*				
HF		7.918 84	0.187 26	0.041 28
Xα	$\alpha = 1$	7.933 53	0.186 83	0.041 07
	$\alpha = 2/3$	7.877 25	0.188 10	0.041 63
	α_{HF}	7.883 86	0.187 95	0.041 56
	α_G	7.881 58	0.187 99	0.041 59
2*p*				
HF		7.868 44	0.161 88	0.032 00
Xα	$\alpha = 1$	7.927 18	0.160 68	0.031 51
	$\alpha = 2/3$	7.866 26	0.162 02	0.032 06
	α_{HF}	7.873 44	0.161 86	0.031 99
	α_G	7.870 93	0.161 91	0.032 01
3*s*				
HF		2.637 56	0.537 80	0.331 73
Xα	$\alpha = 1$	2.655 90	0.535 56	0.329 27
	$\alpha = 2/3$	2.620 19	0.542 90	0.338 73
	α_{HF}	2.624 30	0.542 04	0.337 61
	α_G	2.622 95	0.542 33	0.337 98
3*p*				
HF		2.522 78	0.542 63	0.343 54
Xα	$\alpha = 1$	2.555 17	0.537 01	0.336 86
	$\alpha = 2/3$	2.514 25	0.546 44	0.349 36
	α_{HF}	2.518 99	0.545 32	0.347 87
	α_G	2.517 44	0.545 69	0.348 36
3*d*				
HF		2.276 94	0.550 88	0.317 48
Xα	$\alpha = 1$	2.331 40	0.538 32	0.355 08
	$\alpha = 2/3$	2.274 53	0.554 86	0.379 33
	α_{HF}	2.281 22	0.552 86	0.376 34
	α_G	2.279 05	0.553 51	0.377 31
4*s*				
HF		0.804 19	1.629 38	3.040 32
Xα	$\alpha = 1$	0.858 89	1.551 35	2.751 59
	$\alpha = 2/3$	0.833 03	1.603 68	2.955 21
	α_{HF}	0.835 98	1.597 51	2.930 78
	α_G	0.835 02	1.599 54	2.938 83
4*p*				
HF		0.669 22	1.951 61	4.455 08
Xα	$\alpha = 1$	0.720 58	1.835 61	3.926 51
	$\alpha = 2/3$	0.677 37	1.970 73	4.595 63
	α_{HF}	0.682 56	1.953 17	4.504 42
	α_G	0.680 84	1.958 98	4.534 54

Table IV

Expectation values $\langle r^2 \rangle$ calculated with the total radial density of the atom for Ne, Ar and Kr. (Expectation values are in atomic units.) The experimental results are obtained from diamagnetic susceptibility [11]

		Ne	Ar	Kr
HF		9.372	26.034	39.530
Xα	$\alpha = 1$	8.509	23.306	35.569
	$\alpha = 2/3$	10.038	26.810	40.331
	α_{HF}	9.701	26.154	39.694
	α_G	9.591	26.226	39.904
Exp [11]		8.4—9.8	14.2—24.9	34.8—36.9

References

1. R. Gáspár, Acta Phys. Hung., **35,** 213, 1974.
2. R. Gáspár, Acta Phys. et Chim. Debr., XIX, 7, 1974.
3. R. Gáspár, Acta Phys. et Chim. Szeged., XX, 321, 1974.
4. J. C. Slater, Phys. Rev., **81,** 385, 1951.
5. R. Gáspár, Acta Phys. Hung., **3,** 263, 1954.
6. W. Kohn and L. J. Sham, Phys. Rev., **140A,** 1133, 1965.
7. F. Herman, J. P. Van Dyke and I. B. Ortenburger, Phys. Rev. Lett., **22,** 807, 1969.
8. F. Herman, J. P. Van Dyke and I. B. Ortenburger, Int. J. Quant. Chem., **S3,** 827, 1970.
9. K. Schwarz, Phys. Rev., **B5,** 2466, 1972.
10. C. Froese-Fischer, Atomic Data, **4,** 301, 1972.
11. Landolt-Börnstein, Zahlenwerte und Funktionen aus Physik, Chemie, Astronomie, Geophysik und Technik (Springer-Verlag, Berlin, 1950), 6 Auflage, Band I.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 255—262 (1982)

MATHEMATICAL MODELLING OF SEGREGATION PROCESSES IN MICROBIAL POPULATIONS CONTAINING A SINGLE PLASMID SPECIES*

G. MÜLLER, D. NOACK, R. SCHORCHT

CENTRAL INSTITUTE FOR MICROBIOLOGY AND EXPERIMENTAL THERAPEUTICS, JENA, GDR

and

S. GÁSPÁR, L. HERÉNYI

INSTITUTE OF BIOPHYSICS, SEMMELWEIS MEDICAL UNIVERSITY, BUDAPEST, HUNGARY

This paper presents two mathematical models for plasmid segregation processes of bacteria containing a single plasmid species with various copies. A Markov chain model allows to study the general properties of these processes for stable inheritance of the plasmid and for plasmid loss during continuous culture. The other, more specialized model is based upon the following assumptions: (1) random replication of the plasmid copies concerning the replicated copy number, (2) equal number partitioning of the copies after replication into the daughter cell.

Introduction

The stability of inheritance of a multi-copy-plasmid in a chemostat culture depends on the type of plasmid, its host cell and the culture conditions (Adams et al [1], Godwin and Slater [2], Roth et al [3], Jones et al [4], Noack et al [5]). There are two possible alternatives: either stable inheritance of the plasmid for a long time of cultivation or the loss of the plasmid. The kinetics of plasmid loss in a chemostat and the conditions for their stable maintenance were studied in several mathematical models (Baumberg [6], Stewart and Levin [7], Levin et al [8], Levin and Stewart [9], Levin and Rice [10]). These models describe the dynamic behaviour of the whole population without consideration of the various mechanisms of the segregation process. Other authors discussed the problems of plasmid segregation in relation to these mechanisms (Novick et al [11], [12], [13], Uhlin and Nordström [14], Cullum and Broda [15], Molin and Nordström [16], Nordström et al [17] I, II, Hashimoto-Gotoh and Timmis [18]).

In addition to the experimental studies stochastic models allow a quantitative comparison between theoretical and experimental data and their interpretation.

Such models were based on various assumptions for the basic mechanisms of segregation in the process of inheritance of the plasmid copies to the daughter cells at

* Dedicated to Prof. I. Tarján on his 70th birthday.

cell division. Mathematical models of the mechanisms of replication (rep), replication control (cop) and partitioning (par) of plasmid copies at cell division were discussed quantitatively only in connection with problems of plasmid incompatibility (Ishii et al [19], Novick and Hoppensteadt [20], Cullum and Broda [15], Nordström et al [17] I, II). All discussed mathematical models for plasmid incompatibility have in common the dissection of segregation process of two different plasmid types into two processes: replication of the plasmid copies and their partitioning into the daughter cells, as well as they are common in the assumption of a constant whole copy number of both plasmid types in all generations and in each cell after replication. The assumption of a $N \rightarrow 2N \rightarrow N$ cycle in the course of replication and partitioning allows to calculate the alteration of relative frequencies of mixed plasmid states and the averaged segregation probability. Conclusions can be drawn on the mode of replication and partitioning of the plasmid.

In this paper we present two stochastic models describing the loss of a single multi-copy-plasmid during continuous culture. In both models we assume a random replication of the plasmid copies and a random partitioning, for instance a binomial partitioning, or an equal number partitioning, of the replicated plasmid copies. The mathematical description of the plasmid segregation in agreement with genetic results is a difficult task, because the average number of plasmid copies per cell of the whole population decreases in time in contrast to the mentioned models of plasmid incompatibility. We have solved this problem in two ways. In the first model the basic properties of the segregation process of a single multi-copy-plasmid are discussed. By means of special matrices R and Q for the replication and partition the second mathematical model describes in detail the segregation rate of the plasmid and the proportion of plasmid-free and plasmid-containing cells in the whole population.

A general segregation model

In the first model we study the basic properties of the segregation process described as a Markov chain. Let τ be the number of generations ($\tau = 0, 1, 2, 3, \ldots$) and ξ_τ a random variable which describes the copy number per cell in the generation τ. ξ_τ has possible values $k = 0, 1, 2, \ldots, N$. We call

$$p(\xi_\tau = k) = p_\tau(k)$$

the probability that $\xi_\tau = k$. The vector $(p_0(k))$ $k = 0, 1, 2, \ldots N$ is the initial distribution vector of the plasmid copy numbers in the population. A Markov chain with discrete time is defined by the conditional probability:

$$\pi(\xi_{\tau+1} = k | \xi_0 = j_0, \xi_1 = j_1, \ldots, \xi_\tau = j_\tau) = \pi(\xi_{\tau+1} = k | \xi_\tau = j_\tau) = \pi_{jk} .$$

π_{jk} is the transfer probability of the j-th state to the k-th state after one generation. We assume that π_{jk} is independent of the number of generations. Therefore, the Markov

chain is homogeneous and we have the transfer matrix (π_{jk}) $j, k = 0, 1, \ldots, N$. This is a stochastic matrix with the properties:

$$\sum_{k=0}^{N} \pi_{jk} = 1\,, \qquad j = 0, 1, \ldots, N$$

and

$$\pi_{jk} \geqq 0\,, \qquad j, k = 0, 1, \ldots, N\,.$$

Writing $p_\tau(k)$, $\tau = 0, 1, 2, \ldots$ as a line vector P_τ and (π_{jk}) as the square matrix Π, we have, for example,

$$P_\tau = P_{\tau-1} \cdot \Pi\,. \tag{1}$$

This recursive formula allows to calculate the distribution vector of plasmids in the generation τ

$$P_\tau = P_0 \cdot \Pi^\tau\,. \tag{2}$$

Π^n is the n-fold matrix product. In our model we consider the segregation process from one generation to the next, described by the stochastic matrix Π, as a result of two processes, replication and partition of the plasmid copies. We represent the processes with two matrices R and Q which have the form:

$$R = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & R_{11} & \cdot & \cdot & \cdot & \cdot & R_{1N} \\ 0 & 0 & & R_{22} & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ 0 & \cdot & \cdot & \cdot & \cdot & 0 & R_{NN} \end{bmatrix}$$

and

$$Q = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ P_{10} & P_{11} & & & & 0 \\ \cdot & & P_{22} & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & 0 \\ P_{N0} & & & & & P_{NN} \end{bmatrix}.$$

Then the transfer matrix Π may be written

$$\Pi = R \cdot Q = \begin{bmatrix} 1 & 0 & . & . & . & . & . & . & . & 0 \\ \pi_{10} & \pi_{11} & . & . & . & . & . & . & . & \pi_{1N} \\ \pi_{20} & \pi_{21} & \pi_{22} & . & . & . & . & . & . & \pi_{2N} \\ . & & & & & & & & & . \\ . & & & & & & & & & . \\ . & & & & & & & & & . \\ \pi_{NO} & . & . & . & . & . & . & . & . & \pi_{NN} \end{bmatrix} . \tag{3}$$

a) Non-segregation case

The case of stable inheritance of the plasmid during continuous culture is described by the stochastic matrix Π with the condition for the first column

$$\pi_{i0} = \begin{cases} 1, & i=0 \\ 0, & i=1, 2, \ldots, N \end{cases}$$

(no plasmid containing cells change over into the plasmid free state).

Eq. (2) shows that in this case the first element of the initial distribution vector P_0 does not change during the process of cultivation. From the matrix Π we can obtain the submatrix Π' without the first column and first line of Π. The properties of Π' usually satisfy the conditions of the ergodic theorem for Markov chains and therefore there exists in each case a stationary limit distribution P'_∞ independent of the initial probability distribution P'_0:

$$\lim_{n \to \infty} P'_0 \Pi'^n = P'_\infty .$$

b) Segregation case

The case of continuously decreasing proportion of plasmid containing cells in the population is characterized by the special form of matrix Π:

$$\pi_{00} = 1$$

and

$$\pi_{i0} > 0 \text{ at least for one of } i .$$

The study of the asymptotic behaviour of these homogeneous Markov chains is in direct connection with the existence of eigenvalues of the matrix Π'. With the known notation we get the matrix equation:

$$\lambda P'_\tau = P'_\tau \Pi' \tag{4}$$

with λ as an eigenvalue of Π' and P'_τ a corresponding eigenvector. If $\lambda = 1$, the corresponding eigenvector P'^*_τ in the equation $P'^*_\tau = P'^*_\tau \Pi'$ is the stationary limit distribution of case a) independent of P'_0.

From the standard Perron—Frobenius theory for positive matrices (Canning [21]) and from the detailed classification of the various possibilities of Markov chains and their asymptotic behaviour (Bartlett [22], Feller [23]) we can conlude for the segregation case that eigenvalues $\lambda < 1$ of the reduced matrix equation (4) will appear.

From this result and Eq. (1) it can be concluded that the plasmid copy distribution functions of the plasmid containing cell population are not changed in the course of long term cultivation. For instance the mean value of copies per cell of the plasmid containing cell population is constant in the process of segregation.

Specialized segregation model

In the first model the processes of replication and partitioning are not specified, in the second one, however, explicit calculations will be carried out. We assume a random replication process in the following sense (not in the sense of genetic replication models) and describe this process with the replication factor $\varkappa$, $0 \leqq \varkappa \leqq 1$.

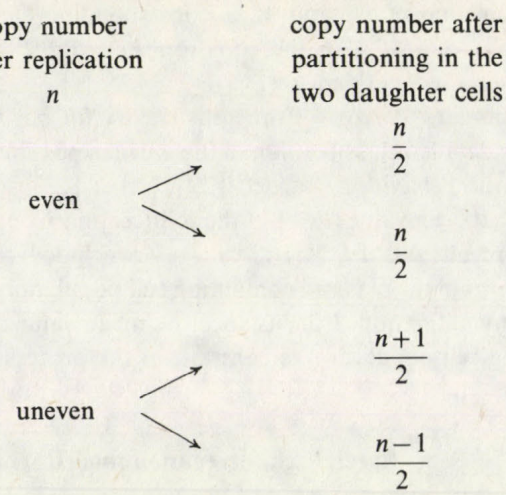
generation	copy number before replication	mean value of copy number after replication
τ	z_τ	$z_\tau + \varkappa z_\tau$

Assumption: after the process of replication the copy number per cell undergoes a binomial distribution. Let P_{ks} be the transfer probabilities that k copies result in s copies after replication. Then

$$p_{ks} = \binom{k}{s-k} \varkappa^{s-k}(k) \, [1 - \varkappa(k)]^{2k-s} \tag{5}$$

with $k = 1, 2, \ldots, N$ and $s = k, \ldots, 2k$. The dependence $\varkappa = \varkappa(k)$ on the copy number makes a more general formulation possible.

The initial distribution vector P_0 can be chosen arbitrarily. We could show that the asymptotic behaviour of the model parameters are independent of P_0. The partitioning of the replicated copies is assumed to be equally distributed and is illustrated in the following scheme:



The equal number assortment model is in agreement with the experimental results of Hashimoto—Gotoh and Sekiguchi [24] for a low copy number plasmid (pSC 101, with 10—14 copies per cell). Novick et al [12] also interpreted experimental results with a high copy number plasmid (about 32 per cell) as supporting equal number partitioning. In contrast to these cases a lack of partitioning (par$^-$) leads to a random distribution of the plasmid molecules between the daughter cells at all divisions (binomial distribution of plasmid copy number in the process of assortment, Nordström et al [17] Part I).

With these assumptions we can calculate the series of distribution vectors of the copy numbers in each generation

$$\{P_\tau\}_{\tau=1,2,\ldots}$$

Further it is possible to calculate the segregation rate δ_τ for each generation as the transfer probability of cells to the plasmid free state:

$$\delta_\tau = \frac{1-\varkappa(1)}{2-(1-\varkappa(1))\sum_{i=0}^{\tau-1} B_i} \cdot B_\tau . \tag{6}$$

The terms B_i, $i=0, 1, \ldots, \tau$ are multiple sums, dependent on the initial distribution P_0, the replication factors $\varkappa(n)$, $n=1, 2, \ldots$ and the transfer probabilities p_{ks}. The experimental measurable proportion of plasmid free and plasmid containing cells can be determined as follows:

$$P(z_\tau = 0) = \frac{1-\varkappa(1)}{2} \sum_{i=1}^{\tau} B_i , \tag{7}$$

$$P(z_\tau > 0) = 1 - P(z_\tau = 0) , \qquad (z_\tau \leqq N) .$$

The segregation process can be described with the replication factor $0 \leqq \varkappa < 1$. The limiting value of segregation rate has been calculated. If the number of generations $\tau \to \infty$, then the segregation rate $\delta_\tau \to \frac{1-\varkappa(1)}{2}$. The function δ_τ depends on τ, the replication factor $\varkappa$ and the maximum copy number N. Our calculations have shown that δ_τ rapidly converges to the limit $\frac{1-\varkappa(1)}{2}$.

Discussion

A favourable assumption previously made in the mathematical descriptions of segregation processes concerning plasmid incompatibility was the constant whole number of plasmid copies per cell after replication. These models described the change in relative frequencies of mixed plasmid states and studied in detail the kinetics of the segregation process. With the aid of the first mathematical model proposed in this paper we analysed the more general properties of the structure of models for the segregation of a single plasmid type if in the segregation case the whole copy number decreases. The two important results are that in the non-segregation case the segregation models must have the property that the plasmid copy number has a limit distribution which is independent of the initial distribution and that the proportion of plasmid free cells is constant during cultivation. These results are in agreement with experimental results and are not trivial.

On the other hand, theoretical results of the real segregation process show that the proportion of plasmid containing cells decreases within the population during cultivation, but the plasmid distribution functions of this part of population remain unchanged.

We considered these general conditions of segregation models in the detailed calculations of the specialized segregation model in the second part of the paper. We have chosen special mechanisms for replication and partitioning, but the dependence of the replication factor $\varkappa$ on the appropriate copy number in the cell gives much more possibilities for use. If one assumes a decreasing function $\varkappa(k)$ as the copy number increases it is possible to describe the loss of the plasmid during continuous culture as well as to replace the failing assumption of a constant copy number used in models for incompatibility.

Acknowledgements

The authors would like to thank Dr. M. Roth for his valuable discussions and for useful comments on the manuscript.

References

1. J. Adams, Th. Kinney, S. Thompson, L. Rubin and R. B. Helling, Genetics, **91,** 627, 1979.
2. D. Godwin and J. H. Slater, J. Gen. Microbiol., **111,** 201, 1979.
3. M. Roth, G. Müller and D. Noack, Loss of the multi-copy resistance plasmid pBR 325 from Escherichia coli GY 2354 pBR 325 during continuous cultivation. In: S. Mitsuhashi, L. Rosival and V. Kromery, (eds.) Antibiotic resistance, Avicenum and Springer Verlag, Prague, Berlin, Heidelberg, New York, 1980, p. 143.
4. I. M. Jones, S. B. Primose, A. Robinson and D. C. Ellwood, Mol. Gen. Genet., **180,** 579, 1980.
5. D. Noack, M. Roth, R. Geuther, G. Müller, K. Undisz, Ch. Hoffmeier and S. Gaspar, Mol. Gen. Genet., **184,** 121, 1981.
6. S. Baumberg, Genet. Res. Camb., **28,** 11, 1976.
7. F. M. Stewart and B. R. Levin, Genetics, **87,** 209, 1977.
8. B. R. Levin, F. M. Stewart and V. A. Rice, Plasmid, **2,** 247, 1979.
9. B. R. Levin and F. M. Stewart, Genetics, **94,** 425, 1980.
10. B. R. Levin and V. A. Rice, Genet, Res. Camb., **35,** 241, 1980.
11. R. P. Novick and R. Brodeky, J. Mol. Biol. **68,** 285, 1972.
12. R. Novick, I. Wyman, D. Bonanchand and E. Murphy, Plasmid life cycles in Staphylococcus aureus. In "Microbiology — 1974" (D. Schlessinger ed.) pp. 115—129. Amer. Soc. Microbiol., Washington, D. C., 1975.
13. R. P. Novick and M. Schwesinger, Nature (London), **262,** 623, 1976.
14. B. E. Uhlin and K. Nordström, J. Bacteriol., **124,** 641, 1975.
15. J. Cullum and P. Broda, Genet. Res. Camb., **33,** 61, 1979.
16. S. Molin and K. Nordström, J. Bacteriol, **141,** 111, 1980.
17. K. Nordström, S. Molin and H. Aagaard-Hansen, Plasmid, **4,** 215, (Part I), 332, (Part II), 1980.
18. T. Hashimoto-Gotoh and K. N. Timmis, Cell, **23,** 229, 1981.
19. K. Ishii, T. Hashimoto-Gotoh and K. Matsubara, Plasmid, **1,** 435, 1978.
20. R. P. Novick and F. F. C. Hoppensteadt, Plasmid, **1,** 421, 1978.
21. C. Cannings, Advan. Appl. Probab., **6,** 260, 1974.
22. M. S. Bartlett, An Introduction to Stochastic Processes, University Press, Cambridge, 1955.
23. W. Feller, An Introduction to Probability Theory and its Applications, Vol. I, John Wiley and Sons, Inc., New York, London, Sydney, 1968.
24. T. Hashimoto-Gotoh and M. Sekiguchi, J. Bacteriol., **131,** 405, 1977.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 263—277 (1982)

COMPUTER MODELS FOR SOME ASPECTS OF EVOLUTION*

F. DEÁK and G. MARX

DEPARTMENT OF ATOMIC PHYSICS, ROLAND EÖTVÖS UNIVERSITY
1088 BUDAPEST, HUNGARY

> « L' esprit va dans son travail, de son désordre à son ordre. Il importe qu'il se conserve jusqu'à la fin des ressources de désordre, et que l'ordre qu'il a commencé de se donner ne se lie pas si complètement, ne lui soit pas un si rigide maître, qu'il ne puisse le changer et user de sa liberté initiale. »
>
> (Paul Valéry)

In this model a certain sequence of characters is supposed to have a "survival advantage" compared to any random sequence. The multiplication happens with errors of a fixed probability. By making use of this model, the evolutionary role of sequence length, life expectancy and mutation rate has been studied.

Evolution

Lifeless objects wear out, they decay, their manifold heads towards complete disorder. A population of living beings does not wear out, it is even able to improve itself. If one looks for the origin of this qualitative difference, it turns out that the essential threshold of life is the ability of self reproduction. In a noisy environment not all offsprings of a living being are identical, so there is always a spectrum of genetic information present in the population. In the actual environment the allele (the variant) with the highest survival value and the highest fecundity multiplies the fastest. Other alleles are decimated or eliminated by natural selection, but they are always reproduced by the mutations of the fittest allele. If the environment changes, another allele of the spectrum will fit the new environment in the best way, and this allele then takes over. In this way the population will adjust itelf to new conditions. In this information-theoretical sense "life is the undertaking of an information carrier, to produce as much copies as possible". This is an updated version of the saying that "the hen is an undertaking of an egg to make more eggs". From this point of view three layers of natural phenomena can be distinguished:

* Dedicated to Prof. I. Tarján, pioneer of exact approach to biology, on his 70th birthday.

I. The behaviour of material objects with few degrees of freedom can be described by deterministic equations of motion (equations of Newton, Maxwell, Schrödinger, Dirac...). These are symmetric with respect to time reversal, the motions described by their solutions are reversible. (E.g a mass point, a mathematical pendulum, a rigid body, a lone electron.)

II. The behaviour of an aggregate of matter with very many degrees of freedom is irreversible due to statistics. It is characterized by increasing disorder, by wearing out, by dissipating any concentrated energy from a single degree of freedom to many degrees of freedom of the aggregate. (Second Law of thermodynamics.) Examples: the universe or any piece of real matter.

III. A population of self reproducing structures is able to improve itself in a noisy environment, to increase the survival chances and multiplication rate of the member structures spontaneously. Natural selection eliminates the deficient copies, so the population of self reproducing structures does not wear out. In fact just the opposite is true: by making use of random mutations, the population is able to adjust itself to a changing environment. Its fate is irreversible as well, but now in a positive sense. (Darwinian evolution.) In accordance with the Second Law, the necessary condition for spreading and developing is a steady flow of free enthalpy to the self reproducing structures, so that the increase of entropy in the environment overcompensates the increase of organisation and information within the population.

In this information-theoretical sense the main task of life science is to understand the spontaneous origin, spreading and increase of the genetic information. No wonder this challenge has attracted the attention not only of biologists, but of chemists, physicists, mathematicians and computer scientists as well [1].

The first computer model

Since the discovery of the genetic code, the mechanism for the evolution of genetic information has been sought by a number of authors [2]. To understand the evolution of genetic information, Manfred Eigen and Peter Schuster have introduced a very simple, but very illustrative model [3]. Let us start with a random sequence of characters. (E.g. BAK GEVLNT GUPIF LESTKKM.) The given environment has been supposed to have the highest preference for a "sensible sequence" of these many characters (e.g. TAKE ADVANTAGE OF MISTAKE). If the characters of an actual sequence agree with the ideal sequence in several places, this actual sequence will have a higher multiplication rate than any other random sequence with a lesser degree of coincidence. In this way the "sensible sequence" will be selected from the manifold of random sequences within a few generations. Evidently, if the copying were 100 percent faithful, a population not containing the ideal sentence from the beginning would not have a possibility of evolution towards the "most sensible sentence". But the interplay of a slight error rate per character with natural selection makes such a spontaneous

increase of sensible information possible. A very high mutation rate, however, is harmful: it may wipe out any achievements of the evolution, before its fixation in the population with the help of natural selection. Eigen has worked with sentences of definite length, with a fixed preference for a unique "ideal sentence".He used the computer model to find the maximum size of inheritable information at a given mutation rate and to explore the advantages of a cyclic information structure (like *TAKE* ADVANTAGE OF MIS*TAKE*).

By generalizing Eigen's model, the present paper explores the influence of different internal parameters on the speed of evolution [4]. Such parameters are the length of words, their mutation rate, their life expectancy, the capability of changing the information length, etc. In this paper also we shall restrict ourselves to a fixed "target word". The case of "changing environment", that of "divergent evolution" with different target words, the possibility of a "symbiosis of information" will be explored in a subsequent paper.

Let us use an alphabet with 16 characters (e.g. the Hawaiian alphabet). A word may contain L characters. It will be compared to a previously fixed target word (e.g. LIFE). If only R letters are right ones at right places (e.g. in the case of LOTE one has $L=4$, but only $R=2$), the word is less "fit", so it will produce 2^{R+1} offsprings per generation. The "sensible word" ($R=L$) has the most offsprings: 2^{L+1}. The parent perishes after one generation, after having produced the offspring.

In general the offspring are copies of the parent word, but there is a chance for a "typing error" (mutation) at each copy. With a mutation probability μ any character of the parent word may be replaced by another, randomly chosen character of the alphabet. (E.g. an offspring of LIFE may be LOTE, another may be LIFE. The value of R may remain, may decrease or may increase.)

The number of words in the population will grow very fast from generation to generation, at the end like a fast diverging geometrical sequence. In order to control the number of digits, we express the composition of the population in percentages.

Influence of the mutation rate

Let us consider a population of words with the fixed length L. The number of the words, containing R right characters at right places is given by $y(R)$. These numbers make the $L+1$ dimensional population vector $\mathbf{y}$. The population vector will change from generation to generation due to selective multiplication and due to mutations. Its value is $\mathbf{y}_n$ in the n-th generation. For the next generation, any R-word will be replaced by its 2^{R+1} offspring (this is described by the spreading matrix $\mathbf{S}$). Some of the offspring are mutants, because any character of the parent word has a chance μ of being miscopied (this is described by the mutation matrix $\mathbf{M}$). So the population vector of the next generation can be obtained by a linear transformation from the former generation:

$$\mathbf{y}_{n+1} = \mathbf{M}\mathbf{S}\mathbf{y}_n \,. \tag{1}$$

The multiplication is described by the $(L+1)\times(L+1)$ sized diagonal spreading matrix

$$\mathbf{S}=\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 2 & 0 & \dots & 0 \\ 0 & 0 & 2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 2^{L+1} \end{bmatrix}. \tag{2}$$

A single mutation may replace either a right character with a wrong one $(R\to R-1)$, with the probability

$$P_{R-1,R}=\mu\cdot\frac{R}{L}\cdot\frac{16-1}{16-1}.$$

(Mutation rate times the chance of affecting a right character times the chance of choosing a wrong character from the alphabet.) Or the mutation may transform a wrong character into another wrong one $(R\to R)$, with a probability

$$P_{R,R}=\mu\cdot\frac{L-R}{L}\cdot\frac{16-2}{16-1}.$$

Or the mutation may transform any wrong character into the corresponding right one $(R\to R+1)$, with a probability

$$P_{R+1,R}=\mu\cdot\frac{L-R}{L}\cdot\frac{1}{16-1}.$$

Substituting a right character with the right one (at a given place it is the same!) is not mutation, it is faithful copying. So the total mutation probability of one character is

$$P_{R-1,R}+P_{R,R}+P_{R+1,R}=\mu.$$

A single-character mutation can be described by the following matrix:

$$\mathbf{P}=\mu\begin{bmatrix} \frac{14}{15}\frac{L}{L}, & \frac{15}{15}\frac{1}{L}, & 0, & 0, & \dots \\ \frac{1}{15}\frac{L}{L}, & \frac{14}{15}\frac{L-1}{L}, & \frac{15}{15}\frac{2}{L}, & 0, & \dots \\ 0, & \frac{1}{15}\frac{L-1}{L}, & \frac{14}{15}\frac{L-2}{L}, & \frac{15}{15}\frac{3}{L}, & \dots \\ 0, & 0, & \frac{1}{15}\frac{L-2}{L}, & \frac{14}{15}\frac{L-3}{L}, & \dots \\ \dots, & \dots, & \dots, & \dots, & \dots \end{bmatrix} \tag{3}$$

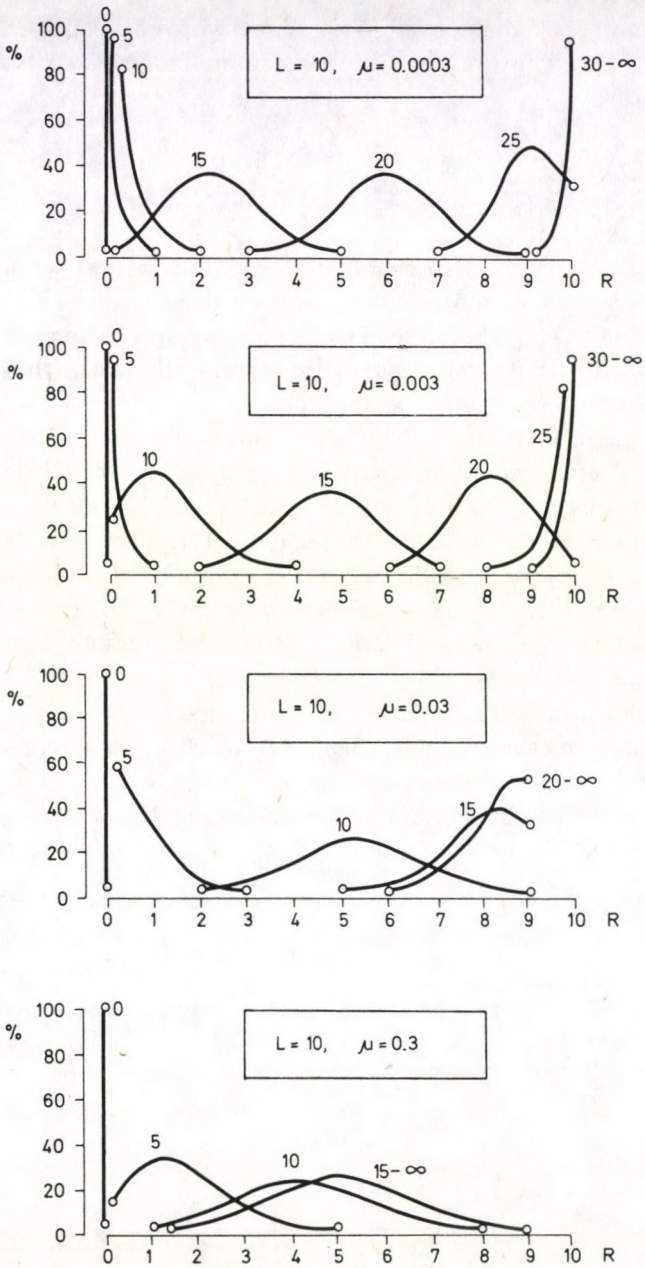


Fig. 1

The expected number of mutations is $\langle N \rangle = \mu L$ in a word of L characters, they may be described by a Poisson distribution. So the overall mutation matrix is constructed in the following way:

$$\mathbf{M} = \sum_{N=0}^{\infty} e^{-\mu L} \frac{(\mu L)^N}{N!} \mathbf{P}^N . \tag{4}$$

Let us now start with a 10 character word, all characters wrong ($R=0$). The evolution of the population towards the "target word" is shown in Fig. 1 for different mutation probabilities (μ). The curve of the population distribution is drawn for each fifth generation. (The serial number of the generation is indicated on the corresponding curve.) Table I shows the final population distribution. This limit is essentially the eigenvector of the product matrix $\mathbf{M} \cdot \mathbf{S}$, belonging to the highest eigenvalue.

For a fixed "target word" (in a fixed "environment") the "best fitting population" would consist of only perfect words ($R=10$), in this population the multiplication rate would be the highest: $2^{11}=2048$. This could be realized only with an exact reproduction, $\mu=0$, but this would make any evolution, any adjustment to a new environment impossible. A similar multiplication rate ($\simeq 2000$) can be realized by any value $\mu < 1\%$, so there is no practical advantage to suppress the mutations below that in this specific model.

If, however, the environment cannot be expected to last longer than 20 generations, if the target word will be changed at the 20th generation, there is no time enough to build up the limiting population by natural selection. Fig. 2 shows that within 20 generations the mutation probability $\mu=1\%$ builds up the highest multiplication rate. If the available time is so limited, a more faithful reproduction were

Table I

μ	R									
	1	2	3	4	5	6	7	8	9	10
0.	–	–	–	–	–	–	–	–	–	100
0.000 3	–	–	–	–	–	–	–	–	1	99
0.001	–	–	–	–	–	–	–	–	2	98
0.003	–	–	–	–	–	–	–	–	6	94
0.01	–	–	–	–	–	–	–	2	7	92
0.03	–	–	–	–	–	–	2	10	34	54
0.1	–	–	–	–	2	8	19	30	29	12
0.15	–	–	1	2	8	17	26	26	15	4
0.2	–	–	2	416	24	25	17	7	1	
0.3	1	4	11	20	25	21	12	5	1	–

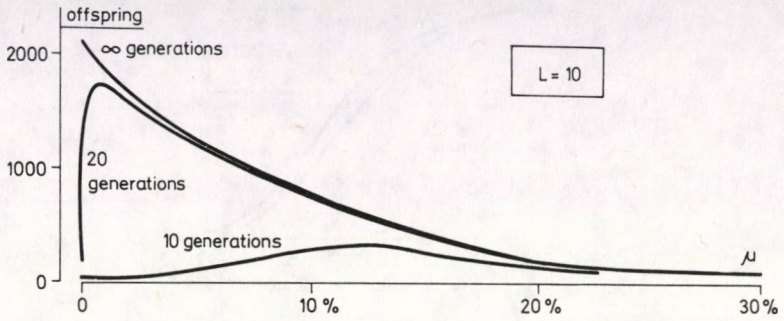


Fig. 2

even disadvantageous. In a fast changing world, in the case when the target word will change after 10 generations, the highest multiplication rate can be achieved by a larger copying error: $\mu = 10\%$. So the moral of this chapter is the following: under strictly controlled environmental conditions a small mutation rate is preferred, because this results in a narrow population spectrum. (This is the idea behind the hybrid chicken.) In a fast changing environment, however, a higher multiplication rate has a definite selective advantage. (This is a lesson learned well by the flu virus.)

Influence of the multiplication advantage

In this calculation the multiplication advantage per correct characters was taken to be 2. If one uses a higher multiplication advantage (Eigen used 2.718 and 10 per bit), it will make the final spectrum narrower. So a higher multiplication rate may compensate a larger mutation probability μ.

Influence of the longevity

In our model it has been assumed up to now that the parent word disappears immediately after having produced the first generation of her offspring. Let us change this condition: the life time of an individual word be 3 generations, i.e. any individuum bears three times. This certainly increases the overall population number, but the increase becomes negligible if the multiplication rate $\gg 1$. During the fast change of the population (R growing from 0 towards L) the surviving parents act as conservative elements, slowing down the increase of the average multiplication rate. (Table II gives the increase of the average multiplication rate in the first generations, if $L = 10$, $\mu = 0.000\,3$, the life time of a word is 1 or 3 generations. Moral: the longevity of words

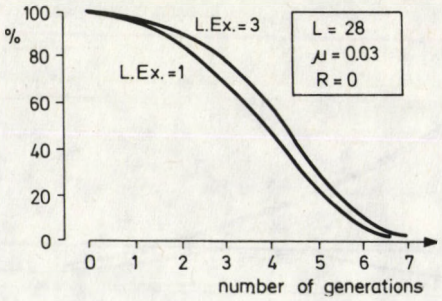


Fig. 3

Table II

Number of generations	Lifetime	
	1	3
0	2	2
1	2	2
2	2	2
3	2	2
4	2	2
4	2.02	2
6	2.02	2.02
7	2.04	2.04
8	2.10	2.08
9	2.18	2.14
10	2.40	2.32
11	2.74	2.56
12	3.75	3.24
13	4.76	4.34
	identical	

has a special value mainly for a population of small multiplication rate. (Fig. 3 shows the delayed decrease of the number of completely wrong words, $R = 0$ for a special case.)

In the following chapters we shall go back to the "one word — one generation" assumption.

Influence of the length of the genetic information

Eigen has claimed [3] that in the case of $\mu L < 1$ the correct words will dominate the final population. This does not seem to be valid in our computer model. The reason for this is simple. Let $\mathbf{y}_1$ describe an ideal population, consisting exclusively of the

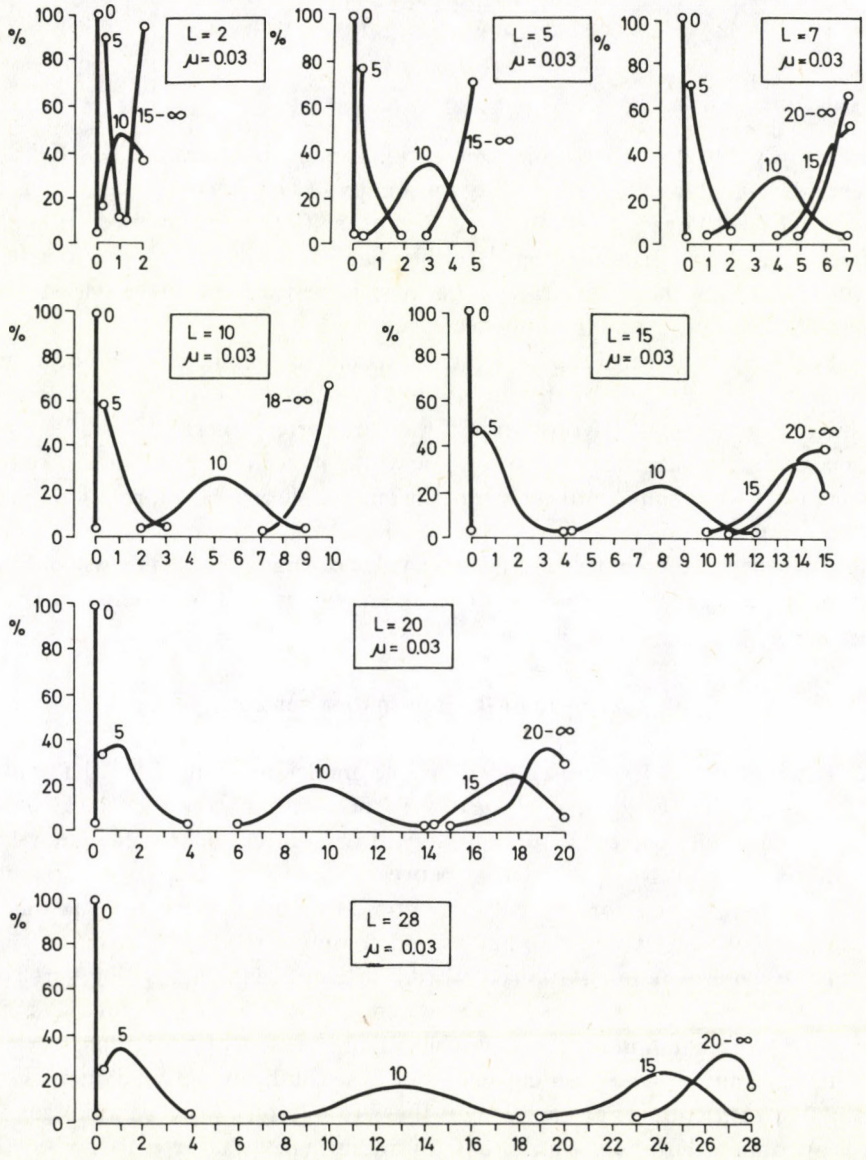


Fig. 4

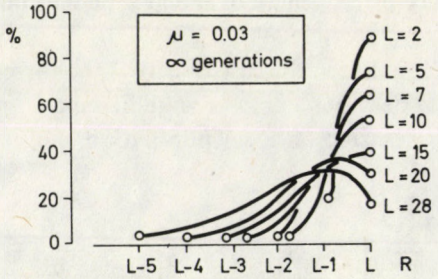


Fig. 5

correct words with $R = L$. Now a new generation cannot be better than $\mathbf{My}_1$. If the error per character is μ, then the probability for a copy to be incorrect is $\simeq \mu L$. E.g. for $\mu = 1/15$ and $L = 10$ one gets $\mu L = 2/3$, i.e. a good half of any generation will contain incorrect characters! It is true on the other hand that the new generation will be produced mainly by the perfect fraction, because the perfect words have at least twice as high multiplication rate as the imperfect ones.

Fig. 4 depicts the evolution of words with different lengths in time. The mutation rate per character is the same for all: $\mu = 3\%$. The speed of evolution increases with growing L. But for longer words, even if they are correct, there is a small chance to produce correct copies. Fig. 5 shows that the width of the final population increases by growing L. This certainly limits the size of the inheritable genetic information. It is only the "selection pressure", not inheritance, which prevents a complete randomization of the text. As Eigen has emphasized: for a fixed value of μ natural selection will optimalize the size of the genetic information. This phenomenon will be studied in the next Chapter.

Growth of the information content

Peter Schuster [5] presented a simple model to show how the mutation probability per character limits the size of the genetic information. Here he assumes for simplicity that only correct words multiply, any incorrect character is lethal. This assumption is useful to get a relation between μ and L_{max}. But to understand the increase of the genetic information to its optimum length one has to relax this strict condition. That is what we do in our second computer model.

Let us consider a population of "words" with different lengths. (E.g. we shall put $L = 2, 3, 4, 5$. Lone characters are not considered to be sensible words, but only building blocks available in unlimited quantity. The maximum value of L is limited by the long build-up time and by computer capacity. Let us see, if the evolution dynamics makes any further restriction.) For each length there is a "target word", the longer target words are obtained from the shorter ones by adding a new character to the end. (E.g. a series of target words may be AB, ABC, ABCD, ABCDE, or in a more English way: TO,

TOO, TOOL, TOOLS. An arbitrary word made of L characters may contain R characters identical with those of the target word with the same length. (There are R identical characters on the correct place). The number of such words will be denoted by $y(L, R)$ in the population. Evidently $0 \leqq R \leqq L$, as before. The number $y(L, N)$ will change in time.

A longer chain of characters needs more time to be copied, so the time gap between two generations is supposed to be proportional to the number L of characters in the word. — A word carrying a more sophisticated information may have a longer life time. ("More enzymes offer a more elaborate defence system against wearing out.") In our second computer model the average life time is assumed to be

$$T(L, R) = 3^R \tau .$$

If one has a population of "correct" words of different lengths ($R = L$) and no error copies are allowed, the time dependence of $y(L, L)$ will be described by the differential equation

$$\dot{y}(L, L) = \left(\frac{1}{L} - \frac{1}{3^L \tau} \right) y(L, L) .$$

The optimum value of L, offering the fastest multiplication, can be obtained from the algebraic equation

$$\frac{\mathrm{d}}{\mathrm{d}L} \left(\frac{1}{L} - \frac{1}{3^L \tau} \right) = 0 ,$$

i.e. from

$$L^2 3^{-L} \lg 3 = \tau . \tag{5}$$

So for any τ one can estimate the optimum length of words (see Fig. 6).

If the word contains also "wrong" characters, its multiplication rate is suppressed by a factor of 1/2 per wrong character. So an (L, R) word has the multiplication value

$$S(L, R) = \frac{1}{L} \frac{2^R}{\sum_{R'=0}^{L} 2^{R'}} . \tag{6}$$

These are the elements of the diagonal spreading matrix **S**.

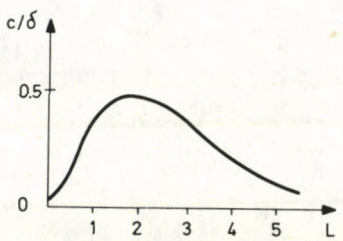


Fig. 6

Not all offspring are true copies. The mutation matrix (4) can be approximated for a small value of μ by the expression

$$M = 1 + \mu LP\,. \tag{7}$$

Now the population contains words of different lengths, so both **S** and **M** are composite matrices containing diagonal submatrices indicated by (6) and (7).

Let us allow that a random character may stick to the end of any word with a probability γ per unit time. (γ for "growth"). There is a 1/16 chance that the new added character is a "right" one (increasing L to $L+1$ and increasing R to $R+1$). There is a 15/16 chance that the new added character is a "wrong" one (increasing L to $L+1$ but leaving R unchanged). This effect can be described by a nondiagonal matrix G.

$$\mathbf{G} = \begin{bmatrix}
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\
\frac{15}{16} & 0 & 0 & -1 & 0 & 0 & 0 & 0 & \dots \\
\frac{1}{16} & \frac{15}{16} & 0 & 0 & -1 & 0 & 0 & 0 & \dots \\
0 & \frac{1}{16} & \frac{15}{16} & 0 & 0 & -1 & 0 & 0 & \dots \\
0 & 0 & \frac{1}{16} & 0 & 0 & 0 & -1 & \dots & \\
0 & 0 & \frac{1}{16} & 0 & 0 & 0 & -1 & \dots & \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & &
\end{bmatrix}. \tag{8}$$

It will be assumed in the new model that a word at the site of a wrong character is more vulnerable than elsewhere. There is a probability β for a break of the word at the wrong character. The wrong character can stick to the first or to the second fragment with equal probability. If both fragments consist of at least two characters, we have got two shorter words instead of the original long one. The $L-R-1$ other wrong characters of the original word distribute in average uniformly between the two fragments, so the first fragment (containing the first L_1 characters of the original word) will consist of

$$R_1 = L_1 - \frac{1}{2}\left[L_1 \frac{L-R-1}{L-1}\right] + \frac{1}{2}\left[(L_1-1)\frac{L-R-1}{L-1} + 1\right]$$

right characters. The right characters of the original word will become wrong certainly in the second fragment, but some of the wrong characters of the original word may turn out to be right, with a probability 1/16 per character. All these changes can be taken into account algebraically with the help of a nondiagonal matrix **B**.

The decay of the words can be described with a mean life time

$$T(L, R) = 3^R \tau ,$$

i.e. with the expression $\tau^{-1}\mathbf{D}$, where

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & 0 & =0 & 0 & 0 & \dots \\ 0 & 3^{-1} & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 3^{-2} & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 3^{-1} & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 3^{-2} & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & 3^{-3} & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} .$$

Finally, the differential equation describing the fate of the population reads

$$\mathbf{y} = [(1 + \mu L\mathbf{P})\mathbf{S} + \gamma\mathbf{G} + \beta\mathbf{B} - \tau^{-1}\mathbf{D}]\mathbf{y} .$$

This takes into account several aspects influencing the information content of the words. To see a specific example, the numerical parameters were chosen in the following specific way:

$$\mu = 0.10 , \qquad \gamma = 0.02 , \qquad \beta = 0.02 , \qquad \tau = 0.4 .$$

The computation started with a single nonsense word of two characters ($L = 2$, $R = 0$, like FZ). The emergence of longer and sensible words is shown in Fig. 7. This Figure gives a motion picture, how the two character word improved (e.g. from FZ to TO), then one experiences a very slow transition to three characters and improving of the meaning of the three characters (e.g. making TOO). In the final steady population about 3/4 of the population is made of a sensible four-character word (e.g. TOOL), coexisting with a few sensible three-character expressions (like TOO) and with a few sensible five-character expressions (like TOOLS). Copies with one character error are also present. The evolution stops at this stage. There would be no advantage to try to build up lengthy words (like TOOLSMITH), because they need too much time for formation and too high a chance to pick up misprints.

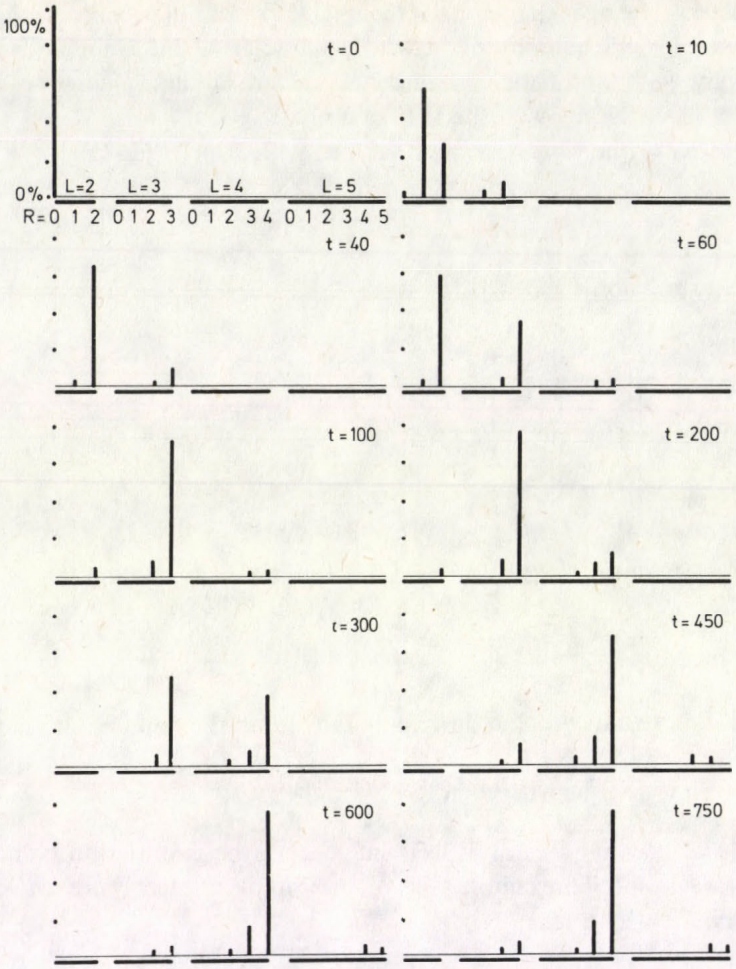


Fig. 7

Evidently the above set of characters has been used only as illustration. The computer program works with other parameter values as well, it can be used as a tool to model more complex situations concerning the evolution of genetic information.

The authors are very much indebted to Prof. Vilmos Csányi and to Prof. Peter Schuster for inspiring and critical discussions.

References

1. See e.g. K. G. Dengibh: An Inventive Universe. Hutchinson, London, 1975.
Manfred Eigen-Ruthild Winkler: Das Spiel. Piper, München, 1975.
2. See e.g. Hubert P. Yockey, Journal of Theoretical Biology, **91,** 13., 1981. M. Coates and S. Stone, Journal of Molecular Evolution, **17,** 311, 1981. U. Niesert, D. Harnasch and C. Bresh, Journal of Molecular Evolution, **17,** 348, 1981.
3. Manfred Eigen, Berichte der Bunsen-Gesellschaft für physikalische Chemie 1059, 1976. Manfred Eigen, Peter Schuster, Die Naturwissenschaften, **64,** 541, 1977.
4. G. Marx, Biológia (Budapest) **28,** 177, 1980. (in Hungarian)
5. Peter Schuster: Prebiotic Evolution, in Biochemical Evolution, Cambridge University Press (1982), in print. See Figure 17.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 279—289 (1982)

MYSTERIES OF NOISE*

L. PÁL

HUNGARIAN ACADEMY OF SCIENCES, BUDAPEST, HUNGARY

The importance of random fluctuations is shown in obtaining significant information about the dynamics of large systems. Several examples are discussed including the branching processes.

Introduction

It is difficult to find any natural phenomena whose numerical parameters show absolutely no random fluctuations. These random fluctuations contain valuable information about the dynamic behaviour of the phenomena and in many instances their study may lead to the discovery of qualitatively new relations.

In scientific practice and in engineering, such fluctuations are known as 'noise'. The everyday concept of noise refers to unpleasant and undesirable sounds that tend to interfere with speech thereby making it difficult to perceive what is being said as well as making people nervous. The acoustic notion of noise has been generalized by physicists and engineers to relate to thermal, electromagnetic and other noise. Noise reduces the useful information or may even make it impossible for information to be received. To characterize the noise level in engineering practice, a ratio is used to relate noise (the interfering effect) to the signal (the useful information). This particular meaning of noise (the signal-to-noise ratio) can be generalized for all phenomena showing random fluctuations.

The investigation of noise, i.e. noise analysis, has lately come to have a decisive role in nearly every field. It is known that the basic issue in telecommunications is how to convey information in the presence of noise in the most effective way. The strong reduction of the 'noise-to-signal' ratio is often too expensive. The facts show that in many instances noise analysis is cheaper; moreover, in certain cases it offers a unique method of gaining information diluted in the 'sea of noise'. This procedure has proven to be almost indispensable in receiving information arriving from the transmitters of space satellites and artificial celestial bodies.

In telecommunications, as has just been indicated, the purpose of noise analysis is to filter out the useful information. In many cases, however, noise analysis is used as a

* This modest contribution, drawing attention to several factors playing an important role in the world of random events, was motivated by an old conversation with my friend Prof. I. Tarján and is dedicated to his 70th birthday.

means of gaining information on the fine details of the noise-producing phenomena themselves. For instance, from the fluctuations of the brain currents or, to be more exact, from analysis of the fluctuations, conclusions can be drawn on some kinds of brain function defects. The fluctuations in the brain currents of an epileptic surprisingly resemble the power fluctuations of a boiling water reactor in an unstable state. As a matter of fact, fluctuations indicate with a great sensitivity the instabilities of the investigated process, therefore their analysis results in useful information and may even be of importance so far as safety is concerned.

In the foregoing — in accordance with convention — the disturbing random fluctuations were called noise. If, however, our task is not to get rid of the 'noise' but simply to reveal the nature of the 'noise', possibly it is more correct to use the word "fluctuations" — as a synonym of noise. In such cases it is clear that the greater the 'noise' or the greater the fluctuations, the easier the task to be solved. It may happen, however, that a system does not exhibit great fluctuations and delicate experimental techniques are needed to discover the nature of the fluctuations.

Perhaps it is also worth mentioning that certain changes in condensed systems are related to strong fluctuations of the physical parameters (density, magnetization, electric polarization, etc.). These fluctuations indicate that there is a transition point at which the substance is in a strongly unstable (critical) state. Analysis of fluctuations enables the reasons for instabilities to be revealed which may thereby help to establish a more accurate foundation of the theory of phase transition.

Some general remarks

Fluctuations are natural concomitants of random phenomena. The randomness of a given phenomenon asserts itself in the fact that the outcome of one and the same phenomenon may vary though it may have been subjected to seemingly the same conditions. The reason for this is that every phenomenon depends on an infinite number of other phenomena and so the fixing of a finite number of conditions (and in practice only the latter can always be realized) does not usually make the behaviour of the phenomena rigorously unambiguous.

As for the fixing of the conditions, this is needed simply so that we can select the phenomenon to be investigated from among the infinity of phenomena. The fixing of the conditions should be carried out in a way that the effects of the essentially important factors influencing the investigated phenomenon are separable from the effects of the inessential factors. The actual behaviour of the phenomenon is substantially influenced by the relationship between the essential and inessential factors developing after the fixing of the conditions. The greater the effect of the inessential factors the greater the randomness that can be observed.

To select the measure of chance we have to turn to what we see in practice. When repeating something several times under the same conditions we can observe that

certain selected events sometimes occur, sometimes not. To characterize the probability of the event in question, a numerical value called frequency can be used, i.e. the ratio of the number of occurrences to the number of repetitions.

By increasing the number of repetitions and determining the frequency of the given event after each repetition, we find that the value of the frequency will become stable around a certain value with fluctuations of decreasing amplitude. The stabilization of the frequency of the events on increasing the number of repetitions is a fundamental law of random events. The value around which the frequency of the investigated event becomes stable characterizes the event itself and shows how probable the occurrence of the event will be. That is precisely why we choose this numerical value as the measure of the probability of a phenomenon.

To illustrate stabilization, there is an interesting computerized 'experiment'. The computer simulates the process of coin tossing by considering the individual tossings as independent from each other and by giving each tossing the value of $+1$ or -1 depending on whether the result of tossing is 'heads' or 'tails'. Let η_k be the numerical value for the k-th tossing and let us designate the following sum by μ_n:

$$\mu_n = \frac{\eta_1 + \eta_2 + \ldots + \eta_n}{n}.$$

It is instinctively expected that the value of μ_n will fluctuate around zero since the frequency of 'heads' will hardly differ from that of 'tails'.

If n is large then $\mu_n \sim \langle \eta \rangle = 0$. For instance, if $n = 50{,}000$ then $-\frac{1}{50} < \mu_{50{,}000} < \frac{1}{50}$, or when $n = 200{,}000$ then $-\frac{1}{100} < \mu_{200{,}000} < \frac{1}{100}$. We can say that the value of μ_n becomes stable around zero and this refers to the probability of both the 'heads' and the 'tails' tossing being the same value, namely it is $\frac{1}{2}$. This behaviour illustrates the so-called law of large numbers.

It is also important to mention that the determination of the probability of elementary events is not a mathematical task. Only the actual investigation of the phenomenon can result in the determination of the probability of elementary events. The probability of a complex event composed of elementary events, however, can be calculated by the methods of probability theory.

Behaviour of macro-systems

After this temporary digression aiming at the understanding of the essence and objective nature of chance and the inherent manifestation of the relation of cause and effect, let us now turn to the fluctuations of the characteristic physical parameters of the 'macro-world'.

The formation of a material system is called macroscopic if its dimensions are multiples of atomic sizes, i.e. if it is built up of a large number of atoms and molecules. The set of macroscopic systems constitutes the 'macro-world'. This designation and particularly its content cannot, of course, be free from the influence of anthropocentric concept-building, since man cannot separate himself from the frame of reference relating to his own sizes and time cycles when willy-nilly creating the 'macroscopic' attributive adjective "in his own likeness and image". I should like to emphasize this now because later on we will deal with the complexity of macroscopic systems and with the problem that the multiplicity of the properties not even interpretable at atomic level is due to the limitless interplay of some 10^{25} interacting atoms in these systems. By this I should not like to support the belief that macroscopic systems are more complex and more inexhaustible in the variety of their properties than the atoms from which they are constructed or than the subatomic particles which are still far smaller in size and persist for a shorter or longer time. No, this is not true; any formation of material can be extremely complicated — we only need think of the newest results in particle physics — and this means that in the dialectics of simple and complex what is simple can become complex and what is complex can become simple depending on the circumstances. The collision of two protons of extremely high energy is at least as perplexingly complex a process as crystallization or the coding of genetic information, but at the same time the collision of two protons of low energy ("low" in terms of the binding energy per nucleon) scarcely differs from the collision of two electrically charged balls. Complexity is always the result of the most different interdependences stemming from interactions and — as we shall see — this is the fundamental source of the probability laws resulting in the various manifestations of the phenomena.

In what follows, we return to the behaviour of macroscopic systems and the fluctuations of the processes arising in these systems. The laws governing the behaviour of the atoms building up macroscopic systems can reliably be derived from quantum theory. The characteristics of electromagnetic forces representing the interaction between atoms is well known. Fundamentally, it is these forces that determine the properties in the everyday meaning (such as phase state of matter, crystal structure, mechanical, electrical, magnetic behaviour, chemical bond, etc.). Gravitational interaction is important in systems of cosmic size, whereas subatomic interactions control the behaviour of nuclei and particles. From this one might think, in princple, that any property of macroscopic systems can be derived from laws determining the world of atoms. This is not so, however. First of all it is not so, not because it is practically impossible to handle the equations describing the state of some 10^{25} interacting atoms and to solve them — even by a possible future computer of fantastic capacity and speed —, but because such a complex of interacting atoms brings forth the appearance of such new — moreover, qualitatively new — properties which cannot even be interpreted for individual atoms nor for systems consisting of a few atoms.

Let me show you a simple example illustrating what I mean. Imagine a closed vessel of volume V_0 with a single molecule moving about inside. This molecule

occasionally collides with the wall of the vessel and transfers a certain momentum to the wall of the vessel depending on the conditions. If there are several molecules in the vessel the momentum transmitted to the wall of the vessel per unit time will increase, but one can hardly speak of pressure since the momentum transmitted to unit surface in unit time fluctuates considerably as a function of time. If, however, the number of molecules per cubic cm reaches the value of say, 10^{18}, then the fluctuations strongly decrease, the momentum transmitted to unit surface per unit time becomes stable around a well-defined constant value and, by this stage, the macroscopic behaviour of the gas enclosed in the vessel will already be characterized by a new quantity — the pressure — which, otherwise, has no meaning in the case of a few molecules.

In a similar way, one can speak of the density of gas only if the number of gas molecules per unit volume fluctuates to a small extent only around a well-defined value. Let V_0 again be the volume of a vessel with walls causing the gas molecules to rebound elastically and let ΔV be its subvolume ($\Delta V \ll V_0$). With only a single molecule moving in the vessel, although we would very rarely have a 'state' when the gas molecule just happens to be in volume element ΔV, it would still be impossible to assign a 'number density of molecules' to the volume element in a way that this parameter could be considered more or less constant. What is more, even in the case of many hundreds of molecules, one could have long periods of time with not a single molecule in volume ΔV. If, however, there is a very large number of molecules $\left(\text{e.g. } 10^{22}\ \frac{\text{molecule}}{\text{cm}^3}\right)$ in the vessel then the number of molecules in volume ΔV deviates only slightly, from time to time, from the numerical value that can uniquely be determined under the given conditions: that is, from the expected value of the number of molecules in volume ΔV.

One might think it would suffice to give the coordinates and velocity components of the molecules in volume V_0 at a given instant t_0 and then, in the knowledge of the forces between the molecules and that of the interaction with the wall of the vessel, one would be able to calculate the number of molecules in the selected subvolume ΔV for any arbitrary instant $t \geqq t_0$. Nevertheless, it would make no sense to follow this line since in the case of many molecules the expected value of the number of molecules in volume ΔV is hardly influenced by the coordinates and velocity values of the molecules at a previous instant. Namely, it is not the individual initial values that determine the macroscopic properties but the distribution of the initial values that play the essential role. This means to say that the macroscopic quantities are randomly connected with the characteristic quantities of microprocesses.

The density, pressure and temperature of the gases represent a qualitatively new resultant of the superposition of a great many microprocesses. On the basis of what has been said, it seems to be natural that the macroscopic properties can be derived by means of the same probability laws applied to microprocesses. It is needless to say that not only do the statistical laws macroscopic systems make easier the handling and solution of 10^{25} equations but they enable one to reveal that a significant increase in the number of interacting atoms and molecules results in the appearance of qualitatively

new, macroscopic features that behave as random variables and, in general, exhibit small or large but nonetheless definite fluctuations. It is also obvious that the fluctuations of macroscopic quantities should contain valuable information on the micro-behaviour of the system. In a simpler way, we can say that the smaller or larger, but continuous fluctuations of the physical quantities are the natural consequences of 'atomism' in the sense that the macroscopic properties are built up as a result of interactions at atomic level.

It is worth mentioning that it was first Clausius and Maxwell and especially Boltzmann who understood this to its fullest extent. They also realized that different physical quantities (density, pressure, temperature, magnetization, etc.), which are actually expected values of random variables, can play a decisive role in the description of physical processes only if the relative standard deviations of the random variables in question are very small. Any changes in the strongly fluctuating quantities should be handled with great care since the regular changes can be masked by spontaneous fluctuations. However, exploration of the nature of spontaneous fluctuations is important not only because it enables us to follow the behaviour of regular changes but also because it enables us to draw conclusions on the dynamic properties of the phenomenon investigated.

Discussion

Turning back to what was said regarding the density of gases, to obtain a clearer concept let us denote by $\xi(t)$ the number of molecules in volume ΔV at time t. With certain assumptions we can calculate the probability of $\xi(t)$ being equal to a given whole number, let us say to 'n'. Let this probability be

$$\text{Prob}\,\{\xi(t)=n\}=p_n(t)\,.$$

The number of molecules in volume ΔV continually varies, fluctuates, since this subvolume is not isolated from its surroundings. Let the volume $V=V_0-\Delta V$ enclosing the subvolume ΔV be sufficiently large so that the processes in ΔV are not able to modify the state of V. In this case the frequency that a molecule can get from V to ΔV can be considered to be constant: let us denote this frequency by λ. Obviously, the more molecules there are in ΔV the more frequently one of them escapes from ΔV to V. If there are n molecules in volume ΔV at a given moment then the frequency with which a molecule escapes to V is expected to be given as some function $\varphi(n)$ of n. In the simplest case this is $\varphi(n)=n\mu$. In the stationary state the expected value of the number of molecules in ΔV can be expressed by the frequencies λ and μ. We find that

$$n_\infty=\lim_{t\to\infty} E\{\xi(t)\}=\frac{\lambda}{\mu},$$

and the density is

$$\rho_\infty=\frac{\lambda}{\mu\Delta V}.$$

Let us suppose that at an arbitrary instant — let this be the zero point of our time scale — we suddenly and drastically alter the number of molecules in volume ΔV. The obtained number of molecules n_0 differs significantly from the $\frac{\lambda}{\mu}$ value corresponding to the stationary value. After a certain time, the deviation substantially decreases and volume ΔV will again contain an average number of molecules $\frac{\lambda}{\mu}$. The expected value of the number of molecules — according to our calculations — depends on time in the following way:

$$n(t) = n_0 e^{-\mu t} + \frac{\lambda}{\mu}(1 - e^{-\mu t}).$$

It can be seen that whatever definition is created at time $t = 0$, the number of particles will always approach the stationary value $\frac{\lambda}{\mu}$ with a 'velocity' determined by parameter μ. Parameter μ stands for the velocity of 'forgetting' the initial state (the past), its reciprocal is the so-called relaxation time whose magnitude depends on the intensity of interaction between the molecules.

A quite remarkable property of the dynamic behaviour of our system has come to light, namely, that its time-dependent processes are unidirectional, they always tend towards the most probable stationary state. To be more precise, however, we ought to say that it is the expected value of the number of particles that shows such unidirectional, irreversible changes, whilst the number of particles itself fluctuates and, among its deviations, fluctuations acting contrary to the approach of the stationary state also occur. This behaviour cannot follow from the laws of classical mechanics and electrodynamics, it completely departs from them and this was — especially at the beginning of the century — a question which occupied many renowned physicists and mathematicians.

We can quite generally state that in closed systems containing many particles, the processes being enacted are always in favour of the most probable state.

The most probable state is at the same time the one exhibiting the greatest disorder. A closed system in the most probable state is regarded as a system in a state of equilibrium whose characteristic macroscopic parameters (density, pressure, temperature, magnetization, etc.) exhibit only small fluctuations and their expected values do not depend on time. The state of equilibrium is independent of the initial state of a closed system. In other words, a system in equilibrium completely forgets its 'past'. A direct consequence of this is that, for any macroparameter, in equilibrium the expected value of the product of two values at two different moments is dependent only on the absolute value of the time difference of the two moments. If $\xi(t)$ denotes the number of molecules in volume ΔV at time t and $\xi(t+\tau)$ at time $(t+\tau)$, then in equilibrium the function

$$E\{\xi(t)\xi(t+\tau)\} - E\{\xi(t)\}E(\xi(t+\tau)\} = K(\tau)$$

can only depend on τ. The dynamic behaviour of the random variable $\xi(t)$ is characterized by the correlation function $K(\tau)$.

What is especially worth considering is that by means of the correlation function one can gain information on the dynamic parameters of the system even from small fluctuations in equilibrium. Therefore in many cases there is no need to disturb the equilibrium by external intervention if we want to determine the nature of the dynamic processes leading to equilibrium. The spontaneous fluctuations — if they are indeed observable — reveal everything. According to the model of the density fluctuations treated earlier, for the correlation function the formula

$$K(\tau) = \frac{\dot{\lambda}}{\mu} e^{-\mu t}$$

is obtained from which it is seen that for $\tau = 0$

$$K(\tau = 0) = \frac{\dot{\lambda}}{\mu} = D^2\{\xi(0)\},$$

and that the extinction of a particular density fluctuation takes place at the same rate as that of the transition to equilibrium from a non-equilibrium state.

When dealing with equilibrium, I would not like to miss the opportunity of mentioning how important the observation of spontaneous fluctuations is when investigating chemical reactions. Consider the simplest reaction in which molecule A changes to molecule B and, conversely, molecule B reverts to molecule A, that is,

$$A \underset{k_{21}}{\overset{k_{12}}{\rightleftarrows}} B.$$

In equilibrium, the same average number of molecules A are converted to molecules B in unit time as molecules B are converted to molecules A. This is expressed by the mass action law, according to which

$$\frac{\langle n_A \rangle}{\langle n_B \rangle} = \frac{k_{21}}{k_{12}} = K_0,$$

where

$$\langle n_A \rangle = E\{\xi_A\} \qquad \text{and} \qquad \langle n_B \rangle = E\{\xi_B\}.$$

In reality, the actual numbers of molecules A and B, even in equilibrium, continually deviate from the values $\langle n_A \rangle$ and $\langle n_B \rangle$; however, this deviation is negligible if the number of molecules is large.

Let us suppose that molecules A are capable of carrying electric charge, whereas molecules B are not. It his case in the system comprising molecules A and B the electric charge is transported by molecules A and the more molecules A there are in the system at a given voltage, the higher the current flowing through the system. It is also clear that small fluctuations in the number of molecules A around the equilibrium value give rise

to current fluctuations, thus, by measuring these current fluctuations the kinetic coefficients k_{12}, k_{21} of the reaction equation can — at least in principle — be determined. Actually we have to calculate the correlation function

$$\langle n_A(t+\tau)n_A(t)\rangle - \langle n_A(t+\tau)\rangle\langle n_A(t)\rangle = K(\tau)\,.$$

With simple considerations it can be shown that

$$K(\tau) = K(0)e^{-(k_{12}+k_{21})\tau}\,,$$

where

$$K(0) = \langle(\delta n_A)^2\rangle = D^2\{\xi_A\}\,.$$

It is the current fluctuations that we can measure but it is obvious that the correlation function $K_i(\tau)$ of the current density

$$i = n_A q v_A$$

differs only by a constant factor from the correlation function $K(\tau)$ of the number of molecules A, since

$$K_i(\tau) = q^2 v_A^2 \langle(\delta n_A)^2\rangle e^{-(k_{12}+k_{21})\tau}\,.$$

In this way the sum $k_{12}+k_{21}$ can be determined from the function $K_i(\tau)$. If we combine this with the value of the equilibrium constant $K_0 = k_{21}/k_{12}$, we can calculate the factors k_{12} and k_{21} separately, too.

In practice, however, frequently it is not the correlation function that is determined but the function $\Phi(\omega)$ describing the spectral density of the fluctuations. Nevertheless, $\Phi(\omega)$ is in a unique relation with $K(\tau)$, since

$$\Phi(\omega) = \int_{-\infty}^{+\infty} K(\tau)e^{-i\omega\tau}\,\mathrm{d}\tau\,.$$

Thus, in our case

$$\Phi(\omega) = q^2 v_A^2 \langle(\delta n_A)^2\rangle \frac{2(k_{12}+k_{21})}{(k_{12}+k_{21})^2+\omega^2}$$

from which $k_{12}+k_{21}$ can also be determined.

There is often the need to determine the kinetic parameters of chemical reactions in the equilibrium state especially if the non-equilibrium state cannot be maintained because of the sensitivity of the molecules. This situation is frequently encountered in biomolecular processes.

The importance of investigating fluctuations around the equilibrium state can be illustrated by many examples. Above, we have shown how the kinetic coefficients controlling non-equilibrium processes can be determined from the fluctuations. Now we quote a well-known example in which the fluctuations give away the discrete character of quantities at the atomic level.

In consequence of electric charge carriers transporting discrete quantities of charge only, fluctuations are expected to occur in current density. From these fluctuations, the magnitude of the elementary charge can be estimated. For instance, vacuum diode can be used for practical measurements: the electrons are emitted at random from the cathode and are caused to travel towards the anode by the electric field between the anode and cathode. It can be shown by a simple calculation — neglecting the effect of the field charge and the initial velocity of the electrons — that the ratio of the variance of current density to the expected value of the current can be given by the expression

$$\langle(\delta i)^2\rangle/\langle i\rangle = 4q/3\tau\,,$$

where τ is the time of flight of the electrons and q is the electric charge. τ can be determined by a separate measurement, thus the value of q can be estimated from the fluctuations in current. Of course this estimation is somewhat inaccurate but the essence here is not in the accuracy but in the fact that from the smaller or larger fluctuations of a macroscopic quantity one can conclude on the atomic characteristics.

Finally, I should like to make brief reference to the nature of the fluctuations exhibited by a particular process — known in the literature as the branching process. Its essence is that in a given system any unit either becomes extinct or produces new units at random moments. The number of new units produced in such an event can be a random variable. Let W_k be the probability of the multiplication event when just k new units are produced. Moreover, let Q_b and Q_d be the frequency of birth and death events, respectively. With relatively simple considerations, it is possible to determine the function $p_n(t)$ that gives the probability of having exactly n units at time $t>0$ in the system, given that at $t=0$ there was only one present. We are now first concerned with the expected value of the number of units at some time $t>0$. We obtain that this expected value is of the form

$$\langle n\rangle = e^{-1(1-\nu)Qt}\,,$$

where $Q=Q_b+Q_d$ and

$$\nu = \nu_1 \frac{Q_b}{Q}\,.$$

In this expression ν_1 is the expected value of the number of new units originating from one multiplication event, that is,

$$\nu_1 = \sum_{k=1}^{\infty} kW_k\,.$$

It can be seen that if $\nu<1$ then the number of units eponentially decreases and if $\nu>1$ then it exponentially increases. Consequently, the critical value is $\nu=1$ for which the expected value of the number of units remains constant. Of course, the actual number

of units is not constant: it fluctuates around the expected value. The nature of these fluctuations is quite specific. The variance of the number of units is

$$\langle(\delta n)^2\rangle = \left(1 + \frac{Q_b}{Q}\frac{\nu_2}{1-\nu}\right) e^{-(1-\nu)Qt}\left[1 - e^{-(1-\nu)Qt}\right],$$

where ν_2 is a quantity characterizing the standard deviation of the new units originating during one multiplication event, that is,

$$\nu_2 = \sum_{k=1}^{\infty} k(k-1)W_k .$$

The peculiarity in this is that in the critical state — when the expected value of the number of particles is constant — the variance gradually becomes greater, since in the case $\nu = 1$

$$\langle(\delta n)^2\rangle = \frac{\nu_2}{\nu_1} Qt .$$

This shows that the critical state itself is unstable since larger and larger fluctuations arise around the expected value. The large fluctuations can give rise to violation of the condition $\nu = 1$; for instance the influence of the fluctuations may modify the initial characteristics of the multiplicative system and thereby cause either a sudden increase or a decrease leading to an extinction in the number of units.

Final remark

In conclusion, it is mentioned that it was not intended to offer new results nor to provide a detailed account of the fluctuation problem. The aim here has been to help the reader to view the disturbing or non-disturbing, easily observable or masked fluctuations of various phenomena and processes with slightly different eyes in the future. Regard them as coded messages from which we can conclude on the deep-seated laws of nature.

Acta Physica Academiae Scientiarum Hungaricae, 53 (1—2), pp. 291—298 (1982)

GAS CARRIER AND DEGASSED MAGMATIC ROCKS AND THE PRIMORDIAL ATMOSPHERE*

A. SZALAY

INSTITUTE OF NUCLEAR RESEARCH OF THE HUNGARIAN ACADEMY OF SCIENCES
DEBRECEN, HUNGARY

The gas contents of granites and other continental plutonic rock samples were investigated and compared with a number of mantle rock samples and particularly with samples of the ophiolite sequence of the Troodos massif (Cyprus). The total volatile loss at 1000 °C was investigated. The gas composition was analysed by means of a quadrupole mass spectrometer after releasing the gases by crushing the samples in high vacuum at room temperature. In further experiments, release of CO_2 and volatile loss were determined as a function of temperature by means of derivatography. The high (about 9—15%) H_2O and gas content of many mantle samples is compared to the always lower (1—2%) volatile content of crustal samples which seem to have been at least partially degassed in the course of geological evolution of the Earth's crust. It is assumed that the primordial atmosphere originated in the degassing of the crust. Gas carrier magmatic rocks in the deeper layers of the Earth, in the mantle still contain the components of the primordial atmosphere. Some of them are accessible and were analysed in this work.

Introduction

The classification of igneous rocks in the title of this paper according to the gas content seems to be unusual compared with the conventional geological classification, according to which the silicate content, chemical composition, mineral structure, etc. are mainly considered in the highly complex magmatic differentiation. The science of igneous petrogenesis developed from the pioneering studies of Norman L. Bowen [1] considers these problems, many of which are still not properly understood.

The investigations reported here are concerned with the probable composition of the primordial atmosphere, a very interesting problem related to the origin of life. The rocks are regarded and investigated here but mainly as sources of the gases.

The title covers a hidden simple model theory which guided these experimental investigations. It is assumed that in the initial geological history of the Earth and the solar system, the aggregation of the Earth and terrestrial planets occurred from cold micrometeorites, chondrules, solid cosmic dust particles, etc. which contained adsorbed and confined gases brought from the cold cosmic space. At first, all material of the cold proto-Earth carried occluded gases and there was no hydrosphere or significant primordial atmosphere. After the aggregation , however, material on the

* Dedicated to Prof. I. Tarján on his 70th birthday.

surface of the proto-Earth became degassed and the liberated gases retained by gravitation built up the primordial atmosphere.

Of course it must be emphasized that for two centuries very many theories have been proposed for the origin of the solar system. Any model theory, necessary for guidance in planning experimental measurements, is not entirely new since it contains at least parts of the earlier theories.

Selection of rock samples

Igneous rock samples from different parts of the world were selected according to a global point of view, neglecting highly differentiated, fractionated samples of a composition very different from the average composition of the crust. The material of the continental crust, about 95% of whichis made up of granites, is best represented by granite samples. Some very old granites were included as well, assuming that they suffered perhaps less differentiation and so are probably less degassed. Some other igneous continental rocks were investigated as well. Basalts originate from and are derived by partial melting of the upper mantle. Only a limited number of ophiolites have been obtained. The Troodos (Cyprus) ophiolite sequence was very particularly investigated, because it represents probably the least disturbed and best investigated mantle material, together with the overlying sequence of crustal rocks.

In the course of geological evolution the Earth's crust suffered very many distortions, (subductions, orogenesis, metamorphosis, etc.). Its thickness under the continents is varied (tens of kilometres), however, the crust under the ocean bottom, amounting to about 70% of the total surface of the Earth, is about 6—7 km thick and more uniform.

The investigation of ophiolites shows that they are ancient parts, slabs of the oceanic crust and upper mantle uplifted to the surface by tectonic movements thus they are easily accessible for observations and sampling for laboratory investigations. The ophiolite sequence of the Troodos massif of Cyprus is one of the least deformed ophiolites and also the one most extensively studied and understood [2]. The lowest layer of oceanic crustal rock, gabbro is well exposed there and so are the uppermost layers of the mantle, the serpentinized ultramafic rocks of dunite and harzburgite. The Mohorovicic discontinuity [6] is the separating boundary layer between the deepest crustal layer of gabbro (or gabbroids) and the uppermost layer of the mantle dunite, or harzburgite, if the sequence is ideally undisturbed.

The investigations briefly reported here about the gas content of these rocks demonstrate that there is a discontinuity in it too, the gas content of crustal rocks being about one order of magnitude lower than that of the mantle rocks on the underlying side of the Moho. Besides the Troodos sequence, a kimberlite and a number of continental granites and some other igneous rocks (e.g. some ultramafic rocks from Hungary) were studied. A more detailed report will be published later [10].

Experimental results

The total volatile content was determined simply by powdering the sample in an agate mill to a grain size of less than 200 μ and after drying at 100 °C, heating for 2 hours at 1000 °C. (The derivatographic investigation demonstrates that 1000 °C is enough for the rapid release of the volatiles from a powdered sample). The weight loss is expressed in per cent and includes practically all the water and real gas volatiles content.

Fig. 1 demonstrates the distribution of the rock samples regarding total volatile loss. Each square represents a sample.

The investigated number of granites and other plutonic rocks from several continents, including some of the oldest granites of the Earth, have a low volatile content of 1.0 weight % mean value, with min. 0.3% and max. about <3%. The few gabbros including the Troodos sample fall into this interval. However serpentinized peridotite rocks from Troodos have a high volatile content (between 8.6—15%) and so has a fresh (unweathered) serpentinized kimberlite from de Beers mine, Kimberley, a high volatile content of 13.2%. Ultramafic rocks from Hungary contain about similarly high volatiles. The volatile contents of a number of basalts range between these two extreme groups. Basalts situated now in the crust derived and rose actually from the upper mantle. We can conclude from Fig. 1 that the volatile content of the serpentinized peridotite rocks from the mantle investigated here is high and that of all investigated igneous rocks of the crust including the gabbro immediately above the Moho is low. Basalts are somewhat less degassed than granites.

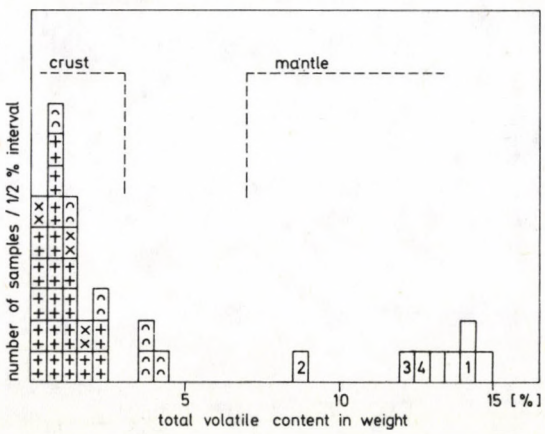


Fig. 1. Comparison of the total volatile loss in weight per cent between crustal rocks and ultramafic mantle rocks. Each square represents a sample. Signs: 1 Dunite (Troodos mts), 2 Harzburgite (Troodos mts), 3 Kimberlite fresh (Kimberley), 4 Serpentinized ultramafic (Lotus Colona, Calif. USA), ▯ Serpentinized ultramafics (Hungary), ▯ Granite, gneiss, amphibolite gneiss, granodiorite, ▯ Andesites, basalt, lavas, ▯ Gabbro

The abrupt change of the gas content from the upper mantle (dunite, harzburgite) to the crustal gabbro is very significant. A strong degassing occurs when hot mantle substance migrates across the Moho into the crust and differentiates by magmatic events.

The composition of the gas content was investigated by means of a quadrupole mass spectrometer (QMS), in a different type of experiment. The high sensitivity of the QMS enabled us to measure just a small sample of the gases released from the rocks without heating, by crushing them in a high vacuum mortar between hard tungsten-carbide steel dies. When crushing a 0.6 g sample of rock of about 0.6 mm grain size to small grains, enough volatiles are liberated to determine the gas content in the original composition as it was in thermochemical equilibrium in the rock without any chemical reaction between the gases by heating. The obtained composition justified the cold crushing, because much of H_2 and CH_4 strongly reducing gases have been liberated, together with H_2O, N_2 and CO_2. They would have been partly reacted at 1000 °C. The molecular composition between $M = 1$—50 was investigated. The whole analysis was carried out in an ultra high vacuum apparatus, at 10^{-6} mbar. The background of the equipment was about $1—3 \cdot 10^{-8}$ mbar.

The measurements were carried out under dynamic conditions. After liberating the gases by crushing, a needle valve of the mortar was opened and regulated so that a pressure of $1 \cdot 10^{-6}$ Torr was established in the QMS under continuous evacuation by means of a turbomolecular pump. Full description of the apparatus and measuring technique will be published elsewhere [3].

Fig. 2 shows the mass spectrum of the gases of dunite from Troodos.

The following can be said concerning the composition of gas content liberated by cold crushing: the QMS mass spectra demonstrate that the amount of water is overwhelming; free oxygen is never present; H_2, CH_4, CO_2, N_2 are rather abundant;

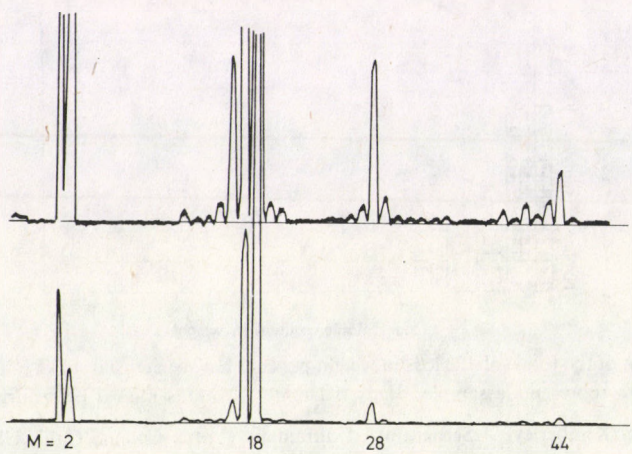


Fig. 2. QMS spectrum of gases in dunite (Troodos mts, Cyprus), released by cold crushing. Ordinate: Electrometer readings, with two (1:8) sensitivities; Abscissa: Molecular weight

argon and some traces of organic components can be observed within the investigated range of molecular weight M = 1—50. Evaluation after the necessary calibration of the QMS and corrections for molecular fragments by ionization result in the following approximate gas composition as the mean value ± standard deviation of 16 dunite samples, in volume units, related to $N_2 = 100$: $CO_2 = 80 \pm 10$; $CH_4 = 50 \pm 4$; $H_2 > 100$. (N.B.: the scattering between the samples is considerable because probably at least a part of the gases is contained in microbubbles within the mineral grains and so they are not evenly distributed.)

The gas is a reducing one and perhaps might be considered to approximate qualitatively the ancient reducing primordial atmosphere. Similar investigations were carried out by the author on granites, etc. earlier [7, 8]. The determination of the hydrogen content was more inaccurate in these measurements because the overhelming amount of H_2O molecules gave a great number of fragmented ions, in addition to $(H_2O)^+ : H^+, (OH)^+, O^+, H_2^+$. The hydrogen inclusion in the rocks is liberated by the crushing in the form of H_2 molecules which are ionized in the QMS to H_2^+ and the fragment ion H^+. The disturbing background of hydrogen fragments originating from water molecules was eliminated in another series of measurements. Water vapour was frozen out by liquid nitrogen trap so it appeared only in a small extent in the spectrum. Of course CO_2 was frozen out as well. The correction due to the water ion fragments became much smaller and less disturbing. This was very favourable for the determination of hydrogen content and for the determination of methane content as well. We obtained in these measurements the following values for H_2 and CH_4 in the dunite sample related to $N_2 = 100$:

$$CH_4 = 47 \pm 5,$$
$$H_2 \sim 100 - 500,$$

There is still a source of uncertainty in the value of hydrogen content because of the higher diffusing speed of H_2 molecules in a high vacuum. When the needle valve of the crushing mortar is opened the H_2 is diffusing much quicker into the QMS than the other gases. In the first seconds its abundance seems to be about 500 related to $N_2 = 100$. It is, however, very quickly extracted from the system by the turbomolecular pump and then after about 90 seconds its abundance is only about equal to N_2. A more accurate determination of H_2 would be possible only in a closed vacuum system in diffusion equilibrium. The QMS used in this experiment was not suitable for long measurements with a closed system since, because of background outgassing problems, the ultrahigh vacuum cleanliness was limited.

In addition the Troodos and other mantle samples were investigated by derivatography, developed by F. Paulik et al [4, 5]. About 300 mg powdered samples were heated gradually by 10 °C/min up to 1000 °C in a slow stream of N_2 (free of CO_2 and O_2). The change of weight and release of heat (change of enthalpy) were continuously and automatically recorded. The automatic quantitative analysis of liberated CO_2 as a function of temperature offered further significant information.

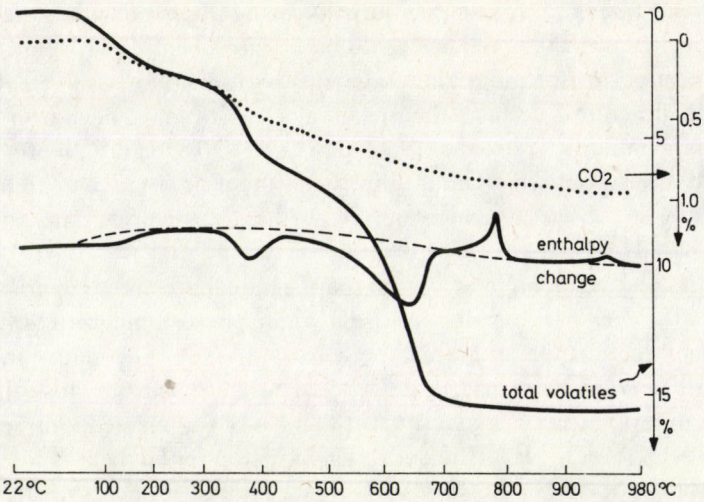


Fig. 3. Derivatographic investigation of dunite (Troodos mts, Cyprus). Ordinate: Loss of total volatiles in weight %, loss of CO_2 in weight %, and change of enthalpy; Abscissa: temperature, rate of increase 10 °C/min

This method gives information about the manner of fixation of H_2O, CO_2 and mineral transformations in the rocks, further about the total volatile loss as a function of temperature.

In Fig. 3 the results of derivatographic investigation of a Troodos dunite are demonstrated. It can be seen that the CO_2 was not chemically fixed as a carbonate. It is gradually released by increasing the temperature, no definite decomposition temperature is observable. The CO_2 content amounts to 0.94% of weight. However, at least a part of the water is fixed into the mineral structure (serpentines). Liberation of H_2O occurs at about 340 °C with disappearance of heat. This water belongs to $Mg(OH)_2$ which is transformed into MgO by liberating one molecule of water which amounts to about 2.2% weight. Another release of H_2O of about 7.2% of weight occurs at a temperature of about 660 °C again with significant heat disappearance. However, if we follow the change of enthalpy curve, we see a sudden liberation of heat at 780 °C. This is the well known transformation of the crystal structure of chrysotile into olivine. A similar transformation was observed on chrysotile asbestos (Troodos), mined on a large scale in the Troodos massif, and on the other peridotites as well. Kimberlite liberates CO_2 from 500—600 °C on, so this seems to be chemically fixed as carbonate, probably carbonatite.

About 95% of the continental Earth crust consists of granites. The significant number of granite and other plutonic rock samples investigated here are obtained from several continents, from different parts of the Earth's crust and have different ages. Their volatile content is always small varying between tight limits of 0.3% and 3%, the mean value of 1.0% approximates the average total volatile content of the continental

magmatic crust. Few gabbro samples have been investigated but they fall into this range. Only a single ophiolite sequence was investigated but the Troodos massif is the least disturbed among the known not numerous ophiolite occurrences. Concerning the mantle rocks, the number of samples of dunite and harzburgite, further chrysotile asbestos from Troodos and fresh kimberlite from Kimberley (South Africa) and some serpentinized ultramafic rocks from Hungary is small but it is generally assumed that the composition of the upper mantle is more uniform, better averaged than that of the very differentiated crustal rocks. Dunite and harzburgite represent the oceanic mantle from a depth of about 6 km to about some ten km. From the well known phase changes of diamond-graphite and quartz it is known that kimberlite must originate at a depth of between 100—300 km [9]. The geographical distance between Cyprus and Kimberley is about 7000 km. The few investigated mantle samples originate from depths between 6 and >100 km, and a distance of 7000 km and have a similar total volatile content of the same magnitude.

Although there are some rocks from the upper mantle which are degassed and some xenolithes (inclusions of some small peaces of mantle rocks into magma) carried quickly to the surface by molten magma are degassed, it seems that the undisturbed mantle can be regarded as a gas carrier. Explorations of the sub-oceanic mature crust and mantle by the originally planned bore-hole (Mohole) project or a much easier exploratory bore-hole into the uplifted Troodos complex will justify this supposition. It can be assumed that the present crust was in the earlier period of the evolution of the Earth a gas carrier like the mantle still is today. Its degassing produced the present hydrosphere and atmosphere.

The composition of the primordial atmosphere was comparable to the gas content of the gas carrier mantle samples. It did not contain any free O_2 at all and it was strongly reducing because of its abundant H_2 and CH_4 content. The H_2 content might have gradually escaped into space because gravitation could not retain it. This composition of the primordial atmosphere was suitable for the evolution of organic molecules and finally life by means of the free energy of the ultraviolet radiation of the sun.

Acknowledgements

The author expresses his best thanks for collaboration to Colleagues of this Institute and for valuable discussion to geologists, further to the donators of some of the samples. The author is particularly grateful to Dr. J. Paulik and Dr. F. Paulik of the Institute for General and Analytical Chemistry, Technical University of Budapest, who carried out the derivatographic determinations of Troodos mantle rocks and other peridotites by their home-developed highly precise automatic derivatographs.

References

1. Norman L. Bowen, in The Evolution of the Igneous Rocks, H. S. Yoder, Ed., Princeton Univ. Press, 1979.
2. I. G. Gass, Proc. Int. Ophiolite Symposium, Cyprus 1979, Geol. Survey Dept. Cyprus (1980).
3. A. Szalay et al., ATOMKI Közlemények (to be published).
4. F. Paulik, J. Paulik and L. Erdey, Talanta Review, **13,** 1405, 1966.
5. F. Paulik and J. Paulik, Journ. of Thermal Analysis, **5,** 253, 1973.
6. S. Müller, The Earth's Crust, pp. 289—317, American Geophys. Union, Washington, D. C. (1977).
7. A. Szalay, Magyar Tudomány **XXII,** 736 1977 (in Hungarian).
8. A. Szalay, Fizikai Szemle, **25,** 460, 1975 (in Hungarian).
9. K. G. Cox, Sci. Am., **238,** 120, 1978.
10. A. Szalay, Acta Phys. Hung., to be published.

Manuscript received by Akadémiai Kiadó: 8 June 1982
Manuscript received by the Printers: 14 June 1982
Date of publication: 6 June 1983

PRINTED IN HUNGARY

Akadémiai Nyomda, Budapest

NOTES TO CONTRIBUTORS

I. PAPERS will be considered for publication in Acta Physica Hungarica only if they have not previously been published or submitted for publication elsewhere. They may be written in English, French, German or Russian.

Papers should be submitted to

Prof. I. Kovács, Editor
Department of Atomic Physics, Technical University
1521 Budapest, Budafoki út 8, Hungary

Papers may be either articles with abstracts or short communications. Both should be as concise as possible, articles in general not exceeding 25 typed pages, short communications 8 typed pages.

II. MANUSCRIPTS

1. Papers should be submitted in three copies.

2. The text of papers must be of high stylistic standard, requiring minor corrections only.

3. Manuscripts should be typed in double spacing on good quality paper, with generous margins.

4. The name of the author(s) and of the institutes where the work was carried out should appear on the first page of the manuscript.

5. Particular care should be taken with mathematical expressions. The following should be clearly distinguished, e.g. by underlining in different colours: special founts (italics, script, bold type, Greek, Gothic, etc.); capital and small letters; subscripts and superscripts, e.g. x^2, x_3; small *l* and 1; zero and capital O; in expressions written by hand: *e* and *l*, *n* and *u*, *v* and *v*, etc.

6. References should be numbered serially and listed at the end of the paper in the following form: J. Ise and W. D. Fretter, Phys. Rev., **76**, 933, 1949.

For books, please give the initials and family name of the author(s), title, name of publisher, place and year of publication, e.g.: J. C. Slater, Quantum Theory of Atomic Structures, I. McGraw-Hill Book Company, Inc., New York, 1960.

References should be given in the text in the following forms: Heisenberg [5] or [5].

7. Captions to illustrations should be listed on a separate sheet, not inserted in the text.

8. In papers submitted to Acta Physica all measures should be expressed in SI units.

III. ILLUSTRATIONS AND TABLES

1. Each paper should be accompanied by three sets of illustrations, one of which must be ready for the blockmaker. The other sets attached to the copies of the manuscript may be rough drawings in pencil or photocopies.

2. Illustrations must not be inserted in the text.

3. All illustrations should be identified in blue pencil by the author's name, abbreviated title of the paper and figure number.

4. Tables should be typed on separate pages and have captions describing their content. Clear wording of column heads is advisable. Tables should be numbered in Roman numerals (I, II, III, etc.).

IV. RETURN OF MATERIAL

Owing to high postage costs, the Editorial Office cannot undertake to return *all* materia not accepted for any reason for publication. Of papers to be revised (for not being in conformity with the above Notes or other reasons) only *one* copy will be returned. Material rejected for lack of space or on account of the Referees' opinion will not be returned to authors outside Europe.

Periodicals of the Hungarian Academy of Sciences are obtainable at the following addresses:

AUSTRALIA
C.B.D. LIBRARY AND SUBSCRIPTION SERVICE
Box 4886, G.P.O., *Sydney N.S.W. 2001*
COSMOS BOOKSHOP, 145 Ackland Street
St. Kilda (Melbourne), Victoria 3182

AUSTRIA
GLOBUS, Höchstädtplatz 3, *1206 Wien XX*

BELGIUM
OFFICE INTERNATIONAL DE LIBRAIRIE
30 Avenue Marnix, *1050 Bruxelles*
LIBRAIRIE DU MONDE ENTIER
162 rue du Midi, *1000 Bruxelles*

BULGARIA
HEMUS, Bulvar Ruszki 6, *Sofia*

CANADA
PANNONIA BOOKS, P.O. Box 1017
Postal Station "B", *Toronto, Ontario M5T 2T8*

CHINA
CNPICOR, Periodical Department, P.O. Box 50
Peking

CZECHOSLOVAKIA
MAD'ARSKÁ KULTURA, Národní třida 22
115 66 Praha
PNS DOVOZ TISKU, Vinohradská 46, *Praha 2*
PNS DOVOZ TLAČE, *Bratislava 2*

DENMARK
EJNAR MUNKSGAARD, Norregade 6
1165 Copenhagen K

FEDERAL REPUBLIC OF GERMANY
KUNST UND WISSEN ERICH BIEBER
Postfach 46, *7000 Stuttgart 1*

FINLAND
AKATEEMINEN KIRJAKAUPPA, P.O. Box 128
SF-00101 Helsinki 10

FRANCE
DAWSON-FRANCE S. A., B. P. 40, *91121 Palaiseau*
EUROPÉRIODIQUES S. A., 31 Avenue de Versailles, *78170 La Celle St. Cloud*
OFFICE INTERNATIONAL DE DOCUMENTATION ET LIBRAIRIE, 48 rue Gay-Lussac
75240 Paris Cedex 05

GERMAN DEMOCRATIC REPUBLIC
HAUS DER UNGARISCHEN KULTUR
Karl Liebknecht-Straße 9, *DDR-102 Berlin*
DEUTSCHE POST ZEITUNGSVERTRIEBSAMT
Sraße der Pariser Kommüne 3–4, *DDR-104 Berlin*

GREAT BRITAIN
BLACKWELL'S PERIODICALS DIVISION
Hythe Bridge Street, *Oxford OX1 2ET*
BUMPUS, HALDANE AND MAXWELL LTD.
Cowper Works, *Olney, Bucks MK46 4BN*
COLLET'S HOLDINGS LTD., Denington Estate
Wellingborough, Northants NN8 2QT
WM. DAWSON AND SONS LTD., Cannon House
Folkstone, Kent CT19 5EE
H. K. LEWIS AND CO., 136 Gower Street
London WC1E 6BS

GREECE
KOSTARAKIS BROTHERS INTERNATIONAL BOOKSELLERS, 2 Hippokratous Street, *Athens-143*

HOLLAND
MEULENHOFF-BRUNA B.V., Beulingstraat 2,
Amsterdam
MARTINUS NIJHOFF B.V.
Lange Voorhout 9–11, *Den Haag*
SWETS SUBSCRIPTION SERVICE
347b Heereweg, *Lisse*

INDIA
ALLIED PUBLISHING PRIVATE LTD. 13/14
Asaf Ali Road, *New Delhi 110001*
150 B-6 Mount Road, *Madras 600002*
INTERNATIONAL BOOK HOUSE PVT LTD
Madame Cama Road, *Bombay 400039*
THE STATE TRADING CORPORATION OF INDIA LTD., Books Import Division, Chandralok
36 Janpath, *New Delhi 110001*

ITALY
INTERSCIENTIA, Via Mazzè 28, *10149 Torino*
LIBRERIA COMMISSIONARIA SANSONI, Via Lamarmora 45, *50121 Firenze*
SANTO VANASIA, Via M. Macchi 58
20124 Milano
D. E. A., Via Lima 28, *00198 Roma*

JAPAN
KINOKUNIYA BOOK-STORE CO. LTD.
17-7 Shinjuku 3 chome, Shinjuku-ku, *Tokyo 160-91*
MARUZEN COMPANY LTD., Book Department, P.O. Box 5050 Tokyo International, *Tokyo 100-31*
NAUKA LTD. IMPORT DEPARTMENT
2-30-19 Minami Ikebukuro, Toshima-ku, *Tokyo 171*

KOREA
CHULPANMUL, *Phenjan*

NORWAY
TANUM-TIDSKRIFT-SENTRALEN A.S., Karl Johansgatan 41–43, *1000 Oslo*

POLAND
WĘGIERSKI INSTYTUT KULTURY, Marszałkowska 80, *00-517 Warszawa*
CKP I W, ul. Towarowa 28, *00-958 Warszawa*

ROUMANIA
D. E. P., *Bucureşti*
ILEXIM, Calea Grivitei 64–66, *Bucureşti*

SOVIET UNION
SOJUZPECHAT – IMPORT, *Moscow*
and the post offices in each town
MEZHDUNARODNAYA KNIGA, *Moscow G-200*

SPAIN
DIAZ DE SANTOS, Lagasca 95, *Madrid 6*

SWEDEN
ALMQVIST AND WIKSELL, Gamla Brogatan 26
101 20 Stockholm
GUMPERTS UNIVERSITETSBOKHANDEL AB
Box 346, *401 25 Göteborg 1*

SWITZERLAND
KARGER LIBRI AG, Petersgraben 31, *4011 Basel*

USA
EBSCO SUBSCRIPTION SERVICES
P.O. Box 1943, *Birmingham, Alabama 35201*
F. W. FAXON COMPANY, INC.
15 Southwest Park, *Westwood Mass. 02090*
THE MOORE-COTTRELL SUBSCRIPTION AGENCIES, *North Cohocton, N. Y. 14868*
READ-MORE PUBLICATIONS, INC.
140 Cedar Street, *New York, N. Y. 10006*
STECHERT-MACMILLAN, INC.
7250 Westfield Avenue, *Pennsauken N. J. 08110*

YUGOSLAVIA
JUGOSLOVENSKA KNJIGA, Terazije 27, *Beograd*
FORUM, Vojvode Mišića 1, *21000 Novi Sad*

HU-ISSN 0001-6705 **Index: 26.022**

ACTA PHYSICA

ACADEMIAE SCIENTIARUM HUNGARICAE

EDITORIAL BOARD
R. GÁSPÁR, K. NAGY, L. PÁL, A. SZALAY, I. TARJÁN

EDITOR-IN-CHIEF
I. KOVÁCS

VOLUME 53 NUMBERS 3–4

AKADÉMIAI KIADÓ, BUDAPEST
1982

ACTA PHYS. HUNG. APAHAQ 53 (3–4) 299–468 (1982)

ACTA PHYSICA

A JOURNAL OF THE HUNGARIAN ACADEMY OF SCIENCES

EDITED BY

I. KOVÁCS

Acta Physica publishes original papers on subjects in physics. Papers are accepted in English, French, German and Russian.

Acta Physica is published in two yearly volumes (4 issues each) by

AKADÉMIAI KIADÓ

Publishing House of the Hungarian Academy of Sciences
H-1054 Budapest, Alkotmány u. 21

Subscription information

Orders should be addressed to

KULTURA Foreign Trading Company
1389 Budapest P.O. Box 149

or to its representatives abroad.

CONTENTS

I. Kirschner: Preface 303

CONDENSED MATTER

K. Fischer, S. Rohr, J. Mönch, W. Holzhäuser and Ch. Frenzel: V_3Ga Superconductors with a High Critical Current Density 305

I. Kirschner and K. Sajó: Determination of Microscopic Parameters by Macroscopic Measurements in Superconductors 313

A. Pevala: Application of the "Boson Method" to the Study of Superconducting Order Parameter 321

J. Hoellrich and G. Kath: Investigation of the Influence of the Heat Treatment and Forming Regimes on the Critical Current and Pinning Force of NbTi-Superconductors 331

S. Takács: Statistical Theory of Pinning in Superconductors Including Correlations in the Defect Distribution 337

I. Kirschner and K. Martinás: The Pinning of Vortex Lines in Very Dilute Superconductors 347

J. Bánkuti and É. Farkas: Experimental Determination of Pinning Forces 353

P. Debreceni, I. Kirschner, M. F. Kotkata, G. Kovács and T. Porjesz: On the Exact Calculation of the Specific Electrical Conductivity of Semiconducting Slabs 361

A. Handstein, J. Schneider and J. Kadlecová: Temperature and Magnetic Field Dependence of Magnetization and Magnetoresistance in Amorphous Fe–Ni Based Alloys 371

A. Gladun, C. Gladun and H. Vinzelberg: Thermal Conductivity of Niobium and Molybdenum Single Crystals at Low Temperatures 383

P. Debreceni, I. Kirschner and T. Porjesz: Crystal Symmetry Effects of the Conduction Tensor .. 395

G. Pompe, M. Gaafar, M. Falz and P. Büttner: Thermal Conductivity of Amorphous Iron Alloys at Low Temperatures 401

T. Porjesz: New Perspective in Semiconductor Physics Based on Advanced Technology 411

K. Fröhlich, D. Machajdik, S. Benacka and V. Cernusko: On the Relation between T_c Measurements and Structure Inhomogeneity of Nb_3Ge Superconductor 419

J. Kruzliak, P. Hutka, P. Kovác and P. Setina: Some Physical Properties of Multifilamentary Nb_3Sn Superconductors Prepared by Bronze Technology 425

F. Gömöry: AC Loss Measurement of Superconducting Magnets. Influence of the Compensating Scheme 433

J. Bánkuti and L. Lászlóffy: Superconducting Laboratory Magnets 445

A. Juhász, I. Főzy, P. Tasnádi, I. Kovács and I. Vitányi: Grain Boudary Motion in Vanadium .. 455

The final manuscript was received by the Editorial Office on 30th April 1982.

CRYOSEM '81

IInd International Seminar on
Low Temperature Physics

Visegrád, Hungary

October 8—10, 1981

PREFACE

Representatives of low temperature physicists of Central Europe met for the second time between 8th and 10th October 1981 to discuss the research aims and achievements of their Institutes. The conference, that we named CRYOSEM '81 (Second International Seminar on Low Temperature Physics and Technology) was hosted by Visegrád, a small Hungarian holiday resort, rich in historical monuments. The purpose of this second meeting, too, was to provide participants an opportunity for informal and free discussions, concerning any "cryogenic" problem, that is any problem connected with low temperatures, whether theoretical, experimental or technical.

It is the spirit of these discussions that is reflected by the papers published in this issue, giving an illustrative survey of the low temperature physical research of neighbouring Central European countries.



Prof. I. Kirschner

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 305—312 (1982)

V_3Ga SUPERCONDUCTORS WITH A HIGH CRITICAL CURRENT DENSITY

K. FISCHER, S. ROHR, J. MÖNCH, W. HOLZHÄUSER and CH. FRENZEL

CENTRAL INSTITUTE OF SOLID-STATE PHYSICS AND MATERIALS RESEARCH
ACADEMY OF SCIENCES OF THE GDR
8027 DRESDEN, GDR

The effect of In, Zn and Mn additions to the bronze matrix on formation rate, critical current density, and critical temperature of V_3Ga layers in bronze-processed V_3Ga wires has been investigated. The V_3Ga layers were produced by heat treatment at 600...650 °C within a period of 20...200 h. For a given heat treatment condition the critical current density (j_c) of the V_3Ga phase and the critical current (I_c) of In-containing conductors with a suitable composition of bronze (e.g. 16 at % Ga, 1.2 at % In) are considerably higher (by the factor $\geqq 1.7$) than those of conductors with a usually applied, binary Cu – Ga alloy containing 19 at % Ga. The critical temperature of In-containing samples is relatively high and approaches values up to 15.4 K (for which the specimens showed half its normal-state resistance). Too high Ga and In contents of the matrix are unfavourable for obtaining high critical parameters. An addition of Mn (0.6 at %) to the bronze leads to considerably thicker V_3Ga layers, however, j_c and T_c down to 13 K are decreased and thus, depending on heat treatment conditions and Ga content of bronze, unchanged or reduced I_c values were obtained.

Substitution of Zn for 2 at % Ga in the matrix causes a reduction of V_3Ga layer thickness and an increase of j_c.

Introduction

Up to now mainly Nb_3Sn and V_3Ga have been used to construct superconducting magnets for fields above 10 Tesla. The current carrying capacity of multifilamentary Nb_3Sn has now been raised to such a level that recently magnets for up to 15 T could be presented [1].

With V_3Ga very high critical current densities in fields above 10 T were obtained [2]. Moreover theoretical considerations show a possibility to enhance the H_{c2}-value. So V_3Ga seems to be an especially suitable material for high field magnets.

In recent years some experience has been gathered at our Institute with the production of V_3Ga multifilament wire by bronze process, and its application in magnets. The basic material is a wire of 0.18 mm diameter containing 109 filaments, (see Fig. 1).

As described by Furuto et al [3] six of the wires are stranded around a central tungsten wire serving as a mechanical stabilizer. After annealing the strand at 600 °C for 100 h it is soldered with pure indium, (see Fig. 2). This type of conductor was used in a NbTi/V_3Ga magnet system which produces 11 T in 30 mm free bore [4].

The V_3Ga coil of a larger system for 12.4 T in 80 mm free bore was wound from a double stranded cable containing 6 of the above mentioned strands and a central copper wire.

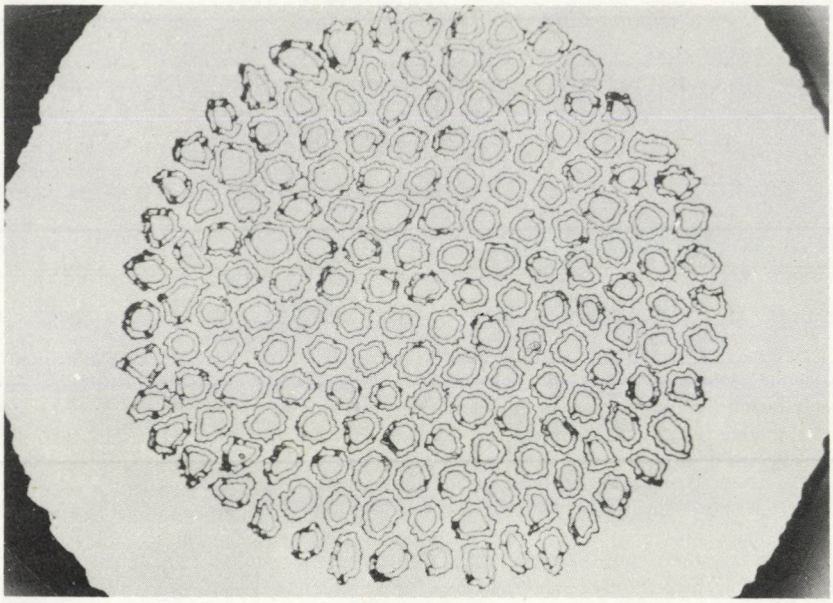


Fig. 1. Cross section of single multifilamentary wire (× 630).

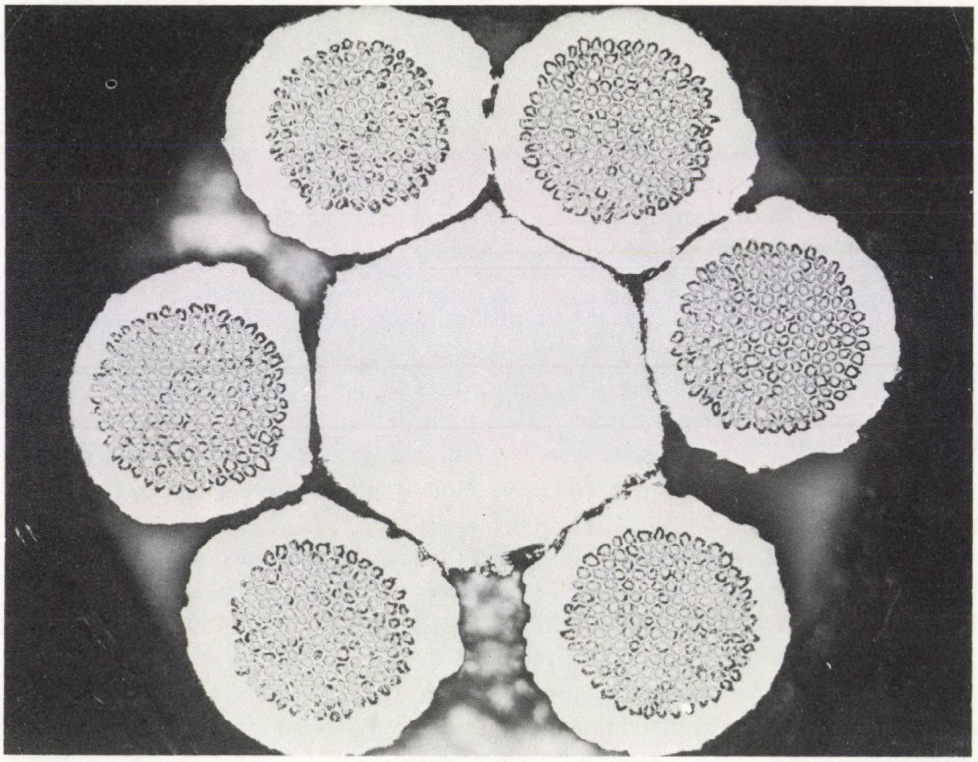


Fig. 2. Cross section of strand (× 250).

Higher magnetic fields can be economically produced only by raising the overall critical current density in the winding. We tried to increase the current carrying capacity of V_3Ga by adding third elements to the bronze, a method successfully applied by several authors. In most known cases the Sn and Ga content, respectively, which is frequently used for binary bronze, must be reduced in order to maintain room temperature workability. This means that the layer thicknesses of superconducting phase are equal or lower in the conductors with additions than in conductors with a binary Cu-7 at % Sn and Cu-19 at % Ga bronze, respectively. An increase of critical current (I_c) and the overall critical current density, respectively, therefore, is exclusively due to the increase of the critical current density (j_c) of superconducting phase. In the case of V_3Ga conductors especially in strong fields j_c is increased by a partial substitution of Al or Mg for Ga in the bronze [5, 6]. An increase of j_c in Nb_3Sn conductors is obtained by Al additions (by adding Al simultaneously to the Nb core material) and Ga additions to the Cu – Sn bronze [7, 8]. However, the Nb_3Sn-layer growth is considerably decreased by Ga additions.

The paper shows the influence of low In, Zn or Mn additions to the Cu – Ga bronze on superconducting properties and V_3Ga-layer growth of bronze processed V_3Ga conductors. The results are compared with those of Al additions which are the most efficient according to Yoshida et al [5].

Experimental procedures

For the production of single core wires by the bronze route a V rod having a diameter of 10 mm was inserted into the cavity of a cylinder consisting of the corresponding bronze material. This cylinder was vacuum-soldered and extruded at about 650 °C.

Employing intermediate anneals (625 °C) this composite rod was reduced by wire drawing to 0.2 mm wires. The final diameters of the V cores of conductors among each other differed by about 20% and ranged between 0.045 and 0.055 mm. For the formation of the V_3Ga phase the wires have been annealed for some periods at 600 °C and 650 °C. With the exception of In the additional elements were alloyed to the Cu – Ga bronze directly in the vacuum-induction furnace during melting. Indium was electrolytically precipitated on to the Cu – Ga/V wires of final dimensions, i.e. immediately before the subsequent heat treatment leading to V_3Ga formation. This method allows the production of wires with bronze sheaths, showing high Ga and In contents, and therefore they are not easily workable. A solution of 57 g/l In, 100 g/l NaCl and 1 g/l gelatine in diluted hydrochlorid acid (pH $\approx$ 2) was used as electrolyte. Current densities up to 2.10^{-2} A/cm^2 in the quiescent electrolyte are possible to separate suitable layers (bath temperature: 25 °C). The In concentrations of bronze mentioned below were calculated from the amount of bronze, determined from the sheath thickness of conductors, and the precipitated amount of In.

The critical temperature was determined resistively at a relative accuracy of ± 0.1 K. The temperature at which the specimen showed half its normal-state resistance was taken as T_c. The measurement of critical current in the field of 5 T was carried out in a NbTi magnet. For the I_c measurements in fields >8 T a Bitter magnet of ILMT Wroclaw was used. The current where a voltage of 5 μV occurred over a length of 1 cm was defined as I_c.

The area of V_3Ga phase, necessary for the calculation of critical current density of V_3Ga, was determined from micrographs of the cross section of conductors by means of the image analyzer "Quantimet 720".

The concentration distribution of In along the cross section of conductors was investigated by means of the SEMQ of ARL.

Results and discussion

Fig. 3 shows that the critical current of a conductor, the bronze sheath of which contains 16 at % Ga and 1.2 at % In, is considerably higher than the value of a conductor produced from a binary Cu – Ga bronze with the usual content of about 19 at %. The In-containing conductor supersedes also that with Al addition. It should be noted that I_c of the In-containing conductor also after long periods of reaction even increases with reaction time, while it remains nearly constant in the case of a conductor with binary bronze. The same effect was observed earlier by Yoshida et al [5] also for Al-containing conductors and confirmed by the present investigations.

Fig. 4 illustrates the effect of ternary additons to the bronze on critical current density of V_3Ga phase in some conductors, the I_c values of which are presented in Fig. 3. Indium and zinc additions to the bronze can cause a considerable increase of j_c of the V_3Ga phase, while j_c is decreased by Mn additions. The decreasing effect of Mn on j_c has

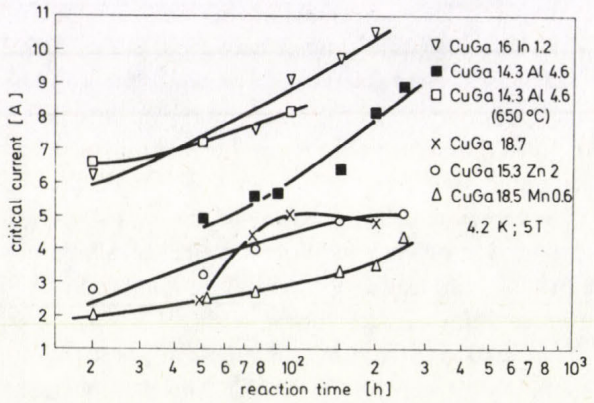


Fig. 3. Critical current vs reaction time of conductors with different compositions of the bronze (reaction temperature 600 °C, for one conductor (□): 650 °C; the concentrations of bronzes are at %).

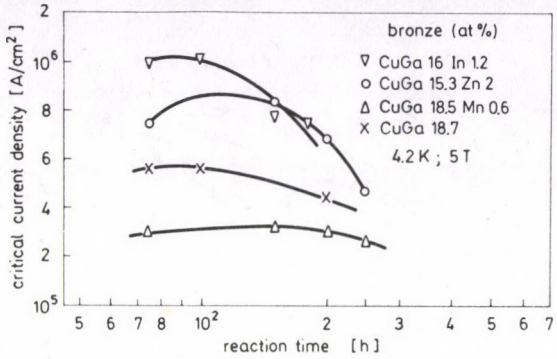


Fig. 4. Critical current density of the V_3Ga phase vs reaction time for conductors with different compositions of the bronze (reaction temperature: 600 °C).

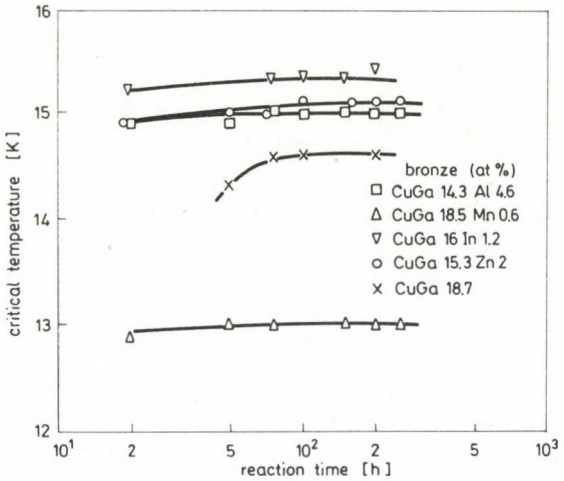


Fig. 5. Critical temperature vs reaction time of conductors with different compositions of the bronze (reaction temperature: 600 °C).

also been observed in the case of a conductor with a Cu – 20 at % Ga – 0.6 at % Mn bronze. This may be due first of all to the strongly decreased temperatures of Mn-containing conductors (Fig. 5). Probably the Mn diffuses into the V_3Ga layer and substitutes for a part of the vanadium of the V_3Ga lattice leading to the mentioned T_c decrease [9].

According to the investigations by Savitzkij et al [10] a similar effect on T_c can be expected also for In, Al and Zn, if a large amount of these diffuses from the bronze into the V_3Ga and substitutes there for a certain amount of Ga. In contrast to this, Fig. 5 shows that the conductors with the mentioned additions have relatively high T_c values. Investigations by means of X-ray microanalysis showed that only about 0.3 at % In are dissolved in the V_3Ga layers. Similar low Al contents were found by Yoshida et al in the V_3Ga Al-containing conductors [5].

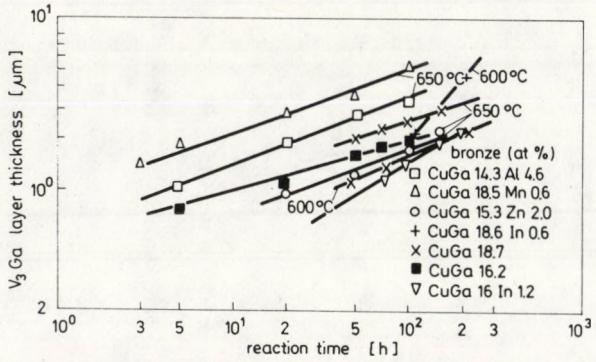


Fig. 6. V_3Ga layer thickness vs reaction time of conductors with different compositions of the bronze (reaction temperature: 600 °C and 650 °C).

The relatively high T_c values of the Al- or In-containing conductors in comparison with the value of the conductor without third elements are reasonable, if one assumes that the composition of V_3Ga phase in the conductor with the binary bronze is off-stoichiometric. Third elements diffusing in small amounts from the bronze into the V_3Ga layer approach the V:(Ga+X) ratio of the V_3Ga phase to the stoichiometric composition, thus increasing T_c (X = In, Al, Zn). Another possibility to approach the stoichiometric composition is that the addition elements (In, Al) increase the chemical activity of Ga in the bronze, thus increasing the Ga content of the V_3Ga phase.

Table I contains the values of critical current, critical current density and critical temperature for some In-containing conductors with different Ga or/and In contents of bronze. These values lead to the conclusion that the positive influence of In on the mentioned variables decreases for higher Ga and/or In contents of bronze.

The influence of ternary elements of the bronze on the V_3Ga layer growth rate is demonstrated in Fig. 6. In and especially Mn additions accelerate the V_3Ga layer

Table I

Critical parameters of In-containing conductors with different compositions of the bronze heat-treated at 600 °C

Alloy element cont. of bronze [at %]	Reaction time [h]	I_c (5 T) [A]	j_c (5 T) [10^5 A/cm^2]	T_c [K]
Ga 16.1 In 0.6	100	7.1	6.6	15.1
	200	6.0	4.3	15.0
Ga 16 In 1.2	100	9.0	10.0	15.3
	200	10.5	7.2	15.4
Ga 18.6 In 0.6	100	6.5	5.9	14.7
	200	6.7	2.5	14.3
Ga 18.5 In 1.2	100	5.8	3.5	14.6

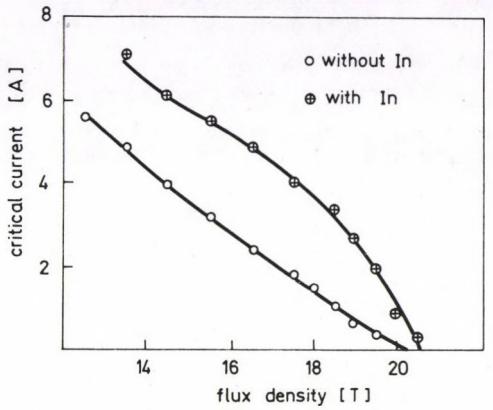


Fig. 7. Critical current vs magnetic flux density of multifilament conductors with and without In added to the bronze (Ga content of bronzes: 16.2 at %, heat treatment conditions: 600 °C/200 h).

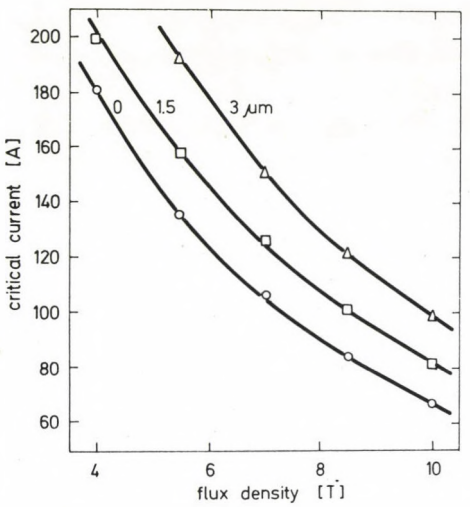


Fig. 8. Critical current vs magnetic flux density of a stranded cable as shown in Fig. 2, without and with In deposition on to the surface before annealing. Annealing temperature: 600 °C, time: 140 h (parameter is the thickness of deposited In layer).

growth. This may be due to an increased activity of Ga in the bronze by these additions. As a comparison between CuGa 16.2 conductors and CuGa 15.3 Zn2 conductors shows, the partial substitution of Zn for Ga in the bronze leads to a retardation of layer growth.

The positive effect of adding In to the bronze has also been confirmed with multifilamentary wire and stranded cable as is seen in Figs 7 and 8.

Conclusions

Additions of In to the bronze are an effective means to increase the critical current of V_3Ga conductors (produced by the bronze route). To obtain high critical superconducting parameters of In-containing conductors Ga contents lower than 19 at % in the bronze are favourable. As a result, the V_3Ga layers of In-containing conductors with high I_c are not much thicker than those of conductors with a binary 19 at % Ga bronze, although In additions accelerate the V_3Ga layer growth. Thus the higher I_c values of In-containing conductors in comparison with a conductor having a binary 19 at % Ga bronze are mainly due to increased j_c values of the V_3Ga phase.

A Zn addition to the bronze considerably increases the critical current density of V_3Ga phase of bronze-processed conductors. As a result, despite their thinner V_3Ga layers the critical current of specimens with a total Ga + Zn content of 17.3 at % in the bronze, which is not an upper concentration level for good workability, is about the same as for specimens with a binary 19 at % Ga bronze. This suggests that Zn-containing conductors with a Ga or Zn content higher than that investigated here have I_c values superior to best conductors with a binary bronze.

Mn additions to the bronze are not suitable to improve the I_c of bronze-processed conductors. Mn increases the V_3Ga amount of the conductor for a given heat treatment, but the critical temperature and thus the critical current density of V_3Ga phase is considerably decreased.

Acknowledgement

The authors wish to thank Ms. Becher and Mr. Bitterlich for the XMY investigations and Dr. Fuchs for the I_c measurements in high fields.

References

1. H. J. Weiße, M. Wilhelm, K. Wohlleben and E. Springer, IEEE Trans. Magn. MAG-**17,** 1628, 1981.
2. D. U. Gubser, T. L. Francavilla, D. G. Howe, R. A. Muessner and F. T. Ormand, IEEE Trans. Magn. MAG-**15,** 385, 1979.
3. Y. Furuto, T. Suzuki, K. Tachikawa and Y. Iwasa, Appl. Phys. Lett., **24,** 34, 1974.
4. D. Eckert, K. Fischer, C. Frenzel, A. Gladun, W. Holzhäuser, M. Knorn, F. Lange, A. Möbius and P. Verges, IEEE Trans. Magn. MAG-**17,** 1632, 1981.
5. Y. Yoshida, K. Tachikawa and Y. Iwasa, Appl. Phys. Lett., **27,** 632, 1975.
6. K. Tachikawa, Y. Tanaka, Y. Yoshida, T. Asano and Y. Iwasa, IEEE Trans. Magn. MAG-**15,** 391, 1979.
7. D. Dew Hughes, T. S. Luhman and M. Suenaga, Nucl. Techn., **29,** 268, 1976.
8. D. Dew Hughes, IEEE Trans. Magn. MAG-**13,** 651, 1977.
9. K. Girgis, W. Odoni and H. R. Ott, J. Less Comm. Met., **46,** 175, 1976.
10. E. M. Savizkij: Sverhprovodyashchie Soedineniya Perehodnih Metallov, Nauka, Moskva, 1976, 78—89.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 313—319 (1982)

DETERMINATION OF MICROSCOPIC PARAMETERS BY MACROSCOPIC MEASUREMENTS IN SUPERCONDUCTORS

I. KIRSCHNER and K. SAJÓ

DEPARTMENT FOR LOW TEMPERATURE PHYSICS, ROLAND EÖTVÖS UNIVERSITY
BUDAPEST, HUNGARY

Macroscopic magnetic measurements were performed on superconducting alloys. From the results of these experiments microscopic parameters namely the energy gap and the vortex creation energy could be calculated. Their dependence on the additive material content (or on the GL-parameter) was determined. It was shown that the BCS and the GLAG theory are applicable for small-κ superconductors in wide temperature interval.

Introduction

In the phenomenon of superconductivity the magnetic properties of the specimen, which are equivalent to the electrical properties, play a fundamental role.

The investigation of the magnetic characteristics of superconductors is very suitable for gaining knowledge on some microscopic and macroscopic parameters. These are highly essential in the investigation of superconductivity.

Experiments were performed to study small-κ superconducting alloys. Beside our experimental data, some measured results of other authors were also applied to construct the main characteristic curves.

Experiments

The following functions and quantities were determined experimentally:

1. the dependence of the critical temperature T_c on the alloying material content C;
2. the dependence of the residual resistance ρ_n on the additive material concentration;
3. magnetic moment – magnetic field ($M-H$) characteristics at different temperatures T;
4. temperature dependence of the critical magnetic fields H_{c1}, H_c and H_{c2};
5. concentration dependence of critical fields extrapolated to zero $H_{c1}(0)$, $H_c(0)$ and $H_{c2}(0)$;
6. dependence of the GL parameter κ on the temperature;
7. dependence of GL parameter extrapolated to zero $\kappa(0)$ on the additive material content.

All these curves are demonstrated in [1, 2] for the alloys of In – Bi, In – Tl, In – Pb, Sn – Bi and Sn – Sb. Of these the typical $T_c - C$, $\kappa(0) - C$, $H_{ci}(0) - C$ and

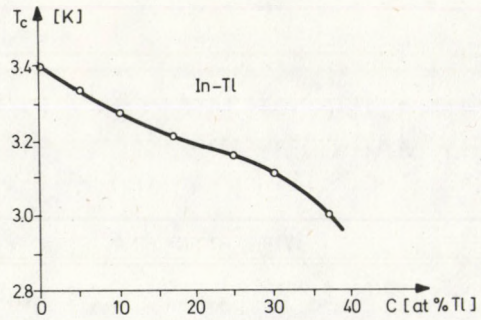


Fig. 1. Dependence of the critical temperaure on the alloying material content.

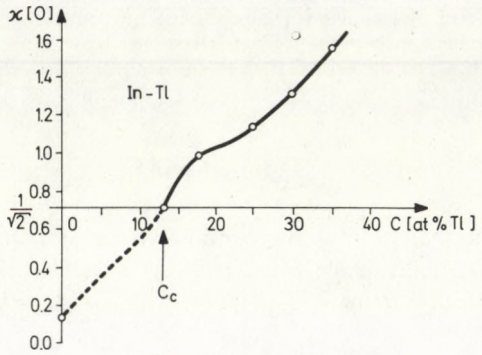


Fig. 2. Dependence of the GL parameter on the additive material concentration.

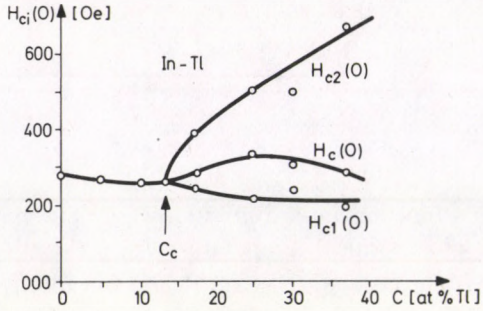


Fig. 3. Dependence of the critical magnetic fields on the alloying material content.

$\gamma - C$ characteristics are shown in Figs 1, 2, 3 and 4 for the alloys of In—Tl, on the basis of the direct critical temperature and magnetization measurements of Love et al [3], where γ is the electronic specific heat coefficient. In the Figures the arrow with symbol C_c indicates the phase transition [4, 2] at which type I superconductivity goes over into type II due to the increasing additive material content.

The magnetic moment — magnetic field measurements performed at different temperatures and concentrations give a possibility to determine the $\kappa - C$ function.

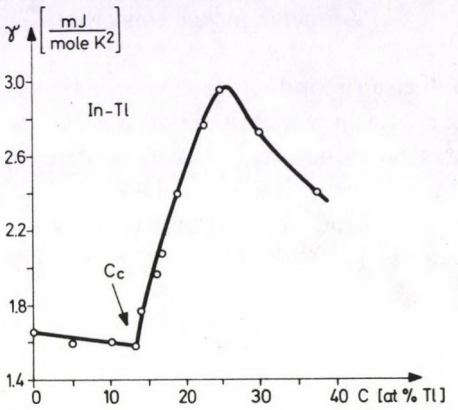


Fig. 4. Dependence of the electronic specific heat coefficient on the additive material concentration.

There are three ways of determining the GL parameter of the samples. Of these, the Gor'kov – Goodman relation [5, 6]

$$\kappa = \kappa_0 + 7.5 \times 10^3 \gamma^{1/2} \rho_n \tag{1}$$

uses the value of the residual resistance, where κ_0 is the GL parameter of the starting material. The corresponding relations of the GLAG theory [7, 8, 5]

$$\sqrt{2}\kappa = \frac{H_{c2}}{H_c} = \frac{(\mathrm{d}H_{c2}/\mathrm{d}T)_{T_c}}{(\mathrm{d}H_c/\mathrm{d}T)_{T_c}}, \tag{2}$$

$$4\pi \left(\frac{\mathrm{d}M}{\mathrm{d}H}\right)_{H_{c2},\, T_c} = \frac{1}{1.16(2\kappa^2 - 1)} \tag{3}$$

are based on the shape of the curves $H_{ci} - T$ and $M - H$. In these formulas $\kappa_0(T_c) = 0.11$ for In, $\kappa_0(T_c) = 0.12$ for Sn, $\gamma = 1.66$ mJ/mole grad^2 for In and $\gamma = 1.74$ mJ/mole grad^2 for Sn. The other essential parameters are $T_c = 3.40$ K, $H_c(0) = 278$ Oe for In, and $T_c = 3.72$ K, $H_c(0) = 303$ Oe for Sn. The three methods give values slightly differing from each other.

Of the formulas given in the literature the Gor'kov equation

$$\kappa(t) = (1.25 - 0.30t^2 + 0.05t^2)\kappa(T_c), \tag{4}$$

where $t = T/T_c$, describes the temperature dependence of the GL parameter the most precisely. According to this relation, at $T = 0$ K, for In, $\kappa_0(0) = 0.14$ and for Sn, $\kappa_0(0) = 0.15$.

Similarly, Gor'kov was able to describe the temperature dependence of H_{c2} in agreement with experiments in the form

$$H_{c2}(t) = H_c(0)\,(1.77 - 2.20t^2 + 0.50t^4 - 0.07t^6)\kappa(T_c), \tag{5}$$

while for the $H_c - t$ a quadratic function was supposed near T_c.

Parameter for the condensation

As it is known, the superconducting (condensed) phase can be characterized by an energy gap in the spectrum, which is very typical of the superconductivity.

This is a microscopic quantity, but can be determined on the basis of our experimental work by an appropriate method.

The concentration dependence of the electronic specific heat coefficient can be calculated from the curves $T_c - C$ and $H_c(0) - C$ by the help of the formula

$$\gamma = 0.17 \frac{H_c(0)^2}{T_c^2}, \tag{6}$$

which is obtained in the BCS theory [9] in the form of a similarity law.

This was used for the evaluation of the measured data to calculate the electronic specific heat coefficient showing a picture on the electrical and magnetic properties of alloys.

Since the expression (6) used for determining γ is based on the same approximation as the relation

$$\frac{2\Delta(0)}{kT_c} = 3.52 \tag{7}$$

in the BCS theory, it is this that will be compared first with the experimental results to determine the energy gap. The possibility to do it is given by Toxen's empirical formula [10]

$$\frac{T_c}{H_c(0)} \left[\frac{\mathrm{d}H_c(T)}{\mathrm{d}T}\right]_{T_c} = \frac{\Delta(0)}{kT_c}, \tag{8}$$

where $\Delta(0)$ is the energy gap and k is the Boltzmann constant.

In this calculation we did not use the theoretical function for $H_c(t)$

$$H_c(t) = H_c(0)\,(1 - t^2) \tag{9}$$

but the real dependence of H_c on t, which deviates from it to a certain extent. The genuine shape of $H_c(t)$ originates from our experiments [1] and from the results of Kinsel et al [11] and Noto et al [12].

As an example, the temperature dependence of H_{ci} is demonstrated in Fig. 5 for In – 2.5 at % Bi alloy.

The data obtained for In – Bi and In – Pb alloys show that the ratio $\frac{2\Delta(0)}{kT_c}$ is independent of the additive material content in this concentration interval and has the value 3.7, which is very near to the magnitude of 3.6 obtained for pure In. It must be noted that these experimental values (see Fig. 6) also lie in the vicinity of the magnitude 3.5 of the BCS theory.

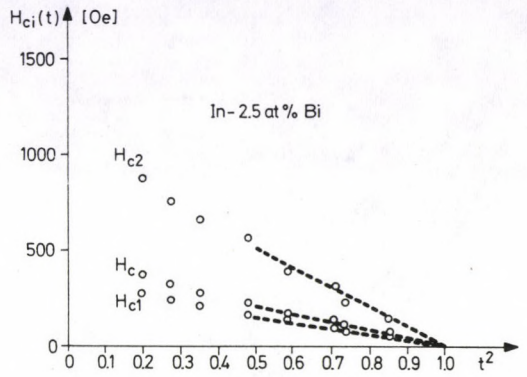


Fig. 5. Dependence of the critical magnetic fields on the temperature.

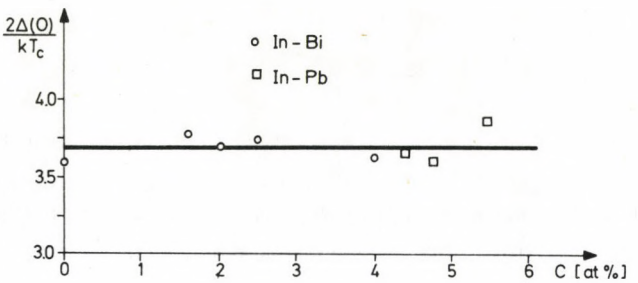


Fig. 6. Dependence of the ratio of the energy gap and actual critical temperature on the alloying material content.

Parameter for magnetic vortices

In type II superconductors in the presence of an appropriate external magnetic field a magnetic vortex lattice comes into being in which the vortices correspond to the normal electrons. This process requires some energy, that is just the vortex creation energy E_v, referring to the unit length of the vortices.

E_v can be determined by the experimental quantities existing in the expression

$$E_v = \frac{H_c}{2\sqrt{2}\kappa} H_{c1} = \frac{1}{2} \frac{H_c^2}{H_{c2}} H_{c1} . \tag{10}$$

For the evaluation of the measured values the GLAG theory is applicable to study of the vortex creation energy.

In Fig. 7 our experimental results have been compared with the theoretical results of Tholfsen and Meissner [13], and of Harden and Arp [14], who calculated the vortex creation energy by numerically integrating the Ginzburg – Landau equations.

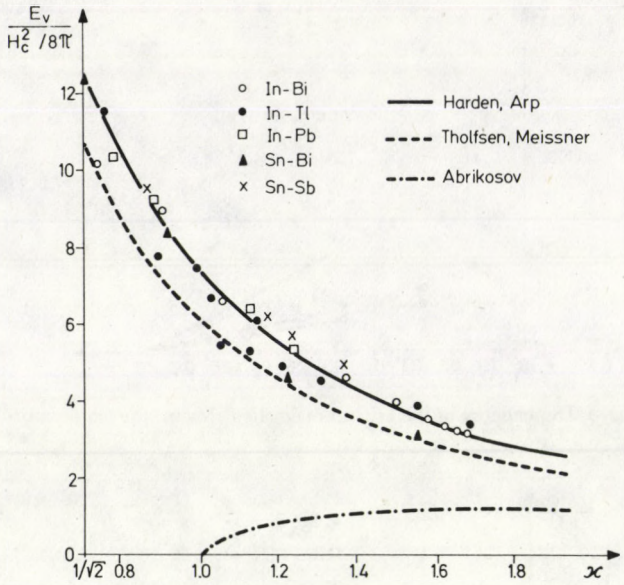


Fig. 7. Dependence of the vortex creation energy on the additive material concentration.

Tholfsen and Meissner carried out this operation for $\kappa = 0.2$, 0.5, 1.0, 5.0 and 20. The points obtained by them form a curve with a shape described by the formula

$$E_v = \frac{6.2}{4\pi} H_c^2 \lambda^2 \kappa^{-1.6}, \tag{11}$$

were λ is the penetration depth.

Harden and Arp represented their results graphically. The fundamental meaning of their work is that E_v is proportional to H_{c1}.

The Figure shows that the experimental data coincide well with the curve $E_v - \kappa$ obtained by Harden and Arp while the results of Tholfsen and Meissner show a deviation of about $10-20\%$ decreasing towards the higher values of κ. This is due to their taking numerical vortex solutions into account even for $\kappa(0) < 1/\sqrt{2}$ supposing that vortices might be formed in the thin surface layers of the superconductors.

Neglecting this point the two solutions are in agreement with each other.

It must be noted that the experimental points in the Figure originate from data extrapolated for $T = 0$ K, thus the good agreement indicates the applicability of the GLAG theory.

Abrikosov [8] determined the vortex creation energy analytically for $\kappa \gg 1$ and he obtained the formula

$$E_v = \frac{H_c^2 \lambda^2}{2\kappa^2} (\ln \kappa + 0.08). \tag{12}$$

As it was expected, this relation produces good results only above $\kappa > 10$.

Conclusions

The investigations we have carried out lead to some interesting conclusions on small-κ superconductors [15]:

1. It is clear that the energy gap and the critical temperature change together in the studied concentration range. It means that the ratio of $\Delta(0)/T_c$ does not depend on the contaminations in superconducting alloys under investigation. It can be expected that other small-κ superconductors also have such a behaviour.

2. These investigations show that the vortex creation energy depends on the alloying material content so that it decreases with the contamination (with increase of the GL parameter), which means that the vortex creation energy is the bigger the more the state is type I superconductivity, i.e. the less it is type II.

3. It was found that the BCS and the GLAG theory can be applied to describe the magnetic features of small-κ superconductors in the whole $0-T_c$ temperature region.

References

1. I. Kirschner, J. Bánkuti, G. Kiss, I. Kovács, L. Lászlóffy, G. Reményi and K. Sajó, Zh. Eksp. Teor. Fiz., **66,** 2141, 1974. (Sov. Phys.–JETP, 39, 1054, 1974.).
2. I. Kirschner and K. Sajó, J. Low Temp. Phys., **39,** 235, 1980.
3. W. F. Love, E. Callen and F. C. Nix, Phys. Rev., **87,** 844, 1952.
4. I. Kirschner, Phys. Lett., **A47,** 139, 1974.
5. L. P. Gor'kov, Zh. Eksp. Teor. Fiz., **37,** 833, 1959.
6. B. B. Goodman, IBM J. Res. Dev., **6,** 63, 1961.
7. V. L. Ginzburg and L. D. Landau, Zh. Eksp. Teor. Fiz., **20,** 1064, 1950.
8. A. A. Abrikosov, Zh. Eksp. Teor. Fiz., **32,** 1442, 1957.
9. J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev., **108,** 1175, 1957.
10. A. M. Toxen, Phys. Rev. Lett., **15,** 462, 1965.
11. T. Kinsel, E. A. Lynton and B. Serin, Rev. Mod. Phys., **36,** 105, 1964.
12. K. Noto, Y. Muto and T. Fukuroi, J. Phys. Soc. Japan, **21,** 2122, 1966.
13. P. Tholfsen and H. Meissner, Phys. Rev., **169,** 413, 1968.
14. J. L. Harden and V. Arp, Cryogenics, **3,** 105, 1963.
15. I. Kirschner and K. Sajó, Phys. Lett., **A88,** 205, 1982.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 321—330 (1982)

APPLICATION OF THE "BOSON METHOD" TO THE STUDY OF SUPERCONDUCTING ORDER PARAMETER

A. PEVALA

INSTITUTE OF ELECTRICAL ENGINEERING, SLOVAK ACADEMY OF SCIENCES
809 32 BRATISLAVA, CZECHOSLOVAKIA

A brief outline of the so called "boson method" in the theory of superconductivity is presented. The computation of the space dependence of superconducting order parameter by means of this method has some advantages compared to standard techniques based on Gor'kov equations. The possibility of obtaining the space dependence of this quantity in the general form is shown.

1. Introduction

In the last few years great attention has been paid to various inhomogeneous structures in the field of superconductivity – such as layered superconductors, surface and interface effects on superconductivity, effects of impurities, etc. All those situations, which are related to the presence of a non-homogeneous structure of the ground state are usually solved by means of Gor'kov equations. It is well known that these equations are not easy to handle. Usually, it is only a perturbative solution which gives some analytic results [1]. Interfaces are simulated by step-function changes in BCS interaction constant. The exact solutions of Gor'kov equations require a great amount of numerical calculations which, however, could be performed only under special assumptions. These integro-differential equations can lead to the Ginzburg – Landau equations in some cases (e.g. near the critical temperature T_c). However, the solutions of the Ginzburg – Landau equations are not very simple to obtain, either.

The "boson method" is a new formulation of the theory of superconductivity which differs significantly from the usual methods. It was developed in connection with dynamical rearrangements of symmetries [2] and since the mid-seventies it acquired a number of applications [4 – 8].

A keynote of the method is the introduction of the so-called boson transformation, under which the field equations remain invariant, but the expectation values of various observables may change. By means of boson transformation we may pass from space-independent solutions to space-dependent ones without solving the Gor'kov equations. The practical consequence of this is that if we know the situation where the order-parameter is space (and time) independent, the space dependence is obtained by this transformation. We shall use this fact in Section 3.

In the following, we shall give a brief summary of the boson method. Although the technique required for practical use is based on a set of rather simple equations, the derivation of them is not simple. One has to use a special technique of quantum field theory [3]. For a full description see [9], where many applications are given, too.

2. The outline of the boson method

For the sake of simplicity we shall consider the case of zero temperature and we shall not take into account the Coulomb interaction between electrons which, however, can be built into the theory in a simple way [9, 10]. We start from the well-known BCS Hamiltonian

$$H=\int \mathrm{d}^3x[\psi_\uparrow^+ \in(\partial)\psi_\uparrow+\psi_\downarrow^+ \in(\partial)\psi_\downarrow]+\lambda\int \mathrm{d}^3x\,\psi_\uparrow^+\psi_\downarrow^+\psi_\downarrow\psi_\uparrow, \tag{2.1}$$

where $\psi_\uparrow, \psi_\downarrow$ are Heisenberg operators of electron fields with a given orientation of spins ($\uparrow, \downarrow$). λ is the coupling constant of electron–phonon interaction and

$$\in(\partial)=-\frac{1}{2m}(\nabla^2+k_F^2), \tag{2.2}$$

where k_F is the value of Fermi momentum and m means the electron mass (units: $\hbar=c=1$). The electron fields $\psi_{\uparrow,\downarrow}$ satisfy the equations of motion, derived from Hamiltonian (2.1). The equilibrium system of superconductor can be described according to the self-consistent field method [9] by a set of free fields:

$$\{\varphi_\uparrow, \varphi_\downarrow, B\}. \tag{2.3}$$

These are the "free" fields of quasifermions ($\varphi_\uparrow, \varphi_\downarrow$) and the boson field B:

$$\varphi_{\downarrow,\uparrow}(x)=\int \mathrm{d}^3k(2\pi)^{3/2}\exp(i\mathbf{k}\cdot\mathbf{x})\{\alpha_{\mathbf{k}\downarrow,\uparrow}\cos\Theta_{\mathbf{k}}\exp(-iE_{\mathbf{k}}t)\pm \pm\alpha^+_{-\mathbf{k}\uparrow,\downarrow}\sin\Theta_{\mathbf{k}}\exp(iE_{\mathbf{k}}t)\}, \tag{2.4}$$

$$B(x)=\int\frac{\mathrm{d}^3l}{(2\pi)^{3/2}}\frac{1}{(2\omega_{\mathbf{l}})^{1/2}}(B_{\mathbf{l}}e^{i\mathbf{l}\cdot\mathbf{x}-i\omega_{\mathbf{l}}t}+B_{\mathbf{l}}^+e^{-i\mathbf{l}\cdot\mathbf{x}+i\omega_{\mathbf{l}}t}), \tag{2.5}$$

where $\alpha_{\mathbf{k}\uparrow,\downarrow}$ and $B_{\mathbf{l}}$ play the role in the construction of Fock space by defining the vacuum state as

$$\alpha_{\mathbf{k}\uparrow,\downarrow}|0>=0; \qquad B_{\mathbf{l}}|0>=0. \tag{2.6}$$

The operators $\alpha_{\mathbf{k}}$ and $B_{\mathbf{l}}$ are the annihilation operators of quasifermions and of quasibosons, respectively. They satisfy the usual anticommutation and commutation relations; mixed commutators are equal to zero.

In (2.4) and (2.5) $\Theta_{\mathbf{k}}$ is the Bogoljubov's parameter given by the relations

$$\cos 2\Theta_{\mathbf{k}}=\frac{\in_{\mathbf{k}}}{E_{\mathbf{k}}}; \qquad \sin 2\Theta_{\mathbf{k}}=-\frac{\Delta}{E_{\mathbf{k}}}, \tag{2.7}$$

where

$$\in_{\mathbf{k}}=\frac{1}{2m}(k^2-k_F^2), \tag{2.8}$$

$$E_{\mathbf{k}}=\sqrt{\in_{\mathbf{k}}^2+\Delta^2} \tag{2.9}$$

and Δ is the gap parameter, satisfying the integral equation

$$\frac{\lambda}{2}\int\frac{\mathrm{d}^3k}{(2\pi)^3}\frac{1}{E_{\mathbf{k}}}=1, \tag{2.10}$$

$\omega_{\mathbf{l}}$ is the energy spectrum of the boson field. It turns out that the electron operators ψ must be written in terms of "free" operators:

$$\psi=f(\varphi, B). \tag{2.11}$$

An expression obtained in that way is called "dynamical map". Since the observables are created by bilinear products of electron operators it is more suitable to construct dynamical maps for such products. We shall give here only the formula for the product $\psi_\uparrow\psi_\downarrow$:

$$T[\psi_\uparrow(x)\psi_\downarrow(y)]=\chi^{(1)}(x-y)+\int\mathrm{d}^3p\int\mathrm{d}^3q\,[F^{(1)}(x,y,\mathbf{p},\mathbf{q})\alpha_{\mathbf{q}\uparrow}\,\alpha_{\mathbf{p}\downarrow}- \tag{2.12}$$

$$-F^{(2)*}(x,y,\mathbf{p},\mathbf{q})\alpha^+_{\mathbf{p}\downarrow}\,\alpha^+_{\mathbf{q}\uparrow}]+\int\mathrm{d}^3l\,[G^{(1)}(x,y,\mathbf{l})\,B_{\mathbf{l}}-G^{(2)*}(x,y,\mathbf{l})\,B^+_{\mathbf{l}}]+\ldots$$

For other bilinear products (see e.g. [11]), the expressions are similar. Here, T is Wick's chronological operator. (2.12) is an expansion into the normal products of free operators; it can be shown that higher order terms in (2.12) can be neglected. The coefficients appearing here may be written as

$$\begin{aligned}\chi(x-y)&=\langle 0\mid T[\psi(x)\psi^+(y)]\mid 0\rangle,\\ F(x,y,\mathbf{p},\mathbf{q})&=\langle 0\mid T[\psi(x)\psi^+(y)]\alpha^+_{\mathbf{p}\downarrow}\,\alpha^+_{\mathbf{q}\uparrow}\mid 0\rangle,\\ G(x,y,\mathbf{l})&=\langle 0\mid T[\psi(x)\psi^+(y)]\,B^+_{\mathbf{l}}\mid 0\rangle.\end{aligned} \tag{2.13}$$

One can show that they obey the Bethe – Salpeter equations. ψ means electron doublet

$$\psi=\begin{pmatrix}\psi_\uparrow\\ \psi^+_\downarrow\end{pmatrix}$$

and therefore χ, F, G, are 2×2 matrices (their elements stand in expressions for other bilinear products).

Once the solutions of Bethe – Salpeter equations are known we may write all physical operators in terms of free fields. So, e.g. the Hamiltonian is given as

$$H=\int\mathrm{d}^3k\,E_{\mathbf{k}}(\alpha^+_{\mathbf{k}\uparrow}\,\alpha_{\mathbf{k}\uparrow}+\alpha^+_{\mathbf{k}\downarrow}\,\alpha_{\mathbf{k}\downarrow})+\int\mathrm{d}^3l\,\omega_{\mathbf{l}}\,B^+_{\mathbf{l}}\,B_{\mathbf{l}}, \tag{2.14}$$

$$+c-\text{number}$$

from where we immediately see the nature of the term "free fields". The current and charge densities take the form of two-term sum:

$$\rho(\mathbf{x},t)=\rho^{(1)}(\mathbf{x},t)+\rho^{(2)}(\mathbf{x},t), \tag{2.15}$$

$$\mathbf{j}(\mathbf{x},t)=\mathbf{j}^{(1)}(\mathbf{x},t)+\mathbf{j}^{(2)}(\mathbf{x},t), \tag{2.16}$$

where terms indexed by (1) are created only by quasifermion operators and those indexed by (2) by boson operators. We are interested in the last ones:

$$\rho^{(2)}(\mathbf{x}, t) = -\eta \frac{\partial}{\partial t} B(\mathbf{x}, t), \tag{2.17}$$

$$\mathbf{j}^{(2)}(\mathbf{x}, t) = v_1 \eta \nabla B(\mathbf{x}, t). \tag{2.18}$$

Each component of ρ and $\mathbf{j}$ conserves separately:

$$\nabla \mathbf{j}^{(i)} + \frac{\partial}{\partial t} \rho^{(i)} = 0; \qquad i = 1,2. \tag{2.19}$$

In (2.17) v_1 is the boson velocity ($\omega_1 = \mathbf{v}_1 \cdot \mathbf{l}$ in approximation) and η is a constant.

One of the main results of boson theory is the fact that

$$\int \mathrm{d}^3 x \rho^{(1)}(\mathbf{x}, t) = 0. \tag{2.20}$$

As it is known, the generator of the phase transformation of electron field ψ is a number operator

$$N = \int \mathrm{d}^3 x \, \rho, \tag{2.21}$$

because

$$e^{-i\Theta N} \psi_{\uparrow,\downarrow} e^{i\Theta N} = e^{i\Theta} \psi_{\uparrow,\downarrow}, \tag{2.22}$$

(here Θ means the phase). From (2.17) and (2.20) it is immediately seen that

$$N = -\eta \int \mathrm{d}^3 x \frac{\partial}{\partial t} B(\mathbf{x}, t). \tag{2.23}$$

In the boson theory one deals instead of N with a modified operator

$$N_f = \int \mathrm{d}^3 x f(\mathbf{x}) \rho(\mathbf{x}) = -\eta \int \mathrm{d}^3 x f(\mathbf{x}, t) \frac{\partial}{\partial t} B(\mathbf{x}, t), \tag{2.24}$$

where $f(\mathbf{x})$ is a certain function which must satisfy the Laplace equation (in the stationary case)

$$\nabla^2 f(\mathbf{x}) = 0 \tag{2.25}$$

and is equal to the half of the superconductor phase. The order parameter is just written as

$$\Delta(\mathbf{x}) = e^{2if(\mathbf{x})} |\Delta(\mathbf{x})|. \tag{2.26}$$

The transformation of free fields induced by N_f now takes the form

$$\begin{aligned} &\varphi_{\uparrow,\downarrow}(x) \to e^{-iN_f} \varphi_{\uparrow,\downarrow} e^{iN_f} = \varphi_{\uparrow,\downarrow}(x), \\ &B(x) \to e^{-iN_f} B e^{iN_f} = B(x) + \eta \int \mathrm{d}^3 y \, c(\mathbf{x} - \mathbf{y}) f(\mathbf{y}). \end{aligned} \tag{2.27}$$

In the last expression appears the function $c(\mathbf{x} - \mathbf{y})$ which is called "boson characteristic function" and plays a crucial role in the boson method. We shall use its simplified form

in the next Section. It is δ-function-like and has the property that

$$\int \mathrm{d}^3 y\, c(\mathbf{x}-\mathbf{y})=1\,. \tag{2.28}$$

From (2.27) we see that quasielectrons do not change and that the original transformation of electron field is now carried by certain transformation of boson field, which is called boson transformation. As it was already mentioned, we can obtain any space dependence of the order parameter through this transformation, knowing the situation where the order parameter is space independent.

Closing this Section, we summarize the procedure, which is used in solving the special space dependent problems:

1. We have to solve the Laplace equation (2.25) under specific boundary conditions (they depend on the problem).

2. We can compute persistent currents and magnetic fields associated with them, the energy, etc.; the expressions are given e.g. in [15]. All these quantities are functions of $c(\mathbf{x})$ and $f(\mathbf{x})$.

We see the extreme importance of $f(\mathbf{x})$; the order parameter as a whole does not play such an important role as its phase, because all observables are its functions. However, there are situations where the knowledge of the whole order parameter is useful – e.g. in studying various interfaces, inhomogeneities, etc. The next Section is devoted to the space dependent order parameter computation.

3. Space dependent order parameter problem

It can be shown [9] that the boson transformation (2.27) makes ψ to take the form

$$\psi = \exp\left[i(B/\eta + f(x)\right] F\left(\varphi, \nabla B, \frac{\partial B}{\partial t}\right), \tag{3.1}$$

where F is a certain function of given arguments. Then the order parameter is given by

$$\Delta(x)=\langle 0 \mid e^{\frac{2iB}{\eta}+2if} F^2\left(\varphi, \nabla B, \frac{\partial B}{\partial t}\right) \mid 0\rangle = e^{2if(x)} \mid \Delta(x) \mid . \tag{3.2}$$

This quantity satisfies the Ginzburg–Landau equations [10]. The space dependence of $\Delta(\mathbf{x})$ is carried out by f, F, B in (3.2). Although we know in principle the function f which is given by (2.25), we know nothing about the function F. It could be computed by expansion in powers of its arguments which we may restrict to small values, but this way seems to be very difficult. In our method we shall start from the dynamical map (2.12). The order parameter is defined as (we consider only the stationary case)

$$\Delta(\mathbf{x})=\lambda\langle 0 \mid T[\psi_\uparrow(\mathbf{x})\,\psi_\downarrow(\mathbf{y})] \mid 0\rangle \mid_{\mathbf{x}=\mathbf{y}}\,. \tag{3.3}$$

According to the scheme described in Section 2 we must first perform the boson transformation in (2.12) and then compute the expectation value of the expression

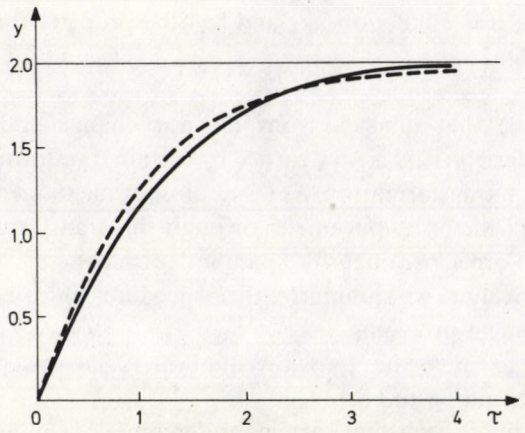


Fig. 1. Dependence of boson spectrum $y=\omega_l/\Delta$ on the relative momentum $\tau=l\cdot\xi_0$. The full line represents the exact numerical computations [13] for the value $\lambda N(0)=0.4$ ($\lambda N(0)$ being the coupling parameter). The dotted line is the approximation (3.6) for the same value of $\lambda N(0)$.

obtained in this way. Because the quasifermion field φ remains unchanged under this transformation, we have to deal only with boson operators $B_\mathbf{l}, B_\mathbf{l}^+$. Because the transformation (2.27) is done on the level of field operator B, we may express $B_\mathbf{l}, B_\mathbf{l}^+$ from (2.5):

$$B_\mathbf{l}=\frac{1}{2}\int\frac{\mathrm{d}^3x}{(2\pi)^{3/2}}\sqrt{2\omega_\mathbf{l}}\,B(\mathbf{x})e^{-i\mathbf{l}\cdot\mathbf{x}}. \tag{3.4}$$

After the transformation one has the form

$$B_\mathbf{l}^f=B_\mathbf{l}+\frac{1}{2}\int\frac{\mathrm{d}^3x}{(2\pi)^{3/2}}\sqrt{2\omega_\mathbf{l}}\,\eta\int\mathrm{d}^3y\,c(\mathbf{x}-\mathbf{y})f(\mathbf{y})e^{-i\mathbf{l}\cdot\mathbf{x}}, \tag{3.5}$$

where $B_\mathbf{l}^f$ means the transformed operator.

We shall make some remarks about the boson characteristic function $c(\mathbf{x}-\mathbf{y})$ and boson energy spectrum $\omega(\mathbf{l})$. Both functions can be computed exactly [12, 13] for all values of momentum, although they need numerical calculations. It turns out that for values $\mathbf{l}\cdot\xi_0>5$ (where ξ_0 is the coherence length of superconductor) $\omega_\mathbf{l}\approx 2\Delta$. The boson spectrum is also function of λ. $N(0)$. ($N(0)$ means the density of electron states at the Fermi surface). For the purpose of practical numerical computations we have fitted the boson spectrum by the function

$$y=-2e^{-\frac{\tau}{\lambda N(0)}}+2, \tag{3.6}$$

where

$$y=\frac{\omega_\downarrow}{\Delta};\qquad \tau=l\cdot\xi_0 \tag{3.7}$$

which is in good agreement with the exact curve for values of $\lambda\cdot N(0)\approx 0.4$ (Fig. 1) and the qualitatively correct dependence on $\lambda\cdot N(0)$ is achieved. For small values of l we

use in this Section also the approximation

$$\omega_{\mathbf{l}} = v_{\mathbf{l}} \cdot \mathbf{l}$$

$$v_{\mathbf{l}}^2 = \frac{v_F^2}{3}[1+\lambda N(0)]\,(1-2v_F^2 l^2/45\Delta^2) \tag{3.8}$$

given in [9], where v_F is the Fermi velocity. Theoretical computations of the boson characteristic function give for small values of momentum the dependence

$$c(l) = \frac{1}{1+\alpha\xi_0^2 l^2}, \tag{3.9}$$

where $\alpha = 2\pi^2/45$ [9]. For our purposes (space dependence) we need its Fourier transform, which gives

$$c(\mathbf{x}) = \frac{1}{4\pi\alpha\xi_0^2}\,\frac{1}{|\mathbf{x}|}\,e^{-\frac{|\mathbf{x}|}{\alpha\xi_0^2}}. \tag{3.10}$$

Now, let us go back to (3.3) and (2.12). After the boson transformation we have

$$\begin{aligned}\Delta(\mathbf{x}) = \{&\lambda\langle 0\,|\,\chi^{(1)}(\mathbf{x}-\mathbf{y})\,|\,0\rangle + \lambda\langle 0\,|\int\int \mathrm{d}^3p\,\mathrm{d}^3q[F^{(1)}(\mathbf{x},\mathbf{y},\mathbf{p},\mathbf{q})\cdot\\ &\cdot\alpha\mathbf{q}_\uparrow\,\alpha\mathbf{p}_\downarrow - F^{(2)*}(\mathbf{x},\mathbf{y},\mathbf{p},\mathbf{q})\alpha^+_{\mathbf{p}\downarrow}\,\alpha^+_{\mathbf{q}\uparrow}]\,|\,0\rangle + \lambda\langle 0\,|\int \mathrm{d}^3l\cdot\\ &\cdot[G^{(1)}(\mathbf{x},\mathbf{y},\mathbf{l})B_{\mathbf{l}}^f - G^{(2)*}(\mathbf{x},\mathbf{y},\mathbf{l})B_{\mathbf{l}}^{f\,+}]\,|\,0\rangle\}\,|_{\mathbf{x}=\mathbf{y}}.\end{aligned} \tag{3.11}$$

The vacuum state is that of $\alpha_{\uparrow,\downarrow}$ and B_l operators. Therefore, nonzero terms give only the first and the third term in (3.11). We shall use also the fact that this vacuum state is normalized. So we get the following expression:

$$\begin{aligned}\Delta(\mathbf{x}) = \{&\lambda\langle 0\,|\,\chi^{(1)}\,|\,0\rangle + \lambda\int \mathrm{d}^3l[G^{(1)}(\mathbf{x},\mathbf{y},\mathbf{l})\cdot\frac{1}{2}\int\frac{\mathrm{d}^3x'}{(2\pi)^{3/2}}\sqrt{2\omega_{\mathbf{l}}}\cdot\\ &\cdot\eta(\mathbf{l})\int \mathrm{d}^3y' c(\mathbf{x}-\mathbf{y}')f(\mathbf{y}')e^{-i\mathbf{l}\cdot\mathbf{x}'} - G^{(2)*}(\mathbf{x},\mathbf{y},\mathbf{l})\cdot\frac{1}{2}\int\frac{\mathrm{d}^3x'}{(2\pi)^{3/2}}\sqrt{2\omega_{\mathbf{l}}}\cdot\\ &\cdot\eta(\mathbf{l})\int \mathrm{d}^3y' c(\mathbf{x}'-\mathbf{y}')f(\mathbf{y}')e^{i\mathbf{l}\cdot\mathbf{x}'}]\}\,|_{\mathbf{y}=\mathbf{x}}.\end{aligned} \tag{3.12}$$

In this formula there are some quantities which have not been mentioned up to now. The function $\eta(\mathbf{l})$ appeared in the simplified constant form in (2.17) and (2.18). In general, it is dependent on the momentum $\mathbf{l}$. We shall use its approximate value (for small $\mathbf{l}$) – see e.g. [9]

$$\eta(\mathbf{l}) = 2\Delta\sqrt{\frac{2R_0}{\lambda}}\,\frac{\bar{\omega}_{\mathbf{l}}}{\omega_{\mathbf{l}}}, \tag{3.13}$$

where

$$R_0 = \frac{\lambda N(0)}{4\Delta^2} \tag{3.14}$$

and

$$\bar{\omega}_{\mathbf{l}}^2 = l^2 v_F^2/3(1-2v_F^2 l^2/45\Delta^2), \tag{3.15}$$

which again is an approximation. In [9] it is shown that if we derive the integral equation for $\chi^{(1)}$ and then solve it in the pair approximation, the first term in (3.12) gives the constant Δ, which is the solution of (2.10).

The most complicated in its analytical form are the Bethe – Salpeter amplitudes $G^{(1)}$ and $G^{(2)*}$. They depend on rational functions of complicated integrals through momentum space [11, 14]. The main trouble in computing of these integrals is the fact that integrations in momentum space are not translationally invariant. Physically, it is the consequence of the cutoff of electron – phonon interaction in the Hamiltonian (2.1) to those values of momentum which lie in the narrow region around the Fermi momentum, such that

$$||\mathbf{k}|^2 - |\mathbf{k}_F|^2|/2m < \omega_D,$$

where ω_D is the Debye frequency. However, if we consider only small values of momentum, we can assume the validity of this translational invariance. In [11, 14] it is shown that the analytical formulas for $G^{(1)}$ and $G^{(2)*}$ are much simplified. We shall give here only the final result for both G's:

$$G^{(1)}(\mathbf{l}) = -G^{(2)*}(\mathbf{l}), \tag{3.16}$$

$$G^{(1)}(\mathbf{l}) = \frac{\Delta\omega_{\mathbf{l}}}{\omega_{\mathbf{l}}^2 - \bar{\omega}_{\mathbf{l}}^2} g(\mathbf{l}), \tag{3.17}$$

$$g(\mathbf{l}) = 2i(2\pi)^{3/2}\Delta\sqrt{\frac{2R_0}{\lambda}}\frac{\bar{\omega}_{\mathbf{l}}}{\sqrt{2\omega_{\mathbf{l}}}} \tag{3.18}$$

and due to the translational invariance, $G(\mathbf{x}, \mathbf{x}, \mathbf{l})$ has the form

$$G(\mathbf{x}, \mathbf{x}, \mathbf{l}) = e^{i\mathbf{x}\cdot\mathbf{l}} G(\mathbf{l}). \tag{3.19}$$

Now we need only to know the functions $c(\mathbf{x}-\mathbf{y})$ and $f(\mathbf{x})$. As it was already pointed out, the characteristic function has some features of Dirac δ-function; so we consider it as the first approximation. Further, we assume a very simple solution of (2.25):

$$f(\mathbf{x}) = \kappa = \text{const.} \tag{3.20}$$

It means, that in our "inhomogeneous situation" (which has to be modelled by space dependence of the interaction constant λ) the phase of the superconductor does not change and remains constant.

After all these remarks made about the important quantities and functions we can compute (3.12). The result is surprisingly simple:

$$\Delta(\mathbf{x}) = \Delta + 4i\kappa\Delta, \tag{3.21}$$

which is a complex constant. The same result follows even if we consider the more general form (3.10) of $c(\mathbf{x})$. One would expect some λ-dependence of $\Delta(\mathbf{x})$ and, therefore, the possibility of applying various models of inhomogeneities, interfaces, etc.

4. Discussion

We may interpret the result (3.21) as follows: The absolute value of $\Delta(\mathbf{x})$ is

$$|\Delta(\mathbf{x})| = \Delta\sqrt{1+16\kappa^2}$$

which is identical with "homogeneous" solution. This is simply the well known energy gap of the Bardeen—Cooper—Schrieffer theory, i.e.

$$\Delta_{\mathrm{BCS}} = \Delta\sqrt{1+16\kappa^2}\,.$$

So, if we have any "model" space dependence of λ, this is exclusively given by Δ_{BCS} dependence, i.e. by the well known relation

$$\Delta_{\mathrm{BCS}} = 2\omega_D e^{-\frac{1}{\lambda N(0)}}.$$

However, the more exact numerical calculations of the amplitudes $G^{(1)}$, $G^{(2)}$ which are in progress show that even under the assumptions mentioned in the preceding Section the relation (3.16) does not hold. We may therefore expect really some space dependence in $\Delta(\mathbf{x})$,

$$\Delta(\mathbf{x}) = \Delta + h(\mathbf{x}),$$

where $h(\mathbf{x})$ is some space dependent "perturbation". Then we may compute e.g. the critical temperature of a given structure, which is the mean value:

$$\bar{\Delta}(T=0) \sim T_c\,.$$

Our result for $\Delta(\mathbf{x})$ may be also interpreted as some spatial mean value for the order parameter in the presence of the great inhomogeneity (compared with the coherence length ξ_0 of superconductor). This is the consequence of our approximations in the momentum space, mainly for the characteristic function.

Acknowledgement

I should like to thank Dr. Takács for critically reading the manuscript and many valuable comments.

References

1. A. E. Jacobs, J. Low Temp. Phys., **11,** 623, 1973.
2. L. Leplae and H. Umezawa, Nuovo Cim., **XLIV B,** 410, 1966.
3. R. Hagedron, Introduction to Field Theory and Dispersion Relations, Pergamon Press, New York, 1964.
4. L. Leplae, F. Mancini and H. Umezawa, Phys. Rev., **B 2,** 3594, 1970.
5. L. DeCesare and F. Mancini, Phys. Stat. Solidi (b) **501,**65, 1974.
6. L. Leplae, F. Mancini and H. Umezawa, Nuovo Cim., **9B,** 233, 1972.
7. F. Mancini, G. Scarpetta, V. Srinivasan and H. Umezawa, Phys. Rev., **9,** 130, 1974.
8. S. DeLillo and F. Mancini, Physica, **87A,** 391, 1977.
9. L. Leplae, H. Umezawa and F. Mancini, Phys. Rev., **10,** 151, 1974.
10. L. Leplae and H. Umezawa, J. Math. Phys., **10,** 2038, 1969.
11. F. Mancini, Physica, **77,** 311, 1974.
12. F. Mancini, Physica, **81B,** 119, 1976.
13. G. F. DeAngelis and F. Mancini, Physica, **77,** 332, 1974.
14. S. Takács and A. Pevala, Internal Report, Electrotechnical Institute, Slovak Academy of Sciences, Bratislava, 1980.
15. F. Mancini, Renormalization and Invariance in Quantum Field Theory, Plenum Publ. Corp., New York, 159, 1975.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 331—336 (1982)

INVESTIGATION OF THE INFLUENCE OF THE HEAT TREATMENT AND FORMING REGIMES ON THE CRITICAL CURRENT AND PINNING FORCE OF NbTi SUPERCONDUCTORS

J. HOELLRICH and G. KATH

VEB HALBZEUGWERK AUERHAMMER, AUE, GDR

The results obtained in the optimization of NbTi superconductors are set forth. The intermediate annealing temperature is discussed and the influence of final forming on the critical current density and the reduced pinning force is studied. For the types of conductors investigated (1, 19, 55-core conductors), the j_c-maximum occurs at different degrees of final forming; it drops as the filament diameter decreases. As to the reduced pinning force, an increase in the number of cores or a reduction of the filament diameters results in a displacement of the maximum and, for optimized conductors, an extension of the maximum towards higher fields is obtained. A final heat treatment makes the j_c-values drop.

1. Introduction

At present, NbTi conductors are the most universally and most frequently used superconductors in the construction of superconducting magnetic systems.

They are used for the construction of small laboratory magnets (with one to several kilograms of material used) as well as for the operation of bubble-chamber magnets (with the material used amounting to several tons).

Although NbTi conductors have lower critical values of T_c, H_{c2} and I_c as compared to the A15-conductors Nb_3Sn and V_3Ga, they are universally used because their production is less laborious and, therefore, more economical, and because their workability is better. The production technology for the manufacture of NbTi conductors though considered to be completed for years, still faces problems of application which necessitate further efforts to improve the NbTi conductors with regard to H_{c2}, I_c, degradation, radioresistance and stress. Investigations into the dependence of the current density and pinning force upon the magnetic field and the heat treatment and forming regimes were carried out on single- and multi-core conductors made by VEB Halbzeugwerk Auerhammer.

2. Optimization of NbTi superconductors

The NbTi superconductors produced consist of 52 percent by weight of Nb and 48 percent by weight of Ti. Their alloy is such that they lie within the optimum range for attaining high critical superconductor parameters [1], [2]. With this composition, H_{c2}

passes a maximum with T_c dropping but slightly. The critical current I_c and the critical current density J_c can be influenced considerably by metallurgical and production-technological measures.

The metallurgical measures, i.e. the influencing of j_c by varying the composition of the alloy will not be discussed in detail here. The deliberate incorporation of inhomogeneities in the NbTi structure allows to considerably increase the current-carrying capacity of the superconductor. For an NbTi alloy of the composition NbTi52/48, this suggests that the intercalation of intermediate heat treatments causes normally-conducting α-Ti-precipitations to be produced in the $(\alpha+\beta)$-NbTi-mixed crystal. These precipitations in combination with the dislocations caused by heavy cold forming represent effective pinning areas for the lines-of-flux lattice [3], [4], [5], the pinning force for the lines-of-flux lattice being determined by the precipitation density and distribution as well as by the particle size [6].

A plot of the critical current density j_c or the pinning force F_p against the magnetic field H applied will give information about the current-carrying capacity and stability of the superconductor. The $j_c = j_c(H)$ curves of optimized conductors with high critical current densities drop slightly, the pinning force F_p having its maximum at medium magnetic field strengths. With the degree of optimization increasing, this maximum is shifted towards higher field strengths.

3. Results

Measurements of the critical current intensity I_c were carried out at the Zentralinstitut für Festkörperphysik und Werkstoff-forschung Dresden (Central Institute of Solid State Physics and Materials Research of Dresden) using short-wire specimens bent to form a U shape. In this case, I_c correspond to the current where the voltage in the specimen drops by 0,1 μV/cm. The pinning force F_p and the reduced pinning forces were calculated from the $j_c = j_c(H)$ curves.

The specimens were taken from batches manufactured in the production process. They have been specially treated for the individual test programmes.

Table I

Material	Number of filaments	Cu/supercon-ductor relation	Filament diameter
Supra NT	1	1,0	0,354 mm
Supra NTM 19	19	1,6	0,071 mm
Supra NTF 55	55	1,2	0,045 mm

Table I gives the geometrical data for the specimens used. The diameter of the specimens is always 0,5 mm.

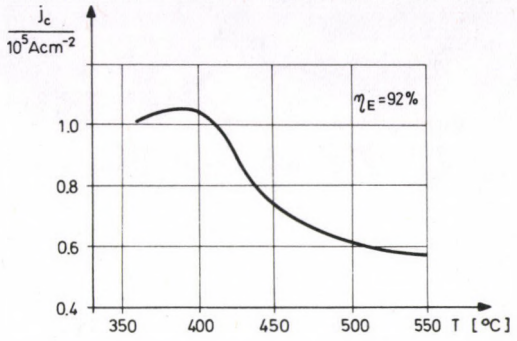


Fig. 1. Dependence of the critical current density j_c on the temperature of intermediate annealing $T(H = 4.0\ \mathrm{MAm^{-1}})$ Material: Supra NTM19 (0.5)

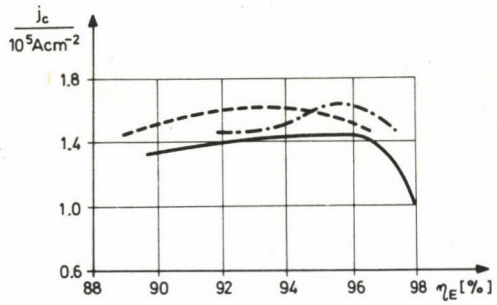


Fig. 2. Influence of the degree of final forming η_E on the critical current density j_c $(H = 4.0\ \mathrm{MAm^{-1}})$ Material: Supra NT (0.5) ——; Supra NTM19 (0.5) — · — · ; Supra NTF55 (0.5) ------

First, tests were conducted to determine the appropriate intermediate-annealing temperature. Fig. 1 shows the curve of $j_c = j_c(T)$. The j_c-maximum obtained is between 380 and 400 °C. On the basis of these data, all the wires used for further investigations were subjected to intermediate annealing at $T = 390$ °C for a time of $t = 3$ h.

Figs 2 to 4 give information about the influence of the heat treatment and forming regime on the j_c-value of different types of conductors in the end phase of conductor production. Fig. 2 shows the influence of the degree of final forming η_E on j_c. η_E, in this case, is shifted from 96% for the one-core conductor through 95.9% for the 19-core conductor to 93% for the 55-core conductor. This allows to conclude that the optimum degree of final forming is shifted towards somewhat smaller values which, however, are still in the range of $\eta_E > 90\%$, when the diameter of the filament decreases as a result of a stronger overall forming of the NbTi material.

Figs 3 and 4 are intended to show, on the basis of two test series, the dependence of the reduced pinning force $\frac{F_p}{F_{p\,\max}}$ on the reduced magnetic field $h = \frac{H}{H_{c2}}$ $(H_{c2} = 9{,}2\ \mathrm{MAm^{-1}})$. The difference in the heat treatment and forming regimes of the series A and B is shown in Table II.

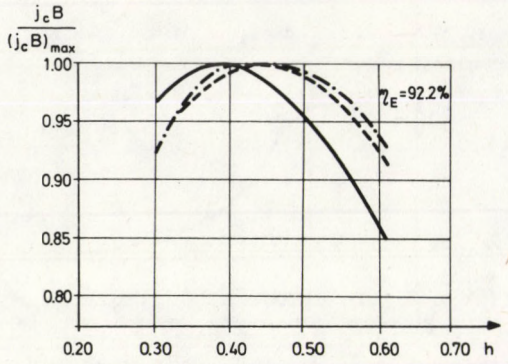


Fig. 3. Dependence of the reduced pinning force on the reduced magnetic field h
Material: Supra NT (0.5) ——; Supra NTM19 (0.5) — · — · ; Supra NTF55 (0.5) ------

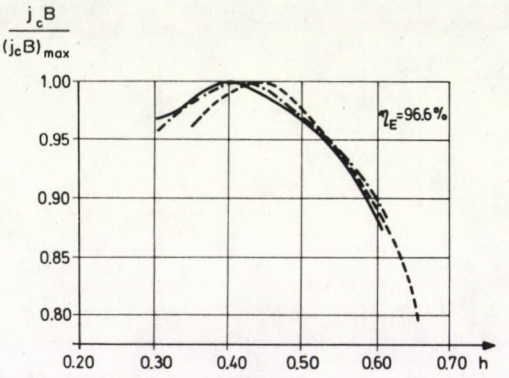


Fig. 4. Dependence of the reduced pinning force on the reduced magnetic field h
Material: Supra NT (0.5) ——; Supra NTM19 (0.5) — · — · ; Supra NTF55 (0.5) ------

Table II

	Degree of final forming	Number of intermediate annealing procedures
Series A	92.2%	6
Series B	96.6%	5

From Fig. 3 as well as Fig. 4, a displacement of the maximum of the reduced pinning force with respect to the reduced magnetic field for the different types of conductors can be seen. Thus, the maximum of $\frac{F_p}{F_{p\,max}}$ for the one-core conductor is at $h \approx 0.39$, that for the 19-core conductor at $h \approx 0.43$, and that for the 55-core conductor at $h \approx 0.46$. This means that an increase in the number of cores and, thus, the reduction of the filament diameter involved result in a displacement of the maximum towards higher fields.

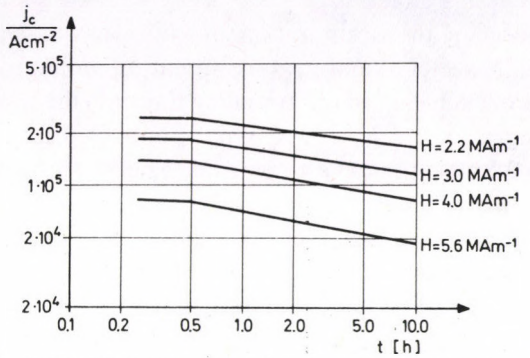


Fig. 5. Dependence of the critical current density j_c on a final heat treatment ($T=390$ °C)
Material: Supra NTM19 (0.5)

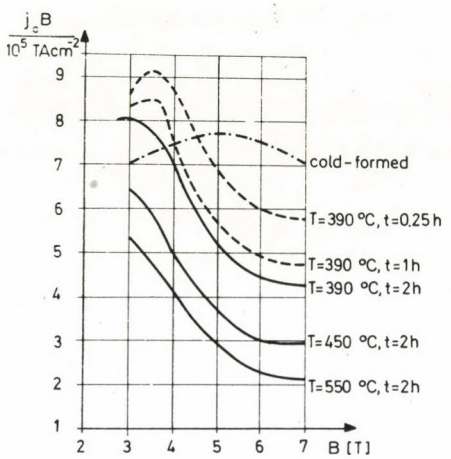


Fig. 6. Influence of final heat treatments on the pinning force
Material: Supra NTM19 (0.5)

Furthermore, it is typical that the curves of the specimens of series B are considerably steeper than those of series A. Towards the higher h-values of the series A a wide maximum is formed. These results must be considered in connection with the information given in Fig. 2.

The series B with $\eta_E=96.6\%$ for all types of conductors lies in the dropping portion of the $j_c=j_c(\eta_E)$ curve. The optimum degree of final forming and thus the optimum setting of the structure by precipitations and dislocations have already been exceeded. The series A with $\eta_E=92.2\%$ lies within the rising portion of the curve. Figs 3 and 4 demonstrate in terms of the extension of the maximum towards higher magnetic fields that the behaviour of the conductors of the series A is much more favourable. Thus, filament conductors in higher fields ($H=4.0$ MAm^{-1}) show a much more favourable behaviour than one-core conductors which are more suitable for us with lower fields ($H=4.0$ MAm^{-1}).

Another aspect of the optimization of NbTi superconductors is the determination of the influence of a final heat treatment on an improvement of j_c.

Fig. 5 shows the influence of the annealing time t on the critical current density j_c for a final annealing process at $T = 390$ °C. Up to about $t \leqq 0.5$ h, j_c drops slightly, while at times exceeding 0.5 h a continual dropping takes place. This is slightly contradictory to publications saying that final annealing makes the j_c-values rise [6], [8]. It should be mentioned that the specimens used there were merely cold-formed specimens whereas we tested conductors subjected to a relatively high number of intermediate annealing treatments. Our specimens have precipitations and dislocations in orders of magnitude that are necessary to obtain a high j_c-value. A final heat treatment of such optimized conductors does not yield a further increase in j_c. In each case, it is necessary to carry out a cold-forming operation of $\geqq 90\%$.

This becomes even more obvious when considering the pinning-force curves plotted against the magnetic field in Fig. 6. With the annealing time increasing, the maximum of F_p changes considerably both as regards its height as well as the dependence on the magnetic field. In addition to Fig. 5, the curves for specimens annealed at 450 °C and 550 °C are shown. The position of the maximum is shifted towards smaller magnetic fields and smaller pinning forces as the annealing time and temperature increase. This is in conformity with other publications [6], [5], [7].

The results of our attempts to optimize NbTi superconductors allow the manufacturers and users of semi-finished products to draw conclusions for their work.

References

1. H. Hillmann and K. J. Best, IEEE Trans. Magn. MAG-**13,** 1568, 1977.
2. P. R. Critchlow, E. Gregory and B. Zeitlin, Cryogenics, **11** (1), 3, 1971.
3. J. Pfeiffer and H. Hillmann, Acta Met., **16,** 1429, 1968.
4. H. Hillmann, Metal, **27,** 797, 1973.
5. H. Hillmann, Proc. of International Discussion Meeting on Flux Pinning in Superconductors, St. Andreasberg, 1974.
6. J. Willbrand, R. Arndt, R. Ebeling, Mohs, Concluding Report of BMFT, NT 106, 1974.
7. R. G. Hamshire and M. T. Taylor, taken from [4].
8. G. Rondelli and E. Olzi, Metallurgia ital., Milano **71,** 531, 1979.
9. B. P. Strauss, R. H. Remsbottom and P. J. Reardon, Applied Polymer Symposium No. 29, p. 53 – 59, 1976.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 337—345 (1982)

STATISTICAL THEORY OF PINNING IN SUPERCONDUCTORS INCLUDING CORRELATIONS IN THE DEFECT DISTRIBUTION

S. TAKÁCS

INSTITUTE OF ELECTRICAL ENGINEERING, SLOVAK ACADEMY OF SCIENCES
809 32 BRATISLAVA, CZECHOSLOVAKIA

The statistical theory of pinning on randomly distributed weak pinning centres with small densities is developed. In addition to the randomness in the spatial distribution of the defects, some correlation effects in the defect distribution are included. The results are in good agreement with experiments not only qualitatively (the volume pinning force proportional to the defect density), but also quantitatively.

1. Introduction

The basic principles of carrying the electric current in type II superconductors — the compensation of pinning forces against the Lorentz force on the flux lines — are known already for many years. In spite of this, a useful theory for explaining the experimental results sufficiently is lacking. The most difficult problem is connected with the proper summation of the individual forces from the pinning centres. The first "compact" theory in this direction was given by Labusch [1], who has calculated the volume pinning force resulting from randomly distributed point defects including the elastic properties of the flux line lattice.

His result was in the lattice approximation

$$F_p = n\, k_m^2\, C, \tag{1}$$

where n is the density of the pinning centres, k_m is the maximum interaction force exerted by the defects on the single flux lines and C is the effective elastic constant of the flux line lattice, containing the elastic moduli C_{ii}. The quadratic $F_p(k_m)$ dependence in (1) can also be seen from calculations with model interaction potentials [2, 3].

There is also some experimental evidence on the $F_p \sim k_m^2$ dependence [4 – 6]. The $F_p \sim n$ dependence is clearly visible from the results on Nb (see [7]). No determined $F_p(n)$ dependence was found in V_3Si at small defect densities [8].

At high defect densities [7], saturation effects appear for very high defect densities, only statistical fluctuations from the mean defect density act as pinning centres [9]. The second important result of Labusch's theory — the existence of the threshold value for the flux line distortion — is a more problematic task (see [7, 10, 11]). Nevertheless, it seemed that the theory of Labusch shows the right way for solving the complicated statistical summation problem in pinning. The "competing" theory of

Dew-Hughes [12], the direct summation of pinning forces (leading thus to the linear $F_p(k_m)$ dependence), was held to be right only for very strong pinning centres. Recently, very strong objections were raised against the basic assumptions and approximations of Labusch [7, 10, 11, 13]. Here are the main points.

The periodicity of the interaction potential defect-flux line was neglected in [1]. This changes mainly the $F_p(k_m)$ dependence appreciably (however, not so drastically at lower magnetic fields [14]).

The improper way to calculate the so-called Labusch parameter $\alpha_L = \langle \nabla^2 U \rangle$ for the pinning potential U was another shortcoming. If the averaging is performed properly, not only α_L, but also $F_p = \langle -\nabla U \rangle$ vanishes below the threshold.

The theoretical results of F_p with the known experimental values for k_m are below the threshold, i.e. F_p should be zero.

For larger specimens, the condition for the dilute limit does not seem to be satisfied, as the displacement field caused by one pin acts also on positions of the other pins.

Therefore, the threshold condition for the pinning forces cannot exist, either.

For some cases, the pinning force cannot be derived from a potential [15, 16, 17].

For strong (or even intermediate) pinning centres, plastic deformations seem to play an important role, which restricts the possibility of using linear elasticity theory, as was made usually in nearly all pinning theories.

Some attempts were made to improve the theory of Labusch, e.g. the more precise calculations of the elastic properties of the flux line lattice [18], pinning by flux bundles [19], etc., without achieving great success.

For the very unsatisfactory state of our present understanding of pinning in type II superconductors, some additional problems are characteristic, which are far from being solved at the present time (the peak effect, the differences in the magnetic field dependence of F_p near the upper critical magnetic field B_{c2}, the direct summation of pinning forces, the importance of the critical velocity model [20], mainly for A-15 superconductors in the vicinity of B_{c2}, etc.).

The theory of Larkin and Ovchinnikov [21] is based on somewhat different principles, which are explained in the next Section, as well as the generalized theory of the author [22] for small defect densities.

In this paper, we give a more exact theory of the influence of the "gathering" of defects in superconductors with small defect densities. This theory can explain, why — in spite of theoretical results that the dilute limit in pinning theories cannot exist — the volume pinning force is proportional to the mean defect density in the low concentration limit.

2. Calculation principles

The magnetic flux enters into homogeneous type II superconductors in the form of periodically arranged flux lines (vortices). The structure of the vortex lines is a hexagonal one, with known elastic constants $C_{11}, C_{12}, C_{44}, C_{66}$ [23]. Each defect in the crystal lattice interacts with the flux lines through a number of interaction mechanisms [24], leading to defects also in the flux line lattice (e.g. flux line dislocations [25]). If the number of the defects increases, the long-range order in the flux line lattice is loosed [26]. However, due to the strong nearest-neighbour and next-nearest-neighbour interactions between the flux lines, one can find some characteristic volumes (the so-called correlation volumes V_c) where the flux line lattice is nearly regular.

It is supposed [27] that these correlation volumes are moved independently under the influence of the applied Lorentz force. Due to the randomness of the defect positions in these volumes, the total force per volume unit is

$$F_p = k_m \frac{\sqrt{N}}{V_c}, \tag{2}$$

where k_m is the maximum force between single defects and single flux lines and N is the number of defects in the correlation volume V_c ($n = N/V_c$ is thus the defects concentration). Therefore, in the critical state

$$Bj_c = k_m \frac{\sqrt{N}}{V_c} = \left(\frac{n}{V_c}\right)^{1/2} k_m. \tag{3}$$

The determination of the correlation volume V_c in these equations is an open question. It was calculated in [28] from some energy considerations due to the elastic distortions of the flux line lattice:

$$V_c = \frac{256\, c_{44}^2 c_{66}^4 a^6}{n^3 k_m^6}. \tag{4}$$

With this value, we obtain:

$$F_p = k_m^4 \frac{n^2}{16 c_{44} c_{66} a^3}. \tag{5}$$

One can see from these results that the performed calculations must be undoubtedly wrong for several reasons. First, the value of V_c from (4) is too large, because e.g. for $n = 10^{-3}\,\xi^{-3}$ and ξ^{-3} (which is already an extreme case, as then all "parts" of flux lines are pinned!), we have $V_c = 10^{14}\,\xi^3$ and $10^5\,\xi^3$, respectively. Further, neither the strong density dependence ($\sim n^2$), nor the k_m^4 dependence of F_p has been (even approximately) detected in experiments. In addition, the correlation lengths in superconductors should be restricted by the largest characteristic length of the interactions which causes the pinning. This length in type II superconductors is the magnetic field penetration depth λ. Therefore, the maximum correlation length has to

be of the order of λ and all theories which give larger values for the correlation length are incorrect or incomplete.

There is also some experimental evidence for this: the maximum distance, on which the disturbances of the flux line lattice in type II superconductors could be detected, is about $p\lambda$ with $p \approx (5 \div 10)$ [27].

To calculate the correct value of the correlation volume, one needs the exact evaluation of the interaction range in disturbed flux line lattice [26, 28]. This evaluation is in addition complicated by the strong anisotropy of the problem, as the flux line distortions in the lattice are essentially two-dimensional, i.e. the distortions are not spread out in the flux line direction. Therefore, one could expect that the realistic value of V_c lies between $(p\lambda)^3$ and $\xi(p\lambda)^2$ for small defect densities.

The second problem is the calculation of the elementary interaction force. This can be performed relatively easily for defects (or defect structures) the characteristic dimension of which is smaller than the coherence length ξ. The interaction energy for a flux line portion of the length is then proportional to the number of defects in the volume ξ^3, because ξ is the length of the direct interaction range defect-flux line (this is true for the elastic pinning mechanisms, for the magnetic interaction mechanisms this lenght is near λ). By dividing the superconductor into such characteristic volumes (we take for simplicity cubes) $\approx \xi^3$, we define the "effective" pinning centres, in analogy with the pinning on statistically distributed dense defects [10].

Due to the statistical grouping of a larger number of defects in these volumes, we obtain an appreciable effect also in the low concentration limit [22], although the probabilities for finding two or more defects in $V_0 = \xi^3$ are very small.

We take into account also the effects corresponding to "clustering" the defects in the neighbouring volumes in the flux line direction. As calculated by Kramer [29], the point pins with mutual distance

$$l < l^* = \left(\frac{C_{44}\varphi_0}{C_{66}B}\right)^{1/2} \tag{6}$$

behave like line forces (φ_0 is the flux quantum). From experiments on Nb $l^* = M\xi$ with $M \approx (10 \div 18)$ for $b = B/B_{c2} \approx (0.2 \div 0.8)$. Therefore, if we have e.g. one defect in each of the neighbouring characteristic volumes V_0 in the flux line direction, we have the same elementary interaction force $2k_0$ as in the case of two defects in one volume V_0 (and no defects in the neighbouring volumes) [21]. The same is valid for more defects in volumes with distances smaller than l^*.

Due to the finite interaction range of two defects, we include an effective interaction length in the flux line direction $M_{\text{eff}} = h \approx M/2$.

As in all theories of pinning, we take the volumes with one, two, three, ... defects as independent pinning centres with elementary interaction forces k_0, $2k_0$, $3k_0$, ... Then, the elementary interaction force is given by the differences in the defect numbers in the neighbouring characteristic volumes V_0 in the Lorentz force direction [22].

We have then to introduce randomly L defects in $K = V_c/V_0$ positions in the superconductor ($\bar{n} = L/KV_0$ is then the mean defect density). As the defects are indistinguishable, one has to apply the Bose–Einstein statistics to obtain the defect distributions.

The main parameters are the probabilities $\psi_1, \psi_2, \psi_3, \ldots$ for finding one, two, three... defects in the given volume. They are given by

$$\psi_1 = \frac{S(K-1, L-1)}{S(K,L)} = \frac{L(K-1)}{(K+L-1)(K+L-2)}, \tag{7}$$

with

$$\psi_2 = \psi_1 \frac{L-1}{K+L-3},$$

$$\psi_3 = \psi_2 \frac{L-2}{K+L-4},$$

with

$$\vdots$$

$$S(K, L) = \frac{(K+L-1)!}{L!\,(K-1)!}. \tag{8}$$

The probability for a volume element to be empty is $(1-\psi)$, where

$$\psi = \psi_1 + \psi_2 + \psi_3 + \ldots + \psi_L \tag{9}$$

is the probability for a volume to be occupied by an arbitrary number of defects.

3. Determination of the volume pinning force

The probability for changing the defect number by $+1$ is then given by

$$g_1 = \psi_1 (1-\psi)(1-\psi)^{h-1} + \tag{10}$$

$$+ \psi_2 \psi_1 (1-\psi)^{h-1} + \psi_3 \psi_2 (1-\psi)^{h-2} + \ldots,$$

where we have already included that the $(h-1)$ volume elements in the flux line direction are not occupied by defects. The right hand side of the Eq. (10) gives only the half of the contribution to the pinning, because e.g. the combination "01" defect gives the same contribution as the combination "10" defects.

The determination of higher g_i's is somewhat tedious, because the occupation of larger number of defects in the flux line direction can be realized with many possibilities. We give only some results here:

$$g_2=\psi_2(1-\psi)^h+\psi_3\psi_1(1-\psi)^{h-1}+\psi_4\psi_2(1-\psi)^{h-2}+\ldots+\frac{h-1}{2}\psi_1^2(1-\psi)^h,$$

$$g_3=\psi_3(1-\psi)^h+\psi_4\psi_1(1-\psi)^{h-1}+\psi_5\psi_2(1-\psi)^{h-2}+\ldots$$
$$+(1-\psi)^h\left\{(h-1)\psi_1\psi_2+\frac{h-1}{2}\frac{h-2}{3}\psi_1^3\right\},$$
$$\vdots$$
$$g_6=\psi_6(1-\psi)^h+\psi_7\psi_1(1-\psi)^{h-1}+\psi_8\psi_2(1-\psi)^{h-2}+\ldots$$
$$+(1-\psi)^h\left\{(h-1)(\psi_1\psi_5+\psi_2\psi_4+\frac{1}{2}\psi_3^2)+\right.$$
$$+(h-1)(h-2)\left(\frac{1}{2}\psi_1^2\psi_4+\frac{1}{2.3}\psi_2^3\right)+$$
$$+\frac{(h-1)(h-2)(h-3)}{2.3}\left(\psi_1^2\psi_2^2+\psi_1^3\psi_3+\right.$$
$$\left.+6\psi_1\psi_2\psi_3+\frac{h-4}{4}\psi_1^4\psi_2\right)+$$
$$\left.+\frac{h-1}{2}\frac{h-2}{3}\frac{h-3}{4}\frac{h-4}{5}\frac{h-5}{6}\psi_1^6\right\}.$$

4. Results and discussion

The defect distribution (7) leads then to the mean volume pinning force [22]

$$\bar{F}_p=2k_0(V_0\,V_c)^{-1/2}g=\frac{2k_0g}{\sqrt{K}\,V_0},\tag{11}$$

with

$$g=\sqrt{g_1}\left(1+2\sqrt{\frac{g_2}{g_1}}+3\sqrt{\frac{g_3}{g_1}}+4\sqrt{\frac{g_4}{g_1}}+\ldots\right).$$

In this way, we have included different statistical effects into our considerations (the randomness of the defect positions with respect to the flux line positions, the randomness of the defect distributions in the characteristic volumes V_0, correlation effects in the defect densities of neighbouring volumes in the flux line direction).

As the results are very insensitive to changes of the K values (like in our previous theory [22]),we give here the results for $K=1\,000$. The function $g(L/K)$ is given in Figs 1 and 2 for different values of M_{eff}. We can see an extended linear region (mainly for smaller values of M_{eff}), as well as the saturation effects for larger L/K. However, our theory does not contain the possible influence from parameters which determine the

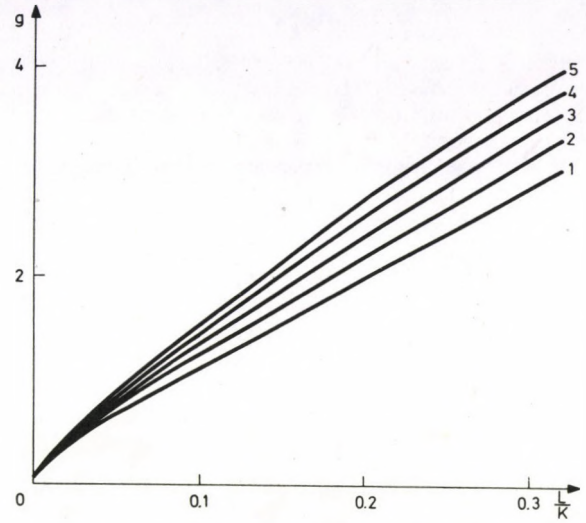


Fig. 1. The function $g \sim F_p$ vs $L/K \sim \bar{n}$ ($\bar{n}$ is the mean defect density) for $K \geqq 200$ and different values of M_{eff}.

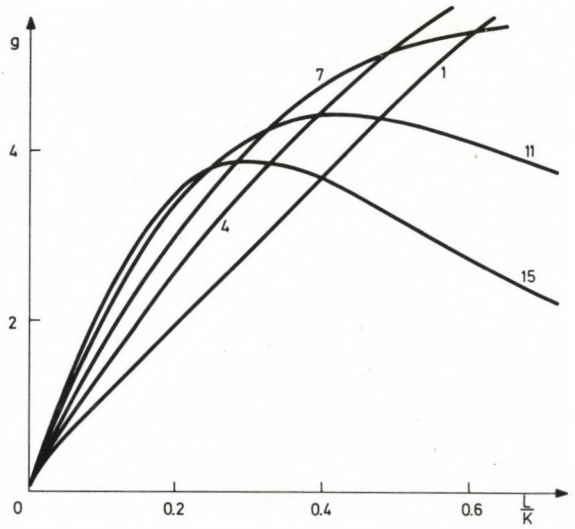


Fig. 2. The same function as in Fig. 1. The saturation effects for larger M_{eff} and L/K values are clearly visible.

elementary interaction force (e.g. $\bar{n}-l-B_{c2}-k_0$, $\bar{n}-l-T_c$, where l is the density dependent electron mean free path).

Although the $\bar{F}_p(\bar{n})$ dependence does not change with the K values considerably, the absolute values of F_p depend on K (and thus on the correlation volume V_c), as it can be seen from Table I. Here, the theoretical results are compared with the experimental

Table I

The quantitative comparison of the calculated volume pinning forces with the experimental ones. In the columns a) and b), the correlation volumes $\xi(p\lambda)^2$ and $(p\lambda)^3$ were assumed, respectively. The volumes in brackets are with $M_{\text{eff}}=7$, without brackets with $M_{\text{eff}}=1$.

Sample	mean defect density $\bar{n}$ $[\times 10^{22}\text{m}^{-3}]$	calculated volume pinning force		measured pinning force
		$F^{(a)}_{p,\text{theor}}$ $[\times 10^{8}\text{Nm}^{-3}]$	$F^{(b)}_{p,\text{theor}}$ $[\times 10^{8}\text{Nm}^{-3}]$	$F_{p,\text{exp}}$ $[\times 10^{8}\text{Nm}^{-3}]$
Nb [30] neutron irradiated *voids*	0.01	46(50)	13(14.5)	27
	0.035	32(45)	10(14)	60
	0.17	18(31)	6(10)	39
	3.8	5(4)	1.5(1.2)	6
Nb [31] Ni^+ irradiated *voids*	0.14	24(48)	9(12.5)	50
	0.34	18(27)	6.8(10)	25
	0.36	15(22)	5.7(8.5)	18
	3	8.7(7.5)	3.3(2.8)	12
Nb [32] neutron irradiated *dislocation loops*	0.7	1.7(2.5)	0.7(1)	0.62
V_3Si [7] neutron irradiated *dislocation loops*	0.02	8(10)	0.6(0.8)	13.5

ones, where the case a) corresponds to $K=(p\kappa)^2$ and the case b) to $K=(p\kappa)^3$ with $p=5$, respectively. The best agreement with the experiments in this Table is reached for the correlation volume $V_c=(p\kappa)^2$ and $M_{\text{eff}}=7$.

References

1. R. Labusch, Crystal Lattice Defects, **1,** 1, 1969.
2. J. Lowell, J. Phys., **F2,** 547, 1972.
3. S. Takács, Czech J. Phys., **B25,** 1155 1975.
4. G. Antesberger and H. Ullmaier, Phys. Rev. Lett., **35,** 59, 1975.
5. S. Takács, Phys. Stat. Solidi (a), **32,** 485, 1975.
6. H. C. Freyhardt and V. Graeger, J. Nucl. Mat., **72,** 34, 1978.
7. E. J. Kramer, J. Nucl. Mat., **72,** 5, 1978.
8. H. Küpfer and A. A. Manuel, Phys. Stat. Solidi (a), **54,** 153, 1979.
9. H. Ullmaier, K. Papastaikoudis, S. Takács and W. Schilling, Phys. Stat. Solidi, **41,** 671, 1970. S. Takács and H. Ullmaier, Phys. Stat. Solidi (a), **19,** K35, 1973.
10. T. Matsushita and K. Yamafuji, J. Phys. Soc. Japan, **47,** 1426, 1979.
11. E. H. Brandt, Phys. Lett., **77A,** 484, 1980.
12. D. Dew-Hughes, Phil. Mag., **30,** 293, 1974.
13. A. M. Campbell, Phil. Mag., **B37,** 149, 1978.

14. S. Takács, Czech. J. Phys., **B29,** 1064, 1979.
15. E. H. Brandt, Phil. Mag., **B37,** 293, 1978.
16. K. Martinás, Phys. Lett., **82A,** 369, 1981.
17. I. Kirschner and K. Martinás, J. Low Temp. Phys., **47,** 105, 1982.
18. R. Schmucker and E. H. Brandt, Phys. Stat. Solidi (b), **79,** 479, 1977.
19. R. Labusch, in Discussion to [7].
20. S. Takács, Phys. Lett., **46A,** 121, 1973; Czech. J. Phys., **B24,** 457, 1974.
21. A. I. Larkin and I. N. Ovchinnikov, J. Low Temp. Phys., **34,** 409, 1979.
22. S. Takács, The Summation of Pinning Forces in Superconductors with Small Defect Densities, Primärbericht, KfK Karlsruhe, 1980 and to be published in Phys. Stat. Solidi.
23. R. Labusch, Phys. Stat. Solidi, **32,** 439, 1969.
24. A. M. Campbell and J. E. Evetts, Adv. Phys., **21,** 199. 1972.
25. H. Träuble and U. Essmann, Phys. Stat. Solidi, **32,** 337, 1969.
26. A. I. Larkin, Zh. Eksp. Teor. Fiz., **58,** 1466, 1970.
27. H. Küpfer, private communication (to be published in Proc. LT 16).
28. E. H. Brandt, J. Low Temp. Phys., **28,** 263, 1977.
29. E. J. Kramer, J. Appl. Phys., **44,** 1360, 1973.
30. C. C. Koch, H. C. Freyhardt and J. O. Scarborough, IEEE Trans. Magn. MAG-**13,** 828, 1977.
31. H. C. Freyhardt, Radiation-Induced Flux Pinning in Type II Superconductors, Institut für Metallphysik, Universität Göttingen, 1976; see [7].
32. D. C. Agrawal, E. J. Kramer and B. A. Loomis, Phil. Mag., **33,** 343, 1976.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 347—351 (1982)

THE PINNING OF VORTEX LINES IN VERY DILUTE SUPERCONDUCTORS

I. KIRSCHNER and K. MARTINÁS

DEPARTMENT FOR LOW TEMPERATURE PHYSICS, ROLAND EÖTVÖS UNIVERSITY
BUDAPEST, HUNGARY

In a one dimensional model the possibility of synchronization of the flux line system to the randomly distributed pinning centres is shown. The result yields direct summation without threshold in the very dilute case.

Introduction

The determination of the total pinning force is a problem of considerable interest, still imperfectly understood.

The generally accepted assumptions about the flux line lattice and pinning centres are as follows [1, 2, 3]:

1. Non-interacting defects;
2. One defect in the interaction range of the flux line;
3. The defects are in randomly distributed positions;
4. The vortex lattice can be treated as an elastic continuum;
5. Thermal activation is negligible.

Generally, these conditions are accepted in the different ways of the calculation of pinning forces.

Determination of flux line pinnings

Lowell [4] introduced the following method for the calculation of pinning force. In linear approximation the elastic response is given by an effective elastic constant C. The equilibrium of forces on a single flux line can be obtained by

$$F_{\text{defect}} = F(x) = C \cdot (x - d), \tag{1}$$

where $F(x)$ is the pinning force, x is the equilibrium position of the flux line, and d is the position which would have the flux line without the presence of defect. For randomly distributed defects, all distances d of flux lines from the defects exist with equal probability [4, 5]. In such a way the volume pinning force is described by the integral of $F(d(x))$.

Campbell [6] calculated the volume pinning force in case of piecewise linear periodic interaction potential, when the interaction range equals the flux-line spacing. His result is

$$F_t = N_p \frac{(P-C)}{(P+C)}, \tag{2}$$

where F_t is the total pinning force density, N_p is the defect density and P is the maximum pinning force with what the centre can effect. This result leads to the threshold problem, as in case of $P<C$ there is no effective pinning force.

Takács [7] calculated the total pinning force for different interaction potentials, and has shown that there are differences in the dependence of the volume pinning force on the maximum elementary interaction forces, nevertheless the threshold problem is inherent.

The usual approach regards the flux lattice as a one-dimensional elastic continuum in which the local tension can be treated as a quantity proportional to the local strain. The elastic force on an element of the flux lattice of length dx is then

$$F_e = -C \frac{\partial^2 \xi}{\partial x^2} dx, \tag{3}$$

where ξ is the displacement at x from the undisturbed position.

This force must be balanced by the force due to the pinning centres in the element dx. This force in case of a harmonic potential can be written as

$$F_p = f_p \Theta(\omega/x) \cdot \omega(x) dx, \tag{4}$$

where $\omega(x) = \min \{x - x_\alpha\}$ and x_α is the position of the αth pinning centre, and

$$\Theta(\omega_\alpha) = \begin{cases} 1 & \text{if } |\omega_\alpha| < L, \\ 0 & \text{if } |\omega_\alpha| > L. \end{cases}$$

Here L is the interaction length of the potential and $f_p L$ is the maximal elementary pinning force.

The boundary condition contains the total transport current I_{tot} in the form of

$$\left[\frac{\partial \xi}{\partial x}\right]_l - \left[\frac{\partial \xi}{\partial x}\right]_0 \equiv \frac{I_{\text{tot}}}{B_0}. \tag{5}$$

where B_0 is the external magnetic field and l is the length of the sample.

The results based on the above assumptions are not adequate for the description of pinning phenomena. Recently, very strong objections were raised against these assumptions:

1. The threshold criterion is not verified by the experiments [6, 8].
2. In some cases the pinning force cannot be derived from a potential [9]. The summation for this case gives quadratic dependence in case of $\frac{P}{C} < 1$, and direct summation of $\frac{P}{C} > 1$ [10, 11].

Discrete model

In this paper we examine the flux line system in "discrete approximation" instead of the linear elasticity description. In the one dimensional model the flux line is characterized by the position vector r_i and the deformation force arises then, when the flux lattice spacing changes. Such change can occur only at the pinning centres, where the pinning force balances the deformation force.

In the discrete description of the flux line system the position of the ith flux line is

$$r_i = r_0 + \sum_{k=0}^{i-1} a_k,$$

where a_k is the distance between the $(k+1)^{\text{th}}$ and k^{th} flux lines.

The elastic force acting on the ith flux line in first neighbour approximation is

$C\,\delta\left(\frac{1}{a}\right)_i$

$$C\,\delta\left(\frac{1}{a}\right)_i = 2C\,\frac{r_{i+1}+r_{i-1}-2r_i}{r_{i+1}-r_{i-1}}. \tag{6}$$

In equilibrium

$$C\,\delta\left(\frac{1}{a}\right)_i = f_p(r_i), \tag{7}$$

where $f_p(r_i) = f_p\Theta(\omega_i)\ \omega_i$ and $\omega_i = \min\{r_i - x_\alpha\}$.

For the sake of simplicity we introduce the quantity ϑ_i:

$$\vartheta_i = \delta a_i = -\delta\left(\frac{1}{a}\right)_i \cdot a_{i-1}^2 = \frac{f_p(r_i)}{C}\, a_{i-1}^2, \tag{8}$$

by the help of which the position of the $(i+1)^{\text{th}}$ flux line is given in the form of a recursion formula:

$$r_{i+1} = r_i + a_i$$

and (9)

$$a_i = a_{i-1} + \vartheta_{i-1}.$$

To examine the structure of Eq. (6) we suppose that the ith vortex is pinned by the αth pinning centre, and the distance to the next pinning is $l_{\alpha 0}$.

The position of the ith vortex is r_i, while ω_i is the distance to the αth pinning, so

$$r_{i+k} = r_i + k \cdot a_i \tag{11}$$

and

$$a_i = a_{i-1} - \frac{1}{C}\, a_{i-1}^2 f_p \omega_i.$$

If $|r_{i+k} - x_{\alpha+1}| < L$ the vortex will be pinned and from this equation the value of k can be determined

$$k = \operatorname{Int}\frac{1+\omega_i}{a_i} \tag{12}$$

and

$$\omega_{i+k}=1-ka_{i-1}+\omega_i\left(1+\frac{1}{C}ka_{i-1}^2 f_p\right). \tag{13}$$

Changing the position of the ith flux line this causes a

$$\delta(\omega_{i+k})=\left(1+ka_{i-1}^2\frac{f_p}{C}\right)\delta\omega_i \tag{14}$$

change in the position of the $(i+k)^{\text{th}}$ flux line.

The relation (14) means, that there is a possibility for both of the ith and $(i+k)^{\text{th}}$ flux lines to be in a bounding position in case of randomly distributed centres, if $\delta\omega_i$ yields a change in bounding positions of the αth pinning results in a $\delta(\omega_{i+k})=a_i$. The distance of pinning centres in the above case has to be larger than

$$L_0=\frac{C}{f_p}L\,. \tag{15}$$

The physical meaning of this is that in case if the pinning density N_p is less than

$$N_{p0}=\frac{1}{L_0}=\frac{f_pL}{C} \tag{16}$$

every pinning centre can be active. Assuming a harmonic potential the total pinning force can be expressed by the relation of

$$F_p=f_p\left(1-\frac{N_p}{2N_{p0}}\right)N_p \tag{17}$$

in the very dilute limiting case ($N_p<N_{p0}$).

Conclusions

Taking into account the discrete nature of the flux lines there is a possibility for self-adjustment of flux line lattice to the randomly spaced pinning centres in the very dilute case

The results give direct summation in the extreme dilute case ($N_p\ll N_{p0}$), and linear dependence on the individual pinning forces and $\left(1-\frac{N_p}{2N_{p0}}\right)N_p$ dependence on the pinning density in the $N_p<N_{p0}$ domain.

References

1. A. M. Campbell and J. E. Evetts, Adv. Phys., **21,** 199, 1972.
2. K. Yamafuji and F. Irie, Phys. Lett., **A25,** 387, 1967.
3. R. Labusch, Crystal Lattice Defects, **1,** 1, 1969.
4. J. Lowell, J. Phys.,**F2,** 547, 1972.
5. S. Takács, Phys. Stat. Solidi, **A32,** 485, 1975.
6. A. M. Campbell, Phil. Mag., **B37,** 149, 1978.
7. S. Takács, Czech. J. Phys., **B29,** 1064, 1979.
8. E. J. Kramer, J. Nucl. Mat., **72,** 5, 1978.
9. E. H.Brandt, Phys. Lett., **77A,** 484, 1980.
10. K. Martinás, Phys. Lett., **82A,** 369, 1981.
11. I. Kirschner and K. Martinás, J. Low Temp. Phys., **47,** 105, 1982.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 353—360 (1982)

EXPERIMENTAL DETERMINATION OF PINNING FORCES

J. BÁNKUTI and É. FARKAS

DEPARTMENT FOR LOW TEMPERATURE PHYSICS, ROLAND EÖTVÖS UNIVERSITY
BUDAPEST, HUNGARY

Four different methods are given for the determination of the pinning forces: critical current measurements, magnetization measurements and two A. C. methods. The measured data of the critical current density obtained by these methods are published here for the samples Nb-10 at % Ti.

Introduction

From the industrial point of view one of the most important characteristics of superconductors is the critical current.

In a type I superconductor the critical current J_c can be determined by Silsbee's rule, according to which the thermodynamical critical field determines J_c definitely, and so the critical current is very small. In type II superconductors the situation is more complicated. This can be seen from the comparison of the magnetization curves of the two different kinds of superconductors in Fig. 1.

The initial parts of the curves are the same and the superconductors are in a so-called Meissner state. The essential difference between the two materials appears only in fields larger than the lower critical field. In this range type II superconductors are in mixed state, that is the magnetic field partially penetrates into the sample in the form of flux vortices.

A vortex consists of a normal core and a circular supercurrent. If any external current passes through a type II superconductor being in mixed state the current-vortex interaction can be described with a Lorentz type force [1]. The form of this driving force F_D per unit volume of the superconductor is as follows:

$$F_D = J \times B, \tag{1}$$

where J and B are the transport current density and the magnetic induction, respectively. Unless there are no other forces, the flux lines will move in the direction of this force and an electrical field E is induced in the normal core given by the following expression:

$$E = v \times B, \tag{2}$$

where v is the velocity of the flux lines. This electrical field induces a current in the normal core and a dissipation occurs. Thus the critical current equals zero in this case.

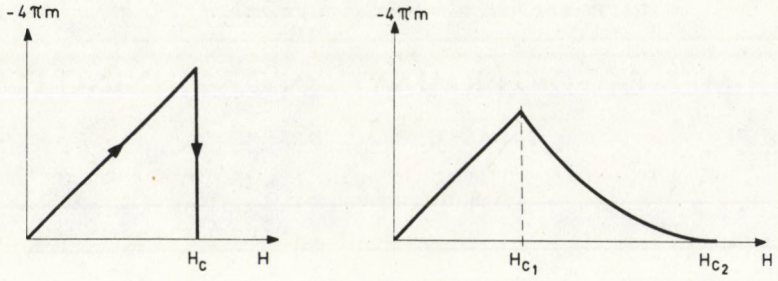


Fig. 1

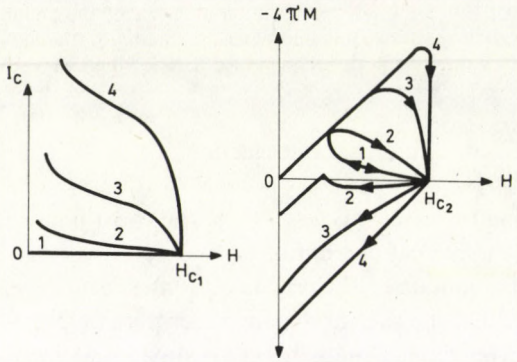


Fig. 2

Fortunately, there is no such an "ideal" type II superconductor. Materials always contain inhomogeneities, lattice defects and other imperfections. These so-called pinning centres prevent the flux lattice from moving. In Fig. 2 some magnetization curves and the corresponding critical current curves are plotted against the external field for some superconductors. Materials with the most irreversible magnetization curves produce the largest critical currents.

The pinning centres mean local energy minimum for flux lines and there are pinning forces on the flux lines to oppose the driving force. It is evident from the existence of the critical current that the pinning force has a maximum value. This maximum pinning force can be determined on the basis of the critical state model introduced by Bean [2].

It is assumed that the superconductor either carries the critical current density or no current at all. Flux vortices are nucleated at the surface and move into the sample until the driving force due to the flux density gradient is balanced by pinning forces. The flux density is usually assumed to possess its equilibrium value defined by a reversible magnetization curve at the surface. The local critical current is related to the maximum of average pinning force F_p, which will be a function of the local value of the induction

B. The critical state is represented by the force equilibrium:

$$J_c \times B = -F_p(B), \tag{3}$$

where pinning force $F_p(B)$ is related to unit volume of the superconductor. If any surplus force appears, for example the transport current increases, the flux lines rearrange until equilibrium is restored. Assuming a dependence for the critical current or pinning force upon the magnetic induction and integrating the critical state equation with an appropriate boundary condition, the flux density distribution throughout the specimen and so the magnetization can be calculated. The relation between the maximum pinning force and the magnetic induction inferred from the measured magnetization or critical current curves can be compared with the theoretical prediction of the given pinning model. Campbell [3] concluded a general form of the dependence of maximum pinning force on magnetic induction from earlier experimental results and their theoretical discussions. He found the following expression:

$$F_p(B) = k\, b^m (l-b)^1, \tag{4}$$

where $b = B/B_{c2}$ is the reduced magnetic induction, m and l are positive numbers. Fig. 3 shows the previous expression for various values of m, l. It can be seen that m and l affect the general shape of these curves.

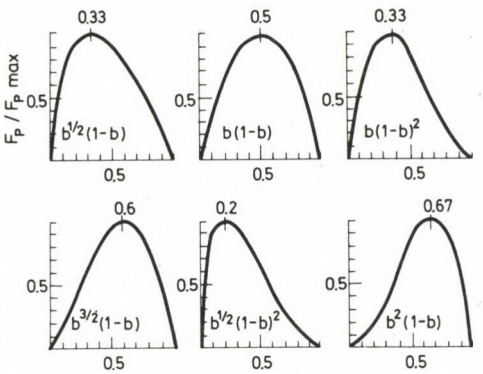


Fig. 3

The place of maximum characterizes the type of defect causing pinning. For many materials containing second phases or plane boundaries the maximum occurs at about $b = \frac{1}{3}$, for dislocation pinning the place of the maximum can vary between $b = 0.25$ and $b = 0.85$, depending on the deformation structure.

Measurements of pinning forces

Critical current measurement

Measurement of the pinning force in most cases is performed in an indirect way, and the evaluation of the experimental data is accomplished on the basis of the critical state equation.

One of these methods is the measurement of the critical current. Now we examine the critical state equation in the special configuration when the magnetic induction is perpendicular to the current. Then the form of the equation is very simple:

$$J_c(B) \cdot B = F_p(B) . \tag{5}$$

The critical current is usually determined by the increase of the transport current passing through the wire or foil sample until a certain small voltage (typically 1 μV) appears along the specimen in a certain external transverse field. This measurement gives correct results only when the flux density and hence $F_p(B)$ do not vary significantly in the cross section of the sample. This condition is fulfilled for specimens of a small cross section and high κ-value and at an external field not too close to the lower critical field. In this case the magnetization is very small and the induction can be easily determined. For a low external field this method cannot be applied since the distortion of the external field due to the self-field of the current alters the simple connection between the current and the induction. It is very difficult to solve the general critical state equation even with a precise boundary condition.

Measurement of the magnetization curves

The other frequently used method is the determination of the magnetization curve. When the critical current measurements become very impractical and inaccurate they are replaced by magnetization measurements. For low fields samples with large cross section present many problems. It is too difficult to produce and control the required large currents and to feed them into the cryostat. All these difficulties can be avoided by measuring the magnetization of a cylindrical sample. The sample axis is parallel to the external field. The external transport current is replaced by "induced" circular currents for which the critical state equation with the cylindrical symmetry can be applied. We may use the following form of the critical state equation:

$$\text{rot}\, H(B) \times B = -F_p(B), \tag{6}$$

where $H(B)$ is thermodynamically determined. In the case of the cylindrical symmetry the following simple form is valid:

$$B \frac{dH}{dr} = F_p(B). \tag{7}$$

Rewriting it we get

$$B\left(\frac{dH}{dB}\right)_{rev}\frac{dB}{dr}=F_p(B), \tag{8}$$

where $\left(\frac{dH}{dB}\right)_{rev}$ is the slope of the $H(B)$ reversible curve. If we assume a concrete dependence of $F_p(B)$ and know the $H(B)$ reversible curve we can calculate the flux profile and the magnetization by the following expression:

$$M(H)=\frac{1}{r^2\pi}\int_{r(B'=0)}^{r(B'=B)} B'(r')2\pi r'\,dr'-\mu_0 H. \tag{9}$$

We can measure the magnetization by a coil wound around the specimen to pick up the voltage induced by flux entering. We integrate the voltage-time function electronically. A more accurate result can be reached by the cancellation of the flux in the gap between the specimen and the search coil with a compensating coil. Comparing the measured results with the calculated ones the dependence of pinning force on induction can be determined. The problem of this method is that it is difficult to determine the reversible magnetization curve for very irreversible materials.

The situation is very much simplified at a small hysteresis of the magnetization curves, i.e. when F_p in the sample is small. In this case the relative difference between the maximum and the minimum flux density in the sample is also small and we may assume the pinning force density to be constant. Using the cylindrical symmetry the critical state equation leads to the expression:

$$F_p=\frac{3}{2}\frac{B}{R}\left(\frac{dH}{dB}\right)_{rev}[<M(H\uparrow)>-<M(H\downarrow)>], \tag{10}$$

where the $<M(H\uparrow)>$ and $<M(H\downarrow)>$ symbols are the average magnetization in an increasing and a decreasing field, respectively. The slope of the reversible magnetization curve can be obtained from the relation:

$$B_{rev}(H)=\frac{1}{2}[<M(H\uparrow)>+<M(H\downarrow)>]+\mu_0 H. \tag{11}$$

Another method with the same experimental configuration but much more simple is a tube sample with thin wall where we may assume the flux density gradient to be constant within the material. In this case the pinning force per unit volume is given by

$$F_p=BJ_c=B_m\frac{H^0-H^i}{\Delta r}, \tag{12}$$

where Δr is the wall thickness of the tube, H^0, and H^i are the values of magnetic field at the outer and the inner surface of the tube measured by a small Hall probe, and B_m is the mean value of the induction in the sample calculated by the equation:

$$B_m = \mu_0 \frac{H^0 + H^i}{2}. \tag{13}$$

A. C. methods

The most powerful methods used in determining pinning forces are the A. C. techniques. The experimental arrangement is the same as the method used in magnetization measurements with a compensating coil. The external field is a large D. C. field on which a small amplitude A. C. field is superimposed. Several A. C. techniques differ from each other in the manner of obtaining the information from the induced voltage.

In the so-called Campbell method the voltage is measured by a phase sensitive detector. When the phase is set correctly the detector output S is proportional to the integral of the input signal over a half-cycle of the reference waveform:

$$S(h_0) = \int_0^{T/2} U(t)dt = kN[\Phi(h_0) - \Phi(-H_0)], \tag{14}$$

where $U(t)$ is the induced voltage, k is a constant depending on the sensitivity of the apparatus, N is the number of the turns of pick-up coil and h_0 is the amplitude of the small A. C. field. The only assumptions are that the critical current density is constant over the amplitude of A.C. signal and the flux profile is cylindrically symmetric, but J_c may vary with the distance measured from the axis. We slightly increase the amplitude of the A.C. field increasing induction inside the specimen by δb_0 at all points in critical state. When the critical state reaches a radius r in the specimen with radius R, the increase $d\Phi$ of the flux is

$$d\Phi = \pi(R^2 - r^2)\delta b_0 = \mu_0 \pi(R^2 - r^2)\delta h_0. \tag{15}$$

Rewriting this we obtain

$$\frac{d\Phi}{dh_0} = \mu_0 \pi(R^2 - r^2). \tag{16}$$

Hence by differentiating the amplifier output with respect to the amplitude of the A.C. field, we obtain the penetration distance of the field b_0, thus we have a flux density distribution function. The gradient of this flux density profile gives the critical current at any point in the specimen and it is easy to calculate the average pinning force. We must use the calculation technique mentioned in the Section on magnetization curves.

The other A.C. method is the harmonic analysis of the voltage-time function. Assuming the validity of the critical state model, and that the critical current density

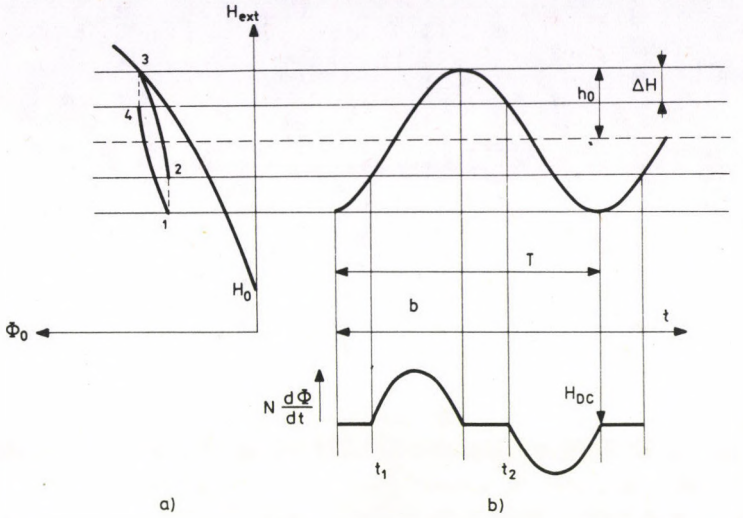


Fig. 4

does not vary over the cross section of the specimen, Bean [4] derived a relationship between the odd harmonics and the critical current density as a function of the applied D.C. field. Some measurements [5], however, showed that the application of Bean's formulas lead to much higher critical current density than it can be expected from other measurements.

Theoretical considerations [6, 7] show that in type II superconductors there is a thin vortex free surface sheath which is capable of carrying a supercurrent with the effect that the external field can be changed within certain limits ΔH without influencing the bulk flux distribution. Ullmaier [8] has derived modified expression for the relationship between J_c, h_0, ΔH and the Fourier coefficients of the voltage-time function. If the A.C. field has a sinusoidal time dependence the voltage-time function has the following form:

$$0 \leqq t \leqq t_1, \qquad U = 0, \tag{17a}$$

$$t_1 \leqq t \leqq T/2, \qquad U = -\frac{2\pi^2 R \mu_0 N h_0^2}{J_c T}\left(1 - \cos\frac{2\pi t}{T} - \frac{\Delta H}{h_0}\right)\sin\frac{2\pi t}{T}, \tag{17b}$$

$$T/2 \leqq t \leqq t_2, \qquad U = 0, \tag{17c}$$

$$t_2 \leqq t \leqq T, \qquad U = -\frac{2\pi^2 R \mu_0 N h_0^2}{J_c T}\left(1 + \cos\frac{2\pi t}{T} - \frac{\Delta H}{h_0}\right)\sin\frac{2\pi t}{T}, \tag{17d}$$

where R is the radius of the sample, N is the number of turns of the pick up coil, h_0 is the amplitude of the small A.C. field. The shape of this voltage is schematically shown in Fig. 4.

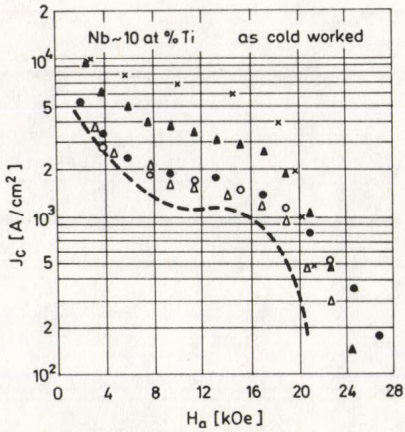


Fig. 5 ○ Campbell method; ● harmonic analysis; ▲ magnetization; --- magnetization corrected for surface barrier; △ magnetization after etching to remove surface barrier; × magnetization (heavily cold worked).

Results and conclusions

The methods described above are based on several assumption. These assumptions sometimes complete, sometimes exclude each other and sometimes they are identical for the various kinds of materials. For the sake of obtaining really good results it would be desirable to make measurements by more than one method.

Fig. 5 shows results interpreted from the different measurements for Nb ~ 10% Ti material [9, 10].

It can be seen that the critical current measurements are in good agreement with the A.C. methods, but critical currents calculated from the magnetization measurements are higher than the currents obtained from other methods at lower fields. If we take into account the surface shielding in the interpretation of the magnetization measurements, the agreement is better than in the previous case.

References

1. A. G. van Vijfeijken, Phil. Res. Rep. Suppl., 1968. (8)
2. C. P. Bean, Phys. Rev. Lett., **8,** 250, 1962.
3. A. M. Campbell and J. E. Evetts, Critical Current in Superconductors, Taylor and Francis Ltd., London, 1972.
4. C. P. Bean, Rev. Mod. Phys., **36,** 31, 1964.
5. M. Cardona, J. Gittleman and B. Rosenblum, Phys. Lett., **17,** 92, 1965.
6. H. J. Fink, Phys. Rev. Lett., **14,** 309, 1965.
7. J. G. Park, Phys. Rev. Lett., **15,** 352, 1965.
8. H. Ullmaier, Phys. Stat. Solidi, **17,** 631, 1966.
9. W. A. Fietz and W. W. Webb, Phys. Rev., **178,** 657, 1969.
10. D. M. Kroeger; C. C. Koch and J. P. Charlesworth, J. Low Temp. Phys., **19,** 493, 1975.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 361—370 (1982)

ON THE EXACT CALCULATION OF THE SPECIFIC ELECTRICAL CONDUCTIVITY OF SEMICONDUCTING SLABS

P. DEBRECENI, I. KIRSCHNER, M. F. KOTKATA, G. KOVÁCS and T. PORJESZ

DEPARTMENT FOR LOW TEMPERATURE PHYSICS, ROLAND EÖTVÖS UNIVERSITY
BUDAPEST, HUNGARY

A model is proposed for the exact calculation of the specific electrical conductivity of a semiconducting slab between two circular metallic electrodes. The assumed mixed boundary conditions and the derived dual integral equations lead to a Fredholm integral equation of the second type for infinite slabs. The latter is solved numerically and the results are discussed in terms of a correction function plotted against the ratio of the slab thickness to the electrode radius in the range of 0.01 – 10. A resistor network has been used to verify the calculations. A good agreement has been found with previously published results in some range of thickness-radius function.

Introduction

The specific conductivity of semiconducting materials gives a large amount of information about the band structure, the transport parameters and the properties of the materials in electric and magnetic fields. Moreover, the knowledge of the specific conductivity has great importance when considering amorphous semiconductors [! – 4], since methods like measuring the Hall effect cannot be easily applied. This is due mainly to the high resistivity (10^{14}ohm. cm) and the low mobility [10^{-7} cm^2 V^{-1} s^{-1}] of these materials, see e.g. Mott and Davis [1].

Precautions should be taken when choosing a measuring arrangement to determine the specific conductivity through resistivity measurements. Further calculations are made on the assumption that there is no contact resistance arising at the metal – semiconductor interface. If an unsuitable sample geometry is chosen it will be difficult to find a mathematical relation between specific conductivity and spreading resistance [5 – 10]. Calculations of the thermal and electric conductivity coefficients σ found in numerous research works accepted the general expression

$$\sigma = \frac{1}{R} \cdot \frac{W}{F} \tag{1}$$

in a simple way regardless of all the sample volume. That is, σ is calculated from the measured spreading resistance R according to the arrangement shown in Fig. 1, with the sample thickness w. It is known, however, that the current lines cover a bigger area than the space determined by the contact area as schematically illustrated in Fig. 2.

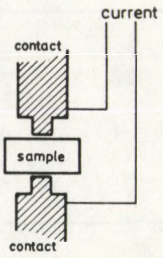


Fig. 1. Measuring arrangement for the resistivity of a semiconducting slab situated between two coaxial metallic cylinders.

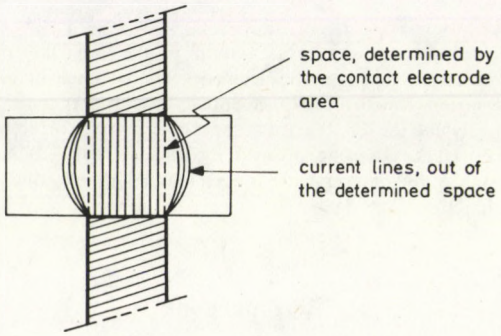


Fig. 2. Demonstration of current lines crossing a semiconducting slab.

This means that Eq. (1) would be satisfied under the condition that the contact area F is equal to the sample interface area. On the other hand, from an experimental point of view, the surface currents on the side walls of the sample would affect the results considerably.

For circular electrodes with contact radius a, Eq. (1) takes the form:

$$a R \sigma = \Pi \frac{W}{a}. \tag{2}$$

Because of the non-fulfilment of the validity conditions for Eq. (1), a correction function, depending on the thickness to contact radius ratio, should be taken into account. Also, when a point contact is used, i.e. $a+0$, Eq. (2) tends to

$$4aR\sigma = 1, \tag{3}$$

which is a well-known result. Such an arrangement can lead to false results as well. This is due to the fact that the possible high current density causes charge carrier injection [11, 12].

In this report, an exact calculation of the specific conductivity for infinite semiconducting slab, with the arrangement of Fig. 1, is presented. A resistor network was used to verify the results.

Description of the model

a) The calculation of conductivity

The semiconductor arrangement is presumed to be homogeneous and isotropic material, i.e. it is either uniformly dopped amorphous material or a single crystal with spherical band structure. Owing to the symmetrical arrangement (Fig. 1) we can use cylindrical coordinates. So, from Maxwell's equations; div $j=0$. Using Ohm's law; $j=\sigma E$, we can get the Laplace equation in the form $\nabla^2 \Phi(r)=0$, where Φ is the potential function of the system, j is the current density, and E is the electric field. Accordingly,

$$\frac{U_0}{R}=\sigma \int_F \text{grad } \Phi \, df, \tag{4}$$

where U_0 is the voltage between the electrodes.

b) The boundary conditions

To solve Eq. (4) for obtaining the relationship between the specific conductivity (σ) and the measured spreading resistance (R), the potential distribution from the Laplace equation should be found. However, the complexity of the mathematical problem encountered in the potential function calculation can be reduced by the aid of the elctrical reflection principle having a mirror plane, (Fig. 3). That is, taking into account only the half space.

In the model shown in Fig. 3, it is considered that the horizontal plane in the middle of the infinite semiconducting slab has a constant potential. This arrangement can be replaced — in principle — by two equivalent slabs both of the thickness (w) and

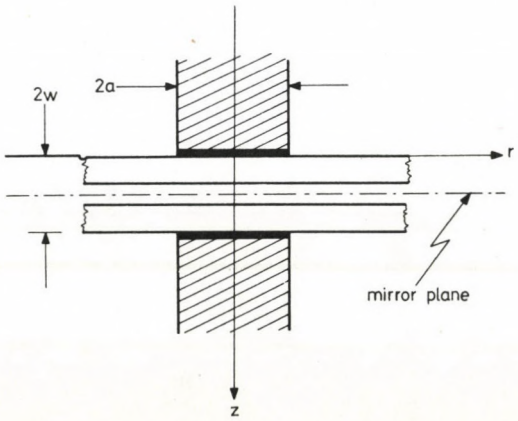


Fig. 3. A mirror plane for two-dimensional representation of the measuring arrangement.

entirely covered with metal contacts: an infinite one at the mirror plane and a finite one of radius (a). For such a model, the following boundary conditions can be introduced:

I — the metal plane is of zero potential and has infinite dimensions, i.e. $\Phi(r, w)=0$;

II — the disc potential is U_0, i.e. $\Phi(r, 0)=U_0$, and,

III — the normal component of the current density vanishes at the surface of the semiconductor not covered with metal, i.e.

$$\left.\frac{\partial \Phi}{\partial Z}\right|_{z=0}=0, \quad r>a\,.$$

c) Consequences of the potential theory

The solution of the Laplace equation for a system having cylindrical symmetry is expected to be of the following form [13].

$$\Phi=\int_0^\infty [A(k)\,e^{-kz}+B(k)\,e^{kz}]J_0(kz)\,dk, \tag{5}$$

where J_0 is the Bessel function of zero order. From (I),

$$B(k)=-A(k)\,e^{-2wk}. \tag{6}$$

On considering the two other boundary conditions (II) and (III), we can write the following Fredholm integral equations

$$\int_0^\infty A(k)[1-e^{-2wk}]J_0(kz)\,dk=U_0, \qquad r\leqq a, \tag{7}$$

and

$$\int_0^\infty kA(k)[1+e^{2wk}]J_0(kZ)\,dk=0, \qquad r>a. \tag{8}$$

Using the following new variables:

$$\begin{aligned} x&=r/2\\ u&=ka,\\ w&=w/a \end{aligned} \tag{9}$$

Eqs. (7) and (8) take the forms:

$$\int_0^\infty A_0(u)[1+f(u)]\frac{1}{u}J_0(ux)\,du=aU_0, \qquad x\leqq 1 \tag{10}$$

and

$$\int_0^\infty A_0(u)J_0(ux)\,du=0, \qquad x>1, \tag{11}$$

where

$$A_0(u)=A(w/a)\,[1+e^{-2uw}]u, \tag{12}$$

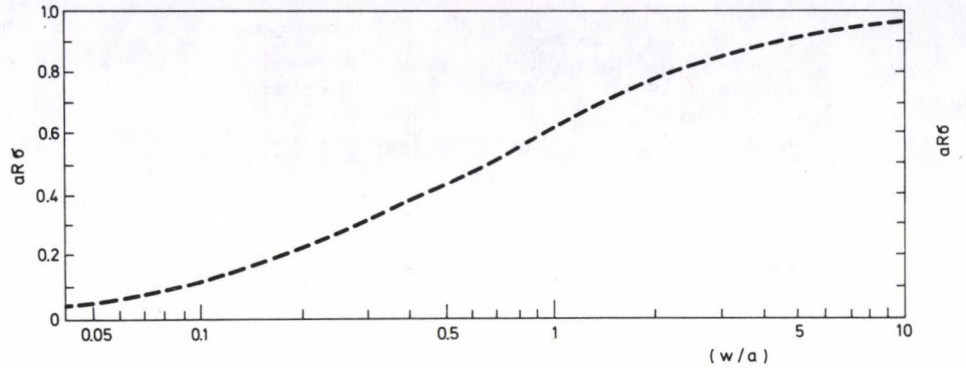


Fig. 4. Dependence of the calculated correction function aR on the sample arrangement represented as a ratio of (w/a). (w) is the slab thickness, and (a) is the contact electrode radius.

and

$$f(u)=\frac{-2\,\mathrm{Exp}\,(-2uw)}{1+2\,\mathrm{Exp}\,(-2uw)}. \tag{13}$$

The solution of the above dual linear and homogeneous integral equations is

$$A_0(u)=aU_0\frac{f(u)}{\Pi}u\int_0^1 S_0(t)\cos(ut)\,dt. \tag{14}$$

For $S_0(t)$, we have

$$S(x)+\int_0^1 S_0(u)\,K(x,u)\,du=1, \tag{15}$$

where the Kernel function is

$$K(x,u)=\frac{2}{\Pi}\int_0^\infty K(t)\cos(ut)\cos(xt)\,dt. \tag{16}$$

These equations can be solved numerically using a computer program. The obtained results are represented in Fig. 4, where a correction function is plotted against the ratio of sample thickness (w) to contact radius (a). It is concluded that: using the general expression for (σ) the deviation from the correct value can be expected to be as high as one order of magnitude, depending on the sample arrangement.

Experimental verification of the results

Since the above calculations are of numerical character, it is plausible to check the results in some way. One of the most practical ways to test the results is to model the electrode arrangement by a resistor network [12, 13]. In the present case, with a

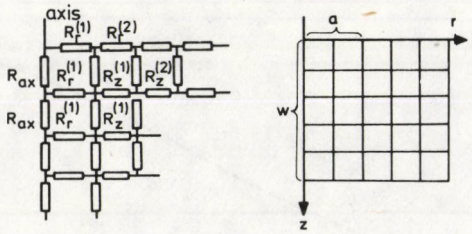


Fig. 5. The resistor network model with its coordinate representation.

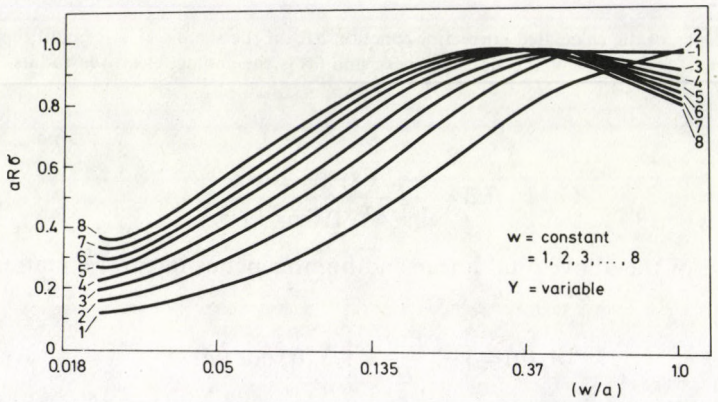


Fig. 6. Dependence of the measured correction function ($aR\sigma$) on (w/a) at different iso-cross-section area values.

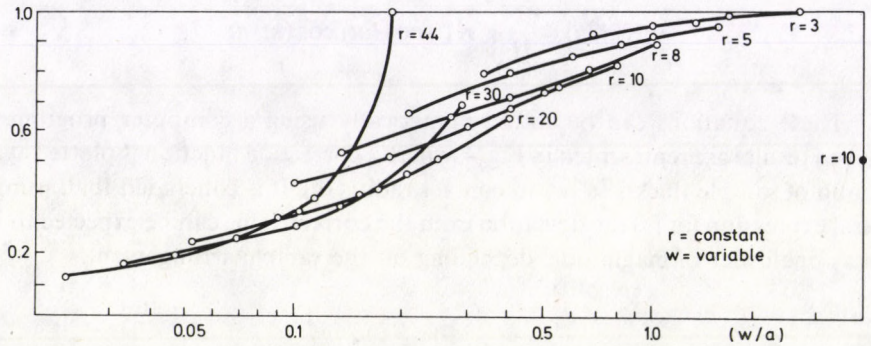


Fig. 7. Dependence of the measured correction function ($aR\sigma$) on (w/a) at different iso-thickness values.

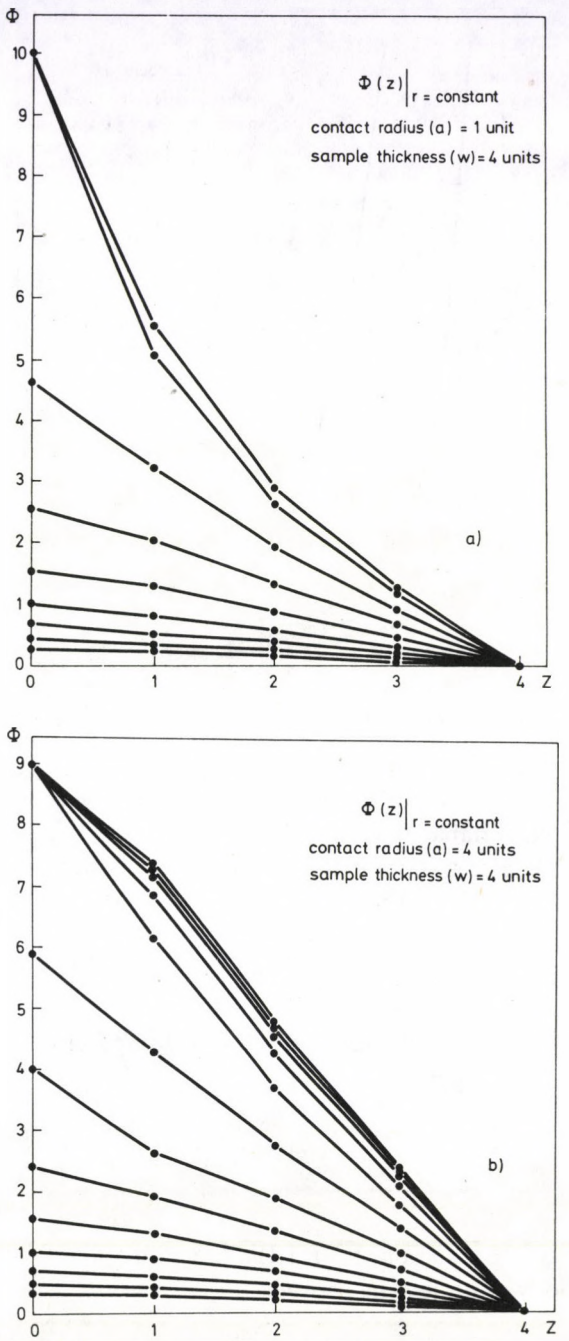


Fig. 8. The potential distribution along the z-axis at different distances from the cylindrical axis. (*a*) contact radius (a) = 1 unit, and (*b*) contact radius (a) = 4 units.

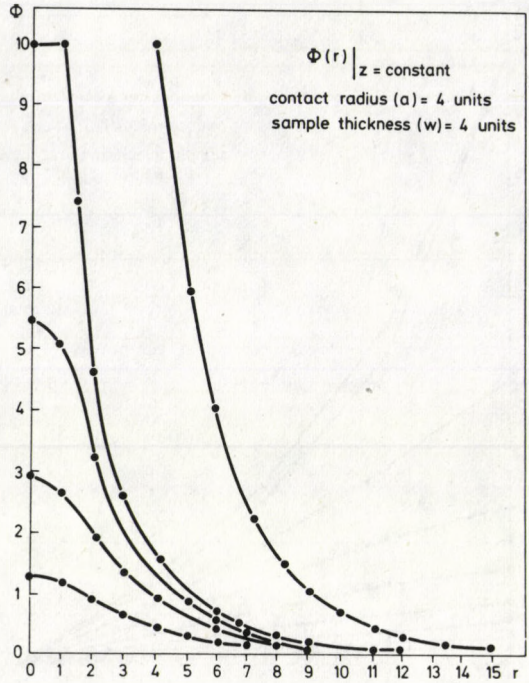


Fig. 9. The potential distribution along the r-axis at different values of z.

potential distribution of cylindrical symmetry, a two-dimensional resistor network model seemed to be suitable, (Fig. 5).

Resistors in Z-direction were chosen over the distances $r = nh$ according to

$$R_z(r) = \frac{R_0}{n + \frac{1}{2}}, \tag{17}$$

where $n = 1, 2, 3, \ldots$ and (h) is the scale and dimensioning factor. Similarly, the resistor values in r-direction were given by

$$R_r(r) = \frac{R_0}{n + \frac{1}{2}} \tag{18}$$

in the intervals between $r = nh$ and $r = (n+1)h$. Along the axis of the cylinder, Eq. (17) is not applicable. So, we can use the formula

$$R_{\text{axis}} = 8R_0. \tag{19}$$

In this case, metal film resistors were used and the characteristic resistance was chosen $R_0 = 12.5 K\Omega$. A great number of resistors are required for high resolution and

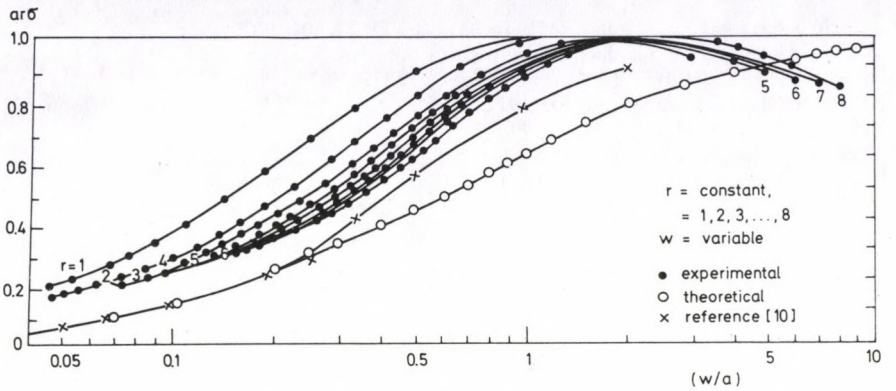


Fig. 10. Comparison between results of the present measurements represented as $aR\sigma$ versus ln w/a, and those calculated according to the data given in [10].

accuracy. For practical reasons (such as measuring time, dimensions, preparation of the measurements, etc.), a network containing 44×80 resistors were used in which the electrode arrangement was simulated as seen in Fig. 5. The resistance can be measured directly on the network. Fig. 6 shows the plots of $aR\sigma$ vs $\ln(w/a)$ for different values of w. If w is changed continuously at different (a) values, the curves will be as shown in Fig. 7. In Figs 8 and 9, the measured potential is shown as a function of z for different values of r. Comparing our calculations and measurements on the network with the results given by theoretical and experimental works of different authors we obtain Fig. 10. On the other hand, our measurements have proved that not even the correct theoretical calculations can be used for samples with arbitrary dimensions. So, one can affirm that in most of the practical cases the finiteness of the sample should be taken into account. However, as this calculation process is too complicated, we think that the present resistivity network model gives a good tool for measuring and evaluating the correction function necessary in the calculation of the spreading resistivity. Moreover, for controlling the applicability of such a network model, the functions $\Phi(z)$ and/or $\Phi(r)$ can be used. It implies, however, that the apparent irregularity in any of those functions refers to the limit of the given network dimension.

References

1. N. S. Mott and E. A. Davis, Electronic Processes in Noncrystalline Materials, Clarendon Press, Oxford, 1971.
2. J. Tauc, Ed. Amorphous and Liquid Semiconductors, Plenum, London, 1974.
3. M. F. Kotkata, S. R. Atallah and M. K. Mously, Proc. 14th. Int. Conf. on Thermal Conductivity, Strorrs, USA, 1975.
4. T. Ishikawa, M. Kitao, H. Akao and S. Yamada, Phys. Stat. Solidi, (a) **57,** 373, 1980.
5. G. F. Foxhall and J. A. Lewis, Bell System Techn. J., **43,** 1609, 1964.

6. P. A. Schumann and E. E. Gardner, IMP Techn. Rep., TR **22,** 404, 1967.
7. P. A. Schumann and E. E. Gardner, J. Electrochem. Soc., **116,** 87, 1969.
8. O. Brooks and H. G. Mattes, Bell System Techn. J., **50,** 775, 1971.
9. S. M. Hu, Solid State Electron., **15,** 809, 1972.
10. L. Gútai, Acta Techn. Acad. Sci. Hung., **80,** 251, 1975.
11. P. Schnabel, Z. Angew. Phys., **22,** 136, 1967.
12. L. Gútai, Publ. of the Res. Inst. for Techn. Phys. of the Hungarian Acad. of Sciences, 1974.
13. O. V. Widder, The Laplace Transform, McGraw–Hill Book Company, Inc., New York, 1959.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 371—382 (1982)

TEMPERATURE AND MAGNETIC FIELD DEPENDENCE OF MAGNETIZATION AND MAGNETORESISTANCE IN AMORPHOUS Fe–Ni BASED ALLOYS

A. HANDSTEIN, J. SCHNEIDER

CENTRAL INSTITUTE OF SOLID STATE PHYSICS AND MATERIAL RESEARCH
ACADEMY OF SCIENCES OF THE GDR
8027 DRESDEN, GDR

and

J. KADLECOVÁ

INSTITUTE OF PHYSICS, CZECHOSLOVAK ACADEMY OF SCIENCES
18040 PRAGUE, CZECHOSLOVAKIA

The behaviour of magnetization M and magnetoresistance $\Delta\rho^H/\rho^0$ of amorphous $(Fe_xNi_{1-x})_{80}P_{10}B_{10}$ alloys is reviewed at low temperatures and high magnetic fields. The behaviour of $M(T, H, x)$ characterizes these alloys over a large concentration range as inhomogeneous and/or heterogeneous magnetic materials. The magnetoresistance data support this picture.

Introduction

The basic problem of magnetism is the determination of the magnetization M as function of temperature, magnetic field and composition of substitutes. Amorphous magnetic alloys are of particular interest because they may yield fundamental information concerning the influence of topological and chemical disorder on the formation of magnetic moments and their interaction. In this review the temperature and magnetic field dependence of magnetization and magnetoresistance at low temperatures of amorphous $(Fe_xNi_{1-x})_{80}P_{10}B_{10}$ alloys are discussed.

Magnetic properties

A great deal of experimental information on the onset of magnetic ordering in amorphous alloys is now available. The situation for Fe–Ni based alloys is illustrated in Fig. 1, where the phase diagram of Bhagat et al [1] for a Fe–Ni based system is supplemented by our results concerning a related system richer in transition metal content. Increasing the concentration of the magnetic component in the alloys results in various magnetic behaviours: paramagnetism, superparamagnetism, inhomogeneous and homogeneous ferromagnetic ordering. Below the critical concentration

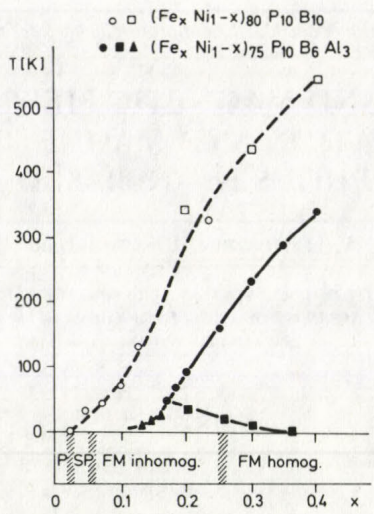


Fig. 1. Magnetic phase diagram of two Fe–Ni based systems. The insert at the bottom shows the states of $(FeNi)_{80}P_{10}B_{10}$ at 0 K, open signs denote T_c (values for FeNiPBAl from [1]).

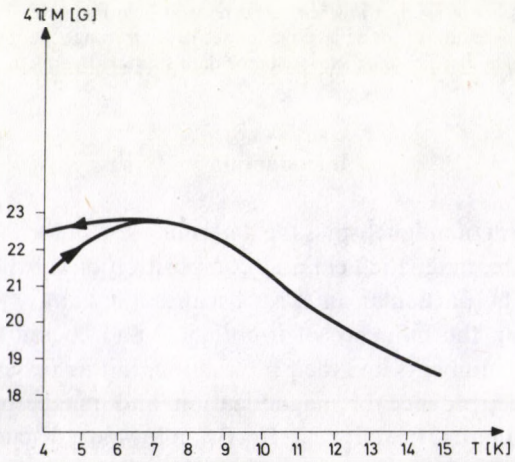


Fig. 2. Magnetization vs temperature of a $(Fe_{0.04}Ni_{0.96})_{80}P_{10}B_{10}$ sample, started from the zero field cooled state ($H = 140$ A/cm).

x_{crit}, where the system becomes ferromagnetic at 0 K, a spin glass state may appear. Above x_{crit} a mictomagnetic-like behaviour (mixed magnetic state) has been observed [2] (see Fig. 2).

The observed phase diagrams for amorphous alloys are similar to those for crystalline alloys, e.g. AuFe [3]. But, like in crystalline alloys, there is some doubt about a phase transition at a temperature T_f below the Curie temperature T_c from the ferromagnetic state to a spin-glass state near x_{crit} (see Fig. 1). More probably the

freezing of finite magnetic clusters at T_f, which coexist with the ferromagnetic matrix above T_f, may cause the observed anomalous magnetic behaviour below T_c [2, 3].

Complex magnetic structures including also a magnetic cluster glass or a mixture of ferromagnetic and mictomagnetic behaviour are expected quite generally in the case of local fluctuations in the exchange field and the anisotropy fields, depending sensitively on short-range chemical and topological order. The amorphous alloys which are heterogeneous in their crystalline equilibrium states, appear to be excellent candidates for such a behaviour.

Much evidence has been given that the magnetic properties of these amorphous magnetic materials can be well described within the itinerant electron model [4 to 7]. The evidence originates from the systematic variation of the Curie temperature similar to crystalline transition metal alloys, the temperature dependence of the magnetization, which obeys the Bloch's law, as well as from the observed ratio q_c/q_s vs T_c, the Arrott – Belov – Kouvel plots (ABK-plots) and the magnetovolume properties, as discussed below. Finally, the itinerancy is manifested by the typical dependence of the magnetization and T_c on concentration x.

For magnetically homogeneous systems one finds the following expression for $M(T, H)$ applying the Landau theory of second order phase transition

$$M^2(H, T) = M^2(0, T) + 2M^2(0, T)\chi(0, T)\, H/M(H, T), \tag{1}$$

where $\chi(0, T)$ and $M(0, T)$ are the zero field susceptibility and magnetization, respectively. Finally, one obtains within the Stoner – Wohlfarth model for homogeneous weak itinerant ferromagnets the following relations

$$M^2(0, T) = M^2(0, 0)\,(1 - T^2/T_c^2), \tag{2}$$

$$M^2(0, 0) \sim T_c^2 \sim \chi(0, 0)^{-1} \sim |\, x - x_{\text{crit}}\,|\,, \tag{3}$$

and

$$q_c/q_s \sim T_c^{-1}\,, \tag{4}$$

where q_c and q_s are the magnetic moments per atom deduced from the Curie – Weiss constant and from saturation magnetization, respectively.

The T^2-variation of $M^2(0, T)$ for Ni-rich FeNiPB alloys [2, 7], the increase of q_c/q_s with decreasing T_c support a description of these materials within the itinerant model. However, there is a systematic increase of q_c/q_s with increasing T_c above that for crystalline alloys (Rhodes – Wohlfarth plot). An explanation of this fact has been recently given by Wohlfarth et al [8], regarding atomic scale fluctuations of the magnetization.

The observed dependence of $M(0, 0)$ on x as well as of T_c on x [7], which are not parabolical with $(x - x_{\text{crit}})$ as expected from Eq. (3) suggest an inhomogeneous magnetization, too. The T_c vs x curve of FeNiPB is similar to that in giant moment systems as e.g. crystalline NiRh alloys [9], where the existence of nonaligned large magnetic clusters above x_{crit} has been proved.

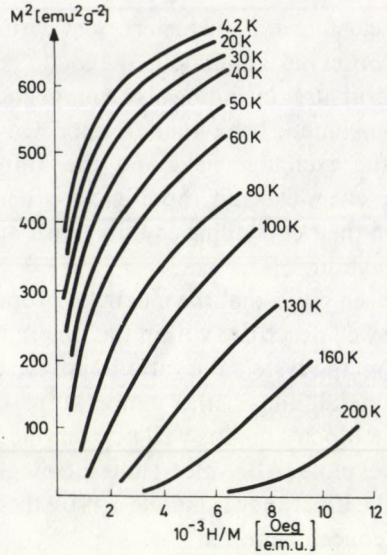


Fig. 3. M^2 vs H/M (ABK-plots) for $(Fe_{0.07}Ni_{0.93})_{80}P_{10}B_{10}$ in magnetic fields up to 14 T.

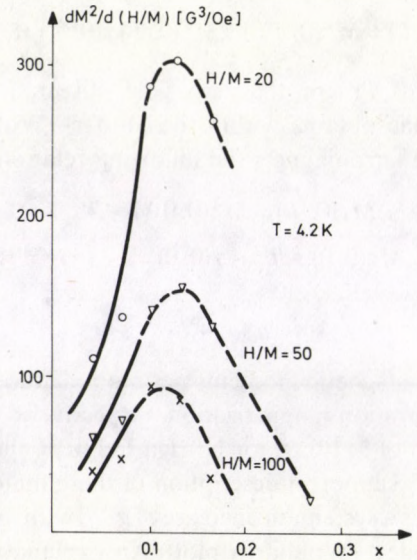


Fig. 4. Curvature of the ABK-plots $dM^2/d(H/M)$ vs Fe concentration x in $(Fe_xNi_{1-x})_{80}P_{10}B_{10}$.

According to Eq. (1) the M^2 vs H/M curves (ABK-plots) should be straight lines as observed for instance for crystalline NiPt alloys. For these amorphous $(Fe_xNi_{1-x})_{80}P_{10}B_{10}$ alloys a rather pronounced curvature of the ABK-plots is observed (see Fig. 3). In comparison with inhomogeneous ferromagnetic alloys, where the curvature of the curves changes from convex to concave as T passes from above to

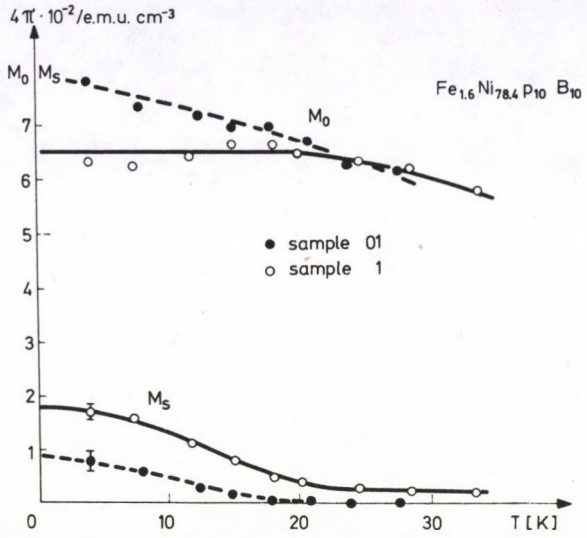


Fig. 5. Spontaneous magnetization M_s and saturated moment M_0 of the clusters as function of temperature.

below T_c [7], an apparent asymmetry of the ABK-plots about T_c is found like in crystalline NiV or NiCu alloys. As is well known, magnetism appears mostly in a largely inhomogeneous way in disordered alloys. For the amorphous FeNi-based alloys large deviations from linearity persist to much higher concentrations. Fig. 4 shows the slope $dM^2/d(H/M)$ as a measure of the curvature vs x (see also [7]). The data reveal the importance of the knowledge of structure and metallurgical state to understand the magnetic behaviour.

Our analysis of the field and temperature dependence of M in fields up to 14 Teslas for the Ni-rich FeNiPB alloys indicated an anomalously large paraprocess and showed the existence of spontaneously magnetized regions in addition to superparamagnetic particles (10 to 12). Figs 5 and 6 give the temperature dependence of M_s in the ferromagnetic phase and of the superparamagnetic moment $(M - M_s)$, respectively, measured with two samples of $(Fe_{0.02}Ni_{0.98})_{80}P_{10}B_{10}$.

The study of $M(T, H)$ after annealing [12, 13] also demonstrates that for the Ni-rich FeNiPB alloys ferromagnetic phases as well as superparamagnetic clusters contribute to the total magnetization (magnetically heterogeneous materials). Therefore the observed magnetic isotherms and the rather curved ABK-plots can be well explained within a model assuming a weak ferromagnetic homogeneous matrix and superparamagnetic clusters described for instance by a Brillouin function including an interaction between them [14, 15]. Fig. 7 shows the measured and calculated magnetic isotherms for $(Fe_{0.07}Ni_{0.93})_{80}P_{10}B_{10}$, assuming that the total magnetization may be written as [14]

$$M(H, T) = M_0(H, T) + M_s(H, T) + \Delta M(H, T). \tag{5}$$

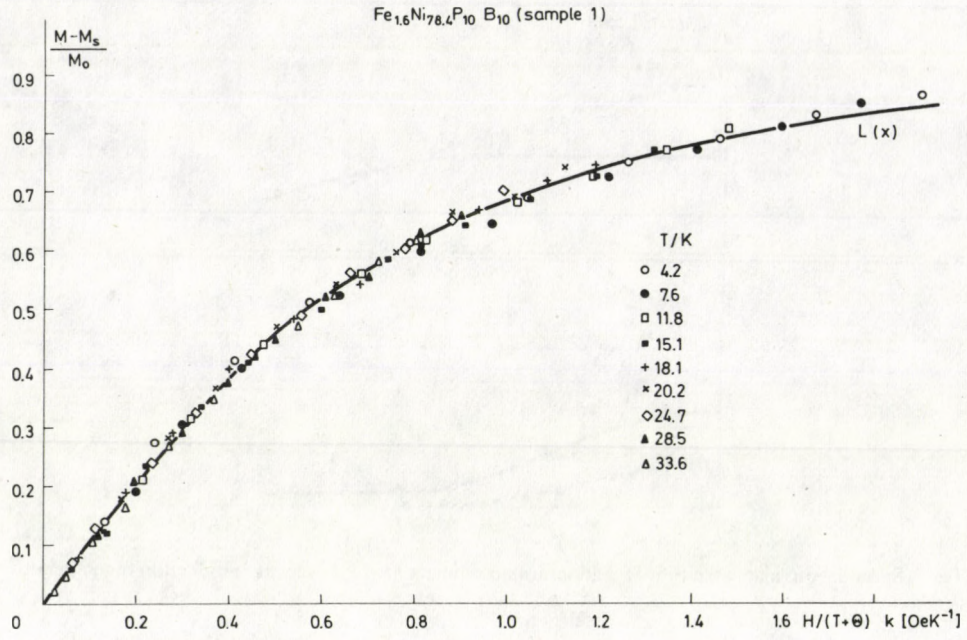


Fig. 6. $(M-M_s)$ as function of $H/(T+\Theta)$ for $(Fe_{0.02}Ni_{0.98})_{80}P_{10}B_{10}$, $\Theta = 14$ K.

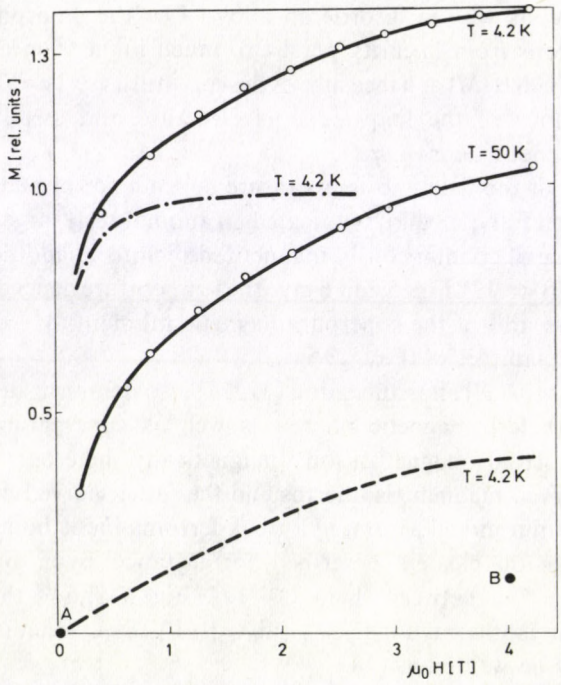


Fig. 7. Magnetic isotherms for $(Fe_{0.07}Ni_{0.93})_{80}P_{10}B_{10}$ (∘ – experimental data, —— calculated curve, $\cdot - \cdot - \cdot$ $M_0(T,H) = M_0(T)(1-a(T)/H)$, — — — $M_s(T,H) + \Delta M(T,H) = b(T)H - c(T)H^3$; cf. [14].

Magnetoresistivity

Recently, the transport properties of amorphous metallic alloys have been widely investigated. The considerable interest is due to the high degree of compositional as well as topological disorder in such systems and the rather short mean free path of the electrons. Further the investigation of the transport properties gives useful information about thermal magnetic excitation as well as about the magnetic structure, the present magnetic phases changing with the chemical composition in the amorphous magnetic alloys.

The electrical resistivity ρ of many amorphous magnetic transition metal — metalloid alloys shows a minimum at a temperature T_{min}. Resistivity minima have been the subject of many recent investigations (16 to 24). A logarithmic rise of ρ below T_{min} and a dependence $\sim T^n$ above it give a characteristic temperature dependence of the form

$$\rho(T)=\rho_0-a\ln T+bT^n. \tag{6}$$

An explanation of the effect has been given by spin-flip Kondo scattering, i.e. by a magnetic origin (see, for instance [25]). Another model has been suggested by Cochrane et al [26] assuming scattering due to two tunneling levels.

Regarding the dependence of the resistivity minimum on the composition of the alloys and on the applied magnetic field [16 to 20, 27] one finds some glasses which clearly support the structural model as well as some glasses which are well explained by a modified Kondo mechanism.

Our investigations of the resistivity of the system $(Fe_xNi_{1-x})_{80}P_{10}B_{10}$ as a function of temperature and magnetic field indicate a magnetic and a nonmagnetic contribution to the resistivity minimum at the first time. Fig. 8 demonstrates the temperature dependence of the resistivity for $(Fe_{0.07}Ni_{0.93})_{80}P_{10}B_{10}$ at different values of an applied longitudinal field. Fig. 9 shows the concentration dependence of the minimum temperature T_{min} at zero field and at $\mu_0 H=14\,T$ in $(Fe_xNi_{1-x})_{80}P_{10}B_{10}$. The coefficient A in Eq. (6) as a function of the iron concentration x at different applied fields is given in Fig. 10 (A is the coefficient of the logarithmic term in $(\rho-\rho_{4.2})/\rho_{4.2}$ analogous to a in Eq. (6)). The observed concentration dependence of T_{min} at $H=0$ is similar to that in $(Fe_xNi_{1-x})_{79}P_{13}B_8$ [24].

The maximum value of A, of T_{min} at $H=0$ and of $\Delta T_{min}=(T_{min}(H=0)-T_{min}(H=14\,T))$ appears at about the critical concentration x_{crit} for the onset of long range magnetic order [2]. With increasing the external magnetic field, A and T_{min} decrease at a chosen value of x and seem to saturate. The observed behaviour of the resistivity minimum may be discussed in the light of the magnetic phase diagram (cf. Fig. 1). The magnetic contribution ΔT_{min} to the resistivity minimum decreases drastically with x and becomes zero about $x=0.18$ to 0.20. ΔT_{min} appears to be nonzero only in the case of a mixed magnetic state or for a magnetically inhomogeneous or heterogeneous phase. ΔT_{min} becomes larger especially near x_{crit}, (cf.

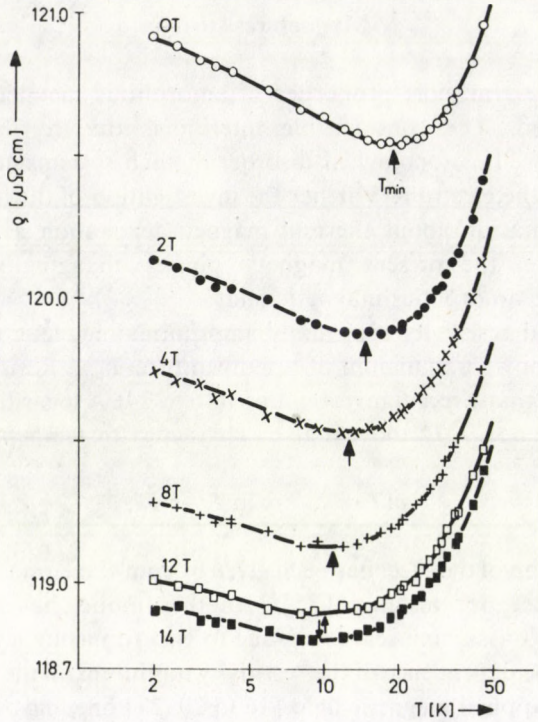


Fig. 8. Temperature dependence of the resistivity of amorphous $(Fe_{0.07}Ni_{0.93})_{80}P_{10}B_{10}$ ribbon in longitudinal magnetic field up to 14 T (the arrows indicate T_{min}; ○ – 0 T, ● – 2 T, x – 4 T, + – 8 T, □ – 12 T, ■ – 14 T).

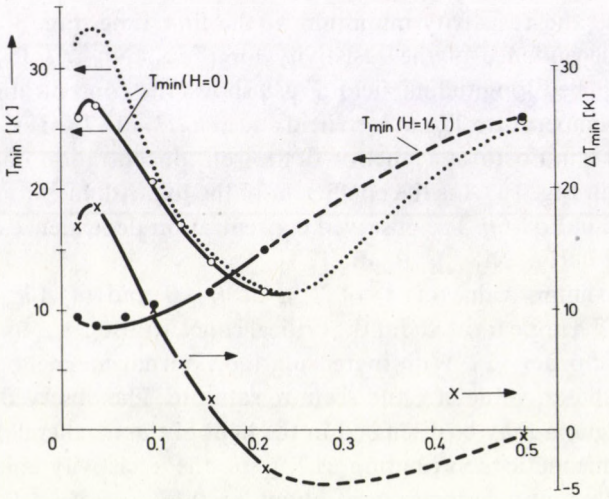


Fig. 9. Concentration and magnetic field dependence of T_{min} for $(Fe_xNi_{1-x})_{80}P_{10}B_{10}$. Dotted line — results for $(Fe_xNi_{1-x})_{79}P_{13}B_8$ from [24]. $\Delta T_{min} = T_{min}(H=0) - T_{min}(H=14\,T)$.

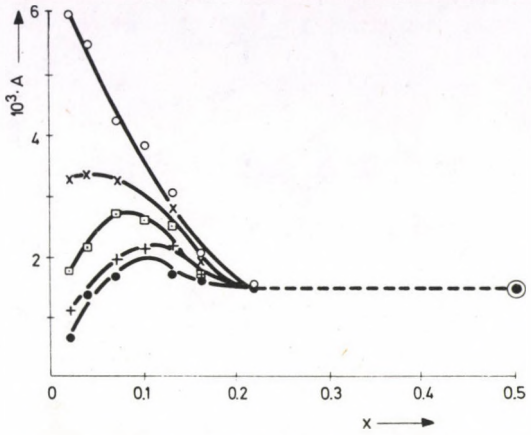


Fig. 10. Coefficient A of the logarithmic term in Eq. (6) vs Fe concentration x for $(Fe_xNi_{1-x})_{80}P_{10}B_{10}$ (○ – zero field, x – 2 T, □ – 4 T, + – 8 T, ● – 14 T).

also [24]) where a large number of independent rather local magnetic clusters may exist. With increasing x the overlap of the magnetic clusters coexisting with an already ferromagnetic phase increases and more and more giant clusters seem to appear like in AuFe [3]. Finally, a mictomagnetic behaviour of the magnetic clusters may also result due to local fluctuations in the exchange field and anisotropy field.

While for the magnetic part ΔT_{min} and A one observes a decrease with increasing x ($\Delta T_{min} \sim A$ from $x \approx 0.04$ to 0.25) the nonmagnetic contribution gives a constant logarithmic coefficient A (see Fig. 10). T_{min} at $H = 14$ T itself, the nonmagnetic part, shows a remarkable dependence on x, which is difficult to explain at present by the structural model proposed by Cochrane et al [26].

So far, neither a satisfactory theory of transport properties of amorphous metallic alloys, nor a generally acceptable model elucidating the resistivity minima observed in these materials exists. For the FeNiPB alloys, however, a qualitative agreement can be achieved between our results on the magnetization and resistivity behaviour, the magnetic phase diagram, and the observed ΔT_{min} dependence using a modified Kondo mechanism as proposed by Hasegawa, Tsuei, Sharon, Grest and Nagel [25, 27, 28]. The model is based on the assumption that not all spins suffer d – d exchange and some of them occur in a very weak exchange field. A distribution $P(H_{ef})$ of the effective internal field H_{ef} with the property $P(0) \neq 0$ was suggested by Sharon and Tsuei [28] on the basis of Mössbauer measurements in $Fe_xPd_{80-x}P_{20}$ amorphous alloys. The same suggestion was made on the basis of Mössbauer studies for $Fe_{32}Ni_{36}Cr_{14}P_{12}B_6$ by Chien [29], for $(Fe_{1-x}Mn_x)_{75}P_{15}C_{10}$ by Wieser et al [30], and for $(Fe_xMn_{1-x})_{75}P_{16}B_6Al_3$ by Chien et al [31]. In [31] the coexistence of ferromagnetic and spin-glass like ordering (mixed magnetic state) was also proved as we likewise observed from magnetic measurements in Ni-rich $(Fe_xNi_{1-x})_{80}P_{10}B_{10}$ alloys. The $P(H)$ distribution and the magnetic structure seem to depend very sensitively on the metallurgical state of the samples.

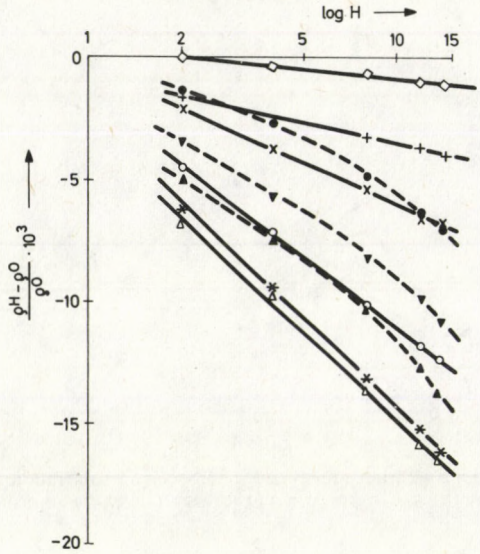


Fig. 11. Dependence of the longitudinal magnetoresistance on log H for $(Fe_xNi_{1-x})_{80}P_{10}B_{10}$($x-x=0.02$ ○ -0.04, ✳ -0.07, △ -0.1, ▽ -0.13, + -0.16, ◇ -0.22 at 4.2 K; solid signs at T 35 K).

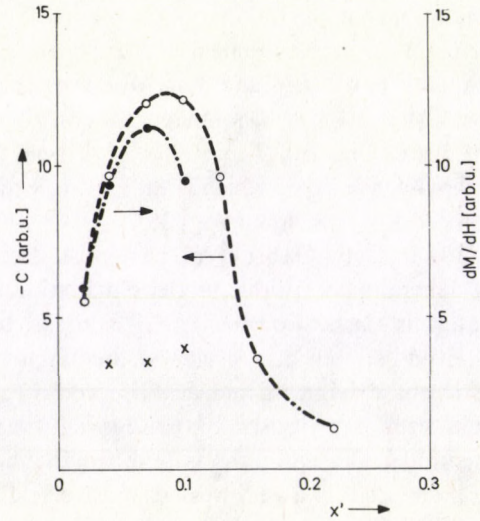


Fig. 12. Concentration dependence of the coefficient C from $\Delta\rho^H/\rho^0 \sim -C \log H$(○) and the slope of the magnetization curve $dM/dH\,|_{14T}$ (●) at 4.2 K for $(Fe_xNi_{1-x})_{80}P_{10}B_{10}$. (x – indicate $(M-M_0)$ for different concentrations in arbitrary units at 4.2 K and 14 T, cf. [20]).

The strong connection between the observed magnetic structure and the resistivity behaviour is also seen from the magnetoresistance data. Fig. 11 gives the longitudinal magnetoresistance $\Delta\rho^H/\rho^0=(\rho^H-\rho^0)/\rho^0$ vs log H for different x at 4.2 K and 35 K. While a log H-dependence holds for very low temperatures T, deviations are observed for $T > T_{min}$ as indicated in Fig. 11. Fig. 12 together with Fig. 4 show the

concentration dependence of the slope C in $\Delta\rho^H/\rho^0 \sim -C \log H$ (Fig. 11) in comparison with the concentration dependence of the high field susceptibility dM/dH, the difference $(M - M_0)$, and $dM^2/d(H/M)$ (see also above, M_0 is the magnetization at $T=0$ and $H=0$).

The macroscopic effect of the field is a decrease of the resistivity. It can be clearly seen that the negative magnetoresistance becomes especially large in the concentration range, where the FeNiPB alloys exhibit a magnetically inhomogeneous spatial distribution of the magnetization (coexistence of large magnetic clusters with a ferromagnetic matrix). The effect in $\Delta\rho^H/\rho^0$ is negligibly small near x_{crit}, where T_{min} is large due to the independent local magnetic moments. The different concentration dependence of T_{min} and $\Delta\rho^H/\rho^0$ seems to be strongly connected with the type of the existing magnetic moments—local moments, finite spatial extended magnetic clusters (giant moments), or infinite magnetic clusters—and their interaction. The ratio $(\Delta\rho^H/\rho^0)/M^2$ at different temperatures is nearly constant (straight lines $\Delta\rho^H/\rho^0 \sim M^2$ in [20]). Small deviations may reflect the importance of the interaction between the clusters and the matrix.

Conclusions

We have shown that in Ni-rich amorphous $(Fe_xNi_{1-x})_{80}P_{10}B_{10}$ alloys there is a magnetic as well as a nonmagnetic contribution to the minimum of the electrical resistivity. The ABK-plots, the temperature dependence of the resistivity at low temperatures, and the magnetoresistivity appear to be useful quantities to differentiate between homogeneous or weakly inhomogeneous ferromagnetic and rather heterogeneous (mixed magnetic phases) magnetic materials. The transport behaviour corresponds to the magnetic structure of the alloys.

A complete microscopic theory for $M(T, H)$ of magnetically inhomogeneous or heterogeneous alloys and for the electrical resistivity $\rho(T, H)$ of amorphous ones does not exist at present. New experimental work in the field of chemical, structural characterization will be of great interest in order to better understand the nature and the degree of heterogeneity in different systems and their influence on $M(T, H, x)$ and $\rho(T, H, x)$.

References

1. S. M. Bhagat, J. A. Geohegan, M. L. Spano and H. S. Chen, J. Appl. Phys., **52,** 1741, 1981.
2. J. Schneider et al, Proc. Int. Conf. Soft Magnetic Mats., 3, 1977, Bratislava, p. 551; Phys. Stat. Solidi (a **45,** K 47, 1978.
3. B. R. Coles et al, Phil. Mag. B, **37,** 489, 1978.
4. E. P. Wohlfarth, IEEE Trans. Magn. MAG-**14,** 933, 1978; Int. Symp. High Purity Mats., Dresden, 198(Vol. 1, p. 28.
5. M. Goto et al., Jap. J. Appl. Phys., **19,** 51, 1980.
6. J. Schneider, A. Handstein, R. Hesske, and K. Zaveta, Physica, **86–88B,** 301, 1977.
7. J. Schneider and K. Zaveta, Proc. Int. Conf. Rapidly Quenched Metals 4, Sendai, 1981, to be publishec

8. D. Wagner und E. P. Wohlfarth, J. Mag. Mag. Mats. **15–18,** 1345, 1980.
9. W. C. Muellner and J. S. Kouvel, Phys. Rev., **B11,** 4552, 1975.
10. J. Schneider, A. Handstein, J. Henke, K. Zaveta, and T. Mydlarz, J. Physique (Paris), **41,** C8 – 682, 1980.
11. K. Zaveta, J. Schneider, A. Handstein and F. Zounova, Proc. Int. Conf. Rapidly Quenched Metals 3, B. Cantor, edit. The Metals Society, London, 1978, p. 156.
12. J. Schneider, K. Zaveta and A. Handstein, Proc. Int. Conf. Amorphous Metallic Mats., (Smolenice 1978), Bratislava, 1980, p. 285.
13. A. Handstein, J. Schneider and K. Zaveta, Proc. Int. Conf. Rapidly Quenched Metals 4, Sendai, 1981, to be published.
14. J. Schneider, K. Zaveta, A. Handstein, F. Zounova and V. Nekvasil, Phys. Stat. Solidi, (a), **50,** K 141, 1979.
15. Z. Kalva, K. Zaveta, J. Schneider, A. Handstein and F. Zounova, Proc. 14. Conf. on Metals, Dresden, 1981.
16. S. J. Poon, J. Durand and M. Yung, Solid State Commun., **22,** 475, 1977.
17. N. I. Marzwell, J. Mag. Mag. Mats., **5,** 67, 1977.
18. A. Hamzic, E. Babić and Z. Marohnić, J. Physique (Paris), **41,** C8 – 694, 1980.
19. V. Gregor and J. Kadlecova, Phys. Stat. Solidi (b), **97,** K 49,1980; Conf. Metallic Glasses: Science and Technology, Budapest, 1980, Vol. 1, p. 417.
20. A. Handstein, J. Kadlecova and V. Gregor, Phys. Stat. Solidi (a), **66,** 657, 1981.
21. R. Asomoza, I. A. Campbell, A. Fert, A. Lièнard and J. P. Rebouillat, J. Phys. F, **9,** 349, 1979.
22. K. V. Rao et al., J. Appl. Phys., **50,** 1592, 1979; IEEE Trans. Magn. MAG-**16,** 1980; Phys. Rev., **B22,** 3373, 1980.
23. G. Böhnke and M. Rosenberg, J. Physique (Paris), **41,** C8 – 481, 1980.
24. J. Durand, Conf. Metallic Glasses: Science and Technology, Budapest, 1980, Vol. 1, p. 399.
25. R. Hasegawa and C. C. Tsuei, Phys. Rev., **B2,** 1631, 1981.
26. R. W. Cochrane, R. Harris, J. O.Strøm-Olsen and M. J. Zuckermann, Phys. Rev. Lett., **35,** 676, 1975; R. W. Cochrane and J. O. Strøm-Olsen, J. Phys. F, **7,** 1799, 1977.
27. G. S. Grest and S. R. Nagel, Phys. Rev., **B19,** 3571, 1979.
28. T. E. Sharon and C. C. Tsuei, Phys. Rev., **B5,** 1047, 1972.
29. C. L. Chien, Phys. Rev., **B19,** 81, 1979.
30. E. Wieser, A. Handstein, J. Schneider and K. Zaveta, Phys. Stat. Solidi (a), **66,** 607, 1981.
31. C. L. Chien, J. H. Hsu, P. J. Viccaro, B. D. Dunlap, G. K, Shenoy and H. S. Chen, J. Appl. Phys., **52,** 1750, 1981.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 383—394 (1982)

THERMAL CONDUCTIVITY OF NIOBIUM AND MOLYBDENUM SINGLE CRYSTALS AT LOW TEMPERATURES

A. GLADUN, C. GLADUN and H. VINZELBERG

CENTRAL INSTITUTE OF SOLID STATE PHYSICS AND MATERIALS RESEARCH
ACADEMY OF SCIENCES OF THE GDR
8027 DRESDEN, GDR

The thermal conductivity of Nb and Mo single crystals was measured in the temperature range of 0.05 – 20 K and 2 – 71 K, respectively.

1. Introduction

The electronic band structure of transition metals has been the subject of many experimental and theoretical investigations in recent years. Much of the success in understanding their properties is due to the improved purification techniques for single crystals. It seems of interest to carry out further experimental investigations of the transport properties at low temperatures. In the present paper measurements of thermal conductivity on pure niobium and molybdenum single crystals over a wide temperature range are reported.

2. Niobium

The production and preparation of the samples have been described elsewhere [1, 2]. A characterization of the samples is given in Table I. The thermal conductivity was measured by a stationary method [1] in the temperature range 0.05 to 20 K.

Table I

Orientation of rod axis, length, diameter, residual resistivity ratio $RRR = \rho_{293}/\rho_0$, residual resistivity ρ_0, saturation factors of thermal and electrical magnetoresistivity, coefficients A, B, $\tilde{B}$, C for electron – defect, electron – phonon intraband, electron – phonon including interband, and electron – electron scattering, respectively, see (2), (4), of investigated niobium samples.

	Nb [110]	Nb [100]
orientation	[110]	[100]
length [mm]	90	90
diameter [mm]	3.3	3.8
$RRR = \rho_{293}/\rho_0$	22700	19200
ρ_0[Ω cm]	6.34×10^{-10}	7.48×10^{-10}
$\lambda(T, B=0)/\lambda(T, B_s)$	1.92	1.27
$r(T, B_s)/r(T, B=0)$ [9]	1.93	1.24
$A = \rho_0/L_0$[cm K^2/W]	2.58×10^{-2}	3.05×10^{-2}
B[cm/KW]	3.8×10^{-4}	3.8×10^{-4}
$\tilde{B}$[cm/KW]	4.4×10^{-4}	4.4×10^{-4}
C[cm/W]	8.0×10^{-4}	8.0×10^{-4}

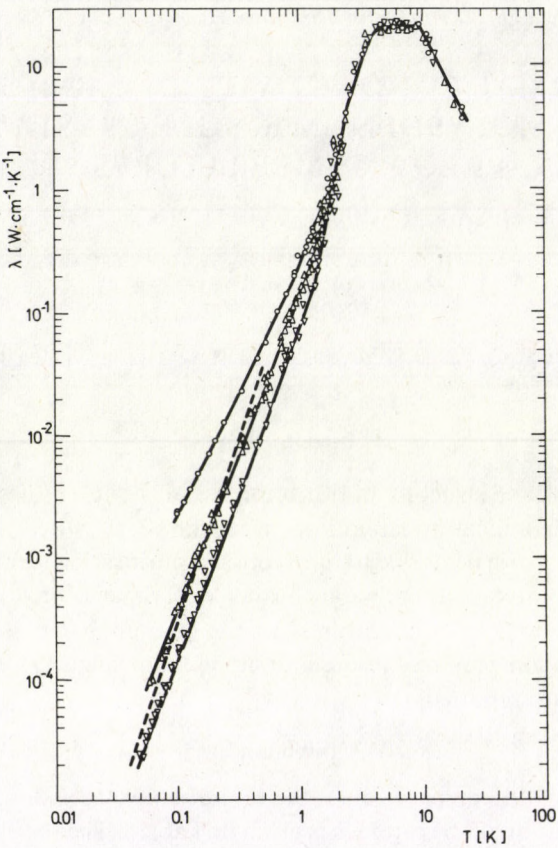


Fig. 1. Thermal conductivity of Nb [110] vs temperature ○ —as received, ▽ —sandblasted, first run, △ —sandblasted, last run. The dashed line represents the calculated T^3 lattice conductivity according to [8].

In Fig. 1 the thermal conductivity of the single crystal Nb [110] vs temperature is shown. Below 2 K the influence of the sample surface may be recognized, while above 2 K the results for the specular and sandblast sample surface coincide. The critical temperature is seen as a sharp bend in the λ-curve.

2.1. *Superconducting state*

In the range below the critical temperature ($T_c = 9.25$ K) down to about 0.4 T_c the heat conduction of pure superconductors is determined by thermally excited conduction electrons. The number of these quasiparticles is reduced continuously with decreasing temperature resulting in a decrease of thermal conductivity.

Based on the BCS-theory Bardeen and coworkers [3] and Gejlikman [4, 5] evaluated the relation of thermal conductivities in the normal and superconducting states. This relation is a function of the reduced temperature T/T_c and the energy gap at

zero temperature only. The fit of the measuring points to this relation gives for the energy gap at zero temperature 1.9 kT, which is in accordance with the results from tunneling investigations [6].

Since the number of thermally excited conduction electrons decreases the heat is transferred by phonons at very low temperatures. The lattice thermal conductivity is influenced by several scattering processes, the scattering by electrons, dislocations, and on the surface of the sample.

Dominating boundary scattering results in a T^3-dependence of the thermal conductivity and can be estimated according to the theory of Casimir [7, 8]. The theoretical curve is plotted in Fig. 1. With incompletely diffuse boundary scattering the temperature dependence is lower and the heat conduction is higher than the theoretical value ($\lambda \sim T^{2.1}$). After sandblasting we measured first the lowest curve in Fig. 1, but then the surface of the sample recovered and the thermal conductivity approaches the theoretical curve. For the temperature dependence we measured $T^{2.7}$ instead of T^3. The smaller increase indicates an additional scattering of the phonons by dislocations.

2.2. *Normal state*

In very pure single crystals the maximum of thermal conductivity in the normal state lies below T_c. To investigate this maximum it is necessary to suppress the superconductivity by means of a magnetic field. But magnetic fields necessary for this suppression already cause a considerable magneto-resistance effect. This effect must be taken into account for extrapolating the thermal conductivity to zero field.

By investigating the ideal electrical resistivity Berthel [9] found a saturation of the longitudinal electrical magnetoresistivity. He found different, temperature-independent saturation factors for the different crystal directions. By means of these factors the extrapolation of the electrical magnetoresistivity to zero field may be performed.

In this work we investigate the problem, whether the thermal conductivity in a longitudinal magnetic field shows a saturation like the electrical resistivity [9]. By means of temperature-independent saturation factors it would be possible to determine the thermal conductivity of niobium in the normal state in zero field. The longitudinal magnetic field was provided by a superconducting solenoid. The maximum field amounted to 5 T.

In Fig. 2 the thermal conductivity as a function of temperature with the magnetic field as parameter is shown for the sample Nb [110]. By a magnetic field the thermal conductivity decreases only to a limiting curve. The magnetic field which is necessary for the saturation of thermal conductivity decreases with decreasing temperatures. At temperatures above 11 K magnetic fields higher than 5 T are necessary and we could not determine the limiting curve, whereas at 5 K the limiting curve is attained already at a field of 1.5 T. We get for the ratio of the thermal conductivities in zero field and in

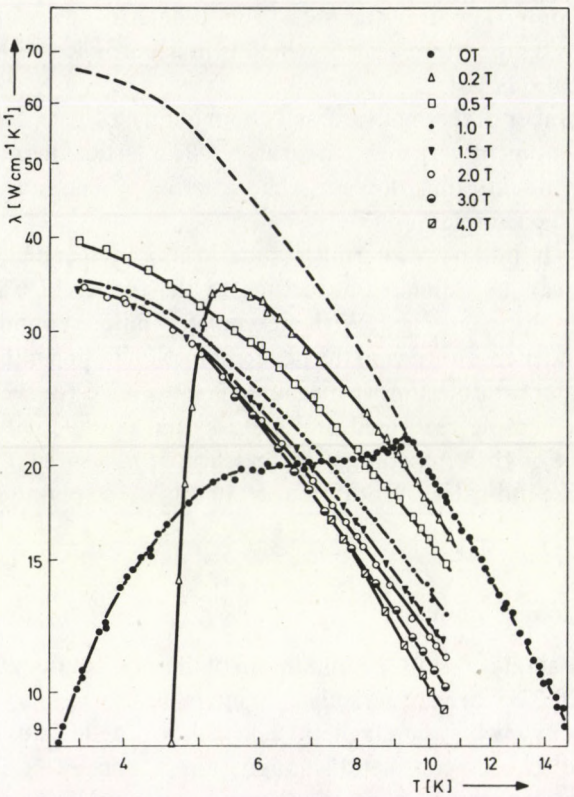


Fig. 2. Thermal conductivity of Nb [110] vs temperature in several longitudinal magnetic fields: ● $B=0$, △ 0.2, □ 0.5, ● 1.0, ▼ 1.5, ○ 2.0, ◒ 3.0, ▨ 4.0 T. The dashed line represents the thermal conductivity in the normal state below T_c extrapolated onto zero field.

saturation field, $\lambda(T, B=0)/\lambda(T, B_s)$, in the temperature range between T_c and 11 K the constant value 1.92.

The thermal conductivity at 0.2 T very steeply decreases below 5.3 K because the crystal is in the mixed state where additional scattering of the electrons by fluxoids occurs [10, 11].

In the plot of the thermal conductivity as a function of the field at constant temperature (Fig. 3) at all temperatures saturation of the thermal conductivity is observed. But below T_c the extrapolation to zero field is not possible because the $\lambda(B)$-curves are very steep at small fields and for fields smaller than H_{c2} values of the thermal conductivity in the normal state do not exist.

In Fig. 4 the ratio $\lambda(T, B)/\lambda(T, B_s)$ as a function of $B\lambda(T, B_s)$ is plotted. This plot corresponds to a Kohler diagram of the electrical magnetoresistivity. Independently of the temperature a common curve results. From this we conclude the ratio of $\lambda(T, B=0)/\lambda(T, B_s)=1.92$ at temperatures below T_c, too. Thus it is possible —as in the determination of the ideal electrical resistivity of niobium below T_c [9] — from the

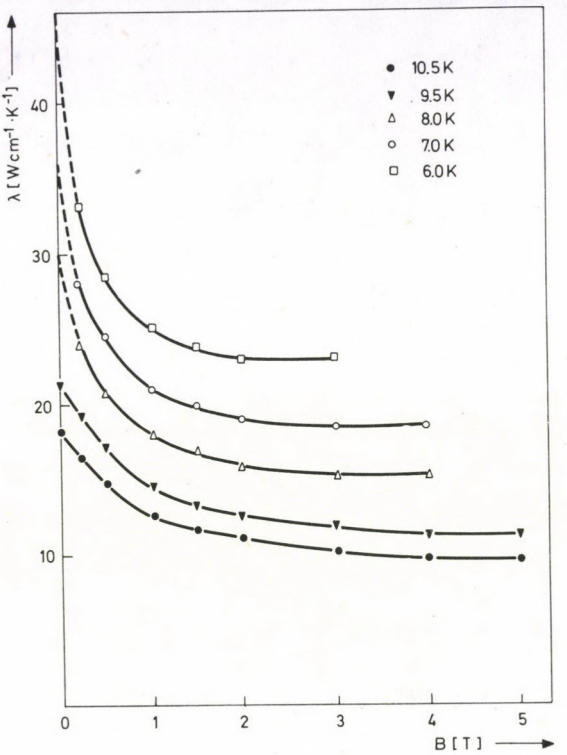


Fig. 3. Thermal conductivity of Nb [110] vs longitudinal magnetic field at several temperatures: ● $T=10.5$, ▼ 9.5, △ 8.0, ○ 7.0, □ 6.0 K. The dashed lines connect the last measured value and the value extrapolated onto zero field by saturation factor 1.92.

measurement in a longitudinal field to extrapolate with the aid of saturation factors to the thermal conductivity in zero field (dotted line in Fig. 2).

In Fig. 5 the thermal conductivity as a function of temperature with the magnetic field as parameter of the sample Nb [100] is shown. In a zero field values like those of sample [110] result, whereas the limiting curve is attained already at smaller fields. We get for the ratio $\lambda(T, B=0)/\lambda(T, B_s)$ the constant value 1.27.

The saturation factors determined for the two crystal directions [110] and [100] are in good agreement with those found for the electrical magnetoresistivity in the temperature range of liquid hydrogen and at fields to 15 T [9] (Table I).

In the following the thermal conductivity in the normal state in zero field is discussed.

In pure normal metals the heat transport by electrons is dominating. For the thermal resistivity at temperatures below 0.1 Θ_D(Θ_D is the Debye temperature) the equation

$$\lambda_e^{-1} = W_e = W_{ed} + W_{eph} = AT^{-1} + BT^2 \qquad (1)$$

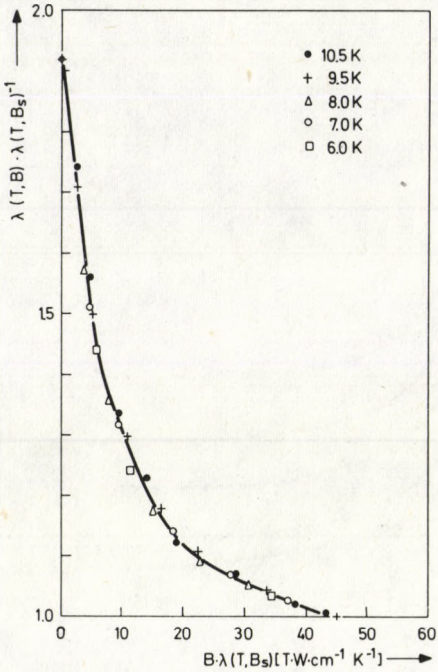


Fig. 4. Kohler plot of the thermal conductivity of Nb [110]. ● $T=10.5$, $+9.5$, △ 8.0, ○ 7.0, □ 6.0 K.

is used in general. The first term describes the scattering of the electrons by defects and the second term the scattering of the electrons by phonons. In pure transition metals additional scattering processes are to be considered, leading to the equation

$$\lambda_e^{-1} = W_e = W_{ed} + W_{eph} + W_{ee} = AT^{-1} + \tilde{B}T^2 + CT. \tag{2}$$

The thermal resistivity W_{ee} due to the electron–electron scattering is proportional to T. The term W_{eph} in (2) additionally contains the phonon induced interband scattering which gives rise to a thermal resistivity also proportional to T^2.

From the electrical resistivity Berthel [9] has found that above a temperature $\Theta_{min}/2$ the interband scattering is dominating ($\rho \sim T^3$, $\Theta_{min} = 35$ K). Below $\Theta_{min}/2$ the interband scattering processes decrease exponentially and may be neglected at temperatures below 6 K. Then the term W_{eph} contains only the intraband scattering: $W_{eph} = BT^2$. At these temperatures it is not possible to determine the constants A, B, and C only from the measured values of thermal conductivity since the investigated temperature range is too small. Therefore, the constant A was determined with the aid of the measured electrical residual resistivity ρ_0 and the Wiedemann–Franz law,

$$W_{ed} = AT^{-1} = \rho_0 L_0^{-1} T^{-1}, \qquad L_0 = 2.45 \times 10^{-8} \Omega W K^{-2}. \tag{3}$$

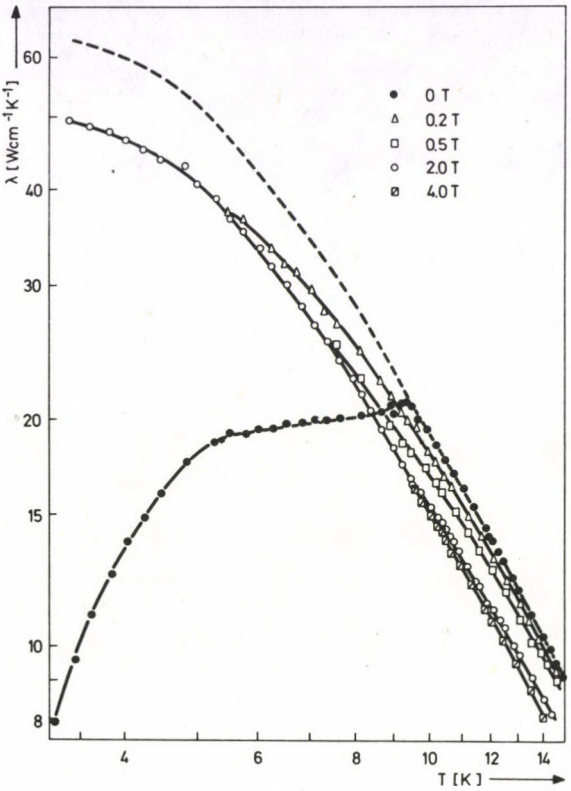


Fig. 5. Thermal conductivity of Nb [100] vs temperature in several longitudinal magnetic fields: ● $B=0$, △ 0.2, □ 0.5, ○ 2.0, ⊠ 4.0 T.The dashed line represents the thermal conductivity in the normal state below T_c extrapolated onto zero field.

The constant A for both crystals is represented in Table I. From a plot of the equation

$$T\lambda_e^{-1} - A = BT + C \tag{4}$$

we determined for both crystals equal values for B and C (Table I).

The Lorenz ratio for the electron – electron scattering contributions is defined by the expression

$$L_{ee} = \rho_{ee} W_{ee}^{-1} T^{-1} = \rho_{ee} C^{-1} T^{-2}. \tag{5}$$

Theoretical calculations by Herring [12] and Bennett and Rice [13] have given for L_{ee} values between 1.0 and 1.58×10^{-8} ΩWK^{-2}. With the value $\rho_{ee} = 23.6 \times \times 10^{12} T^2 \Omega$ cm K^{-2} obtained by Berthel [9] and our result for W_{ee}, L_{ee} becomes $2.9 \times 10^{-8} \Omega WK^{-2}$. The agreement is not good. Differences between theory and experiment were found in other metals, too. For example, for nickel [14], platinum [15], palladium [16], and tungsten [17] values of L_{ee} between 0.1 and 1.3 $\times 10^{-8} \Omega WK^{-2}$ have been determined.

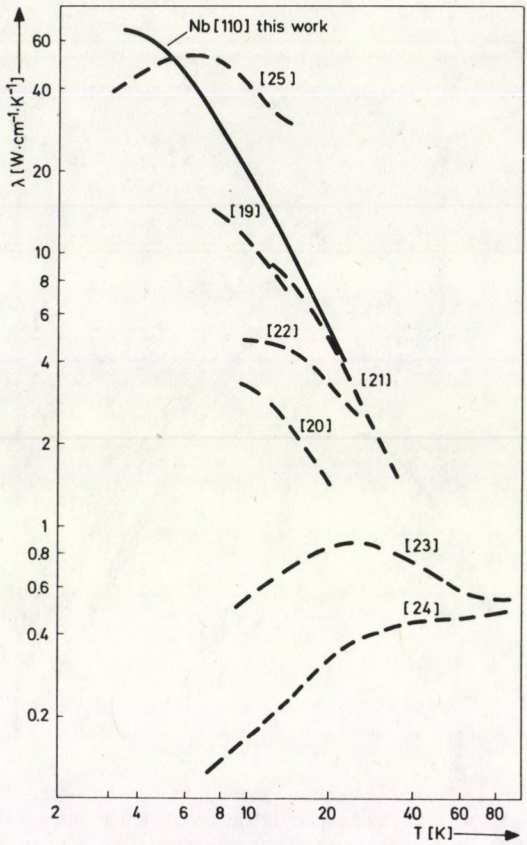


Fig. 6. Thermal conductivity of different Nb samples in the normal state vs temperature. ——— this work, – – – results from different papers.

At temperatures above 15 K we obtained according to (2) for $\tilde{B}$ the value 4.4×10^{-4} cm K^{-1} W^{-1}. This value is higher than the value B, determined for temperatures below 6 K, owing to the additional interband scattering.

For the case of dominating phonon-induced interband scattering Gladun et al [18] derived the expression

$$L_{\mathrm{eph}} = \rho_{\mathrm{eph}} W_{\mathrm{eph}}^{-1} T^{-1} = 0.36\, L_0$$

for the Lorenz ratio of the electron – phonon scattering contributions. With $\rho_{\mathrm{eph}} = 4.12 \times 10^{-12} T^3 \Omega$ cm K^{-3} [9] and our value for W_{eph} we obtain $L_{\mathrm{eph}} = 0.38\ L_0$ in good agreement with the results of Gladun et al [18].

In Fig. 6 the thermal conductivity of different Nb samples in the normal state is plotted. Our sample Nb [110] has the highest purity, therefore the maximum of the thermal conductivity has the highest value and appears at the lowest temperature. At

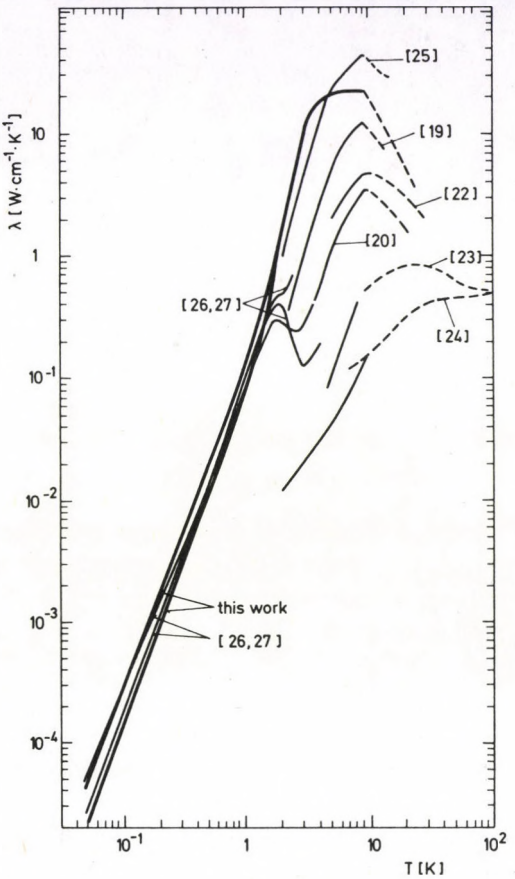


Fig. 7. Thermal conductivity of different Nb samples vs temperature in the superconducting (———) and normal (– – –) states.

high temperatures the measured curve adapts well to the behaviour of the curves in [19 to 24] for less pure Nb, on the other hand, there are deviations from the curve in [25]. In Fig. 7 the thermal conductivity of different Nb samples in the superconducting state (only above T_c in the normal state) is plotted. Below 1 K the influence of impurities on the thermal conductivity disappears.

3. Molybdenum

We investigated molybdenum single crystals purified by electron beam zone melting repeated three times after a special chemical precleaning process [28]. The crystals had the form of a rod nearly 100 mm long and of nearly 4 mm diameter. The orientation of rod axis, the residual resistivities ρ_0, and the residual resistivity ratios

Table II

Orientation, tungsten impurity content, residual resistivity ratio RRR, residual resistivity ρ_0, coefficients A, B, C for electron – defect, electron – phonon, and electron – electron scattering, respectively, of investigated molybdenum samples.

Sample	Orientation	W impurity [at ppm]	RRR	ρ_0 [10^{-11} Ω cm]	A [10^{-3} cm K^2/W]	B [10^{-5} cm/KW]	C [10^{-4} cm/W]
Mo 1074	[110]	9	68000$^{+)}$	7.75	3.2	1.7	1.2
Mo 1053	[111]	18	35000$^{+)}$	15.0	6.1	1.7	1.2
Mo W9		35	18000$^{+)}$	29.2	12	1.7	1.2
Mo [34]	[111]	–	3500$^{*)}$	158	64.0	5.13	–
Mo [35]	polycrystal	–	36.2$^{**)}$	–	6700	7.5	–

$^{+)}$ ρ_{293}/ρ_0
$^{*)}$ ρ_{300}/ρ_{12}
$^{**)}$ ρ_{293}/ρ_{20}

RRR [29] are given in Table II. The major impurity element is tungsten. The mass-spectrometric analysis has given for the concentrations of interstitial impurities C < 3, N < 0.4, O < 3 wt ppm, but only the soluted part of these impurities affects the low-temperature transport properties. The remaining impurities had a total concentration of about 2 at ppm.

The thermal conductivity was measured in the temperature range 2 to 71 K. The temperature dependence of the thermal conductivity of molybdenum is plotted in Fig. 8. In the temperature range 2 to 18 K the measured values of thermal conductivity may be described by Eq. (2), with $\tilde{B} = B$, owing to the small interband interaction in molybdenum. As in nontransition metals the electrical resistivity has a T^5-dependence [29].

Within the limits of error a fit of the measured curves in the temperature range 2 to 18 K gives equal values for the coefficients B and C for all three crystals, whereas the values for the coefficient A are related to the residual resistivity ρ_0 according to the Wiedemann – Franz law ($A = \rho_0/L_0$). The coefficients A, B, C are given in Table II. At the lowest investigated temperatures the scattering of the electrons by defects is dominating. The contribution of this scattering to the total thermal resistivity e.g. at 2.7 K amounts to 55, 70, and 82% for the samples Mo 1074, 1053, and W 9, respectively. At this temperature the ratio of electron – electron and electron – phonon scattering contributions to the remaining thermal resistivity approximatively is 3:1.

The Lorenz ratio $L_{ee} = \rho_{ee}/W_{ee}T$ for the electron – electron scattering contributions is determined from the results of the measurements of the electrical resistivity [29] $\rho_{ee} = 1.5 \times 10^{-12} T^2 \Omega\,\mathrm{cm/K^2}$ and the value of the coefficient C from (2) by $L_{ee} = 1.27 \times 10^{-8} V^2/K^2$ in good agreement with theoretical calculations [13].

In the temperature range $18 < T < \Theta_D/10$ ($\Theta_D = 430$ K, Debye temperature) the thermal resistivity increases more rapidly with temperature than T^2. In this

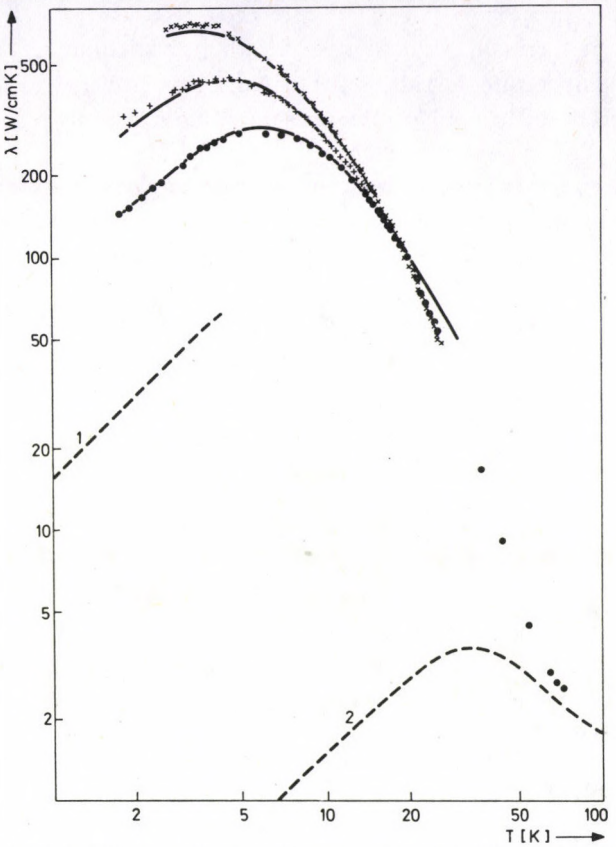


Fig. 8. Thermal conductivity of Mo as a function of temperature. × Mo 1074, + Mo 1053, ● Mo W9, 1 Mo [34], 2 Mo [35]. The solid lines are the curves evaluated according to (2) with parameters fitted to the measurements in the temperature range 2 to 18 K (see Table II).

temperature range in all investigated molybdenum crystals the electron – phonon scattering is dominating.

The electrical resistivity at temperatures near 30 K also increases more rapidly than corresponds to the T^5-law for dominating electron – phonon intraband scattering, but the deviation is smaller than the deviation of the temperature dependence of the thermal resistivity from the T^2-law. So for example the deviation $\delta\rho/\rho = (\rho_{\text{measured}}-\rho_{\text{evaluated}})/\rho_{\text{measured}}$ amounts to 2.6% at 30 K [29] but the deviation $\delta W/W$ amounts to 50%.

An increase of the thermal resistivity with temperature more rapid than expected according to the standard theory was observed already in other very pure metals, as silver [30], tungsten [17, 31], niobium [32], and cadmium [33].

There exist several models which may explain the observed more rapid increase of the thermal resistivity with temperature in the range of dominating electron-phonon scattering [17, 30]. But none of them seems to explain both the measured thermal and electrical conductivity satisfactorily.

In Fig. 8 the thermal conductivities of both molybdenum crystals investigated hitherto at low temperatures are also represented. Curve 1 refers to a single crystalline sample [34], and curve 2 to a polycrystalline sample of 99.95% purity [35]. The residual resistivity ratios of the crystals are given in Table II. In both papers [34, 35] the authors fitted the measured curves by (1) without the term for the electron – electron scattering, the resulting constants A and B are also shown in Table II. As expected, at high temperatures the measured curves for all molybdenum samples run together.

Acknowledgements

The authors wish to thank Dr. D. Elefant for the performance of the measurements of residual resistivities of the samples and Dr. H. Eschrig for valuable discussions.

References

1. A. Gladun, C. Gladun, M. Knorn and H. Vinzelberg, J. Low Temp. Phys., **27,** 873, 1977.
2. J. Barthel, K. H. Berthel and K. Fischer, Fiz. Metallov i Metallovedenie, **35,** 922, 1973.
3. J. Bardeen, G. Rickayzen, and L. Teword, Phys. Rev., **113,** 982, 1959.
4. B. T. Gejlikman, Zh. Eksp. Teor. Fiz., **34,** 1042, 1958.
5. B. T. Gejlikman, M. J. Dusenat, V. R. Cecetkin, Zh. Eks. Teor. Fiz., **73,** 2319, 1977.
6. B. Schöneich, Thesis, ZFW Dresden, 1982.
7. H. B. G. Casimir, Physica, **5,** 495, 1938.
8. C. G. Winterheimer and A. K. McCureley, Solid State Commun., **14,** 919, 1974.
9. K. H. Berthel, Thesis, TU Dresden, 1976.
10. A. Oota, T. Mamiya and Y. Masuda, J. Phys. Soc. Japan, **37,** 279, 1974.
11. A. Oota, T. Mamiya and Y. Masuda, J. Phys. Soc. Japan, **38,** 1362, 1975.
12. C. Herring, Phys. Rev. Lett., **19,** 167, 684, (E) 1967.
13. A. J. Bennett and M. J. Rice, Phys. Rev., **185,** 968, 1969.
14. G. K. White and R. J. Tainsh, Phys. Rev. Lett., **19,** 165, 1967.
15. A. C. Anderson, R. E. Peterson and J. E. Robichaux, Phys. Rev. Lett., **20,** 459, 1968.
16. J. F. Schriempf, Phys. Rev. Lett., **19,** 1131, 1967.
17. A. B. Batdalov, N. P. Katrich, N. A. Redko, V. J. Tamarchenko and S. S. Shaluit, Fiz. tverd. Tela, **19,** 672, 1977.
18. A. Gladun, C. Gladun, A. Hofmann and B. Pietrass, Phys. Stat. Solidi, (a) **58,** 149, 1980.
19. Y. Muto, K. Nato and T. Fukuroi, Proc. XI. Internat. Conf. Low Temp. Phys., Vol. 2, St. Andrews 1968, Ed. J. F. Allen, University of St. Andrews, 1968, p. 930.
20. W. Wasserbäch, Phys. Stat. Solidi, (b), **84,** 205, 1977.
21. K. Reiche and G. Pompe, J. Low Temp. Phys., **36,** 467, 1979.
22. C. D. Eger, Thesis, TU Dresden, 1975.
23. G. K. White and S. B. Woods, Canad. J. Phys., **35,** 892, 1957.
24. K. Mendelsohn and H. M. Rosenberg, Proc. Phys. Soc. London, **A65,** 385, 1952.
25. T. Mamyia, A. Oota, and Y. Masuda, Solid State Commun., **15,** 1689, 1974.
26. A. C. Anderson and S. C. Smith, J. Phys. Chem. Solids, **34,** 111, 1973.
27. A. C. Anderson and S. G. O'Hara, J. Low. Temp. Phys., **15,** 323, 1974.
28. H. Oppermann, G. Weise, J. Barthel and K. H. Berthel, Phys. Stat. Solidi, **22,** K151, 1967.
29. D. Elefant, Thesis, AdW Berlin, 1970.
30. A. C. Ehrlich and J. T. Schriempf, Solid State Commun., **14,** 469, 1974.
31. B. Schumann, Thesis, TU Dresden, 1976.
32. K. Reiche, Thesis, TU Dresden, 1978.
33. R. Bogaard and A. N. Gerritsen, Phys. Rev., **B3,** 1808, 1971.
34. A. Waleh and N. H. Zebouni, Phys. Rev., **B4,** 2977, 1971.
35. H. M. Rosenberg, Phil. Trans. Roy. Soc. (London), **A247,** 441, 1955.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 395—400 (1982)

CRYSTAL SYMMETRY EFFECTS OF THE CONDUCTION TENSOR

P. DEBRECENI, I. KIRSCHNER and T. PORJESZ

DEPARTMENT FOR LOW TEMPERATURE PHYSICS, ROLAND EÖTVÖS UNIVERSITY
BUDAPEST, HUNGARY

Some considerations are given about the Onsager–Kohler relations with respect to the linear dependence or independence of the magnetoresistivity (MR) effects. The symmetry properties of the heat conductivity tensor are investigated to elucidate a possible experimental proof of the Onsager–Kohler relations.

1. Introduction

We have measured magnetoresistivity effects on semiconductor samples at very low temperatures. We expected the MR to be a quadratic function of the magnetic field strength. But any number of experiments were carried out, we have always found a linear term in **B**, too [1, 2], which, of course, seemed to be a strange behaviour, because it must not be so according to the theoretical calculations. In the theoretical calculations the linear terms have been dropped, first owing to the effect of the crystal symmetries and secondly the unaffected parts have been ordered out by the Onsager relations or, better to say, in the case of MR by the Kohler–Onsager relations [3–4]. Onsager in his paper [3] restricts himself to heat conduction. The classical theories of heat conduction, viscosity and diffusion may be set upon a thermodynamic framework. If we introduce the entropy production σ:

$$\sigma \equiv \rho \cdot \delta - \mathbf{h}\boldsymbol{\gamma}$$

where ρ is the mass-density, δ is the internal dissipation, $\mathbf{h}$ is the heat flux, and $\boldsymbol{\gamma}$ is the gradient of the coldness per unit coldness, i.e.:

$$\boldsymbol{\gamma} = \frac{1}{\Theta} \operatorname{grad} \Theta = -\frac{1}{T} \operatorname{grad} T,$$

then:

$$\sigma \equiv \begin{cases} -\boldsymbol{\gamma}\mathbf{h} & \text{for heat conduction alone,} \\ \operatorname{tr}(\mathbf{VD}) & \text{for viscosity alone,} \\ -\rho \Sigma m \dot{\mathbf{x}} & \text{for diffusion alone,} \end{cases}$$

where $\mathbf{V}$ is the viscosity operator and $2\mathbf{D} = \operatorname{grad} \dot{\mathbf{x}} + (\operatorname{grad} \dot{\mathbf{x}})^T$ that is:

$$\sigma = \boldsymbol{\xi} \cdot \mathbf{1}.$$

The constitutive equations are of the form:

$$\mathbf{\imath} = \mathbf{\Lambda} \cdot \xi,$$

where $\mathbf{\Lambda}$ is a constant tensor. Hence σ is given by a quadratic form in ξ

$$\sigma = \sigma(\xi) = \xi \cdot \mathbf{\Lambda}\xi.$$

Onsager in his paper claimed to prove "a general class of reciprocal relations in irreversible processes from the assumption of microscopic reversibility". For the class of processes considered above, his result is

$$\mathbf{\Lambda}_{-} = 0, \qquad \text{where} \qquad \mathbf{\Lambda}_{-} \equiv \frac{1}{2}(\mathbf{\Lambda} - \mathbf{\Lambda}^T).$$

He chose the heat conductivity as a particular phenomenon and he said also that from experimental facts it is true that

$$\mathbf{K} = \mathbf{K}^T.$$

Let us see now what it really means, i.e. what kind of experiment would be suitable to prove the above relation.

2. Heat conductivity

The classical theory of heat conduction proposed by Fourier rests on the constitutive assumption that the heat flux $\mathbf{h}$ is a linear function of the temperature gradient

$$\mathbf{h} = \mathbf{K}(T, \mathbf{F}) \operatorname{grad} T. \tag{1}$$

The tensor $\mathbf{K}$ is the thermal conductivity, T is the temperature and $\mathbf{F}$ is the deformation tensor. The differential equation for the temperature field $T(\mathbf{x}, t)$ is obtained from the balance of energy

$$\rho\dot{\varepsilon} = w + \operatorname{div} \mathbf{h} + \rho s, \tag{2}$$

where ε is the specific energy, w is the stress power, and s is the body heating. This s is to be regarded as a given function of position; in most special problems it is assumed to vanish. The classical theory refers mainly to a material obeying a caloric equation of state

$$\varepsilon = \varepsilon(T, \mathbf{F}, \mathbf{x}) \tag{3}$$

and it restricts itself to static deformation. Then

$$\dot{\varepsilon} = c\,\partial_t T, \qquad \text{where} \qquad c = c(T, \mathbf{F}, \mathbf{x}) \equiv \partial_T \varepsilon,$$

$$w = 0. \tag{4}$$

If we substitute the Fourier law into the equation of balance of energy where $s = 0$, it yields

$$\rho c\,\partial_t T = \operatorname{div}(\mathbf{K} \operatorname{grad} T) = \operatorname{tr}(\mathbf{K}_{+} \operatorname{grad}^2 T) + (\operatorname{div} \mathbf{K}^T) \operatorname{grad} T, \tag{5}$$

where the symmetric part of **K** is

$$\mathbf{K}_+ = \frac{1}{2}(\mathbf{K}+\mathbf{K}^T).$$

In a homogeneous configuration of a body, c and **K** reduce to function of T alone, and (5) becomes a differential equation for the determination of T

$$\rho c \partial_t T = \operatorname{tr}(\mathbf{K}_+ \operatorname{grad}^2 T) + \operatorname{grad} T \cdot (\partial_T \mathbf{K}_+) \operatorname{grad} T,$$

$$c = c(T), \tag{6}$$

$$\mathbf{K}_+ = \mathbf{K}_+(T),$$

since F and ρ are constants. In the common applications, c and $\mathbf{K}_+$ are taken as independent of T, so that the second term vanishes (but this simplification need not be made). The differential equation as it stands, is unaffected by $\mathbf{K}_-$, the skew part of conductivity.

Therefore, if the boundary conditions are likewise independent of $\mathbf{K}_-$, we can simply assume that **K** is symmetric.

In particular, for a homogeneous rigid conductor of heat subject to prescribed temperatures upon its surface, no measurement of internal temperatures can determine whether **K** is symmetric or not.

Stokes noted in one of his papers [6] that for many types of crystalline symmetry, the skew part of **K** can be proved to vanish "and that it is only among crystals which possess a peculiar sort of asymmetry, that we should expect to find traces of their existence". Therefore he suggested that in all bodies $\mathbf{K}_- = 0$ must be true.

Resting upon the basic assumptions of the principle of material frame-indifference, which Stokes used only implicitly, it is possible to show that the principle of material symmetry is sufficent to prove that **K** is symmetric, without any further assumptions. Coleman and Noll [7] in their 'Non-Linear Field Theories' showed, that the condition which allows us to determine the restrictions imposed upon **K** is that

$$\mathbf{H}\mathbf{K}_0 = \mathbf{K}_0\mathbf{H},$$

i.e. every member **H** of the symmetry group of the material relative to a fixed reference commutes with $\mathbf{K}_0$. ($\mathbf{K}_0$ is the value of **K** in the same reference system.)

When a class of crystals is fixed, the most general symmetries of **K**'s may be determined; there are thirteen types of crystals. (The 32 elements of the crystal point group may be grouped into 13 types.)

Table I

Symmetry properties of the heat conduction tensor

Type of the crystal	Symmetry of $\mathbf{K}_0$
triclinic (C_1, C_i)	K_{ij} is arbitrary
monoclinic (C_2, C_3, C_{2h})	$(K_{ij})=\begin{bmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & e \end{bmatrix}$
tetragonal (S_4, C_4, C_{4h}) hexagonal (C_3, C_{3x}, C_{3h}, C_6, C_{6h}) transversely isotropic	$(K_{ij})=\begin{bmatrix} a & b & 0 \\ -b & a & 0 \\ 0 & 0 & c \end{bmatrix}$
rhombic (D_2, C_{2v}, D_{2h}) tetragonal (D_{2d}, C_{4v}, D_4, D_{4h}) cubic (T, T_h, T_d, 0, 0_h) hexagonal (C_{2v}, D_3, D_{3d}, D_{3h}, D_6, D_{6v}, D_{6h}) isotropic	$\mathbf{K}_0$ is necessarily symmetric

In particular, from the 32 optical crystal classes 19 are such as to force **K** to be symmetric, and only two classes, those of the triclinic system, let **K** be arbitrary; in all the others in which $\mathbf{K}_-$ need not vanish entirely, at least one of its off-diagonal components with respect to a preferred set of axes must vanish.

Now, what kind of experiment should we make in order to prove these results, or the opposite of them. Recalling Eq. (5), we shall see that by the steady distributions of temperature in a body with constant conductivity **K** satisfy

$$\operatorname{tr}(\mathbf{K}_+ + \operatorname{grad}^2 T)=0.$$

A particular solution, considered to represent a "source of heat at the origin", is

$$T=\frac{\text{const}}{(\mathbf{r}\cdot\mathbf{K}_+^{-1}\mathbf{r})^{1/2}},$$

where **r** is the position vector with respect to the origin. For this solution, the isotherms are the concentric ellipsoids

$$\mathbf{r}\cdot\mathbf{K}_+^{-1}\mathbf{r}=\text{const}$$

and

$$\operatorname{grad} T \,||\, \mathbf{K}_+^{-1}\mathbf{r}$$

and the heat flux is

$$\mathbf{h}=(\mathbf{K}_+ + \mathbf{K}_-)\operatorname{grad} T \sim \mathbf{r}+\mathbf{K}_-\mathbf{K}_+^{-1}\mathbf{r}.$$

If $\mathbf{K}_-=0$, the lines of heat flux are radial, if $\mathbf{K}_-\neq 0$ they are not; and the heat will be conducted along curved streamlines. At a calorically insulated boundary $\mathbf{n}\cdot\mathbf{h}=0$ (**n** is

the unit normal), so

$$\mathbf{n}\cdot(\mathbf{K}_+ + \mathbf{K}_-)\,\mathrm{grad}\,T=0\,.$$

In principle, the experiment goes on as follows. First we determine $\mathbf{K}_+$ (say for example by symmetrical boundary conditions, i.e. a given temperature distribution), then we insulate the boundary and determine the isotherms near it, so that the direction of grad T is known. Thus we can see whether or not

$$\mathbf{n}\cdot\mathbf{K}_+\,\mathrm{grad}\,T=0$$

there. If it is so, also

$$\mathbf{n}\cdot\mathbf{K}_-\,\mathrm{grad}\,T=0.$$

But in order to conclude that $\mathbf{K}_-=0$, we must assure that the grad T has non-zero components in at least two tangential directions for one boundary surface.

The only experiments to test the symmetry of the heat conductivity tensor were carried out by Soret, Curie and Voigt [8–12]. They did not find evidence of a symmetry. (Some comments on the experiments and what they really measured will be published later [14]). Since their measurements $\mathbf{K}_-=0$ has been regarded as an experimental fact. Although, Soret in his article [8] recognises that "... the above presented method for the experimental verification of the existence or absence of the rotational coefficients is a difficult and doubtful application ...", only Voigt declared unambiguously that the experiments had proven the theory.

3. Magnetoresistivity

Now let us return to our case, the MR effects. In our measurements the current density was maintained constant, and the electric field, which is permitted to adjust itself, was determined from measurements of potential differences. In such a case, it is desirable to express the electrical transport effects in terms of the resistivity tensor $\boldsymbol{\rho}(\mathbf{H})$ and we can write

$$\mathbf{E}=\boldsymbol{\rho}\cdot\mathbf{J}.$$

Now if we decompose the conductivity tensor into a symmetric and an antisymmetric part

$$\boldsymbol{\rho}=\boldsymbol{\rho}_+ + \boldsymbol{\rho}_-$$

and then in the light of the properties of the tensors we write down the dual tensor of the antisymmetric part of $\boldsymbol{\rho}$ say $\mathbf{b}$, we get for the electric field

$$\mathbf{E}=\boldsymbol{\rho}_+\mathbf{J}+\mathbf{b}\times\mathbf{J}.$$

This result is not a surprise at all, because the second term is simply the Hall field. But carrying out the calculations for the MR effects by this $\mathbf{E}$, and not only by the symmetric part of it, which is the usual procedure in the literature, one can get the linear

terms in the magnetic field strength for the expressions searched. Seeger also had linear terms in the expressions for the resistivity and for the mobilities, but he had to introduce a certain "inner Hall field" to achieve the results [13]. (See for example eq. 7.6.7 in Chap. 7). Examining this Hall field one can see that it is the above mentioned antisymmetric part of the conductivity tensor.

References

1. T. Porjesz, Phys. Stat. Solidi, **A22,** K145, 1974.
2. I. Kirschner, M. F. Kotkata and T. Porjesz, Ind. Journ. Techn., **18,** 112, 1980.
3. L. Onsager, Phys. Rev., **37,** 405, 1931; **38,** 2265, 1931.
4. M. Kohler, Ann. Physik, **40,** 601, 1941.
5. C. A. Domenicalli, Rev. Mod. Phys., **26,** 237, 1954.
6. Encyclopedia of Physics III/1 ed. S. Flügge, 1960.
7. ibid Vol. III/3.
8. Ch. Soret, Arch. des Sc. Phys. et Nat., **27,** 373, 1892.
9. P. Curie, Arch. des Sc. Phys. et Nat., **29,** 355, 1893.
10. P. Curie, Arch. des Sc. Phys. et Nat., **29,** 353, 1893.
11. W. Voigt, Gött. Nachr., p. 87–89, 1903.
12. W. Voigt, Lehrburh der Kristallphysik, p. 397–404., 1910.
13. K. Seeger, Semiconductor Physics, Springer Verlag, Berlin, 1973.
14. P. Debreceni, to be published.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 401—409 (1982)

THERMAL CONDUCTIVITY OF AMORPHOUS IRON ALLOYS AT LOW TEMPERATURES

G. POMPE, M. GAAFAR, M. FALZ and P. BÜTTNER

PHYSICS SECTION, DEPARTMENT OF LOW TEMPERATURE PHYSICS, TECHNICAL UNIVERSITY DRESDEN
8027 DRESDEN, GDR

The thermal conductivity of ferromagnetic amorphous iron alloys ($Fe_{40}Ni_{40}P_{14}B_6$, $Fe_{80}B_{20}$, $Fe_5Co_{75}B_{20}$) was measured in the temperature range 2 – 100 K. It was shown that there is a significant dependence of the phonon thermal conductivity on structure relaxation.

1. Introduction

This paper presents measurements of thermal conductivity of ferromagnetic amorphous iron alloys ($Fe_{40}Ni_{40}P_{14}B_6$, $Fe_{80}B_{20}$, $Fe_5Co_{75}B_{20}$) in the temperature range from 2 to 100 K [1, 2, 3].

Particularly, the influence of structural relaxation by low temperature annealing on thermal conductivity has been investigated. The thermal conductivity of the iron-boron alloy $Fe_{80}B_{20}$ has been measured in the recrystallized state.

For the investigation of the thermal conductivity of amorphous metals the phonon part of thermal conductivity is important, because the measurement of its temperature dependence can be considered as a kind of phonon spectroscopy. The main part of thermal conductivity in carried by the dominating phonons of a frequency $\omega_d \approx 4k_B T/\hbar$ (k_B is Boltzmann's constant, $\hbar$ is Planck's constant). In Table I the

Table I

Frequency and phonon wave lengths of the dominating phonons for various temperatures

T/K	ω_d/s^{-1}	$\lambda_{d/nm}$
1	$5 \cdot 10^{11}$	60
10	$5 \cdot 10^{12}$	6
100	$5 \cdot 10^{13}$	0.6

frequency of these phonons is given for various temperatures. E.g. for amorphous silica the dominating phonon wave lengths are given in the last column. Corresponding to the phonon wave length, defect structures of different dimensions determine the scattering processes at different temperatures.

From investigations of the thermal conductivity of amorphous dielectrics three peculiarities are known [4]:

1. For $T<1$ K it is observed that λ varies as $\lambda=aT^n$ ($n=2.0$). For various amorphous dieletrics the constant a is of the order $10^{-3}\ldots10^{-4}$ W/cm K^3.

2. For $5\,\mathrm{K}<T<20\,\mathrm{K}$ a so-called "plateau" of the thermal conductivity is observed.

3. For $T>20$ K the thermal conductivity increases slowly similarly as in the case of specific heat. For $T<1$ K the anomalous dependences of thermal conductivity, specific heat, velocity, and observed damping of ultrasonic waves can be explained by additional local excitations of small energy (Anderson, Halperin, Varma model [5, 6]). For the theoretical discussion of the plateau-like behaviour, a number of theoretical models exist (a modified Anderson model [7], density fluctuations after Joshi [8], the model of a constant free path after Choy [9], [10]). A common feature of all models is that they do not involve any restriction on the type of atomic interaction. Therefore it may be assumed that the phonon part of thermal conductivity of amorphous metals shows a similar temperature dependence as that of dielectrics, so far as the quasifree electrons do not strongly influence the phonon transport by phonon–electron scattering. The first investigations of the phonon thermal conductivity (also Fig. 6) metallic glasses by Matey and Anderson (1977) [11, 12, 13] yield for some alloys (Pd—Si, Ni—P) a plateau at 10 K and for some alloys only a slight increase in this temperature range (Fe—P, Fe—Ni—P).

For $T<1$ K the temperature dependence of thermal conductivity is described by $\lambda_{\mathrm{Ph}}\sim T^{1.7}$ similar to the thermal conductivity of amorphous dielectrics. It can be shown that for $T<1$ K the phonon–electron scattering can be neglected. Matey and Anderson measured no change in the thermal conductivity of titanium–beryllium alloy; the superconducting transition temperature is 0.3 K. That means that the phonon–electron scattering mechanism has no influence in this temperature range.

2. Experiments

The samples $Fe_{80}B_{20}$ and $Fe_5Co_{75}B_{20}$ were prepared by rapid quenching from the melt at the Central Institute of Solid State Physics (ZFW) Dresden, GDR. The sample $Fe_{40}Ni_{40}P_{14}B_6$ was obtained from a commercial source [14]. We used samples of a thickness of 20 to 40 μm and a width of 2 to 4 mm. The conductivity was measured by a stationary method.

$$\dot{Q}=\lambda\frac{A}{l}\Delta T=\frac{\Delta T}{W},$$

where A is the area of the samples, l is the length between the thermoelements, ΔT is the temperature difference, and W is the thermal resistivity of the sample. Owing to the

small thermal conductivity of these substances and the geometry of the samples, their thermal resistivity is very high (about 10^4 K/W) and the error from the heat flow by radiation has to be considered for $T > 80$ K ($\Delta\lambda/\lambda\,(T = 100\text{ K}) \approx 4\%$).

3. Evaluation of the phonon part of thermal conductivity

The thermal conductivity of metallic glasses can be described by

$$\lambda = \lambda_e + \lambda_{\text{Ph}}.$$

In the following discussion the magnon influence is neglected. The electron part of thermal conductivity can be separated by the well-known Wiedemann – Franz law. Under the assumption of elastic scattering processes it yields the relation

$$\frac{\lambda_e \cdot \rho}{T} = L_0 = 2.45 \cdot 10^8\, W\Omega/\text{K}^2. \tag{1}$$

The specific electrical resistivity ρ of the samples was measured in the temperature range from 2 to 300 K. A more detailed discussion of the minimum of resistivity and the temperature dependence above the minimum is given in [2, 3]. However, all these temperature effects on ρ are small in the temperature range from 2 to 100 K. $((\rho(T) - \rho(4.2\text{ K}))/\rho(4.2\text{ K})) \approx 0.7\%$ for $Fe_{80}B_{20}$, $\approx 0.4\%$ for $Fe_{40}Ni_{40}P_{14}B_6$ and $\approx 1.4\%$ for $Fe_5Co_{75}B_{20}$). Therefore, we introduce in Eq. (1) the electrical resistivity at $T = 4.2$ K and calculate the electronic part of thermal conductivity for 2 to 100 K. The high specific electrical resistivity as well as its small temperature dependence justify the use of the Wiedemann—Franz law. In Table II the parameters of the heat treatment and its influence on resistivity are presented.

Table II

Parameter of the heat treatment and its influence on resistivity

	Annealing condition		T_{cryst}/K	$\frac{\rho_{\text{ann}} - \rho}{\rho}(T = 4.2\text{ K})$
	T_a/K	t_a/min		
$Fe_{40}Ni_{40}P_{14}B_6$	373	120	≈ 655 [20]	+4%
	(in argon gases)			
$Fe_{80}B_{20}$	570	30 b)	660 [16]	
			700 [19]	≈ 0
(recryst.)	(770)	(30)		a)
$Fe_5Co_{75}B_{20}$	620	30 b)	≈ 790 [21]	−4.5%

a) As a result of recrystallization of the sample $Fe_{80}B_{20}$ the specific electrical resistivity is decreased by a factor 4 [2]

b) in vacuum

4. Phonon part of thermal conductivity

In Fig. 1 the experimental values as well as the electron and phonon part of thermal conductivity are pictured for the Fe—Ni alloy. Up to 10 K, λ_{Ph} increases with temperature. At about 20 K a plateau-like branch is visible. The phonon and electron part of thermal conductivity before and after annealing are plotted in a log—log representation in Fig. 2. By low temperature annealing the phonon conductivity increases markedly in the range of lower temperature and decreases above 40 K.

The increase of phonon conductivity at low temperature after thermal treatment was also observed for Pd—Si alloy by Matey and Anderson [11]. This behaviour corresponds to a decrease of extended structural defects connected with a relaxation of internal stresses. Within the frame of the Anderson model this means a decrease of the density of states of low energy.

The change of thermal conductivity above 40 K can be explained in the following manner. From the literature [15] it is known that during heat treatment, phosphorus segregates to clusters of the dimension of about 3 nm. At 40 K the wave length of the dominating phonons is in the order of 1 nm. It can be assumed that the scattering of these phonons at the phosphorus precipitates yields a decrease of thermal conductivity.

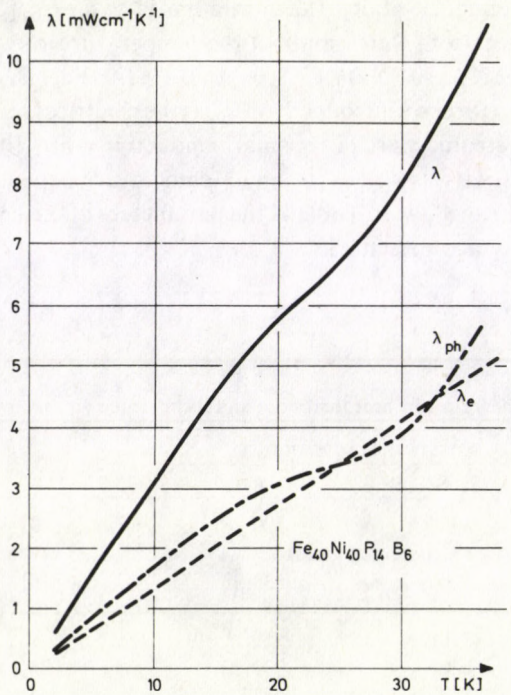


Fig. 1. Temperature dependence of experimental values, electronic and phonon part of thermal conductivity for $Fe_{40}Ni_{40}P_{14}B_6$.

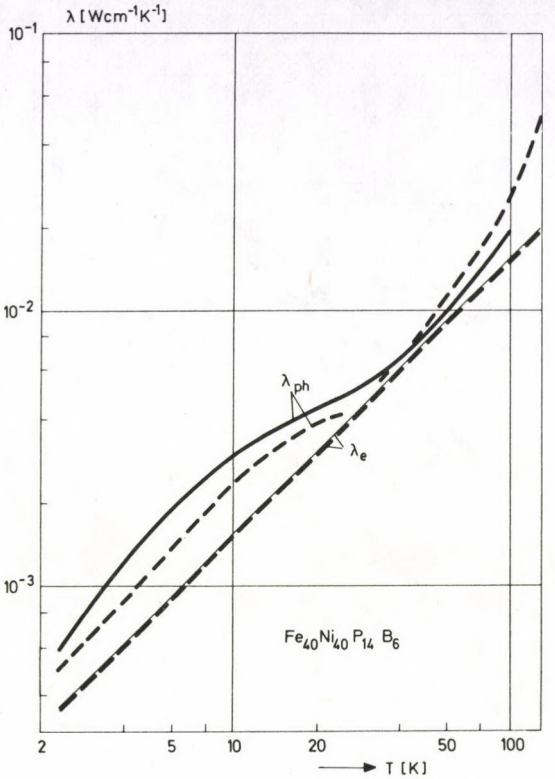


Fig. 2. Temperature dependence of electronic and phonon part of thermal conductivity before and after annealing for $Fe_{40}Ni_{40}P_{14}B_6$ – – – before and ——— after annealing.

In order to exclude the influence of phosphorus precipitates we investigated amorphous alloys with boron. In Fig. 3 the temperature dependences of λ_e and λ_{Ph} are represented in a log–log plot. These values are higher than those of the four-component Fe—Ni—P—B alloy by a factor of about 3. A branch of reduced temperature dependence is visible in the range of 40 K. At the lower part λ_{Ph} increases with T. Fig. 4 shows the change of thermal conductivity owing to heat treatment. The low temperature annealing of $Fe_{80}B_{20}$ at 570 K does not cause changes in electrical resistivity and thermal conductivity.

The phonon thermal conductivity of recrystallized materials is changed considerably, as compared with that of the initial state. For temperatures smaller than 20 K, λ_{Ph} decreases. The decrease of λ_{Ph} by recrystallization can be explained as a consequence of phonon scattering at dislocations but also as a result of the phase mixture, which is produced by recrystallization ($Fe_{80}B_{20} \rightarrow Fe_3B + \alpha - Fe$ [19]). Above 20 K the thermal conductivity of the recrystallized state is higher than that of the amorphous state. The increase shows that the mean free path of phonons with wavelengths of about 1 nm is enlarged by recrystallization. That means that the

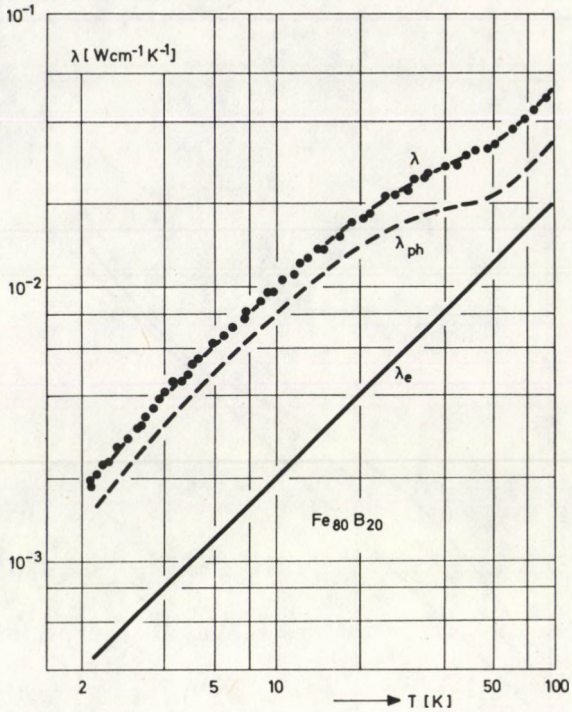


Fig. 3. Temperature dependence of experimental values, electronic and phonon part of thermal conductivity for $Fe_{80}B_{20}$.

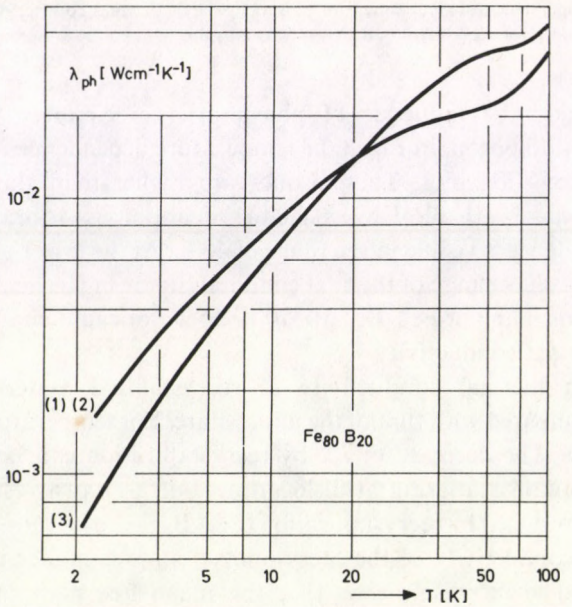


Fig. 4. Influence of the heat treatment on phonon thermal conductivity for $Fe_{80}B_{20}$: (1) before low-temperature annealing; (2) after low-temperature annealing; (3) after recrystallization.

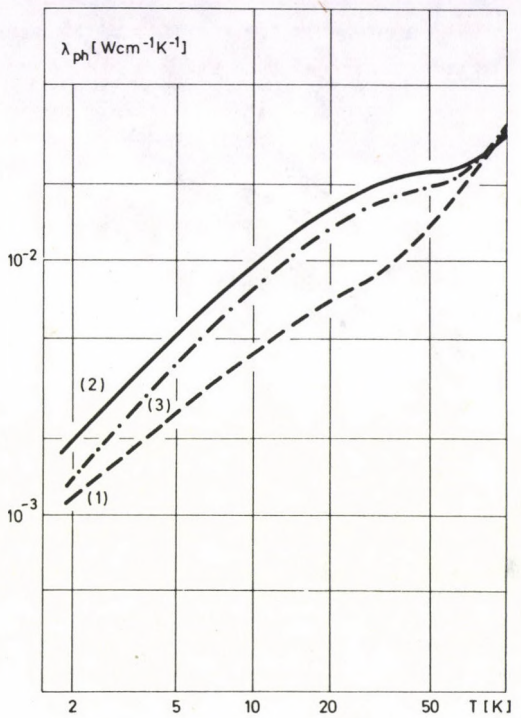


Fig. 5. Temperature dependence of phonon thermal conductivity for $Fe_5Co_{75}B_{20}$ and, for comparison, for $Fe_{80}B_{20}$: (1) $Fe_5Co_{75}B_{20}$ before low-temperature annealing; (2) $Fe_5Co_{75}B_{20}$ after low-temperature annealing; (3) $Fe_{80}B_{20}$.

number of point defects has been reduced. In Fig. 5 the influence of low temperature annealing on phonon thermal conductivity of $Fe_5Co_{75}B_{20}$ is represented. The thermal conductivity of the relaxed state shows a temperature dependence similar to the thermal conductivity of amorphous $Fe_{80}B_{20}$. That means that the structural state of alloy $Fe_{80}B_{20}$ is a stable state of amorphous structure in the case of amorphous alloys with boron as metalloid. Oshima and Fujita [22] conclude also from Mössbauer spectroscopy that in the amorphous alloy Fe—B with boron content below 20 at % exhibited well-developed short range order with the bcc-like near-neighbour configuration. Especially $Fe_{75}B_{25}$ shows a considerable fraction of chemical order of the Fe_3B type and this alloy was therefore the thermally most stable one.

In Fig. 6 all results published on the thermal conductivity of amorphous alloys with metalloid are plotted against our own results. All systems in which phosphorus is the main metalloid constituent have similar values of thermal conductivity at temperatures higher than 20 K. Below 10 K down to 1 K, the Ni—P alloy, which is not ferromagnetic, has a different temperature dependence. It is not quite sure to which extent the phonon–magnon scattering is important, and responsible for this

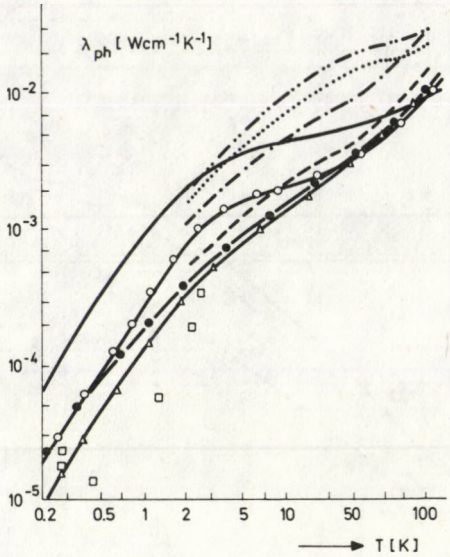


Fig. 6. Phonon thermal conductivity for metallic alloys with constituent $(TM)_{1-x}(B, P)_x$ (TM-transition metal).

—·—·—	$Fe_5Co_{75}B_{20}$ (annealed)	[3]	□□□□	$Co_{88}P_{12}$	[13]
–·–·–	$Fe_5Co_{75}B_{20}$	[3]	–○–○–	$Ni_{86}P_{14}$	[12]
·····	$Fe_{80}B_{20}$	[2]	–●–●–	$Fe_{32}Ni_{36}Cr_{14}P_{12}B_6$	[12]
———	$Pd_{77.5}Si_{16.5}Cu_6$	[11]	–△–△–	$Fe_{84}P_{16}$	[12]
– – – – –	$Fe_{40}Ni_{40}P_{14}B_6$	[1]			

difference. Measurements of the specific heat [17, 18] have shown that the magnons contribute a great part to the specific heat at temperatures 3 – 5 K. However, to clarify the influence of magnons on the heat flow in the ferromagnetic amorphous alloy requires measurements in high magnetic fields. The influence of boron is considerable, as it can be seen if we compare the alloys Fe- and Fe/Co-boron with various Fe-phosphorus alloys. The investigations of the specific heat described by Krause et al [18] also show a strong influence of the metalloid on the Debye temperature. The Debye temperatures decrease from 420 K down to 275 K if x varies from zero up to 14 in the $Fe_{80}B_{20-x}P_x$ alloy.

5. Conclusion

In conclusion we may state two facts:

— There exists a significant dependence of the phonon thermal conductivity on structural relaxation. Thus the thermodynamically more stable systems ($Fe_{80}B_{20}$) show a weaker influence of low temperature annealing than less stable systems.

— The influence of the metalloid on thermal conductivity is strong. Amorphous alloys with boron as metalloid have higher hardness, higher Debye temperature and higher phonon thermal conductivity than the alloys with phosphorus.

Acknowledgement

The authors thank Prof. Dr. habil. E. Hegenbarth and Dipl.-Phys. B. Hüttner for many useful discussions.

References

1. M. Falz. Thesis, TU Dresden, 1979.
2. M. Gaafar, Dissertation, TU Dresden, 1981.
3. P. Büttner, Thesis, TU Dresden, 1981.
4. R. Zeller and R. O. Pohl, Phys. Rev., **B4,** 2029, 1971.
5. P. W. Anderson, B. I. Halperin and C. M. Varma, Phil. Mag., **25,** 1, 1972.
6. W. A. Phillips, J. Low Temp. Phys., **7,** 351, 1972.
7. M. P. Zaitlin and A. C. Anderson, Phys. Stat. Solidi, (b), **71,** 323, 1975.
8. Y. P. Joshi, Phys. Stat. Solidi (b), **95,** 317, 1979.
9. C. L. Choy, Polymer, **18,** 984, 1977.
10. R. G. Pompe, B. Hüttner and M. Mertig, Wiss. Zeitschrift der TU Dresden, to be published.
11. J. R. Matey and A. C. Anderson, J. Non-Cryst. Sol., **23,** 129, 1977.
12. J. R. Matey and A. C. Anderson, Phys. Rev., **B16,** 3406, 1977.
13. J. R. Matey and A. C. Anderson, Phys. Rev., **B17,** 5029, 1978.
14. Allied Chemical Corp., Morristown, N. J.
15. J. L. Walter, D. G. Legrand and F. E. Luborsky, Mat. Sci. and Eng., **29,** 161, 1977.
16. U. Köster and U. Herold, Scripta Metall., **12,** 75, 1978.
17. D. G. Onn, W. D. Johnson, P. F. Gleeson, T. A. Donnelly, T. Egami and H. H. Liebermann J. Phys. C, **10,** L639, 1977.
18. J. K. Krause, T. C. Long and D. G. Onn, J. Appl. Phys., **50,** 7665, 1979.
19. T. Kemény, I. Vincze and B. Fogarassy, Phys. Rev., **B20,** 470, 1979.
20. W. Teoh, N. Teoh and S. Arajs, Proceed. Amorph. Magnet. II, p. 327, 1976.
21. S. Roth, private communication.
22. R. Oshima and F. E. Fujita, Jap. J. Appl. Phys., **20,** 1, 1981.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 411—417 (1982)

NEW PERSPECTIVE IN SEMICONDUCTOR PHYSICS BASED ON ADVANCED TECHNOLOGY

T. PORJESZ

DEPARTMENT FOR LOW TEMPERATURE PHYSICS, ROLAND EÖTVÖS UNIVERSITY
BUDAPEST, HUNGARY

Up-to-date technology provides the possibility to produce submicron semiconductor devices and highly dense device structures. The classification of the devices is based on the characteristic durations of the processes taking place in them. The dominant or the perturbative nature of the effects depends on the sizes of the system in question. The Boltzmann transport theory can fail with the decrease of dimensions. After some criticism of the fundamental assumptions the limits of the validity of the theory are marked out. The main lines of the quantum transport theory are pointed out and with the help of the superlattices very dense device arrays are described, with indications of the further development of theory and practice.

We are now in a very interesting stage of the progress of technology and science. In this paper we attempt to demonstrate this on some exciting examples, which may seem some kind of science-fiction, but they are realistic ones, based on existing experiments and technology. These examples have been selected from the topics of semiconductor physics but they can be generalized to all branches of solid state physics.

The main trend of technological progress is now to produce smaller and smaller devices. These devices are economically useful, they have important military applications and of course they are very interesting from the technical and scientific point of view as well [1, 2]. The electron-beam, the X-ray, the molecular- and the ion-beam litography provide the possibility to produce very fine device structures on semiconducting materials. First, it must be defined what is meant by very small devices or very fine structures. These are the simple experimental structures of 0.1 to 0.01 micrometers [3] as compared to bulk solid state or macroscale devices, the sizes of which are of the order of a few micrometers.

What can we do to understand what is going on inside these devices? First of all we take our old classical tool, the Boltzmann transport theory. This theory arises from the statistics of dilute gases, treating the independent particles as a dilute gas. And we assume that the electronic states are stationary in time and the momentum is well defined. Let us label with τ_D the transit time required by the electron to pass through the sample, with τ the mean free time of the moving carriers, and with τ_c the collision time, the duration of an elementary collision event.

If the inequality

$$\tau_c \ll \tau \ll \tau_D \tag{1}$$

is valid, we can use the Boltzmann transport theory. But what can happen, when the dimensions of the devices decrease further? All of the size effects, the scales and the

associated processes to be considered within these limits can be treated as if they were perturbations.

If the sizes are decreased still further, these effects lose the character of perturbation but they will dominate. In the medium scale devices the mean free time may be a significant part of the transit time and what is even more important, it may be comparable to the duration of the collision

$$\tau_c \lesssim \tau < \tau_D.$$

In this case the Boltzmann picture is still valid. If we take into account the effects to be considered later, we can use the perturbation theory.

In the very small devices of the order of about 20 nm, the transit time of the electron through the device is very short. The concept of the mean free time is meaningless because many of the electrons can pass through the device without any collision.

Before explaining how to treat these effects, I shall refer to the Kronig – Penney model, a very simple mathematical tool of growing interest in solid state physics.

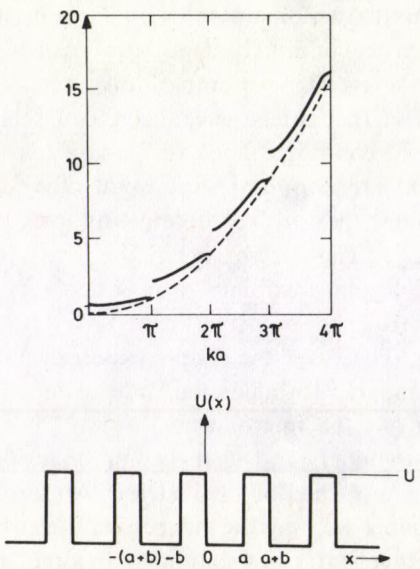


Fig. 1. The periodical square-ware potential where the atomic distance is $a+b$ and the corresponding energy-wavefunction diagram with the assumption $b=0$.

If there are atoms at the distance of K, we can apply a square-wave like periodic potential and with very simple mathematics, even if only in one dimension, the result is that we find energy bands and gaps. In the past we used this model because of its simplicity, but now it gained more importance.

In the last years a rather sophisticated equipment was built at the IBM laboratory [4]. It consists of an ultrahigh vacuum chamber, containing a sample holder and several guns for producing different nuclear beams, as well as an Auger electronmicroscope, a secondary ion mass spectroscope and an integrating mass spectrometer. The cost of this equipment is very high but it may be well worth it, if we take into account what it can produce. Epitaxial layers can be grown by means of the nuclear beam guns on a monocrystalline material. Between these layers the transition is very steep. Repeating this procedure with different impurities or different impurity concentrations results in a sandwich-structure. The thickness of one layer can be in the order of 2 – 20 nm. The potential energy profiles of the sandwich-structure of two types of semiconductors are shown in Fig. 2.

In Fig. 2 the one-dimensional Kronig – Penney model can be recognized. The dimensions of the potential energy profile and both the widths and the heights of the peaks can be chosen arbitrarily by selecting the adequate impurities or impurity concentrations. This sample gives a very good possibility to examine our theoretical tools based on the Kronig – Penney model.

An example of practical use is to take a very high impurity concentration in one of the sandwich materials resulting in a low mobility and to select the other material to be very pure. In the latter, if there are charge carriers in it at all, the mobility is very high. Because of the small dimensions, the wavefunctions of the electrons are overlapped.

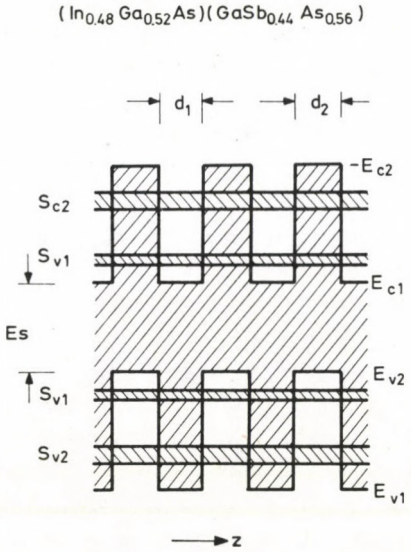


Fig. 2. Potential energy profile of InGaAs—GaSbAs superlattice.
d_1 and d_2 are the thicknesses of the two layers E_{c1} and E_{c2};E_{v1} and E_{v2} are the conduction and valence band edges of the original material;
S_{c1} and S_{c2}; S_{v1} and S_{v2} are the conduction and valence bands of the resulting superlattice; E_s is the energy overlap parameter.

That is, the carriers of the low mobility layer can populate the high mobility layer. In this way we can have a two-dimensional good conductor made of semiconductors. Its conductivity is very high because of the high purity of the first material and the high carrier concentration from the other material. This structure is called superlattice. The theoretical importance of the superlattice, of course, exceeds that of the examples mentioned here.

To find the limit of the Boltzmann theory the fundamental assumptions must be examined:

I. the collisions take place locally in time;
II. the collisions take place locally in space;
III. the electrons and the phonons are treated independently;
IV. low frequencies are considered.

The classical Boltzmann equation is

$$-\frac{f(t)}{t} = \frac{f(t)}{t_{\text{diff}}} + \frac{f(t)}{t_{\text{fields}}} + \frac{f(t)}{t_{\text{collision}}}, \tag{2}$$

where $f(t)$ is the distribution function.

With further downscaling, when the dimensions are comparable with the atomic distances, it can be expected that the features neglected in the Boltzmann theory will be considerable:

1. The scattering events are non-local in time and space.

2. The driving forces may be strong, which means that the electron can gain from the driving field an energy higher than the equilibrium thermal energy.

3. The strong scattering invalidates the assumption that the electronic states have a long life time and free-electron like character.

4. Considering a device with dimensions of an order of 10 nm at least about hundred electrons are required to have some possibility to use statistics at all, so the minimum impurity concentration must be $10^{18} - 10^{19}/\text{cm}^3$. And in a system having such high concentration, that is in a dense system, many-body effects may appear and the single-carrier description fails.

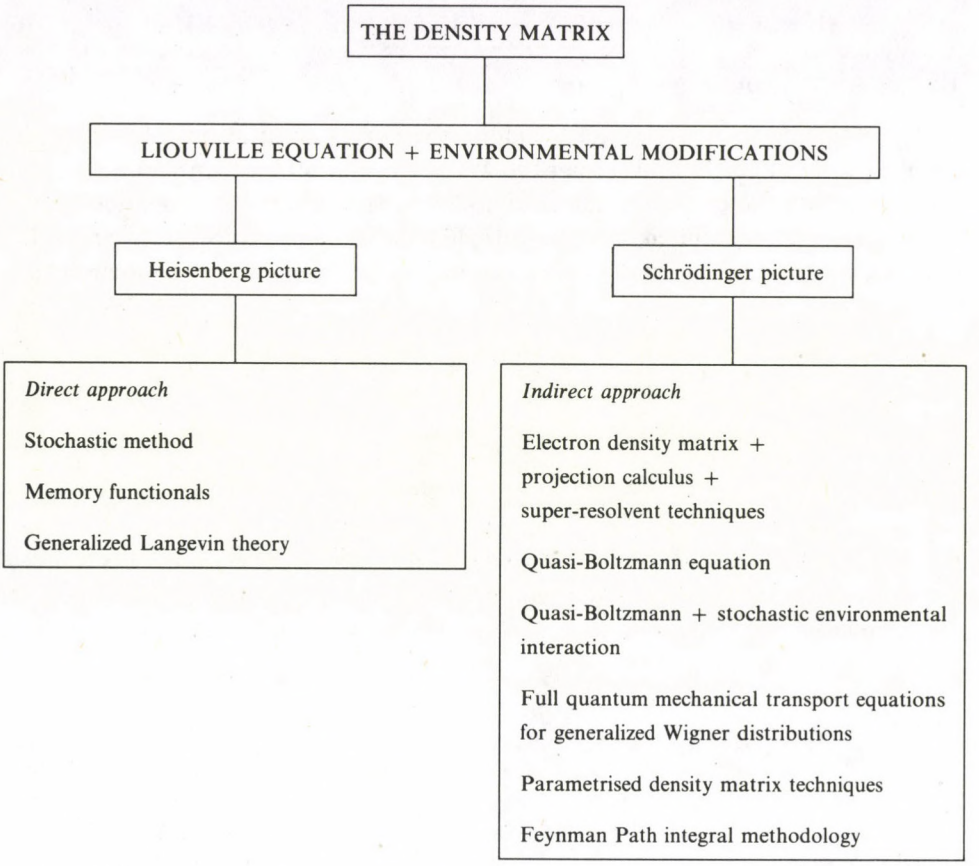
5. In small systems the influence of the size and the surface limited transport effects have to be taken into account.

6. Finally, the non-classical influence of the driving fields has to be taken into account. Sufficiently strong electric or magnetic fields deform the electronic states causing Stark and Landau quantum effects.

In the literature [5] two different ways are proposed to treat these effects. For both of them the main tool is the density-matrix method based on the Liouville equation, taking into account the environmental modifications. It means that if many devices are grown side by side, it is not only the devices that must be taken into account one by one, but the structure or lattice as a whole. The two ways for theoretical treatment are to use the Heisenberg or the Schrödinger picture, as is shown in Table I.

Table I

Two possible ways to treat the quantum transport theory



The equation for the density matrix is based on the Liouville – von Neumann equation:

$$i\hbar \frac{\partial \rho(t)}{\partial t} = \hat{H}_F \rho(t) \equiv [\hat{H}_F, \rho] \equiv \hat{H}_F \rho - \rho \hat{H}_F, \tag{3}$$

where $\hat{H}_F = \hat{H} + \hat{F}$. The Hamiltonian $\hat{H}$ describes the system in absence of any external influence. $\hat{F}$ describes the external driving forces acting from $t = 0$.

The solution for $t < 0$:

$$\rho(t < 0) = \rho, \tag{4}$$

where ρ is the thermal equilibrium solution, the grand canonical density matrix. If we want for example the current density, we have to calculate it as a quantum statistical expectation value, as the trace of a matrix

$$\langle J_i(t) \rangle = \mathrm{Tr}[\rho, J_i] \equiv \sum_{\lambda, \lambda'} \langle \lambda(J_i)\lambda' \rangle \langle \lambda'(\rho)\lambda \rangle. \tag{5}$$

Usually the Hamiltonian $\hat{H}_f$ is partitioned into "free carrier", "free scatterer", "carrier-scatterer" Hamiltonians and an operator describing the driving force:

$$\hat{H}_F = \underset{\substack{\text{free}\\\text{carrier}}}{\hat{H}_e} + \underset{\substack{\text{free}\\\text{scatterer}}}{\hat{H}_s} + \underset{\substack{\text{carrier-}\\\text{scatterer}}}{\hat{V}} + \underset{\substack{\text{driving}\\\text{force}}}{\hat{F}} \equiv \hat{H}_0 + \hat{V} + \hat{F}. \tag{6}$$

This partitioning is of course not a unique one, for example $\hat{H}_e$ might include the coupling to a magnetic field. Using the Wigner one-electron distribution function, which preserves the classical concept of a phase-space, a transport equation for the time-dependent generalized electron distribution may be constructed from the Liouville equation. In the super-operator picture the general master equation will have the form:

$$\frac{\partial f(t)}{\partial t} + i[\hat{H}_e f(t)] + i[\hat{F}, f(t)] = i\int_0^t d\tau \hat{C}_F(\tau) \cdot f(t-\tau) + \hat{M}_F(t). \tag{7}$$

collision free diffusion	acceleration of the carriers	collision effects	memory effect

If we want to describe not only one device but an array of many devices coupled into one system, we must take into account the Hamiltonians of the single devices and the Hamiltonians describing the interaction between them. That is the full Hamiltonian in the coupled N-element system is:

$$\hat{H}(1, 2, \ldots N) = \sum_{i=1}^{N} \hat{H}(i, F^i) + \sum_{i>j=1}^{N} \hat{H}'(i, j, F^k). \tag{8}$$

$\{i\}$ refer to complete sets of dynamical variables for the i-th device. $\hat{H}'(i, j, F^k)$ is the coupling between the devices i and j. Substituting the operators into (7) we have:

$$\begin{aligned} i\partial_t \rho(i) = \{\hat{H}(i) + \langle \hat{H}'(i,j)\rangle_t\}\rho(i) - \langle \hat{H}'(i,j)\rho^0(i,j,t)\rangle_t + \\ + i\int_0^t d\tau \langle \hat{H}'(i,j,t) \exp\left[-i\int_{t-\tau}^{t} d\tau' \hat{H}(i,j,\tau')\right] \hat{H}'(i,j,t-\tau)\rangle_{t-\tau} \cdot \\ \cdot \rho(i, t-\tau), \end{aligned} \tag{9}$$

where

1. $\{\hat{H}(i) + \langle \hat{H}(i,j)\rangle_t\}\rho_i$ replaces $[\hat{H}_e, f(t)]$ the dynamics is renormalised by the average interaction, if it has periodic components $\hat{H}(i)$ over-ridden by superlattice effects;

2. $\rho^v(i, j)$ describes the memory of the intial state

$$\Omega^{-1} \underset{j\ (j)}{\Sigma'} TR \ldots \rho(j, t) \equiv \langle \ldots \rangle_t;$$

3. the third term describes the generalised inter-device "scattering".

This is of course only a formal theory of coupled device-arrays, but offers an idea how to try to describe cooperative phenomena. It seems to be necessary to develop an analogue of quasi-particle theory to extract from strongly coupled device arrays an effective renormalised device array in which the individual quasi-devices are weakly-linked, making it convenient for simple analysis.

References

1. J. M. Ballantyne (Editor), Proc. NSF Workshop on Opportunities for Microstructure Science, Engineering and Technology, Cornell University Press, 1979.
2. N. G. Einspruch, (Editor) Microstructure Science, Engineering and Technology, National Academy of Science, Washington D. C. 1979.
3. A. N. Broers, J. M. E. Harper, and W. W. Molzen, Appl. Phys. Lett., **33**, 392, 1978.
4. G. A. Sai-Halasz, Proc. of 14th Int. Conf. on the Phys. of Semiconductors, p. 21. a Edinborough, 1978.
5. J. R. Barker, and D. K. Ferry, Solid State Electronics, 23, (preprint), 1980.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 419—423 (1982)

ON THE RELATION BETWEEN T_c MEASUREMENTS AND STRUCTURE INHOMOGENEITY OF Nb_3Ge SUPERCONDUCTOR

K. FRÖHLICH, D. MACHAJDÍK, Š. BEŇAČKA and V. ČERNUŠKO

ELECTROTECHNICAL INSTITUTE, SLOVAK ACADEMY OF SCIENCES
80932 BRATISLAVA, CZECHOSLOVAKIA

Transition curves of Nb_3Ge samples prepared by chemical vapour deposition were obtained by resistive and inductive methods. The differences between the values of T_c midpoints of both methods as well as the width of superconducting transition measured inductively are found to be dependent on the degree of inhomogeneity. As a measure of inhomogeneity the half-width of diffraction peak was chosen. Using the resistive method T_c onset of 23.2 K was obtained for one of the samples.

1. Introduction

Nb_3Ge is known as a material having the highest critical temperature. Its high magnetic field properties and current carrying capability are also remarkable. Some problems with preparation have arisen because of the metastable nature of the compound. Up to now, the high T_c material was synthesized successfully and reproducibly by low energy sputtering, evaporation and chemical vapour deposition (CVD) only.

We have recently reported on development of Nb_3Ge tapes by CVD method [1]. It was pointed out that the temperature gradients as well as the gradients of concentrations of gaseous species have occurred along the relatively long deposition chamber. In such a way different phases from the Nb-Ge system were deposited in adjacent parts of the reactor. As a result, a layered multiphase deposit has been formed on the moving tape. Also the distribution of lattice parameter across the A15 phase layer indicating inhomogeneity inside this phase was observed. The degree of this inhomogeneity depends on the technological conditions of tape preparation.

In this article we want to show that the inhomogeneity causes the spreading of the width of the transition from superconducting to normal state and may be responsible for the differences between the curves of transition measured resistively and inductively, respectively. The differences between both methods have been reported by several authors [2, 3, 4].

2. Experimental

Samples were prepared by chemical vapour deposition on stainless steel type substrate. The CVD process, the apparatus, and the effect of technology on the properties of Nb_3Ge superconductor were described in more detail previously [1]. The

Table I

Summary of properties of Nb_3Ge tapes: d — thickness of the layer, H — content of hexagonal Nb_3Ge_2 phase, A15 — content of A15 superconducting phase, Nb — content of Nb and NbH_x, T_{cmr} — midpoint of transition measured resistively, T_{cmi} — midpoint of transition measured inductively

Sample	d [μm]	H	A15 [wt%]	Nb	T_{cmr} [K]	T_{cmi} [K]
E-40-5	6.0	—	50	50	18.3	14.3
E-43-10	6.5	—	100	—	17.3	17.1
E-46-4	3.3	40	60	—	20.8	17.8
E-54-2	11.0	50	50	—	19.4	17.6
E-55-1	3.0	30	70	—	20.0	18.5

total average thicknesses of the deposited layers were determined from their weights. The phase analysis of samples has been carried out by X-ray diffraction using an automatic diffractometer HZG 4. The critical superconducting temperature was measured by standard four-probe d.c. resistive method [5]. Current 1 mA was passed through 15 mm long samples in this measurement. The distance between potentials was about 10 mm. A silicon diode, calibrated against a platinum thermometer has been used as a temperature sensor. Transitions to the superconducting state were studied also by the inductive method [6]. Samples of 10 mm length were put on the special planar coil of the inductance bridge. In order to ensure high sensitivity the bridge was operated at a frequency of 100 kHz. The smallest volume detectable by this type of measurement was about 350 μm^3. The maximum value of the magnetic field generated by the measuring coil was less than 10^{-8} T. The accuracy of temperature measurements by both methods was better than 0.2 K.

The results of phase analysis and T_c measurements of some samples prepared at various experimental conditions are summarized in Table I. A special technological process was used to prepare the sample E-54-2. The tape was pulled throughout the deposition chamber four times so that the layers grew one on another subsequently.

3. Discussion

Superconducting transitions of the sample E-43-10 containing only the A15 phase (within the detection limit of X-ray diffraction method) are illustrated in Fig. 1. It is seen that the curves match quite well. The differences between transition curves as well as between T_c onset values are characteristic for the sample E-54-2 (Fig. 2). Unexpectedly high T_c onset of 23.2 K was obtained for this sample using the resistive method. Both resistive and inductive measurements of this sample were carried out under the same conditions in the same probe. We suppose that the sensitivity of the resistive method may be higher whenever the thickness of the high T_c film is smaller than the penetration depth of high frequency field in this high T_c A15 phase.

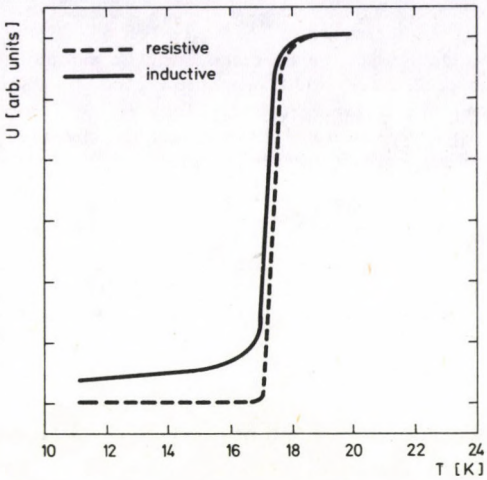


Fig. 1. Transitions from normal to superconducting state of the sample E-43-10. U — normalized output signal.

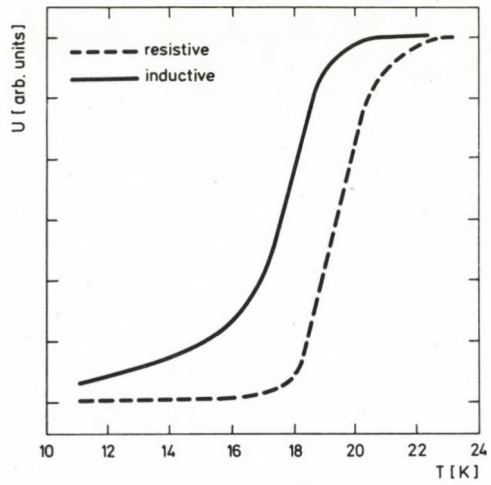


Fig. 2. Transitions from normal to superconducting state of the sample E-54-2. U — normalized output signal.

The matching of the transition curves for the sample E-43-10 implies the good homogeneity of superconducting phase in the sample. This was supported by the following study of correlation between the homogeneity of the structure and the character of transitions into superconducting state:

The analysis of structure inhomogeneity was performed by X-ray diffraction. Although the deconvolution of diffraction profile should give the best information on homogeneity of structure, the width of the diffraction could also be used for this

Table II

Correlation between structure inhomogeneity and the character of the transition into superconducting state β — half-width of diffraction peak, ΔT_{ci} — width of inductive transition defined as difference between temperature values at 90 and 10% of signal amplitude, $T_{cmr} - T_{cmi}$ — difference between midpoints of transition measured resistively and inductively

Sample	β [°]	ΔT_{ci} [K]	$T_{cmr} - T_{cmi}$ [K]
E-43-10	0.425	2.1	0.2
E-55-1	0.762	0.5	1.5
E-46-4	0.925	2.6	3.0
E-40-5	1.163	9.0	4.1

purpose. Several factors, such as the inhomogeneity of structure (A15 phase with different lattice parameter), grain size, stresses, etc., could influence the width of the diffraction peak. We suppose, with a great deal of approximation, that all these factors except inhomogeneity are constant in the studied samples so that their contribution to the changes of diffraction peak width could be neglected. Thus, the half-width β of the chosen 004 diffraction peak has been used as the measure of structure inhomogeneity. The higher values of the half-width of the peak correspond to the higher degree of structure inhomogeneity.

The character of the transition into superconducting state could be expressed by the width of inductive curve ΔT_{ci} and by differences between resistive and inductive transitions curves. These quantities are believed to reflect inhomogeneity of superconducting phase and therefore would be related to β.

The results of analysis are given in Table II. It is evident that the transition width ΔT_{ci} spreads with increasing β. One can observe that the differences between midpoints of resistive and inductive methods also arise. A deviation from this trend can be seen at the value ΔT_{ci} for the sample E-43-10. In this case the width of transition is distorted by the low temperature tail caused by the presence of some little amount of Nb on the surface of the sample. The presence of Nb was confirmed only by inductive T_c measurements.

4. Conclusion

The differences between resistive and inductive measurements of CVD prepared Nb_3Ge superconductors are pointed out. We have shown that for a sample with non-homogeneous superconducting phase not only different transition curves could be obtained but also the T_c onsets might differ. A correlation was found between the parameter of structure inhomogeneity β and the transition width of inductive method ΔT_{ci} as well as between the differences of resistve and inductive methods.

A more detailed study of the microstructure of the deposited A15 phase and the measurement of the inductive transition curve at different frequencies are necessary to

explain these discrepancies. We believe that the inductive method gives information about the homogeneity of the sample while the resistive one is convenient in searching for the highest superconducting critical temperature of a thin layer in the sample. The high T_c onset value of the sample E-54-2 demonstrates this ability.

Acknowledgements

The authors are grateful to A. Seifertová, M. Herda, A. Hrodegh and P. Štrauch for technical assistance with the experiments.

References

1. V. Černuško, K. Fröhlich, D. Machajdík and M. Jergel, IEEE Trans. Magn. MAG-**17,** 2051, 1981.
2. A. Müller, Z. Metallkde, **71,** 507, 1980.
3. E. L. Haase and O. Meyer, IEEE Trans. Magn. **MAG-17,** 541, 1981.
4. J. J. Engelhardt and G. W. Webb, Solid State Commun., **18,** 837, 1976.
5. M. Jergel, L. Janšák and M. Bláhová, Acta Phys. Slovaca, **27,** 73, 1977.
6. Š. Beňačka and V. Shabratov, to be published.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 425—431 (1982)

SOME PHYSICAL PROPERTIES OF MULTIFILAMENTARY Nb_3Sn SUPERCONDUCTORS PREPARED BY BRONZE TECHNOLOGY

J. KRUZLIAK, P. HUTKA, P. KOVÁC and P. SETINA

CENTER OF ELECTRO-PHYSICAL RESEARCH, INSTITUTE OF ELECTRICAL ENGINEERING
SLOVAK ACADEMY OF SCIENCES
84239 BRATISLAVA, CZECHOSLOVAKIA

The critical current densities and critical temperatures of multifilamentary Nb_3Sn superconductors prepared by the bronze technology were studied. It is shown that the growth rate of Nb_3Sn layer under given conditions of heat treatment depends also on the number of filaments in the wire. The critical current densities obeying 10^9 A/m^2 at 5T were obtained. The J_c values in the cross section of the Nb_3Sn layer are not constant, because the grain diameter of Nb_3Sn is not homogeneous. The influence of growth rate on T_c was also observed. The lower T_c values are obtained for higher growth rates of the Nb_3Sn layer.

1. Introduction

The compound Nb_3Sn has become increasingly important, e.g. in developing high field magnets for high energy or fusion reactor application. Filamentary Nb_3Sn superconductors are usually processed by the bronze method in which niobium rods are coreduced with copper-tin bronze matrix. In the forming process of these composites intermediate heat treatments are needed to eliminate work hardening of the bronze. The final heat treatment takes typically 50 to 200 h at 600 to 800 °C to form Nb_3Sn [1, 2, 3]. The overall critical densities of commercial Nb_3Sn superconductors usually reach the value 10^9 A/m^2 in the background magnetic field $B = 5$ T. The number of Nb_3Sn filaments of such superconductors is usually 10^3 to 10^5 with a filament diameter $2-8$ μm [4, 5].

In this paper the time dependence of the layer thickness growth upon the number of filaments was studied. The dependence of J_c vs B, heat treatment conditions and number of filaments in the wire is also shown. The values of J_c in the Nb_3Sn layer and also the T_c values were obtained. The influence of Nb_3Sn growth rate on the values of T_c is discussed.

2. Sample preparation

The 7 to 19 filaments core was fabricated using pure niobium rods which were inserted in the bronze matrix containing 11 wt% of tin. The overall bronz/niobium ratio was 3:1. These composites were then extruded by hot forward extrusion and

drawn to the wire with a final diameter of 0.4 mm. After each 40 to 60% reduction intermediate annealing was used to eliminate work hardening. The samples with 703 and 1159 filaments were bundled up from 19 filamentary cores of suitable size in the bronze container and processed by cold working in the same manner as 19 filaments conductors. The samples with a final diameter of 0.4 mm were then heat-treated to obtain Nb_3Sn in the range of 650 to 800 °C for 5 to 150 h. Short samples of a length of the order 10 to 100 m were prepared.

The thickness of Nb_3Sn layers was observed with an optical microscope and a scanning electron microscope. The Nb_3Sn layer thickness was measured by drawing a line perpendicular to a tangent at the outer periphery of the layer. The obtained thickness values were averaged over several such measurments for each sample.

The J_c measurements of Nb_3Sn superconductors were carried out by the resistive method at a temperature of 4.2 K in the applied magnetic field 3 to 6 T. A voltage drop of 1 μV/cm was used as a criterion for superconducting to normal transition. The overall critical current density and the critical current density in the Nb_3Sn layer were then calculated.

Critical temperatures of the prepared samples were measured by the resistive method using standard four-probe technique. T_c was determined as a midpoint of the resistive transition. The T_c values were obtained from $T_{c\,on} - T_{c\,low}$, where $T_{c\,on}$ and $T_{c\,low}$ are the temperatures of transition from normal to the superconducting state, respectively.

3. Results

3.1 *The growth of Nb_3Sn layer*

All annealed specimens showed Nb_3Sn formation. For a given reaction temperature, the layer thickness is given by:

$$h = \text{const.} \cdot t^n,$$

where h is the layer thickness and t is the time of reaction. Exponent n ranging from 0.35 to 1.00 has been found for the growth process of Nb_3Sn [6]. If the diffusion rate during formation of Nb_3Sn is controlled by the diffusion of tin through the previously formed compound, the growth rate should follow the parabolic law and n is 0.5 [7]. If the diffusion of tin is performed via the grain boundaries and the grain size of Nb_3Sn increases during the reaction, the growth rate will be reduced and $n < 0.5$. The time exponent n is less than 0.5 also when the tin content is lowered in the bronze matrix. The exponent n was calculated for the same 0.4 mm outer diameter of samples containing a different number of filaments for the diffusion temperature of 700 °C. The values of n for superconductors with various numbers of filaments are shown in Table I. The formation of Nb_3Sn occurs primarily by the diffusion of tin into niobium at the Nb/Nb_3Sn interface. The decrease of the time exponent n with an enlarged number of

filaments may be explained by a significant increase of the niobium-bronze interaction surface area. The amount of tin in the bronze is faster exhausted in 703 and 1159 filament superconductors and the slower growth of Nb_3Sn layer is then observed.

Table I

The dependence of time exponent n on the number of filaments

Number of filaments	7	19	703	1159
Diameter of filaments, μm	80	40	10	6
Diameter of wire, mm	0.4	0.4	0.4	0.4
time exponent n	0.36	0.36	0.29	0.25

3.2 *The measurement of critical current and critical current densities*

The dependence of overall critical current density for 7, 19, 703, 1159 filament superconductors vs magnetic field value is plotted in Fig. 1. The dependence $J_c(B)$ for 1615 filament superconductors fy Siemens is plotted too. The critical current densities linearly decrease with the increase of field. J_c of our samples are comparable to the values of superconductor fy Siemens.

The dependence of the critical current value vs reaction time and 19 filaments superconducting wire is plotted in Fig. 2. The maximum values of critical currents and the greatest increase of critical current values as a function of the reaction time were obtained on samples heat treated at a temperature of 720 °C. For 800 °C the values of I_c

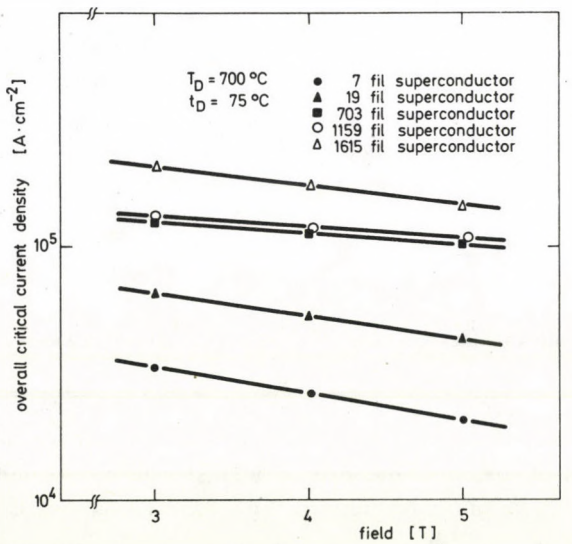


Fig. 1. Dependence of critical current densities of multifilamentary superconductors on B.

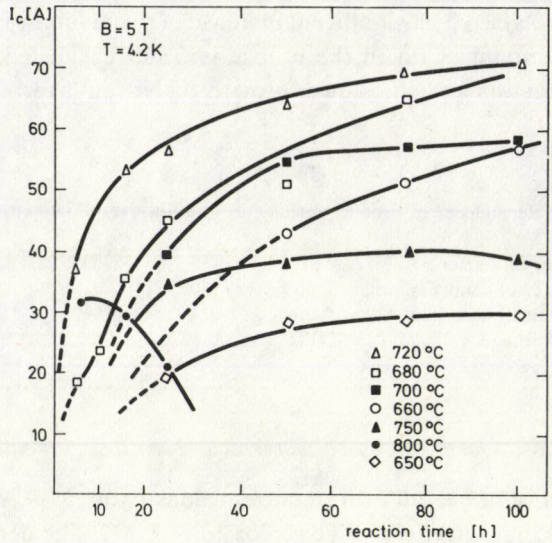


Fig. 2. Dependence of critical current 19 filaments superconductors on heat treatment time.

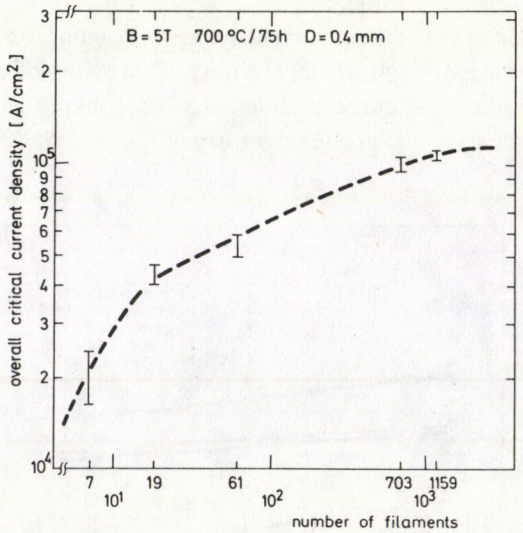


Fig. 3. Dependence of critical current densities on the number of filaments in the wires.

decrease after a short period of diffusion, which may be attributed to a significant growth of the Nb_3Sn grains and as a result a drop of pinning forces and I_c is observed.

The influence of various numbers of filaments in the wires on the overall critical current density is shown in the diagram of Fig. 3. All samples were heated under the same conditions: 700 °C – 75 h. To obtain the maximum values of J_c in our samples more than 2000 filaments are necessary.

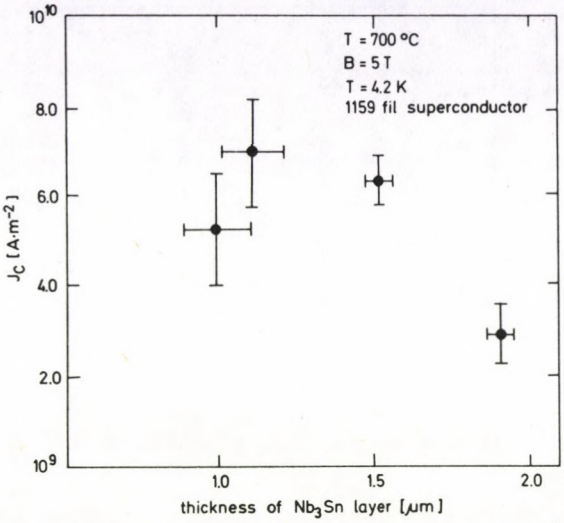


Fig. 4. Dependence of critical current densities Nb_3Sn on the layer thickness.

The dependence of critical current density in the Nb_3Sn layer on the average thickness of the layer is shown in Fig. 4. The highest values of critical current densities have layers with thicknesses about $1.0-1.5$ μm. Over this layer thickness the critical current densities decay. It is generally believed that the grain boundaries of Nb_3Sn are mainly responsible for flux-pinning in the Nb_3Sn. It was shown in [8] that the grain diameters in the cross-section of Nb_3Sn is not homogeneous. From this fact we may draw the conclusion that in the layers with thicknesses of more than 1.5 μm the grain diameter grows considerably, and, consequently, the pinning forces are lower and the J_c value decreases.

3.3 *The measurements of critical temperatures*

It was found recently that the T_c values are strongly influenced by the layer thickness. The highest T_c was observed for a Nb_3Sn layer more than 2 μm thick [9, 10]. Thinner layers, presumably less perfect, have lower T_c.

The influence of internal strains on T_c was discussed by Kwasnitza [11]. By growing the Nb_3Sn layer within the bronze matrix internal strains are developed, because the tin ions are pumped into Nb filaments. The diffusion reaction which produces the Nb_3Sn layers gives rise to as much as 38% volume expansion of Nb. As more and more Nb filaments are converted into Nb_3Sn, the layer is continually being pushed outward against the surrounding bronze matrix and thus it continues to develop increasing comprehensive strains until the ultimate yield strength of bronze is reached. This idea was supported by T_c measurements. Shorter annealing times produce a thin layer of Nb_3Sn with less internal strain and a lower number of grains, which results in higher T_c and lower ΔT_c.

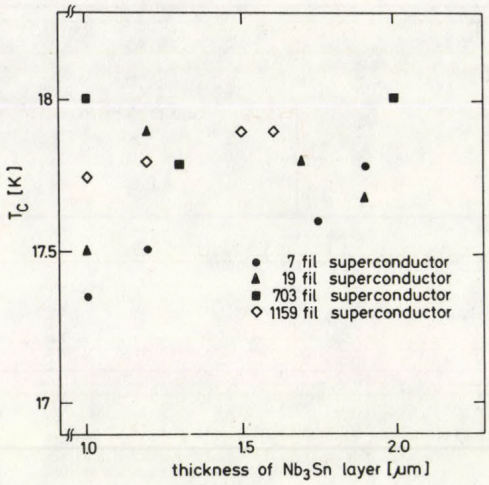


Fig. 5. Dependence of critical temperatures on the Nb_3Sn layer thickness.

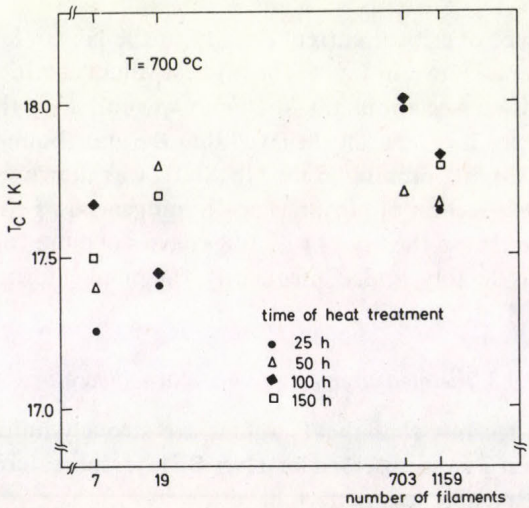


Fig. 6. Dependence of critical temperature on the number of filaments in the wire.

The transition temperature T_c as a function of average thickness Nb_3Sn is plotted in Fig. 5. The superconducting critical temperature values show no systematic dependence on the layer thickness. High T_c was observed both in the thin and in the thick layers. In addition the Nb_3Sn layers of equal thickness have various T_c. The dependence of transition temperature on the number of filaments for the preparation 700 °C is shown in Fig. 6. It is interesting that for our samples, by rising the rate of the growth layer, the critical temperatures tend to decrease slightly.

The lower T_c obtained for higher growth rate for 7 and 19 filament samples may be attributed to an insufficient integrity of the ordered structure during fast growth of Nb_3Sn layer, and/or to different influence of strains on T_c in the samples with various numbers of filaments and various growth rates of the Nb_3Sn layer. The weak variation of lattice parameter and T_c during the growth of Nb_3Sn was observed.

From our result we may conclude that the resulting T_c value of multifilament superconductors is influenced by all factors mentioned above.

4. Conclusions

Nb_3Sn multifilamentary superconductors with overall critical current densities 10^9 A/m^2 at B 5 T and T_c being 17 to 18 K were prepared.

The growth rate of Nb_3Sn for given conditions of preparation depends on the number of filaments within the wire.

It was observed that the J_c value generally decreases with increasing layer thickness.

The T_c value showed no systematic dependence upon the layer thickness.

The authors wish to thank Dr. M. Jergel for reading the manuscript and comments.

References

1. D. W. Deis, IEEE Trans. Magn. MAG-**13,** 447, 453, 1977.
2. D. L. Martin et al., IEEE Trans. Magn. MAG-**15,** 185, 1979.
3. P. E. Madsen et al, IEEE Trans. Magn. MAG-**15,** 182, 1979.
4. Prospect fy SIEMENS, 1977.
5. E. Adam and E. Gregory, IEEE Trans. Magn. MAG-**13,** 319, 1977.
6. H. H. Farell et al, Thin Solid Films, **25,** 253, 1975.
7. H. H. Farell et al, J. Appl. Phys., **45,** 4025, 1974.
8. W. Schelb, KfZ, Karlsruhe, 1980.
9. T. S. Luhman and M. Suenaga, Adv. Cryogenic Egn., **22,** 356, 1976.
10. O. Horigami et al, Appl. Phys. Lett., **28,** 738, 1976.
11. K. Kwasnitza et al., Cryogenics, 715, Dec. 1980.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 433—444 (1982)

AC LOSS MEASUREMENT OF SUPERCONDUCTING MAGNETS. INFLUENCE OF THE COMPENSATING SCHEME

F. GÖMÖRY

INSTITUTE OF ELECTRICAL ENGINEERING, SLOVAK ACADEMY OF SCIENCES
809 32 BRATISLAVA, CZECHOSLOVAKIA

The paper deals with the errors, involved in the loss measurement of superconducting magnets by the compensating scheme in Wilson's method. The expressions for determining the experimental error, depending on the network parameters, have been derived. Some results of calculations are presented for various values of compensating scheme elements.

1. Introduction

One of the basic characteristics of the superconducting magnets for pulsed purposes (e.g. particle accelerators) is the energy dissipated during the cycle. This energy (named AC losses) determines the essential load of the cooling system. It is necessary to check this quantity in the assembled magnet experimentally.

This paper deals with the errors, involved in the loss measurement by the compensating scheme in the method proposed by Wilson [1]. Assuming that the apparatus is a linear system, the transform analysis has been carried out. The expressions for determining the experimental error, depending on the network parameters, have been derived. Some results of calculations are presented for various values of compensating scheme elements.

2. Analysis of the measuring circuit

The arrangement for AC loss measurement [1] is shown in Fig. 1. During the magnet current cycle the loop is recorded on the XY plotter, the area of this loop is proportional to the AC loss per cycle:

$$\alpha = \frac{F}{S_x S_y \tau_I} \int\limits_{0 \to I_{max} \to 0} di \int\limits_0^t u(t')dt' = \frac{F}{S_x S_y \tau_I} W, \tag{1}$$

where the meaning of the symbols is the following:

α: area of the loop [cm^2],

S_x, S_y: sensitivities of the recorder on the X and Y axes [V/cm], [A/cm],

τ_I: time constant of the integrator [s],

F: voltage divider coefficient,

i: magnet current [A],

u: voltage across the magnet [V],

w: AC loss per cycle [J].

Substituting the real elements of the circuitry by the ideal ones is essential for our further analysis. In this paper, the following substitutions have been made:

a) The magnet is represented by its self inductance L_M.

b) The compensating coil has mutual inductance to the magnet current circuit M, self inductance L_{cc}, resistance R_{cc} and capacitance C_{cc}.

c) The integrator consists of the input resistance R_I, operational amplifier with the constant gain — A and integrating capacity C_I in the feedback loop. The operational amplifier input current is assumed to be zero and the gain is taken as a large negative number.

The analysis is carried out by means of the Laplace integral transform [2], where the time function $x(t)$ has the operator representation

$$X(p) = \int_0^\infty x(t) \exp(-pt) dt$$

and the inverse transform for obtaining $x(t)$ is

$$x(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} X(p) \exp(pt) dp,$$

where j is the imaginary unit.

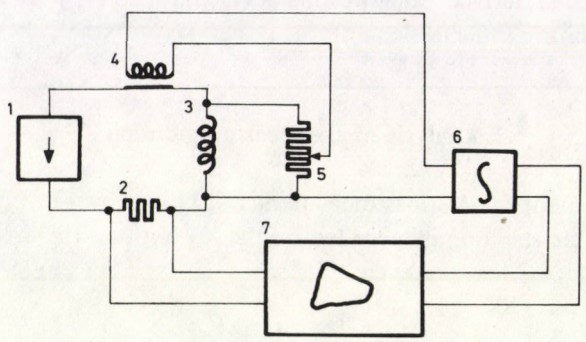


Fig. 1. Apparatus for AC loss measurement in superconducting magnets [1]. 1 — power supply, 2 — shunt, 3 — superconducting magnet, 4 — compensating coil, 5 — voltage divider, 6 — electronic integrator, 7 — XY plotter.

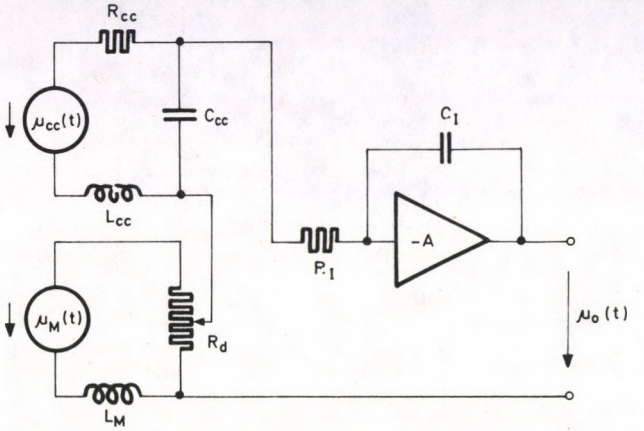


Fig. 2. Equivalent scheme for integrating the magnet voltage.

In the following we shall use assignments for operator representation and time function similarly as in the case of $X(p)$, $x(t)$.

For the part of circuitry made up of items 3, 4, 5, 6 in Fig. 1 the equivalent network is shown in Fig. 2. R_d is the magnet voltage divider with the coefficient

$$F = \frac{R_1}{R_1 + R_2}.$$

There are two sources of signal in the network: the magnet and the compensating coil. The voltages, generated during a cycle in these two elements, are designated by $u_M(t)$ and $u_{cc}(t)$. The output voltage of the integrator is $u_0(t)$. The relation between the operator representations of the voltages can be found in the form:

$$U_0(p) = \frac{U_{cc}(p)\,(p\tau_M + 1) + F\,U_M(p)\left(\dfrac{p^2}{\omega_{cc}^2} + p\tau_{cc} + 1\right)}{a_4 p^4 + a_3 p^3 + a_2 p^2 + a_1 p + a_0}. \tag{2}$$

The denominator polynomial coefficients are given as

$$a_4 = -\frac{\tau_M}{\omega_{cc}^2}(F\,\tau_d + \tau_I),$$

$$a_3 = \frac{\tau_M}{A\omega_{cc}^2} - \frac{F(1-F)\tau_d + \tau_I + \beta_c\tau_M}{\omega_{cc}^2} - \tau_{cc}\tau_M(\tau_I + F\,\tau_d),$$

$$a_2 = \frac{1}{A\omega_{cc}^2} + \frac{\tau_{cc}\tau_M}{A} - \tau_{cc}[F\tau_d(1-F) + \tau_I + \beta_c\tau_M] - \frac{1}{\omega_{cI}^2} - \tau_M(\tau_I + F\,\tau_d),$$

$$a_1 = \frac{\tau_{cc} + \tau_M}{A} - \tau_I(1 + \beta_R) + \tau_d F(F-1),$$

$$a_0 = \frac{1}{A},$$

where

$$\tau_{cc}=R_{cc}C_{cc}, \qquad \omega_{cc}^2=\frac{1}{L_{cc}C_{cc}},$$

$$\tau_M=\frac{L_M}{R_d}, \qquad \omega_{cI}^2=\frac{1}{L_{cc}C_I},$$

$$\tau_I=R_IC_I, \qquad \beta_R=\frac{R_{cc}}{R_I},$$

$$\tau_d=R_dC_I, \qquad \beta_c=\frac{C_I}{C_{cc}}.$$

In the ideal case, when the integrator output signal is just the integral of the voltages from the compensating coil and the magnet (the latter multiplied by F), the relation between $U_0(p)$, $U_{cc}(p)$ and $U_M(p)$ has to be

$$U_0(p)=\frac{U_{cc}(p)+F\,U_M(p)}{-\tau_I p}. \tag{3}$$

The last expression is a particular case of (2) for $\tau_M\to 0$, $\omega_{cc}^2\to\infty$, $\tau_{cc}\to 0$, $A\to\infty$, $\omega_{cI}^2\to\infty$, $\beta_R\to 0$, $\tau_d\to 0$. Some of these conditions can be fulfilled simultaneously, but some of them are controversial. Therefore, we have to use (2) for every particular case.

3. The evaluation of the measurement error

In following we suppose that:

a) The integrator output voltage drift is negligible.

b) The integration with respect to i in (1) is perfect. These assumptions are not so strange, because

ad a) there exist several types of chopper stabilized amplifiers with very low drift, and a sufficient correction can be made supposing the drift of the output signal to be linearly increasing during the cycle,

ad b) recorders with good dynamic properties are available, or the integration without an XY plotter can be used [3], [4].

Then the measured losses are:

$$W^R=\int_{0\to I_{\max}\to 0} u_0(t)di. \tag{4}$$

The time dependence of the magnet current

$$i(t)=\begin{cases}\dfrac{I_{\max}}{T}t & 0\leqq t\leqq T\\ \left(2-\dfrac{t}{T}\right)I_{\max} & T\leqq t\leqq 2T\end{cases}$$

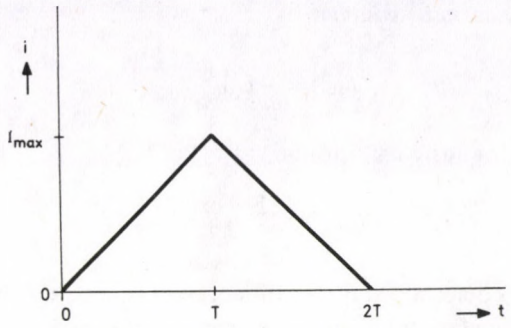


Fig. 3. Magnet current versus time plot.

is shown in Fig. 3. Then, the Eq. (4) can be rewritten as

$$W^R = \frac{I_{max}}{T}\left\{\int_0^T u_0(t)dt - \int_T^{2T} u_0(t)dt\right\}. \tag{5}$$

For solving (5) we need to insert expressions for $U_{cc}(p)$ and $U_M(p)$ into (2) and find $u_0(t)$ by the help of inverse transform.

3.1 *Compensating coil voltage* $u_{cc}(t)$

The transform of magnet current is

$$I(p) = \int_0^\infty i(t)\exp(-pt)dt = I_{max}\frac{(1-\exp(-pT))^2}{p^2T}.$$

The voltage generated across the compensating coil is

$$u_{cc}(t) = M\frac{di}{dt}$$

and thereby

$$U_{cc}(p) = MpI(p) = M\frac{I_{max}}{T}\frac{(1-\exp(-pT))^2}{p}. \tag{6}$$

3.2. *Magnet voltage* $u_M(t)$

The voltage signal from the magnet has the inductive and loss components

$$u_M(t) = u_i(t) + u_L(t).$$

The inductive component $u_i(t)$ is given by

$$u_i(t) = -L_M \frac{di}{dt}$$

and its transform analogously as in (6) is

$$U_i(p) = -L_M \frac{I_{\max}}{T} \frac{(1-\exp(-pT))^2}{p}. \tag{7}$$

The loss voltage $u_L(t)$ is determined by the dissipative processes in the magnet and is to be found for every particular case using an appropriate theoretical model. To deal with the effect of compensation, we introduce preliminarily $U_L(p)$ as the transform of $u_L(t)$.

The compensating coil is used to generate the inductive signal, which can cancel the inductive component of magnet voltage. Because the inductive to loss voltage ratio is of the order of 1000, the sensitivity of the measurement is increased in this way. To avoid the awkward construction of the coil, only a fractional part $\left(\text{approximately } \frac{1}{10}\right)$ of the magnet voltage is taken.

The condition of perfect compensation is

$$M = F L_M. \tag{8}$$

Comparing (6), (7) and (8) one can show that

$$U_{cc}(p) = -\frac{M}{L_M} U_{Mi}(p) = -F U_{Mi}(p)$$

and (2) can be written as

$$U_0(p) = F \frac{U_{Mi}(p)\left[\frac{p^2}{\omega_{cc}^2} + p(\tau_{cc}-\tau_M)\right] + U_L(p)\left[\frac{p^2}{\omega_{cc}^2} + p\tau_{cc} + 1\right]}{a_4 p^4 + a_3 p^3 + a_2 p^2 + a_1 p + a_0}. \tag{9}$$

For the inverse transform we need to know the roots of the denominator. It is not difficult to find them numerically and the denominator can be taken in the form:

$$a_4 p^4 + a_3 p^3 + a_2 p^2 + a_1 p + a_0 = D(p-p_1)(p-p_2)(p-p_3)(p-p_4), \tag{10}$$

where p_1, p_2, p_3, p_4 are the roots.

Then, it is useful to divide (9) into two parts:

$$U_0(p) = U_{0i}^R(p) + U_{0L}^R(p) =$$

$$= F \frac{U_{Mi}(p)\left[\frac{p^2}{\omega_{cc}^2} + p(\tau_{cc}-\tau_M)\right]}{D(p-p_1)(p-p_2)(p-p_3)(p-p_4)} + F \frac{U_L(p)\left[\frac{p^2}{\omega_{cc}^2} + p\tau_{cc} + 1\right]}{D(p-p_1)(p-p_2)(p-p_3)(p-p_4)}. \tag{11}$$

That means, the voltage $u_0(t)$ can be divided into two components $u_{0i}^R(t)$ and $u_{0L}^R(t)$, and according to (4) we consider the measured loss to consist also of two parts:

$$W^R = W_i^R + W_L^R .$$

The first term on the right-hand side represents the value, which would occur in the measured losses even in the lossless magnet, meanwhile the second one is referred to the loss processes themselves.

For the determination of W_i^R, we find the voltage $u_{0i}^R(t)$ by the inverse transform. Inserting the result into (5) then yields

$$W_i^R = -\frac{F L_M}{D}\left(\frac{I_{\max}}{T}\right)^2 \sum_{i=1}^{4} \frac{\dfrac{1}{\omega_{cc}^2} + \dfrac{\tau_{cc}-\tau_M}{p_i}}{\prod\limits_{\substack{j=1 \\ j\neq i}}^{4} (p_i - p_j)} \left[-\exp(2p_i\tau) + 4\exp(p_i\tau) - 3 \right]. \quad (12)$$

In the ideal apparatus (3), this term should be zero.

The area of the loop, displayed on XY plotter, is proportional to W^R. The actual losses are given as

$$P^R = W^R \frac{\tau_{\text{int}}}{F},$$

where the time constant of the electronic integrator is usually taken as

$$\tau_{\text{int}} = R_I C_I .$$

Therefore, the error in the result, caused by $u_{0i}^R(t)$, is

$$P_i^R = W_i^R \frac{\tau_{\text{int}}}{F}, \quad (13)$$

with W_i^R given by (12). After inserting the parameters of the analysed apparatus into (10), we can compute P_i^R using the expressions (12) and (13).

Now, we shall determine the error, referred to the second term in (11). It can be carried out by a similar procedure, but we have to set $u_L(t)$ at the beginning. This voltage is composed of several signals $u_{L\,1}(t)$, $u_{L\,2}(t)$, ... $u_{L\,k}(t)$ depending on the number of the loss processes. For every considered process, we determine $u_{L\,j}(t)$, its transform $U_{L\,j}(p)$, then $u_{0Lj}^R(t)$ from the inverse transform of the second term in (11). Finally, we find $W_{L\,j}^R$ by using (5) and compare it with the theoretical value $W_{L\,j}^I$, which should be measured on the ideal apparatus (3). The relative error in measured value of the losses, corresponding to the considered process, is

$$\Delta P_{L\,j} = \frac{W_{L\,j}^R - W_{L\,j}^I}{W_{L\,j}^I}. \quad (14)$$

As an example, we show the producere outlined here for the case of eddy current losses in a multifilamentary twisted composite. In the notation it is labelled by index e instead of general L_j. The volume unit of such conductor corresponds [5] to the secondary circuit in Fig. 4, where the magnet winding represents the primary circuit,

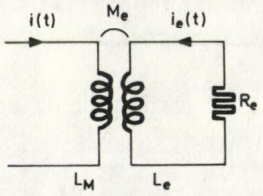


Fig. 4. Model circuit for the simulation of eddy currents.

and the time constant of the secondary circuit equals the time constant of the magnetic flux diffusion in the multifilamentary conductor [6]

$$\tau_e = \frac{u_0}{\rho_{\text{eff}}}\left(\frac{l_p}{2\pi}\right)^2,$$

$$u_0 - 4\pi \cdot 10^{-7}\ \text{H/m},$$

where l_p is the twist pitch of the filaments and ρ_{eff} is the effective matrix resistance. The voltage, induced back to the primary winding by the secondary current, is

$$u_e(t) = \frac{M_e^2 I_{\max}}{L_e T}\left\{1_{(t)}\exp\left(-\frac{t}{\tau_e}\right) - 1_{(t-T)}2\exp\left(-\frac{t-T}{\tau_e}\right) + \right.$$

$$\left. + 1_{(t-2T)}\exp\left(-\frac{t-2T}{\tau_e}\right)\right\}, \tag{15}$$

where

$$1_{(t)} = \begin{cases} 1 & t \geqq 0, \\ 0 & t < 0. \end{cases}$$

In the case of the ideal apparatus (3), we should measure the eddy current losses

$$W_e^I = \frac{F}{\tau_{\text{int}}}\frac{M_e^2}{L_e} T\tau_e\left(\frac{I_{\max}}{T}\right)^2\left\{2 + \frac{\tau_e}{T}\left(-3 + 4\exp\left(-\frac{T}{\tau_e}\right) - \exp\left(-\frac{2T}{\tau_e}\right)\right)\right\}. \tag{16}$$

Setting the Laplace transform of (15) to (11) in the place of $U_L(p)$ and solving the inverse transform yields the expression for measured value of eddy current losses

$$W_e^R = \frac{F M_e^2}{DL_e}\left(\frac{I_{\max}}{T}\right)^2 \sum_{i=1}^{5} \frac{\dfrac{p_i}{\omega_{cc}^2} + p_i\tau_{cc} + 1}{\prod\limits_{\substack{j=i \\ j\neq i}}^{5}(p_i - p_j)} \cdot$$

$$\cdot \frac{1}{p_i}(3 - 4\exp(p_i T) + \exp(2p_i T)), \tag{17}$$

where p_1, p_2, p_3, p_4, have the same significance as in (11) and the fifth root is $p_5 = -\frac{1}{\tau_e}$.

The relative error of measuring the eddy current losses for a process with time constant τ_e is given by comparing (16) with (17):

$$\Delta P_e = \frac{W_e^R - W_e^I}{W_e^I}. \tag{18}$$

4. The selection of the apparatus parameters

Knowing the magnet self inductance L_M, we can compute the error for given current cycle and measurement circuitry parameters, caused by the first term in (11); we call this error "inductive voltage error", because the inductive voltage is responsible for its appearance. However, one might have in mind that it is a value in Joules, added to or subtracted from (depending on the sign) the correct value. The "inductive voltage error" dependence on the magnet current ramp rate for several parameters of compensating scheme is shown in Fig. 5. The dependence is linear in all presented cases,

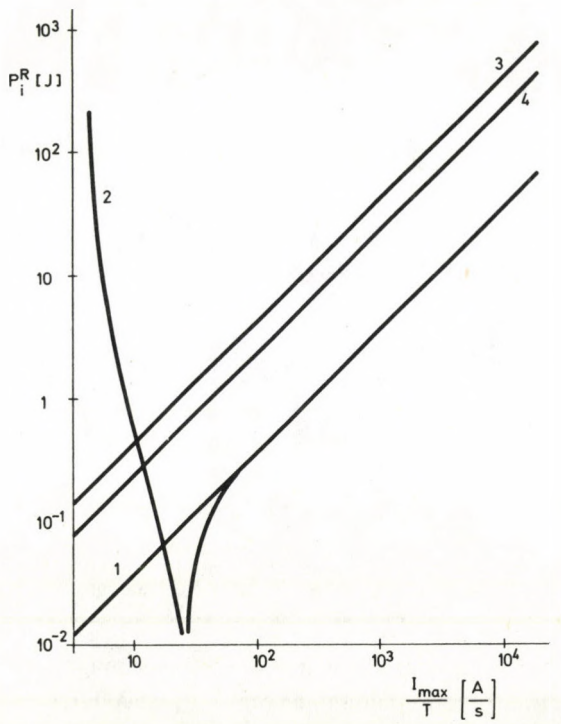


Fig. 5. The ramp rate dependence of "inductive voltage error"
1 — $C_{cc} = 2.10^{-8}$ F, $L_{cc} = 1$ H, $R_{cc} = 10^3$, $C_I = 10^{-6}$ F, $R_I = 2.10^5$, $R_d = 10^3$, $A = 10^6$; 2 — the same as 1, except $A = 10^3$; 3 — the same as 1, except $R_{cc} = 10^4$; 4 — the same as 1, except $C_{cc} = 10^{-7}$ F.

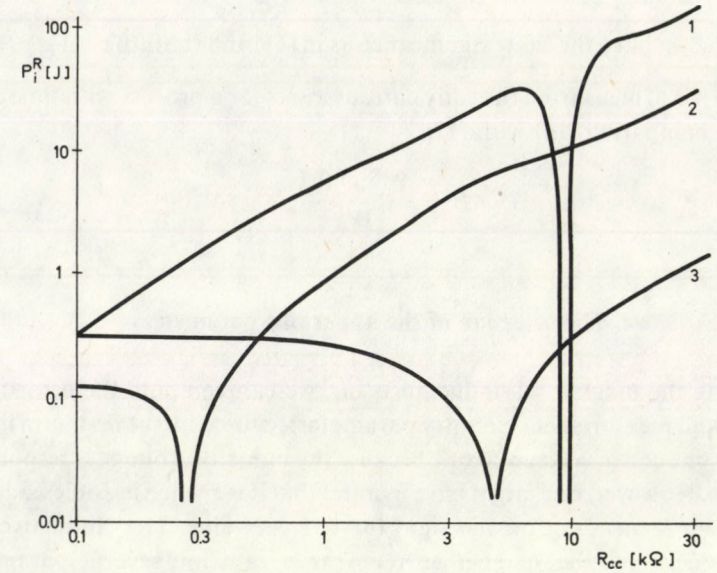


Fig. 6. The "inductive voltage error" dependence on R_{cc}. $1 - C_{cc} = 10^{-7}$ F, $2 - C_{cc} = 2.10^{-8}$ F, $3 - C_{cc} = 10^{-9}$ F, ramp rate 250 A/s, other parameters the same as in Fig. 5 case 1.

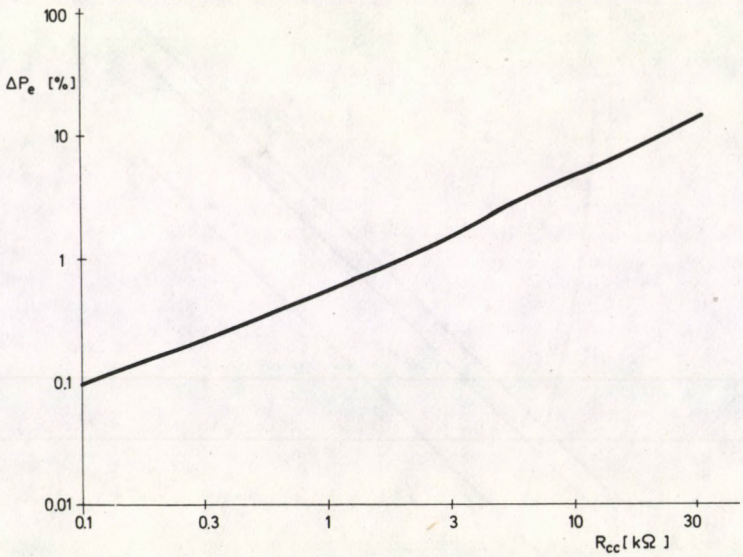


Fig. 7. The eddy current loss measurement error dependence on R_{cc}, ramp rate 250 A/s, other parameters the same as in Fig. 5 case 1.

except the case of insufficient amplifier gain $A = 1000$. The magnet self inductance is set to be $L_M = 5.10^{-3}$ H and the maximum current $I_{max} = 5000$ A.

The dependence on the compensating coil resistance R_{cc} is given in Fig. 6 for three different coil capacitances C_{cc}. The ramp rate is considered to be 250 A/s. As it can be seen from this Figure, in principle the error increases with the growth of R_{cc} and C_{cc}.

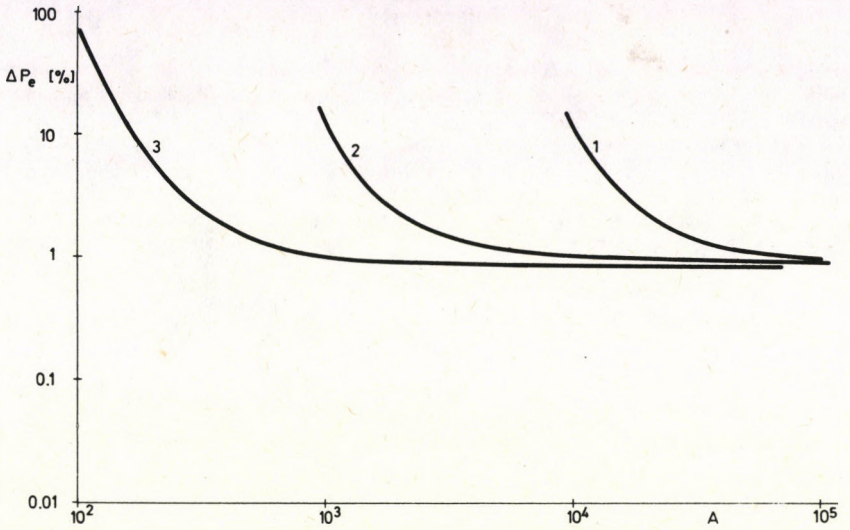


Fig. 8. The eddy current loss measurement error dependence on amplifier gain A. Ramp rate 250 A/s, other parameters the same as in Fig. 6.

The calculations of the error (18) involved in the measured eddy current losses show that it practically does not depend on the time constant in the wide range of τ_e from 10 to 10^{-8} s. In Fig. 7, the dependence of this type of error on the compensation coil resistance R_{cc} is presented. It is close to linear and does not depend on the current ramp rate. On the other hand, the dependence on the amplifier gain A exhibits large changes with respect to the ramp rate, as shown in Fig. 8.

5. Conclusions

From several examples in the previous Section it is clearly seen that the estimation of the error is of great importance for real apparatuses. Since there may be a strong dependence on some parameters like τ_e in our example in several cases of the loss processes, these parameters can be determined from short sample measurements. In analogy to the eddy current losses, it is possible to estimate the resulting error in measuring the AC losses of the magnet by a relation similar to (16).

Acknowledgement

The author would express his thanks to Dr. L. Cesnak and Dr. S. Takács for helpful discussions.

References

1. M. N. Wilson, Cryogenics, **13,** 361, 1973.
2. E. V. Bohn, The Transform Analysis of Linear Systems, Addison-Wesley Publishing Company, Inc., Reading, 1963.
3. L. Jansak and F. Chovanec, Cryogenics, **20,** 125, 1980.
4. F. Gömöry, A. N. Jerokhin, S. J. Jershov, L. S. Shirshov and G. Enderlein, Prib. Techn. Eksp., 1982, No. 3. p. 70.
5. S. S. Shen and R. E. Schwall, Proceedings of the 7th ICEC, Guilford, 1978, p. 659.
6. M. N. Wilson, C. R. Walters, J. D. Lewin and P. F. Smith, J. Phys. D, **3,** 1517, 1970.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 445—453 (1982)

SUPERCONDUCTING LABORATORY MAGNETS

J. BÁNKUTI and L. LÁSZLÓFFY

DEPARTMENT FOR LOW TEMPERATURE PHYSICS, ROLAND EÖTVÖS UNIVERSITY
BUDAPEST, HUNGARY

Superconducting magnets were constructed for the different investigations of magnetic phenomena. The magnetic field of this series ranges from 1,5 k0e to 72,6 k0e. By the volume minimalization of each magnet we have developed these with little helium consumption. In some cases our aim was to reach high homogeneity using a graphical interpolation technique with computer.

Introduction

In our research work homogeneous superconducting laboratory magnets were needed for the investigations of the properties of superconducting materials, the analysis of the new, type I—II phase transition of superconducting alloys with small Ginzburg – Landau parameter, the study of the behaviour of chemical reactions in a high magnetic field and for other purposes.

The superconducting magnets designed and constructed by us during the last few years are listed in Table I.

In the design of a superconducting magnet the size of the necessary working chamber and the available cryostat volume were taken into account. The consumption of liquid helium was reduced by the volume of each magnet [1].

Special requirements were a horizontal field direction, high homogeneity, a solenoid bore for high pressure multiplicator and a chemical working chamber, at the Helmholtz's coils, a highly homogeneous solenoid, a magnet for high pressures and a magnet for oriented reactions.

Low field and high field solenoids were also produced for the investigations which do not require homogeneity and a cryomagnet made of copper was constructed for different educational purposes.

Construction of superconducting magnets

1. *High field solenoid*

The high field superconducting solenoid (Fig. 1) was made of a Nb—Ti—Zr superconducting wire of a diameter of 0.35 mm. The magnetic field can be regulated from 0 up to the 48.5 k0e maximal field intensity at a current 29.6 A [2]. The

Table I

Constructed superconducting magnets

	H_{max}[k0e]	Homogeneity [cm^{-1}]	Material	Diameter [mm] Outer	Inner	Length [mm]	Year
High field solenoid	48.5	$5 \cdot 10^{-3}$	Nb—Ti—Zr	70	25	100	1969
Magnet for high pressures	45.2	10^{-3}	Nb_3Sn	66.5	25.6	88.2	1971
Helmholtz's coil	5.5	$5 \cdot 10^{-2}$	Nb—Ti—Zr	46	25	55	1973
Homogeneous solenoid	10.9	$2 \cdot 10^{-5}$	Nb—Ti	33	25	105	1975
Low field solenoid	1.5	$5 \cdot 10^{-3}$	Nb—Ti	10.5	7	98	1976
Cryomagnet	25.4	10^{-5}	Cu	108	28	172	1977
Magnet for oriented reactions	72,6	10^{-5}	Nb—Ti	180	40	146	1978

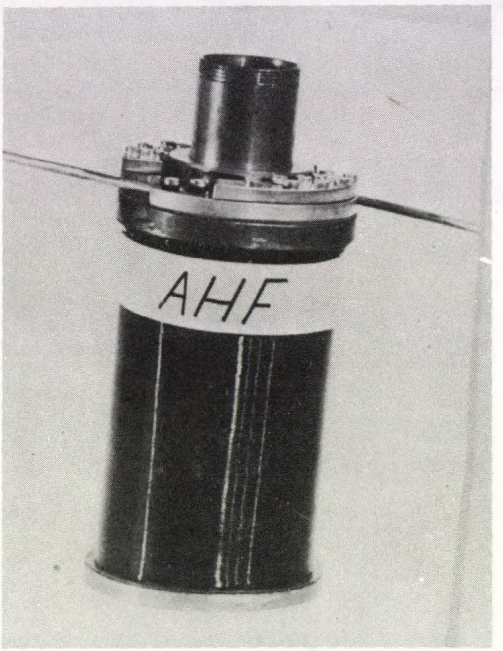


Fig. 1

homogeneity was better than 5×10^{-3}/cm along the axis of the solenoid. The superconducting core was surrounded by copper shieldings and between two layers there were plastic foils of a thickness of 25 μm to ensure insulation. The magnet has an effective inner diameter of 25 mm, an outer diameter of 70 mm and an axis length of 100 mm.

2. *Magnet for high pressures*

The high pressure solenoid (Fig. 2) is suitable to hold a high pressure multiplicator. The good mechanical stability and the minimal geometric sizes of the solenoid were achieved by an appropriate wall thickness to compensate the radial magnetic forces. The axial rigidity of the magnet was ensured by outer stiff (iron) sticks. Using the solenoid at a moderate field the sticks are not needed and a more favourable helium consumption can be achieved [3].

The magnet was made of Nb_3Sn multifilament superconducting wire. A maximal magnetic field of 45.2 k0e can be reached at a homogeneity of 10^{-3}/cm in the central part. The insulation between the layers is ensured by tereftalat foils of a thickness of 20 μm. For the protection against the transition to normal state we applied Al-foil after every second layer to decrease inductivity and to prevent rapid changes in the magnetic field. The geometrical sizes of the magnet are: length: 88.2 mm, outer diameter: 66.5 mm, bore: 25.6 mm.

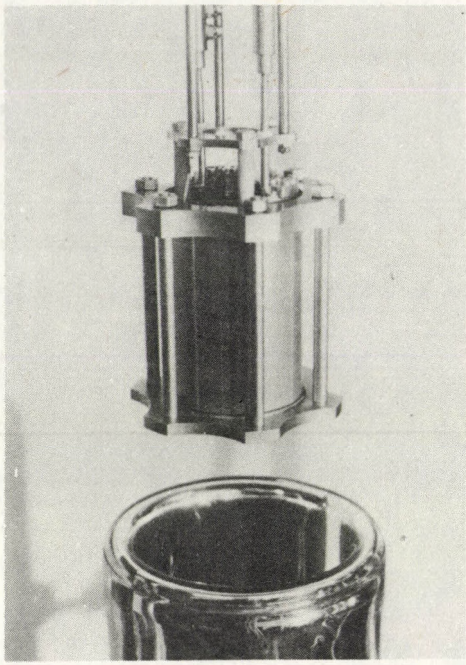


Fig. 2

3. *Helmholtz's coil*

The Helmholz-type coil pair (Fig. 3) was designed for the horizontal field direction. The coils were made of Nb—Ti—Zr superconducting material. The coils have a diameter of 46 mm with 840 turns. At the maximum current of 34 A the magnet produces a magnetic field intensity of 5.5 k0e with a homogeneity of 5×10^{-2}.

4. *Homogeneous solenoid*

The measuring solenoid (Fig. 4) was built for different laboratory aims. It was wound from a multifilament Nb—Ti conductor with 60 cores. The maximal current is 44.5 A at 4.2 K, which produced a magnetic field of 10.9 k0e with better than 2×10^{-5}/cm homogeneity in a 5 cm section. The overall length of the solenoid is 105 mm, the inner bore is 25 mm and the outer diameter of the magnet is 33 mm. The coil consists of 1927 turns in seven layers. In the case of our fourth magnet the high homogeneity was achieved the overwinding of the solenoid ends. The correction coils have 147 turns in seven layers and are 7 mm long. The appropriate configuration was designed by the minimalisation of the volume by computer programme [4].

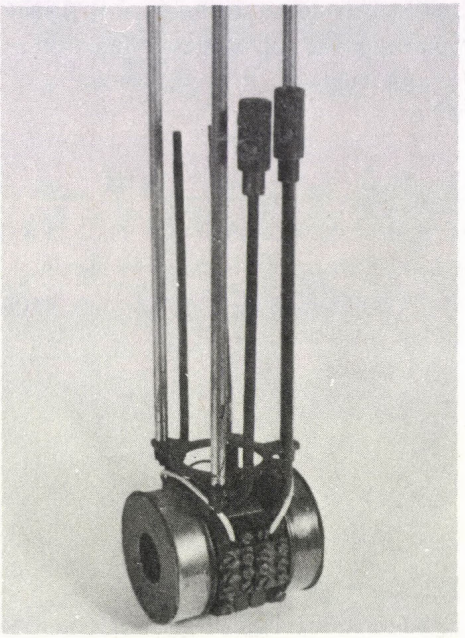


Fig. 3

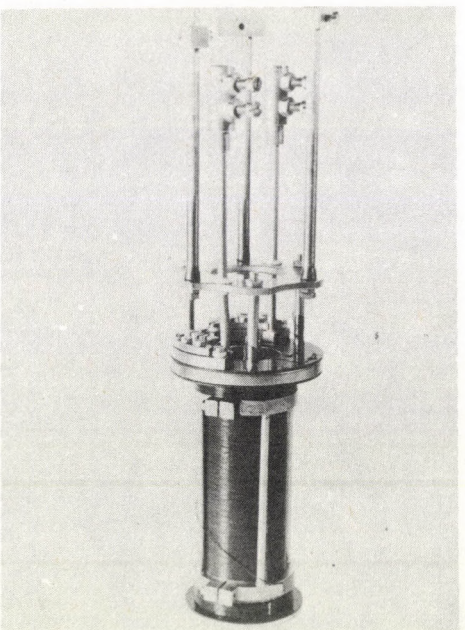


Fig. 4

The result optimalised by computer is presented in Fig. 5, which shows the homogeneity along the axis of the solenoid as a function of the relative correction windings. b, N and N_c are the length of solenoid, the number of turns and the number of correction turns, respectively.

5. *Low field solenoid*

For the investigation of the properties of small samples a low field solenoid (Fig. 6) has been built with very small volume. The maximal magnetic field intensity was 1.5 k0e with a homogeneity of 5×10^{-3}/cm. This solenoid has been wound from superconducting multifilament Nb—Ti wire.

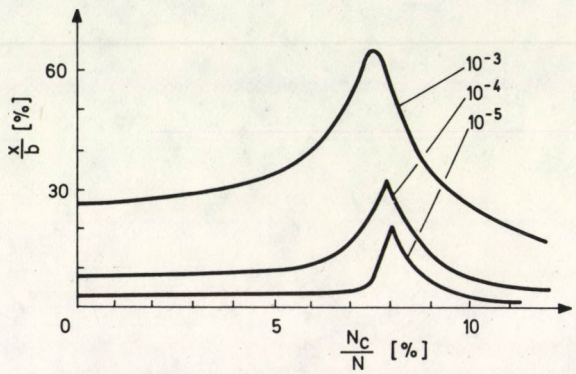


Fig. 5

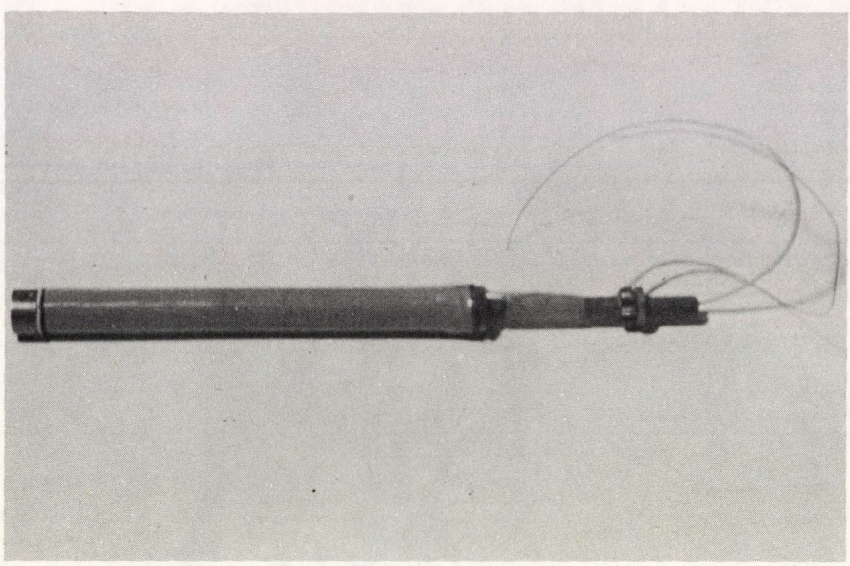


Fig. 6

6. *Cryomagnet*

A model magnet (Fig. 7) with two continuous winding correction coils has been built for different cryogenic experiments, educational purposes and demonstration of the values of calculating magnetic parameters by computer. The magnet produces

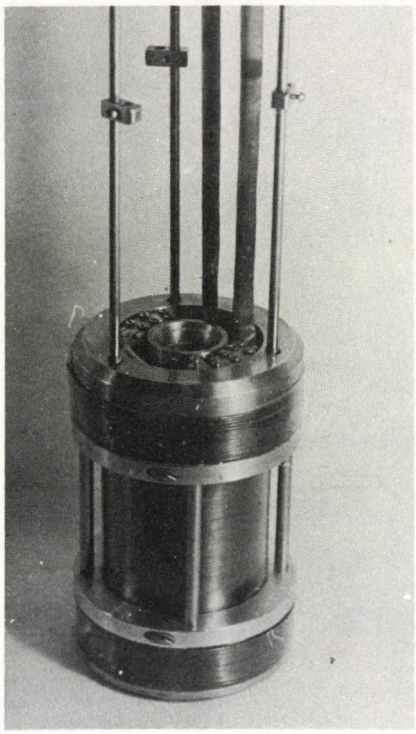


Fig. 7

25.4 k0e at a supply current of 100 A with a homogeneity of 10^{-5}/cm. This solenoid is operated with liquid nitrogen, at room temperatures. Its material is Cu wires, its outer diameter is 108 mm, the inner bore is 28 mm and its overall length is 172 mm.

7. *Superhomogeneous solenoid system*

Certain chemical reactions in a highly homogeneous magnetic field show interesting orientation properties. The homogeneity required along one direction in a rather large volume is about 10^{-5}/cm. For this reason a solenoid type superconducting magnet was built with a rigorous homogeneous field along its axis. The minimaiization of the volume at the appropriate parameters of solenoid was performed by using a graphical interpolating technique [5].

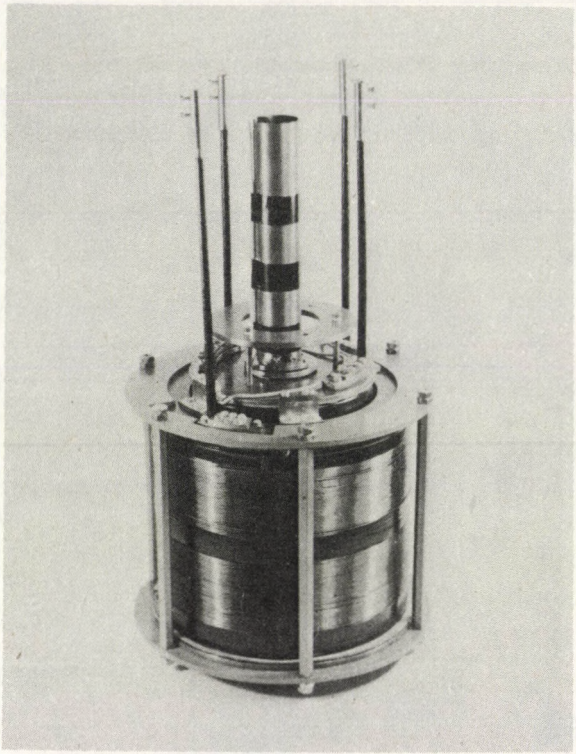


Fig. 8

Table II

Main data of solenoid system

	A	B	C
Number of layers	25	38	5
Number of windings	3650	6688	540
Outer diameter (mm)	180	120	42
Length of solenoid (mm)	136	146	129

The solenoid (Fig. 8) was made in a three-layer form: the outer solenoid (A), the inner solenoid (B) and the symmetrical correction coils (C). The outer solenoid and the upper part of correction coil is removable. The inner symmetrical layer in an adequate form served as the correction coils to produce the homogeneity required. The appropriate profiles were calculated by computer. In our case for a six-order solenoid [6] [7] construction with 50 × 80 mm working chamber the magnetic field which can be reached is plotted against the calculation parameters in Fig. 9. This Figure shows the

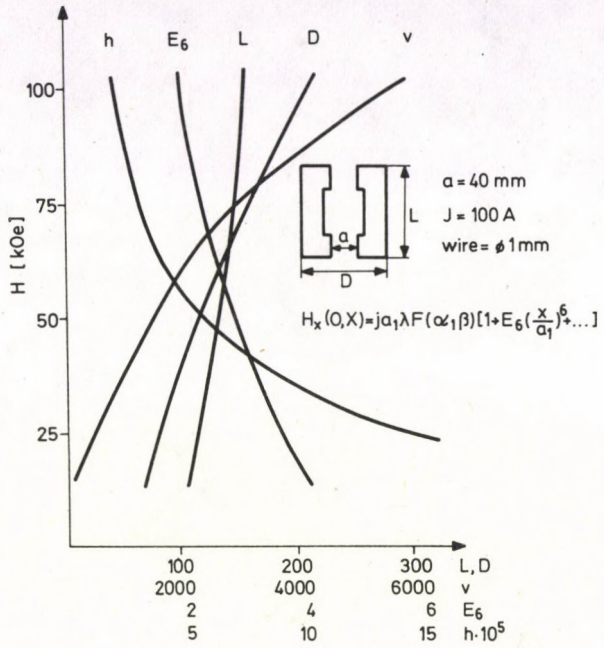


Fig. 9

magnetic field as a function of the calculation parameters: the current intensity j, the relative active volume-coefficient λ, the inner radius a_1, the geometrical factor $F(\alpha, \beta)$ and the characteristic factor of the degree of homogeneity E_6.

The main data of the highly homogeneous solenoid (Fig. 10) system are shown in Table II. The wire of the coils is a multifilamentary Nb—Ti alloy.

The maximum magnetic field strength is 72.6 k0e at a transport current of 100 A, and without the outer part it is 43.8 k0e. The homogeneity in the middle working chamber of 40 mm diameter is better than 10^{-5}/cm [8] [9].

The magnetic field of our magnets was measured by Hall probe calibrated by the proton resonance method.

References

1. J. Bánkuti, I. Kirschner, L. Lászlóffy and J. Lukács, Proc. 6th Magnet Techn. Conf., p. 390., Bratislava, 1977.
2. G. Fricsovszky, I. Kirschner, I. Kovács and E. Papp, Proc. 3rd Intern. Cryogenic Eng. Conf., p. 362., West-Berlin, 1970.
3. I. Kirschner, I. Kovács and L. Lászlóffy, Proc. 4th Intern. Cryogenic Eng. Conf., p. 372., Eindhoven, 1972.
4. I. Kirschner, T. Porjesz and G. Reményi., Cryogenics, **17,** 565, 1977.
5. D. Atherton, Canad. J. Phys., **47,** 753, 1969.
6. M. W. Garrett, J. Appl. Phys., **22,** 1091, 1951.
7. D. B. Montgomery, IEEE Spectrum, **1,** 103, 1964.
8. J. Bánkuti, I. Kirschner and L. Lászlóffy, Abstr. 7th Intern. Cryogenic Eng. Conf., p. 87., London, 1978.
9. J. Bánkuti, I. Kirschner and Gy. Reményi, IEEE Trans. Magn. MAG-**17,** 1999, 1981.

Acta Physica Academiae Scientiarum Hungaricae, 53(3—4), pp. 455—468 (1982)

GRAIN BOUNDARY MOTION IN VANADIUM

A. JUHÁSZ, I. FŐZY, P. TASNÁDI and I. KOVÁCS

INSTITUTE FOR GENERAL PHYSICS, ROLAND EÖTVÖS UNIVERSITY
BUDAPEST, HUNGARY

and

I. VITÁNYI

ALUTERV-FKI, BUDAPEST, HUNGARY

Grain boundary motion has been investigated "in situ" in 99,9% purity V sheets in a microscope. The individual grain boundary motion has been followed by the grooves etched thermally on the surface of the sample. The velocity of individual grain boundaries has been measured and compared to the average velocity determined from the normal grain growth law. An analysis of the relation between the velocity of curved boundaries and their curvature is given. The results obtained prove that in the applied experimental circumstances the driving force of the grain boundary motion is the surface energy of the boundaries. Finally, a simple geometrical model is given for explaining the spasmodic migration of the boundaries joined to the surface, as well as the observed sudden breaking of certain grains.

1. Introduction

The grain growth during annealing of polycrystalline metals takes place by high angle grain boundary motion. The rate of boundary motion, v is given by

$$v = m(p - p^*), \tag{1}$$

where p is the driving force, p^* is the dragging force, m is a parameter which depends on the grain boundary structure and the atomic mechanism of the boundary motion, including therefore the grain boundary mobility and the grain boundary energy [1].

The driving force rises from the decrease of the free enthalpy of the sample which takes place by the decrease of the amount of the grain boundary surfaces. The driving force can be characterized therefore by the ratio of the average grain surface and the average grain volume, that is:

$$p = \frac{\frac{D^2\pi}{2}\gamma}{\frac{1}{6}D^3\pi} = \frac{3\gamma}{D}, \tag{2}$$

where D is the average grain diameter and γ is the specific surface energy [1]. If the driving force satisfies Eq. (2), then the growth of average grain diameter obeys a

parabolic law

$$D - D_0 = a\sqrt{t}, \tag{3}$$

where D_0 is the initial grain size, t the time and a is a temperature dependent constant [2].

According to direct observations the curved boundaries migrate in general toward their curvature centre, so they tend to straighten, and during straightening the surface of the boundaries decreases. By the analogy of soap films the velocity of a curved boundary can be taken as proportional to the forces due to its surface tension:

$$v \sim \gamma \left(\frac{1}{R_1} + \frac{1}{R_2} \right), \tag{4}$$

where R_1 and R_2 are the principal radii of curvature [3]. This velocity depends also on hindering effects [4]. In sheet samples the most important hindering effect arises from the thermal groove system developed along the lines of the intersection of the grain boundaries and the free surface [1]. In equilibrium the shape of the grooves at the free surface is such that the resulting surface tension vector is zero (Fig. 1). The formation of grooves and their hindering effect on boundary migration were discussed by Mullins [5]. He found theoretically that the hindering effect depends on the surface energies and on the angle between the free surface normal and that of the wall of the groove. It was also shown that a boundary will become stuck at the surface if this angle is less than a critical value, Θ_c.

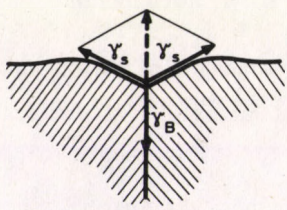


Fig. 1

It is well known that the specimen thickness has a characteristic effect on grain boundary motion. If the grain size is much less than the specimen thickness, then the grooves influence the grain growth only in a narrow layer at the specimen surface [5]. In the case of thin sheets the grains extend through the sheet, so the grooves can influence the grain structure of the whole sample therefore retarding considerably the grain growth process [6]. During "in situ" observations spasmodic grain boundary migration can be seen, which is a consequence of the grooves. Namely, the groove developed along a grain boundary can fix its position at the surface for a time, after which the boundary jumps out of its old groove and moves a certain distance along the surface at high speed. Because of the decrease of the driving force the boundary motion

will be slowed down and will be fixed again by the development of a new groove, and so on. So the moving boundary can leave a row of observable grooves behind it because their healing time is much longer than their formation time.

2. Experimental

The investigations were made on 99,9% purity V which was produced by ALUTERV FKI, Budapest [7]. The metal contained the following main impurities (concentrations in at%):

$$6 \cdot 10^{-2}\% \, O_2, \; 2.9 \cdot 10^{-2}\% \, N, \; 5 \cdot 10^{-3}\% \, H, \; 6.6 \cdot 10^{-3}\% \, Ce.$$

Ingots were reduced 70% thickness by cold rolling to sheets of 1.5 mm thickness and subjected to an annealing treatment in $1.33 \cdot 10^{-4}$ Pa vacuum at 900 °C for 4 h. After this treatment the surface of the samples was polished electrolytically. The "in situ" grain boundary investigations were made in a UNION HM-4 type high temperature microscope in $8 \cdot 10^{-3}$ Pa vacuum at 1200—1350 °C.

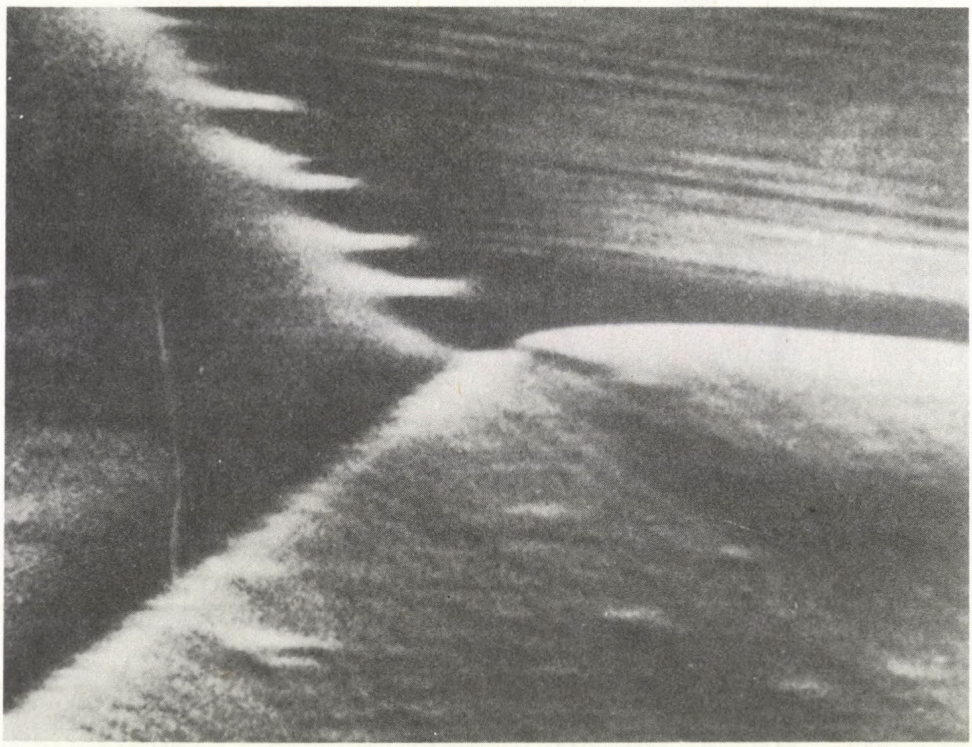


Fig. 2

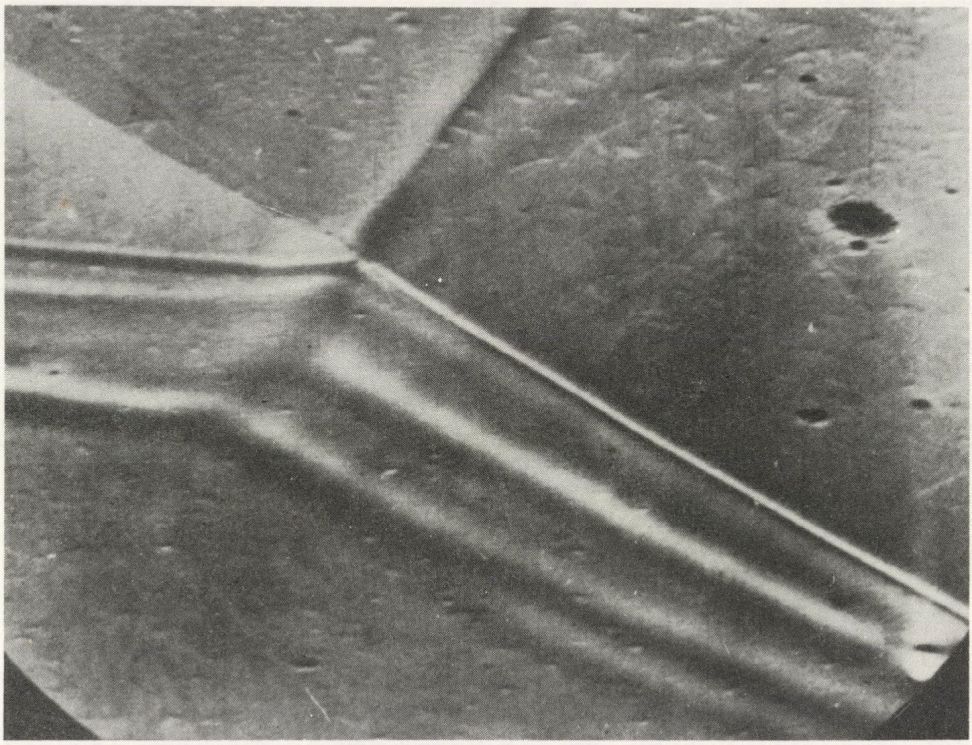


Fig. 3

3. Results

10—12 minutes after heating up the samples groove formation occurred along the lines where grain boundaries intersected the free surface. The grain boundary motion could be observed by the spasmodic appearence of the grooves. To investigate the changes in the grain structure and the process of individual grain boundary motion, different series of photographs and also movie films were taken. Besides the "in situ" observations, scanning electronmicrographs of the grooves were taken, too. Fig. 2 shows the scanning electron micrograph of a groove. In Fig. 3 a marked and a few less distinct lines of a boundary can be seen which correspond to the last and previous positions of the grooves, respectively.

3.1. *Motion of individual boundaries*

a) *Straight boundaries.* Individual grain boundary velocity can be studied "in situ" by means of the spasmodic appearance of the thermal grooves. Fig. 4 shows one from a series of pictures taken from a contracting grain. The internal groove shows the

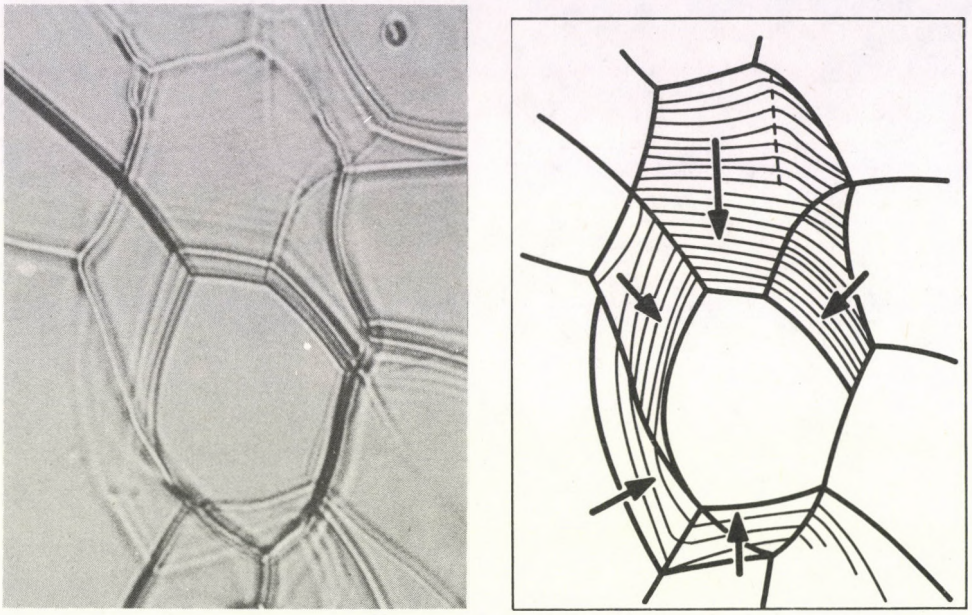


Fig. 4

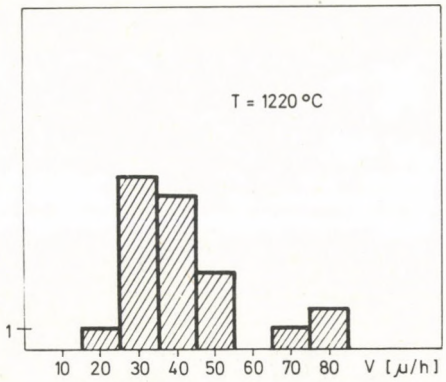


Fig. 5

momentary position of the boundaries, the others mark previous positions. Measuring the distance between the groove positions, and the time between the appearance of two grooves, one can calculate the average velocity of the moving boundary. Average grain boundary velocities obtained in this way are shown by the histogram in Fig. 5.

b) Curved boundaries. Figs 6 and 7a show examples of two different series of pictures taken during the migration of a curved boundary. It can be seen on both pictures that the boundary moves toward the inner part of the grain and that the larger

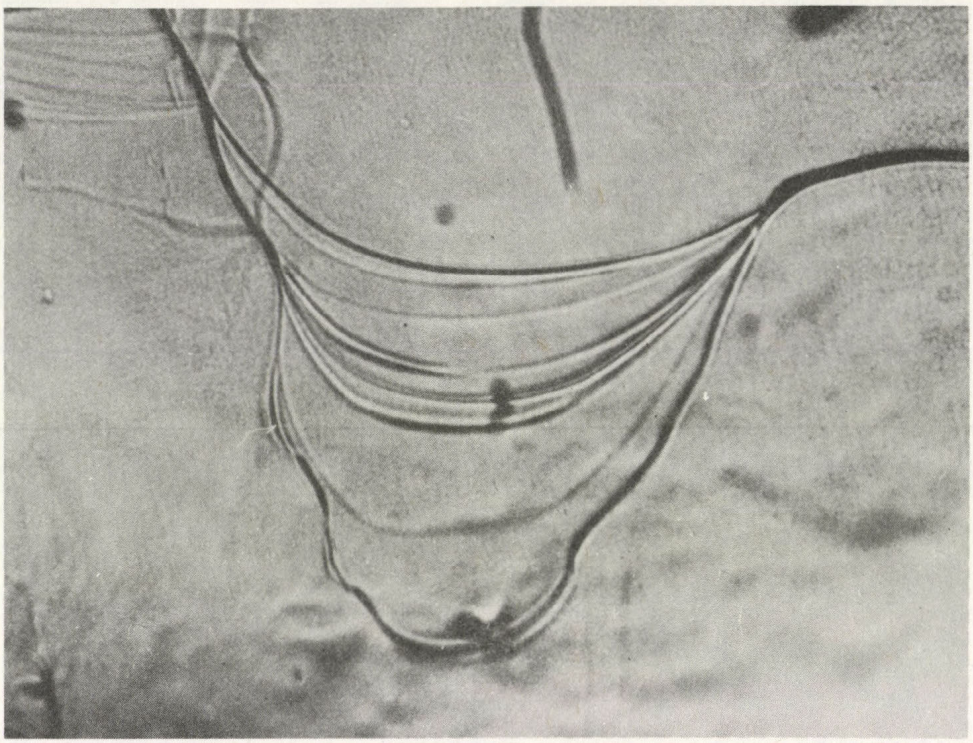
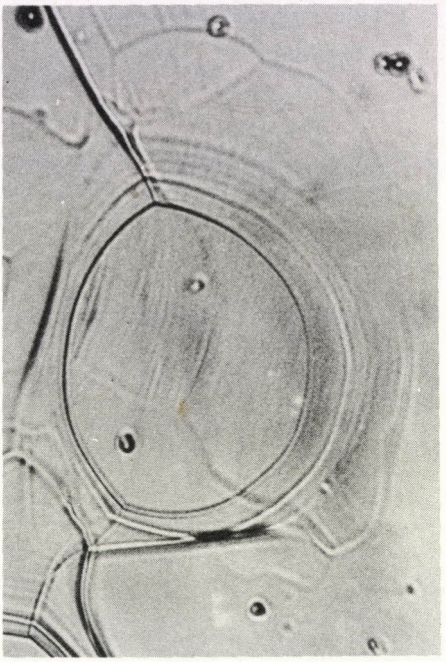


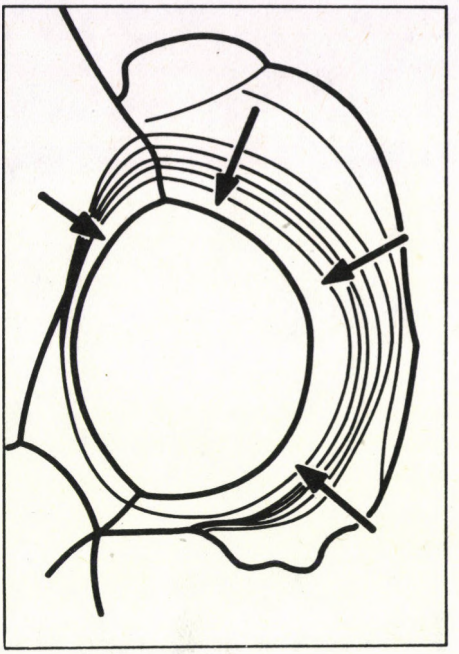
Fig. 6

the curvature, the bigger the distance between the neighbouring positions of the grooves. According to Fig. 6 the boundaries are straightening during the motion. In contrast to this, Fig. 7a shows a case where the boundary curvature increases continuously, which must be a conseqence of the motion of the other joining boundaries. It was observed in this case that after reaching a minimum size this grain rapidly disappeared.

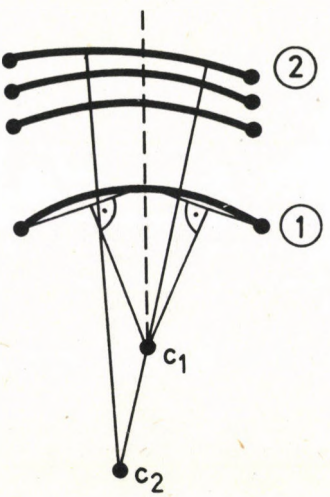
On the basis of a series of pictures it was possible to determine the relation between the curvature and the grain boundary velocity. The approximate radius of the investigated boundary was determined by a simple geometrical method as follows. The picked part of the boundary was regarded as an arc of a circle, and secants were drawn through the end points and the middle point of the arc. The point of intersection of the normal bisectors of the two secants can be considered as the centre of curvature (Fig. 7c). Applying this method the rate of motion of the same grain boundary was determined at three different radii of initial curvature shown by the arrows in Fig. 7b. The average velocities obtained are plotted as a function of curvature in Fig. 8.



a)



b)



c)

Fig. 7

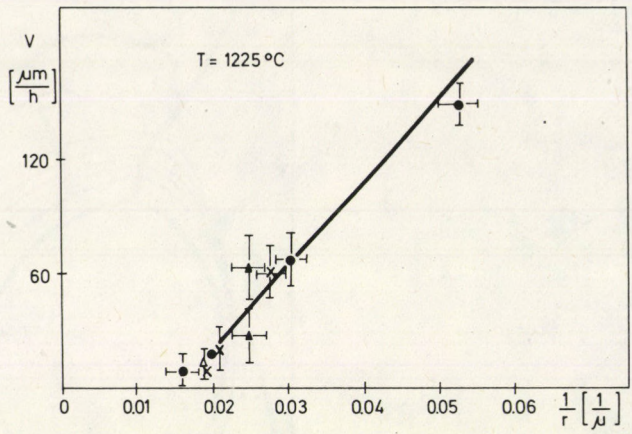


Fig. 8

3.2. *Breakdown of grains*

A series of micrographs (Fig. 9) shows a characteristic process of structural changes observed many times in the high temperature microscope. The grain shown on the photos becomes gradually smaller with time. At the beginning its boundaries moved relatively slowly, while after reaching a critical size the process accelerated so much that the grain, almost in a moment, had disappeared and a triple boundary junction appeared instead.

3.3. *Variation of the mean grain size*

Besides the motion of individual boundaries the variation of the mean grain size in time was measured, too. From time to time pictures were taken from three parts of the sample by a low magnification objective. The mean grain size vs the square root of time is plotted in Fig. 10. It can be seen that the grain coarsening can be described well by the normal grain growth law, in contrast to previous observation when a time dependence of $D \sim t^{1/3}$ was obtained [8].

4. Discussion and conclusions

The grain boundary migration has been extensively studied in the literature and it was found that the migration rate depends on many different parameters [9, 10, 11, 12]. Recently it became clear that the velocity of grain boundaries depends strongly not only on the temperature and impurities but on the orientation and structure of the boundary as well [12, 13].

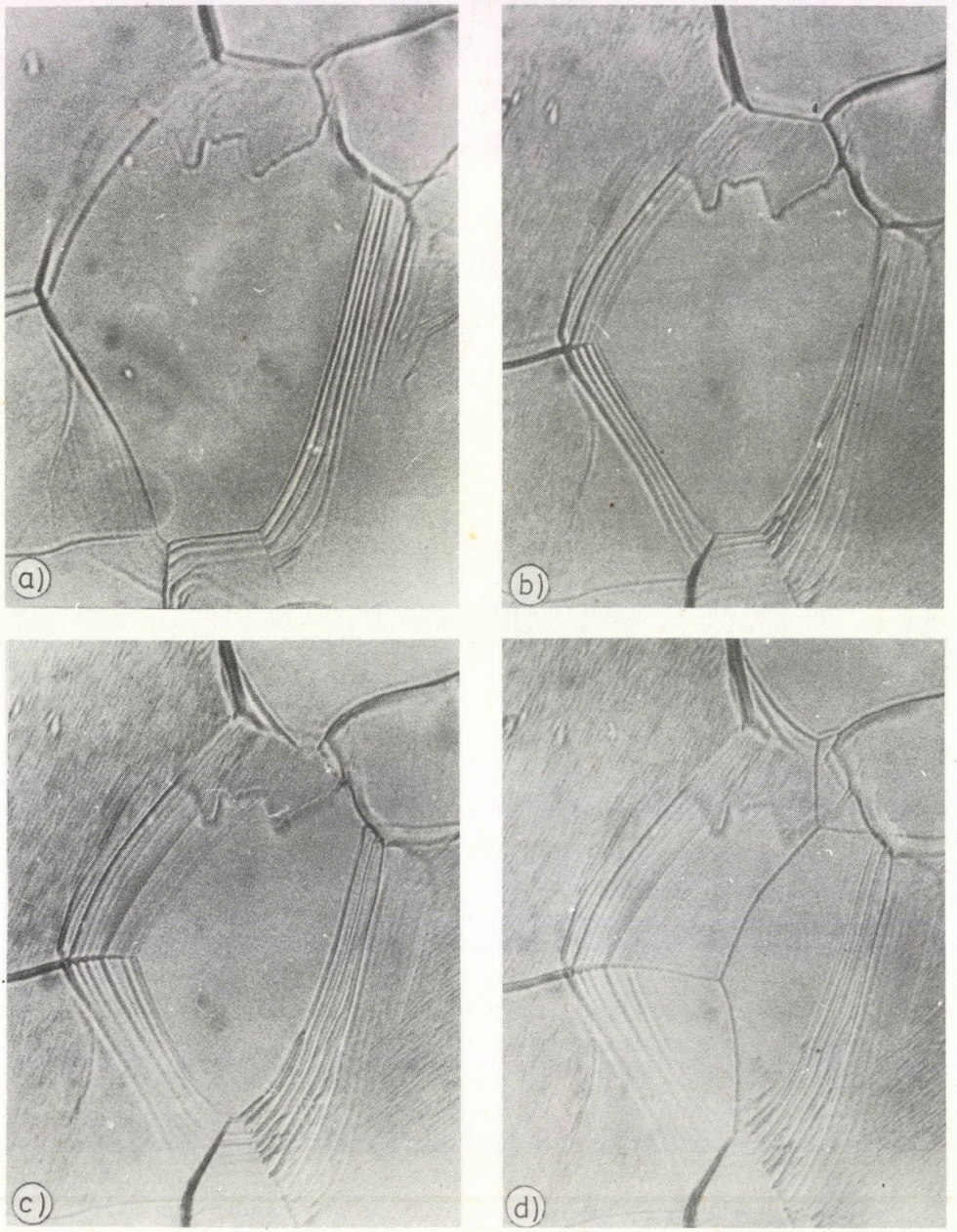


Fig. 9

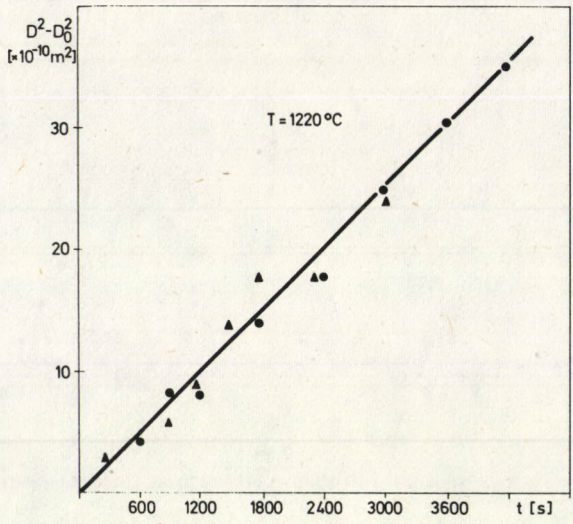


Fig. 10

It seems that the role of grain boundary structure in grain boundary migration is much less understood at present than the role of other parameters. The reason for this fact is on the one hand that the simultaneous experimental investigation of the grain boundary migration and the grain boundary structure is very difficult [11, 13], and on the other hand no universal model is known at present for describing the different grain boundary structures, although significant advances appeared during the last years in the theoretical understanding of the grain boundary structures [12, 14, 15, 16, 17].

Recently, a connection has been found between the geometric models of grain boundary structures and dislocation theory [16, 17]. By applying the concept of dislocation density tensor [18, 19, 20] and that of the surface dislocations [16, 21], powerful methods were developed to describe the elastic field of dislocation networks [17, 20]. The simultaneous development of electronmicroscopic investigations and the computer modelling of grain boundaries made it possible to prove experimentally the basic ideas of these theoretical models [20, 22, 23, 24]. However, most of the papers deal with the structure of static boundaries, and there are only a few articles in which the connection between grain boundary migration and structure is discussed [13, 25]. Nevertheless, it seems to be well supported experimentally that the migration rate depends on the boundary structure, or more exactly, the migration can be interpreted by the motion of grain boundary dislocations and grain boundary steps [11, 13].

With the help of the usual metallographic methods the change of the mean grain size, that is the average migration rate can be determined. The specific methods of high temperature metallography make the optical observation of the motion of individual boundaries possible, but it is rather difficult to relate the velocity data obtained in this way to the theoretical models of grain boundary structures. The orientation of a

boundary, the orientation misfit and the dislocation density in the boundary cannot in general be determined by optical methods.

Because of these difficulties we shall discuss our experimental results only on phenomenological basis.

It is supposed that the grain boundary motion is controlled by the decrease of the free enthalpy through the decrease of the amount of grain boundary surfaces. Although the processes in the whole crystal are governed by thermodynamical laws, because of local effect the motion of the individual boundaries does not always reduce locally the free enthalpy. Consequently, to describe the migration of an individual boundary one has to take into account the forces exerted at grain boundary junctions, at the free surface as well as the the forces originating from the boundary curvature.

4.1. *The driving force of boundary motion*

The theory of normal grain or subgrain growth is based on the fact that the driving force arises from the grain boundary interfacial free energy. Several authors [2, 3, 6, 26, 27] have derived a parabolic growth law for the average grain size, using slightly different arguments.

As it is to be seen in Fig. 10 the normal grain growth law is satisfied for polycrystalline vanadium, consequently the driving force of boundary motion is the decrease of the grain boundary free energy. On the basis of the normal grain growth law the average velocity of grain boundary motion can be determined by using the following approximation:

$$\bar{v} = \frac{dD}{dt} = \frac{1}{2D} \frac{D^2 - D_0^2}{t} \approx \frac{1}{2D_0} \frac{D^2 - D_0^2}{t}. \tag{5}$$

With the data of the plot in Fig. 10 a mean velocity of 21 μm/h was obtained. This value agrees well in order of magnitude with the ones obtained for the velocities of the individual boundary motion, although the latter ones are greater. However, it is easy to understand this deviation, because not every grain boundary moves at the same time during the annealing applied for the mean grain size measurements.

It can be seen from Fig. 8 that in the case of curved boundaries their velocity is to a good approximation inversely proportional to the radius of curvature. This proves once again that the driving force of the motion is the decrease of the surface energy. If the curvature is small, the velocity of the grain boundary motion can be taken as constant which can be considered the velocity of straight boundaries.

4.2. *The spasmodic migration of boundaries at the free surface*

As we have seen in Section 3.1. the motion of boundaries intersecting the free surface of the sample can be observed by using discrete series of grooves thermally etched along the line of the boundary. In the following a simple geometric model is

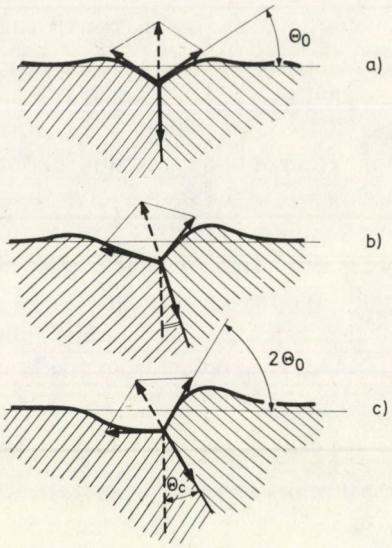


Fig. 11

given based on some results of Mullins [5] and Shewmon [6] to interpret qualitatively the jerky motion of thermal grooves.

The basic principles of the model are: 1) The appearance of the grooves in discrete lines is a typical surface effect, but inside the matter the grain boundary itself moves in a continuous way [5]. The grain boundary is considered as an ideal geometric surface and we disregard the structure of the boundary and the atomic process causing the migration. 2) The boundary moving continuously inside the sample is stuck at the surface by the groove formed by rapid surface diffusion. The groove root remains in the same position so far as the magnitude of the angle between the normal of the walls of the groove and the surface normal is smaller than a critical angle, Θ_c.

The existence of the critical angle can be interpreted in a simple manner by the equilibrium condition valid at the intersection of the boundary and the free surface. Figs 11a—c represent a perpendicular segment of the groove when the boundary lies at a right angle to the surface, and in a general case, respectively.

In equilibrium the resultant of the surface tension vectors has to be zero. We can see that this condition leads to the turning of the groove if the boundary is not normal to the surface. Micrographs taken with interference microscope show that the grooves are asymmetrical with their steepest wall in the direction of the boundary migration [5].

On the basis of these assumptions it can be supposed that the grain boundary which migrates continuously inside the matter turns away the groove walls in such a way that the balance of surface tension vectors remains fulfilled. Naturally, this rigid like turn of the groove in respect to the free surface is the result of surface diffusion,

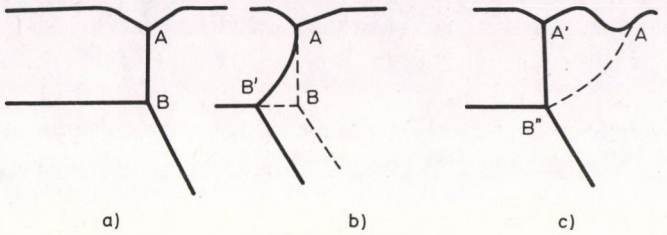


Fig. 12

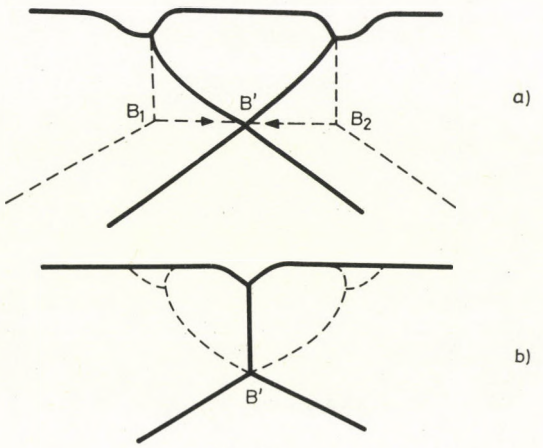


Fig. 13

which carries the matter from one part of the groove to the other. If one of the groove walls becomes parallel with the free surface, then after further movement of the joining boundary the surface tension vectors cannot be balanced, so the boundary breaks away from the groove and jumps in a new position.

The critical angle, Θ_c, defined in this way is demonstrated in Fig. 11c, which agrees with the one obtained by Mullins [5].

The grain boundary motion inside the matter can result, for example, from an unbalanced boundary junction inside the sample. So the migration of a boundary terminated at the surface can be interpreted as follows. Fig. 12 represents a boundary subtending a right angle with the surface at point A. This boundary ends in a triple point (point B). Suppose that at point B the surface tension vectors are not in equilibrium. For simplicity let us take one of the joining boundaries parallel to the surface, and suppose that the right hand side grain is large. Since the junction at point B is unbalanced, the boundary AB moves left. However, at point A the boundary is fixed, so it becomes curved. The curved boundary enhances its further motion to the left direction. As a result of this process at point A the angle between the boundary and the normal of the outside surface reaches the critical angle, Θ_c. At this moment the moving boundary escapes from the point A and jumps to the new position in which it is again

normal to the surface (Fig. 12). In this new position the length of the AB boundary is the shortest, but the triple junction at point B is unbalanced again, so the process will be repeated.

By this simple model we can explain the sudden collapse of certain grains, too. It is possible that the migration of the boundaries of a small grain which intersect the free surface can lead to the formation of an unbalanced quadruple point junction (Fig. 13). From this position a stable triple point junction can be formed by the disappearance of the B_1B' and B_2B' boundary segments, which removes quickly the small grain and a triple point boundary junction appears at the free surface.

Acknowledgement

The authors are grateful to the Diffractional Department of the Research Institute for Technical Physics, Budapest, for taking the scanning electron micrographs.

References

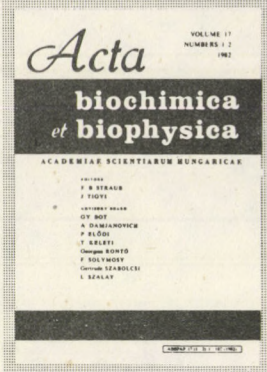
1. H. P. Stüwe, Recrystallization of Metallic Materials, Ed. F. Haessner, Rieder-Verlag Gmbh., Stuttgart, 1978.
2. P. Feltham, Acta Met., **5,** 97, 1957.
3. M. Hillert, Acta Met., **13,** 227, 1965.
4. K. Lücke and H. P. Stüwe, Recovery and Recrystallization of Metals, Ed. L. Himmel, John Wiley and Sons, New York, 1963, p. 171.
5. W. W. Mullins, Acta Met., **6,** 414, 1958.
6. P. G. Shewmon, Diffusion in Solids, McGraw-Hill, New York, 1963.
7. P. Vitányi, Gy. Várhegyi, M. Miskei and L. Iván, Conf. of Rare Met., Ajka, 1974.
8. A. Deschanvres, G. Nouet and J. Vicens, Mater Sci. Engng., **19,** 43, 1975.
9. H. Gleiter and B. Chalmers, Progress in Material Sci., Pergamon Press, Oxford, 1972.
10. C. J. Simpson, W. C. Winegard and K. T. Aust, Grain Boundary Structure and Properties, Eds. G. A. Chadwick and D. A. Smith, Academic Press, London, 1976.
11. P. Gordon and R. A. Vandermeer, Recrystallization, Grain Growth and Texture, Ed. H. Margolin, ASM Cleveland, 1966.
12. K. T. Aust and B. Chalmers, Met. Trans., **1,** 1095, 1970.
13. C. M. F. Rae and D. A. Smith, Phil. Mag., *A***41,** 477, 1980.
14. W. Bollmann, Crystal Defects and Crystalline Interfaces, Springer-Verlag, Berlin, 1970.
15. K. Sadananda and M. J. Marcinkowski, Canonical Theory of Grain Boundaries I—II., E. M. G. Report No. CXXII 1973.
16. M. J. Marcinkowski, Unified Theory of the Mechanical Behavior of Matter, John Wiley and Sons, New York, 1979.
17. G. Saada, Acta Met., **27,** 921, 1979.
18. E. Kröner, Kontinuumstheorie der Versetzungen und Eigenspannungen, Springer-Verlag, Berlin, 1958.
19. B. A. Bilby, R. Bullough and E. Smith, Proc. R. Soc., *A***231,** 263, 1955.
20. C. Rey and G. Saada, Phil. Mag., **33,** 825, 1976.
21. K. Jagannadham and M. J. Marcinkowski, Crystal Lattice Def., **8,** 81, 1979.
22. R. C. Pond, D. A. Smith and V. Vitek, Acta Met., **27,** 235, 1979.
23. L. M. Clareborough and C. T. Forwood, Phys. Stat. Solidi (a), **58,** 597; **59,** 263; **60,** 51, 1980.
24. V. Vitek, D. A. Smith, and R. C. Pond, Phil. Mag., *A***41,** 649, 1980.
25. M. S. Masteller and C. L. Bauer, Phil. Mag., *A***38,** 697, 1978.
26. G. F. Bolling and W. C. Winegard, Acta Met., **6,** 283, 1958.
27. R. Sandström, Acta Met., **25,** 905, 1977.

Manuscript received by Akadémiai Kiadó: 28 July 1982
Manuscript received by the Printers: 17 August 1982
Date of publication: 30 June 1983

PRINTED IN HUNGARY
Akadémiai Nyomda, Budapest

Acta Biochimica et Biophysica

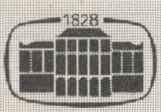
ACADEMIAE SCIENTIARUM HUNGARICAE



EDITORS
F. B. STRAUB and J. TIGYI

Acta Biochimica et Biophysica publishes original papers on biochemistry and biophysics in English. Papers on proteins (structure and synthesis), enzymes, nucleic acids, regulatory and transport processes, bioenergetics, excitation, muscular contraction, functional structure and ultrastructure will be accepted.

Published in one volume (4 issues) annually
Subscription rate per volume $ 40.00



AKADÉMIAI KIADÓ
Publishing House of the Hungarian Academy of Sciences
Budapest

Distributed by KULTURA
Hungarian Foreign Trading Co.
PO Box 149, H-1389 Budapest

CONTENTS

VOLUME 17, NUMBERS 3–4

Bot, Gy., Kovács, E., Tóth, B., Dombrádi, V., Gergely, P.: Role of fructose-1-phosphate in the regulation of the dephosphorylation of glycogen phosphorylase a

Tóth, B., Gergely, P.: Role of pyridoxal 5'-phosphate in the regulation of the interconversion of phosphorylase a into phosphorylase b

Vereb, Gy., Szőcs, E., Bot, Gy.: Immunological and allosteric identity of heart-specific glycogen phosphorylase isoenzymes from various mammals

Édes, I., Dózsa, E., Sohár, I., Guba, F.: Effect of plaster cast immobilization on the turnover rates of soluble proteins and lactate dehydrogenase isoenzymes of rabbit m. soleus

Csillag, A., Kálmán, M., Csató, Zs.: Adenosin-5'-diphosphate penetration into synaptosomes isolated from rat celebral cortex

Misik, S., Masszi, G.: Microwave method for determining dielectric parameters of living biological objects II. Study of ionic water binding

Koszorús, L., Masszi, G.: Investigation of hydration of macromolecules II. Study of ethylene-glycol and 1-4-dioxane solutions by dielectric method

Vető, F.: Is the difference of vapour pressure the driving force of real osmotic water transport?

Book Reviews

NOTES TO CONTRIBUTORS

I. PAPERS will be considered for publication in Acta Physica Hungarica only if they have not previously been published or submitted for publication elsewhere. They may be written in English, French, German or Russian.

Papers should be submitted to

Prof. I. Kovács, Editor
Department of Atomic Physics, Technical University
1521 Budapest, Budafoki út 8, Hungary

Papers may be either articles with abstracts or short communications. Both should be as concise as possible, articles in general not exceeding 25 typed pages, short communications 8 typed pages.

II. MANUSCRIPTS

1. Papers should be submitted in three copies.

2. The text of papers must be of high stylistic standard, requiring minor corrections only.

3. Manuscripts should be typed in double spacing on good quality paper, with generous margins.

4. The name of the author(s) and of the institutes where the work was carried out should appear on the first page of the manuscript.

5. Particular care should be taken with mathematical expressions. The following should be clearly distinguished, e.g. by underlining in different colours: special founts (italics, script, bold type, Greek, Gothic, etc.); capital and small letters; subscripts and superscripts, e.g. x^2, x_2; small *l* and 1; zero and capital O; in expressions written by hand: *e* and *l*, *n* and *u*, *v* and *v*, etc.

6. References should be numbered serially and listed at the end of the paper in the following form: J. Ise and W. D. Fretter, Phys. Rev., 76, 933, 1949.

For books, please give the initials and family name of the author(s), title, name of publisher, place and year of publication, e.g.: J. C. Slater, Quantum Theory of Atomic Structures, I. McGraw-Hill Book Company, Inc., New York, 1960.

References should be given in the text in the following forms: Heisenberg [5] or [5].

7. Captions to illustrations should be listed on a separate sheet, not inserted in the text.

8. In papers submitted to Acta Physica all measures should be expressed in SI units.

III. ILLUSTRATIONS AND TABLES

1. Each paper should be accompanied by three sets of illustrations, one of which must be ready for the blockmaker. The other sets attached to the copies of the manuscript may be rough drawings in pencil or photocopies.

2. Illustrations must not be inserted in the text.

3. All illustrations should be identified in blue pencil by the author's name, abbreviated title of the paper and figure number.

4. Tables should be typed on separate pages and have captions describing their content. Clear wording of column heads is advisable. Tables should be numbered in Roman numerals (I, II, III, etc.).

IV. RETURN OF MATERIAL

Owing to high postage costs, the Editorial Office cannot undertake to return *all* material not accepted for any reason for publication. Of papers to be revised (for not being in conformity with the above Notes or other reasons) only *one* copy will be returned. Material rejected for lack of space or on account of the Referees' opinion will not be returned to authors outside Europe.

Periodicals of the Hungarian Academy of Sciences are obtainable at the following addresses:

AUSTRALIA
C.B.D. LIBRARY AND SUBSCRIPTION SERVICE
Box 4886, G.P.O., *Sydney N.S.W. 2001*
COSMOS BOOKSHOP, 145 Ackland Street
St. Kilda (Melbourne), Victoria 3182

AUSTRIA
GLOBUS, Höchstädtplatz 3, *1206 Wien XX*

BELGIUM
OFFICE INTERNATIONAL DE LIBRAIRIE
30 Avenue Marnix, *1050 Bruxelles*
LIBRAIRIE DU MONDE ENTIER
162 rue du Midi, *1000 Bruxelles*

BULGARIA
HEMUS, Bulvar Ruszki 6, *Sofia*

CANADA
PANNONIA BOOKS, P.O. Box 1017
Postal Station "B", *Toronto, Ontario M5T 2T8*

CHINA
CNPICOR, Periodical Department, P.O. Box 50
Peking

CZECHOSLOVAKIA
MAD'ARSKÁ KULTURA, Národní třida 22
115 66 Praha
PNS DOVOZ TISKU, Vinohradská 46, *Praha 2*
PNS DOVOZ TLAČE, *Bratislava 2*

DENMARK
EJNAR MUNKSGAARD, Norregade 6
1165 Copenhagen K

FEDERAL REPUBLIC OF GERMANY
KUNST UND WISSEN ERICH BIEBER
Postfach 46, *7000 Stuttgart 1*

FINLAND
AKATEEMINEN KIRJAKAUPPA, P.O. Box 128
SF-00101 Helsinki 10

FRANCE
DAWSON-FRANCE S. A., B. P. 40, *91121 Palaiseau*
EUROPÉRIODIQUES S. A., 31 Avenue de Versailles, *78170 La Celle St. Cloud*
OFFICE INTERNATIONAL DE DOCUMENTATION ET LIBRAIRIE, 48 rue Gay-Lussac
75240 Paris Cedex 05

GERMAN DEMOCRATIC REPUBLIC
HAUS DER UNGARISCHEN KULTUR
Karl Liebknecht-Straße 9, *DDR-102 Berlin*
DEUTSCHE POST ZEITUNGSVERTRIEBSAMT
Sraße der Pariser Kommüne 3–4, *DDR-104 Berlin*

GREAT BRITAIN
BLACKWELL'S PERIODICALS DIVISION
Hythe Bridge Street, *Oxford OX1 2ET*
BUMPUS, HALDANE AND MAXWELL LTD.
Cowper Works, *Olney, Bucks MK46 4BN*
COLLET'S HOLDINGS LTD., Denington Estate
Wellingborough, Northants NN8 2QT
WM. DAWSON AND SONS LTD., Cannon House
Folkstone, Kent CT19 5EE
H. K. LEWIS AND CO., 136 Gower Street
London WC1E 6BS

GREECE
KOSTARAKIS BROTHERS INTERNATIONAL BOOKSELLERS, 2 Hippokratous Street, *Athens-143*

HOLLAND
MEULENHOFF-BRUNA B.V., Beulingstraat 2,
Amsterdam
MARTINUS NIJHOFF B.V.
Lange Voorhout 9–11, *Den Haag*
SWETS SUBSCRIPTION SERVICE
347b Heereweg, *Lisse*

INDIA
ALLIED PUBLISHING PRIVATE LTD 13/14
Asaf Ali Road, *New Delhi 110001*
150 B-6 Mount Road, *Madras 600002*
INTERNATIONAL BOOK HOUSE PVT LTD
Madame Cama Road, *Bombay 400039*
THE STATE TRADING CORPORATION OF INDIA LTD., Books Import Division, Chandralok 36 Janpath, *New Delhi 110001*

ITALY
INTERSCIENTIA, Via Mazzè 28, *10149 Torino*
LIBRERIA COMMISSIONARIA SANSONI, Via Lamarmora 45, *50121 Firenze*
SANTO VANASIA, Via M. Macchi 58
20124 Milano
D. E. A., Via Lima 28, *00198 Roma*

JAPAN
KINOKUNIYA BOOK-STORE CO. LTD.
17-7 Shinjuku 3 chome, Shinjuku-ku, *Tokyo 160-91*
MARUZEN COMPANY LTD., Book Department, P.O. Box 5050 Tokyo International, *Tokyo 100-31*
NAUKA LTD. IMPORT DEPARTMENT
2-30-19 Minami Ikebukuro, Toshima-ku, *Tokyo 171*

KOREA
CHULPANMUL, *Phenjan*

NORWAY
TANUM-TIDSKRIFT-SENTRALEN A.S., Karl Johansgatan 41–43, *1000 Oslo*

POLAND
WĘGIERSKI INSTYTUT KULTURY, Marszałkowska 80, *00-517 Warszawa*
CKP I W, ul. Towarowa 28, *00-958 Warszawa*

ROUMANIA
D. E. P., *Bucureşti*
ILEXIM, Calea Grivitei 64–66, *Bucureşti*

SOVIET UNION
SOJUZPECHAT — IMPORT, *Moscow*
and the post offices in each town
MEZHDUNARODNAYA KNIGA, *Moscow G-200*

SPAIN
DIAZ DE SANTOS, Lagasca 95, *Madrid 6*

SWEDEN
ALMQVIST AND WIKSELL, Gamla Brogatan 26
101 20 Stockholm
GUMPERTS UNIVERSITETSBOKHANDEL AB
Box 346, *401 25 Göteborg 1*

SWITZERLAND
KARGER LIBRI AG, Petersgraben 31, *4011 Basel*

USA
EBSCO SUBSCRIPTION SERVICES
P.O. Box 1943, *Birmingham, Alabama 35201*
F. W. FAXON COMPANY, INC.
15 Southwest Park, *Westwood Mass. 02090*
THE MOORE-COTTRELL SUBSCRIPTION AGENCIES, *North Cohocton, N. Y. 14868*
READ-MORE PUBLICATIONS, INC.
140 Cedar Street, *New York, N. Y. 10006*
STECHERT-MACMILLAN, INC.
7250 Westfield Avenue, *Pennsauken N. J. 08110*

YUGOSLAVIA
JUGOSLOVENSKA KNJIGA, Terazije 27, *Beograd*
FORUM, Vojvode Mišića 1, *21000 Novi Sad*

HU-ISSN 0001-6705 **Index: 26.022**